Applied machine learning and the role of dynamic time-series data in optimising clinical management of acute febrile illnesses

A thesis submitted to:

Imperial College London for the degree of Doctor of Philosophy

Damien Keng Yen Ming
Department of Infectious Disease
2023

Statement of originality

I confirm that all work described within this thesis was carried out by myself unless otherwise stated, under the supervision of Professors Alison Holmes, Sophie Yacoub and Pantelis Georgiou.

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC).

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work.

This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law

Abstract

Background

Accounting for time-series dependencies inherent in patient healthcare data is important but optimal approaches are not well understood. This thesis investigates the role, and implementation of machine learning models which utilise dynamic data to improve clinical decision-making in acute febrile illnesses.

Methods

Mixed methods involving data science, qualitative and bioengineering techniques were used. Real-world exemplars including datasets of bacterial bloodstream infections (n=4,198), of dengue (n=4,131) and undifferentiated fever (n=8,100) were analysed using data and machine learning approaches. Clinical implementation of dengue models developed in the thesis was investigated through use of process mapping and thematic analysis. The issues of enhancing patient data acquisition to support future model development were examined through minimally-invasive biosensing techniques.

Results

Machine learning models which account for time-series dependencies provided better predictive ability compared with time-invariant versions in a range of acute febrile illnesses. Applying such models to healthcare settings required understanding of decision-making processes and patient pathways. Continuous biosensing methods could offer an effective and acceptable approach to enhancing the patient healthcare data collection across different settings.

Conclusion

The thesis demonstrates the importance of time-series data in supporting clinical decision-making in acute febrile illnesses. Use of novel machine learning methods to enhance existing and routinely collected healthcare information can improve care and is likely to be cost-effective. It is crucial that factors relating to healthcare implementation and data acquisition are addressed to ensure research findings result in clinical impact.

Acknowledgements

I am grateful for the guidance, advice and attention provided by my supervisors, Alison Holmes, Sophie Yacoub and Pantelis Georgiou throughout the duration of the work.

Many thanks to my colleagues at both Imperial College London and Oxford University Clinical Research Unit (OUCRU), Ho Chi Minh City Vietnam who have offered me much of their time and help during my research. These include (but are not limited to), Sorawat Sangkaew (Gunn), Bernard Hernandez, Richard Wilson, Tim Rawson, Stefan Karolcik, David Freeman, Paul Vasikasin, Ho Quang Chanh, Huynh Trung Trieu and many others. To Kerri Hill-Cawthorne and Rakhee Parmar whose organisational skills have been invaluable.

Acknowledgements to the patients and study participants at Imperial College Healthcare NHS Trust, OUCRU and NIHR-Imperial Clinical Research Facility. Thanks also to the Welcome Trust Flagship Project: Innovative biomedical engineering and computational science to improve the management of critical illness in resource-limited settings which has provided funding for the project.

I am grateful to Toju Sillo for her friendship and indebted to my wife, Tehmina for her absolute patience and unshakable love. To my daughter Annabel for keeping life and everything so very exciting during the PhD.

The thesis is dedicated to my family.

Table of contents

CHAPTER 1 - INTRODUCTION 17

1.1 AIMS OF THE THESIS 17
1.2 MEDICAL DECISION-MAKING IN ACUTE FEBRILE ILLNESSES 17
1.3 HEALTHCARE DIGITALISATION AND THE USE OF DYNAMIC PATIENT DATA 20
1.4 MACHINE LEARNING TO UNDERSTAND COMPLEX HEALTHCARE DATA 22
1.5 DENGUE AND BLOODSTREAM INFECTIONS AS IMPORTANT EXEMPLARS OF ACUTE FEBRILE ILLNESSES 23
1.6 ADDRESSING OPTIMAL USE OF HEALTHCARE DATA AND ACQUISITION 24
1.7 RESEARCH HYPOTHESIS AND THESIS AIMS 25
1.8 THESIS STRUCTURE AND CHAPTER OUTLINES 27

CHAPTER 2 - METHODS 2

2.1 STATISTICAL METHODS AND MACHINE LEARNING 2
2.1.1 Dataset handling 2
2.1.2 Data inconsistencies, cleaning, and missing value imputation 3
2.2 OVERVIEW OF MACHINE LEARNING ALGORITHMS USED IN THE THESIS 6
2.2.1 Supervised machine learning approaches 7
2.2.2 Unsupervised machine learning approaches 9
2.2.3 Deep learning approaches 9
2.2.4 Model optimisation and hyperparameter tuning 11
2.2.5 Evaluation metrics of machine learning models 15
2.2.6 Model Interpretability 16

CHAPTER 3 – APPLIED MACHINE LEARNING AND THE ROLE OF DYNAMIC TIME-SERIES DATA IN THE PREDICTION OF BACTERIAL BLOODSTREAM INFECTIONS 18

3.1 CHAPTER OVERVIEW AND SUMMARY 18
3.2 BACTERIAL BLOODSTREAM INFECTIONS AS A CAUSE OF ACUTE FEBRILE ILLNESS 20
3.2.1 Aetiology and existing laboratory diagnosis 21
3.2.2 Clinical decision-making for bloodstream infections 23
3.3 AIMS AND RESEARCH HYPOTHESES 31
3.4 STATEMENT OF CONTRIBUTION 31
3.5 ETHICS STATEMENT 32
3.6 METHODS 32
3.6.1 Description of dataset 32
3.6.2 Feature selection 33
3.6.3 MODEL APPROACHES AND ASSUMPTIONS 34
3.6.4 Data pre-processing 35

3.6.5 Algorithms and optimisation 36
3.7 RESULTS 37
3.7.1 Baseline description of data 37
3.7.2 Data completeness 40
3.7.3 Descriptive data patterns over time 41
3.7.4 Prediction of bloodstream infections through time-invariant approaches 43
3.7.5 Prediction of bloodstream infections using deep learning 46
3.7.6 Predicting hospital-onset bloodstream infections 47
3.7.7 Prediction of bacterial gram stain results 48
3.7.8 Windowed approaches for prediction of bloodstream infections 49
3.7.9 Deep learning for continuous prediction of bloodstream infections 50
3.7.10 Importance of sequence order in the deep learning models 52
3.7.11 Comparator model approaches through logistic regression and time-invariant models 54
3.8 INTERPRETABILITY 56
3.8.1 Change in model risk prediction of bloodstream infection over time 56
3.8.2 Interpretability analysis of deep learning models 59
3.8.3 Error analysis and model bias of prediction models 61
3.9 DISCUSSION 64

CHAPTER 4 – APPLIED MACHINE LEARNING AND THE ROLE OF DYNAMIC TIME-SERIES DATA IN THE RISK PREDICTION OF DENGUE DISEASE PROGRESSION IN HOSPITALISED PATIENTS 71

4.1 CHAPTER OVERVIEW AND SUMMARY 71
4.2 GLOBAL BURDEN OF DENGUE 73
4.2.1 Clinical features of dengue 73
4.2.2 Priorities for clinical management of dengue in an endemic setting 74
4.2.3 Risk factors and the clinical prediction of dengue progression 76
4.2.4 Incorporating dynamic factors to augment risk prediction 77
4.3 AIMS AND RESEARCH HYPOTHESES 80
4.4 STATEMENT OF CONTRIBUTION 80
4.5 ETHICS STATEMENT 81
4.6 METHODS 81
4.6.1 Description of clinical datasets for model development 81
4.6.2 Dataset pre-processing approaches 82
4.6.3 Definition of prediction outcomes 83
4.6.4 Feature selection, missing data, model development and evaluation approaches 86
4.6.5 Time-series approaches and baseline models 90
4.7 RESULTS 90
4.7.1 Quantitative analysis of data 90

4.7.2 Model performance for prediction of dengue shock using 48 hour aggregated data 90
4.7.3 Model performance for secondary outcomes and time-invariant model..... 94
4.7.4 Model interpretability 95
4.8 Discussion 100

CHAPTER 5 – ACCOUNTING FOR THE DYNAMIC INFLUENCE OF SEASONALITY IN THE DIAGNOSIS OF ACUTE FEBRILE ILLNESSES 103

5.1 Chapter overview and summary 103
5.2 The acute febrile illness syndrome 105
5.2.1 Role of diagnostics in low and middle-income country settings..... 106
5.2.2 The role of adjunctive data to support patient diagnosis..... 107
5.2.3 Impact of non-healthcare factors in clinical diagnosis 108
5.3 Aims and Research hypotheses 108
5.4 Statement of contribution 109
5.5 Ethics statement 109
5.6 Methods 110
5.6.1 Dengue outcome definitions 110
5.6.2 Feature selection of models..... 111
5.6.3 Model development procedures..... 111
5.6.4 Characterising the relationship between outcome and seasonality 112
5.7 Results 114
5.7.1 Description of demographics and seasonal variation 114
5.7.2 Baseline model performance for predicting acute febrile illness diagnosis 117
5.7.3 Performance of the model when applied over the study period 118
5.7.4 Empirical tuning of thresholds over time 119
5.7.5 Climatic model to inform changing thresholds..... 121
5.7.6 Understanding feature distributions of patients diagnosed with other febrile illness .. 125
5.8 Discussion 131

CHAPTER 6 - THE IMPLEMENTATION OF MACHINE LEARNING MODELS UTILISING DYNAMIC TIME-SERIES DATA, THROUGH HEALTHCARE PATHWAYS MAPPING AND UNDERSTANDING OF DECISION-MAKING 134

6.1 Chapter overview and summary 134
6.2 Implementation of machine learning models to the healthcare setting 136
6.3 Aims and research hypotheses 138
6.4 Statement of contribution 139
6.5 Ethics statement 139
6.6 Methods 139

6.6.1 Process mapping and task analysis of existing practice 140
6.6.2 Semi-structured interviews to characterise decision-making 141
6.6.3 Workshops for the design of decision-support interfaces 142
6.6.4 Wireframe and mock ups of decision-support systems 142
6.7 RESULTS 143
6.7.1 Pathway mapping of patients admitted with dengue 143
6.7.2 Task analysis of healthcare staff 146
6.7.3 Clinical decision-making in dengue management and emergent themes 149
6.7.4 Workshops to understand ideal functionality of decision-support systems 153
6.7.5 Wireframe designs and development of computerised interfaces 158
6.8 DISCUSSION 160

CHAPTER 7 - ENHANCING DYNAMIC DATA COLLECTION IN ACUTE FEBRILE ILLNESS TO INFORM CLINICAL CARE: INVESTIGATION OF CONTINUOUS LACTATE MONITORING THROUGH A MINIMALLY-INVASIVE BIOSENSOR 164

7.1 CHAPTER OVERVIEW AND SUMMARY 164
7.2 USE OF DYNAMIC HEALTHCARE DATA IN THE MANAGEMENT OF ACUTE FEBRILE ILLNESSES 166
7.2.1 The role of lactate in acute febrile illnesses and other conditions 167
7.2.2 Continuous patient monitoring using the microneedle platform 169
7.3 AIMS AND RESEARCH HYPOTHESES 171
7.4 STATEMENT OF CONTRIBUTION 172
7.5 ETHICS STATEMENT 172
7.6 METHODS 173
7.6.1 Participant population and recruitment 173
7.6.2 Reporting of findings 174
7.6.3 Technical aspects of the Lactate Oxidase Microneedle Biosensor 174
7.6.4 Study procedures 175
7.6.5 Sample size, data and statistical analyses 179
7.7 RESULTS 180
7.7.1 Baseline description of participants and characteristics 180
7.7.2 Results of microneedle measurement of lactate 181
7.7.3 Participant acceptability of the microneedle biosensor 187
7.8 Public acceptability of the microneedle biosensor 188
7.9 DISCUSSION 190

CHAPTER 8 - DISCUSSION AND CONCLUSIONS 194

8.1 SUMMARY OF THESIS FINDINGS 194
8.2 FINDINGS IN A WIDER CONTEXT 197

8.3 Future perspectives and recommendations 200

REFERENCES..... 202

I THESIS APPENDIX..... 227

A List of peer-reviewed publications arising from the thesis 228
B Ethics approval for use of BSI dataset from Imperial College Healthcare NHS Trust 239
C Study protocol for MIRACLE 241
D Patient information sheet for MIRACLE study 255
E Patient questionnaire 263
F Patient consent form for MIRACLE study 264
G Sponsorship letter from Imperial College London for MIRACLE study 265
H Correspondence with MHRA and notice of no objection for device validation study 267
I IRAS application form for MIRACLE study for Health Research Authority approval 270
J HRA ethics approval letter for MIRACLE study 307
K Correspondence with Imperial College Healthcare NHS Trust 317
L Ethics protocol for process mapping study at Hospital for Tropical Diseases, Vietnam 320
M Ethics approval and correspondence from HTD and OXTREC for process mapping study 332
N Workshop guides 338

II SUPPLEMENTARY DATA AND RESULTS..... 358

A Dengue risk for Hospitalised Patients 358
i Details of individual studies 358
ii Full results for model evaluation in prediction of severe dengue outcomes 365
iii Static models 367
B Seasonality and time on clinical management of acute febrile illnesses 369
i Optimal model configuration 369
ii Window size 369
iii Clustering approaches for other febrile illnesses 369
iv Deep learning parameters 371
C Deep learning for BSI 372
i Correlation between parameters 372
ii Data completeness and differences 372
iii Models used in the chapter 374
iv A regression approach to lactate prediction 375
D Implementation of models and pathway mapping 378
v Coding results of interviews 378
vi Next steps of work for decision-support development 397

List of tables

TABLE 1.1 SELECTION OF CLINICAL SCORES USED FOR USE IN THE MANAGEMENT OF ACUTE FEBRILE ILLNESSES AND INFECTIOUS DISEASES. 19

TABLE 2.1 OVERVIEW OF HYPERPARAMETERS OPTIMISATION FOR NEURAL NETWORKS 13

TABLE 3.1 LIMITATIONS TO CURRENT CULTURE-BASED METHODS FOR THE DIAGNOSIS OF BACTERAEMIA AND BSI 22

TABLE 3.2 CLINICAL PARAMETERS FOR THE SHAPIRO SCORE TO PREDICT SUBSEQUENT BSI 24

TABLE 3.3 SUMMARY OF DATA-DRIVEN MACHINE LEARNING APPROACHES FOR THE PREDICTION OF BSI 26

TABLE 3.4 FEATURES INCLUDED IN MODEL DEVELOPMENT AND FINAL FEATURES 33

TABLE 3.5 MICROBIOLOGICAL DETAILS OF BLOOD CULTURES FROM 4,198 PATIENTS INCLUDED IN ANALYSIS 37

TABLE 3.6 A COMPARISON OF FEATURES USED IN THE FINAL MODELS. 38

TABLE 3.7 DATA COMPLETENESS FOR EACH OF THE FEATURE VARIABLES. 40

TABLE 3.8 TIME INVARIANT BASELINE MODEL TO PREDICT OUTCOME. 45

TABLE 3.9 RESULTS OF 5-FOLD CROSS VALIDATION OF THE LSTM MODEL 47

TABLE 3.10 DETAILS OF THE TRAINING, VALIDATION AND HOLD OUT TEST SET USED FOR MODEL TRAINING 51

TABLE 3.11 RESULTS OF LOGISTIC REGRESSION MODEL THROUGH THREE APPROACHES 55

TABLE 3.12 PREDICTIONS OVER TIME STRATIFIED BY WINDOW FOR THE HOLD OUT TEST SET 59

TABLE 3.13 SUBGROUPS OF THE HOLD OUT TEST SET FOR THE BASELINE UNORDERED LSTM MODEL 62

TABLE 4.1 DENGUE WITH WARNING SIGNS CRITERIA, FROM WHO DENGUE GUIDELINES 76

TABLE 4.2 POOLED ODDS RATIOS OF PROGRESSION TO SEVERE DENGUE FOR RISK FACTORS 77

TABLE 4.3 CLINICAL PREDICTION MODELS FOR PREDICTION OF SEVERE VERSUS NON-SEVERE DENGUE 78

TABLE 4.4 STUDIES (S1-5) INCLUDED IN ANALYSIS AND CORRESPONDING REFERENCES 81

TABLE 4.5 BASELINE COMPARISONS BETWEEN DEVELOPMENT AND HOLD OUT TEST SET 86

TABLE 4.6 - BASELINE CHARACTERISTICS OF INDIVIDUAL STUDIES INCLUDED IN THE FINAL DATASET 92

TABLE 4.7 CHARACTERISTICS OF COHORT DIVIDED BY COMPLICATION OUTCOME DURING HOSPITAL. 93

TABLE 4.8 FINAL MODEL PERFORMANCE AFTER HYPERPARAMETER OPTIMISATION 93

TABLE 4.9 PERFORMANCE METRICS OF THE FINAL MODELS AFTER UNDERGOING CALIBRATION 94

TABLE 4.10 ARGUMENTS FOR EXPLAINABILITY IN MACHINE LEARNING FOR REAL-WORLD APPLICATIONS 97

TABLE 5.1 BASELINE CHARACTERISTICS AND RESULTS OF INVESTIGATIONS AT ENROLMENT 115

TABLE 5.2 RESULTS OF MODEL 10-FOLD CROSS VALIDATION AND HOLD OUT SET EVALUATION. 117

TABLE 5.3 SUMMARY STATISTICS FOR NPV PERFORMANCE FOR THE THREE DIFFERENT MODELS. 125

TABLE 6.1 TEN COMMANDMENTS FOR EFFECTIVE CLINICAL DECISION SUPPORT, FROM [265]. 137

TABLE 6.2 A LIST OF TASKS CAPTURED DURING PATIENT PATHWAY OBSERVATIONS 146

TABLE 6.3 FINDINGS FROM WORKSHOP REGARDING IDEAL FEATURES FOR A DENGUE CDSS 154

TABLE 7.1 EXISTING CLINICAL AND RESEARCH APPROACHES TO LACTATE MEASUREMENT 168

TABLE 7.2 BASELINE CHARACTERISTICS AND RESULTS OF PARTICIPANTS ENROLLED 181

TABLE 7.3 SUMMARISED FINDINGS FROM PUBLIC ACCEPTABILITY OF THE MICRONEEDLE BIOSENSOR PATCH THROUGH INDIVIDUAL AND GROUP CONVERSATIONS IN A PEOPLE'S RESEARCH CAFÉ SETTING. 189

List of figures

Figure 1.1 Thesis outline 1

Figure 2.1 Bias-variance trade off for models. 8

Figure 2.2 Cross validation and grid search procedure. 12

Figure 3.1 Approaches used by models 35

Figure 3.2 Histogram distribution of patients with any biomarker results from the feature list 40

Figure 3.3 A visualisation of final model features and their longitudinal trends over time 42

Figure 3.4 Heatmap of top 10 important features in the XGBoost model 46

Figure 3.5 Confusion matrix of predictions against actual classification for hospital onset bacteraemia 48

Figure 3.6 Confusion matrix of predictions against actual classification for whole cohort 49

Figure 3.7 learning curves for LSTM models for the random split and ordered split datasets. 51

Figure 3.8 Confusion matrices of the ordered LSTM model 53

Figure 3.9 Confusion matrices of the shuffled LSTM model 53

Figure 3.10 Predictions over time for subset of patients resulted in a true positive. 57

Figure 3.11 Predictions over time for subset of patients resulted in a true negative 58

Figure 3.12 SHAP values for each timesteps from a sample of 1,000 predictions 60

Figure 4.1 Flow diagram for patients recruited in original clinical studies 85

Figure 4.2 Data pipeline used for model training 89

Figure 4.3 Confusion matrix displaying performance of the hold out test set for XGBoost 94

Figure 4.4 Calibration curve for the artificial neural network model 97

Figure 4.5 Summary plot of SHAP values for the XGBoost model 99

Figure 4.6 A scatter plot of predictor values against SHAP values 99

Figure 5.1 Expected test results for dengue diagnosis 111

Figure 5.2 Dynamic threshold model. 114

Figure 5.3 Ratio of dengue:OFI over time with average rainfall and temperature 116

Figure 5.4 Feature importance of the XGBoost model 118

Figure 5.5 Calibration curve of the final model applied to the hold out test set 118

Figure 5.6 Empirical dynamic model over study period 121

Figure 5.7 Plot of temperature, precipitation and optimal thresholds 122

Figure 5.8 Warping paths produced by dynamic time warping 123

Figure 5.9 Prediction of future thresholds to achieve a NPV > 90%. 124

Figure 5.10 A plot of negative predictive values achieved in the latter half of the study 125

Figure 5.11 Area under the receiver operating curve of the model over the entire study period 126

Figure 5.12 Parallel coordinates plots showing the relationships between features 128

Figure 5.13 Permutation importance of lymphocyte and white cell count 130

Figure 6.1 Process map of dengue management for adult and children 145

Figure 6.2 Screenshots of online workshop board 157

FIGURE 6.3 A SELECTION OF WIREFRAME DESIGNS DEVELOPED FOR THE DENGUE CDSS 159
FIGURE 7.1 LACTATE MICRONEEDLE BIOSENSOR. 170
FIGURE 7.2 RESPONSE OF LACTATE OXIDASE BIOSENSORS 175
FIGURE 7.3 STUDY OUTLINE AND TIMINGS 178
FIGURE 7.4 STUDY RECRUITMENT. 180
FIGURE 7.5 PLOTS OF INDIVIDUAL MICRONEEDLE CURRENT AGAINST VENOUS LACTATE LEVELS FOR EACH PARTICIPANT..... 183
FIGURE 7.6 RESULTS OF MICRODIALYSIS FOR PARTICIPANT 5 184
FIGURE 7.7 BLAND-ALTMAN ANALYSES OF CALIBRATED BIOSENSOR OUTPUTS 185
FIGURE 7.8 RATE OF CHANGE BETWEEN BIOSENSOR AND VENOUS LACTATE FOR ALL PARTICIPANTS 186
FIGURE 7.9 AGGREGATED RATE OF CHANGE FOR ALL 5 PARTICIPANTS 186
FIGURE 7.10 PHOTOGRAPHS OF SKIN AFTER REMOVAL OF THE MICRONEEDLE BIOSENSOR. 188
FIGURE II.1 PEARSON CORRELATION HEATMAP WITH COLOUR INTENSITY SIGNIFYING STRENGTH OF CORRELATION 372
FIGURE II.2 HEATMAPS SHOWING DATA AVAILABILITY. 373
FIGURE II.3 LSTM MODEL OPTIMISED FOR ROLLING WINDOW PREDICTIONS 375
FIGURE II.4 ACTUAL VS PREDICTED LACTATE VALUES 376
FIGURE II.5 BLAND-ALTMAN PLOT SHOWING DIFFERENCES BETWEEN PREDICTED AND ACTUAL LACTATE VALUES. 377

Abbreviations and definitions

AFI	Acute febrile illness
AUPRC	Area under the precision recall curve
AUROC	Area under the receiver operating curve
BSI	Bloodstream infections
CDSS	Clinical decision support system
CI	Confidence interval
COVID-19	Coronavirus disease 2019
DENV	Dengue virus
DET	Direct electron transfer
DF	Dengue fever
DHF	Dengue haemorrhagic fever
DSS	Dengue shock syndrome
DWWS	Dengue with warning signs
EBM	Evidence-based medicine
EHR	Electronic health records
EU	European Union
FDA	Food and Drug Administration
GRU	Gated Recurrent Unit
HRA	Health Research Authority
HTD	Hospital for Tropical Diseases, Ho Chi Minh City
HTML	HyperText Markup Language
ICHNT	Imperial College Healthcare NHS Trust
IMDRF	International Medical Devices Regulators Forum
IQR	Interquartile range

ISF	Interstitial fluid
LAMP	Loop-mediated isothermal amplification
LASSO	Least absolute shrinkage and selection operator
LMIC	Low- and middle-income country
LOx	Lactate oxidase
LSTM	Long short-term memory
MeSH	Medical subject headings
MHRA	Medicines and Healthcare products Regulatory Agency
MIRACLE	Minimally-invasive real time assessment of continuous lactate in exercise
MLP	Multilayer perceptron
MOH	Ministry of Health
NAAT	Nucleic acid amplification test
NaN	Not a number
NEWS2	National Early Warning Score 2
NIHR	National Institute for Health and Care Research
NMFI	Non-malarial febrile illness
NPV	Negative predictive value
NS1	Non-structural protein 1
OFI	Other febrile illness
OPD	Outpatient department
OUCRU	Oxford University Clinical Research Unit
PCA	Principal component analysis
PPV	Positive predictive value
qSOFA	Quick sequential organ failure assessment
RNA	Ribonucleic acid
RNN	Recurrent neural network

RT-PCR	Real-time polymerase chain reaction
SaMD	Software as medical device
SHAP	Shapley additive explanations
SOFA	Sequential organ failure assessment
SVM	Support vector machine
WHO	World Health Organization

Chapter 1 - Introduction

1.1 Aims of the thesis

The primary aims of my thesis are to examine the role of machine learning and contribution of dynamic, time-series information for the development of clinical prediction models. I will focus specifically on the management of acute febrile illnesses: routinely collected clinical data from patients with bacterial bloodstream infections in the UK and dengue in Vietnam and will be used as exemplars.

The secondary aims are to address issues related to the translation of my thesis findings to the clinical setting. I will use mixed methods approaches to understand key factors in healthcare implementation of machine learning models, with a focus on dengue management in Vietnam. I will address the problems of time-series data acquisition in healthcare through research in minimally-invasive biosensing for continuous patient monitoring.

1.2 Medical decision-making in acute febrile illnesses

Fever is a cardinal feature in infectious diseases, characterised by a raised body temperature greater than 38 °C. Acute febrile illness (AFI) represents a clinical syndrome of fever with, or without localising signs. Worldwide, AFIs are a common clinical presentation to healthcare [1] and major infectious aetiologies of AFI include seasonal epidemics such as dengue and influenza, bloodstream infections, malaria, scrub typhus, leptospirosis, and more recently, COVID-19. Together, these conditions exert considerable morbidity and mortality, placing burdens on health systems globally [2].

Numerous clinical decisions are made during the evaluation and management of AFI. These decisions include the diagnosis of underlying aetiology (if an infective cause is sought), patient risk stratification, selection and dosing of therapy, supportive care, establishing prognosis, decisions on optimal place of care, and the inclusion of adjunctive investigations. These decisions are influenced by the healthcare setting and often need to account for any implications on public health. Specifically for bacterial and fungal infections, treatment should ideally balance immediate individual benefit

with public health, i.e. the risk of antimicrobial resistance [3], and consider infection control precautions where there is risk for patient-to-patient transmission.

Information available to the clinician during the decision-making processes can be abundant. Data utilised in these processes include patient history, physical examination findings, vital signs measurements, results of bedside and laboratory investigations and non-clinical information such as cost and patient preference. These aspects are synthesised and interpreted within a series of cognitive frameworks [4]. Although the mode of reasoning may differ between clinicians, a mix of intuitive and analytical approaches commonly adopted aim to provide a pragmatic (and ideally accurate) representation of patient state, allowing for inferences on disease causation, progression and management to be made [5].

Clinical tools including decision pathways, management algorithms and scoring systems allow for the translation of evidence-based medicine (EBM) research findings into the healthcare setting. The role of such tools is to improve the quality and consistency of clinical decision-making through incorporating research findings [6]. For AFI and the management of infection, clinical tools have been developed to support diagnosis, risk-stratification, prognosis, or treatment choice. The application of these clinical tools for the management of infection is likely to be of the greatest utility when used by the bedside given patients are often acutely unwell and there exists the need for timely decision-making. These tools need to be sufficiently straightforward to implement but often have a trade-off in terms of limited complexity and scope, and tend to focus only on specific aspects of management. The ability of such tools to account for patient subgroups and individual treatment effects is also restricted, which attenuates clinical impact.

Exemplars of clinical tools (specifically clinical scoring systems which assign a numerical value to clinical parameters and provide a recommendation or prediction depending on value) utilised routinely in UK hospital and primary care settings are shown in table 1.1.

Table 1.1 Selection of clinical scores used for use in the management of acute febrile illnesses and infectious diseases.

Name of clinical scoring system	Year of development	Intended use	Criteria	Application in clinical setting
4C mortality score for COVID-19 [7]	2020	Prediction of in-hospital mortality for patients hospitalised with COVID-19	Age, sex, number of comorbidities, respiratory rate, oxygen saturation, Glasgow Coma Scale, urea, C-reactive Protein (CRP) levels	Scoring done at hospital presentation
Centor criteria [8]	1981	Likelihood of bacterial (Streptococcal) pharyngitis	Age, tonsillar exudate, cervical lymphadenopathy, fever, cough	Scoring done on clinical presentation
CURB-65 score [9]	2003	Prediction of mortality in community-acquired pneumonia	Confusion, urea, respiratory rate, blood pressure, age	Scoring done at clinical presentation after confirmation of consolidation on radiology
Duke criteria [10]	1994	Diagnosis for infective endocarditis	Pathological criteria, positive blood cultures, echocardiography results, predisposition, fever, vascular, immunological or microbiological evidence.	Scoring done at timepoint of clinical evaluation
WHO dengue guidelines 2009 [11]	2009	Diagnosis and risk stratification for dengue	Presence of nausea, rash, myalgia, leukopenia, abdominal pain, persistent vomiting, clinical fluid accumulation, lethargy, hepatomegaly, increase	Scoring done at timepoint clinical evaluation. Criteria of "increase in haematocrit and decrease in

			in haematocrit with concurrent decrease in platelets, presence of shock, bleeding or organ impairment	platelet" requires assessment over time although the exact magnitude of this change is not explicit.
Pitt bacteraemia score [12]	1999	Prediction of mortality in bacteraemia	Fever, blood pressure, use of mechanical ventilation, cardiac arrest and mental status.	Scoring done 48 hours before, or on the day of positive blood culture
Sequential organ failure assessment (SOFA) [13]	1996	Prediction of mortality in an intensive care setting.	Oxygen saturation, mechanical ventilation, platelet count, Glasgow coma score, bilirubin, blood pressure, creatinine	Scoring done using the worst parameters in a 24-hour period, daily.

1.3 Healthcare digitalisation and the use of dynamic patient data

There has been a significant and ongoing increase in the availability, quantity and diversity of clinical data (termed 'Big Data'). These changes have been attributed to developments in healthcare digitalisation, widespread adoption of electronic healthcare records (EHR) [14] and advances in high throughput '-omics' research. In parallel, sources of information which will likely to play a growing role in decision-making across healthcare are also increasing in availability. These include: i) secondary data – data not generated or collected for intended purposes of health evaluation, ii) metadata – data which provides information on other data, and iii) adjunctive data – information not directly related to the patient or the disease but may nonetheless provide context [15].

Ensuring that healthcare systems are able to effectively utilise these diverse datasets, and operate in conjunction with other technological innovations is a priority [16–18]. The uptake of digital health innovations has increased in pace globally, with many low- and middle-income (LMIC) country settings adopting technological interventions in healthcare, driven in part by the COVID-19 pandemic. For example, there is increasing acknowledgement that use of wearable devices such as pulse oximeters for inpatient and outpatient management could fill a gap in patient monitoring [19].

Given the opportunities of using data to guide patient care with far greater precision than before [20], it is all the more crucial that such resources can be utilised effectively. Methodologies in applying these data to directly inform decision-making at both a public health and an individual clinical level need to be developed, evaluated and better understood.

There are three specific aspects of healthcare datasets which warrant consideration, and these form the focus of my thesis:

i) **The use of individual longitudinal data** which may span hours to years. Monitoring patient progress and response to treatment over time is fundamental to clinical practice (for example, in AFI a 'watch and wait' strategy might be appropriate). Clinical management is inherently dynamic and responsive to changes to patient state. Time-series information relating to physical status, vital signs and biomarker information and their change is implicitly considered in clinical management. There is support that the explicit consideration of biomarker dynamics and their change over time provides added information for a variety of infections – such as lactate clearance in shock states [21, 22] and change in C-Reactive Protein (CRP) in COVID-19 [23]. Existing clinical tools however do not (or are unable to) adequately account for such time-series relationships: none of the clinical scores and algorithms listed in table 1.1 make use of the longitudinal aspect of clinical information. Risk scores designed for clinical use, such as the National Early Warning Score 2 (NEWS2) [24] or Sequential Organ Failure Assessment (SOFA) have implementations showing repeated evaluation and deriving their change over fixed timepoints can provide clinically-relevant information [25, 26]. However, this does not adequately capture their time-series dynamics and provides a limited understanding of the interactions with other time-dependent healthcare data.

ii) **The inclusion of a diverse range of healthcare data** in clinical decision-making. This includes data related to clinical presentations, vital signs, radiology, laboratory results, as well healthcare administration, all generated as part of routine healthcare. Although much of this information is integrated and implicitly considered as part of clinical decision-making – there exists significant variation on how information is applied in practice [27]. Challenges to data use include understanding the relative importance of each component, and the need to consider large volumes of data in a consistent fashion [28]. Inherent complexities

in healthcare data such as non-linear relationships between variables, missing data and diverse data types (e.g., data in the form of images, signals and measurements) mean that the ability for traditional statistical approaches to modelling may be limited [29]. The development of analytical frameworks which can flexibly interpret such diverse data and provide understanding of their latent relationships is therefore a priority.

iii) **The use of adjunctive, non-healthcare data** to support and inform patient management. Use of non-patient or disease data for example, relating to seasonality, disease vector populations, climate patterns and change may provide useful information, especially for the management of acute infections [30]. Factors such as human movement [31], or effects of climate on disease vector populations [32] in turn have downstream effects on disease incidence and presentation. The increased availability of such data through data-sharing initiatives enables this information to be accessed. Examples of use include forecast modelling of dengue disease incidence on a regional level through use of weather data [33]. However, optimal methods to formally integrate and combine these sources of data specifically with healthcare data for clinical decision-making or their contributions to such models remain unclear.

There is a need to address the conceptual and methodological challenges regarding the use and application of these diverse healthcare and non-healthcare data to support clinical management. Conventional approaches for understanding relationships between patient variables and outcomes include use of randomised clinical trials or cohort studies. These methodologies aim to control for confounders through randomisation, imposing strict inclusion criteria, and examination of only a small set of pre-specified variables. Such hypothesis-driven approaches may be ill-suited to fully utilising available data because of factors including high cost and feasibility [34]. Although the use of retrospective observational data is subject to inherent bias, there remains significant opportunities in the use of novel data methodologies for augmenting the clinical impact of such data for decision-making.

1.4 Machine learning to understand complex healthcare data

Data-driven methods such as machine learning offer complementary approaches to information handling and interpretation. Machine learning refers to a growing set of

algorithmic approaches, which aim to identify patterns and associations in data and generate hypotheses and knowledge [35]. Supervised and unsupervised machine learning methods (defined in Chapter 2) are adept at handling high-dimensional data in real-time, and are well suited for high volume data. Examples of these data relevant to healthcare include time series information, language, images, or combined datasets which are integrated from multiple sources. Such tools are inherently amenable to scale up, particularly in a global context where digitalisation through computers and smartphones is widespread [36].

The presentation of machine learning models through digital clinical decision support systems (CDSS) as clinical tools could allow for healthcare data to be considered in far greater detail than previously possible. The number of machine learning models which are approved by regulatory bodies for clinical implementation are increasing. Examples include those designed to assist in image interpretation for radiology or histology data [37–39]. Data-driven methods are particularly suited for the analysis of large volumes of routine clinical data from electronic health records. However, there is a paucity of work on the use of machine learning to inform clinical decision-making specifically for the management of AFI. Research questions remain on how such diverse healthcare data should be used most effectively, and how to ensure these findings result in clinical impact. Research should also address limitations inherent in existing clinical tools: these include being able to explicitly handle the large quantities and types of individual patient healthcare data and their time-series dependencies, and also provide linkages between sources of information (e.g., from clinical tabular, radiological imaging, continuous signal data).

With recent advances in algorithmic understanding, computing processing power and healthcare data-availability (with emphasis on data-sharing), research in the application of machine learning for healthcare is timely and important.

1.5 Dengue and bloodstream infections as important exemplars of acute febrile illnesses

In this thesis I have chosen two important infections from contrasting healthcare settings as exemplars of AFI to address my research questions (see below):

1. Bacterial bloodstream infections (BSI) and bacteraemia are caused by the presence of viable bacteria in blood. The condition increases the risk of sepsis

which exerts significant mortality and morbidity worldwide. Early recognition and appropriate clinical management are important in optimising patient outcomes. I will examine this subtype of AFI from a UK perspective and utilise routinely collected blood biomarker and clinical data from electronic healthcare records to inform model development.

2. Infection caused by dengue virus (DENV) is common cause of fever worldwide, resulting in 96 million symptomatic infections each year [40]. Treatment is supportive and epidemics exert significant pressure on healthcare systems. The diagnosis of dengue from presenting symptoms (usually an undifferentiated fever) and identification of patients at high risk from severe disease are research priorities. I will examine the development of clinical prediction models from a Vietnam perspective, with focus on the use of basic laboratory and clinical parameters. I will also make use of adjunctive but non-healthcare data including local rainfall and temperature to augment clinical prediction models through providing context.

1.6 Addressing optimal use of healthcare data and acquisition

This thesis will examine how healthcare data can be utilised effectively to support clinical decision making for the management of AFI. Focus is on the use of data which is routinely collected as part of clinical care. These include demographics, history and examination findings and results of investigations – commonly blood biomarkers. The extent to which such data is available (and constituting as 'routinely collected') differs across different healthcare settings; the choice of two exemplars (bacterial bloodstream infections and dengue) managed in contrasting healthcare settings in the UK and Vietnam respectively will increase the generalisability of my findings.

The additional dimension of how such clinical data changes over time, i.e., the dynamic time-series aspect, and the contribution to the performances of clinical prediction models which account for these features are important. Changes in patient state as a function of disease progression or response to treatment will be explicitly considered using a range of machine learning methods (a description of machine learning, and algorithms used in this work is further described in Chapter 2). I will also investigate how machine learning models developed to support dengue management can be introduced and implemented in a healthcare setting.

Opportunities for the acquisition of dynamic time-series data from individuals will be examined. In particular continuous sensing of patient biomarkers over time through novel minimally-invasive approaches could provide significant benefits and address the many limitations in healthcare systems which result in the inadequate acquisition of patient data. The thesis will specifically address this issue through an exemplar of a biosensor modality based on the microneedle platform. The biosensor will be functionalised to detect lactate, which is an important prognostic indicator for a range of infectious diseases including dengue and bloodstream infections.

1.7 Research hypothesis and thesis aims

The hypothesis for this thesis is:

> "The inclusion of dynamic time-series data through the application of machine learning models can enhance clinical decision-making for the management of acute febrile illnesses and offer advantages over existing approaches".

For the clinical management of acute febrile illnesses, the specific research questions which I will address are:

1. Do machine learning clinical prediction models which explicitly account for dynamic changes in data over time offer better predictive performance, robustness and/or clinical utility over conventional static approaches? What are the comparative strengths and weaknesses of such approaches across different healthcare settings? (Chapter 3, Chapter 4 and Chapter 5)

2. How can machine learning models be implemented effectively into a healthcare setting, and what important factors related to clinical decision-making need to be addressed? (Chapter 6)

3. How can the limitations relating to healthcare data availability be addressed? What innovations can be used to improve the acquisition of dynamic clinical data to support decision-making and model development? (Chapter 7)

I will answer these research questions by bringing together a range of clinical research, data science, bioengineering and qualitative techniques. My main focus is on the application of machine learning and statistical approaches for dynamic time-series healthcare in the management of AFI. I will develop and explore prediction models with

the aim of providing clinically relevant information evaluate them against other approaches.

Machine learning work will make use of two exemplar clinical datasets which represent important AFI from contrasting healthcare resource settings. I present research and analyses on patients with dengue from a tertiary hospital in Ho Chi Minh City, Vietnam and those with bacterial bloodstream infection in London, United Kingdom.

The secondary aims of the thesis are to address the key issues relating to the clinical translation and implementation of such machine learning models to real-world healthcare settings. I will employ qualitative and health systems research methods to better understand factors important in decision-making for AFI and use this information to examine the questions around healthcare implementation. Challenges to healthcare data collection in many healthcare settings limited the role of dynamic data available to support healthcare. I will address the issue of optimising healthcare data acquisition through research on novel biosensing technologies, which collect continuous physiological information in real-time from the individual.

The thesis will contribute to knowledge in the field of clinical research through providing better understanding of performance characteristics in the use of machine learning and time-series approaches. Research and models from the thesis have direct relevance to clinical practice: results are amenable to translation as clinical prediction scores for example, within CDSS frameworks.

Work on characterising and addressing the limitations of routine healthcare data use is also important. My thesis findings will aim to provide support for the future use of biosensors or wearable technologies in healthcare and directly address the gaps in patient data acquisition. Characterising decision-making and healthcare processes for implementation also provides important practical, and methodological results which is of utility in this field of research.

1.8 Thesis structure and chapter outlines

Chapter 1 (current chapter) contains the introduction, rationale and research questions to be addressed by the thesis.

Chapter 2 provides common methods, time-series machine learning approaches and considerations utilised throughout the thesis. Specific techniques unique to each chapter are expanded in their relevant locations.

Chapter 3 examines the use of routinely collected patient healthcare data with a focus on their biomarker dynamics. The exemplar of predicting bacterial bloodstream infections at a set of UK teaching hospitals at Imperial College Healthcare NHS Trust in London, United Kingdom will be used. I present my findings on the development and evaluation of deep-learning models which explicitly handle time-series information from a cohort of 4,198 hospitalised patients with bacterial bloodstream infections. I will also investigate model interpretability and perform error analysis to better understand model function.

Chapter 4 provides research findings on the use of dynamic and time-series data within a low- and middle-income country setting, in the context of limited data availability. I focus on the role of machine learning models for predicting dengue shock in hospitalised patients and examine how biomarker dynamics over time can be accounted for. Work will be done using a patient cohort of 4,131 patients hospitalised with confirmed dengue in Ho Chi Minh City, Vietnam.

Chapter 5 explores and provides understanding on the inclusion and role of non-healthcare data which changes over time, to specifically guide clinical decision-making at an individual level. I examine the role of such information as an adjunct to support conventional diagnostics by focusing on AFI in Vietnam in the clinical setting as an exemplar. Specifically, I will investigate and account for the dynamic influence of local non-healthcare factors such as local rainfall and temperature into prediction models. A dataset of 8,100 children presenting with fever to healthcare facilities in Southern Vietnam will be used.

Chapter 6 addresses issues of real-world healthcare implementation of machine learning models, specifically for dengue. I will answer this through utilising health systems research methodologies to map care pathways and better understand clinical decision-making in dengue for a hospital context in Vietnam. I analyse results from semi-structured interviews of clinicians (n=10) to map emergent themes and understand priorities in decision-making in dengue. I present results on group workshops carried out with the aim

of co-design of a clinical decision-support system based on models presented in Chapters 4 and 5.

Chapter 7 addresses the issue of ensuring high-quality, granular patient data is available to the clinician and for supporting development of robust time-series machine learning models. Specifically, I focus on problems of limited patient biomarker data acquisition identified in previous chapters. This is particularly relevant for individuals cared for outside a critical care environment, or those managed in a pre-hospital / resource-limited setting (with limited access to laboratory services). I present results on the use of a novel minimally-invasive biosensor capable of real-time continuous lactate measurement without blood and results of a Phase I clinical evaluation in a set of participants (n=5).

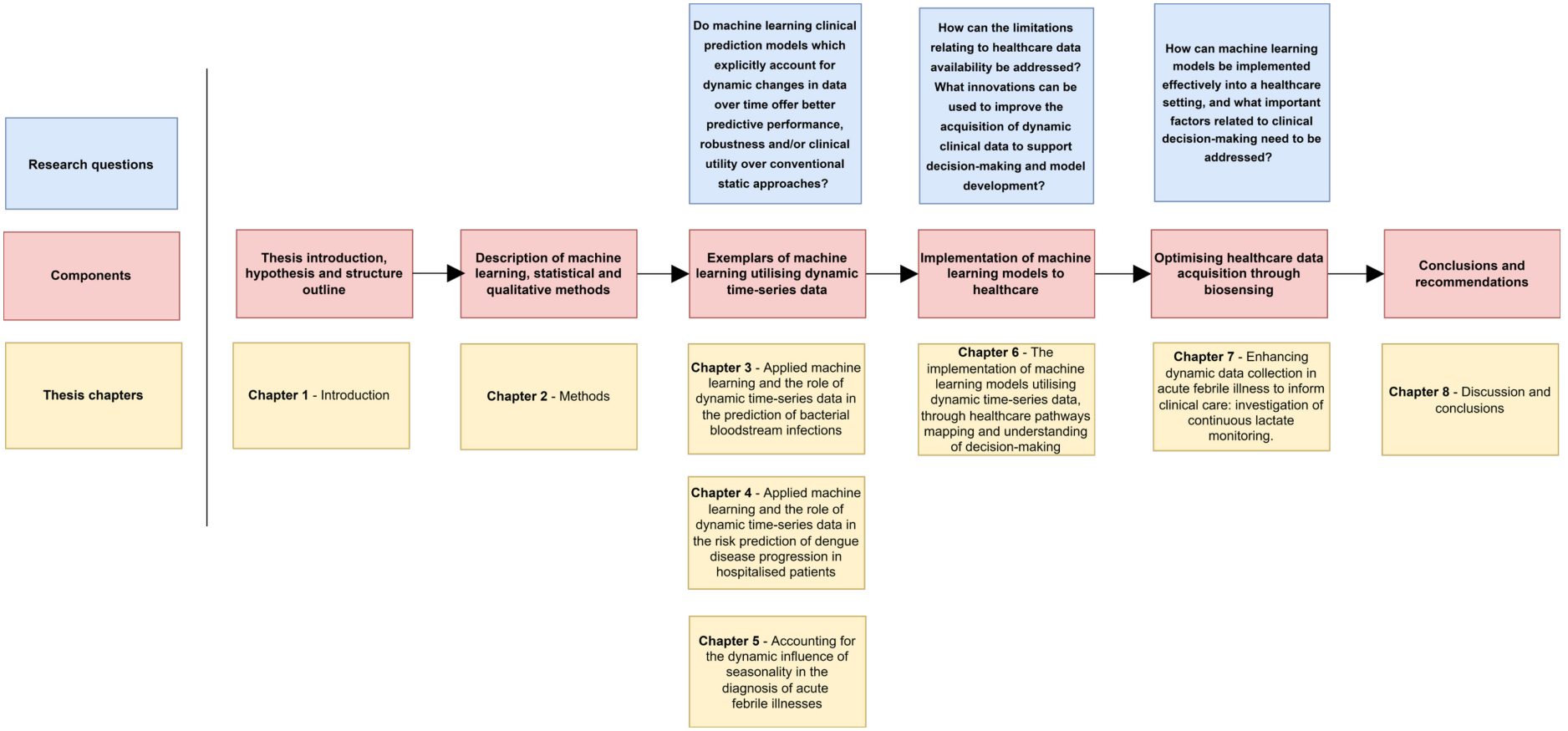
Chapter 8 provides an overview discussion of all the chapters and recommendations which arise from the findings of this thesis.

The appendix contains a list of publications arising from the thesis, correspondence with relevant bodies for permissions granted for the study and materials used for interview and workshops.

Supplementary data are presented and referenced in the main text where relevant, alongside the code appendix attachment which contains my selected workbooks with comments in the Python programming language, created during the analyses of data in this thesis.

An overview of my thesis is presented in Figure 1.1:

Figure 1.1 Thesis outline



Chapter 2 - Methods

Descriptions of common methods I have chosen and utilised throughout the thesis for the purposes of data analysis and model development are presented in this chapter. Specific details of datasets, sources and methodology relevant to each chapter are presented within their relevant subchapters.

2.1 Statistical methods and Machine Learning

All statistical and machine learning data analyses were conducted within a Python 3.7 environment (Python Software Foundation). Analyses were run as .py scripts through the terminal or in notebook format through Jupyter (Project Jupyter) and VSCode (Microsoft, Washington USA) integrated development environments. Core libraries provided the majority of standard functions for data processing and analysis. These libraries included: NumPy [41], pandas [42], Matplotlib [43], seaborn [44] and SciPy [45]. Scikit-learn implementations of supervised machine learning models were used [46] alongside the XGBoost package [47]. Keras [48] / TensorFlow [49] was used to construct deep learning models. All libraries were loaded from PyPI or Anaconda repositories.

2.1.1 Dataset handling

Datasets were imported from comma separated values (csv) or Excel formats (Microsoft, Washington USA) and loaded as DataFrames for notebook analysis. Formatting of data between sources varied; data for analysis were reformatted in either a wide or long/tidy format as a means of standardisation [50].

In the wide data format, each row of the DataFrame refers to a unique identifier or patient, and subsequent features such as observations or results added as additional columns. Each row contains a unique identifier and is not repeated subsequently. Strengths of the wide format are that it is amenable for immediate analysis and represents an intuitive method for data storage. In the long/tidy data format, each row of the DataFrame represents a unique instance or observation and therefore suitable for repeated measurements or time-series data. For the latter, each row represents unique id-timestamp observation followed by features in the subsequent columns.

For specific analyses, data were converted into numpy arrays of specific dimensions – for example, 3-dimensional data used for time-series analysis for Long short-term memory (LSTM) layers in Keras / Tensorflow.

2.1.2 Data inconsistencies, cleaning, and missing value imputation

Some clinical data required a verification step prior to analysis, termed 'data cleaning'. Exact duplicate observations in the dataset were merged, and categorical features re-encoded as necessary through label, one-hot or ordinal encoding. Categories were assigned either an integer value though label encoding method or encoded using one-hot encoding through creation of dummy variables. Certain algorithms the ordinality intrinsic to label encoding means that one-hot encoding would be a more suitable when there is no relationship assumed between features. Where ambiguity in the data existed, I utilised existing clinical judgement for data cleaning decisions.

The extent of data processing involved in this thesis was significant and involved the consolidation of large numbers of datasets which were stored in diverse formats. Notably, for chapters 3 and 4 each dataset had its own properties and limitations: databases in the form of Excel and comma separate values for each component were stored in different formats which had to be harmonised to ensure suitability for common analysis. The inclusion of free text strings required manually cleaning and re-encoding. Separate datasets required merging and inconsistencies managed on a case-by-case basis.

The majority of numerical data handled in the analyses were biomedical in nature and conformed to a population (normal) distribution. Extreme outlier values were defined as those values beyond 3 standard deviations (SD) from the mean. These values were flagged using the three-sigma rule: in which 99.7% of the population is assumed to lie within 3 SD of the mean, provided a normal distribution. The outliers were then handled on an individual basis depending on biological plausibility – values were either retained or excluded from analysis prior to any transformation step.

Missing values were either represented as 'Not a Number' (NaN) as standard in Python, or imputed using the median or mean of column-wise feature values. For certain algorithms such as tree-based classifiers, missing values can be handled as-is and imputation was not performed. All imputation procedures were derived through using values from the training set only in order to avoid data leakage. For meaningful evaluation of model performance it is important that information from the hold out or validation set is not

'leaked' back into the training set – this can occur if values for imputation are affected by those in the test set [51].

For time-series analyses, missing values were forward filled per individual patient. That is, the a 'Last observation carried forward' [52] procedure was used for subsequent timepoints. Timepoints where all values were missing were excluded from data analysis using a masking technique for deep learning models [53]. The first masking layer in the neural would then propagate this mask to all subsequent layers of the model. Other missing values for were replaced again with either median, mean or zero imputation. For zero imputation it is assumed that 0 does not represent a biologically meaningful value and would therefore not contribute to model training or predictions. However, this would not be the case for data which was not missing completely at random [54] and is a limitation of single imputation strategies used. Multiple imputation strategies were not used because of computational cost and in general the priority for model development was to prioritise analysis on complete data.

Data splitting, cross-validation and bootstrapping

Datasets were split into training and hold out test sets. The training set is the subset of data which is used to inform the parameters of the algorithm in order to create a feature representation of the data. The hold out test set represents a subset of data reserved only for model evaluation and not used at any stage of model development: this in turn allows for an independent evaluation of model performance [55].

For most analyses, splitting was done at an arbitrary training:hold out split of 80:20. A stratified process was used to ensure that important groups and outcomes were distributed equally between sets. The assumption is that performance on the hold out test set is reflective of model performance in a real-world testing. Therefore, the hold out test set should ideally be a representative, but randomised sample of the training data. For time-series based analyses splits were also performed temporally with hold out test set data being collected at later dates.

Cross-validation is a resampling method used to provide better understanding of model performance and for hyperparameter optimisation [56]. The procedure involves partitioning data into a k number of folds. The model would then be trained on k-1 folds and evaluated against the remaining unused fold of data. This procedure was repeated k times until each fold has been used once as the validation set, and the average

performance and distribution providing an indicator on overall model generalisability. A stratified k-fold procedure was also used in which groups and outcomes of interest were evenly distributed across folds.

I utilised bootstrap procedures to derive confidence intervals (CI) of models as a robust means to estimating variance [57]. Datasets were sampled with replacement to create multiple data resamples. With sufficient resampling procedures the distribution of the difference in means of each resample set approaches a normal distribution according to the central limit theorem. For each resampled dataset a model was trained and evaluated against out of bag samples (i.e., datapoints not included in the resample procedure). This provides an indicator of the variation of performance and 95% CI of the model performance, when applied to the actual population [58].

Pre-processing and feature selection

Numerical data were transformed through different methods. Standardisation of data was applied to achieve a mean of 0 and standard deviation of 1, or normalised between a specific range (e.g., between -1 and 1) to ensure that features of different ranges can be meaningfully compared with each other. This is particularly important in neural network models in which unscaled features can result in poor performances and stability i.e., the degree of perturbation to model performance in relation to changes to data input. To avoid data leakage related to data transformation, any rescaling process was fitted to the value distribution in the training set only before being applied.

An iterative approach was taken for imputation and pre-processing strategies for certain analyses in which different pre-processing strategies (e.g., mean versus median imputation, standardisation versus normalisation) formed part of the grid search process (see below).

Feature selection refers to the inclusion of feature variables for use in a model. Requirement for a large number of feature variables can limit model performance and result in model overfitting (see below). It can also increase the difficulty of implementing such a model in real world settings, given the need for more data. Conversely too few relevant features can also hinder model performance. Feature selection strategies are therefore a compromise based on data-availability and relevance – selection approaches utilised in the chapters include selection based on expert opinion and clinical plausibility, post-hoc model-based selection based on feature and permutation feature importance

[59], as well as recursive feature elimination [60]. Rationale for feature selection for models is expanded within individual chapters and in many cases were based on my own clinical judgement of plausibility.

Statistical testing

Continuous features were described using median and interquartile ranges unless normally distributed where means and standard deviations are calculated. Baseline comparisons of numerical feature distributions were done through univariate analysis, using Student's t-test and Mann-Whitney test as appropriate. Comparison of categorical variables were done using the Chi squared or Fisher's exact test. A p-value cut-off > 0.05 was taken as support for the null hypothesis. For multiple univariate comparisons, a Bonferroni correction was applied in which the significance level for rejection of the null hypothesis is changed depending on the number of testing procedures performed [61].

Baseline models – logistic regression

Multivariable logistic regression was chosen as a baseline classification model given its widespread use in medical analyses. Input features and coefficients are arranged in a linear equation to provide log odds (probabilities) of output in question [62]. Classification is done through the sigmoid activation function. Through adjustment of individual coefficient values through the maximum likelihood estimation method, the model can be made to fit the training data. Regularisation in this model was done through Least Absolute Shrinkage and Selection Operator (LASSO) whereby the least important features are removed from the model and therefore reduces the tendency to overfit [63].

2.2 Overview of machine learning algorithms used in the thesis

Definitions of machine learning are broad but involve the aim of extracting knowledge from data through constructing computer models which represent the data used for training [64]. For the purpose of the thesis the term refers to an algorithmic approach whereby a model undergoes a training process to 'learn' and provide an empirical inference based on existing data. This process of training makes use of an error function to evaluate differences between the inference and desired states, and uses this to modify model functions and improve predictive performance. Supervised, unsupervised machine

learning methods were used in the thesis although many other approaches such as reinforcement, transfer, self-supervised learning exist [65].

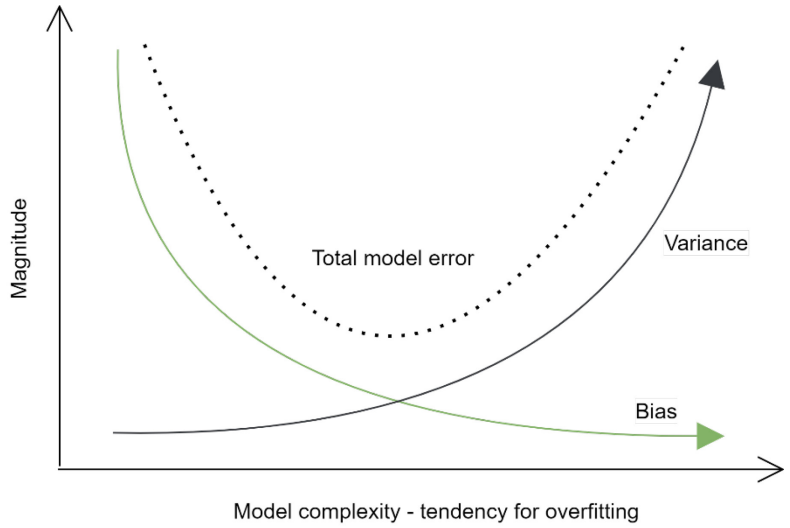
2.2.1 Supervised machine learning approaches

For supervised machine learning, the goal is the prediction of a correct label (classification), or continuous value (regression) for a given set of data. The training dataset which is provided to the model to learn from includes pre-labelled outcomes or values. Prediction involves inductive inference as an extension of the original data.

Models need to capture generalisations regarding the training data to a sufficient degree of accuracy, yet not overfit. Overfitting describes the model mapping of training data to such a high degree that it is subsequently unable to produce meaningful predictions to previously unseen data. This is a function of training data almost always being a sample of the entirety of data, and subject to problems such as noise and selection biases.

There exists a bias-variance trade-off for most predictive models: reduction in bias results in an increase in variance, and vice versa. Models which overfit have low bias (small difference between prediction and actual value by virtue of characteristics such as high model complexity), but high variance (as the variability of model predictions will be significant when applied on a different dataset). The aim of model development is therefore optimising this bias-variance trade-off through different means. This relationship is shown graphically in Figure 2.1. I had chosen the following algorithms based on their differences in mapping training data and these were as follows:

Figure 2.1 Bias-variance trade off for models. When the ability of the model to make predictions which has fewer errors on the training set because of increasing complexity, there is a higher variance in which different datasets will result in very different performances of such a model i.e., model overfitting. The theoretical optimal point is such that total model error is at its minimum.



Random Forest

The random forest model [66] is based on an ensemble of decision trees. Decision trees or Classification and Regression Trees, [67] attempt to map the data based on conditional criteria at decision nodes, providing a set of branching empirical rules to classify data. Random forests utilise bootstrap aggregation to decrease overall model variance compared with decision trees alone. The bootstrap procedure refers to data sampling with replacement (a unique sample in the data can therefore be included more than once through bootstrapping) [58]. Multiple decision trees are subsequently fitted to the bootstrap samples, aggregated and consensus of the trees used for the output of the random forest. One of the strengths of such a tree-based approach is its non-linearity and tendency to not overfit the data. In the scikit-learn implementation of the random forest [68], the number of estimators (number of decision trees), minimum samples for a node split, maximum depth of the tree and criterion to measure the quality of a split were iteratively tuned in the hyperparameter optimisation process.

Boosting approaches

Boosting is an ensemble approach to sequentially fit multiple weak models in an adaptive way to handle bias and variance and create a stronger model [69]. XGBoost, a tree-based gradient boosting model, was utilised as an efficient exemplar of this approach [47] although other boosting algorithms exist. Initial decision trees created by the algorithm

are iteratively improved by taking into account the residuals (errors). Trees are modified in different ways including regularisation, pruning, and splitting in context of missing data to improve performance. Regularisation refers to different approaches for the model to avoid overfitting through controlling the complexity of the model. Hyperparameters changed in the XGBoost algorithm include the minimum child weight, learning rate (eta), decision for node splitting (gamma), max depth of trees and number of decision trees.

Support Vector machines

Support vector machines (SVM) can be used for classification tasks through identifying optimal decision boundaries using hyperplanes in high dimensional data (i.e. data with multiple feature variables) [70]. The hyperplanes attempt to maximise the separation between data clusters in this representation in high dimensional space. These boundaries are constructed through support vectors which are adjacent data points which in turn orientate and define these hyperplanes. Maximising margins between data and hyperplane is done through a hinge loss function during training, and non-linear classification can be done through the kernel method [71].

2.2.2 Unsupervised machine learning approaches

Principal component analysis

Principal component analysis (PCA) was utilised in this thesis. The aim is given a set of data, to reduce the dimensionality by linear transformation into a new coordinate system which retains as much as the original information as possible [72]. This trade off is done through computing eigenvectors and eigenvalues of a covariance matrix to characterise the principal components which have the least correlation with each other, and therefore captures the maximum amount of data variance (i.e. information). The reduction of data dimensionality can reduce model overfitting and its projection into a 2-dimensional space allows for visualisation of data clusters.

2.2.3 Deep learning approaches

In deep learning, the emphasis is on learning a representation of data where successive layers of the model are increasingly meaningful [73]. This is done through the use of neural networks, which are an interconnected network of artificial neurons. Information fed into the network is passed often across layers through connections between artificial

neurons: the input and output layers exist either side of hidden layer(s), and ultimately providing an output. The effect of one neuron on another connected neuron is affected by the weight of the connections. In a standard feed-forward multi-layer perceptron (MLP), layers are in comprised of neurons with respective inputs and outputs. The function of the neuron depends on the weight, a bias, special functions and the intrinsic activation function.

The goal of training is to minimise the loss, or error of the network as a whole, with the labels provided in the data as a reference. The loss function refers to a metric which quantifies the magnitude of difference between the data and the output of the neural network. This can be ascertained through different methods including binary cross-entropy for classification tasks [74], or mean squared error for regression tasks.

In backpropagation the magnitude (i.e. gradient) of the loss function is used to update weights of the network accordingly to match the neural network representation with the training data. Optimisers which are used to update weights include stochastic gradient descent [75], and Adam [76], with the aim of finding optimal solutions for a given loss function. Training occurs over epochs which represent the number of passes over the entire training dataset, and the amount of data which is provided to the network before the gradient is estimated and model weights updated is defined by the batch size [77].

For binary classification models, the output is a single neuron with a sigmoid activation function, which provides a binary classification from a continuous value. Multi-task classification can be done through multiple output neurons and use of a SoftMax function. For regression tasks the numerical output of the final output layer can be converted into relevant continuous values.

A main hypothesis in the thesis is that the sequence of data and past values should be taken into account in predicting outcomes. In order to handle this time-series information, where the ordering of the data is considered, a recurrent neural network (RNN) approach was chosen to explicitly handle sequence data. The RNN allows for information which occurs earlier in the sequence to be recalled and considered during the handling of more recent values [78]. A long short-term memory (LSTM) architecture of the RNN was primarily utilised in the work, although multiple other approaches to time-series data are available [79, 80]. RNNs can experience a vanishing gradient problem as an error signal decreases significantly as backpropagation reaches the parts of the network handling data which is earlier in the sequence. The issue with a low error is that the neuron is not updated meaningfully with each training epoch (due to the small gradient) and therefore

network weights are effectively not updated or done slowly. LSTMs add to RNNs by including several gates (input, output and forget gate) in the neuron itself in order to allow it to carry forward and retain sequence data, or discard it as a means to address the vanishing gradient problem [81]. Transformer-based deep learning architectures which are adept and handling sequence data were also utilised [82] albeit this was only used for a small part of the analyses.

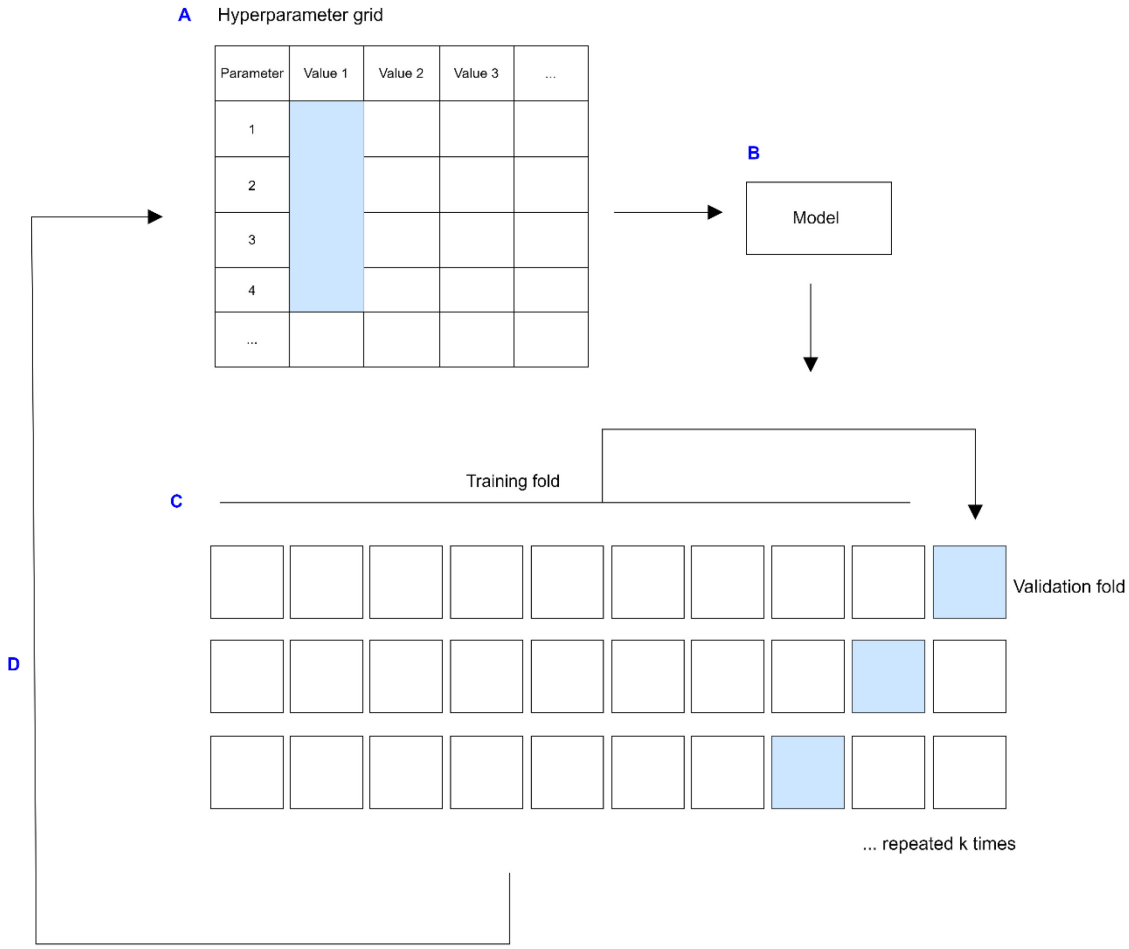
2.2.4 Model optimisation and hyperparameter tuning

The concept of model optimisation is based on configuring each algorithmic approach so that the best performance can be achieved. In turn, methods to quantify the nature of model performance and explicitly characterise the dataset which is being evaluated is crucial. Ideal models should perform consistently well on the training set, any validation or hold out test set (see below) – but more importantly – also on future encounters with hitherto unseen data. Data used to train such models should therefore be representative and possess qualities such as having sufficient variance for the model to learn and provide robust future generalisations.

For non-deep learning approaches, the tuning of model hyperparameters represents an approach to improve future model performance. Changing hyperparameters such as the number of samples utilised in the model, or the rate of learning can in turn lead to more effective model representations of the underlying data, and a better ability to infer on unseen data. Through an iterative grid search approach, where each hyperparameter variable of the model is trialled in combination, this comprehensive – but computationally expensive – method allows for the best performing set of hyperparameters to be identified [83]. Variations in this approach include randomised search and other Bayesian optimisation algorithms [84].

In order to utilise data effectively, the hyperparameter grid search approach was combined with cross-validation. In this procedure, the training data is divided into a number of folds (k), and all but one fold (k-1) is used for model training. This model is then evaluated against the remaining unused fold (termed validation set) using the specific hyperparameter combination in question. This resampling method is repeated with all folds for k iterations, and performance is therefore a generalised average across different training sets, of which each set varies with another to a degree. This procedure is shown graphically in figure 2.2:

Figure 2.2 Cross validation and grid search procedure. A) A hyperparameter grid is defined whereby each parameter is associated with a range of values to be trialled. Each combination of these hyperparameters is then used to set the initial model configuration in B). The particular model is then used in k-fold cross validation in which model trained on the training fold of the data is evaluated against the validation fold. This is repeated k times with a different validation fold each time in C). The average training score for that model configuration is recorded, and the iterative approach repeats in D) with a different hyperparameter combination until all combinations are exhausted. The best performing combination of hyperparameters is then provided as an output.



For deep learning approaches, the underlying architecture of the network itself can be optimised alongside hyperparameters. These include the number, and type of neurons in each layer, their activation function, initial bias states, learning rate, as well as the nature of data provided to the model in terms of batch size. Other aspects of the overall architecture in the neural network can also be altered including the addition of special functions which affect how information is passed between subsequent layers. The modifiable parameters in the neural network approach which were optimised for work presented in the thesis and a short description of their overall effect are summarised in Table 2.1:

Table 2.1 Overview of hyperparameters optimisation for neural networks

Attribute	Parameters	Comment
Network topology	The structure of interconnectivity in the network.	Different network structures may be more suited at handling different forms of data and extracting information. For example, a LSTM network in general handles long sequential data more effectively.
Network size	Number of neurons for each layer.	Larger networks have higher learning capacity for complex data but can be more prone to overfitting.
Loss function	Common loss functions for classification problems include binary or categorical cross-entropy metric (based on Kullback-Leibler divergence). For regression problems the deviation measured through differences between prediction and actual value put through a function is used.	Metrics used to compare model predicted values and data labels, as a quantification of model representation to the data. This value is used to inform the training process.
Learning rate	A value which determines the rate at which a model weight is updated per step / epoch. Different adaptations to the learning rate behaviour during training including scheduling, momentum or mini-batch gradient descent.	With gradient descent approaches, the aim is to empirically optimise a multi-dimensional loss function with unknown characteristics. Higher learning rates can result in model convergence sooner but may result in suboptimal solutions. Lower learning rates can be

		computationally expensive and also result in solutions being stuck in local minima.
Activation function	The function of a node which determines output according to input weights. Examples include sigmoid, Tanh and ReLU.	Activation functions have differing empirical properties which affect model performance.
Dropout	Random dropping out of nodes per layer during training.	Dropout can reduce overfitting to noise. Features in the data which are learnt through co-adaptations between layers are disrupted and therefore allow for the model to generalise better. A dropout rate which is too high however will impair the ability for the model to learn the data meaningfully.
Batch size	The number of training samples presented to the model and used to inform adjustment of the weights in model training.	Larger batch sizes tend to take less time to train as the model weights are updated less frequently per epoch. This results in larger oscillations in weights for the optimal solution in gradient descent. Smaller batches may however fit to noise and take longer to train.

For evaluation of the neural network performance - given cross-validation is computationally expensive, I had adopted a validation set split procedure as a compromise. The loss between training – validation set formed the basis of evaluation. The validation split was done in such a stratified manner that the sample reflects that of the training set. Learning curves were used to evaluate the degree of fit of the neural

network (i.e., whether the model underperformed due to insufficient capacity, or if it overfitted and therefore would lack ability to generalise on unseen data). These learning curves were used to inform subsequent modifications of the training parameters. A more formalised approach was also used through the KerasTuner package. Provided a hyperparameter grid and range of model architectures, the package iterates through each combination to evaluate performance. Subsequent hyperparameters are then chosen on the basis of a Bayesian optimisation algorithm [85] until the specified number of iterations is reached.

An early stopping approach was utilised, given that the loss on the training set invariably decreases with each epoch but validation loss may increase. This occurs as the network learns the values of the training data itself rather than the generalisable features needed for inference (i.e. overfitting). With no improvement on the validation loss after a set number of epochs, training is stopped, and model weights are reset to an earlier state where there is likely higher generalisability to unseen data.

2.2.5 Evaluation metrics of machine learning models

The aim of model evaluation is to quantify performance of a trained model on previously unseen data.

For classification tasks model, the accuracy metric reflects the proportion of predicted labels which match with their corresponding labels in the dataset. However, for heavily imbalanced datasets its utility is limited. For example, in binary classification, if the positive class constitutes 5% of outcomes, an unskilled classifier which predicts all outcomes as the negative class regardless will achieve an accuracy of 95%.

In binary classification, one of four outcomes in prediction can be achieved: true positive, true negative, false positive and false negative. Specificity (True negatives / (True negative + False positive)) and sensitivity (True positive/ (True positive+ False negative)) metrics reflect the ability of the model to detect true negative and true positive cases respectively. The sensitivity metric only considers the population who experience the positive outcome and is a reflection of the performance of the predictor against a reference standard. It is therefore the proportion of predictions correctly identified as belonging to the positive class.

Positive predictive (PPV) and negative predictive values (NPV) take into account population prevalence. PPV is defined as (True positives / (True positives + False

positives)), and NPV as (True negatives / (True negatives + False negatives)). As the background prevalence increases, the likelihood of false positive predictions decreases and therefore PPV increases accordingly. PPV is therefore the probability that those identified to be in the positive class are actually true positives [86].

The area under the receiver operating characteristic (AUROC) score is used as the main metric in classification tasks for evaluating discriminant ability of the model for binary classification. The receiver operating characteristic plot examines the trade-off between model ability to for true positive (sensitivity) predictions against false positive (1-specificity) predictions across the different decision thresholds. As the decision threshold decreases, both rates of true positive as well as false positive predictions will increase.

An ideal classifier has an AUROC of 1.0 where there are no false positives or negatives, whereas an AUROC of 0.5 indicates that the classifier performs no better than chance. Given that the decision threshold is a trade-off between true positives against true negatives, Youden's J-statistic was one approach used to determine optimal cut off points [87] to maximize performance, where J = sensitivity + specificity - 1.

For imbalanced datasets in which one class is significantly greater than the other class(es), area under the precision recall curve (AUPRC) is reported. The AUPRC plots precision (i.e. positive predictive value) against recall (sensitivity) and represents a sensitive discriminator where performance in the identification of the positive class which is usually the minority class is important. The average precision metric summarises the precision achieved at each decision threshold and provides an estimate to the AUPRC.

2.2.6 Model Interpretability

Model interpretability refers to the property of being able to better understand how, and why a model makes its predictions [88]. For complex algorithms including neural networks, complexity of operations involved in creating a model representation is akin to a black box, whereby only inputs and outputs can be characterised. Interpretability has implications for the real-world implementation of such models in healthcare in terms of trust and accountability.

For the thesis, I utilised feature and permutation importance. This post-hoc approach ranks model features in terms of their contribution to the overall prediction. Important features are those which when modified, result in a greater change to model output than

less important features. Feature importance was derived from in-built algorithm functions and Sckit-learn *permutation_importance* function was used [59].

Shapley Additive Explanations (SHAP) values provide additional information regarding feature importance including the effect of changing values on output. This is done through a model agnostic approach which allocates importance amongst all input features [89]. The additive nature of SHAP values mean that the relative contribution for each feature can be quantified and provide better understanding of the magnitude of importance. SHAP values were calculated using the *KernelExplainer* function from the SHAP package [90].

Chapter 3 – Applied machine learning and the role of dynamic time-series data in the prediction of bacterial bloodstream infections

3.1 Chapter overview and summary

This chapter presents research on the application of machine learning models which utilise dynamic time-series data to predict bacterial bloodstream infections (BSI), using only routinely collected and existing patient information.

Bacterial infections and BSI are responsible for a significant proportion of AFI presentations. BSI incur significant patient healthcare burdens and increase risk of sepsis. Clinical features can often be non-specific, and an undifferentiated may be the only clinical feature observed during early illness. Prompt recognition of individuals at risk allows for appropriate treatment and monitoring. Blood cultures remain the gold standard for the diagnosis but are associated with delays in result turnaround. Existing clinical scores available to support management do not fully make use of existing available data collected routinely, nor do they explicitly account for change in biomarker values over time. Where tools do consider the change in patient state over time, these data are handled in a limited fashion. Machine learning models which utilise clinical data for BSI prediction could provide an adjunctive approach to diagnosis.

I used data from a cohort of 4,198 unique adult and paediatric patients with BSI hospitalised at Imperial College Healthcare NHS Trust, United Kingdom between 2016 and 2021. I included existing patient blood biomarker features from routine haematology, biochemistry and ward gas point-of-care panels. I chose deep learning models and a long-short term memory (LSTM) approaches as such models are able to account for the sequence order and dynamics over time. Patients were classified according to whether they were diagnosed with a BSI, or not.

Models were developed for two purposes, to:

i) Predict patient risk of BSI for patients with a blood culture taken. I found that performances of models through a range of time-series approaches were similar, with AUROC being around 0.73 in both cross-validation and hold out set testing.

ii) Provide a daily risk assessment for BSI. I used a windowed approach in which up to 7 days of preceding data for each individual was considered in the model. These models were able to predict BSI with an AUROC of at least 0.80 and AUPRC of between 0.25 and 0.45, when evaluated against an independent hold out test set. In general, for true positive classifications, predicted risk increased over time and crossed the positive threshold 1-2 days before the date of blood culture sampling.

I carried out sensitivity and interpretability analyses to better understand how the model functioned. These results support the importance of both the sequence order of each biomarker feature as well as their absolute values in contributing to a prediction. Error analysis to understand model bias showed that the model had lower performance in specific groups of patients, for example those inpatients being managed by haematology and paediatrics. A time-invariant implementation of the prediction model using the first available biomarker value for each window was developed and resulted in significantly lower discrimination: AUROC was 0.82 (IQR 0.82-0.83); AUPRC of 0.45 vs AUROC of 0.74 (0.71-0.75); AUPRC of 0.19 for the time-series and static models respectively.

Chapter conclusion: I developed and evaluated two complementary time-series approaches with the aim of individual prediction of BSI in a UK hospital setting. Through accounting for existing blood biomarker levels and their change over time in a preceding 7-day period, I found that such dynamic risk scores had improved performance over static time-invariant models. These models could support clinical decision-making for further patient monitoring, antimicrobial treatment and may represent a cost-effective intervention for healthcare systems.

3.2 Bacterial bloodstream infections as a cause of acute febrile illness

Bacterial infections are responsible for a significant proportion of AFI presentations. Presence of viable bacteria in the bloodstream, termed bacteraemia, occurs from the dissemination of localised infection, translocation from skin and mucosa and/or introduction from the external environment into the circulation such as during invasive medical procedures. Transient presence of bacteria in the bloodstream is often asymptomatic as it is rapidly cleared in the immunocompetent host [91]. However, interplay of diverse factors including underlying immunity, bacterial virulence and inoculation dose can result in persistence of bacteria in the bloodstream and development of an inflammatory response. Bacterial bloodstream infections (BSI)[1] represent the subset of bacteraemia whereby patients develop systemic clinical and laboratory features in as a response to infection [92]. These individuals may present with an AFI presentation with, or without localising symptoms. For the purposes of the thesis the two terms 'bacteraemia' and 'BSI' will be used interchangeably.

BSI are associated with significant health burdens globally. For example, *Staphylococcus aureus* is a common and important cause of BSI, associated with a one-year mortality rate of 30.2% (95% CI 22.4-39.3%) despite appropriate therapy [93]. Overall, BSI were attributable to an estimated 2.91 million (95% CI 1.74-4.53 million) deaths in 2019 [94]. Further complications such as including meningitis, endocarditis and spondylodiscitis can result from haematogenous dissemination, and these all pose additional challenges to individual patient management. There exists a close relationship and overlap between patients with BSI and sepsis, the latter representing a clinical syndrome which is the main cause of death from infection. In sepsis, dysregulation of the host immune response from infection (by bacteria, viruses, fungi and protozoa) manifests as end-organ impairment and critical illness. Although BSI, or a positive microbiological culture is not a prerequisite for the diagnosis of sepsis - with around only half of cases having a positive microbiological diagnosis in the US healthcare setting [95] and around 39% of patients with BSI exhibiting signs of end organ dysfunction compatible with sepsis in a large paediatric cohort study [96] - infection with BSI greatly increases the risk of sepsis [97]. Patients with culture-positive and culture-negative severe sepsis exhibit significant clinical differences [98], culture status alone has not been shown to be predictive of mortality [99]. Globally there were an estimated 48.9 million (95% uncertainty interval

[1] BSI can be caused by other organisms such as fungi – for the purposes of the thesis this term is limited to represent bacterial BSI only.

38.9 – 62.9 million) cases of sepsis in 2017 resulting in around 11 million deaths [100]. This syndrome is associated with high impacts on subsequent quality of life in those patients who survived hospitalisation [101].

The prompt recognition of patients with BSI is therefore crucial in allowing early and appropriate medical management. Delayed treatment of BSI is associated with higher mortality and worse individual outcomes [102]. Given the overlap in the proportion of patients with BSI who also. have sepsis – interventions directed at improving management of BSI will have downstream benefits from a sepsis perspective [103]. The accurate identification of patient cohorts who have BSI as a cause of the AFI presentation is important also given the need for judicious antimicrobial use. The requirement for prompt, empirical broad spectrum antibiotic treatment of patients at high risk of BSI has to be balanced with the selection pressure exerted on the microbiome of the patient and downstream effects on the wider community [3].

3.2.1 Aetiology and existing laboratory diagnosis

BSI are commonly categorised by their gram stain properties under microscopy. Bacteria which numerically represent the most common causes of BSI in the United Kingdom include gram-positive bacteria: *Staphylococcus aureus*, *Streptococci* and *Enterococci* spp. and gram-negative bacteria: *Escherichia coli, Klebsiella pneumoniae, Klebsiella oxytoca and Pseudomonas* spp. [104]. Identification of the causative species is important for clinical management and provides information regarding the underlying infection source, typical antibiotic resistance patterns and patient prognosis. Different bacteria species are implicated in community-acquired versus hospital-acquired BSI, and exhibit distinct clinical and patient properties [105].

The diagnosis of BSI remains predominantly culture-based [106]. Where a clinical presentation is compatible with BSI, whole blood can be inoculated into sterile growth media and incubated at body temperature. Growth of bacteria in the media, which can be monitored through automated methods such as CO_2 production and colourimetric change can then prompt an initial gram stain examination and subculturing in selective bacteria media, enabling further speciation and antibiotic susceptibility testing in the laboratory.

Although blood cultures for the diagnosis of BSI represent the gold standard for diagnosis, a range of pre-analytical, analytical, and post-analytical factors represent barriers to real-world clinical utility. These are summarised in Table 3.1. Furthermore, these factors are

underpinned by the need for specialist microbiology laboratory infrastructure and access to these services. Developments in molecular microbiology in the last two decades have led to the introduction of diagnostics based on nucleic acid amplification which can bypass the culture step and directly utilise patient body fluid for a targeted rapid diagnosis [107–109]. False positive results as a result of cell-free pathogen (non-viable) DNA or contamination of commensal organisms from non-sterile sites and the inability to provide detailed phenotype and antimicrobial susceptibility testing information crucial for treatment optimisation in the era of antimicrobial resistance limit the role of such culture-free diagnostics in current clinical practice.

Table 3.1 Limitations to current culture-based methods for the diagnosis of bacteraemia and BSI through blood cultures

Category	Barrier	Comment
Pre-analytical	Volume and frequency of blood inoculation	Test sensitivity is dependent on the initial inoculation concentration of pathogen. For BSI the concentration of bacteria is typically between 1 to 10^4 colony forming units / mL, and therefore around 10-20 mL of blood per set is ideally required. Large blood volume sampling can be challenging in paediatric and neonatal patients.
	Sample timing	Intermittent bacteraemia can occur with infections such as abscesses [110] therefore timing of sampling affects test sensitivity.
	Prior antimicrobial use	Antimicrobial exposure prior to blood culture sampling can affect the viable concentration of bacteria and lead to reduced test sensitivity and false negatives [111].
	Aseptic technique for sampling	Inoculation of skin commensals into the blood culture medium can result in false positive results. This can pose considerable issues as early identification of such organisms is typically limited to the gram stain level only.
	Fastidious organisms	Organisms belonging to the HACEK group (*Haemophilus*, *Aggregatibacter, Cardiobacterium*, *Eikenella*, and *Kingella spp.*) and *Brucella* spp. can be slow to grow or grow poorly without specific nutrient supplementation.

	Laboratory access and transport	Delays to time to incubation after blood culture is taken and ambient temperature exposure can result in loss in bacteria viability and lower test sensitivity.
Analytical	Culture turn-around time	Culture and growth to initial positivity typically takes 12-24 hours from start of incubation. Further delays associated with subculture and species level identification mean that the test turnaround time can be up to 24-48 hours. With fastidious or slower growing organisms this process can take up to 5 days.
	Limits of detection for sampling methods	As above – false negative results inherent to the culture method can be enhanced
	Species identification versus contamination	Subculturing methods and identification of contaminated samples affects test interpretation.
	Availability of modern incubators and species identification techniques e.g., Matrix-Assisted Laser Desorption/Ionization – Time of Flight (MALDI-TOF)	Improved yields from blood cultures
Post-analytical	Reporting of results, antimicrobial stewardship	Availability of pathways to report results to clinician in timely manner, and stewardship interventions to improve use of antimicrobials and optimise treatment.

3.2.2 Clinical decision-making for bloodstream infections

The decision to undertake blood sampling for BSI is affected by the pre-test likelihood and findings from clinical assessment. Common indications for obtaining blood culture samples include presentations compatible with such infections (for example presence of fever, derangement in vital signs, biochemical or haematological markers compatible with infection). These criteria however are not standardised [112], are subject to differences in individual clinical practice, and features of infection may be attenuated in patients at extremes of age [113, 114] or those with an altered immune status. The typical yield for

blood cultures is low – reported to be around 7.5% in one large European cohort study [115].

The use of rapid screening tools for sepsis – such as the use of the quick Sequential Organ Failure Assessment (qSOFA) score have been utilised as a means to initiate the sepsis care bundle (which includes blood culture sampling) [116]. It is crucial to note that not all BSI will present with features of sepsis and end-organ impairment, and primary validation for qSOFA has been done for patient prognosis and may not be predictive of microbiological culture positivity [117]. This is reflected in the performance of qSOFA in predicting BSI with an AUROC of around 0.59 (95% CI 0.55-0.64) [118].

Clinical decision rules which formalise the prediction of BSI in high-income setting contexts have been developed for an emergency department [119] inpatient context [120] and for the elderly [121, 122] through incorporating a range of clinical, haematological and biochemical parameters. The Shapiro scoring system (Table 3.2) stratifies patient risk of BSI as low, moderate and high risk based on presence of 12 major and minor criteria on presentation in the emergency department. Discrimination based on AUROC for this score in predicting subsequent BSI (based on moderate or high scores) was 0.80 and 0.75 for the derivation and validation cohort respectively with the authors proposing its use could result in a reduction of otherwise unnecessary blood cultures. Other rule-based systems such as the use of the neutrophil to lymphocyte count ratio have also shown promising results in small patient cohorts [123]. However, for all clinical decision rule-based tools reported in the literature, inconsistencies exist as to *when* during patient illness these should be used. For example, such tools might be used at initial patient evaluation or at the point of hospital admission which might vary significantly between healthcare settings. The impact of early or late presentation of BSI on the performance of these tools is therefore unclear and hinders external validity. There are also limitations in applying these systems for patients who are already hospitalised and subsequently develop hospital-acquired BSI - which exerts a far greater mortality and morbidity.

Table 3.2 Clinical parameters for the Shapiro score to predict subsequent BSI on admission to the emergency department. Any major criteria or 2 or more minor criteria are indications for blood culture based on risk. Table adapted from [119].

Major Criteria	**Minor Criteria (1 point each)**
Suspicion of endocarditis (3 points)	Temperature 38.3–39.3°C
Temperature > 39.4°C (3 points)	Age > 65 years

Indwelling vascular catheter (2 points)	Chills
	Vomiting
	Hypotension (systolic blood pressure < 90 mm Hg
	White blood cell count > 18×10^9 cells/L
	Bands > 5%
	Platelets < 150×10^9/L
	Creatinine > 175 μmol/L

Data-driven approaches for the prediction of BSI have focused on utilising electronic medical records to include clinical, vital signs and laboratory features of hospitalised patients. Articles describing machine learning-based approaches to predict BSI onset in a hospital setting were collated through Medline using Medical Subject Headings (MeSH) and Google Scholar using terms: (bacteraemia OR "bloodstream infection" OR "BSI") AND prediction AND artificial intelligence; entries between 1996 to 2022 were included.

The majority of studies utilise a fixed preceding window of data prior to blood culture acquisition to derive aggregated features (e.g., initial value, minimum, maximum) [124–128]. These were then used to predict the outcome of a blood culture through a supervised machine learning approach, although in several studies descriptions of how the time-series nature of repeated feature measurements were not described in the text or supplementary materials provided [126, 129–131]. In two studies [132, 133] a time-series prediction approach was taken using a recurrent neural network based architecture – but limitations in the article include the unclear reporting of methods, lack of hold out test set testing beyond model development and unclear classification of outcomes. The approaches, feature variables, patient population and performances of these models are summarised in Table 3.3.

Table 3.3 Summary of data-driven machine learning approaches for the prediction of BSI in hospitalised patients

Author	Algorithm(s)	Features included in model	Approach	Population	Performance	Comment
Lee et al., 2019 [129]	Range of algorithms including Naïve Bayes, MLP, RF, SVM	20 features including: Demographics, laboratory measurements, antibiotic use, hospital stay, vital signs	Prediction of blood culture using clinical features measured within 72 hours before blood culture.	Retrospective data from 1,260 hospitalised adult patients with BSI (13,402 admitted patients) in South Korea between 2008 and 2012	AUROC 0.73 in hold out set testing of 210 patients	The handling of repeated measurements and the time-series dependency of data was not mentioned in manuscript text. Antibiotic use as a feature variable in model could be confounder.
Roimi et al., 2020 [124]	RF and gradient boosting	Demographics, medication, vital signs, Laboratory values, antibiotic treatment, presence of lines, duration of stay, clinical features including bowel frequency	Prediction of BSI using features available from ICU admission until sampling time of blood culture. Feature variables were summarised as minimum, maximum, median and $25^{th}/75^{th}$ percentiles and rate of change.	Retrospective data from 313 adult patients with BSI (1021 admitted patients) hospitalised in intensive care in US and Israel	AUROC of 0.89 in hold out set	Antibiotic use as a feature variable in model could be confounder. Requires a large number of clinical features.

			Models were developed using MIMIC III data and externally validated in another centre			
Park et al., 2020[132]	RNN	Demographic data, ICD-10 codes, operative details incl. transfusion volume, spirometry, vital signs, laboratory results	Prediction of BSI at different time steps between 8 and 24 hours prior to positive culture	Retrospective data from 720 adult patients with BSI (36,023 admitted patients) who had undergone surgery in South Korea	Model performance of AUROC 0.96 to predict BSI 8 hours prior to onset.	Details of whether model performance results apply to an independent hold out test set is unclear from manuscript text; neural network-based models are prone to overfitting. Methods to generate training / validation / test sets from the cohort are not mentioned in the text as results not reported according to TRIPOD criteria.
Bhavani et al., 2020 [125]	Gradient boosting	Demographic, vital signs, laboratory, culture history between admission and time of blood culture summarised as maximum and	Prediction of BSI positivity using prior data.	Retrospective data from 76,688 adult patients of which 19,527 blood culture days positive for BSI in a US hospital.	AURUC 0.78 in the 20% validation cohort. External validation in another	Done for patients across settings – Emergency department, wards and intensive care

		minimum where appropriate.			institution showed a AUROC of 0.73.	
Mahmoud et al., 2021 [130]	Neural network	Demographics, clinical features, antibiotic use, vital signs, SIRS, qSOFA scores	Prediction of blood culture positivity.	Retrospective data for 4,242 adults >14 years old with BSI hospitalised between 2017 and 2019 in a Saudi Arabia.	AUROC of 0.53 reported – with an accuracy of 0.82	Imbalanced problem as only 6.6% had positive cultures – leading to high accuracy but final model performed only slightly better than chance (AUROC 0.53). Unclear reporting on how the data was used in the model and whether time dependencies accounted for. Hold out or validation test set results unclear. Antibiotic use as a feature variable in model could be confounder.
Pai et al., 2021 [131]	XGBoost, SVM, MLP, RF, Logistic regression	Feature window between 7 days and 24h hours prior to the blood culture collection timeframe	Prediction of pathogenic BSI using	Retrospective data from 1,478 episodes of bloodstream infections from 4275 hospitalised patients in an	AUROC of 0.82 in hold out test set of 1,006 patients	The handling of repeated measurements and the time-series dependency of data was not mentioned in manuscript text.

		included. Thirty-two features including demographics, vital signs, laboratory data and clinical information including day of admission, APACHE II scoring, presence of central line and urinary catheter used.		intensive care unit in Taiwan between 2015 and 2019		
Boner et al., 2022 [133]	RNN and CNN	Use of 1-hourly timesteps up to 48 hours prior to blood culture sampling. In total 33 features laboratory and vital signs features included.	Prediction of positive blood cultures using a time-series approach.	Retrospective data from 3,969 patients admitted to intensive care and 4,582 blood culture episodes between 2011 and 2015.	AUROC of 0.84 on hold out test set of 655 episodes of blood cultures.	Unclear if positive blood culture outcome definition included only pathogenic results, or contaminants as well. A fixed 48-hour window of data was to predict blood culture positivity.
Lee et al., 2022 [126]	MLP and gradient boosting	Use of 34 clinical variables including age, sex, vital signs and laboratory values at 12 or 24 hours	Prediction of pathogenic BSI	Retrospective data from 38,752 episodes of BSI in adult hospitalised inpatients in Korea between 2007 and 2018	AUROC of 0.76 on hold out test set	Handling of time-series dependency of data not mentioned in manuscript text – mention of merge hour suggest that these data series were summarised.

		before blood culture acquisition				Only variables collected within last 24 hours before blood culture used in models.
Choi et al., 2022 [127]	XGBoost, RF	Demographics, clinical features and symptoms, vital signs, laboratory findings summarised as initial values, minimum and maximum	Presence of pathogenic BSI	Retrospective data from 2,654 adult patients with BSI (24,768 patients total) attending emergency room with at least 2 sets of blood cultures between 2016 and 2018 in South Korea.	AUROC of 0.72 in hold out test set in 8,571 patients	Short window period of data from emergency department used for prediction.
Goh et al., 2022 [128]	SVM, RF	Data available in the emergency department including demographics, comorbidities, vital signs at triage and laboratory data transformed as categorical variables	Prediction of pathogenic BSI	Retrospective data from 4,058 adult patients with pathogenic BSI (40,395 total) attending the emergency department study between 2015 and 2019 in Taiwan.	AUROC of 0.73	Short window period of data from emergency department used for prediction.

*MLP – multilayer perceptron, RF – random forest, SVM – support vector machine, RNN – recurrent neural network, . – convolutional neural network

3.3 Aims and Research Hypotheses

The prospective identification of patients who undergo blood culture sampling and subsequently return a positive result for pathogenic BSI is of significant clinical utility. However, existing models listed above do not address the complementary issue of missed, or delayed acquisition of blood cultures, all of which can impact patient care. Reliance on empirical antimicrobial treatment of patients with a BSI without information on antibiotic susceptibility or species identification can lead to higher rates of treatment failure, poor outcomes [134], and drive emergence of antimicrobial resistance [3].

A system which can provide risk-stratification of patients across their hospital admission in terms of BSI risk could allow for a personalised approach to infection treatment. The identification of patients at high, or low risk of BSI can be coupled with clinical pathways and enable further clinical evaluation or laboratory investigations.

The hypothesis for this chapter is that it is possible to develop a model which provides risk stratification of hospitalised patients in the UK, through the use of patient biomarker information over time. Such a model which also explicitly accounts for the time-series nature of the feature variables i.e., the dynamics and interrelationship over time might provide a more robust prediction compared with models which utilise static data. For example for prediction of positive microbiology in patients with COVID-19, there is support that both concentrations and the dynamics of C-reactive protein (CRP) and procalcitonin levels over admission are useful [23].

The specific research questions for this chapter are:

- Is it possible to predict the risk of pathogenic BSI and non-BSI states for patients admitted to hospital in the UK using blood biomarker data alone?
- Does accounting for the time-series dependencies of patient biomarkers collected over a time period in the machine learning model improve its performance?
- What patterns of biomarker information are associated with pathogenic BSI?

3.4 Statement of contribution

All work presented in this chapter is of my own.

3.5 Ethics statement

This is a retrospective analysis of de-identified, pseudonymised data from patients admitted to Imperial College Healthcare NHS Trust, UK. Ethics approval was granted under the research project 'Enhanced, Personalized, and Integrated Care for Infection Management (IRAS 204949) by London-Chelsea REC (reference 17/LO/0047)[2]. Confirmation of project approval by the Health Research Authority is included in the thesis appendix.

3.6 Methods

3.6.1 Description of dataset

The dataset used for model development and evaluation comprises of adult and paediatric patients who underwent blood culture sampling at Imperial College Healthcare NHS Trust (ICHNT), UK between 3rd March 2014 and 17th January 2021. The Trust comprises of a set of teaching hospitals located in West London providing medicine, surgery, paediatrics, obstetrics, intensive care, and other specialist services to a population of 1.5 million.

Inclusion criteria for analysis were:

- Patients who had a blood culture undertaken that flagged positive after incubation, with gram positive or negative bacteria identified on microscopy at North-West London Pathology in ICHNT.
- Isolation of bacteria from subsequent culture:
    - Pathogenic BSI were defined for the purposes of this cohort as growth of one or more of the following bacteria: *Escherichia coli, Klebsiella spp., Enterococcus spp., Pseudomonas spp., Proteus spp., Serratia spp., Citrobacter spp., Streptococcus spp., Staphylococcus aureus*.
    - Blood culture isolates classified as non-pathogenic (for example as a result of contamination) and therefore not belonging to the BSI group were defined as growth of one or more of: *Coagulase negative Staphylococcus group, Micrococcus spp., Corynebacterium spp*.
    - Growth of other bacteria species were not included in analyses

[2] https://www.hra.nhs.uk/planning-and-improving-research/application-summaries/research-summaries/enhanced-personalized-and-integrated-care-for-infection-management/

Relevant and routinely available blood biomarkers available for this patient cohort including full blood count, urea and electrolytes, liver function tests, CRP, and ward-based arterial blood gas results from the 14 days preceding the date of blood culture sampling were extracted. Demographics including age and sex were used alongside as time-invariant features.

The outcomes for each patient in the study were binary, defined as the presence or absence of a BSI for each specific date.

The date on which the blood culture was sampled, and subsequently resulted in the growth of a pathogenic bacteria on blood culture was defined as the positive (BSI) class. Preceding dates for these patients, and any dates where non-pathogenic bacteria were grown on blood culture were defined as the negative non-BSI class. For patients who underwent repeated blood cultures – the first instance of a BSI was used as the date belonging to the positive class, with all preceding dates defined as non-BSI. Instances of fungal bloodstream infections were not included in this study.

3.6.2 Feature selection

An inclusive approach was taken for feature selection given that the available biomarker data were already routinely collected as part of patient care at ICHNT. Feature selection steps were subsequently performed based on either permutation importance or SHAP values for the model. Features included in the initial models are shown in Table 3.4, and features included in the final models marked with an asterisk which I had chosen based on biological and clinical plausibility:

Table 3.4 Features included in model development, and final (35) features incorporated into models denoted by an asterisk (*).

Point of care blood gas analysis:	**Full blood count:**
- Blood pH - pO_2 - pCO_2 - HCO_3 - Ward Ionised Ca - Oxygen saturation (sO_2)* - Lactate* - Glucose* - Carboxyhaemoglobin level - Methaemoglobin level	- Haemoglobin* - White blood cell count, blood* - Platelets* - Mean cell volume, blood* - Mean cell haemoglobin concentration* - Mean cell haemoglobin level* - Red blood cell count* - Red blood cell distribution width* - Neutrophil count* - Lymphocyte count* - Monocyte count* - Eosinophil count*

	- Basophil count - Mean platelet volume, blood - Nucleated RBC count
Biochemistry: - Sodium* - Potassium* - Creatinine* - Urea* - Chloride* - Albumin* - Alkaline Phosphatase* - Alaine Transaminase* - Bilirubin* - Total Protein* - Globulin* - Point of care glucose* - Magnesium	**Inflammatory:** - C-Reactive Protein* - Procalcitonin* - D-Dimer* - Ferritin*
Coagulation: - Activated Partial Thromboplastin Clotting Time (APTT) - Prothrombin time* - Fibrinogen*	**Demographics:** - Sex* - Age*

3.6.3 Model approaches and assumptions

The emphasis of the model approaches taken in this thesis is to account for the dynamic time-series nature of patient features.

1. For prediction of blood culture outcome (BSI versus non-pathogenic), provided the condition that a blood culture has been taken – a window of data up to 7 days (days 0 to -6 inclusive) was used as features to classify blood culture outcome. The role of such a model could be suitable for application at the point of obtaining a blood culture and inform decisions such as the need for empirical antimicrobial treatment at the time of blood culture acquisition. A sub-analysis also examined the ability for such models to identify the broad category of bacteria (gram-positive, gram-negative or non-BSI). Performances were compared with an approach which summarised feature values as minimum/median/maximum.

2. In order to understand the nature of biomarker patterns associated with BSI, a window approach was utilised. Sequences of biomarkers and demographics data within fixed windows which were associated with BSI or non-BSI states using outcome definitions above were used for model training. The aim of such a model

is classify any particular series of biomarkers and predict whether or not these are associated with a BSI, or not.

The following assumptions were made for the purposes of the model:

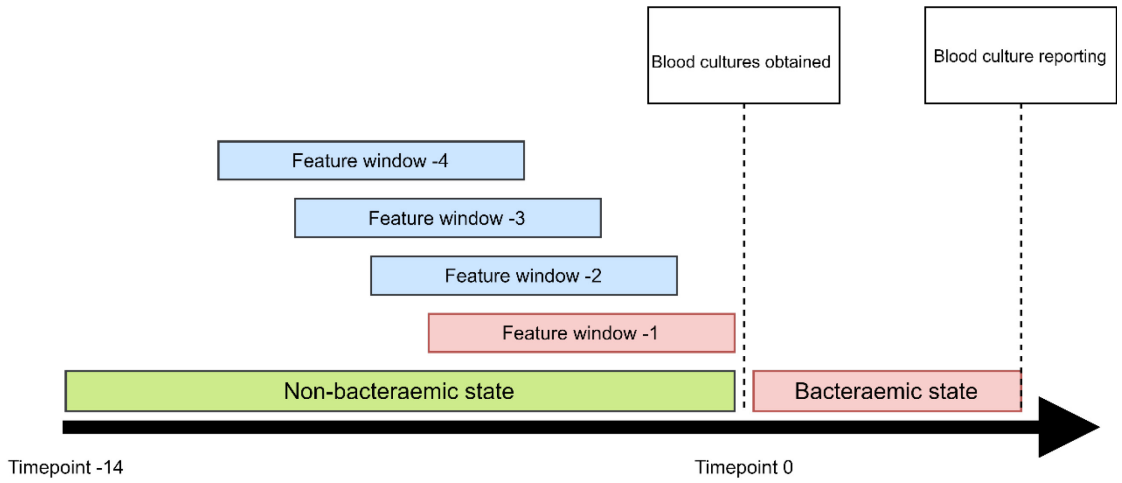
i) The bacteraemic state for the patient started at the point of blood culture sampling (timepoint 0) and did not extend earlier in time,

ii) Earlier periods where no blood cultures were drawn were representative of a non-bacteraemic state,

Organisms classified as non-pathogenic such as *Coagulase negative Staphylococci* were derived from skin flora and were considered bone fide contaminants. These would be representative of the non-BSI state and provide appropriate comparison-controls to the BSI group. These points are discussed further in the discussion and the approaches illustrated in

iii) Figure 3.1:

Figure 3.1 Approaches used by models: the passage of time for the individual patient is represented by the horizontal arrow from left to right. Blood cultures are taken at timepoint 0, and the time period leading up to it is assumed to be representative of a non-bacteraemic state. Feature windows of fixed sizes and sequence data are used to predict the outcome of the blood culture report: for prediction of blood culture outcome, only feature window -1 is used. For training of a window model, features windows spanning back until -14 (7 windows) are used, with feature window -1 denoted as BSI and the rest non-BSI



3.6.4 Data pre-processing

For pre-processing, data was formatted using a single day (24 hours) as the unit of time increment, given the majority of biomarkers are not performed more than once daily, and demographic data are fixed. For features from point of care ward blood gas tests – which may be taken several times day the mean values of repeated measurements were used. Day 0 was defined as the date of blood culture sampling.

As feature numerical data was of different units, these were either transformed and scaled so all values have a mean of 0 and a standard deviation of 1 or scaled to a range of -1 to 1 depending on performance. This was done through fitting the scaler to the training set only and then applying it to the validation and hold out test set.

Missing data was forward filled in time where possible i.e., features for any particular day will reflect the last known (and available) value, using patient results from up to 30 days prior to day 0. All remaining missing values were either masked and not considered in model training in deep learning models or left as NaN (Not a Number). Given that patients had different length data sequences, all sequences were padded to equal lengths (14 days).

Datasets were split randomly using a stratified split to training, validation and hold out test set (between 20-35% of data), with model development utilising the training and validation set only and evaluated against the hold out test set.

3.6.5 Algorithms and optimisation

Recurrent neural network (RNN)-based approaches were used for processing longitudinal, time-series sequence data. The RNN approach effective adds loops to the neural network architecture and allows information to be passed sideways (when unrolled conceptually) as well as onwards to the next layer. Past information such as earlier sequence values can then be considered and inform the evaluation of later features. Iterations of the RNN, including Long short-term memory (LSTM) and Gated recurrent unit (GRU) models address the limitations of the RNN – namely the vanishing gradient problem whereby as the sequence grows in length, reduction of gradient between the error function against the weight to a small value means that the weight of the neuron is effectively prevented from change and learning [135]. The LSTMs [81] and GRUs [136] include additional mechanisms to discard information (forget and reset gates respectively) so that longer sequences can be meaningfully processed as these layers can effectively retain relevant information downstream of the sequences.

Optimisation of the neural network architecture was done systematically using a Bayesian optimisation framework [84] through KerasTuner [85]. This approach iteratively constructs models defined through a search space in terms of topology, network size, learning rate, activation function, batch size amongst other parameters – using previous performance on validation set (minimising validation loss) to inform future parameters.

Adam (adaptive moment estimation) was used as the optimisation algorithm [137] and binary or categorical cross entropy was used as the loss function, given the classification tasks. A call back system was set up in which if the validation loss score does not decrease after a specific number of training epochs, the weights of the model which provided the best performance are restored and taken as the optimised model. Detailed descriptions of methods are provided in the methods section, and specific information regarding model topology, parameters and learning curves are listed in the supplementary results.

A comparator gradient boosting algorithm used in previous chapters, XGBoost, was used as a time-invariant approach to modelling, and grid search for hyperparameter optimisation.

3.7 Results

3.7.1 Baseline description of data

Between 8th August 2016 and 13th January 2021, there were 4,198 unique instances of BSI at ICHNT which fulfilled the inclusion criteria and were included in analysis. The median age of the cohort was 67 years old (IQR 51-79). Paediatric patients below the age of 18 years old comprised of 7% (297/4198) of the cohort – median age 2 months (IQR 10 days to 5 years old). Overall, 48% (2006/4198) of patients were female.

Bacteria species isolated from blood cultures in descending order of frequency included: Coagulase negative Staphylococci (34%), Escherichia coli (25%), Streptococci spp. (11%), Staphylococcus aureus (8%) and others (22%). Using classification scheme above – 63% (n=2642) of isolates belonged to the pathogenic BSI category and 37% (n=1556) were non-BSI. This is shown in Table 3.5:

Table 3.5 Microbiological details of blood cultures from 4,198 patients included in analysis. Bacteria species were grouped accordingly to a pathogenic BSI and non-BSI category.

Category	**Species**	**Frequency (%) n= 4,198**
Non-pathogenic	*Coagulase negative Staphylococci*	1432 (34.1%)
	Micrococcus spp.	73 (1.7%)

	Corynebacterium spp.	51 (1.2 %)
Pathogenic	*Escherichia coli*	1046 (25.0%)
	Streptococcus spp.	464 (11.1%)
	S. aureus	326 (7.8 %)
	Klebsiella spp.	249 (5.9%)
	Enterococcus spp.	192 (4.6%)
	Pseudomonas spp.	158 (3.8%)
	Enterobacter spp.	70 (1.7%)
	Proteus spp.	69 (1.6%)
	Serratia spp.	35 (0.8%)
	Citrobacter spp.	33 (0.8%)

Patients who had a BSI tended to be older (median age 53 vs 49, $p < 0.001$). The admitted locations of blood culture sampling were: medicine (54%), surgery (12%), intensive care (9%), renal (7%), haematology (4%), paediatrics (2%) and others (12%). A baseline comparison of features on day 0 at the time of blood culture, between patients with or without BSI is shown in Table 3.6, and correlation relationships between feature variables to inspect for collinearity is available in the supplementary results:

Table 3.6 A comparison of features used in the final models at the time of blood culture sampling for patients with pathogenic BSI or non-BSI. Median values are provided with interquartile ranges. A Mann-Whitney univariate comparison was done for each feature, and a Bonferroni correction for multiple testing indicates whether differences in distribution retained their statistical difference defined as p<0.05 (1 = yes, 0 = no).

Feature	Non-BSI - median (IQR)	Pathogenic BSI - median (IQR)	p-value	Bonferroni significance
Ward sO2 (%)	92.5 (78.5-96.9)	85.6 (66.5-95.5)	<0.001	1
Ward Glucose (mmol/L)	7.5 (6.0-9.7)	7.1 (5.8-9.6)	<0.001	1
Ward Lactate (mmol/L)	1.4 (1-2.1)	1.8 (1.2-2.7)	<0.001	1
Haemoglobin (g/L)	108 (90-126)	110 (94-126)	<0.001	1
White blood cell count, blood (x 10^9 cells/L)	9.5 (6.4-13.5)	11.3 (7.2-15.8)	0.986	0

Platelets (x 10^9 cells/L)	210.5 (153.0-292.5)	191.0 (126.8-262.0)	<0.001	1
Mean cell haemoglobin, blood (g/L)	325 (315-335.5)	328 (318-338)	<0.001	1
Mean cell haemoglobin level, blood (pg)	29.3 (27.3-30.9)	29.5 (27.5-31.1)	<0.001	1
Red blood cell count, blood (x 10^12/L)	3.72 (3.12-4.34)	3.76 (3.2-4.29)	<0.001	1
Red blood cell distribution width (fL)	14.5 (13.1-16.5)	14.5 (13.1-16.1)	0.153	0
Neutrophils (x 10^9 cells/L)	7.4 (4.5-10.6)	9.4 (5.8-13.5)	0.002	0
Lymphocytes (x 10^9 cells/L)	1.05 (0.7-1.6)	0.75 (0.4-1.2)	<0.001	1
Monocytes (x 10^9 cells/L)	0.7 (0.5-1.05)	0.7 (0.4-1.1)	<0.001	1
Eosinophils (x 10^9 cells/L)	0.05 (0-0.15)	0 (0-0.1)	<0.001	1
Sodium (mmol/L)	138 (135-141)	136 (133-139)	<0.001	1
Potassium (mmol/L)	4.1 (3.7-4.6)	4.1 (3.7-4.5)	<0.001	1
Creatinine (µmol/L)	76 (60-133)	94 (67-168.63)	<0.001	1
Urea level, blood (mmol/L)	6.53 (4.2-11.5)	7.9 (4.9-12.6)	0.032	0
Chloride (mmol/L)	103 (100-107)	102 (98-105)	<0.001	1
Albumin (g/L)	28 (23-33)	28 (23-33)	<0.001	1
Alkaline Phosphatase (IU/L)	102 (75-149)	109 (75-175)	0.001	0
Bilirubin (µmol/L)	11 (7-18)	15 (10-26)	<0.001	1
C-Reactive Protein (mg/L)	64.5 (22.2-137.5)	111.4 (40.6-205.6)	<0.001	1
Procalcitonin (ng/mL)	0.2 (0.07-0.51)	1.89 (0.44-6.06)	<0.001	1
D-Dimer (ng/mL)	1884 (908-3257)	3057 (1245-4412)	<0.001	1

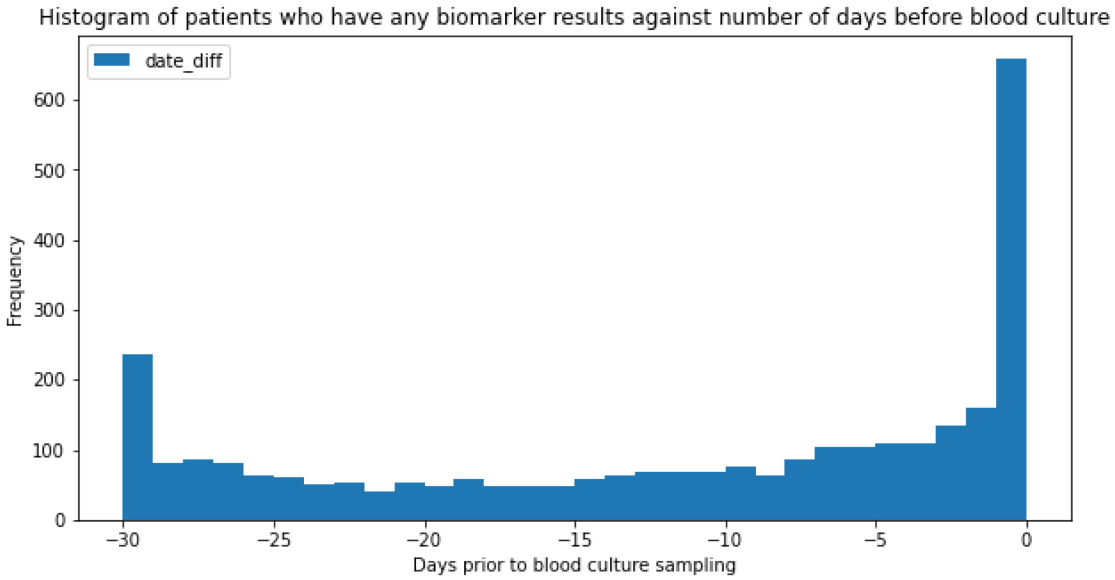
Ferritin (mcg/L)	418 (174-1077)	459 (182-1062)	0.069	0

3.7.2 Data completeness

The timeframe for model development for each patient is 14 days prior to the date of blood culture collection. Data completeness was the ratio of present data for each feature divided by the 14-day window as the denominator, without any forward filling or imputation.

Data was most complete and available for full blood count (46.9%), biochemistry (46.3%), CRP (46.3%) and liver function (45.5%) results. The median number of days prior to the date of blood culture where a patient has had a biomarker result from the feature list is 18 days (IQR 7 to 28 days) – this distribution is shown in Figure 3.2.

Figure 3.2 Histogram distribution of patients with any biomarker results from the feature list, the x-axis denotes the maximum number of days before date of blood culture sampling (timepoint 0) where the patient has had biomarker results recorded. For example, a patient with 10 days of biomarker results prior to the date of blood culture will be added to the height of the bar at -10 on the x-axis.



Data completeness for each of included feature is shown in Table 3.7:

Table 3.7 Data completeness for each of the feature variables in 4,198 patients across a 14 day window without imputation or forward filling.

Feature	Data completeness (%)	Feature	Data completeness (%)

Haemoglobin	46.9	Monocytes	45.5
Haematocrit	46.9	Eosinophils	45.3
Mean cell haemoglobin level, blood	46.9	Alkaline Phosphatase	37.9
Mean cell haemoglobin, blood	46.9	Albumin	37.8
Mean cell volume, blood	46.9	Total Protein	34.4
Red blood cell count, blood	46.9	Globulin	34.3
Red blood cell distribution width	46.9	Bilirubin	34.0
Platelets	46.7	Alanine Transaminase	32.1
Creatinine	46.6	Fibrinogen	17.5
Sodium	46.6	Prothrombin time	17.5
Chloride	46.5	Glucose POCT Strip Blood	15.8
Urea level, blood	46.5	Ward Lactate	14.6
White blood cell count, blood	46.3	Ward Glucose	14.6
C-Reactive Protein	46.3	Ward sO2	14.5
Potassium	45.7	Ferritin	3.9
Neutrophils	45.5	D-Dimer	2.7
Lymphocytes	45.5	Procalcitonin	0.5

When stratified by pathogenic BSI and non-BSI results, the amount of data completeness differed between groups. In general, data were more complete for those who subsequently cultured pathogenic bacteria likely because these patients received higher intensity medical care.

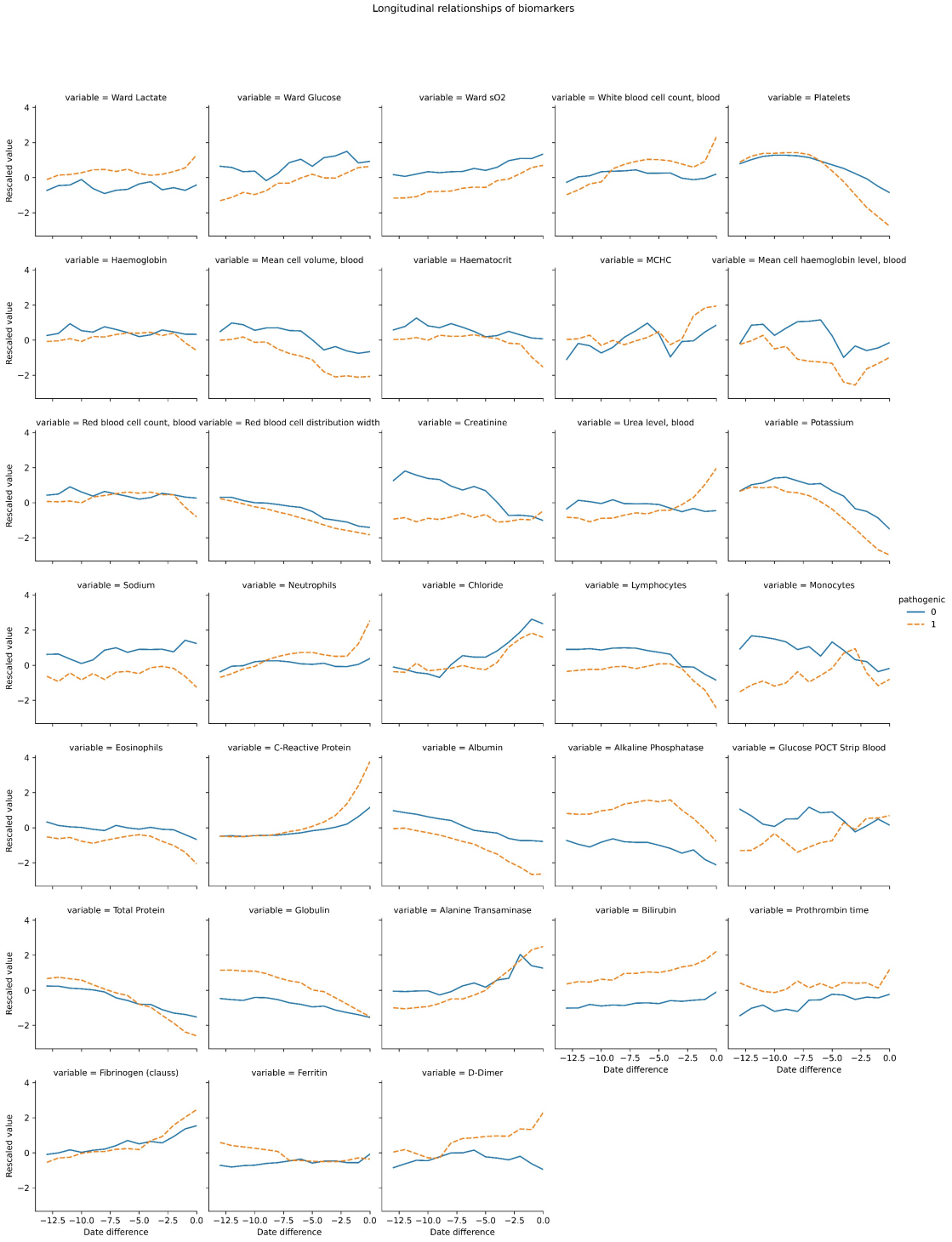
3.7.3 Descriptive data patterns over time

Temporal data trends for each feature stratified by patients with pathogenic BSI and non-BSI blood culture results were plotted. Values were forward filled by patient where

possible from data collected from up to days prior to blood culture sampling. Individual features were then rescaled to result in a mean of 0 and standard deviation of 1 for comparability.

In general trends of individual biomarkers between the pathogenic BSI and non-BSI groups followed each other. There were greater visual deviations closer to the timepoint of blood culture sampling for lactate, platelet count, haematocrit, urea. neutrophil count, CRP, and D-dimer concentrations. There were also differences in concentrations between groups for features across the window. These relationships are shown in Figure 3.3:

Figure 3.3 A visualisation of final model features and their longitudinal trends over time, stratified by pathogenic BSI (1, orange dotted) or non-BSI (0, blue continuous) blood culture results. All values have been scaled and transformed to a mean of 0 and a standard deviation of 1 for comparability. The x-axis denotes the date difference where 0 represents the date of blood culture sampling, y-axis represents the rescaled values. Values for each feature were forward filled without any imputation.



3.7.4 Prediction of bloodstream infections through time-invariant approaches

Baseline predictor models were developed and evaluated with the aim of predicting final outcome in terms of pathogenic BSI or non-BSI.

A time-invariant model was trained using XGBoost: a preceding feature window of data before the timepoint of blood culture sampling (timepoint 0) was used. Each feature was summarised across the window using the minimum, median and maximum values without imputation, and included in the model to classify the final outcomes (see Chapter 4 for similar methodology). The size of this feature window was iteratively changed to examine the effect of including features from different timepoints. i.e., feature window size 1 included only features from the 1 day preceding the timepoint of blood culture, feature size 2 included features from the 2 days preceding. Feature types included were those denoted with an asterisk, and this is shown in Table 3.4. Hyperparameter optimisation[3] was done using grid search. Data was split 80/20 between the training and hold out test set respectively, with the latter not involved in any stage of model development. The proportion of pathogenic BSI classes in the training and hold out sets were maintained through a stratified split, at approximately 35% each (this would vary to a small degree during cross-validation)

In total, all 4,198 patients were included in the analysis. Through 10-fold cross validation, the model performances were similar across feature windows sizes of between 1-14. Shorter window sizes of data, namely inclusion of 1 and 2 days only tended to result in better performance compared with longer window sizes however these differences were small. The median across all window sizes 1-14 was AUROC 0.72 (IQR 0.71 – 0.73). The median AUROC when tested on the stratified hold out set was 0.73, with a specificity of 0.69 and sensitivity of 0.69. The AURPC ranged between 0.60 to 0.73. These results are summarised in Table 3.8:

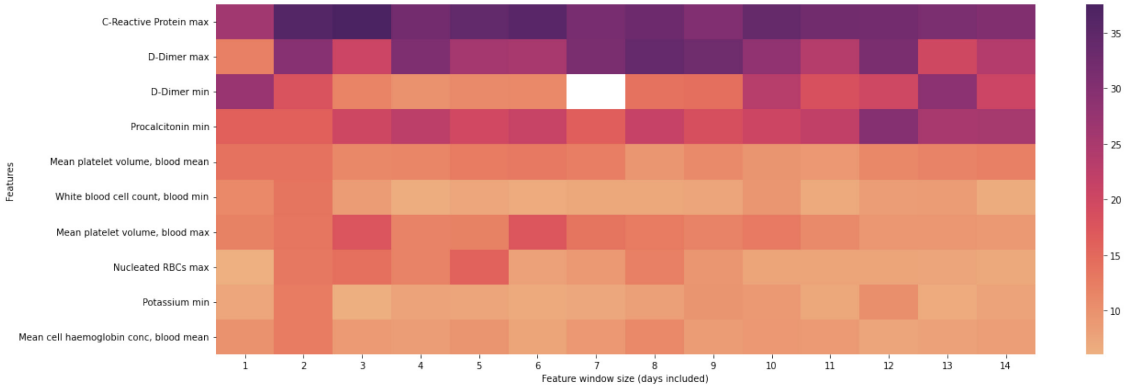
[3] Optimal parameters: subsample = 0.8, n_estimators = 100, min_child_weight = 0.25, max_depth = 3, gamma = 5, eta = 0.1

Table 3.8 Time invariant baseline model to predict outcome (presence or absence of BSI). Varying feature window sizes before timepoint 0 were included in the XGBoost model and median area under the receiver operating curves (AUROC) was derived through 10-fold cross-validation as a metric of discrimination. The model AUROCs are reported with interquartile ranges (IQR). Each version of the model was then tested against the hold out test set.

Feature window size (days included)	Median model AUROC (IQR)	Hold out test set AUROC	Specificity	Sensitivity	AUPRC
1	0.74 (0.73-0.75)	0.74	0.72	0.66	0.72
2	0.73 (0.72-0.74)	0.72	0.72	0.64	0.64
3	0.72 (0.72-0.73)	0.74	0.76	0.64	0.73
4	0.72 (0.71-0.74)	0.73	0.69	0.7	0.68
5	0.72 (0.71-0.73)	0.73	0.72	0.66	0.72
6	0.72 (0.70-0.73)	0.73	0.70	0.67	0.60
7	0.72 (0.71-0.73)	0.73	0.60	0.76	0.67
8	0.72 (0.71-0.73)	0.73	0.61	0.75	0.71
9	0.72 (0.71-0.73)	0.73	0.70	0.67	0.67
10	0.72 (0.71-0.73)	0.72	0.69	0.67	0.68
11	0.71 (0.70-0.72)	0.72	0.62	0.74	0.66
12	0.71 (0.70-0.72)	0.72	0.57	0.78	0.66
13	0.71 (0.70-0.72)	0.72	0.64	0.71	0.66
14	0.71 (0.71-0.73)	0.72	0.61	0.74	0.71

In order to gain better insight and understand the importance of each feature with varying time window, the gain for each feature, for each model was extracted. The gain in XGBoost denotes improvement in the model accuracy which is provided by the inclusion of this feature [47]. In general, across all models maximum CRP, minimum and maximum D-dimer, procalcitonin, white blood cell count represented important features taken in the model for classification. A heatmap depiction is shown in Figure 3.4:

Figure 3.4 Heatmap of top 10 important features in the XGBoost model as measured through gain. The x-axis represents the different models organised by the number of days included from the timepoint of blood culture sampling, and y-axis represents different features. Colours denote the importance of features with darker values denoting increased importance. Where features were not included in the model for that version, this is represented as white.



3.7.5 Prediction of bloodstream infections using deep learning

A complementary LSTM approach was taken for the prediction of BSI using a single preceding window of data of 7 days. Feature values as ordered sequences were provided directly to the model without summarising (minimum, median and maximums were explicitly not calculated as additional features). The 35 features included were those listed in Table 3.4 with an asterisk, rescaled to a range between -1 and 1. Missing values were forward filled and masked. Data was then reformatted as a 3-D array representing [samples, timesteps, features] for input to the LSTM network. The same training / test split at a ratio of 80/20% was done again as previous section, with the latter not used in any of the training process. Training was done until validation loss score did not decrease after 5 epochs.

A 5-fold stratified cross validation process was used to evaluate the model. The training set was further split into training (80%) and validation set (20%) for each fold, repeated 5 times. The proportion of pathogenic BSI class within each set was around 35% as

described above. The median AUROC of the models in cross validation was 0.73 (IQR 0.73-0.74) and the median AUROC when models were tested on the hold out test set was 0.73 (IQR 0.71-0.73). These results are shown in Table 3.9:

Table 3.9 Results of 5-fold cross validation of the LSTM model when trained and evaluated against a validation set. The final version of each of the optimised model was then evaluated against the hold out test set.

Characteristic	Fold	AUROC	Specificity	Specificity
Cross-validation	1	0.73	0.57	0.77
	2	0.74	0.50	0.83
	3	0.75	0.53	0.79
	4	0.71	0.24	0.91
	5	0.73	0.45	0.82
Median (IQR)		0.73 (0.73-0.74)	0.50 (0.45-0.53)	0.82 (0.79-0.83)
Hold out testing	1	0.73	0.57	0.77
	2	0.73	0.47	0.82
	3	0.74	0.55	0.77
	4	0.70	0.23	0.92
	5	0.71	0.41	0.82
Median (IQR)		0.73 (0.71-0.74)	0.47 (0.41-0.55)	0.82 (0.77-0.82)

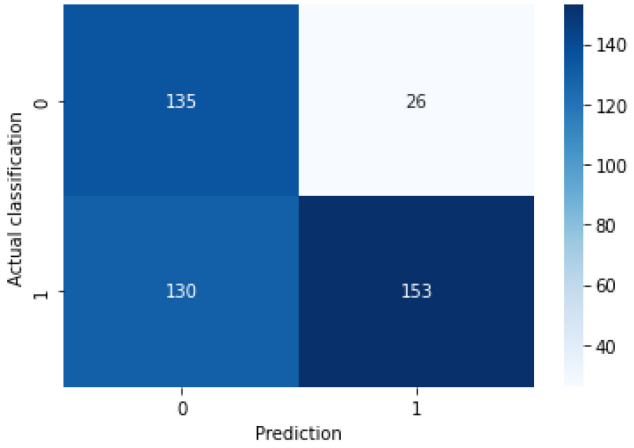
3.7.6 Predicting hospital-onset bloodstream infections

To address if the performance was affected by data availability, those patients with fewer than 7 consecutive days of any biomarker data were removed so that the training data was more complete. This resulted in the dataset consisting only of patients with hospital-onset BSI and assumed to have been a hospital inpatient for 7 or more days (in order to have undergone biomarker measurement). The dataset consisted of 2779/4198 patients (66%) and model training/ testing procedures with a LSTM model as described in the preceding subchapter were undertaken through 5-fold cross validation.

There was a small but statistically insignificant improvement in model performances: the median AUROC of the models against the validation set was 0.74 (IQR 0.74-0.77) and AUROC against hold out set 0.74, specificity 0.84, sensitivity 0.54 in predicting the

pathogenicity of the blood culture results. A confusion matrix of the hold out test set results is shown in Figure 3.5:

Figure 3.5 Confusion matrix of predictions against actual classification (1 = pathogenic BSI , 0= non-BSI) for the LSTM model applied for the hospital onset bacteraemia cohort.



3.7.7 Prediction of bacterial gram stain results

The LSTM prediction models which utilise a single preceding feature window of 7 days were used to predict the outcome of BSI in terms of their gram stain results. The empirical treatment regimens between gram-positive and gram-negative pathogenic BSI differ significantly and such an approach might allow for antibiotic stewardship measures to be implemented at the timing of blood culture sampling.

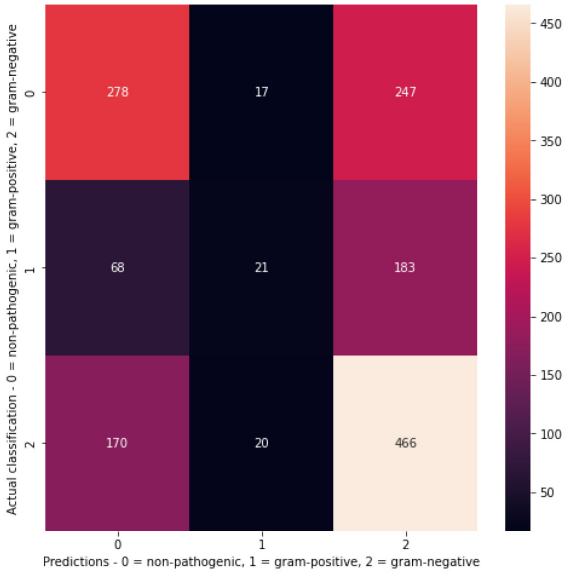
BSI results were classified accordingly to:

i) Non-BSI (*Coagulase negative Staphylococcus group, Micrococcus spp., Corynebacterium spp.*,
ii) Gram-positive (*Streptococcus spp., Staphylococcus aureus*), and
iii) Gram-negative (*Escherichia coli, Klebsiella spp., Enterococcus spp., Pseudomonas spp., Proteus spp., Serratia spp., Citrobacter spp.)*

All 4,198 patients were included – these were split to a training and a hold out test set (35%) with the latter not used for model development. Training and evaluation approach followed that used in preceding subchapters with testing of different model architectures and hyperparameters. The multi-class (3) outcomes were one-hot encoded entered into a LSTM network with a SoftMax activation function as the final layer. The model was trained against a validation set using a batch size of 16.

The results of 5-fold cross validation are as follows: the median AUROC of the models against the validation set using a 'one versus rest' comparison approach (whereby the predictions are evaluated per class as a binary approach e.g., gram-positive versus rest, gram-negative versus rest and non-BSI versus rest) was 0.69 (IQR 0.67-0.70) and performance against the hold out test set was 0.66. In general, the final model was especially poor at identifying non-BSI gram-positive bacteria subgroup although discrimination abilities for the other classes were also weak. A confusion matrix of predictions against the hold out test set is shown in Figure 3.6. When a classifier was used to distinguish between gram-positive versus gram-negative organisms per se (with the former including skin contaminants such as *Coagulase negative Staphylococci*), performance was also poor: median AUROC 0.62 (IQR 0.61-0.62).

Figure 3.6 Confusion matrix of predictions against actual classification for non-BSI (0), gram-positive (1) or gram-negative (2) bacteria in a LSTM model for the whole cohort.



3.7.8 Windowed approaches for prediction of bloodstream infections

A rolling window approach was undertaken whereby fixed window lengths (n) containing biomarker features for each patient were classified as associated with pathogenic BSI or non-BSI state. The research question was: given a particular sequence of biomarker combinations from timepoint (t) to (t-n), where t represents the date of blood culture sampling, what was the likelihood the window would be associated with subsequent diagnosis of BSI? Therefore, for a period preceding the timepoint of blood culture sampling, the model classifies each day as either high-risk (experiencing BSI) or lower

risk (not experience BSI). The model aims to provide a continuous risk stratification of patients for that day, accounting for the preceding biomarker values and trends.

The development of models followed previous sections with inclusion of asterisked biomarker features from Table 3.4. Data from the 14 days preceding blood culture sampling were included, and fixed window size of 7 days were incrementally shifted by each day until timepoint 0 was reached.

3.7.9 Deep learning for continuous prediction of bloodstream infections

A neural network-based approach was utilised with use of LSTM layers to account for the sequence order of the biomarkers. Missing data was forward filled where possible then masked and excluded from use in training.

Training and hold out test sets were split through two different approaches:

i) Random stratified split at a ratio of 80/20%, with the training set split further into a smaller validation set which was used only for model training and optimisation.

ii) Ordered split according to the date of blood culture – all blood cultures were chronologically ordered, and the training/validation set comprised of blood cultures taken between 3rd March 2014 to 20th January 2020, and the test set of blood cultures between 21st January 2020 and 17th January 2021.

All data was rescaled to a range of -1 and +1 fitting on the training set only. An iterative approach using hyperparameter tuning through Bayesian Optimisation was used whereby a range of layer size, dropout, batch normalisation, and activation functions were evaluated using the validation loss as the scoring metric. A callback function was instituted whereby if model performance on validation set did not improve over 10 consecutive epochs, then training will stop early, and optimal weights instituted.

In both cases, there were no overlap of patients between training, validation or test sets. A description of the training, validation and hold out test sets through the two approaches in splitting are shown in Table 3.10 and details of the model development and final parameters are available in the supplementary results. The dataset was split 80/20: training and validation sets consisted of data from 3,358 patients and the test set consisted of data from 840 patients.

Table 3.10 Details of the training, validation and hold out test set used for model training. Two approaches were undertaken: a random split and an ordered split according to date of blood culture. Both splits were stratified so that the pathogenic BSI/non-BSI classes were balanced as much as possible between sets.

Feature	Random split			Ordered split		
	Training	Validation	Hold out test	Training	Validation	Hold out test
Date range	3/3/2014 to 17/1/2021	1/3/2015 to 7/1/2021	2/3/2015 to 17/1/2021	9/5/2014 to 20/1/2020	3/3/2014 to 19/1/2020	21/1/2020 to 17/1/2021
Pathogenic windows n, (%)	1,708/25,152 (6.4%)	414/6,720 (6.2%)	520/8,400 (6.2%)	1,783/27,360 (6.5%)	448/6,850 (6.5%)	411/7,770 (5.3%)

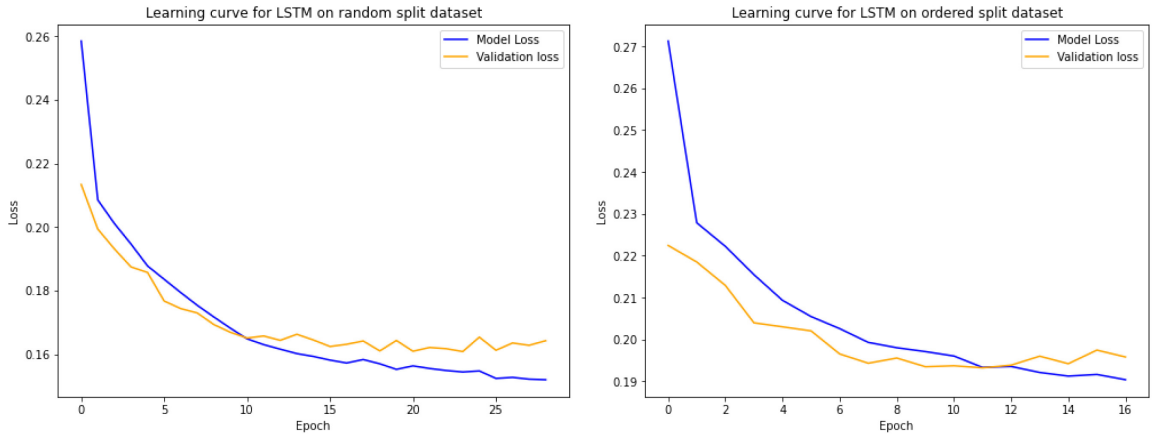
The optimal LSTM/GRU model performance were as follows:

For the random split approach, the best model AUROC against the validation set was 0.88 after 21 epochs. When applied to the hold out test set the AUROC was 0.87, specificity of 0.87, sensitivity of 0.74, positive predictive value of 0.29 and negative predictive value of 0.98. The area under the precision recall curve (AUPRC) measured through average precision was 0.45.

For the ordered split approach, the best model AUROC against the validation set was 0.80 after 16 epochs and when applied to the hold out test set the AUROC was 0.81, specificity of 0.72, sensitivity of 0.74, positive predictive value of 0.13 and negative predictive value of 0.98. The AUPRC was 0.25.

Respective learning curves for model training for both approaches are shown in Figure 3.7:

Figure 3.7 learning curves for LSTM models for the random split and ordered split datasets. Training epochs are represented on the x-axis and loss (binary cross-entropy) on the y-axis



The performance of optimal model on the random split dataset was higher compared with that on the ordered split, with the latter having a lower positive predictive value. Of note is that the period between January 2020 and 2021 was associated with hospitalised COVID-19 co-infections which can itself produce an inflammatory presentation [23] and therefore result in false positive predictions if the cohort was trained on non-COVID cases. The differences in performance between models trained on the two types of training sets and other considerations are expanded in the discussion section.

To better characterise model performance, a grouped 5-fold cross-validation approach was also undertaken whereby the entire dataset was trained on 80% of the data and evaluated against the remaining 20%, repeated 5 times. Using the optimal model settings for the random split dataset above – the overall median AUROC of the model was 0.82 (IQR 0.82-0.83) and median AUROC when evaluated against the hold out test set was 0.82 (IQR 0.82-0.82), specificity of 0.76, sensitivity of 0.74, positive predictive value of 0.19 and negative predictive value of 0.97.

3.7.10 Importance of sequence order in the deep learning models

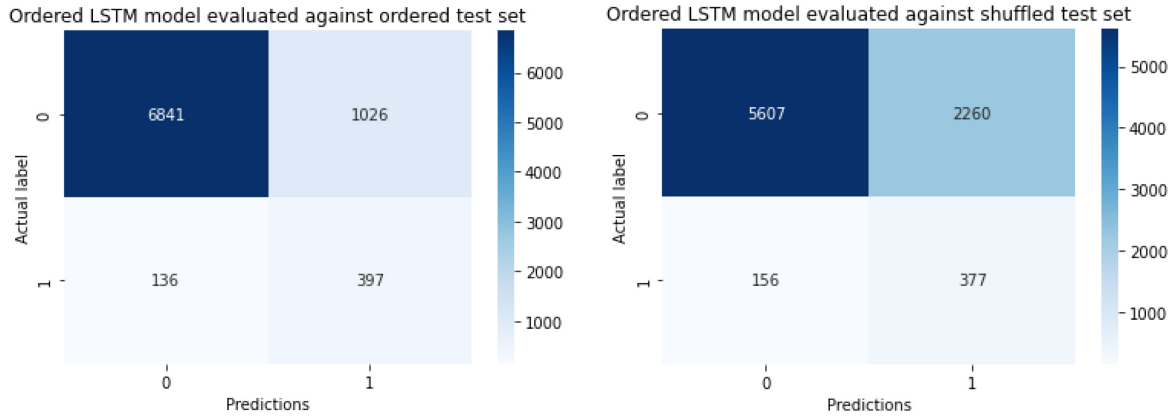
The main hypothesis throughout the thesis is that the actual biomarker sequence order is important – i.e., the time-series element within each feature window contributes to the predictive power of a model which has the capacity to take this into account.

To examine the specific impact of the ordering of biomarkers, a counterfactual approach was taken. Models trained on normal (ordered) windows of biomarkers were evaluated against a hold out test set of either normal data, or data which had been randomly shuffled. In the randomly shuffled hold out test set, the same biomarker values are available within the window, but their sequence relative to timing are randomised in each feature window. This addresses whether input of biomarker values per se would be sufficient to account for model performance, with the null hypothesis being that performance would not differ significantly between hold out test sets. All other procedures and algorithms used were done as per previous section using LSTM layers.

The performance of the model against the normal (ordered) hold out test set was as above: AUROC of 0.87, specificity of 0.87, sensitivity of 0.74, positive predictive value of 0.29 and negative predictive value of 0.98. When evaluated against the shuffled hold out test set, the AUROC was lower at 0.76, sensitivity of 0.71, specificity 0.71, positive predictive value of 0.14, and negative predictive value of 0.97. Confusion matrices of

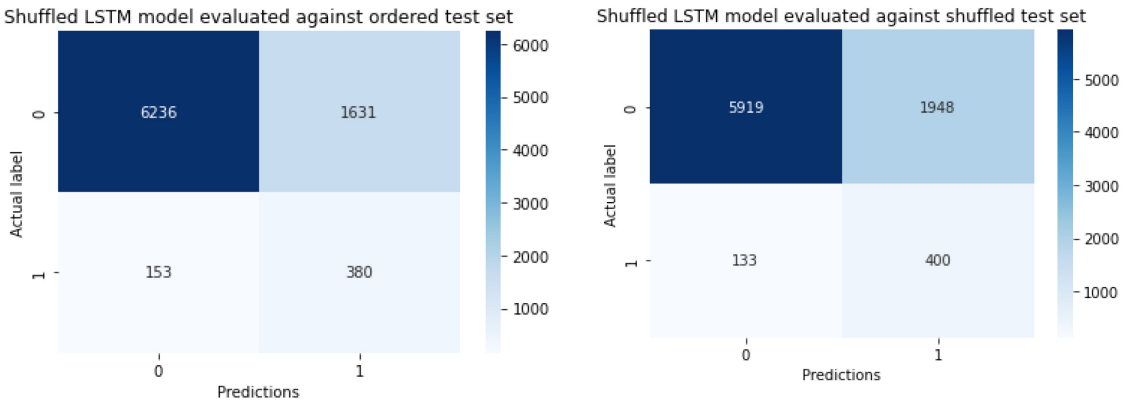
evaluations on both hold out test sets are shown in Figure 3.8, with a greater proportion of false positive predictions in the shuffled hold out test set.

Figure 3.8 Confusion matrices of the ordered LSTM model evaluated against either the ordered hold out test set (left, AUROC = 0.87) or shuffled hold out test set (right, AUROC = 0.76).



For completeness, a LSTM model with the same topology and hyperparameter settings was trained on a training set which was randomly shuffled. This model was then evaluated against either the original (ordered) hold out test set, or a shuffled hold out test set. After 24 epochs training stopped with a validation AUROC of 0.83 (loss 0.19, validation loss 0.19). When evaluated against both hold out test sets, performance was similar with AUROCs for both test sets being 0.82. The other metrics of performance against the ordered and shuffled hold out test sets were: specificity of 0.79 and 0.75, sensitivity of 0.71 and 0.75, positive predictive value of 0.19 and 0.17, negative predictive value of 0.98 and 0.98 respectively. Confusion matrices for this sensitivity analysis is shown below in Figure 3.9.

Figure 3.9 Confusion matrices of the shuffled LSTM model evaluated against either the ordered hold out test set (left), or shuffled hold out test set (right), with both AUROCs being 0.82



One interpretation is that a model trained on a shuffled training set will prioritise biomarker values and interactions over their time-series relationships – and therefore will perform with similar levels of discrimination across both ordered and shuffled test sets. These findings are expanded further in the discussion.

3.7.11 Comparator model approaches through logistic regression and time-invariant models

A direct comparison of the LSTM approach used in Chapter 6.11.1 against other methods is not possible, but insights can be gained through the use of comparator approaches for the same problem. With the use of the same biomarker feature set and data splitting method, a multivariable logistic regression model with Large-scale Bound-constrained Optimization [138] was used. This algorithm provides a linear approach to data classification in contrast with the neural network approach. Feature variables were left unscaled, and missing values were forward filled where possible and imputed with median values. The imputer fitting was applied to the training set only. Three approaches were taken, and the time-series nature of the data was accounted for with descending levels of contribution.

For a given 7-day window of values in each individual:

i) The minimum, median and maximum biomarker values were summarised and used as feature variables (see Chapter 4).
ii) The mean values only of each biomarker value were used as feature variable.
iii) The last values for each biomarker (i.e., the value collected most recently, relative to the day of prediction) were used as feature variables as a static model.

Splitting for the dataset into a training and hold out test set was done as above, and the discriminant performance of each model was evaluated through 10-fold cross validation and hold out set evaluation.

The model which included the summarised values of minimum, median and maximum values for each biomarker performed the best in cross validation and hold out set testing, followed by one which took the mean of values across the 7-day window period. The model which only used the most recent value measured showed the lowest ability to discriminate between high and low risk states. These results are summarised in Table 3.11:

Table 3.11 Results of logistic regression model through three approaches for the classification of patient states in a rolling window approach.

Approach	10-fold cross validation			Hold out test set			
	Median AUROC (IQR) in 10-fold cross validation	AUROC	AUPRC	Specificity	Sensitivity	Positive predictive value	Negative predictive value
Minimum, median and maximum of each feature over 7-day period	0.82 (0.79 – 0.83)	0.79	0.18	0.75	0.74	0.14	0.98
Mean values for each biomarker over 7-day period	0.79 (0.77 – 0.80)	0.78	0.18	0.73	0.75	0.14	0.98
Most recent values for each biomarker over 7-day period	0.74 (0.71 – 0.75)	0.76	0.19	0.77	0.66	0.14	0.98

3.7.12 Alternate model approaches

In addition to the LSTM architecture, I tried using transformer network as an alternative for handling time-series sequence data. Transformers are characterised by the ability for

self-attention, a property which allows the network to focus on specific aspects of data in a flexible fashion during training [82]. Using a 4-block model with an attention head size of either 32 and 256 for the rolling window classification, I achieved an AUROC of 0.83 and 0.84 on the test set, and a AUPRC of 0.33 and 0.34 respectively. These performances are comparable to the LSTM model. Similar performance metrics were also achieved with a convolutional neural network (CNN) architecture although I did not undergo extensive hyperparameter tuning for either transformer or CNN networks because of time constraints.

3.8 Interpretability

3.8.1 Change in model risk prediction of bloodstream infection over time

Model predictions as a probability function were used to describe how predicted risk in an individual changes with time. The pathogenic BSI/non-BSI states are treated as a binary phenomenon in the models, with the assumption that these states transition at the timepoint of blood culture sampling in the absence of other information. However, the clinical assessment of risk in practice is better conceptualised as at least semi-quantitative, or ideally quantitative, so that these properties between different patients can be compared and therefore ranked/prioritised.

Using the LSTM window model from Chapter 6.11, the individual probabilities for each feature windows were examined for a random subset of patients in the hold out test set. Patients who did not have any biomarker data beyond the 14-day timepoint prior to blood culture sampling were excluded from this sub-analysis (e.g., those patients who were directly admitted from the community and underwent blood culture sampling).

For true positive predictions – predicted risk rises over time and crosses the decision threshold a median of 2 days (IQR 1 to 4) before the date of the blood culture sampling at timepoint 0. For true negative predictions in general predicted risk remains below the decision threshold throughout all windows although this was not consistent across all patients. A random selection of individual patient predictions is shown in Figure 3.10 fand Figure 3.11:

Figure 3.10 Predictions over time for subset of patients with biomarker data and resulted in a true positive. The blue continuous lines represent the model predictions with higher values representing greater likelihood of a pathogenic state. The x-axis represents time where timepoint 0 represents date of blood culture sampling. The green dotted line represents the decision thresholds for classification.

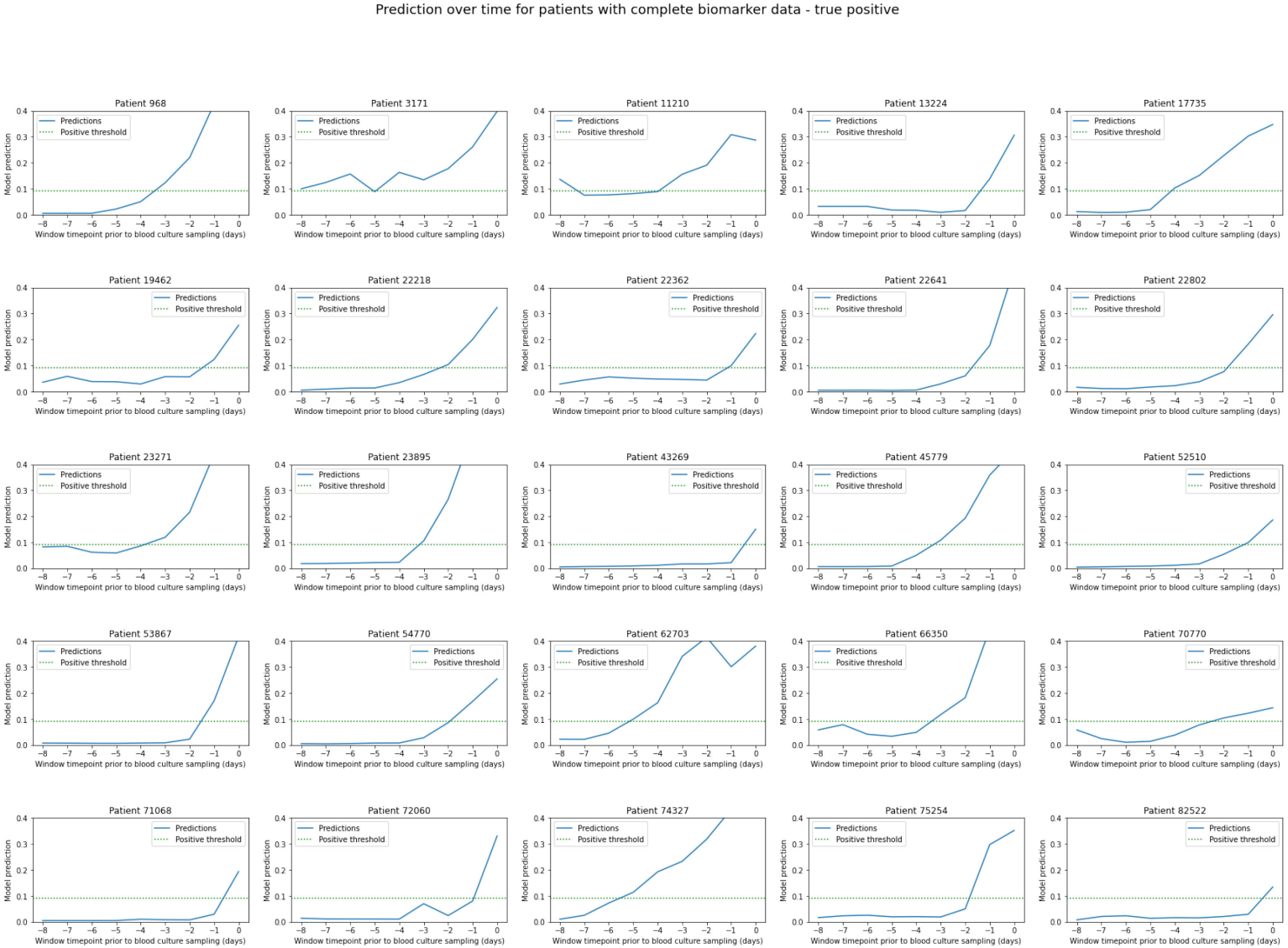
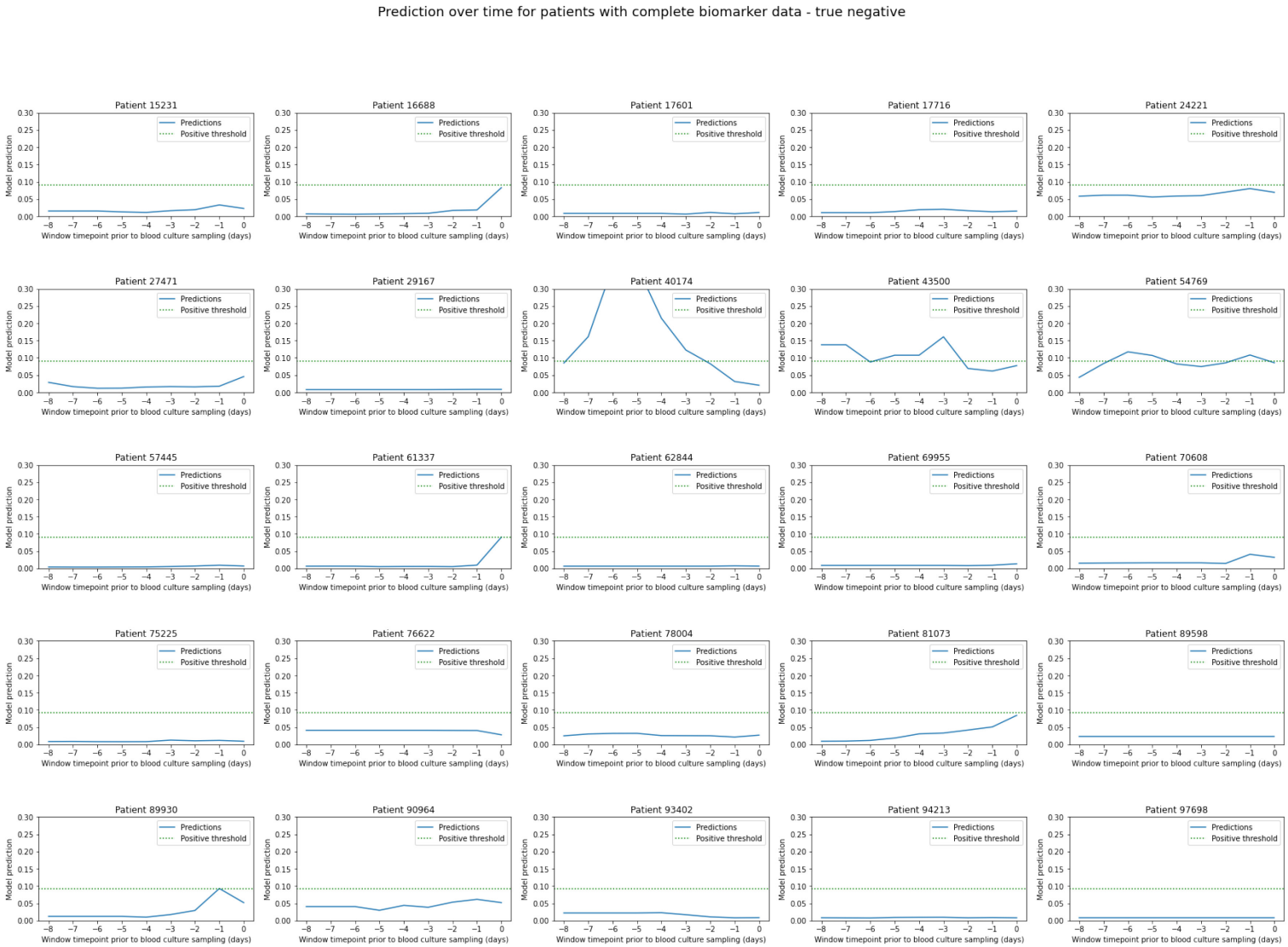


Figure 3.11 Predictions over time for subset of patients with biomarker data and resulted in a true negative. The blue continuous lines represent the model predictions with higher values representing greater likelihood of a pathogenic state. The x-axis represents time where timepoint 0 represents date of blood culture sampling. The green dotted line represents the decision thresholds for classification.



When predictions are aggregated and stratified by prediction window, the rate of false positive predictions increase as window 0 is approached. These results are shown in Table 3.12:

Table 3.12 Predictions over time stratified by window for the hold out test set with positive and negative predictions against the label. The proportion of false positive prediction increases with time up to point 0.

Window	False negative	False positive	True negative	True positive	Proportion of false positives	Sensitivity	Specificity
-7	0	100	740	0	0.12	-	0.88
-6	0	130	710	0	0.15	-	0.85
-5	0	170	670	0	0.20	-	0.80
-4	0	195	645	0	0.23	-	0.77
-3	0	253	587	0	0.30	-	0.70
-2	0	324	516	0	0.39	-	0.61
-1	0	432	408	0	0.51	-	0.49
0	121	129	178	412	0.15	0.77	0.58

One interpretation of these findings includes the possibility that the actual bacteraemic state and its associated change in physiology / biomarker derangement occurs earlier in time, prior to the date of the actual blood culture sampling. These findings are expanded in the chapter discussion.

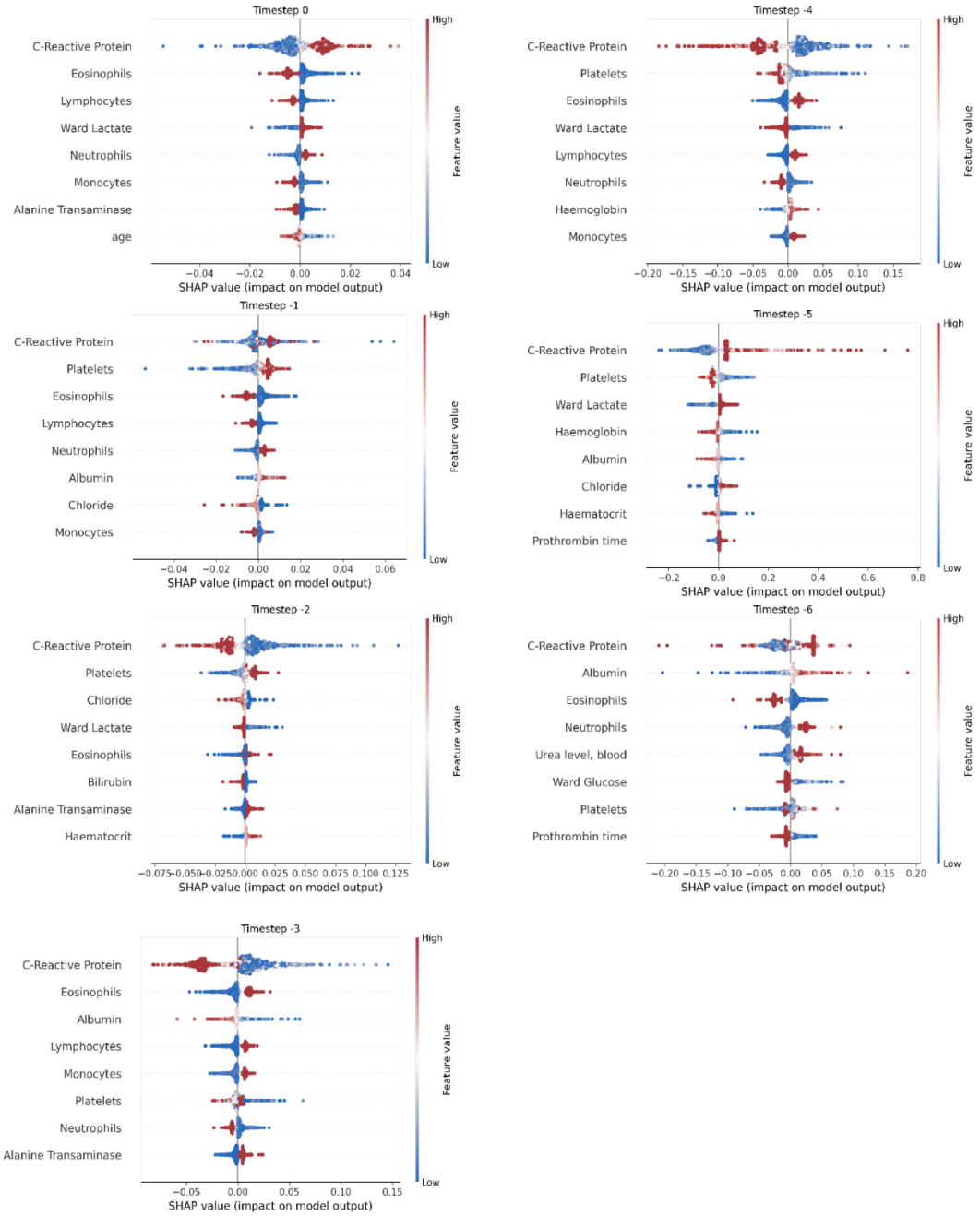
3.8.2 Interpretability analysis of deep learning models

Feature importance for the LSTM model was derived through calculating SHAP values, as a measure of contribution to the model. SHAP values for each timestep was calculated in 1,000 samples from the test set after model training. For all timesteps, the concentration of CRP provided the greatest contribution to model outcome, followed by platelet count, lactate, lymphocyte count and other features.

The magnitude and direction of contribution however was not the same in each timestep: although a higher CRP was associated with a greater contribution to predicting the pathogenic state in timestep 0, this relationship was reversed in timesteps (-2) to (-4)

whereby a lower CRP concentration was associated with the pathogenic state. This supports the hypothesis that it is the acute rise in CRP (from a nadir) which contributes to model prediction of the pathogenic state rather than an elevated CRP concentration per se. These relationships are also seen for lactate (acute rise), lymphocyte count (acute fall). Summary plots by timestep of the SHAP values are shown in Figure 3.12:

Figure 3.12 SHAP values for each timesteps from a sample of 1,000 predictions from the hold out test set. Each feature variable is ranked according to descending importance from top to bottom, and the timestep prediction is represented at the top title. A higher feature value is shown by red colouring, and a lower feature value is shown by blue. The x-axis represents the SHAP value whereby a higher positive value indicates a higher contribution to prediction of the positive class (i.e. the pathogenic state).



3.8.3 Error analysis and model bias of prediction models

Model predictions were examined with respect to patient feature variables including age, sex, type of organisms cultured and metadata proxies for clinical history in order to identify if any subgroups were subject to model bias, i.e., systematic errors or differences in performances.

The baseline LSTM window model was evaluated against the unordered hold out test set for 840 patients (7,560 windows) to predict which windows belonged to the pathogenic BSI vs non-BSI state; predictions were then stratified by each variable using AUROC as a measure of discrimination ability. To allow for comparisons between groups, a bootstrapping approach was used to estimate the 95% confidence intervals for AUROC for each subgroup [139] – this was done through resampling with replacement of label-prediction pairs 2,000 times to generate a set of bootstrapped scores with assumption of normal distribution. The 95% bounds estimated through 1.96 times the standard deviation. In some cases, estimation of AUROCs was not possible due to the lack of predictions belonging to all classes.

In general, model discrimination in terms of AUROC was higher for male sex 0.85 (IQR 0.83-0.88) compared with female sex 0.80 (IQR 0.77-0.83) but this difference did not reach statistical significance. Patient age was binned at the extremes (<1 years, 1 to <18 years, 18 to <75 years, >75 years): predictions for patients <18 years had a lower discrimination score, but this difference again was not statistically significant. AUROC however was significantly lower for *Enterococci* BSI AUROC 0.69 (IQR 0.58 to 0.79) compared with other species. Performance of the model was lower for patients with blood cultures originating under haematology and for paediatric wards (AUROC 0.74 and 0.69 respectively), although the distributions of bootstrapped AUROCs did not significantly differ with those predictions of patients in other specialties. These results are summarised in Table 3.13:

Table 3.13 Subgroups of the hold out test set and proportion of windows in brackets for the baseline unordered LSTM model and respective AUROCs for the prediction of pathogenic vs non-BSI states. The total size of windows was 7,560. Confidence intervals (CI) are displayed in brackets and derived from bootstrapping with replacement. [+] including organisms considered contaminants such as *Coagulase negative Staphylococci.* *not all species were included in the organism list as there were insufficient predictions of all classes to generate a AUROC estimation.

Feature	Variable	AUROC (95% CI)
Sex	Male (52%)	0.85 (0.83-0.88)
	Female (48%)	0.80 (0.77-0.83)
Age	< 1 year (2.3%)	0.73 (0.54-0.91)
	1 to <18 years (3.2%)	0.76 (0.50-0.96)
	18 to <75 years (35.7%)	0.84 (0.82-0.86)
	75 years or above (58.8%)	0.81 (0.77-0.85)
Gram stain result	Gram-positive[+] (55%)	0.85 (0.81 -0.88)
	Gram-negative (45%)	0.81 (0.78 – 0.84)
Organism*	*E. coli* (25.7%)	0.85 (0.82-0.87)
	Streptococcus spp. (12.1%)	0.82 (0.77-0.87)
	Enterobacter spp. (4.3%)	0.82 (0.69-0.93)
	Proteus spp. (2.4%)	0.83 (0.72-0.92)
	Klebsiella spp. (4.9%)	0.8 (0.72-0.89)
	Pseudomonas spp. (4.6%)	0.75 (0.65-0.85)
	Enterococcus spp. (4.3%)	0.69 (0.58-0.79)
	S. aureus (6.4%)	0.81 (0.74-0.88)
	Serratia spp. (0.8%)	0.74 (0.47-0.98)
	Citrobacter spp. (0.5%)	0.80 (0.35-1.00)
Specialty	Medicine (63.0%)	0.83 (0.80-0.85)
	Surgery (12.4%)	0.84 (0.79-0.89)

	Intensive care (9.4%)	0.86 (0.79-0.92)
	Haematology (5.0%)	0.74 (0.62-0.85)
	Renal (7.7%)	0.85 (0.79-0.91)
	Paediatrics (2.5%)	0.69 (0.39-0.95)

3.9 Discussion

Early prediction of BSI at an individual level is of significant clinical importance. Alone, these conditions exert a high degree of mortality and morbidity [134], but also increase the risk of sepsis which poses considerable health burdens worldwide. Patients who are stratified at high risk of BSI, especially those with clinical characteristics associated with poorer outcomes, such as neonatal patients, the elderly and immunocompromised, might in turn benefit from early empirical antimicrobial therapy and increased intensity of clinical monitoring. Conversely, identification of patients at low risk of BSI could allow for empirical treatment to be safely withheld or de-escalated, supporting better antimicrobial stewardship through precision prescribing. Current culture-based methods for BSI used in healthcare have important limitations, including need for skilled healthcare and laboratory staff, requirement for specialist infrastructure and are often associated with a results turnaround time in days. Interventions which make use of existing, routinely collected data to provide additional support to clinical management will also likely be cost-effective and support diagnostic stewardship practices.

In this chapter, I developed machine learning models which utilise routine blood biomarkers and demographics data measured as part of routine clinical care, to: i) predict blood culture result as a discrete binary outcome after blood culture acquisition and ii) continuously evaluate patient risk of BSI through time using a window-feature approach. These analyses were done using a large cohort of 4,198 patients with confirmed BSI between 2014-2021 at a UK teaching hospital, utilising up to 35 blood biomarker features across 14 days prior to the date of blood culture sampling. Findings presented show that patient risk can be predicted to a clinically actionable degree of performance. Models predicting BSI outcome at the point of blood culture sampling were able to discriminate subsequent BSI versus non-pathogenic states with a AUROC of around 0.73 (IQR 0.71-0.74). Using a separate approach involving a neural network model with inclusion of LSTM layers – which explicitly account for time-series ordering of information – an optimised model was able to discriminate between patient pathogenic BSI and non-BSI states through time with an AUROC of 0.82 (IQR 0.82-0.83), specificity of 0.76, sensitivity of 0.74.

Although the positive predictive value of such a system for BSI is relatively low at 0.19, it is comparable to the actual yield of blood cultures in healthcare settings, estimated to be approximately 7.5% in one study, [115] with the assumption that the majority of patients who undergo blood culture sampling also have a working diagnosis of BSI, and are

managed as such. The high negative predictive value of 0.97 supports its role as a reliable rule-out test which might be important in supporting diagnostic stewardship, especially in settings where constraints in healthcare staff time or diagnostic resources exist. In many settings globally, lack of infrastructure for the processing and interpretation of blood cultures are major challenges to patient care. A system which can identify individuals with low probability of BSI would be useful in prioritising resources. Results from blood cultures are also crucial for phenotypic and antimicrobial susceptibility testing for both judicious antimicrobial use, and contribute to global surveillance systems such as GLASS [140] and SEDRIC [141].

Overall performances in discriminating pathogenic BSI versus non-BSI states for models presented in this study were comparable with other published studies which utilised an explicit deep learning time-series methodology [132, 133] or use of score-based decision rules [119], but has the strength of including a larger number of patients over a long period of time, likely increasing generalisability.

Contribution of time-series ordering

The results support the role of time-series information (i.e., the ordering of the biomarker feature data) in providing additional performance in models. In a sensitivity counterfactual analysis in which the order of the feature windows was randomised, the performance of a LSTM model decreased significantly (AUROC 0.87 vs 0.76 for normal and shuffled hold out set testing respectively), suggesting that ordering was important. From SHAP interpretability analysis of these models the contribution of feature biomarkers including CRP, blood lactate and lymphocyte count significantly changes in the direction of effect depending on the feature window in question. For example, a sequence of biomarker associated with a lower nadir in days -2/-3, followed by a higher CRP concentration on day 0 (at timepoint of blood culture sampling) was associated with model prediction of pathogenic states. This is consistent with observations of whereby acute inflammatory changes induced by a bloodstream infection are observed. The importance of lactate dynamics in predicting onset of BSI in the model supports the role of continuous lactate monitoring presented in Chapter 7.

Large differences in performance were not observed when comparing performance of BSI prediction through use of only one single preceding time-series window through AUROC. Comparing a LSTM model against the time-invariant XGBoost model, which utilised features measured on the day of blood culture sampling AUROCs of 0.73 (IQR 0.73-0.74)

vs 0.74 (0.73-0.75) respectively were achieved. The LSTM model which was trained on shuffled training data also performed to a degree level of discrimination against both a normal or shuffled hold out dataset (both AUROCs 0.82). This suggests that in the absence of information on sequence order, inclusion of the feature variables alone can still provide information for prediction to a sufficient degree. This is consistent with other published studies for BSI prediction in the emergency department, where AUROCs of around 0.73 were achieved using static data alone collected from the 8 hour timeframe [127, 128].

This is further supported by findings in which a logistic regression model was applied to a 7-day window summarised through their minimum, median and maximum values. The resulting AUROC of 0.82 in cross validation was similar to that achieved by the LSTM approach (0.82). The limited sequence length of information available (i.e., up to 7 values per biomarker) might mean that a summary approach was adequate in capturing the time-series dependencies of the biomarkers. However, models which utilised data which contained less of the time-series information (through use of mean and most recent biomarker values) performed significantly poorer.

An ideal model for clinical implementation is one which retains optimal performance over time. This will be explored further in Chapter 5 in which external non-healthcare factors such as seasonality affect pre-test probabilities and need to be accounted for in prediction models. In BSI, seasonality is also exhibited as incidences of *E. coli* BSI increase in the summer in the UK [142]. When the hold out test set was split temporally there was a modest decrease in discrimination performance (AUROCs 0.88 vs 0.80; AUPRC 0.45 vs 0.25). The period of January 2020 – January 2021 was associated with high incidences of hospitalised COVID-19 in the UK, and changes associated with patient clinical characteristics (e.g., high inflammatory markers and fever as a result of viral infection [23], reduced representation of the elderly population) or altered healthcare practice (e.g., use of personal protective equipment affecting blood culture contaminants [143], high requirements for invasive ventilation in patients) could introduce complex biases to the dataset, meaning that applicability of the base model is reduced as the population is unrepresentative to the training set.

The findings also highlight the effect of heterogeneity in patient and infection subgroups. The model discrimination performance in those profoundly immunosuppressed such as in haematology settings, or in children was also reduced (albeit these were the smallest subsets in the dataset and differences were not demonstrated to be statistically significant in the analysis). An altered immune response affects biomarker features and model

predictive performance. For species such as *Enterococcus spp.* – these BSI are clinically associated with less acute symptomatology compared with gram-negative infections – predictive performance also differed. Addressing bias in these predictive models, but crucially explicitly identifying these factors in the first instance, is important in ensuring models will be able to account for the variance encountered in clinical medicine and are fair. A system of continuous learning [144], accounting for diverse patient variables, coupled with a post-hoc system of oversight [145] to ensure that model performance does not fall below the minimum level accepted might be a minimum pre-requisite for implementation. This is a non-trivial issue [146] and is a focus of multiple nascent frameworks which cover software as a medical device (SaMD) by the FDA [147], proposals by the EU to ensure trust and excellence in AI [148] and guidance on Good Machine Learning Practice endorsed by the MHRA [149].

Model risk of pathogenic BSI state in true positive predictions increased over time up as the timepoint of blood culture sampling was approached. In this group of predictions, the optimal threshold is crossed a median of 2 days before blood culture sampling, resulting in an increasing rate of false negative predictions over time. The implications of these findings are unclear, and interpretation will ultimately depend on the underlying objective-truth in the patient. These predictions might indeed reflect false positive predictions, but also that the model is able to predict the BSI state at an earlier stage when compared with blood culture sampling.

Clinically, blood culture sampling is often associated with delays, and it is possible that the biomarker patterns associated with bloodstream infection, or bacterial bloodstream invasion occurs at an earlier timepoint. Given a lack of alternative diagnostic modalities, and the relative infrequency of BSI as a presentation – accurate characterisation of these outcome labels is a challenge. This true reference problem also highlights a fundamental issue in clinical supervised machine learning, when available reference standard or data quality may not reflect objective truth in the patient, such as for cancer diagnosis [150] or where the a clinical proxy is used [151]. The goal of ensuring clinical utility, improving patient outcome and increasing the cost-effectiveness of healthcare therefore remain objective measures of success, rather than the need for accuracy of predictions per se. Carefully conducted, prospective evaluation which adhere to standardised methodologies are required [152] but examples of this type of clinical research remain limited [153].

Strengths and Limitations

The strengths of research presented in this chapter include the clinically relevant prediction task, large dataset utilised which span 7 years at a large teaching hospital in London, utilisation of robust approaches to handle time-series and the empirical approach in optimising model hyperparameters. Through providing actionable information on how individual patient risk of BSI changes over time, early clinical interventions including timely blood cultures, patient monitoring and use of early empirical antibiotic therapy. The target label of BSI also provides an objective prediction outcome which is not subject to interpretation.

A main limitation was the inclusion of only patients with a confirmed BSI into study analyses i.e., patients who underwent blood culture sampling whose cultures did not result in any growth were not included. This limitation related to data availability and the scope of the ethical approval, which only allowed use of patients with a positive blood culture. The control (comparator) population used in model development was therefore the group of patients who underwent blood culture sampling, but grew non-pathogenic bacteria – these were assumed to arise from contamination (most commonly from the skin) during the blood culture draw. It was assumed that feature characteristics of these patients with non-pathogenic bacteria will also closely resemble those with a negative blood culture result. Although *Coagulase negative Staphylococci* may be considered pathogenic in specific settings such as central line-associated bloodstream infections or in neonatal populations – these form a minority of the samples in the dataset. Furthermore, inclusion of earlier (negative) time windows in the windowed approach model served as self controls which fulfils the role for negative controls (i.e. in which a blood culture was obtained but was negative) assuming that the healthcare context would be sufficiently responsive to perform blood cultures in patients who exhibit signs of BSI. These support the prospective applicability of the model presented to the broader clinical setting locally at Imperial College NHS Trust. The evaluation of the model to the wider setting for other datasets e.g., MIMIC (Medical Information Mart for Intensive Care) was not done in this thesis given that this repository consisted of only patients who were cared for in intensive care settings in the US. In the training dataset for this study, patients in the intensive care setting comprised of a small proportion (9.4%) and given the very different clinical practices which exist between countries, applicability across settings might be limited.

This highlights the importance of models which are ultimately valid to the local setting and clinical practice but represents a significant limitation.

There was missing data, as not all patients underwent the same biomarker measurements at the same frequency. Some patients were admitted with community-acquired BSI and therefore lack the longitudinal time-series data that patients with hospital-acquired BSI possess. Missing data was forward filled, and masked in the first layer where features were still missing. It is possible that the pattern of missingness still contributed to model prediction. Imputation strategies using averages were trialled in model development but did not result in better performance.

Other model approaches which are able to take into account time-series information such as other ensemble strategies [79] were not evaluated in this work. A convolutional neural network approach was applied for prediction, but this did not result in better performances (results not shown). Given that increasing model complexity for the LSTM model resulted in early overfitting during hyperparameter optimisation, it is likely that other approaches will not significantly change overall performance. Adopting a probabilistic approach to focus on the change in patient risk, or use of unsupervised approaches to cluster patients according to outcome could be avenues for future research. The use of a multivariable logistic regression algorithm and biomarker summaries performed to a level of discrimination which is lower compared to the LSTM model. However, the use of the former requires additional specific feature engineering to achieve these results (other iterations of the logistic regression model resulted in far lower performance – data not shown), whereas the LSTM model allows for the use of an empirical framework for data input and model construction.

The sequence length of biomarkers used (up to 7 days) in the study was chosen on the balance of data availability as well as biological feasibility but might not exploit the full capabilities of neural network layers utilised. Sequences of different lengths between 3 and 14 days were tested but this did not result in significant differences in performances (see supplementary data) – ultimately the effect of including different sequence lengths on performance will depend on the biological basis of BSI. Although possible, it is less likely that the date of blood culture acquisition is significantly delayed relative to the true timing of BSI onset although the true reference problem remains.

The role of blood biomarkers in predicting BSI were the focus of investigation in the chapter. The inclusion of exogenous data including past microbiology, and laboratory data generated during blood culture incubation e.g., time to positivity [154], initial gram stain

results for the individual patient could be incorporated in refining predictions and still be of clinical utility. Addition of data of higher granularity and complexity could improve model discrimination – these include the use of vital sign measurements, clinical evaluation information or continuous signals such as photoplethysmography [155, 156]. The inclusion of clinical data in this analysis was not possible due to the constraints relating to data collection and permissions which only allowed the use of demographics and blood biomarkers. However, I speculate that the inclusion of these important features would improve the robustness of the final model. In addition, given that many inflammatory biomarkers may be well secreted intermittently (for example, interleukins and tumour necrosis factor display a diurnal variation in rheumatological conditions [157]) – measurement at higher frequency or continuously could provide additional information to risk prediction as well as treatment response, providing important substrates for personalised predictions.

One potential substrate amenable for such measurement might be lactate – measurement of lactate concentrations and their rate of change over time provide important information on prognosis in many shock and pathological states. From model interpretability analysis, the biomarker provides additional non-redundant information to the overall prediction of bloodstream infections and was ranked in the top 5 for importance (see supplementary results section for further analyses). Barriers which need to be overcome for better widespread use of the biomarker include need for (repeated) invasive blood sampling, laboratory processing as well as factors associated with measurement initiation by the clinician – these implementation obstacles will be examined and addressed in Chapter 7.

In conclusion, two complementary approaches were developed and evaluated for individual prediction of BSI in a UK hospital setting. Through accounting for blood biomarker levels and their change over time from a preceding 7-day period, a dynamic risk score associated with a clinically acceptable level of discrimination (AUROC ~0.82) can be derived. The use of existing biomarkers to augment the diagnosis of BSI through blood cultures would address limitations related to blood culture acquisition (including low rates of sampling, contamination and issues relating to timing of cultures) and likely represent a cost-effective intervention. The performance of these models would support a role for prospective evaluation and implementation in a hospital setting.

Chapter 4 – Applied machine learning and the role of dynamic time-series data in the risk prediction of dengue disease progression in hospitalised patients

4.1 Chapter overview and summary

The chapter provides the research findings on the development and the evaluation of machine learning models. Specifically understanding of how patient biomarker dynamics can be accounted for in an endemic setting with limited data availability is presented. I use the prediction of dengue shock syndrome (DSS) in hospital patients through supervised machine learning in Vietnam as an exemplar.

Existing clinical tools for dengue management, such as the World Health Organization 2009 dengue treatment guidelines, focus on decision-making prior to hospitalisation. There is a need for clinical tools which can provide better triage and risk prioritisation for hospital inpatients.

To address the research questions, I developed a set of machine learning algorithms and logistic regression with the aim of developing clinical prediction tools for hospital use in the management of dengue. I aggregated and cleaned a dataset pooled from 5 prospective clinical studies from Ho Chi Minh City, Vietnam between 2001 and 2019. Data from 4,131 adult and paediatric patients were included. I selected patient features including age, weight, sex, day of illness as well as platelet and haematocrit measurements collected over the initial 48 hours of hospital admission for the final model. The time-series nature of information was presented as aggregations of laboratory data, namely their minimum, median and maximum values over their admission.

I found that an artificial neural network model provided optimal discrimination in 10-fold cross validation (area under the receiver operating curve, AUROC of 0.83, 95% CI 0.76-0.85, AUPRC of 0.42) and in hold out test set evaluation. Models constructed using multivariate logistic regression provided an AUROC of 0.79 and AUPRC of 0.30. When I included only time-invariant data i.e., the use of haematocrit and platelet values obtained on patient admission, all models performed with significantly lower discrimination (AUROC 0.59–0.69). Interpretability was approached through the use of SHAP values. I found that dynamics of laboratory values during the initial period of patient admission was important in model performance.

Chapter conclusion: I show that machine learning models which account for short term biomarker dynamics were able to predict dengue shock and could provide a layer of safety for dengue management. This could be particularly useful during seasonal epidemics where there are increased healthcare demands and facilitate patient triage. Such models are also amenable to roll out and implementation given requirement of only few clinical features.

4.2 Global burden of dengue

Dengue virus infection is responsible for a significant healthcare burden worldwide. The single-strand RNA virus belongs to the *Flaviviridae* family and is transmitted through the bite of the *Aedes* species mosquito. Like other arthropod-borne viruses, transmission is thought to have originated through sylvatic cycles involving animal hosts [158]. However, dengue is predominantly an urban disease and the main contributory factor to its global distribution is the effectiveness of the main mosquito vector, *Aedes aegypti* and to a lesser extent *Aedes albopictus,* within a peri-domestic environment. Factors including (unplanned) urbanisation and overcrowding [159], human movement patterns [31] and seasonal factors such as rainfall and temperature [30] in the recent decades have affected the mosquito vector in terms of survival, behaviour and viral replication with downstream effects on increased disease incidence. This has facilitated the establishment of the vector in new geographical areas such as in Southern Europe and United States.

Estimates based on modelling indicate that around half of the world population is at risk from infection with dengue [160], with the majority (70%) of individuals located in Asia [40]. Around 390 million (credible interval 284-528) infections occur every year of which around 96 million [40] present symptomatically. Although the number of deaths from dengue is relatively lower (36,100 deaths worldwide in 2019) – numerically the burden exerted on patients in terms of disability adjusted life-years (DALYs) [161], resulting economic burden on patients and their families [162] and corresponding healthcare utilisation costs are significant [163].

4.2.1 Clinical features of dengue

Dengue infection manifests as an acute undifferentiated febrile illness which follows an uncomplicated course in the majority of patients. After inoculation of the virus in humans through the bite of an infected mosquito, there is an incubation period of typically 6 days (95% CI 3-10 days) [164], with a quarter of infected individuals developing symptoms [165]. Fever, headache, retro-orbital pain, myalgia, transient rash, mild bruising and skin petechiae are well-described features of acute dengue. These symptoms are commonly associated with mild thrombocytopenia, leukopenia and a degree of transaminitis [166]. The early illness phase, termed the 'febrile phase' typically lasts for around 3-4 days and is followed by defervescence and recovery. In a small proportion of patients – estimated to be up to 5% of those attending healthcare facilities – develop complications during as

they enter the 'critical phase' of disease [167]. Exact figures vary with setting and healthcare set up which is complicated by case ascertainment and occurrence of asymptomatic disease.

Increased vascular permeability, endothelial dysfunction and disruption to the glycocalyx layer [168] is associated with plasma leakage, manifesting clinically as intravascular fluid depletion and serosal effusions. The syndrome termed, 'dengue shock syndrome' (DSS) can be life-threatening and requires prompt medical treatment in the form of maintaining vital organ perfusion. A narrowing pulse pressure (< 20 mmHg) is observed in which compensatory peripheral vasoconstriction is enacted in response to gradual extravascular fluid leakage [169]. Notably, recurrent DSS can occur in the subsequent 2-3 days despite adequate fluid resuscitation in cases of prolonged vascular leakage [170]. Other manifestations of severe disease during the critical phase include haemorrhage presenting as skin and mucosal bleeding, haematemesis and melaena, as well as organ impairment including hepatitis, encephalopathy and derangements in other systems [167].

The pathogenesis which drives severe dengue is thought to be immune-mediated, including a process of antibody dependent enhancement [171]. Four dengue serotypes (DENV 1-4) exist: infection results in both a homologous antibody effective against that serotype [172] and a heterologous, cross-reactive response which provides immunity against other serotypes for up to 2 years [173]. The latter response wanes over time to a sub-neutralising concentration but instead enhances viral entry via Fc receptors in target cells. This is mediated through an opsonised uptake pathway which results in a higher infection burden and greater pro-inflammatory cytokine response [174]. Levels of pre-existing dengue antibodies in secondary infection are associated with disease severity [175]. However, other determinants including viral non-structural protein 1 (NS1) and other host genomic factors are thought to contribute to disease course but remain poorly understood [176, 177].

4.2.2 Priorities for clinical management of dengue in an endemic setting

No specific treatment for dengue is currently available. Research into antiviral and host-directed therapies have yet to demonstrate clinical benefit [178–182]. Disease treatment is therefore supportive in nature in addressing complications arising from severe disease. For patients presenting to healthcare in endemic settings, key priorities in early dengue management include accurate diagnosis and risk stratification to guide management.

Given the non-specific presentation of dengue, patients need to be differentiated from other illnesses which present with haemodynamic shock, such as sepsis. Identification of patients likely to progress to severe dengue during early illness is of significant clinical utility and allows for prioritisation of healthcare resources. Close monitoring can be provided for patients at high risk of severe dengue, whereas those at lower risk may be managed on an outpatient basis. Sufficient flexibility in the healthcare set-up is required in order to handle rapid surges in caseload such as those during seasonal epidemics.

The main clinical presentations of dengue include an undifferentiated febrile illness, or features consistent with severe dengue such as shock or bleeding which occur in later disease. Health-seeking behaviour in endemic settings affects case presentation and is turn influenced by a range of complex factors relating to disease awareness, access to services and cost [183]. Recent public health strategies have focused on encouraging early presentation in order to increase case ascertainment [184]. World Health Organization (WHO) guidelines for the treatment and diagnosis of dengue plays a central role in contemporary dengue management. These guidelines first published in 1997 and updated in 2009 have standardised approaches to dengue and are widely adopted [167]. The 2009 guidelines were developed through expert consensus and provide a classification of patients with dengue as experiencing: dengue (without warning signs), dengue with warning signs or severe dengue in order of increasing risk of complications. Severe dengue is defined as: i) clinically significant plasma leakage which may lead to shock and/or fluid accumulation (DSS), ii) severe bleeding and/or iii) severe organ impairment – replacing previous 1997 definitions of dengue fever (DF) and dengue haemorrhagic fever (DHF).

The 2009 guidelines introduced the dengue with warning signs (DWWS) category (Table 4.1). Patients who present with features consistent with dengue, but also specific warning signs, symptoms or laboratory findings are at higher risk of disease progression. Conversely the absence of these warning signs during the febrile phase indicates a reduced risk of severe dengue. The intention is that these criteria can be applied during clinical triage and is suited particularly for outpatient use [185]. The guidelines were designed to be generalisable across resource settings – with implementation requiring only basic laboratory facilities (haematocrit and platelet count measurement in blood). However heterogeneity in real-world performance between settings [186] and suboptimal specificity [187] have resulted in criticisms that use of the warning signs system for patient triage may result in an excess of unnecessary hospitalisations [187]. Certain aspects of the criteria and their wording, such as “persistent” vomiting and “rapid”

decrease in platelet count may also be subject to interpretation and result in an inconsistency in its real-world application [188].

Table 4.1 Dengue with warning signs criteria, from dengue guidelines for diagnosis, treatment, prevention and control WHO 2009

Dengue with warning signs is defined as one or more of the following features in a patient with symptom compatible with dengue:
• Abdominal pain or tenderness • Persistent vomiting • Clinical fluid accumulation • Mucosal bleed • Lethargy, restlessness • Liver enlargement >2 cm in adults • Laboratory: increase in haematocrit concurrent with rapid decrease in platelet count

4.2.3 Risk factors and the clinical prediction of dengue progression

Individual patient and disease risk factors measured during the febrile phase, which are associated with increased progression to severe dengue are shown in Table 4.2. These data were derived from a systematic review and meta-analysis performed in collaboration at Imperial College London in 2021 [189]. In total 6,643 studies were screened to identify predictors present during the febrile phase which were amenable for routine clinical measurement in different healthcare settings including in low- and middle-income settings (LMICs). Outcomes were defined according to the WHO 2009 dengue guidelines using the severe dengue definition (shock, clinically-significant bleeding and/or organ impairment).

Identified risk factors were: female sex for adults, presence of vomiting, abdominal pain and tenderness, spontaneous bleeding, clinical fluid accumulation, lower platelet count values, secondary infection and comorbidities including presence of diabetes mellitus, renal and cardiovascular disease. Although haematocrit was important as a marker of haemoconcentration during the critical phase, early measurement alone was not prognostic of severe disease.

Table 4.2 Pooled odds ratios of progression to severe dengue for risk factors measured during the febrile phase of illness using the WHO 2009 classification, table adapted from Sorawat et al. [189]

Risk factor	Number of studies	Pooled participants (n)	Odds ratio (95% confidence interval)
Female	78	45,623	1.13 (1.01-1.26)
Diabetes	12	5,852	4.38 (2.58-7.43)
Hypertension	9	4,380	2.19 (1.36-3.53)
Renal disease	6	1,786	467 (2.21-9.88)
Vomiting	9	6,229	2.25 (1.87-2.71)
Abdominal pain and tenderness	9	7,171	`1.92 (1.35-2.74)
Obesity	5	4,839	1.06 (0.90-1.24)
Clinical fluid accumulation	4	1,520	4.61 (2.29-9.26)
Secondary infection	32	23,912	2.26 (1.65-3.09)

4.2.4 Incorporating dynamic factors to augment risk prediction

Identification of associations between individual risk factors and clinical outcome are important in understanding disease – particularly aetiology – which in turn guides research and therapeutics. However, interactions between individual factors, the resulting magnitude or uncertainties of their effects need to be captured and quantified [190]. Clinical prediction models provide such a statistical framework to incorporate risk factors and produce an evidence-based output. Such systems allow the application of research findings derived from population and cohort studies to the individual, often in the clinic setting [191].

Prediction models should ideally: capture and maximise the utility of relevant data to achieve a clinically-acceptable performance, be feasible to implement for a given resource

setting or workflow, produce predictions which are relevant and valid for the local setting and future observations, and be of actual clinical utility to support decision-making.

In dengue, one key priority is prediction of severe disease or complications early in the disease course, i.e., in patients during the febrile phase. Development of the WHO 2009 dengue guidelines and severe dengue definition allows for a harmonised approach to classification of outcomes. Selected exemplars of such models which are designed for clinical implementation are shown in Table 4.3:

Table 4.3 Clinical prediction models for prediction of severe versus non-severe dengue

Reference	Model approach	Population	Variables	Prediction outcome	Reported performance
Nguyen et al. [192]	Multivariate logistic regression	2060 children in Vietnam with laboratory-confirmed dengue	History of vomiting, platelet count, aspartate aminotransferase, NS1, viraemia in the first 72 hours of illness	Severe dengue (WHO 2009)	AUROC* = 0.95, positive predictive value 10%, negative predictive value 99%
Lam et al. [193]	Multivariate logistic regression	1207 children in Vietnam presenting with dengue shock	Age, day of illness, pulse, temperature, haematocrit, haemodynamic index	"Profound dengue shock" (2 or more episodes of dengue shock overall or requirement for vasopressor therapy)	AUROC = 0.69
Lee et al. [194]	Multivariate logistic regression	1063 adults in Singapore	Age, gastrointestinal bleeding, white cell count, platelet count in the first 4 days of illness	Severe dengue (WHO 2009) and dengue haemorrhagic fever grades III and IV (WHO 1997)	AUROC = 0.859

Pang et al. [195]	Multivariate logistic regression	92 adult dengue patients in Singapore	RNA microarray data (CCL8, VPS13C and uPAR) expression in first 3 days of illness	Development of warning signs and or hospitalisation (WHO 2009)	AUROC = 0.899
Tan et al. [196]	Multivariate logistic regression – with rolling next day prediction	2285 adult patients in Singapore	Neutrophils, platelets, respiratory rate, temperature, pulse rate, eosinophils, haematocrit, monocyte count as a function of minimum and maximum values	Dengue haemorrhagic fever	AUROC = 0.72 – 0.79

*AUROC, Area under the receiver operating curve

Scoring systems displayed in the table above all utilise linear-based models, namely logistic regression to incorporate individual variables. With the exception of [196], static variables were used – measurements taken within a timeframe and treated as time-invariant. However, dengue is an inherently dynamic disease – rapid variations in symptoms, signs and laboratory values are seen in the transition between febrile and critical phase. The criteria of "increase in haematocrit concurrent with rapid decrease in platelet count" as a key warning sign is likewise explicit in the WHO 2009 dengue guidelines. Lack of granular patient clinical data in many healthcare settings where dengue is endemic is a major barrier in development and validation of models which take time-series nature of change into account.

At the time of the study, there are no published dengue models utilising machine learning approaches which can be applied during the febrile phase for prediction of clinical outcomes. The application of WHO guidelines for hospitalised patients (many of which are admitted on the basis of warning signs) is also limited in scope, and systems for stratifying risk in hospitalised cohorts are needed. Forecasting tools which utilise machine learning methodologies focus on predicting future dengue incidence using meteorological and incidence data but their application to clinical decision-making is less clear [33]. The

flexibility offered by non-linear approaches such as machine learning (through tree- or neural network based model construction) could offer distinct advantages in robustly handling complex data such as clinical data [197].

4.3 Aims and Research Hypotheses

The aim of the work presented in this chapter is the development and evaluation of clinical models for dengue which utilises:

> A range of linear and non-linear methods including use of supervised machine learning

and

> Data over time to account for changes in patient variables during early illness.

Focus is on the development of a model for use in patients hospitalised with dengue. Although criteria for hospital admission and monitoring is provided by the WHO 2009 dengue guidelines, there exists variation in practice in terms of risk stratification for patients who are admitted. Identification of patients at higher risk of deterioration is based on clinical evaluation

The hypothesis is that a system which would be able to take in the time-dependent factors of dengue would allow for better standardisation of care compared with static models.

The research question is:

> Do such machine learning models which account for patient change over time result in performances suitable for clinical implementation and which algorithmic approaches are best suited in this context?

4.4 Statement of contribution

All work presented in this chapter is my own.

4.5 Ethics statement

The analyses and use of retrospective clinical data were approved by the scientific and ethical committee of the Hospital for Tropical Diseases (HTD), Ho Chi Minh City and by the Oxford Tropical Research Ethics Committee (OxTREC) with datasets pseudonymised prior to analyses (references 145-0420 and 146-0420). Correspondence and approval letters between Oxford University Clinical Research Unit (OUCRU) on behalf of Imperial College London with the HTD are included in the thesis appendix.

4.6 Methods

4.6.1 Description of clinical datasets for model development

The clinical datasets which were included in model development and evaluation for the purposes of this chapter include data derived from prospective clinical studies conducted by OUCRU in Ho Chi Minh City, Vietnam between 12th April 2001 and 30th January 2018. The studies contain both paediatric and adult patients – consisting of 5 individual studies in total. This included 4 prospective observational studies and 1 randomised control trial (ISRCTN39575233). Broadly, patients were eligible for inclusion if they presented with an undifferentiated illness which was compatible with dengue on presentation to healthcare facilities. A summary of studies is shown in Table 4.4 and description of the clinical studies their inclusion criteria and measured outcomes are provided in the supplementary data:

Table 4.4 Studies (S1-5) included in analysis and corresponding references

No.	Study identifier code	Recruitment dates	Enrolled patients (n)	Title of study
1	S1[198]	12/4/2001-24/7/2009	3,042	Inpatient-based study examining prognostic factors during the febrile phase
3	S2[199]	3/8/2009 – 8/12/2010	225	Randomised control trial comparing use of steroids versus placebo in acute dengue
4	S3*	08/12/2010 – 16/6/2011	88	Prospective study in acute dengue

5	S4[192, 200]	19/10/2010 – 4/12/2014	8,100	Outpatient-based study examining prognostic factors during the febrile phase
6	S5*	20/10/2016 – 30/1/2018	665	Study of acute dengue in pregnancy during the febrile phase

4.6.2 Dataset pre-processing approaches

Datasets from individual studies were obtained in .xlsx format (Microsoft Corporation, USA) and pre-processed individually. Data formats used within each study were bespoke with a combination of tidy format (identifier-timestamp-variable), wide formats (identifier-timestamp-multiple variables) or both. Data from individual studies were concatenated into a single working dataset consisting of 12,119 unique patients in a tidy format.

Errors in data entry were corrected per study for inconsistencies, outlier values and duplicate entries. In order to standardise datasets for analysis the following categories were extracted as they constitute standardised observations common across all studies (note not all extracted data were included in the final models):

1) Baseline data (patient identifier, date of observation, date of illness onset, sex, weight, admission location)
2) Clinical data (symptoms, examination and vital signs findings)
3) Laboratory results (haematocrit, platelet count, white cell count and biochemistry results)
4) Diagnostic information (serology, reverse transcriptase polymerase chain reaction (RT-PCR), serotype data)
5) Interventions (fluid replacement, vasopressor, diuretic, and antibiotic use, transfusions of platelet or packed red cells).
6) Outcome coding (effusion, respiratory distress, bleeding, haemodynamic shock, length of stay and death).

Re-encoding of selected categorical variables was undertaken. For example, where the severity of bleeding was categorised in one study (S5) as ‘bruising’, ‘gastrointestinal bleeding’, ‘gum bleeding’, ‘epistaxis’, ‘vaginal bleeding’, ‘haematuria’ but classified only as ‘skin’ or ‘mucosal’ bleeding in another (S4). Patients in the included studies were screened on the basis of an AFI compatible with dengue. A standard diagnostic criteria employed by dengue clinical trials was employed [201] and confirmation of infection was based on:

An AFI compatible with dengue on clinical assessment,

and one or more of:

i) Detection of dengue NS1 by point of care lateral flow or through NS1 ELISA for dengue
ii) Detection of dengue RNA through RT-PCR
iii) A single positive dengue IgM by ELISA in acute serology obtained during episode of illness
iv) Seroconversion of IgM paired samples

From the entire cohort, patients: without confirmation of dengue infection (n = 6,329), enrolled in the interventional arm of the clinical study (n=150), not admitted for hospital management (n=1,226), who presented to hospital after onset of severe dengue and/or enrolled after 120 hours after illness onset and/or developed dengue within the first 48-hour period were all excluded from analysis. The latter timepoint of 120 hours was chosen as this represented the typical timing for onset of severe dengue to account for survivorship bias – this is in effect right-censoring the data so that direct comparisons can be made between patients who experienced complications and those who did not.

A random stratified split of 80-20% was done to the remaining 4,131 patients in the analysis group to produce a development set (n=3,304) which was used for model training and optimisation and a hold out test set (n=827) which was used exclusively for evaluation only. There were no statistically significant differences between the demographics or features between the development and hold out set through univariate analysis supporting that the hold out test set is representative of the overall population. A CONSORT-format patient diagram is shown in Figure 4.1 and the baseline characteristics with comparisons of the development and hold out set shown in Table 4.5:

4.6.3 Definition of prediction outcomes

The outcome for the model prediction is the risk of DSS at any point after hospital admission as a binary outcome. DSS represents the most common phenotype of severe dengue and is caused by excessive intravascular fluid leakage and haemodynamic collapse. Standardised clinical definitions of DSS were employed across all studies included in the analysis. A diagnosis of DSS was defined as:

Clinical signs consistent with reduced peripheral perfusion

and one of:

i) Low blood pressure for age in children [202],

ii) A pulse pressure equal or less than 20 mmHg in adults [167]

A secondary analysis was carried out using a composite dengue outcome defined as one of: DSS, and/or clinically-significant bleeding and/or plasma leakage.

Figure 4.1 Flow diagram for patients recruited in original clinical studies and processing to derive a development and hold out set for model evaluation

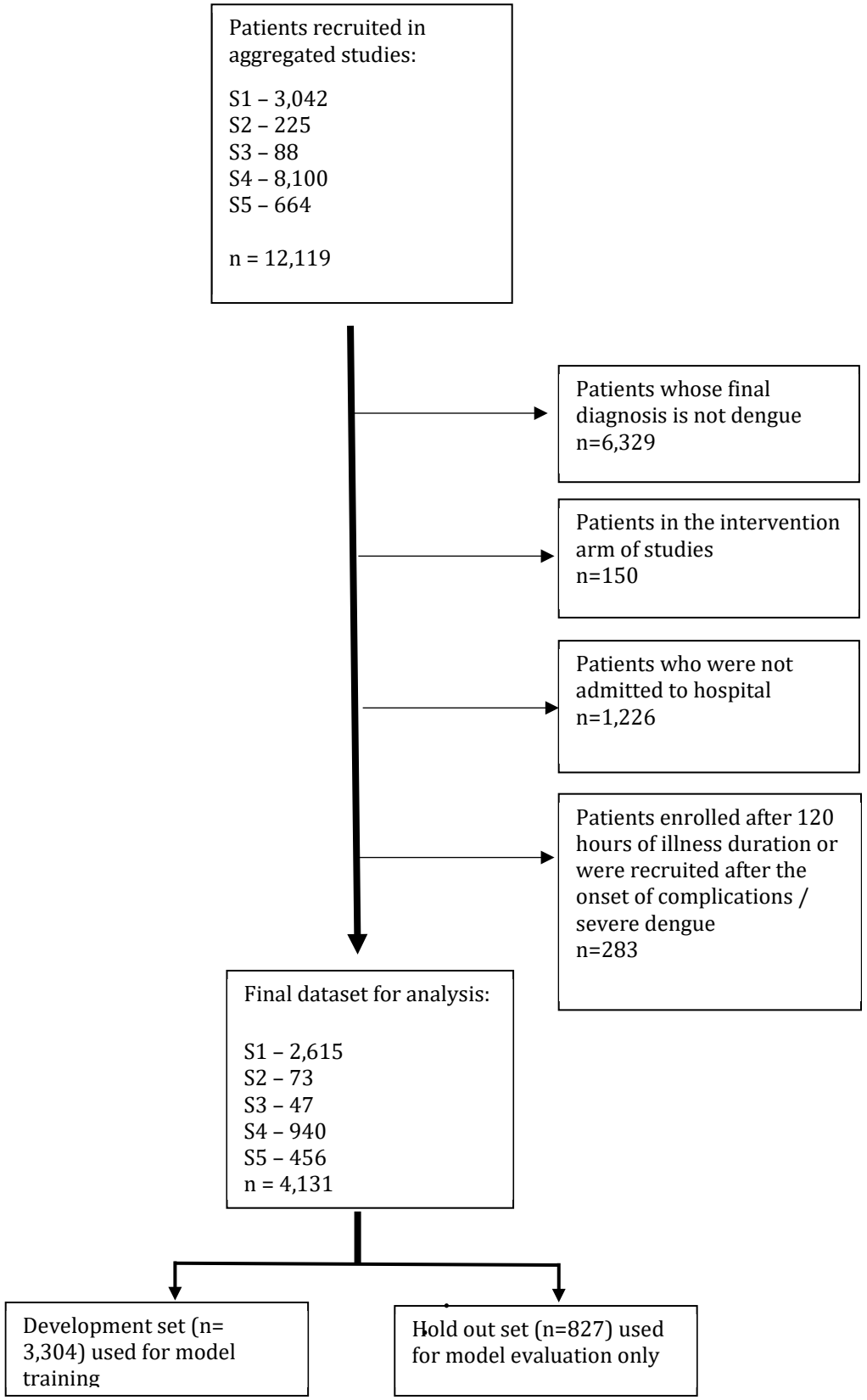


Table 4.5 Baseline comparisons between development and hold out test set generated through random stratified split, univariate comparisons made using chi-squared or Mann-Whitney tests

Parameter	Development set (n=3,384)	Hold out test set (n=847)	p-value
Complications (n, %)	291/3,093 (8.6%)	73/774 (8.6%)	0.959696
Median age (years)	12 (9-14)	12 (9.6-14)	0.30781
Median weight (kg)	35 (27-45)	35 (27-45)	0.374199
Female sex (n, %)	1651/3,384 (48.8%)	407/440 (48.1%)	0.730036
Median haematocrit (%,)	40 (38-42)	40 (38-42)	0.076784
Maximum haematocrit (%)	42 (39-46)	42 (39-46)	0.203267
Minimum haematocrit (%)	38 (35-40)	38 (35-40)	0.210143
Median platelet count (x 10^9/L, IQR)	115 (80-158)	112 (79-153)	0.071596
Maximum platelet count (x 10^9/L, IQR)	157 (117-207)	153 (116-202)	0.083157
Minimum platelet count (x 10^9/L, IQR)	85 (49-134)	80 (46-130)	0.026191

4.6.4 Feature selection, missing data, model development and evaluation approaches

The aim of model development was to produce a model with optimal performance for the prediction of DSS, but one which was amenable to implementation particularly in a low- and middle-income setting. Imputation and any data scaling was undertaken where required and fitted on the training set only to avoid data leakage. For categorical variables, data was imputed using the mode and missing numerical data imputed using either mean or median depending on results in cross-validation model evaluation.

Data scaling was done either by standardisation to achieve a mean value of 0 and a standard deviation of 1 or left unscaled. These different pre-processing parameters and steps were included as part of the grid search process in order to identify the optimal procedures for each algorithm involved. Feature selection was therefore based on expert consensus of predictors, data-availability and pragmatic utility. The final datasets for all

models included predictors with less than 1.5% missing data per feature. The final predictors (features) chosen for the model input were as follows:

i) Age in years
ii) Sex (Male/Female)
iii) Day of illness
iv) Weight in kilograms
v) Haematocrit in %
vi) Platelet count × 10^9 cells/L

Change in haematocrit and platelet over time were included as part of predictors given their central role the WHO 2009 dengue guidelines. Other clinical and laboratory variables including vital signs, patient history, serotype, viral load, liver and biochemical profiles were available in the dataset and are prognostic indicators in dengue. Pre-However, these factors were not included as the different study designs within the aggregated analysis dataset resulted in a significant proportion of missing data (>10%). Specialist tests such as serotype, viral load and patient biochemistry would also limit the implementation of the final model, especially for use in settings without access to specialist laboratory testing.

Four different machine learning algorithms were selected for model development in order to approach the problem – these utilise a mixture of tree and ensemble-based algorithms, network and linear techniques. Algorithms used were: extreme gradient boosting (XGBoost) [47], random forest [203] classifier, artificial neural network (ANN) [46] and support vector machine (SVM) [204]. Multivariate logistic regression with LASSO or ridge regularisation [205] was also utilised to establish a comparison performance. All models were applied as classifiers for a binary outcome, defined as development of outcome (DSS or the composite outcome) or not during the admission period.

Features in the training set were used to train each model. Optimisation of model hyperparameters for each algorithm was done through 10-fold stratified cross evaluation with the AUROC metric used as a measure of performance to discriminate. A hyperparameter grid search method was employed to iteratively examine combinations which would result in optimal performance.

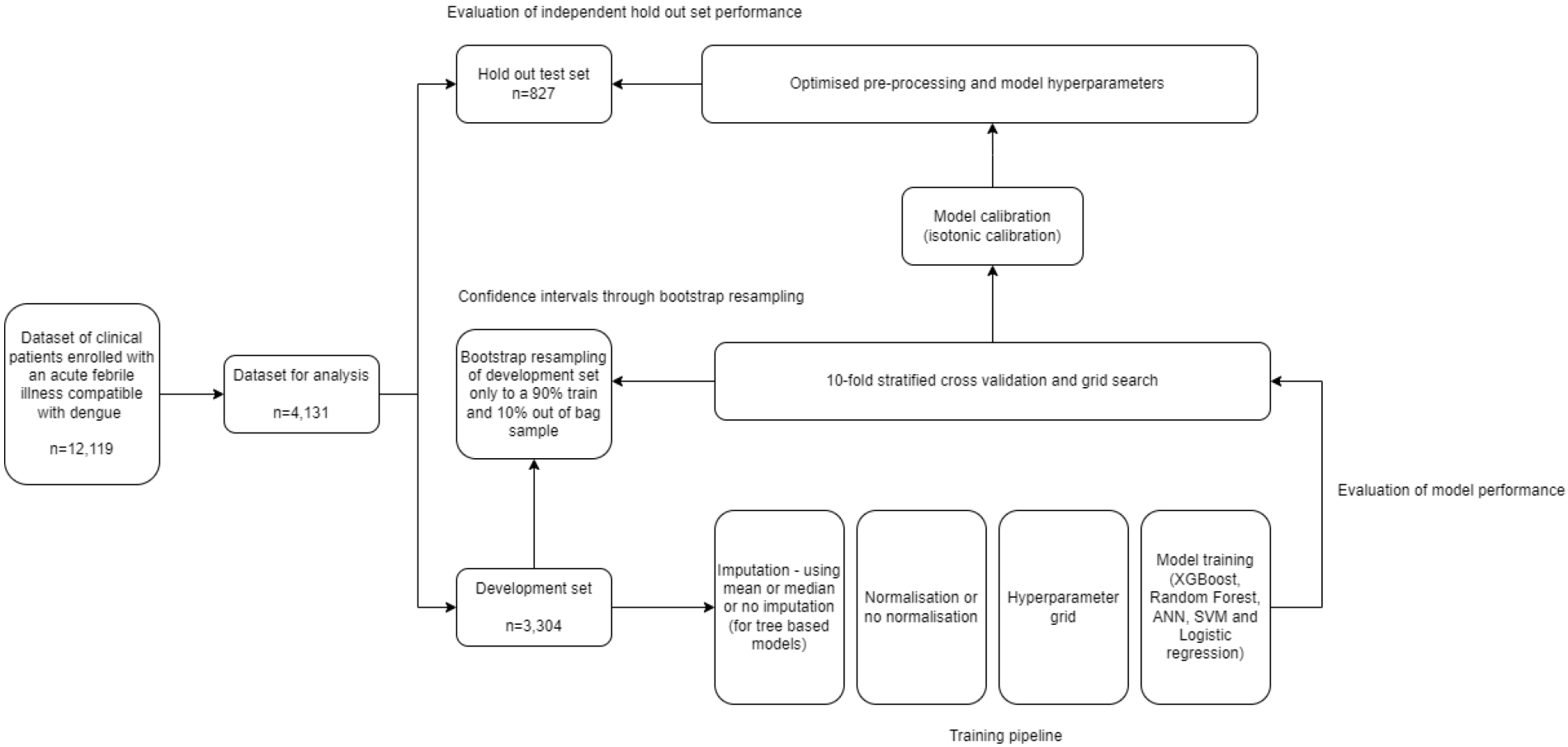
Default values for each model were used to generate a parameter grid for each algorithm, and a randomised search and cross-validation was used to narrow down the ranges of the hyperparameters. Grid search parameters for each model were subsequently derived and are presented in supplementary data. A probabilistic output was produced from each

model and optimal cut-off value was determined using Youden's J-statistic, (sensitivity + specificity – 1) [87] for each algorithm producing optimal sensitivity, specificity, positive and negative-predictive values for each algorithm. In order to establish the confidence intervals for the performance in each model, bootstrap resampling of the training set (90%) and evaluation against the out of bag samples (10%) was done, and repeated 1,000 times [62].

Each of the optimal models was then tested against the hold out test set for an independent evaluation of model performance using classification metrics including AUROC, sensitivity, specificity, positive and negative-predictive value (PPV and NPV).

Details of methods are presented in Chapter 2 and a diagram of the analysis pipeline is shown in Figure 4.2:

Figure 4.2 Data pipeline used for model training involving initial split to a development and hold out test set, with the latter used only for evaluation. All transformations and imputation were fitted to the training set only. In the development set a pipeline process for each algorithm, which included pre-processing, imputation and hyperparameter tuning was set up. Each combination of parameters was evaluated through 10-fold cross validation in an iterative grid search approach. Optimal models for each algorithm were determined by their AUROC in cross validation and underwent isotonic calibration before they were used for model evaluation against the hold out set



4.6.5 Time-series approaches and baseline models

To capture trends in patient variables over the febrile phase, the minimum, maximum and median values of laboratory results taken in the first 48 hours of patient admission were as individual features in the model, provided that these were obtained within 120 hours after the onset of fever and before onset of the outcome in question (DSS or the composite dengue outcome).

Each patient included in the final analysis had at least 2 laboratory values and aggregation will therefore likely produce different values for the minimum, maximum and median in order to summarise trends over that period. This methodology was chosen given the low number of measurements taken within this admission period.

A time-invariant model was also trained as a comparator using patient laboratory data obtained from *admission only* to predict clinical outcomes.

4.7 Results

4.7.1 Quantitative analysis of data

For the final analysis cohort (n=4,131), patients recruited between 12th April 2001 and 30th January 2018 were included. The overall median age of the cohort was 12 (Interquartile range, IQR: 9 to 14 years old), and the proportion of female patients was 49% (n=2,008). Those who experienced dengue shock were slightly younger (11 vs 12 years old). In total, 222 (5.4%) patients experienced DSS during hospitalisation. For the secondary analysis 344 (8.3%) of patients experienced the composite outcome defined as DSS, significant bleeding and fluid leakage.

Characteristics of patients included in the analysis cohort by original study are shown in Table 4.6 and their clinical characteristics displayed in Table 4.7.

4.7.2 Model performance for prediction of dengue shock using 48 hour aggregated data

All machine learning algorithms provided similar performance in 10-fold cross validation after the optimisation process for prediction of DSS in patients with dengue admitted to hospital during the febrile phase. The mean AUROC for the models ranged between 0.77 and 0.83. Performance of logistic regression using a LASSO regularisation method yielded

a AUROC of 0.79 (95% CI 0.74-0.83). The ANN implementation model provided optimal discrimination performance in predicting DSS (AUROC 0.83, 95% CI 0.76-0.85) as well as the highest AUPRC of 0.42. With optimal cut-off thresholds this model provided a NPV of 0.98 (95% CI 0.97-0.99), PPV of 0.24 (0.12-0.33), specificity of 0.88 (0.68-0.93) and sensitivity of 0.66 (0.52-0.81). The results of model training and 10-fold cross validation performance are displayed in Table 4.8.

Table 4.6 – baseline characteristics of individual studies included in the final dataset

Study	Recruitment dates	Patients enrolled (n)	Patients with confirmed diagnosis of dengue (n)	Patients included in analyses (n)	Patients included in analyses					
					Age (median, IQR)	Day of illness onset on hospitalisation (median, IQR)	Female sex (n)	Secondary infection (n)	Shock (n)	Outcome measurement
S1	12/4/2001-24/7/2009	3,042	3,042	2,615	12 (10-13)	4 (3-4)	1079 (41%)	1536/1903 (81%)	170 (7%)	WHO 1997 dengue outcomes
S2	3/8/2009 – 8/12/2010	225	225	73	15 (13-20)	1 (1-2)	18 (25%)	42/64 (66%)	3 (4%)	WHO 2009 dengue outcomes
S3	8/12/2010 – 16/6/2011	88	88	47	23 (21-25)	3 (3-3)	28 (40%)	11/38 (29%)	8 (17%)	WHO 2009 dengue outcomes
S4	19/10/2010 – 4/12/2014	8,100	2245	940	12 (10-16)	2 (1-3)	427 (45%)	687/915 (75%)	31 (3%)	WHO 2009 dengue outcomes
S5	20/10/2016 – 30/1/2018	664	542	456	30 (27-45)	4 (3-5)	456 (100%)	494/539 (92%)	10 (2%)	WHO 2009 dengue outcomes

Table 4.7 Characteristics of cohort (n=4,131) divided by complication outcome during hospital. Haematological values and ranges refer to results taken over the initial 48 hours of hospitalisation only. Data is presented as medians and brackets denote the interquartile range. Univariate analyses were performed using the Mann-Whitney test

Parameter	Shock n = 222	No shock n = 3,909	p-value	Missing data (%)
Median age (years, IQR)	11 (9-13)	12 (9-14)	0.07	1.5
Median day of illness at hospital admission (days, IQR)	3 (2-4)	3 (2-4)	0.05	0
Median weight (kg, IQR)	34 (27-43)	36 (27-45)	0.26	0.9
Female sex (n, %)	90/222 (41%)	1918/3,909 (49%)	0.02	0
Median haematocrit (%, IQR)	41 (39-43)	40 (37-42)	<0.001	0.2
Maximum haematocrit (%, IQR)	48 (44-51)	42 (39-45)	<0.001	0.2
Minimum haematocrit (%, IQR)	37 (35-39)	38 (35-40)	0.05	0.2
Maximum platelet count (x 10^9/L, IQR)	140 (102-192)	159 (118-207)	<0.001	0.2
Median platelet count (x 10^9/L, IQR)	70 (45-99)	116 (83-159)	<0.001	0.2
Minimum platelet count (x 10^9/L, IQR)	29 (18-58)	86 (51-135)	<0.001	0.2

Table 4.8 Final model performance after hyperparameter optimisation through 10-fold cross validation. The 95% confidence intervals (CI) were derived from bootstrap resampling of the development set only to derive a training set (90% of the development set) which was evaluated against an out of bag sample (remaining 10%).

Model	Mean AUROC	Specificity	Sensitivity	Positive Predictive value	Negative Predictive value	Brier score	AUPRC
XGBoost	0.81 (0.74-0.84)	0.90 (0.70-0.93)	0.59 (0.52-0.78)	0.26 (0.12-0.31)	0.97 (0.97-0.98)	0.068	0.36
Random forest	0.80 (0.75-0.84)	0.82 (0.75-0.93)	0.71 (0.54-0.76)	0.18 (0.14-0.31)	0.98 (0.97-0.98)	0.068	0.38
Logistic regression	0.79 (0.74-0.83)	0.89 (0.71-0.91)	0.62 (0.56-0.78)	0.24 (0.13-0.28)	0.98 (0.97-0.98)	0.075	0.30
Artificial neural network	0.83 (0.76-0.85)	0.88 (0.68-0.93)	0.66 (0.52-0.81)	0.24 (0.12-0.33)	0.98 (0.97-0.99)	0.068	0.42

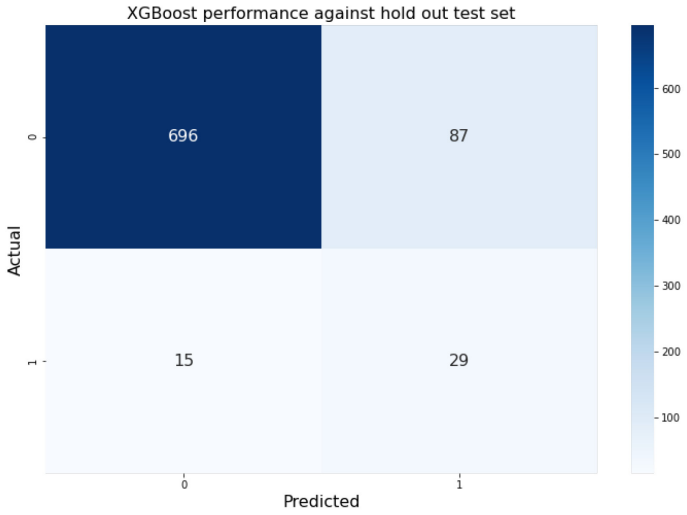
Support vector machines	0.77 (0.71-0.81)	0.80 (0.66-0.94)	0.63 (0.43-0.77)	0.15 (0.11-0.33)	0.97 (0.96-0.98)	0.070	0.33

With evaluation against the independent hold out set of 827 patients, again similar performances were obtained, with the XGBoost model providing a AUROC of 0.84 (Table 4.9). The confusion matrix for prediction for this set is also shown in Figure 4.3:

Table 4.9 Performance metrics of the final models after undergoing calibration when evaluated against the hold out test set (n=827)

Model	Mean AUROC	Specificity	Sensitivity	Positive Predictive value	Negative Predictive value
XGBoost	0.85	0.91	0.64	0.29	0.98
Random forest	0.84	0.87	0.68	0.23	0.98
Logistic regression	0.79	0.90	0.55	0.24	0.97
Artificial neural network (ANN)	0.82	0.84	0.66	0.18	0.98
Support vector machines	0.75	0.84	0.59	0.17	0.97

Figure 4.3 Confusion matrix displaying performance of the hold out test set for XGBoost where 1 represents DSS and 0 represents no complication



4.7.3 Model performance for secondary outcomes and time-invariant model

Model development was done to predict secondary composite outcomes comprised of one of more of: DSS, bleeding and leakage. However, data collection was done in such a way

that the timing of bleeding and leakage events was not recorded (unlike for DSS). Although it is likely that these outcomes (especially bleeding) occur at similar timepoints to onset with DSS at time of defervescence given the natural history of severe dengue [167] results are to be interpreted as a classification model rather than for prediction. Nonetheless there is utility in the identification of those who *do not* develop any of these outcomes as these patients can be stratified as low risk and managed accordingly.

Model construction and training followed previous approaches above – with grid search using 10-fold cross validation encompassing pre-processing and hyperparameter values followed by hold out set testing generated using the same previous split. The best performing model was a random forest classifier which provided a mean AUROC of 0.76 (95% CI 0.71-0.79) in cross-validation and AUROC of 0.77 in hold out set testing. Logistic regression provided a AUROC performance of 0.70 (0.66-0.74) in cross validation and 0.71 In hold out set testing.

A further analysis was performed using static variables as features, The first haematocrit and platelet value for each patient on admission were used as features without including subsequent laboratory values. The hypothesis is that subsequent values and trends do provide additional information to support predictive performance and use of these static variables would result in lower ability to predict outcomes.

The AUROC for all models in 10-fold cross validation after optimisation ranged between 0.59–0.69 with hold out set testing AUROC 0.56-0.73. The best performing supervised machine learning algorithms for model development were the ANN implementation and random forest classifier (AUROC both 0.68), with the AUROC of logistic regression being 0.69. These performances are lower than those obtained by the former approach encompassing time-series trends. Full results are shown in the supplementary data.

4.7.4 Model interpretability

Interpretation of probabilistic outputs

These models provide the most utility when appropriately implemented into clinical practice to support routine decision-making. In turn aspects of *how* these model outputs are to be interpreted by the decision-makers in question, and whether these 'black-box' models require explicit means to aid explainability need to be addressed.

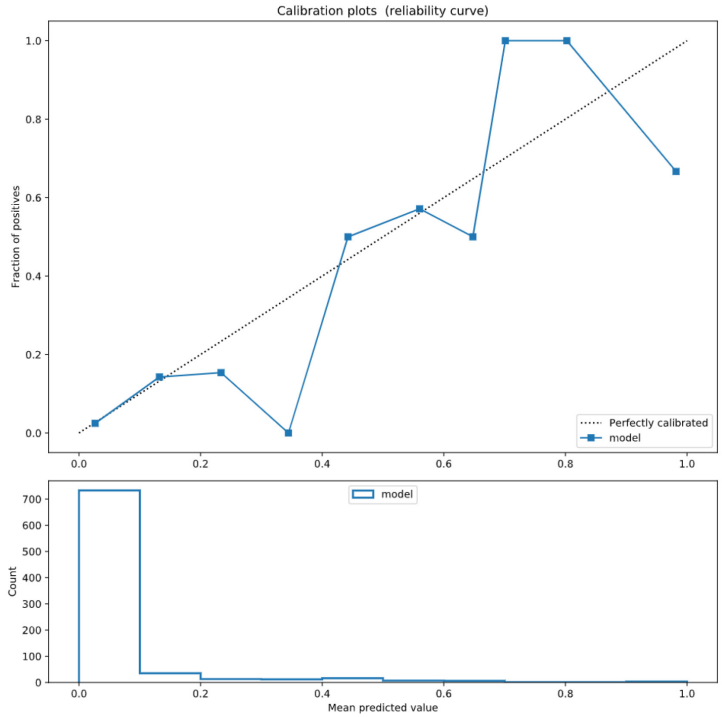
Prediction and classification models provide a discrete dichotomous output e.g., whether DSS is likely (class 1) or DSS is unlikely (class 0) provided a threshold function is applied

to the probabilistic output. The default cut-off for the threshold is commonly taken as 0.5 and this value can undergo optimisation to adjust the trade-off between sensitivity and specificity. Youden's J index was used as the threshold function to optimise this cut-off value in the models in this chapter. However limitations to this approach include the fact that other factors in determining optimal thresholds, including context and cost may not be taken into account [206]. For example, in decision-making in dengue, a model used in patient triage which misses fewer cases of severe dengue at the expense of false positives might be preferential.

The use of probabilistic outputs directly, ideally quantified with CI, is desirable in decision-making. Interpretation of probability and uncertainty in healthcare varies significantly however – although the probabilistic output attempts to quantify the degree of uncertainty in prediction (e.g., an output of 0.9 can be interpreted with more confidence than an output of 0.6), how this value is interpreted in turn by the decision-maker is less clear [207]. Downstream, how information from these outputs should acted is even less certain.

The degree of calibration in a model is important attribute in allowing probabilistic outputs to be used with confidence [208]. In an ideal and well-calibrated model, predicted probabilities should correspond to the actual frequency of the output labels – for example: for a given set of patients with dengue, if the probabilistic output for DSS is 0.75 then 75% of these patients should experience DSS (the true positive event). Through this frequentist interpretation of probability, calibration can be quantified through a calibration curve plot and the Brier score metric [209] which is a reflection of the mean squared difference between predicted probabilities and actual outcome. Calibration can also be applied to the model post training in which outputs are rescaled to better represent actual frequency-probabilities of a given set [210]. The Brier scores for the DSS model are displayed in Table 4.8 above and an example calibration curve for a classifier shown in Figure 4.4. The presentation of machine learning model outputs for clinical application is covered in greater detail in Chapter 6.

Figure 4.4 Calibration curve for the artificial neural network model for prediction of DSS using 48 hour time-series data



Model interpretability and explainability

Machine learning models have the capacity to incorporate large volumes of diverse data into its predictions and when deployed for such purposes are often complex by nature. Understanding *how* such complex models produce predictions may be important for implementation, especially when intended for a clinical context. In the most basic incarnation – many machine learning models can be viewed, and used as 'black boxes' which convert inputs (data) into outputs (predictions). However, having the attribute of explainability i.e. allowing a user to understand how these predictions are derived can be important for a range of different reasons. Main arguments for explainability from a recent commentary [211] are summarised in Table 4.10.

Table 4.10 Arguments for explainability in machine learning for real-world applications

Domain	Purpose	Comment
Technological	Development of robust, safe and fair algorithms	Ensuring that the model does not make predictions based on meta-data or on confounders such as ethnicity, with implications on future performance and biased predictions.

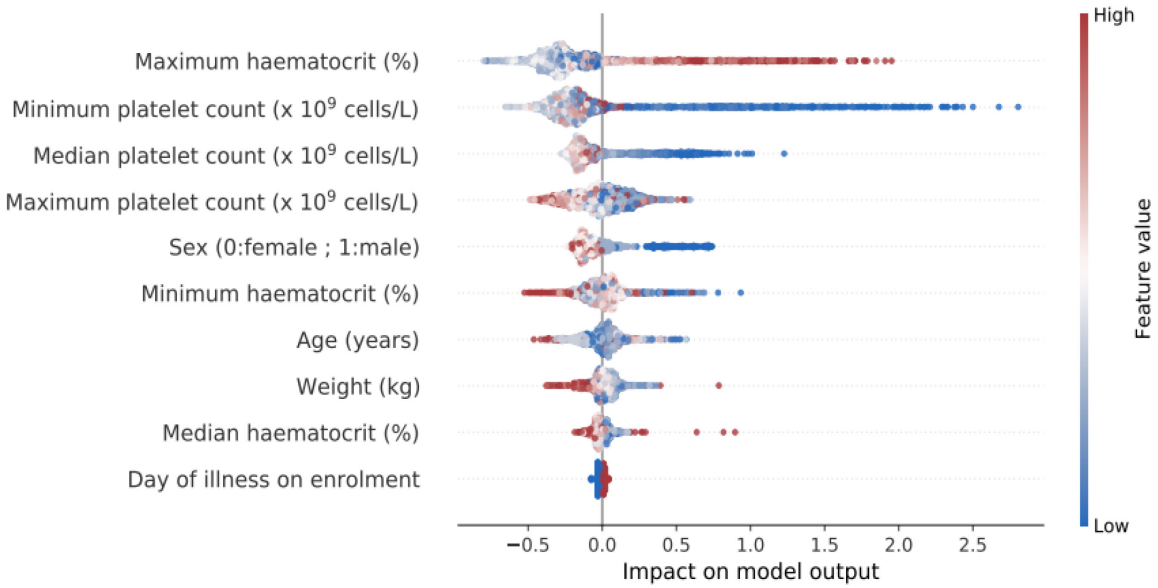
	Monitoring performance	Provides a means to interrogate the model when performance or context changes with time.
Legal	Informed consent	Ability to explain to patients how the machine learning model works as part of the informed consent process, especially how their personal data is used for this process.
	Oversight and accountability	For understanding risks and benefits of using a particular machine learning model, especially in the context of such algorithms making the wrong treatment recommendations. Accountability should lie with the clinician with explainability as a prerequisite (with implications also for litigation).
Medical	Trust and uptake in medical decision making	Understanding model limitations so that these factors can be incorporated into clinical decision-making. A system which is interpretability and explainable is likely to facilitate better quality decision-making
	Provide insights for research	As a means to understand better contributions towards health and disease covered by the model, and could allow for novel research questions to be formulated.
Patient	Patient centred-care	To understand to what extent the model applies to the individual in question in order to allow better shared decision-making.

In order to understand model predictions, a post-hoc approach to model explainability was used for the dengue model. This was done through assigning a contribution for each feature to the predicted model output based on game theory by calculating SHAP values [89]. These values are derived through an independent, model-agnostic process and can be interpreted within a linear mode whereby the sum of the values explains output for a particular prediction, whilst capturing interaction effects [90]. SHAP values were derived for the hold out test set using the XGBoost model and summarised in Figure 4.5: on the x-axis, the impact on the model prediction output is shown whereby a positive value indicates an increased contribution towards the prediction of DSS and a negative value indicates the reverse. The features on the y-axis are arranged in descending importance order from top to bottom, and colours represent the actual values of the features.

The summary plots in figure 4.6 hows that a higher maximum haematocrit, lower minimum platelet and median platelet count, female gender, younger age and lower weight all contributed towards a prediction of DSS, consistent with current understanding

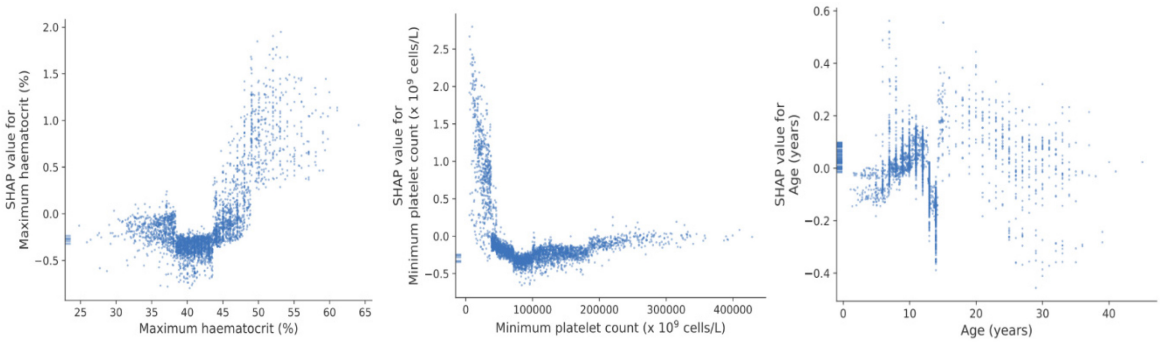
[189]. The trend of platelet count was also an important factor given that minimum, median and maximum haematocrit are all ranked as features.

Figure 4.5 Summary plot of SHAP values for the XGBoost model. The plot shows the contribution of individual predictors and their range of values towards final model output prediction, where shock and no shock are represented by 1 and 0 respectively. The main predictors are arranged in descending importance for the model.



The contribution of each of the variables were mapped through a scatter plot of individual variables against their SHAP values for the XGBoost model. The plot in Figure 4.6f shows that changes in feature values affect predictions in the model through a non-linear fashion: increase in SHAP values is associated with haematocrit levels > 45% likely indicating sub-clinical plasma leakage without shock. Risk of DSS also increases significantly with platelet levels less than 100,000 cells × 10^9/L. Age has a complex effect with regards to DSS risk but in general increases with age until 10 then decreases sharply again consistent with understanding of risk of severe disease and timing of secondary infections in endemic settings [212].

Figure 4.6 A scatter plot of predictor values (x-axis) against SHAP values (y-axis) in the XGBoost model for DSS for haematocrit, platelet count and age. The plot shows a non-linear relationship between predictor values and model output.



4.8 Discussion

Risk stratification remains one of the most important tasks in dengue management. The aim of the chapter was to examine how clinical data collected over time could be meaningfully used to guide management of patients hospitalised with dengue in Vietnam. Given dengue is a dynamic disease and rapid changes in haemodynamic and fluid status occur during the febrile phase, the hypothesis was that explicitly accounting for these dynamic factors in a model would provide better understanding and improve performance. This problem was approached through aggregation of basic demographic and laboratory parameters over the first 48 hours of a patient's hospital admission, and the use of different supervised machine learning methods. The goal was to predict the risk of DSS in patients admitted during the early febrile phase but did not have any signs of severe disease.

Optimal performance was achieved through an artificial neural network (ANN) algorithm. The mean area under the receiver operating characteristic of the ANN model in discriminating patients who did or did not develop DSS after 48 hours was 0.83 in 10-fold cross validation and AUPRC was 0.42 in testing against the independent hold out test. Similar results were also attained when this model was evaluated against the independent hold out test set (AUROC of 0.82). As a binary classifier using prediction thresholds determined by the J-statistic, the model provided a negative predictive value of 0.98 which would be of clinical utility in practice.

Final feature predictors used in the model were patient age, sex, day of illness, weight, haematocrit and platelet values over the first 48 hours of hospital admission. Feature selection was done to minimise missing data across the aggregated datasets and reduce the need for data imputation. Limiting the number of features used in these models will also likely reduce predictive performance, given that variables including serological status and NS1 titres [176], biochemical and endothelial biomarkers [213], viral load [214] and co-morbidity are associated with, or independently predict severe dengue. Future models which are able to leverage these data effectively will likely strengthen predictive performance.

However, there exists a need to balance model performance through inclusion of detailed patient data with the feasibility of roll and implementation, particularly in a low- and middle-income healthcare setting. Access to laboratory services can be limited by availability of infrastructure, transport and cost [215]. A simpler model which can provide decision-support to a broader setting might result in greater clinical impact than a

complex model which is more accurate. The model developed in this chapter was trained using data derived from a > 4,000 patients recruited from a series of prospective real-world observational clinical studies in Vietnam. It will therefore have the most applicability when deployed in such endemic settings.

To account for the dynamic changes seen during early dengue, summaries of features in the form of minimum, median and maximum values over the 48-hour period were used for model input. An alternative approach would have been to include every observation as an additional term in the model, although this would have resulted additional missing data imputation in those patients with fewer observations. Ideally, granular vital signs data in these 48 hours would be included but again electronic capture of these data in our setting has been shown to be challenging [216]. In comparison a static model implementation where the first available datapoint was used for prediction, the final model had better performance (AUROC 0.83 vs 0.69) by presumably taking into account information relating to patient change over time.

Model performance in this chapter was commensurate with other published prognostic models. Whereas other models have been developed for either adult or children, the model presented in this chapter has been trained and validated in both groups. The time-series nature of dengue disease and impact on changing laboratory values were also explicitly captured in this model. Similar performances were observed across different implementations of machine learning and logistic regression methods. The mean AUROC was higher for ANN models compared with linear models (0.83 vs 0.79) however with bootstrap resampling, actual confidence intervals did support a statistically significant difference in performance. This is likely a reflection on data-dimensionality. Machine learning is particularly suited for analysis of large scale, multidimensional data and provide robust methods for including diverse data types such as physiological signals and radiological images within the same framework. At the time of this study there have not been any machine learning implementations for clinical risk prediction although this will likely change with availability of digitalised data.

Post-hoc explainability of models through calculating SHAP values was done to understand contribution of each feature included in the model. Through explainability analysis, a higher maximum haematocrit, a lower overall platelet count, female gender, younger age and lower weight were associated with model prediction for DSS. These characteristics are in line with understanding of dengue risk in the literature [165] and adds validity to the model. Although the set of features used in the model was relatively small, model interpretability is important in the clinical setting, given that clinicians need

to trust model outputs and be accountable for management decisions, although the extent of interpretability of black box systems is ultimately limited, and ethical issues surrounding this are complex [217].

There were limitations to the work. There was a class imbalance in the data in which only around 5% of the cohort developed the outcome of interest. This figure is in line with real world data in which the majority of patients admitted for hospitalisation (based on WHO warning signs criteria and other factors) tend to experience uncomplicated disease. Data augmentation methods to address the class imbalance including SMOTE [218], and oversampling of the minority class were attempted in the pipeline but this did not lead to improvements in overall model performance (results not shown). Secondly, the model does not account for treatment such as fluids which were administered during early hospitalisation – these data were not available as features for the model. Dengue treatment is however standardised at the HTD according to Ministry of Health guidelines and therefore there is unlikely to be significant differences in treatment effects between patients.

In conclusion the models developed in this chapter utilise clinical data collected over time to predict the risk of DSS and possess performance metrics which are suitable for clinical implementation. Although there exists a lack of effective disease-directed therapeutics in dengue, the identification of those at most risk of complicated disease will allow for better use of resources and monitoring. The model if incorporated into a decision-support system could also add a layer of safety for healthcare particularly in settings with greatest need and during seasonal epidemics. Prospective evaluation of the model in the setting of the HTD, Vietnam, as well as other healthcare settings in Vietnam and South-East Asia would be appropriate to understand real-world performance and gather metrics on its clinical utility.

Chapter 5 – Accounting for the dynamic influence of seasonality in the diagnosis of acute febrile illnesses

5.1 Chapter overview and summary

The chapter examines the role and contribution of dynamic non-healthcare factors, including seasonality in prediction models developed to support clinical management. I employed machine learning methods to include basic, routinely collected blood biomarkers with the aim of predicting the final diagnosis of patients presenting with an AFI. I examined the effects of dynamic environmental variables including local rainfall, temperature and seasonality factors on the performance of these clinical prediction models and developed methods to account for their effects on performance.

Many infections exhibit a seasonal pattern as a function of climate and its effects on the disease vector. Such tacit knowledge is commonly used by clinicians in decision-making to guide differential diagnoses and management. However, there is a lack of research which explicitly accounts for these factors within diagnostic clinical prediction models.

I utilised a non-linear machine learning algorithm, XGBoost, to predict the final diagnoses in a cohort of 8,100 children with an AFI. This prospective dataset was collected between 2010 and 2014 in Vietnam. Patients attending a healthcare centre with an AFI were diagnosed with either laboratory-confirmed dengue, or 'other febrile illness' (OFI). I chose features based on data which was most readily-available for the setting, including age, sex, haematocrit, platelet, white cell and lymphocyte count at clinical presentation.

The model discrimination performance between diagnoses was associated with an AUROC of 0.86 (Interquartile range, IQR 0.84-0.86). However, I found predictive values of the model varied significantly throughout the year which would reduce real-world clinical utility. I developed dynamic models to account for these changes through using a rolling window threshold to account for recent cases to maintain predictive values throughout the year.

I observed relationships between the optimal model thresholds, temperature and precipitation data from the region. A secondary dynamic model was developed by directly incorporating temperature and precipitation values to the clinical prediction model. These resulted in consistent predictive values over time values (median negative predictive value 91%, IQR 0.89-0.93). I characterised the features of OFI through

unsupervised machine learning methods to gain better understanding of their characteristics and seasonality effects.

Chapter conclusion: By accounting for the temporal factors associated with AFI and through integrating non-healthcare factors which exhibit seasonality into the model, I was able to strengthen machine learning clinical prediction for the diagnosis of AFI. My findings support the role of data as an adjunct to conventional pathogen diagnostics. Such models could play a role in clinical decision support, and passive surveillance for infectious diseases.

5.2 The acute febrile illness syndrome

Fever is one of the most common presentations to healthcare. Within resource-limited settings and low and middle-income countries (LMICs) [219], communicable diseases including lower respiratory infections, diarrhoeal diseases, malaria, meningitis and whooping cough account for the top ten disability-adjusted life years (DALYs) in children [220]. Historically, malaria has been synonymous with fever across multiple regions in Sub-Saharan Africa and Asia [221]. Global interventions to tackle malaria which have focused on providing prompt access to treatment as artesunate combination therapy, provision of preventive therapy to susceptible groups, vector control strategies and the roll-out of insecticide-treated nets have resulted in significantly lower disease burdens [222]. As a result of decreasing malaria incidence, the contribution of non-malarial febrile illnesses (NMFIs) as a cause of the AFI syndrome have increased proportionately, but also numerically over time as a result of the diverse factors such as ecology and human-induced climate change [223]. As a heterogenous group of infections caused by viruses, bacteria, protozoa and fungi much less is understood regarding regional incidence and distribution of NMFIs [224].

Diagnosis is central in guiding individual clinical management in infections. Individuals with febrile illnesses may present with a syndrome associated with features such as cough, rash or diarrhoea. However, undifferentiated fevers where such adjunctive clinical cues are either missing or non-specific pose challenges to clinical diagnosis and adversely impacts on management [225]. Ideally, key information on causative organisms as well as key properties (e.g., resistance profile, virulence, transmissibility) is available at the point of decision-making. Having a diagnosis available supports risk stratification and care prioritisation, allows assessment on whether empirical antimicrobials are warranted [226], and is key to specific prescribing of agents which best target the aetiology.

Beyond individual care, aggregating information on diagnoses provides crucial understanding of regional epidemiology. Public health interventions such as education campaigns, vector control measures and policies on research and funding are contingent on these data. Knowledge of caseload, and distribution of many infections can direct appropriate resources in prevention and vector control [227]. As highlighted by the COVID-19 pandemic, where such information collated from diverse data sources has been important in guiding the public response [228].

5.2.1 Role of diagnostics in low and middle-income country settings

The diagnosis of NMFIs is inherently challenging given that the group is one of exclusion. These infections are caused by a diverse range of organisms and diagnosis needs to be tackled individually. Many LMIC healthcare settings suffer from a lack of diagnostic capacity for infectious diseases – but these issues exist as part of a wider problem relating to inadequate pathology and laboratory services capacity. Common factors including training, infrastructure, fragile supply chains, and insufficient quality standards drive this problem [229]: solutions to tackle this have to be tailored to the particular setting in question [230]. Where laboratory services are available, factors including cost and prolonged result turnaround times represent additional barriers [231].

Information provided by laboratories are the important substrates for clinical decision-making and allow for improved prognosis, treatment decisions and disease monitoring. In settings without diagnostics, the reliance on empirical antimicrobial strategies can result in treatment failure, but also exacerbate antimicrobial resistance [3]. And where such diagnostics are available and affordable for use – their implementation can be complex with unforeseen consequences. For example, introduction of malaria tests as a standalone intervention paradoxically increased the rate of empirical antibiotic prescribing in patients where the test result was negative [232].

Dengue represents a major cause of AFI globally, with the majority (70%) of cases occurring in South-East Asia [40]. In a major fever study in Laos where the aetiological agent was identified, 21% of participants had serological evidence of recent dengue infection [224]. Accurate diagnosis represents a key priority for clinical management as a minority of patients develop severe dengue, which is life-threatening. For many, diagnosis is syndromic in nature, but this process is ultimately dependent on local experience and training. In terms of laboratory diagnosis, tests considered gold-standards for diagnosis in viral infections including paired serology or other viral isolation. But these techniques are impractical in guiding clinical management, given their long turnaround time. Nucleic acid amplification tests (NAAT) such as reverse transcriptase polymerase chain reaction (RT-PCR) of blood is considered the preferred test for dengue diagnosis and provides good performance. There is a short period of viraemia in peripheral blood (typically up to 4-6 days of illness) and access to testing can be an issue as specialist laboratory infrastructure and training is required to support this process.

A key point-of-care diagnostic used for dengue diagnosis is the dengue specific NS1 lateral flow assay. This test can be performed by the healthcare professional using patient

capillary blood, and has acceptable performances when compared with other diagnostics in dengue [233] . Results are affected by dengue serotype, viraemia, timing of illness and host antibody response [234]. In later stages of dengue (day 4 onwards), blood NS1 testing can result in a false negative – this is important as differential diagnoses of patients who present with undifferentiated fever and shock include sepsis, where management can be very different. Novel approaches which are able to support clinical diagnosis and management at point-of-care are a necessity [223, 235, 236]. Other NAAT-based molecular diagnostics such as loop-mediated isothermal amplification (LAMP) are currently a subject of research given that they do not require laboratory thermocycler equipment and are of generally lower-cost. Nonetheless, implementation of these next-generation diagnostics currently remains out of reach of most LMIC settings.

5.2.2 The role of adjunctive data to support patient diagnosis

With increasing data availability and novel methodologies to process and consolidate Big Data, there exists an opportunity to utilise data-driven approaches to support diagnosis of AFI particularly in settings where access to gold-standard diagnostics is challenging. Disease modelling and forecasts have been used for a range of infectious diseases. In dengue modelling, information from meteorological sources [237], satellite imagery [238], human mobility [239], internet search data [240] and other diverse sources have been incorporated in order to estimate caseloads, future incidences and geospatial distributions.

At a clinical level, much research is focused on the use and evaluation of diagnostic technologies to inform diagnosis at point of care. However, in settings where only basic and non-pathogen specific investigations are available, electronic prediction tools and algorithms offer opportunities to improve the clinical management of AFI [200, 241–244]. These utilise basic clinical or rapid-testing information (such as urine dipsticks and rapid diagnostic tests) to provide pathogen-specific or syndromic diagnosis for the clinician, in turn offering recommendations for management and follow up. With the exception of the WHO Integrated Management of Childhood Illness (IMCI), such tools have been limited to research use only and not widely translated into clinical practice.

5.2.3 Impact of non-healthcare factors in clinical diagnosis

Inherent in clinical management of many infections is seasonality. For many such diseases in South-East Asia such as dengue, influenza, scrub typhus, leptospirosis and typhoid a changing pattern over the year in association with rainfall and temperature is seen. Cases of dengue in Southern Vietnam increase precipitously between May – November where hospitalisation rates can increase by more than ten-fold within the same year [245]. This change in dengue disease incidence is largely attributed as a function of reproduction and behaviour in the *Aedes* mosquito vector, and affected by weather variables including humidity, rainfall and temperature variability [246]. Other diverse factors including human movement, urbanisation and change in land use are also contributory to changing distributions and incidence of dengue [247]. During COVID-19 disruptions to travel and movement, the incidence of dengue reduced dramatically across most endemic settings globally – this was attributable to societal disruptions after controlling for climatic, host factors or disruption to vector control programmes [248].

Such climatic and human factors are inherently dynamic, varying according to time and location. Although their impact on AFI and disease incidence are well-established, how they can be meaningfully incorporated into systems to guide clinical management is unclear. Tacit knowledge of seasonality and local epidemiology are routinely utilised by clinicians to inform diagnosis and management – yet existing frameworks mentioned in the previous section do not take this into account. Pre-test probability of a causative agent in an endemic setting also changes with time. In dengue, during a period of low disease prevalence, a positive dengue NS1 test therefore needs to be viewed with additional scrutiny, given the lowering of its positive predictive value.

With the backdrop of human-induced climate change, new areas with autochthonous transmission and changing seasonal epidemiology of infections over time – understanding how these factors affect disease distribution and how to meaningfully incorporate this information to support clinical management to change is important.

5.3 Aims and Research hypotheses

The chapter addresses the diagnosis of AFI and provides understanding of the role of non-healthcare data for supporting clinical diagnosis through the use of machine learning models. Dengue infections will be used as an exemplar.

The aim is the development of a data-driven approach which accounts explicitly for time-series information to augment diagnosis for patients presenting with a non-malarial AFI within a LMIC setting. The chapter will provide understanding on the contribution of non-healthcare factors in providing context and improving diagnosis.

The research hypotheses are that:

- The use of routine existing clinical and non-healthcare data has value in supporting diagnostic pathways for the management of acute febrile illnesses and diagnosis of dengue.

- Non-healthcare or clinical factors which change over time, including seasonality, play a significant role and affect the performance of such pathways.

- Accounting for changes in these non-healthcare factors in models can provide context and supports better, or more consistent performance of such clinical prediction models.

5.4 Statement of contribution

All work presented in this chapter is of my own.

5.5 Ethics statement

The analyses and use of retrospective clinical data were approved by the scientific and ethical committee of the HTD, Ho Chi Minh City and by the Oxford Tropical Research Ethics Committee (OxTREC) with datasets pseudonymised prior to analyses (references 145-0420 and 146-0420). Correspondence and approval letters between Oxford University Clinical Research Unit (OUCRU) on behalf of Imperial College London with the HTD are included in the thesis appendix.

5.6 Methods

This is an analysis of data from an observational clinical study performed in Vietnam [14,15]. The aim of the work in this chapter was to develop and evaluate machine learning models which could be used during early patient febrile illness (<72 hours onset) in order to predict the final diagnosis, and understand how such an approach is affected by external factors including seasonality.

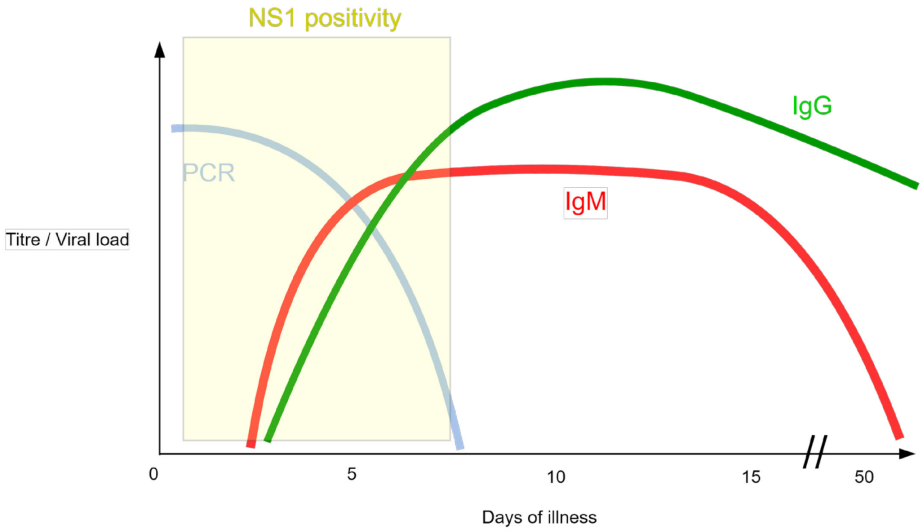
The dataset for model development and evaluation was derived from a paediatric cohort recruited in Southern Vietnam. Children aged between 1 and 15 years old inclusive presenting from the community with an undifferentiated AFI consistent with dengue were enrolled from 7 healthcare facilities around the Ho Chi Minh City area, Vietnam. The study was performed between 12th March 2010 and 10th December 2014, and informed consent was obtained from caregivers.

Enrolled patients underwent recording of their demographic and clinical information through a standardised case record form at clinical presentation. Blood samples for full blood count, biochemistry tests including renal and liver function tests, a NS1 rapid antigen test (NS1 Ag STRIP, Bio-Rad, USA), and dengue-specific reverse-transcriptase polymerase chain reaction (RT-PCR) from blood were collected at enrolment during their initial presentation to the healthcare facility.

5.6.1 Dengue outcome definitions

The diagnostic criteria for acute dengue were in line with those used in previous studies included in Chapter 4. A diagnosis of dengue infection was defined as one or more of the following: i) Positive RT-PCR, ii) Positive NS1 result in blood, or iii) IgM seroconversion (Panbio, Australia) for patients who were admitted to hospital. Patients with an AFI without any positive dengue-specific laboratory tests were diagnosed with an OFI. An outline of expected test results regarding RT-PCR, NS1 detection and serology are shown in Figure 5.1:

Figure 5.1 Expected test results for dengue diagnosis according to the day of symptomatic illness. Viraemia and the presence of NS1 in blood is usually short-lived and may disappear prior to the critical phase (typically days 4-6 of illness).



5.6.2 Feature selection of models

Features included in model development were selected on the basis of availability. These features have to be sufficiently basic and routinely-available to ensure they are available in a variety of LMIC healthcare settings and facilitate subsequent implementation. The final demographic and laboratory features chosen for the models were: patient age, gender, day of illness and full blood count results on enrolment (haematocrit, platelet, white cell and lymphocyte count), with full blood count investigations constituting an essential investigation in healthcare facilities within the WHO essential diagnostics list [249].

5.6.3 Model development procedures

A binary classifier accounting for patient features obtained on enrolment was developed to predict the final diagnosis of either dengue or OFI. A gradient boosting machine learning algorithm, XGBoost was chosen [47] – this is a robust, ensemble tree-based algorithm which accounts for missing values imputation. There was missing data for 8 (0.1%) patients and additional imputation was not performed in this study. The entire dataset underwent an initial stratified random split into 80/20 ratio to form a training and hold out test set respectively, with the latter not used in model development at any stage. Hyperparameter optimisation of the model was done through grid search using 10-fold cross validation, and aspects of the model including child weight, max depth of each

tree, loss reduction for partition (gamma), step size shrinkage (eta) and number of trees constructed (n_estimators) were iteratively tested.

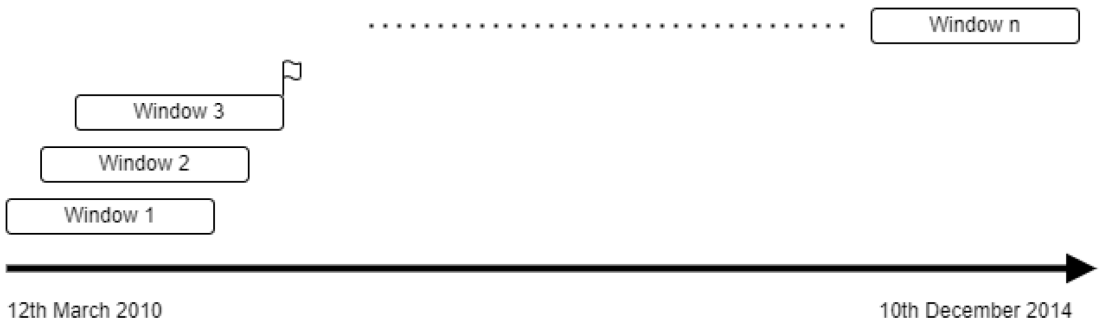
5.6.4 Characterising the relationship between outcome and seasonality

The temporal association between seasonality, diagnosis of OFI:dengue and model performance were examined visually. Average precipitation and temperature values at a country level for Vietnam were obtained from the World Bank and used as a reference [250]. To examine the association between seasonality and model performance, models were constructed whereby these precipitation and temperature values were included as additional features with a varying lag period. Rolling windows of varying size (28, 21, 14 or 7 days) were also generated. These rolling windows include all patients contained within the time period and were advanced forward by a day at each iteration until the end of the study period. Within each window all cases of AFI recruited within this timeframe were included and used as the test set, evaluated against a model trained on all data *not* included in this particular timeframe – this is equivalent to a leave one group out cross validation approach [251].

From this method either a constant, or dynamic threshold model was incorporated. For the constant model positive classes of dengue were defined when probabilistic output for the model is >0.5. For the dynamic model we took the assumption that the qualitative and quantitative nature of AFI presenting temporally closer were more similar to those presenting further in time from each other. Through iterative adjustment of the decision threshold based on the characteristics of the *preceding* data window, the true positive rate (sensitivity) or false positive rate (1/specificity) was tuned to achieve the desired PPV and NPV target values. As the primary clinical role for the model was the safe triage of patients at low risk of dengue for community follow-up and to reduce hospital admission, as well as to prioritise clinical evaluation for those with a higher probability of dengue, the negative predictive value of the model was fixed to achieve a performance above 90%. This process is described and illustrated in

Figure 5.22.

Figure 5.2 Dynamic threshold model: The entire study period was divided up into rolling windows of fixed periods (7, 14, 21 or 30 days) which advanced by 1 day until the end of the study period. Models were trained using data from all available data with exception of observations contained within the particular window through a leave one group out cross validation approach. The trained model was then evaluated against that particular window and specific thresholds for that window were iteratively tuned to achieve a NPV > 90% so that each window was associated with a unique threshold. For any new prediction (flag) the threshold applied to that decision would correspond to that of the immediately adjacent window – window 3 in the case of the figure.



Further methodologies to evaluate and incorporate the contribution of seasonality in terms of model characteristics and model performance are described in their respective subsections.

5.7 Results

5.7.1 Description of demographics and seasonal variation

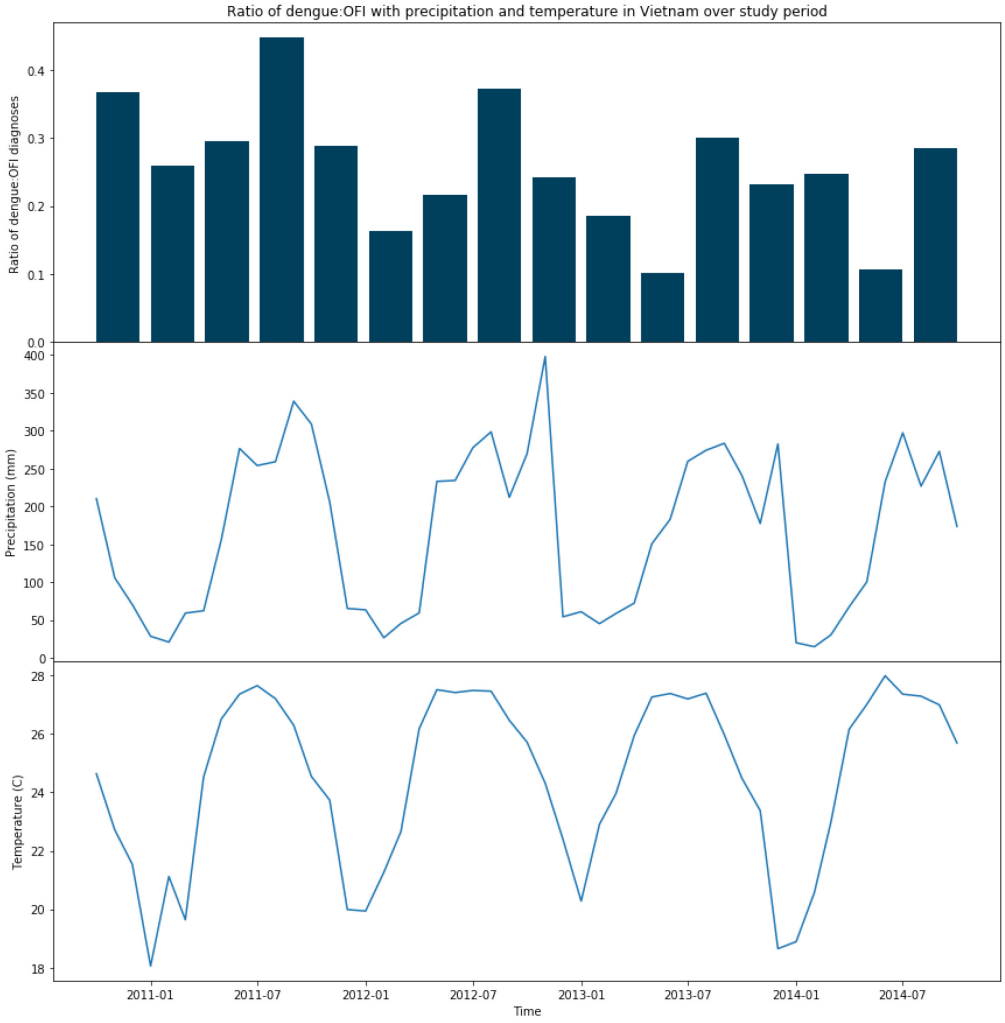
In total 8,100 children recruited between 12th March 2010 and 10th December 2014 were included in the analysis. The median age was 6 years old (Interquartile range, IQR 3 to 9 years), 44% were female (3,541/8,100) and the median duration of fever at enrolment was 3 days (IQR 3 to 4 days). Those with a laboratory-confirmed dengue diagnosis constituted 27.6% (2,244/8,100) of the population: 2,186 had a positive RT-PCR, 1,134 had a positive NS1 and 973 had a positive IgM and/or paired serology. Patients with OFI were significantly younger (5 vs 9 years old), had a shorter duration of illness (3 vs 4 days) and had different full blood count differentials: lower haematocrit, higher platelet count and higher white cell count (p values all <0.01). Results are shown in Table 5.1.

Table 5.1 Baseline characteristics and results of investigations at enrolment of the analysis cohort, separated according to final diagnosis.

Parameter	All patients (n=8100)	Dengue (n=2240)	OFI (n=5860)	p-value
Median age in years (IQR)	6 (3-9)	9 (6-11)	5 (3-8)	< 0.01
Duration of illness (days)	3 (3-4)	4 (3-4)	3 (2-4)	< 0.01
Female sex	3541 (44%)	976 (44%)	2565 (44%)	0.90
Haematocrit (%)	38 (36-40%)	39 (37-41%)	37 (35-40%)	< 0.01
Platelet count (cells x 10^9/L)	227 (181-278)	180 (141-225)	243 (202-292)	< 0.01
White cell count (cells x 10^9/L)	7.7 (5.2-11.2)	4.9 (3.6-6.9)	9.0 (6.4 – 12.5)	< 0.01
Lymphocyte count (cells x 10^9/L)	1.8 (1.1-2.8)	1.1 (0.8-1.6)	2.1 (1.4-3.2)	< 0.01

The ratio between dengue and OFI diagnoses of the cohort demonstrate a cyclical pattern over the study period (between 2010 and 2014). In general, a higher proportion of patients were diagnosed with dengue during periods of higher rainfall consistent with understanding of how climate affects vector behaviour and disease incidence. This relationship is shown in figure 5.3:

Figure 5.3 Ratio of dengue:OFI over time with average rainfall and temperature in Vietnam over the analysis period. In general rises in rainfall and precipitation during May – November are associated with relative increases in diagnosed dengue cases compared with OFI.



5.7.2 Baseline model performance for predicting acute febrile illness diagnosis

A random stratified 80/20 split was applied to the full cohort to generate a training set (n=6,480 with 27.2% dengue) and a hold out test set (n=1,620 with 27.2% dengue). Only the training set was used for model development through 10-fold cross validation. The median performance in cross-validation of the model using XGBoost after hyperparameter optimisation was as follows: AUROC of 0.86 (Interquartile range 0.84-0.86), specificity 0.92, sensitivity 0.56, positive predictive value (PPV) 0.73 and negative predictive value (NPV) 0.84. Isotonic calibration with 10-fold cross-validation was used and the Brier score was 0.13 demonstrating good calibration (Table 5.2).

Table 5.2 results of model 10-fold cross validation and hold out set evaluation.

	Sample size	**Dengue diagnoses**	**AUROC**	**Specificity**	**Sensitivity**	**Positive predictive value**	**Negative predictive value**	**Brier score**
Cross validation	6480	1792 (27.7%)	0.86 (IQR 0.84-0.86)	0.92	0.56	0.73	0.84	0.13
Hold out set testing	1620	441 (27.2%)	0.86	0.78	0.79	0.58	0.91	0.13

This optimal iteration of the model was then evaluated against the independent hold out test set. An AUROC of 0.86, specificity of 0.78 sensitivity of 0.79, PPV of 0.58 and NPV of 0.91 was attained in predicting dengue vs OFI outcomes. Examining feature importance based on f-score (weight), which is metric of how many times a particular variable was split on, the most important features were white cell count, lymphocyte count and platelet in descending importance. These results are shown in Figure 5.4, and the calibration plot indicating the match between probability and frequency of outcome shown in Figure 5.5.

Figure 5.4 Feature importance of the XGBoost model derived from the training set (left) and model receiver operating characteristics (ROC) curve (right)

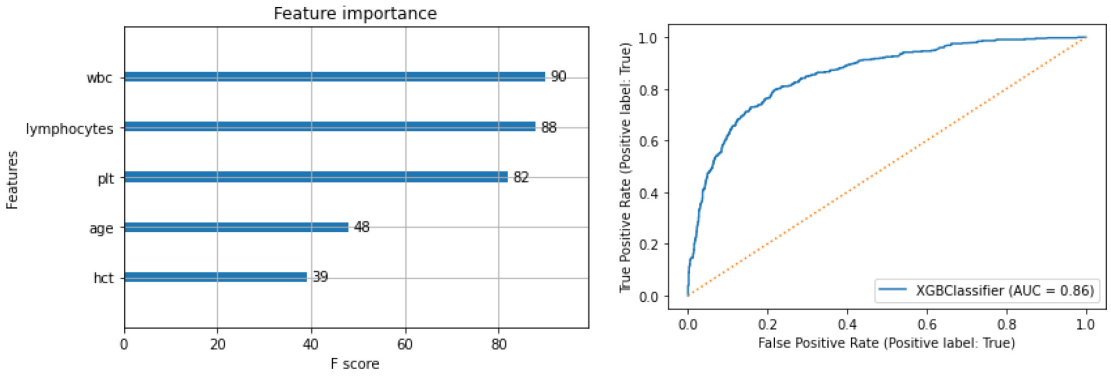
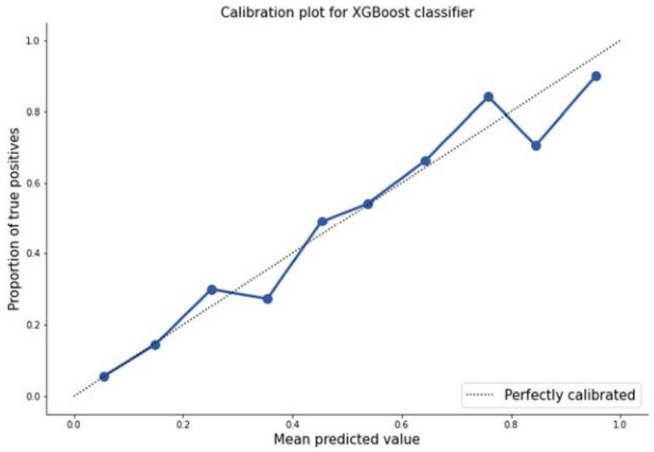


Figure 5.5 Calibration curve of the final model applied to the hold out test set



5.7.3 Performance of the model when applied over the study period

To explore the change in model performance over time, a rolling-window approach was adopted (Chapter 2 – methods). The windows needed to be of sufficient size to include patients of sufficient variance within – applying a 30-day rolling window across the dataset resulted in generation of 1,303 windows for the entire study period with a mean of 180 patients included per window.

When the leave one group out cross validation was applied to this dataset using model hyperparameters derived in the previous section, the median AUROC in differentiating dengue from OFI diagnoses was 0.86 (IQR 0.83-0.88) and significant variation in PPV (median 0.73, IQR 0.60 to 0.83) and NPV (median 0.87, IQR 0.84-0.92) was seen over time. In general, PPV was highest when the prevalence of dengue was highest and NPV highest when prevalence of dengue was at its lowest which fits with the hypothesis that pre-test probability influences model performance in a similar way as a traditional dengue diagnostic.

In order to account for variation in model performance, a varying threshold which changes with time was applied to the model. The predictions produced by the XGBoost classifier used was probabilistic and outputs a continuous prediction value between 0 and 1. The threshold is a value which determines the cut-off point for the binary classification output (i.e., dengue vs OFI). Conceptually where pre-test probability is low, such as during dry season, the threshold for a positive prediction of dengue should be higher to account for an increased number of false positives. The adjustment of thresholds is however a trade-off as a higher threshold would compromise the ability of the model to classify true positives. Using this approach only *one* aspect of the model metric can be maintained so that its performance remains consistent for use throughout the year. The NPV metric was chosen for optimisation as the important clinical role for the model is to reliably exclude dengue in patients with an AFI and support effective triage. A value of >90% throughout the year was chosen as a target which would provide sufficient confidence in the model. These thresholds are dynamic given change over time, and approached empirically or through inclusion of climatic variables as features (Figure 5.5).

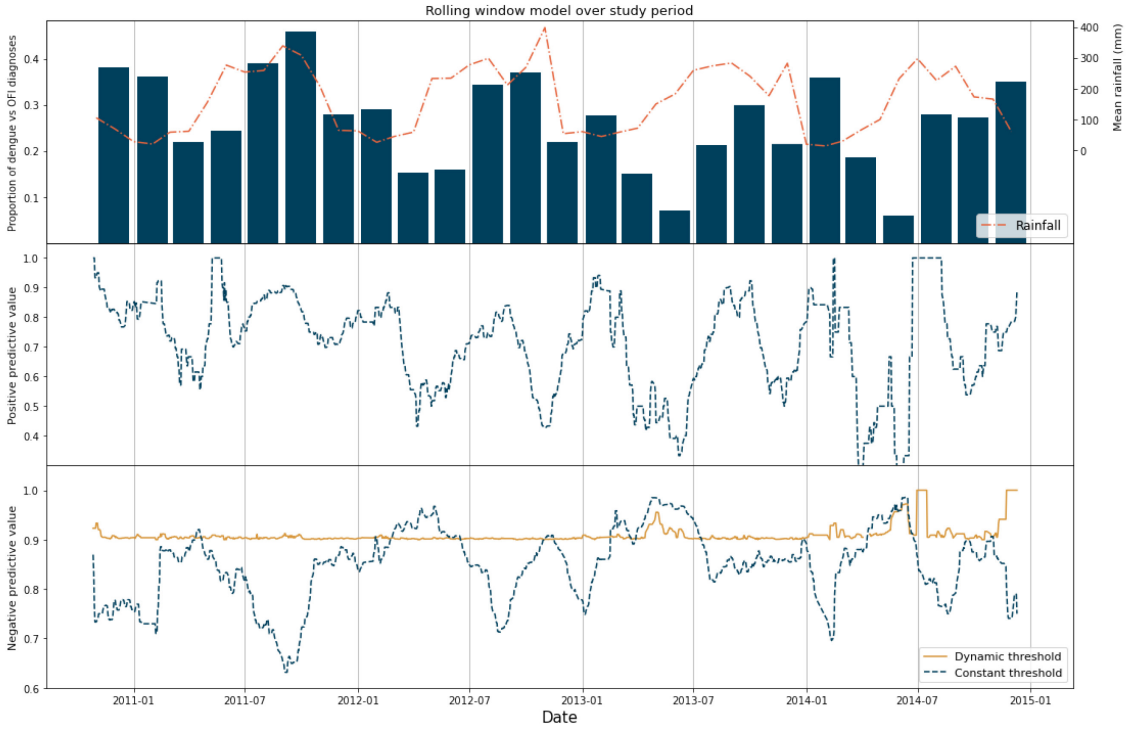
5.7.4 Empirical tuning of thresholds over time

The assumption for this model is that qualitative and quantitative nature of the patients presenting with an AFI presenting closer in time were more similar with each other compared with those presenting farther away. For a given prediction at a point in time within this model, the threshold for its prediction is set by the optimal threshold in the closest previous adjacent window which was in turn tuned to achieve a NPV >90%. This was done iteratively through an empirical search of the threshold range between 0-1 until the desired metrics were achieved. New predictions would therefore account for the case mix of the patients enrolled in the immediate past 30-day window as a form of continuous adjustment.

A consistent NPV was attained using this approach over time using a 30-day rolling window – median NPV 0.9 (IQR 0.90-0.91) at the expense of a changing PPV (median 0.66). The respective NPV and PPVs are shown graphically in

Figure 5.6. Changing the window to 21, 14 and 7 days respectively did not result in statistically significant differences in predictive values compared with 30 days (see supplementary results).

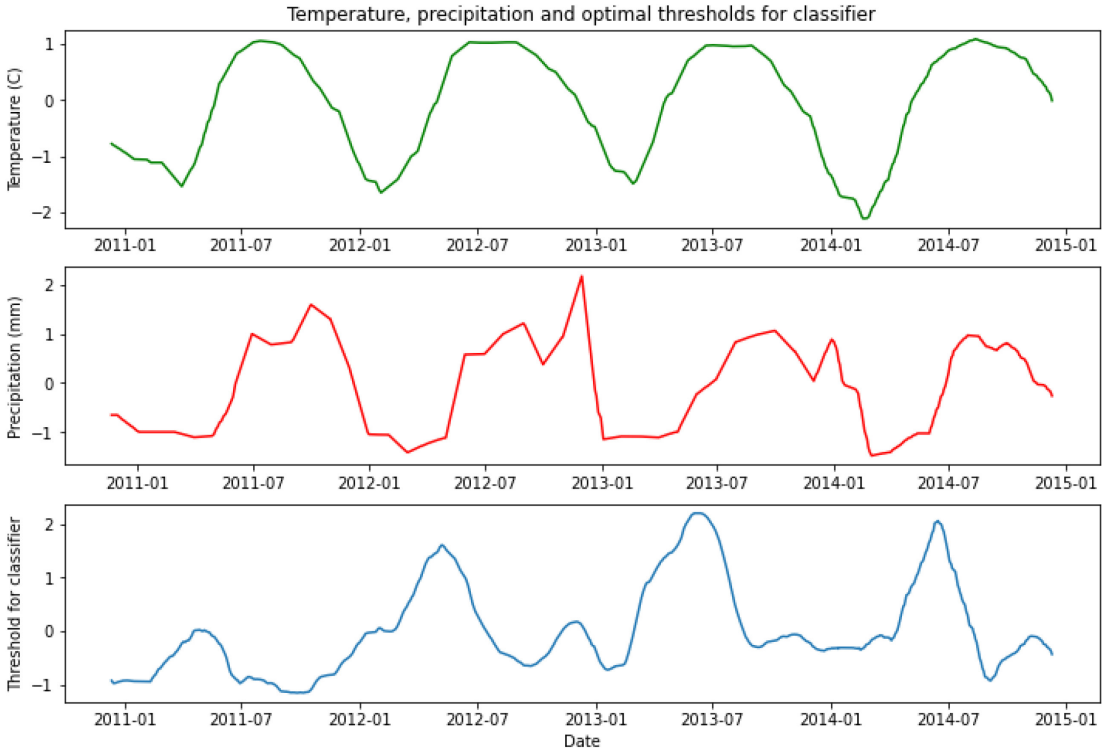
Figure 5.6 Empirical dynamic model over study period. Top: Proportion of dengue diagnoses grouped in 2-month bins plotted against mean rainfall in Vietnam for the entire study period. Middle: Positive predictive value of the rolling window model for the constant threshold model only, with each point on the graph representing the upper range of the 30-day window used as the test set. Bottom: Negative predictive value of the rolling window model for the constant (blue) and dynamic model (orange), with the NPV of the latter model fixed to > 90%.



5.7.5 Climatic model to inform changing thresholds

The resulting optimal thresholds in this model varied with a seasonal pattern over time. Visually, the thresholds exhibit an inverse relationship with climatic features: they were highest during periods of lowest rainfall and temperature and lowest during dry season. A plot of thresholds in relation with temperature and precipitation is shown in 5.7:

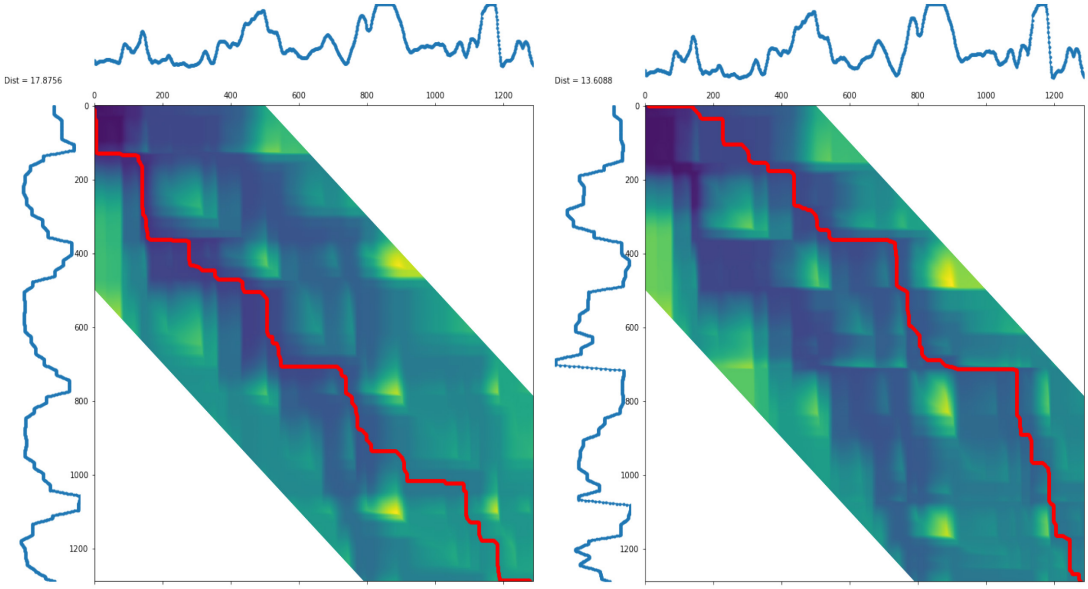
Figure 5.7 Plot of temperature, precipitation and optimal thresholds from the empirical dynamic model. Temperature and precipitation data were derived from the World Bank dataset for the whole of Vietnam. These data have been smoothed with a rolling window average function.



In order to explore and understand the temporal relationships between rainfall, temperature with these optimal thresholds, a dynamic time warping (DTW) method [252] was applied in order to quantify similarities between the data. DTW derives an optimal match between two temporal sequences which might vary in terms of morphology or lag to derive a distance metric. This distance metric is a quantification of cost of matching these sequences and therefore representative of similarity.

The distance between precipitation and thresholds was 13.6 and between temperature and thresholds 17.9 indicating that the former was more similar to each other. The warping paths generated, which allows for transformation of one sequence to another are shown in Figure 5.8:

Figure 5.8 Warping paths (diagonal red line) produced by dynamic time warping (DTW) between temperature and thresholds (left) and between precipitation and thresholds (right). The paths traverse a cost grid and represent the transformations needed to match one sequence with another. The greater the deviation from the diagonal, the greater the cost and resulting overall warping distance.

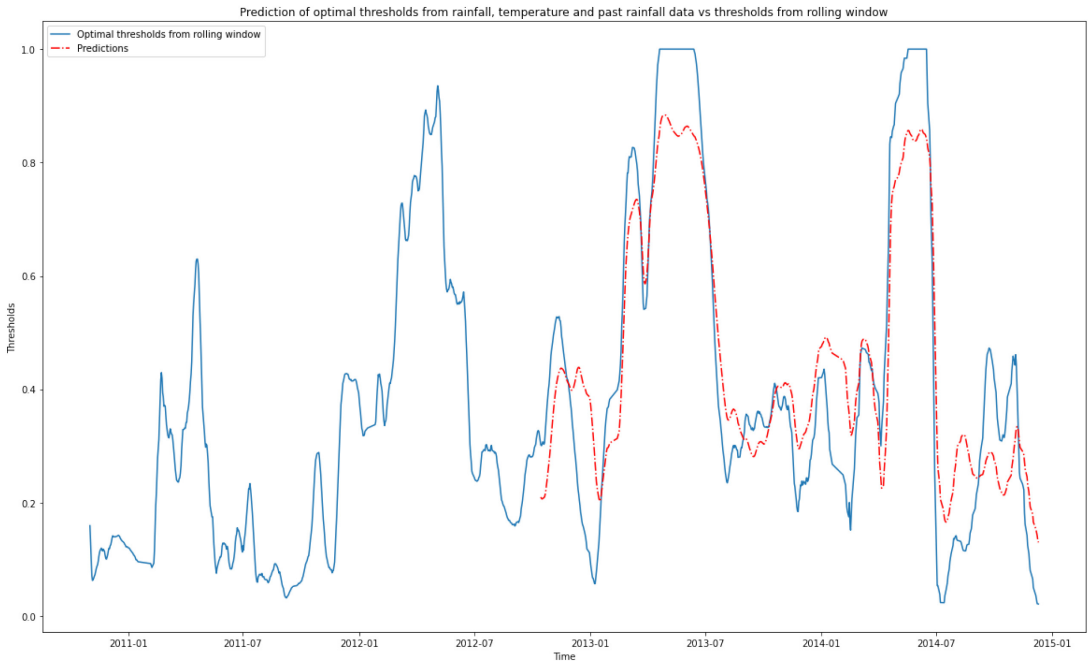


Given that there exists similarity between these data supporting an association (which is likely causal) between climatic factors, a second model which incorporated climatic variables directly to predict these optimal thresholds over time was constructed. One hypothesis is that changes in local rainfall and temperature affect *Aedes* behaviour and reproduction, in turn affecting dengue incidence, or at least the observed dengue:OFI ratios (through its effect on the OFI group). In order to predict threshold sequences, a model which only takes rainfall and temperature as exogenous features to predict future threshold (and maintain a NPV >90%) was developed and evaluated:

The entire data period was split 50/50 chronologically – all rainfall and temperature data between October 2010 and August 2012 in the first half were used for model training. The model was then used to predict optimal thresholds for August 2012 to Dec 2014 which formed the hold out test set. A deep-learning approach (see the methods chapter and Chapter 3 for further details) was adopted for univariate time-series prediction, given that this was bias-free and would represent an empirical approach without requiring underlying assumptions about the data. Using a Long short-term memory (LSTM) model with two layers with 20 neurons each, and a dropout rate of 0.2 and a validation_split of 0.2 (from the training set only) and mean absolute error (MAE) as the loss metric. The validation set used was formed from 20% of the training set. At epoch 163 the training

loss was 0.054 and validation loss was 0.056. The model was then tested on the hold out test set and predicted climatic thresholds are shown graphically in Figure 5.9:

Figure 5.9 Prediction of future thresholds to achieve a NPV > 90%. The optimal thresholds derived from the empirical tuning approach (blue continuous) are shown. Rainfall and temperature data taken from the World Bank during the first half of the study period (March 2010 – August 2012) was used in the LSTM climatic model in order to predict thresholds for the second half (dashed red).



These predicted climatic thresholds were then applied as actual cut off values in the leave one group out cross validation (see 5.5.3). These thresholds showed an improvement over the constant threshold approach in terms of maintaining the NPV > 90% (median NPV for the climatic thresholds 0.91, IQR 0.89 – 0.93 vs 0.87) but were not as consistent as the empirical dynamic model. A further iteration of the model whereby climatic variables were directly incorporated as feature variables was done, but this did not result in any difference in model characteristics compared with the baseline models (data not shown). The results of the thresholds from the constant, empirical and climatic models are shown graphically in Figure 5.10 and summary statistics of NPV shown in Table 5.3

Figure 5.10 A plot of negative predictive values achieved in the latter half of the study (as the hold out test set) using the 3 approaches. The empirical rolling window approach which constantly adjusted the threshold by each window achieved the most consistent NPV (blue continuous). The LSTM climatic model (red dashed) resulted in fewer fluctuations from the target NPV compared with the constant static threshold model (green continuous).

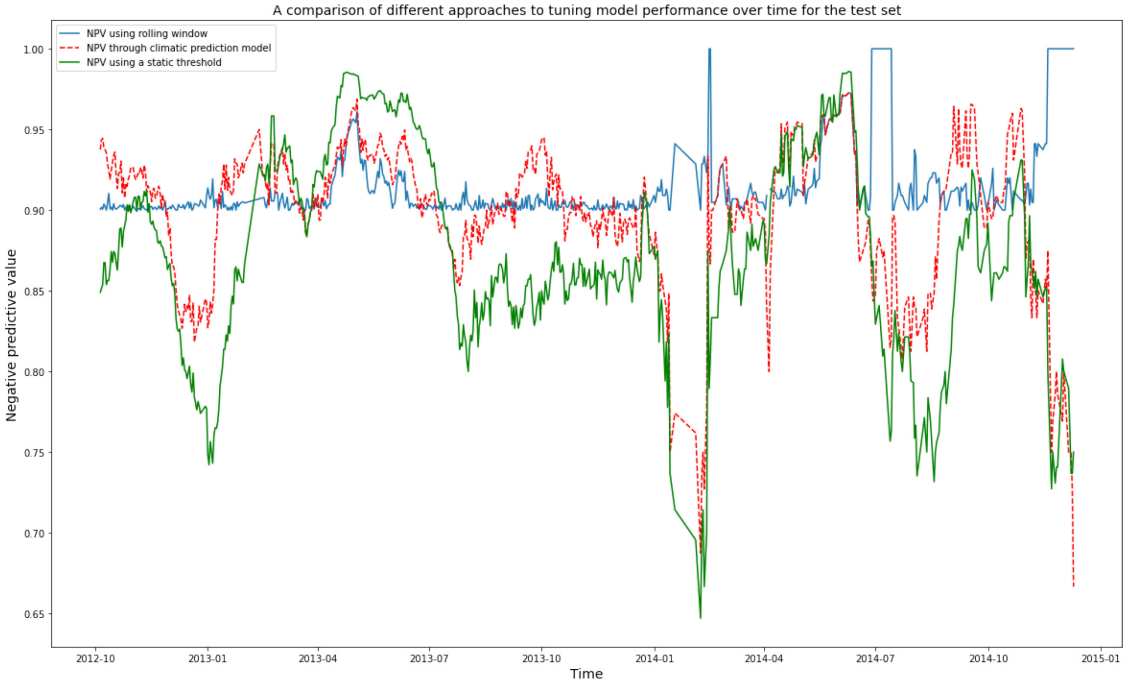


Table 5.3 Summary statistics for NPV performance for the three different models.

	Rolling window adjustment	LSTM climatic model	Constant model
Minimum	0.90	0.67	0.65
Quartile 1	0.90	0.89	0.84
Median	0.90	0.91	0.87
Quartile 3	0.91	0.93	0.92
Maximum	1.00	0.97	0.99

5.7.6 Understanding feature distributions of patients diagnosed with other febrile illness

The ratio of dengue to OFI infections for each time period in the study is assumed be a proxy for dengue prevalence in the region. This value varies with time over the course of the year, and is associated with changes in model characteristics in terms of predictive values. However, it is likely that the epidemiology of OFI i.e., the causative organisms

which make up the OFI group also changes over the course of the year, given many AFI are caused by pathogens which exhibit seasonality in incidence (possible OFIs which are known to result in an AFI in the region include seasonal respiratory viruses. leptospirosis, scrub typhus, bacterial infections etc.).

Changes in OFI distribution might in turn affect model performance: a model trained on, and is adept at differentiating dengue from a bacterial infection may perform differently when in classifying between dengue and another viral infection (e.g., influenza). As these incidences of bacterial and viral infections likely change throughout the year, the result is seasonal variation in the ability for the model to discriminate between dengue and OFI. This is observed when a leave one group out cross validation approach was used, resulting in a median AUROC of 0.86 (IQR 0.83-0.88). There is variation in AUROC over the study period, although this does not exhibit as clear a pattern as the optimal thresholds – this is displayed graphically in Figure 5.11.

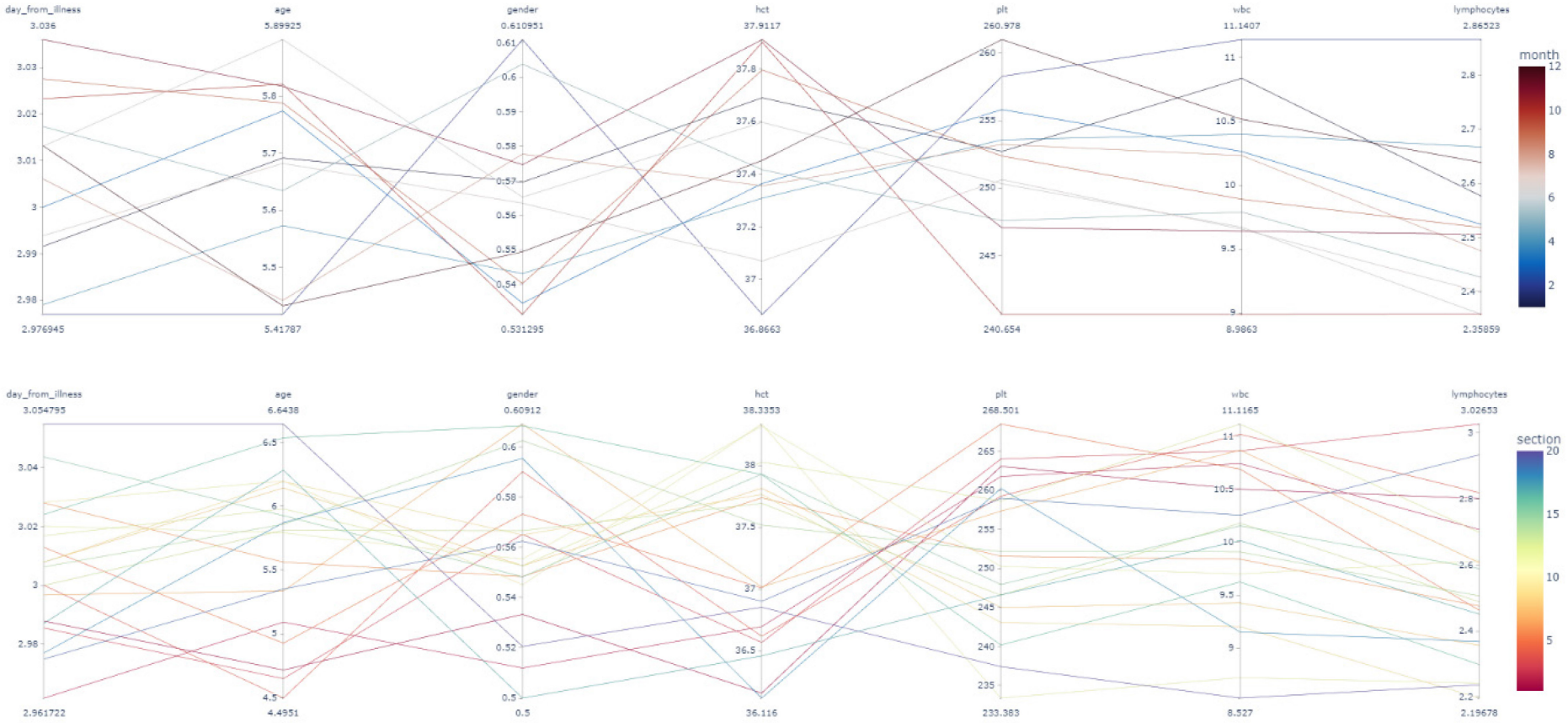
Figure 5.11 Area under the receiver operating curve (AUROC) of the model over the entire study period in classifying dengue and OFI in patients. There is variation in performance with significantly deteriorated scores in Q2 2014.



Given the limitation of the data (OFIs were diagnosed on the basis that they were AFI which were not confirmed to be dengue), feature variables namely age, gender, full blood count and day of illness in this group of patients were used as a proxy to understand change. The assumption taken was that differences in aetiology might also result in different distributions of feature variables.

Feature variables were grouped according to month of the year and into 3-month window sections in relation to the overall study period. The plots of features values show different univariate distributions of each feature variable with month or window sections, although no significant patterns can be inferred visually (see Figure 5.12).

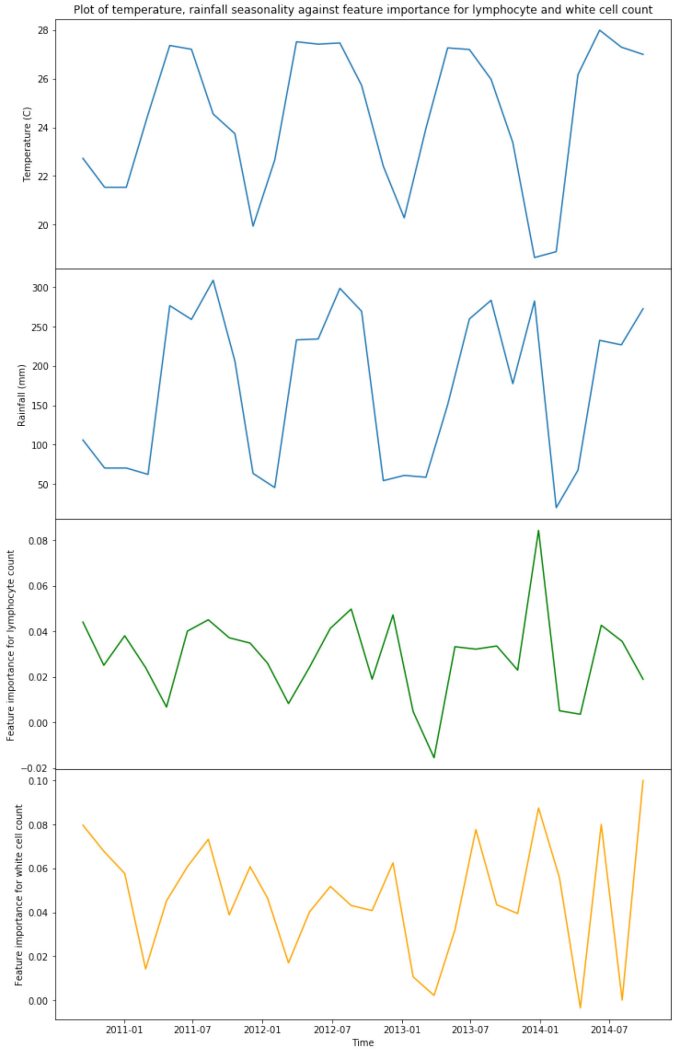
Figure 5.12 Parallel coordinates plots showing the relationships between features in the OFI group of patients. Each line is coloured by month (top) or by one section of the entire study period which has been divided into 20 equal sized periods (bottom).



To account for interactions between features in terms of similarity/dissimilarity, an unsupervised approach was used to see if features from specific months or window sections clustered together. Smaller distances between clusters derived from features temporally closer to each other (months, or window sections contiguous with each other) would support the hypothesis that OFI also exhibits seasonality in terms of aetiology. Through Principal Component Analysis (PCA) [72] feature variables underwent dimensionality reduction and were plotted according to their month or period. Clear differences in clustering patterns were not observed (see supplementary results).

A separate approach in identifying if subsets within OFI group was done through feature importance. Given that the baseline model was trained on features with the aim to discriminate between dengue and OFI labels – were there periods in which specific features contributed to the model to a greater extent compared with other periods? Permutation importance was used as a means to quantify the contribution of each feature [59]. For a trained model, feature values in the test set were randomly shuffled. Decreases in model performance as a result of shuffling, compared with baseline is indicative of feature importance – i.e., features of little importance when altered will not result in changes in performance. The dataset was split into n window sections and with the model trained on n-1 sections, permutation importance was then calculated for the test section not involved for training. The features white cell and lymphocyte count (ranked as the most important features respectively in the model) demonstrated variations with time in terms of importance with higher contributions generally during rainy season (July-November). These results are displayed graphically in Figure 5.13:

Figure 5.13 Permutation importance of lymphocyte and white cell count against time for the study period, displayed alongside temperature and rainfall data in Vietnam



5.8 Discussion

Diagnosing AFI is important for patient management and informs public health interventions. In this chapter, a machine learning model was trained on a large cohort (n=8,100) of paediatric patients presenting with an undifferentiated febrile illness at healthcare facilities in Southern Vietnam. Using basic laboratory and clinical information collected on enrolment and early presentation (median day 3 of illness), the model was able to predict patient diagnosis between dengue or OFI. The baseline model performed well in cross validation and hold out set testing (AUROC 0.86 for both) and demonstrated good calibration (Brier score 0.13) appropriate for clinical use.

Although use of measurement techniques based on antigen, nucleic acid detection or serological measurement are vital in supporting clinical diagnosis for infections – challenges in making such technologies available across many healthcare settings remain numerous. Staff training, quality assurance, equipment maintenance and supply chains relating to laboratory consumables represent barriers to quality pathology and laboratory services. Use of data-driven methods which take in basic routine clinical and laboratory features could augment clinical diagnoses and provide added value without incurring significant costs. Once suitably trained, such models can run on basic computers and mobile smart phones through on-device machine learning approaches. It is important to note that implementation should ideally incorporate prospective evaluation for model performance and clinical utility. This is particularly true where such models are intended for application in another setting where healthcare and patient factors might differ significantly. The chapter supports the role of data as a adjunct, or an alternative to conventional pathogen diagnostics which are based on the detection of specific compounds.

For dengue, a system which provides a high negative predictive value could allow for efficient patient triage through management pathways or a tiered NS1 testing strategy, with the aim of reducing clinician workload whilst maintaining patient safety. The latter aspect is particularly crucial given the high numbers of patients who present to healthcare during rainy season in endemic settings [253]. Intrinsic digital connectivity and interoperability between diagnostics, laboratory and electronic clinical systems is crucial [254] allows for data and model predictions to directly inform public health interventions. A passive-surveillance system to ascertain dengue cases using routine laboratory data would guide specific interventions such as vector control, prepare staffing and other dengue-specific measures. Such systems provide a bridge between data and diagnostics to

support clinical decision-making. Appropriate integration into the clinical pathway could therefore provide an additional layer of safety.

Understanding how the model changes with time is crucial to its clinical implementation. There was a significant drop in model performance between February and August 2014 (

Figure **5.11**). Univariate approaches to examining if the patients enrolled within this period were different from those in other periods did not demonstrate significant differences in their feature variables (data not shown). There could be numerous explanatory factors which might contribute to this, including outbreaks of other viral infections e.g., there was a measles outbreak in Ho Chi Minh City during this period **[255]**, changes in dengue serotype and resulting disease severity or differences in healthcare access during this period. Given the lack of data on OFI these are all speculatory but highlights the importance of understanding changes over time, and the need for continuous learning to maintain consistent model performance. Understanding the make-up of the OFI group would contribute to this.

There was significant variation in model characteristics over the year which was associated with seasonal changes in temperature and rainfall. The working assumption is that these climatic factors change vector behaviour and in turn affect dengue prevalence. The change in prevalence changes the pre-test probability and ultimately the characteristic of the diagnostic (in this case a model). An empirical system which optimises the threshold according to characteristics of the cases from a recent cohort of patients was one approach used in this chapter, to ensuring predictive value remains constant and reliable. A contrasting time-series forecasting method which utilises average temperature and rainfall as features was also able to predict a dynamic optimal threshold which improved performance over a constant model. These findings indicate that taking into account temperature and rainfall, as an example of diverse but important non-healthcare data, is crucial for any predictive task for seasonal AFI. Although models developed for infectious diseases forecasting commonly utilise these data – work presented in this chapter bridges the gap between disease prediction and individual patient treatment and is the only model published to date which is suitable for, and intended for supporting clinical decision-making. In a resource-limited setting in which clinical data availability might be sparse, the inclusion of such adjunctive information and other secondary data might be important in improving the performance of clinical models.

Understanding the relationships between disease and climate will be increasingly crucial in the context of human-induced climate change. Changing disease prevalence, timings of

outbreaks, new geographical areas at risk will likely be observed in the next decade [246]. As an urban disease, local physical environments within pockets of densely-populated areas and human factors such as behaviour and mobility play a significant role in disease transmission. More granular data is likely required to aid understanding of transmission, prevalence and their impact on clinical cases [256].

The strengths of the study include large sample size, small proportion of missing data, use of robust evaluation methods including cross-validation and hold out set testing, which increases the validity of the findings. Good calibration of the model is also crucial to inform decision-making as it provides a true representation of probability. The model features utilised were basic clinical and laboratory results which increases its applicability especially in settings where basic diagnostics and investigations are available. Being able to account for changing model characteristics and adjust thresholds over time ensures the reliability of the model but is also crucial in clinical implementation and fostering trust in the system.

Limitations of the work include its retrospective nature, focus only on paediatric patients and the single-region nature of the study. The study recruited from 7 healthcare centres located in Southern Vietnam and are likely to share similarities in patient presentation and management – it would be important to evaluate its external validity through deployment and testing in another healthcare setting located in another dengue-endemic country. The OFI group was defined as patients with an AFI which did not fulfil dengue diagnosis; it is possible that false negative cases are contained within. Characterisation of OFI cases through specific diagnostics would add to the strength of the model and also allow for a multi-class predictor which could be of greater clinical utility.

In conclusion a supervised machine learning approach was used to predict diagnosis in patients presenting with an AFI. The time-series nature of the predictions was important and associated with changing seasonality. Accounting for these as well as developing more in-depth understanding on how climate interacts with infectious diseases and clinical management is crucial to ensure healthcare systems can adapt to these changing burdens in the future.

Chapter 6 - The implementation of machine learning models utilising dynamic time-series data, through healthcare pathways mapping and understanding of decision-making

6.1 Chapter overview and summary

The chapter presents learnings on the implementation of a machine learning model to the clinical setting within a low-and middle-income context. This work represents the clinical translation of machine learning models developed for the management of dengue in Chapter 4 and Chapter 5. I utilised a set of qualitative methods in order to better characterise the care setting in terms of healthcare processes, needs and features important to clinical decision-making for dengue. These findings were analysed through thematic analysis and used to support the initial stages of developing a CDSS.

Translation of research findings which result in clinical impact is crucial. An approach for machine learning models to be deployed in healthcare settings is through their inclusion in electronic CDSS. These tools provide a digital interface to healthcare staff and display relevant information by the patient bedside with the aim of improving decision-making quality during patient encounter. However, their implementation requires understanding of existing patient pathways, decision-making and design of interfaces which suit needs of end-users for the intervention to be taken up and sustained.

To better understand the context and cognitive processes involved in the management of dengue at the HTD, Vietnam, I made use of methods including ethnography, process mapping, semi-structured interviews, thematic analysis and co-design workshops. A series of CDSS design workshops were then carried out to establish understanding on what constitutes an intervention which would likely be taken up.

My findings included identifying that a triage-based CDSS would be best suited in the outpatient department, where the majority of dengue patients are managed. The work identified the importance of clinical signs, and role of established protocols in clinical decision-making at the hospital, and existing positive attitudes towards digital interventions for clinical management. Workshop activities further clarified adjunctive functions of a CDSS expressed as desirable for clinician use in routine patient care. Wireframe mock-ups and user interfaces were developed which form the basis of ongoing work.

Chapter conclusion: there is significant support that characterisation of context is important for successful implementation in healthcare. I present implementation research findings important to the design and roll-out of a CDSS in this chapter which help the translation of machine learning models to support decision-making in dengue. Such methodologies are also likely transferrable, and applicable to the introduction of other novel digital interventions.

6.2 Implementation of machine learning models to the healthcare setting

In the era of digital health, availability of detailed patient data to support decision-making holds promises for improving healthcare. Models which can meaningfully utilise clinical information, and capture time-series relationships have been presented in preceding chapters. These approaches allow for greater individualisation in clinical management of infections, through taking into account diverse, individual data not otherwise possible, complementing existing EBM medicine frameworks [257].

For any research it is crucial that their findings and model outputs are presented in such a way which ultimately results in clinical impact. The use of risk factors and clinical scoring systems provide a means to apply population findings from clinical studies to the individual patient. But for more individualised decision-making, the integration and presentation of complex information to end-users is crucial. One method for encapsulating these needs is through clinical decision-support systems (CDSS) [258]. These systems are well-suited for the digital medium and delivered through computers, tablets and phones, providing relevant information at point of patient contact. The aim of these systems is to work alongside clinical decision-making to improve the quality, consistency or efficiency of care delivered. Strengths of these approaches are that they make available information which is otherwise not intuitive such as through interpreting large quantities of data, are timely to collate such as through summarising electronic healthcare records and provide a layer of safety for routine tasks for example, by highlighting drug interactions. Such CDSS have shown to improve the quality of healthcare through reduction in prescribing errors [259], costs [260], increase adherence to guidelines [261] and improve targeted aspects of care within local contexts [262]. For infectious diseases specifically, the scope of CDSS include those aimed at augmenting risk stratification, diagnosis, treatment selection, predicting outcome for bacterial and viral infections [263]. However, it is important to note that only a small overall proportion (for example in the review referenced, only 5%, n=3 of studies) underwent prospective evaluation in real-world clinical settings.

This problem of translating research findings into clinical practice is common. Often highlighted in existing literature, the implementation gap is significant [264] and challenging to address for multiple reasons: performance and accuracy of such models, which often utilise curated datasets, may lack external validity. Contextual factors relating to staff training, healthcare pathways, behaviour, societal factors and resource availability

also mean that adoption of interventions do not necessarily result in clinical impact. The uptake of any interventions can be short-lived and not sustained outside a research environment [265]. Although desirable properties of an ideal CDSS for clinical impact have been described in the literature (Table 6.1) it remains important that the context and intended environment of use are fully taken into account to ensure success.

Table 6.1 Ten Commandments for Effective Clinical Decision Support, from [266].

Ten Commandments for Effective Clinical Decision Support
1. Speed Is Everything
2. Anticipate Needs and Deliver in Real Time
3. Fit into the User's Workflow.
4. Little Things Can Make a Big Difference.
5. Recognize that Physicians Will Strongly Resist Stopping.
6. Changing Direction Is Easier than Stopping.
7. Simple Interventions Work Best.
8. Ask for Additional Information Only When You Really Need It.
9. Monitor Impact, Get Feedback, and Respond.
10. Manage and Maintain Your Knowledge-based Systems.

Health systems research, or implementation research represent a set of approaches to support the delivery of research findings. Common to the research themes is the emphasis on better characterisation of context and existing processes at the healthcare setting – ensuring that the interventions in question can address any issues encountered. For many complex interventions this will involve consideration of the interactions between the intervention and the environment, quantifying and addressing uncertainties, involving stakeholders in roll-out as well as anticipating (otherwise) unintended consequences on safety, resource use and outcome [267, 268]. Ideally any model intended for clinical use should incorporate the main elements of health systems research from the offset to ensure that end results are appropriate, and applicable for real-world settings.

The application of machine learning models for use in low- and middle-income country (LMIC) healthcare settings adds to implementation complexity [269]. Direct transfer of technological interventions through a top-down approach is often unsuitable for a range of reasons [216]. Many healthcare pathways and practices which exist in these settings may also not be formally mapped out, or their health metrics such as utilisation poorly quantified. These are some of the major barriers to the implementation of interventions aimed at improving healthcare quality. Yet decision-support interventions might be of

most value in these settings where healthcare resources are often scarce [270]. The opportunities to leap-frog and capitalise on scalable digital health interventions including CDSS are significant and could be cost-effective. Many LMIC settings are also undergoing rapidly increases in digital connectivity in the context of a young population familiar in technological use [36, 271]. Digital or otherwise, better understanding of local context will therefore be a key prerequisite for any healthcare intervention.

6.3 Aims and research hypotheses

Work presented is aimed at supporting clinical implementation of machine learning models for a LMIC healthcare setting. The management of dengue and use of models from Chapter 4 and Chapter 5 will be the focus, with emphasis on development of a CDSS framework.

The aims of this chapter are to:

i) Understand the processes and map key determinants in clinical decision-making which take place at a tertiary hospital in Vietnam for the management of acute dengue infection.
ii) Understand end-user factors which support, or hinder the use of an adjunctive clinical decision support system for dengue management, and describe the ideal features which would promote uptake and sustainability of such an intervention.
iii) Develop basic wireframe designs for a prototype dengue CDSS in collaboration with research colleagues from Oxford University Clinical Research Unit (OUCRU) and evaluate its usability.

Specifically, the research questions are:

1) What main sources of information are used by healthcare workers and what are the processes involved in the clinical management of dengue patients in a hospital setting in Vietnam?
2) What are optimal designs, function, features and place of use for a CDSS which will be of likely clinical utility for dengue management in the above context?

The scope of implementation work presented is exhaustive, and is focused on addressing current practice and needs in decision-making in dengue in Vietnam. Downstream work and findings on the development and testing of the CDSS within this context is important,

but beyond the scope of this thesis. Parts of this work will be referenced but will not be formally presented in this chapter.

6.4 Statement of contribution

I designed, carried out and analysed the work presented in this chapter. COVID-19 restrictions in air travel from UK to Vietnam between 2020-2021 resulted in cancellation of intended field visits for data collection – this included process mapping, and interviews.

As a result, I collaborated with Dr Huy Nguyen Quang, clinician researcher and Ms. An Luu Phuoc, social scientist both based at OUCRU in order to complete this part of the project. HNQ undertook observations for process mapping and task analysis. Both colleagues performed the semi-structured interviews and took part in group discussions. These discussions allowed me to generate the codes required for thematic analysis. The workshops were done in hybrid format, and I chaired the sessions online.

6.5 Ethics statement

Ethical approval for the research was granted by the scientific and ethical committee of the HTD, Ho Chi Minh City dated 14th May 2021 and by the Oxford Tropical Research Ethics Committee (OXTREC) reference 531-21 dated 17th June 2021. All identifiable information was anonymised, and interview participants assigned a study code.

6.6 Methods

Research took place at the HTD, Vietnam. This is a 700-bed tertiary referral hospital for infections which manages approximately 30,000 patients with dengue each year. The healthcare facility accepts referrals from Ho Chi Minh city and the surrounding catchment area which has a combined population of 40 million people. Of the patients who present, approximately 25-50% (8,000-15,000) each year are hospitalised for dengue treatment. All work took place between September 2020 and March 2022, before and during periods where COVID-19 controlling measures were implemented at the hospital.

To address the research questions, an observational qualitative approach was taken. The study was designed in three parts:

i) Process mapping and task analysis of the healthcare pathways for patients with a working diagnosis of dengue,
ii) Semi-structured interviews examining clinician end-user decision-making and their attitudes towards digital CDSS,
iii) Interactive workshops for end-users to engage in co-design of the CDSS tool.

Findings from the above studies were then used to inform a separate process for the design, and development of a CDSS using a human-centred design approach [272] to ensure suitability to needs and local context.

The theoretical basis used to align aspects of this implementation work has its basis from the Consolidated Framework for Implementation Research [273]. This framework consists of five domains (intervention, inner and outer settings, individuals and the implementation process) which need to be addressed in order to better understand and guide implementation of an intervention. Findings from the work will apply to better understanding of the issues relating to the inner setting and intervention itself.

6.6.1 Process mapping and task analysis of existing practice

Patient pathways for dengue management were mapped and roles of healthcare staff in these processes captured. This work was carried out as current actual processes in patient care at the HTD are not formally described and variations in patient flow may exist. The assumption was that through identifying key clinical decision-making steps and patient flow, this information would allow for the design and introduction of interventions in the future to improve healthcare quality.

An ethnographic approach was adopted whereby a clinician-researcher (NQH) was embedded within the hospital clinical teams at the HTD for several weeks, but did not directly intervene in any aspect of patient care. Processes from first presentation to discharge were recorded and compiled using a last-10 patient approach [274]. The scope of sampling included activities carried out by any healthcare staff (doctors, nurses, healthcare assistants) or patients' relatives directly related to dengue management in the hospital departments, including the emergency department, inpatient and outpatient units, adult and paediatric intensive care and the outpatient department. Any ambiguities from observations were addressed through direct discussions with the individual healthcare providers. A task analysis of all personnel involved in dengue patient management was carried out in parallel [4]. The roles, timeframes, use of equipment and

responsibilities for each healthcare worker type involved in patient management were collated and described to complement findings from process mapping.

6.6.2 Semi-structured interviews to characterise decision-making

For more in-depth understanding of the decision-making process in dengue, semi-structured interview guides were developed with focus on describing information use and existing challenges in clinical decision-making. A set of core interview stems were developed, covering the major topics including: differential diagnosis, severity assessment, admission and discharge decisions, therapeutic intervention, disagreements with care, hospital and external factors and attitude towards a computerised CDSS. This work was done to characterise existing practice and identify any areas of need where an intervention would have the most impact.

The resulting interview guide consisted of a set of 10 open-ended question stems accompanied by adjunctive follow up questions. Iterations of the interview guide were presented at OUCRU dengue research team meetings for feedback and amendments. The aim was that each interview could be conducted at the place of work over a 45-minute period. Interviews were conducted online through teleconferencing software given COVID-19 restriction measures in place during the time. All interviews were conducted by two interviewers, consisting of a social sciences researcher and a clinician (LPA and NQH). All interviews were conducted in Vietnamese. The interview guide and question stems used are presented in the thesis appendix.

Potential interview participants were identified through consultation with the department heads and relevant stakeholders in the emergency, inpatient adult and paediatric wards and intensive care departments and a peer network. The aim was to enrol participants with a variety of specialties and clinical experience levels. The target sample size was determined by the range of responses and continued until thematic saturation was reached – typically in the literature this can range between 12-20 participants.

Informed consent was then sought from participants who wished to take part in the study. A study number was allocated and basic information including gender, role, and place of work was captured on the interview. Audio recordings of the interview were transcribed and translated into English by an independent third-party outside the research team, and analysed. All recordings were subsequently deleted after the transcription process.

Transcripts were analysed through a grounded theory approach [275] with emphasis on describing tasks and associated cognitive processes involved [4]. For each transcript coding was done independently alongside two other researchers (LPA and NQH). Open coding was performed, and responses discretised into shorter description and phrases in order to extract salient features contained in the qualitative data [276]. Codes across different respondents were then connected and combined by their meaning through discussions between researchers and the wider dengue research group. Aggregation of codes into overarching (higher-level) themes was done through an iterative process of discussion and re-coding [277].

6.6.3 Workshops for the design of decision-support interfaces

The intention of the CDSS is that it should be designed with specific end-users in mind and integrate with existing workflows, taking into account its roles within hospital departments. Two workshops (one in-person and one online) were held with a focus on discussing the main themes identified from the semi-structured interviews. Information was then used to generate a shortlist of desirable features and guide co-design of the CDSS. The in-person workshop was held over 2 hours at the HTD, Vietnam, and online session a month later.

Activities in the workshop include creation of individual personas frequently used in in User Experience (UX) design [278]. Each participant was asked to create a profile of an imagined typical clinician user – their appearance, roles, experience and preferences. Post-up and affinity diagramming activities were done through pinning up notes on a board to reach a consensus on desirable CDSS functions, user experiences and prioritise the order of development. The online workshop utilised a collaborative message board, Miro whiteboard (www.miro.com) to rank and prioritise CDSS features. Guide developed for the workshops are available in the thesis appendix.

6.6.4 Wireframe and mock ups of decision-support systems

A web-based CDSS based on HyperText Markup Language (HTML) was chosen based on flexibility and widespread compatibility across a range of devices. Most digital devices (computers, tablets, phones) have web browsers and therefore can access the content. Wireframes were created based on previous findings to outline potential CDSS designs: these are rudimentary frameworks which display the appearance, layout and functional

flow of the CDSS. This approach was used to iteratively test and gather feedback on different design layouts in the workshops. Wireframes were done by hand on paper and online, and shown to clinicians at the HTD during the online workshop session.

6.7 Results

6.7.1 Pathway mapping of patients admitted with dengue

Dengue pathways for triage, management and discharge at the HTD were mapped out. Ten consecutive adult and paediatric patients with confirmed dengue were observed throughout their hospital stay. Median length of stay of these patients was 6 days (IQR 3-10). The finalised process map is presented in . The main patient flow is as follows:

- Patients arriving at the hospital would undergo clinical assessment and triage by a doctor. Risk stratification was based on age, clinical features of severe disease, and presence of risk factors (obesity, pregnancy, other co-morbidities), following the Vietnamese Ministry of Health (MOH) dengue treatment guidelines.

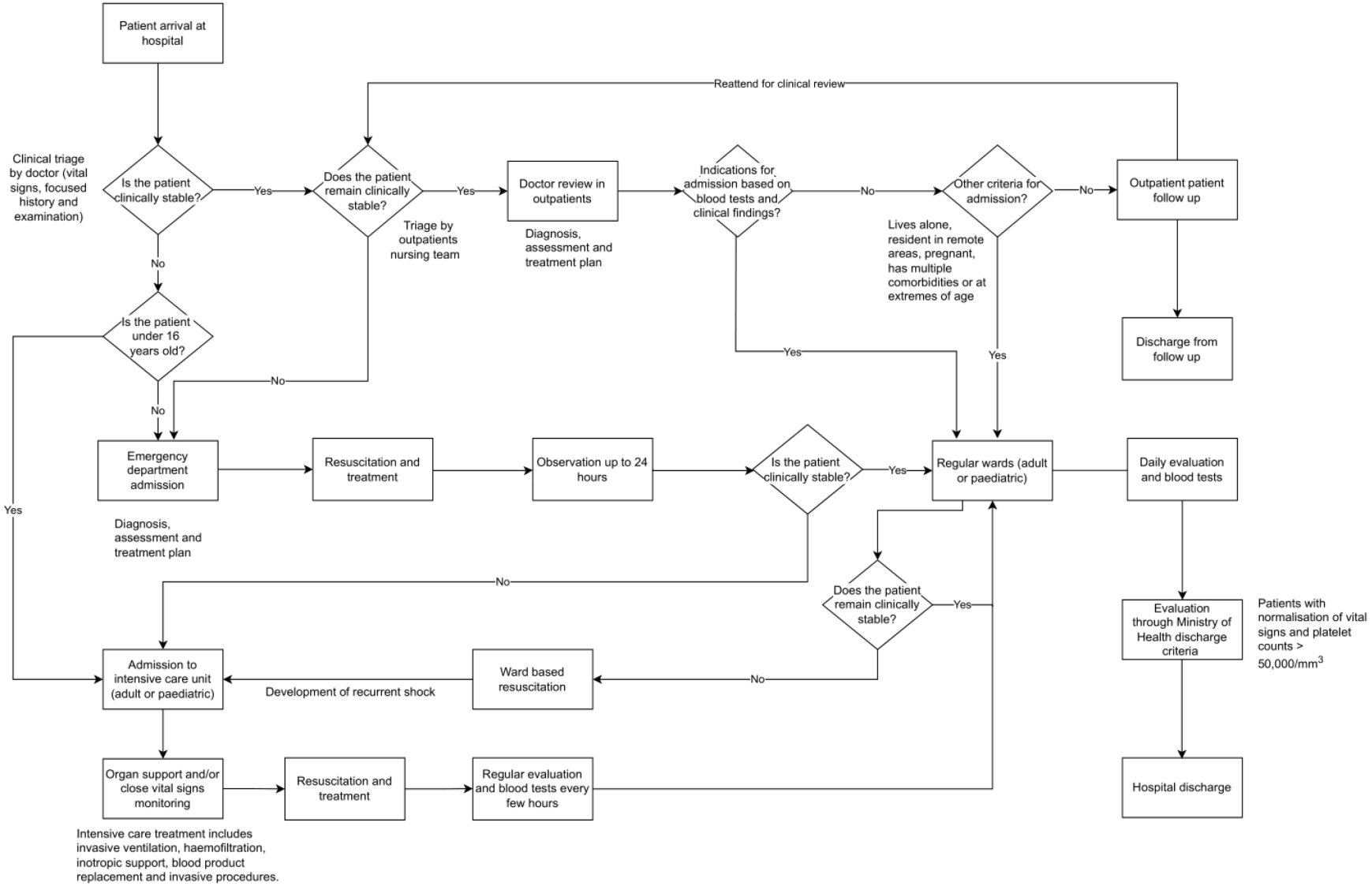
- Diagnosis is mainly clinical and based on blood results. The use of the NS1 point of care assay for dengue was variable (see results from interviews).

- Outpatient management was the default route of care for the majority of patients, with daily follow up until resolution of symptoms. Patients with features of warning signs or severe dengue would be directed to the Emergency Department, or Intensive Care Units in the case of dengue shock syndrome.

- Frequent patient re-assessments were undertaken. In the intensive care departments this would be done every hour with repeat point-of-care haematocrit measurements. For ward-based patients re-assessments were done less frequently especially during dengue epidemics where the ward capacity would be expanded rapidly. Treatment pathways often crossed over as patients require organ support.

- MOH dengue guidelines played a prominent role in dengue management during the early stages of disease, presentation and discharge criteria. Clinical decision-making and triage throughout these processes were almost exclusively doctor-led. Supervision was provided by senior staff in a formalised hospital structure, where explicit approval was required for treatment in difficult cases.

Figure 6.1 Process map of dengue management for adult and children at the Hospital for Tropical Diseases, Vietnam



6.7.2 Task analysis of healthcare staff

Dengue treatment at the HTD was delivered by doctors, nurses, nursing assistants, administrative nurses and paramedic assistants. Results of task analysis showing individual tasks, their technical environment and timeframes observed are presented in Table 6.2.

There were significant similarities in roles between healthcare in Vietnam and in the UK. Vital signs monitoring of patients relied exclusively on manual methods and this was done by both doctors and nursing staff. An electronic health record system (e-Hospital) had been introduced in 2019 for purposes of clinical management and patient billing although the majority of clinical records remained paper-based at the time of observation. A system of medically-trained administrative nurses would transcribe clinical notes daily to capture details of received treatment for purposes of billing.

Table 6.2 A list of tasks captured during patient pathway observations in the management of patient with dengue, classified by hospital staff and activity.

Role	Activities observed	Technical environment	Duration of each event
Doctor	Completing patient admission proforma	Pen, paper, patient clinical notes	5 minutes
	Physical examination of patient	Patient bedside	5 minutes
	Updating orders and patient status on progress notes	Ward, pen, paper, patient clinical notes	5 minutes
	Obtaining consent for additional interventions	Discussion with patient or relatives, pen, consent form	5 minutes
	Set up of vital signs monitors	Vital signs monitors – wall mounted	4 minutes
	Recording patient summary on round	Patient clinical notes, verbal communication with lead doctor/ vice-lead doctor	5-15 minutes
	Update patient on electronic health record system	Desktop computer located on wards	4 minutes
	Placement of invasive central venous catheter	Patient bedside using central venous catheter, monitoring device and set up	30 minutes
	Completion of nutrition form	Nutrition evaluation, pen, patient clinical notes	10 minutes

	Consultation with senior / consultant clinician	Meeting over phone or through clinical rounds	15 minutes
	Manual monitoring of vital signs	Blood pressure monitor, watch, pulse oximeter	2 minutes
	Measurement of haematocrit using bedside testing	Bedside haematocrit testing equipment	5 minutes
	Patient communication	Patient bedside	2 minutes
	Complete discharge summary	Patient clinical notes and pen on the ward	5 minutes
	Clinical examination	Patient bedside	5 minutes
	Order laboratory testing	Patient clinical notes and complete request form	5 minutes
Nurse	Measure vital signs	Blood pressure cuff, watch, pen, patient clinical notes	3 minutes
	Fill general evaluation form and estimate evaluation score	Evaluation form, pen	10 minutes
	Prepare medical equipment	Monitors, ventilator, closed suction catheter	5 minutes
	Prepare IV line for fluids administration	Intravenous fluids and giving set by patient bedside	3-5 minutes
	Filling daily care form	Pen, patient notes	5 minutes
	Measurement of vital signs	Blood pressure cuff, pen, notes	2 minutes
	Clean patient with warm cloth	Warm cloth, warmed kettle	5-10 minutes
	Assist patient getting dressed	Patient's clothes	3 minutes
	Transferring patient from outpatients/emergency department	Patient bed	10 minutes
	Communication with doctor	Phone, in person meeting	1 minute
	Preparing IV fluids	Intravenous fluids and giving set	5 minutes
	Measure vital signs	Blood pressure cuff, notes, pen	3 minutes
	Collect blood sample for CBC	Lab blood tubes, needle and syringe.	2 minutes

	Medical equipment preparation	Procedures trolley, central lines and catheters	2 minutes
	Filling general plan form	Patient clinical notes, pen	5 minutes
	Check identity bracelet	Patient bedside	1 minute
	Remove all IV lines	Patient bedside	5 minutes
	Follow vital signs and fill in daily vital sheet, online system	Blood pressure cuff, notes, pen, desktop computer	10 minutes
	Patient hygiene	Warm cloths, new blanket, washing equipment	5 minutes
	Patient communication	Patient bedside	3-5 minutes
	Medical equipment replacement when needed	Medical equipment	2-4 minutes
	Filling evaluation form	Patient clinical notes, pen	10 minutes
Nursing assistant	Move patient from OPD/ED to AICU	Movable beds	10 minutes
	Bring blankets, warm kettle	ICU blankets, kettle	5 minutes
	Support patient dressing	Patient gowns	3 minutes
	Bring lab samples to laboratory	Collected lab tubes, urine sample	10 minutes
	Support patient to get out from bed	ICU bed, wheel chair if required	2 minutes
	Clean the area after discharge	Cleaning equipment	10 minutes
	Bring fees to planning management department	Fees printed out from administrative nurse	7 minutes
	Take relatives to counter for payment for medical care	Payment counter	10 minutes
Administrative nurse	Patient's insurance health check	Online system E-hospital using computer, discussion with relatives	10-15 minutes
	Calculate overall hospital fees	Online system E-hospital using computer	10-15 minutes
	Refund if necessary	Cash	10 minutes

	Check signatures of doctors	Patient clinical notes	10 minutes
	Check patient notes before discharge	Patient clinical notes	10 minutes
Paramedic	Transfer patients between wards	Movable beds	10 minutes
	Maintain ward hygiene	Cleaning equipment	30 minutes
	Transfer clinical samples to laboratory	Sample container	10 minutes
	Collect testing results from laboratory	Results slip, patient clinical notes	10 minutes

6.7.3 Clinical decision-making in dengue management and emergent themes

Semi-structured interviews were conducted with clinicians who have experience in managing dengue at the HTD. Interviews were done between 20th September 2021 and 6th January 2022. Ten interviews in total were done, with respondents based in paediatric intensive care (n=2), adult intensive care (n=3), adult medicine (n=1) and general paediatrics (n=4). The median years post-qualification was 8 years (IQR 3-12).

From transcripts, a list of the codes were generated and higher level themes extracted. A full list of initial codes for all interviews and examples of responses are provided in the supplementary results. After iterative coding and discussion, the following five themes emerged from analyses and are shown in bold, and excerpts from transcripts in italics:

1. **Information derived from clinical assessment and evaluation represents the main driver in patient management**

There was consistent emphasis on the role of history-taking, patient examination and clinical risk assessment based on social and epidemiological factors to guide management when dengue diagnosis remains unclear.

> *'…clinical speaking, the presentation of dengue fever is similar that of other viral infections. A patient with dengue normally has a high temperature which is very prolonged, and might be initially responsive to antipyretics but will later experience fever again. Management depends on their clinical manifestations, location of care and seasonal factors.'*

Given the similarity between dengue and septic shock, experienced clinicians expanded on the role of clinical assessment in differentiating the two groups of infection:

> *'For patients with septic shock, the first thing that we observe is their pronounced lethargy, and other signs such as having cool arms and legs, blue lips or other related symptoms. Generally speaking it is predictable whether a patient is experiencing septic shock or shock resulting from dengue fever based on their clinical symptoms…'*

> *'In patients with septic shock, their blood pressure would normally drop more dramatically…if we observant, we can figure it (the diagnosis) out.'*

The timing of illness onset was consistently used an important predicting for the onset of the critical phase in dengue, and this aspect of patient history s used to direct intervention and monitoring:

> *'First things first, it's the day of illness. Usually, a case of dengue may turn critical between days four and seven and these patients may require extra attention.'*

Other non-clinical, social risk factors also influence need for hospitalisation:

> *'There are some cases that I request to be hospitalized. For example, those patients from somewhere far away, or those who live alone at home with no one to take care of them.'*

2. Dengue-specific diagnostic tests and non-specific tests are useful in dengue patient management but have limitations

Rapid diagnostics such as including the lateral flow assay detecting dengue non-structural protein 1 (NS1) and reverse-transcriptase polymerase chain reaction (RT-PCR) are available for dengue diagnosis in hospital and research clinic settings. In early disease, NS1 is useful in differentiating Dengue and other infectious disease.

> *'...so during the first three days of illness…. We have tests which can be used for diagnosis such as NS1 or NAAT.'*

Occasionally, the turnaround time for NS1 results and the limited availability of RT-PCR means other laboratory tests may occupy a more prominent role in clinical decision-making. These tests include the use of blood haematocrit and markers such CRP and can guide differentiation between bacterial and viral infections.

> *'…NS1 results may take a while to return, possibly take hours, whilst we need the results of clinical investigation immediately. At times we might not obtain a result until the following day so we rely on rapid tests such as the blood haematocrit, combined with information on epidemiology and clinical signs…'*

> *'…if no source of infection is identified in the patient and their CRP values are not high, it must be a case of dengue fever and not a bacterial infection.'*

> *'Although a patient with dengue may not exhibit critical signs of illness, a low value of platelet count may be sufficient to admit them to hospital.'*

3. Dengue management in Vietnam is largely protocolised by MOH guidelines. Complexities which are not covered in the guidelines can generally arise in later stages of the disease or in complex manifestations.

For early stage (days 3 or 4) illness management, pathways are clearly defined in the MOH guidelines. These include criteria and flowcharts for admission and algorithms for calculating fluid replacement volumes. However treatment decisions during, and after the critical phase of the disease or in specific populations may be deferred to senior clinical staff and experience.

> *'... in patients where the protocol is applied but no responsiveness is seen, some experienced doctors who are considered experts may try a small customization in treatment. This is done based on clinical experience and not part of the hospital's protocol.'*

Recurrent shock refers to the re-development of haemodynamic shock after adequate treatment and stabilisation of a previous episode of shock. This was a commonly cited example of where guidelines were inadequate for management, resulting in a reliance on adjusting fluid balance based on clinical experience.

> *'...but when a patient has repeated episodes of shocks up to three, four, or five times and this is not included in the guideline, we have to rely on the patient's individual clinical state in order to figure out the appropriate treatment.'*

Respondents also describe a lack of a formalised system in order to assess illness severity after hospitalisation and evaluation treatment response. Initial pre-critical risk is well defined (dengue, dengue with warning signs and severe dengue) but subsequent disease course is not.

> *'In the later stage of disease, we don't have any scoring scale for dengue patients. There is no system for the evaluation of a patient's illness severity. Instead, doctors perform examinations and provide their expert opinion on the patient clinical state and prognosis...'*

4. External and non-clinical sources of information play a crucial role in clinical decision- making

Knowledge of peak time periods for seasonal dengue incidence and awareness of the capacity status of the hospital may influence approaches to patient treatment. Respondents describe the challenges in prioritising individual patient management whilst taking into account the wider needs of the community by effective utilisation of hospital resources:

> *'For example, when an epidemic comes – the hospital may become too occupied and there may be no beds available anymore. So sometimes patients have to lie on the floor, which is not comfortable. The staff get overloaded (with work), and they might not be able to complete all their tasks. At those times, I must consider whether to request daily check-ups on a patient as an outpatient, or advise them to go to other hospitals for care…'*

The opinion of the care-giver or relative of the patient is also an important factor in decision-making:

> *'Examination and treatment must be done according to the wishes of a patient. But at times, the relatives might not agree and they might request for a hospital transfer or other aspects of care. During this time the doctors must also take into account the relative's opinion and plan the patient management accordingly…'*

Financial matters relating to the patient ability to pay for treatment can also be crucial. Although the majority of admitted patients experience an uncomplicated disease course and are discharged within a week, the need for dengue organ support in severe disease can be associated with catastrophic healthcare costs.

> *'… they need medicines, treatment methods that require high expenses, especially using ventilator, dialysis machine, it seems that the cost is so high. If the treatment, on average, if patients having both dialysis and ventilator use, it costs about two to three million dongs a day ($120 USD). If the treatment is to go on, within one to two weeks, half a month, it may reach fifty to sixty million. For those*

patients without the medical insurance, being manual laborers, with economic constraints, it'll be a very big number.'

5. Implementation of digital CDSS is welcomed, but specific aspects are viewed with caution

Respondents had a positive outlook regarding the future role of a digital CDSS for their setting in Vietnam:

'A computer system which can produce an explicit recommendation for approaching a problem will be the most helpful. It should be able to tell me how to monitor a patient, how often I should take tests, and how frequently to do this…'

However, issues of interpretability, governance, accountability of decisions, time efficiency, and the adverse impact on deskilling medical staff were raised as possible issues to implementation.

'… if I use the system and the results conflict with my decision then I will have to review the patient again to see if I missed anything, as an individual can be wrong and make mistakes. But I have to understand the results of the system as I may be asked 'why should you believe it?' by my seniors and I won't be able to explain this'

'If the CDSS is used in peak dengue season when there are too many patients, we will not be able to use the system if it is onerous. We prefer to be able to read test results quickly but not spend time inputting data into the system….'

'…, if someone is reliant on a software product to guide clinical management, it is possible that they might eventually lose their clinical sense… some clinicians may not accept the CDSS and rather prefer to rely on their knowledge.'

6.7.4 Workshops to understand ideal functionality of decision-support systems

A workshop with focus on CDSS functionality was held on 6th January 2022 in Vietnam, facilitated in-person and online. Invitations were extended to clinicians who were approached or consented for the interviews, and key decision-makers in dengue identified through peer networks working at the HTD. In total 13 people attended the session.

Using the perspectives of personas created at the start of the workshop, participants were asked to discuss findings from semi-structured interviews, likes and dislikes regarding current responsibilities for patient care, and the potential role for a CDSS for dengue management at the HTD setting.

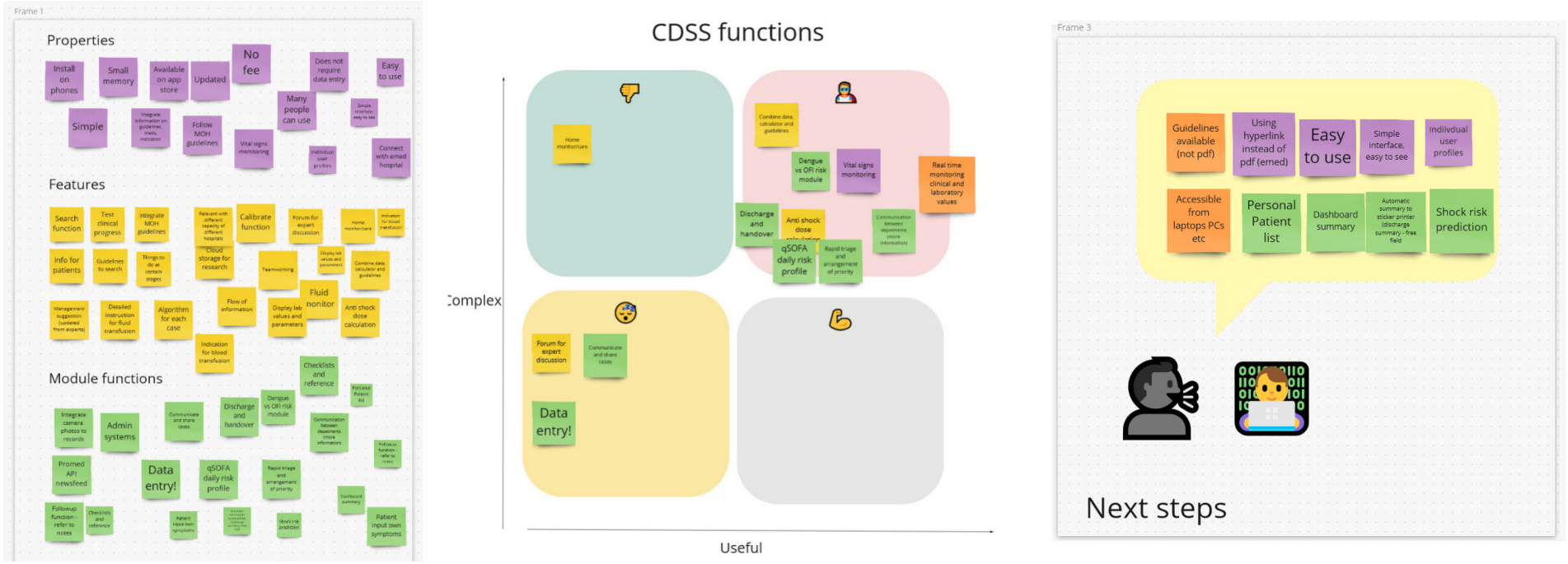
Participants described that they enjoyed teamworking at the hospital, the support provided by colleagues and identity provided as part of a department in caring for patients. However, delays in laboratory results, administrative obligations, high workload and difficult access to digital systems (e-hospital) represented challenges. Desired CDSS functions which would improve the ease or quality of clinical processes were also discussed. Functions for an ideal CDSS are shown in Table 6.3 and full results are shown in the supplementary results.

Table 6.3 Findings from workshop regarding ideal features for a dengue CDSS which could be used at the HTD in Vietnam

Attributes
Provide information also for patients and professionals
Easy to use and understand
Available to download on app stores
Simple design
Interoperability with e-Hospital
Updated automatically
Free of charge
Function
Patient search function
Can view clinical progress
Provide guidelines integrated with the Ministry of Health in searchable format
Provide a forum for expert discussion
Allow for home monitoring and care for use by patients
Indicate hospital capacity
Provide hyperlinks in guidelines
Provide specific calculations for dengue shock Allow for a step-by-step guide to patient management
Provide guidelines on blood product and fluid use
Provide algorithm for optimal patient management from an AI or expert-based consensus
Format
Can be installed on phones with low memory requirements
Secure cloud storage
Can be displayed on tablets

Findings from the first workshop were prioritised further in terms of complexity for implementation against their likely clinical utility in an online setting. Participants in this meeting were able to interact with the meeting board through post-it notes (www.miro.com). Final agreed features prioritised for development were: fluid calculation, rapid patient triage, departmental communication, display of clinical and laboratory values, assist in discharge and handover and provide access to published guidelines. An example screenshot from the online workshop is shown in Figure 6.2.

Figure 6.2 Screenshots of online workshop board. Left – findings from workshop and potential features for the CDSS are categorised and presented to users, with the option of dynamically adding entries. Middle – prioritisation of features in terms of likely clinical use against complexity for implementation. Right – Final features to take forward for inclusion in CDSS development.



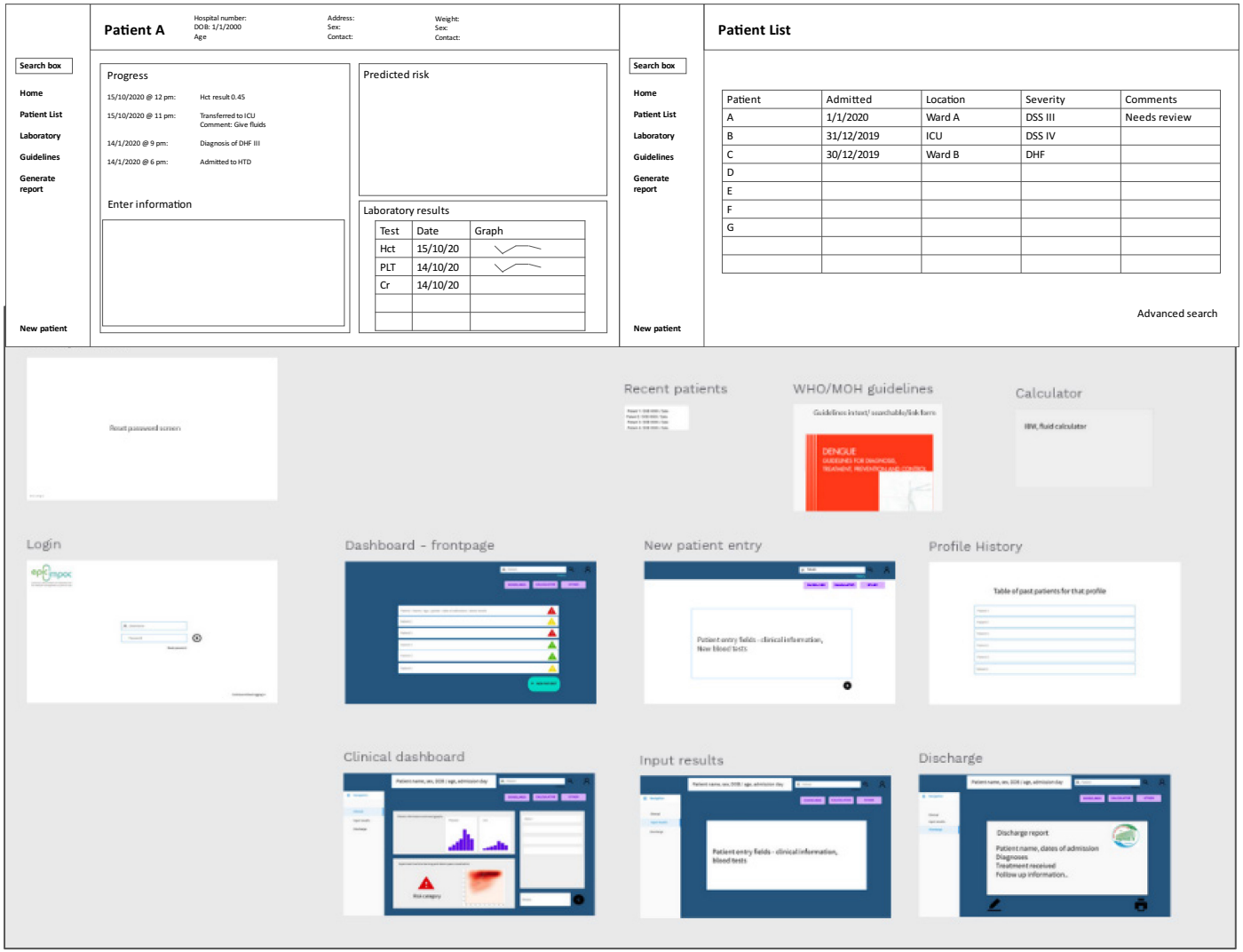
6.7.5 Wireframe designs and development of computerised interfaces

Wireframe mock ups of the CDSS were hand-drawn and subsequently developed using an open-source online tool (Penpot.app). This was done in order to establish draft layouts, map out user flow and experience and to create a template on which further changes or iterations can be built upon [279].

A website-based design with module tab navigation was the basis for the CDS. This design is common to many existing web pages and the aim was to allow for familiar navigation for users. Functions included a patient dashboard, display of guidelines, results display and decision-support [280]. Initial wireframes are shown in Figure 6.3 and presented to end-users at the online workshop. Feedback was used to create further iterations of CDSS with the aim of iterative testing.

Subsequent development work for the CDSS is beyond the scope outlined in this thesis chapter and will not be shown. Descriptions of subsequent work and proposed next steps are available in the supplementary results.

Figure 6.3 A selection of wireframe designs developed for the dengue CDSS used as part of iterative development, The top wireframes were taken from hand-drawn designs to incorporate a side-bar module tab navigation similar to web pages. The bottom series of wireframes were developed in PenPot and allowed for more interaction with individual elements on the page. These designs were shown to end-users as part of the workshop and feedback obtained from discussion and used as templates to inform the next series of developments



6.8 Discussion

Establishing context is crucial in ensuring research findings translate effectively to the real world. Aligning the needs between intended end-users with features provided by a novel intervention, in the context of the healthcare setting, is fundamental for successful uptake. Work and findings presented in this chapter represent the initial components for implementation work for the development of a digital CDSS intended to support dengue management. This work was undertaken given the lack of published implementation research on dengue decision-making and paucity of digital tools for use in LMIC healthcare settings for these purposes. Although these results are specific to the HTD, Vietnam, the methodological approaches and some of the findings – particularly related to clinical decision-making processes- are likely applicable to other healthcare settings in dengue-endemic regions.

Pathway mapping of dengue care and task analysis was done to formally describe existing healthcare processes and understand the key local determinants in dengue management. These results subsequently allowed for characterisation of the important issues relating to clinical decision-making. For the majority of patients presenting to hospital with an illness compatible with dengue, management is done in as outpatient department (OPD) basis. The use of the Ministry of Health (MOH) and World Health Organization 2009 dengue guidelines feature prominently in such management decisions. It is possible that introduction of a CDSS intended for use during triage or management decisions specifically for the OPD could be of most benefit. During seasonal epidemics the dengue caseload increases significantly [253] and places additional strain on the existing workforce. Digitalisation of such systems allow for dissemination of the intervention, and could also provide opportunities for patient involvement as well. Patient engagement in the intervention was not explicitly explored in the work but could be a valuable topic for research and would support existing ambulatory management. Alongside decision-support, availability of data from patient entered data could allow for reallocation of staff resources to focus on other crucial aspects of clinical management. However, it is important to note that all clinical management decisions (including triage) are performed only by doctors in this healthcare system: this might suggest that significant barriers to task shifting exist in this setting and need to be addressed.

Interviews of clinicians provided an in-depth understanding of cognitive processes involved in dengue management. Barriers to the use of dengue specific-diagnostics

were identified: these relate to existing workflow (NS1 testing is done in the laboratory rather than by patient bedside), assay performance (the negative predictive value of the test decreases after the initial febrile phase) or limited availability of tests with better performance such as NAAT-based diagnostics. Adjunctive models aimed at supporting clinical diagnosis of dengue (Chapter 5) might therefore be of utility in this context. Much of clinical management in dengue was also ultimately affected by non-clinical factors including seasonality – which has downstream effects on staff workload, differential diagnoses, and hospital capacity. Clinicians who were interviewed utilised information on local epidemiology, outbreaks and hospital bed-availability as part of the decision-making process in deciding whether patients were admitted or managed as outpatient. It is therefore possible that providing adjunctive non-clinical information at point-of-care through a CDSS might be of use in improving the quality of decision-making in dengue.

A human-centred design approach for CDSS development was adopted for the work in this chapter. The emphasis was to develop an intervention that focuses on the end-user. This is in contrast with many digital health approaches in which existing solutions available from commercial vendors, many of which have undergone external validation, might be appropriate [281]. For decision-support in dengue, within a LMIC context – no such product or research was available. Using workshops, wireframe prototyping and group discussions, the aim was to include functions which would ultimately benefit the particular created (clinician) persona [282]. A subsequent feasibility-utility selection process was to prioritise functions which were taken forward to develop the first-generation version of the CDSS. However, it is anticipated that the CDSS design and function may need to change over time in to remain relevant as end-user needs undergo change. Continual iterative design cycles – common in design and app development – needs to be considered to ensure sustainability of the intervention [265]. This aligns with the need for continual evaluation and monitoring of machine learning models as expanded in previous chapters to account for dataset drifts or changing contexts.

There is emphasis that CDSS interventions should be provided as part of a workflow and be context-relevant with regards to the site of intervention as well as the purpose. There is significant scope for harm in the deployment of CDSS. For example, a multi-centre trial aimed at improving awareness of kidney injury resulted in significantly higher mortality in patient cohorts randomised to the digital intervention [283]. Although in retrospect the authors discuss effects relating to heterogeneity of the sites,

and therefore arguing for more locally-applicable models such effects are inherently complex and impact on clinical outcomes will only be apparent through prospective evaluation. This emphasises the importance of interventions designed with the local context in mind as well as safeguards (such as a system for continual re-evaluation of impact) in-built particularly given the uncertainties inherent with machine learning in the context of a changing setting.

Upcoming regulatory frameworks published by regulatory agencies such as the Food and Drug Administration (FDA) and International Medical Device Regulators Forum (IMDRF) in use of software as medical device addresses this need [147], although how best to ensure these standards are met post-deployment remains challenging [284].

Strengths of the work presented in this chapter include the multi-disciplinary approach and interview findings from a range of clinicians with different experience levels in a LMIC setting. Questions stems in the interview were comprehensive to cover aspects of decision-making to in sufficient detail. The identification of common themes from the coding process also adds to validity of the findings. From the research literature there exists many models and decision aids published for dengue prediction of severe disease, but there is a lack of information on cognitive processes focusing on individual decision-making. Characterising these fundamental aspects of existing practice is crucial in the design of any intervention aimed at improving care. The prioritisation exercises held during the workshops allowed for a focused version of the CDSS to undergo development and testing. It is likely that the methods utilised in this work can also be applicable to other diseases and settings.

Limitations to the work include the research being done only at one tertiary referral hospital, which might limit the future generalisability and external validity of results to other centres. Participants were also identified by the heads of departments and may have imposed a selection bias. The benefit of digital health interventions is their inherent scalability. It is likely however that common features identified in this study in terms of cognitive processes and tasks identified at the HTD will be applicable to other settings. Interviews, process mapping and task-analyses were also done by research colleagues in Vietnamese – this was because of restrictions on travel to conduct the work in person because of COVID-19. Decisions were also made to conduct the semi-structured interviews in Vietnamese as this was the preferred language by most of the interviewees, with the translation of transcripts done by a third-party translator service. Finally – work on the development of the actual CDSS in terms of interface design and implementation has not been presented within this chapter as this was

beyond the scope of work set out in the chapter aims. This is currently being done in collaboration with the wider research group and initial results are available in the supplementary results.

Through better understanding of context, decision-making priorities and needs of clinicians at the Hospital for Tropical Diseases, Vietnam, CDSS development can adopt a human-centred design approach which is likely to result in better intervention uptake and sustainability. To address this hypothesis however, carefully designed future prospective studies will need to be conducted.

Chapter 7 - Enhancing dynamic data collection in acute febrile illness to inform clinical care: Investigation of continuous lactate monitoring through a minimally-invasive biosensor

7.1 Chapter overview and summary

This chapter addresses the issues of ensuring high-quality, granular patient data is available to the clinician outside a critical care environment. In previous chapters a common limitation identified in model development is the need for more complete and granular data in order to apply time-series methods. In AFI lactate concentrations commonly serves as an important prognostic indicator of disease severity and can be used to directly guide therapy. I present research on a novel minimally-invasive biosensor for real-time continuous measurement of lactate and a clinical evaluation in healthy volunteers.

Patient data of sufficient granularity is important for both patient care, and development of clinical models which maximise the role of dynamic time-series data. Measurement of patient physiology and biomarkers are done routinely clinical practice, but the frequency of measurements vary significantly between different healthcare and resource settings. Existing healthcare data from patients tend to be collected at irregular intervals, resulting in sparse datasets. Barriers to acquiring health data of high granularity include need for healthcare staff to initiate patient measurements (often in response to changes to clinical status), requirements for invasive monitoring through modalities such as blood sampling, and limited availability of laboratory and pathology infrastructure in many community and austere healthcare settings.

Lactate is an important biomarker which provides robust prognostic information for infections and other health conditions. Changes in concentration and trend over time provides a proxy of patient physiological status. A novel biosensor based on the microneedle platform can measure lactate in the interstitial fluid (ISF). This sensor is placed on the skin surface minimally-invasively at a depth of 1 mm, and protrusions in direct contact with ISF sense lactate concentrations, providing an electrical output signal in real-time.

To understand the role of biomarker acquisition through continuous sensing, I designed, conducted and analysed a Phase I clinical study to understand and evaluate the

performance of the lactate biosensor. Healthy participants performing aerobic exercise (n=5) underwent lactate measurement with the biosensor as well as through venesection. The median peak venous lactate generated during exercise was 9.25 mmol/L (interquartile range IQR 6.73-10.71). Biosensor signals correlated closely with venous blood lactate in all participants and exhibited a lag time of around 5 minutes. Participant acceptability obtained from post-study feedback was high, and no adverse events occurred.

Chapter conclusion: I provide data supporting the role of the novel biosensor in acquiring clinically actionable healthcare data in real time, without need for repeated healthcare staff intervention or invasive blood sampling. Continuous physiological data from this modality could be incorporated alongside other clinical data in machine learning models to create a real-time early warning system and support clinical management.

7.2 Use of dynamic healthcare data in the management of acute febrile illnesses

Clinical decision-making for infectious diseases is inherently dynamic, as discussed in the previous chapters. Risk stratification, diagnosis and treatment decisions need to account for initial patient evaluation and change in clinical state over time. Up-to-date understanding of patient physiology can be derived directly from physical examination, measurement of vital signs (such as blood pressure, heart rate and respiratory rate) and investigations such as the use of medical imaging and blood biomarkers. Understanding patient status and their change over time is important for AFI where decision-making is often acute, and severe conditions such as malaria or sepsis can result in precipitous patient deterioration over a relatively short period of time. The availability of patient data with sufficient granularity is therefore vital to inform clinical decision-making and support model development.

In many healthcare settings such information can be limited. Issues with inadequate staffing and high workload common to many healthcare settings mean that clinical and vital signs measurements cannot be collected regularly [216]. Where laboratory investigations are available, these may be basic in nature or limited in scope [230]. The measurement of blood biomarkers is commonly used to support healthcare decision-making, but barriers to implementation include: i) the need for venesection, requiring healthcare training and imposes direct time requirements on staff – this is more so for the case of repeated measurements to assess change over time, ii) patient discomfort and/or technical difficulties in obtaining blood in special populations such as neonates and children, iii) requirement for laboratory infrastructure to process blood samples which also increases healthcare costs, iv) often prolonged result turnaround times which mean that treatment decisions are based on delayed information.

Point-of-care technologies based on immunoassays, electrochemistry and lab-on-a-chip methods are promising approaches which enable biomarker measurement at patient point-of-care [285], addressing many of the issues of laboratory access. However, these modalities of testing still rely on the need for testing on blood and are challenging in special populations such as children and neonates. Where trends of biomarkers with relation to disease progression or treatment are useful for patient evaluation, repeated blood measurements are needed. Approaches which directly address these shortcomings would ideally: provide continuous or high frequency measurements of relevant biomarkers, inform patient management in real-time at point-of-care, be minimally-invasive in nature and acceptable to patient end-user and, can be deployed with basic

healthcare training. Such technologies could provide significant clinical utility and also generate rich data to inform models to support decision-making.

This chapter introduces the concept of biosensing as an approach for real-time *in vivo* quantification of biomarkers. Such devices could provide a novel approach to patient evaluation and address many of the limitations associated with blood-based biomarker testing.

Work presented is focused on evaluation of a microneedle biosensor platform used for measurement of lactate as an example of a key biomarker to inform decision-making. The microneedle platform consists of multiple small protrusions placed on the skin surface, where it is in direct contact with interstitial fluid (ISF) and quantifies its components in real-time. Research findings and implications from the in-human evaluation of a microneedle-based biosensor for continuous measurement of lactate done during this period of research is presented.

7.2.1 The role of lactate in acute febrile illnesses and other conditions

Lactate is produced during glycolysis from pyruvate through the enzyme lactate dehydrogenase, and metabolism occurs predominantly in the liver [286]. Normal blood lactate concentrations range between 0.5 – 2.0 mmol/L but transient levels up to 20 mmol/L can be observed after intense exercise. The accumulation of lactate through the interplay of increased generation and decreased clearance is associated with pathological states in humans: persistently elevated concentrations in blood are associated with hypoxia, shock and adrenergic stimulation [287]. These pathophysiological pathways represent common endpoints for many different AFI which affect the cardiovascular or volume status of patients.

Measurements of lactate from venous or arterial blood is used as a biomarker to guide patient management and decision-making. Elevated concentrations of blood lactate are associated with all-cause mortality and poor outcomes for hospitalised patients across a range of diagnoses [288–290]. From studies in sepsis, a venous lactate of greater than 4 mmol/L on hospital admission increases the odds of death twofold compared with low (< 2 mmol/L) levels [291]. Absolute blood lactate concentrations are associated with clinical outcomes in a range of conditions, such as dengue [292], malaria [293], major trauma [294] and intraoperative states [295]. Delays in lactate measurements are also associated with increased mortality in hospital patients [296].

The change in blood lactate concentration over time and dynamics provide additional prognostic information over the measurement of concentrations alone. Reduction of lactate levels within 6 hours of admission in patients with sepsis are associated with improved outcomes [21]. Lactate dynamics in children also predict clinical status in *falciparum* malaria [297]. This concept of lactate clearance and kinetics applies across a range of conditions where lactate concentrations are important [22, 298, 299]. In a clinical trial, intensive care patients with hyperlactatemia were randomised to either a treatment arm where lactate reduction over the initial 8 hour admission period was the end goal, or the control group where lactate information was not available to the clinical team for this period [300]. There were no differences in lactate clearance between groups but hospital mortality adjusted for risk factors was lower in the treatment group (hazard ratio 0.61, 95% CI 0.43-0.87) with differences in use of fluids and vasodilators between groups. The findings suggest that knowledge of lactate dynamics was important – but also that blood lactate might not adequately reflect the true pathophysiology in sepsis. Current and research approaches for lactate measurement for healthcare are summarised in Table 7.1:

Table 7.1 Existing clinical and research approaches to lactate measurement

Sampling modality	**Examples in healthcare use**	**Comments**
Laboratory whole blood (arterial or venous)	Architect platform (Abbott)	Samples collected in fluoride oxalate processed in laboratories which are associated with prolonged result turnaround times associated with sample collection, transport, analysis, and reporting. Venesection and arterial puncture can be uncomfortable and impose direct healthcare costs.
Point of care whole blood as part of blood gas analysis	ABL series blood gas (Radiometer), i-STAT system (Abbott)	As above but can be done at point-of-care, with analysis turnaround time around 5-10 minutes
Capillary blood as point-of-care test	Accutrend Plus (Roche)	Lack accuracy in performance when compared with whole blood methods in clinical studies and not in routine healthcare use [301].
Microdialysis	None	Requires invasive placement of indwelling catheter in tissues. As a

		dialysis technique is not able to directly measure lactate concentrations (results are a function of diffusion equilibrium) but can monitor trends and variations. Used mainly in context of research [302].
Indwelling catheter biosensors	None	Implanted biosensors [303] in access ports or central catheters in development and would allow continuous real-time quantification of lactate – although inherently invasive in nature with risk of infection.
Non-invasive biofluids	None	Lactate measurement from tears [304] , saliva [305], and sweat [306] show promising results in healthy volunteers but healthcare applications unclear. Extraction of these biofluids in a hospital context can be challenging and the lactate relationships between these biofluids and blood poorly defined.
Skin interstitial fluid	None	Microneedle-based sensing for lactate has been previously proposed but no human clinical evaluations have been conducted at the time of the study. Two studies have since been published in 2022 utilising microneedle-based wearables [307, 308]. Another approach utilises an invasive LED-based approach to lactate measurement [309].

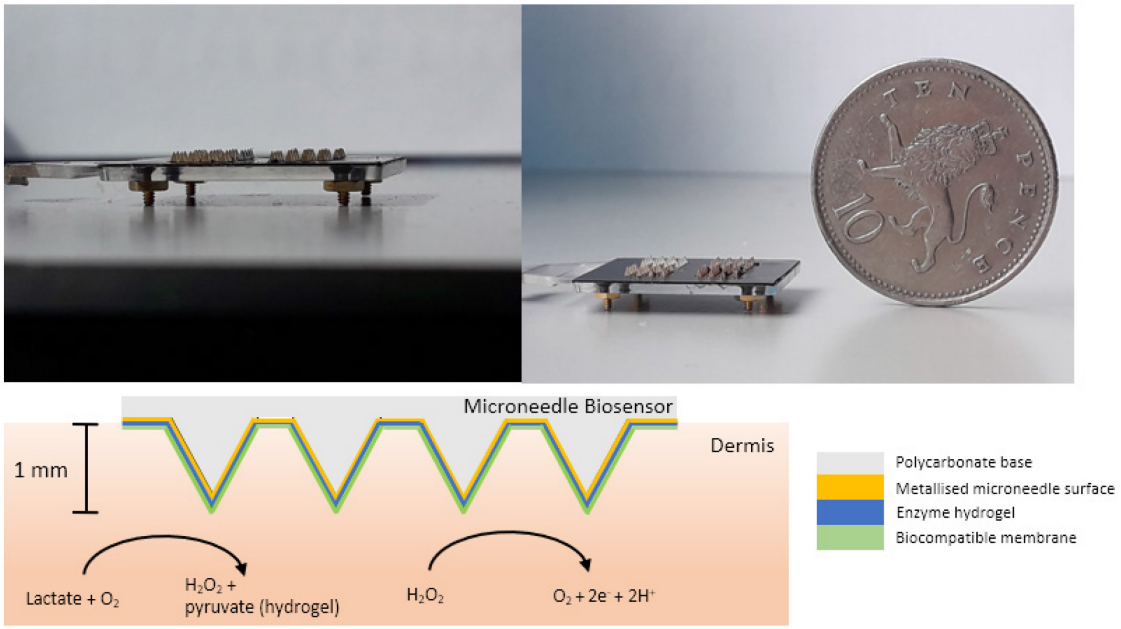
7.2.2 Continuous patient monitoring using the microneedle platform

The microneedle biosensor is a small (2 × 2 cm) device with a polycarbonate base upon which multiple sharp protrusions act as independent electrodes [310]. Placement of the device is done by applying firm (10 N) pressure onto the skin surface – typically the forearm – for a minute in order for embedding to occur in the dermis. A sensor layer on the protrusions mediates specific, and continuous detection of substrates within the ISF.

For lactate measurement, an enzyme such as lactate oxidase is fixed within a biocompatible hydrogel on the surface of these biosensor protrusions. When placed in skin and interfacing directly with ISF, lactate is converted into pyruvate and hydrogen

peroxide, resulting in electrochemical oxidation of the latter and producing a current. The magnitude of the current is proportionate to the original lactate concentration, and is subsequently quantified by a potentiostat reader through amperometry. The end result is that a quantitative measurement which is continuous and available in real-time (Figure 7.1).

Figure 7.1 Lactate microneedle biosensor. Top left and right: the microneedle biosensor measures less than 2 × 2 cm and consists of small 1 mm protrusions which penetrate the stratum corneum in the epidermis to come into direct contact with tissue interstitial fluid. Bottom: lactate in the interstitial fluid is converted to pyruvate and hydrogen peroxide, with the latter being oxidised at the biosensor electrode surface, which is held at +0.7 V vs Ag|AgCl reference electrode. The resulting current to ISF lactate concentration. Figure has been published in [311].



The depth of penetration of the microneedle device is around 1 mm as demonstrated through optical coherence tomography imaging studies [312] – this places the biosensor protrusions superficial to nerve fibres and blood vessels. Clinical studies in using this biosensor platform in healthy volunteers for the continuous measurement of glucose [313] and penicillin [314] have shown good performance and acceptability in terms of prolonged usage (up to 24 hours). Public acceptability of the biosensor as a modality for patient monitoring is also high [315].

Interstitial fluid (ISF) represents the largest component of extracellular fluid in the body. Most small molecular substrates such as lactate and glucose exist in dynamic equilibrium between ISF and blood, passing freely between compartments [316]. Characterisation of ISF has been carried out through skin microdialysis methods, which involve insertion of a subcutaneous probe whereby infused perfusate equilibrates with ISF components, which is then analysed *ex vivo* [317, 318]. There is a consistent relationship between ISF and

blood lactate in healthy participants after exercise [319]. This relationship is also maintained in pathology: in a clinical study utilising microdialysis in 162 hospitalised patients with shock, proportionate changes between ISF and blood lactate were seen throughout [320]. Significant changes in ISF lactate concentrations preceded those in blood by up to four hours, suggesting that continuous dynamics provide useful actionable information. The 24-hour lactate clearance in the adipose ISF was also shown to be higher in survivors versus non-survivors in a cohort of patients with sepsis or septic shock, whereas this relationship was not observed in blood [321]. Although blood lactate has been studied extensive by virtue of accessibility, it is possible that ISF levels provide a closer representation of local tissue perfusion and represents a more sensitive proxy for physiological derangements.

The clinical utility and role of ISF-biosensing is now well established through continuous glucose monitors for the management of diabetes [322]. Technological development has resulted in robust devices which provide actionable data and are suitable for patient self-use. It is possible therefore that ISF monitoring of lactate could also meet these clinical needs for patient monitoring without the need for blood. At the time of writing, there have been no published clinical studies or reports of continuous direct ISF sensing for lactate measurement. Characterising the relationships between ISF lactate measured directly through a microneedle biosensor, lactate ratios measured through microdialysis compared with blood lactate in health and disease will be important in understanding the clinical utility of a biosensing approach.

7.3 Aims and Research Hypotheses

This chapter addresses the thesis question of how patient healthcare data acquisition can be optimised to support the use of dynamic data for clinical decision-making. Specifically, the following questions will be addressed regarding the role of the lactate biosensor in providing a means of acquiring patient data of high granularity:

- When measured directly using a microneedle biosensor, how does ISF lactate relate to blood lactate and microdialysis as a reference?

  And

- Does this relationship provide sufficient information to inform clinical decision-making?

  And

- What are the perceptions of using the lactate microneedle biosensor?

The hypotheses are:

- The lactate microneedle biosensor is able to meaningfully capture changes in ISF lactate and these changes relate to those observed in microdialysis and blood lactate measurement.

- The lactate microneedle biosensor is an acceptable modality for biomarker monitoring in a study setting and is preferable over repeated venesection.

7.4 Statement of contribution

The design of the full study protocol, regulatory approval processes with Imperial College London and NHS Health Research Authority (HRA), all participant recruitment in the study, clinical procedures during the study and all subsequent data analyses presented within this chapter are of my own.

The lactate microneedle biosensors used for clinical testing were provided by the department of Bioengineering (Professor Danny O'Hare, Imperial College London) and department of Chemistry (Professor Tony Cass, Imperial College London). Technical aspects of participant biosensor monitoring during the clinical study were done by members of their team.

7.5 Ethics statement

The study was reviewed and approved by London – Bloomsbury Research Ethics Committee (20/LO/0364) and registered on Clinicaltrials.gov on 23rd January 2020 (NCT04238611). The study was sponsored by Imperial College London and conducted at the National Institute of Health Research/Wellcome Trust Imperial Clinical Research Facility (Imperial College London, UK). All correspondence between Imperial College London and the Health Research Authority, MHRA, Imperial College Healthcare NHS Trust and related applications are included in the thesis appendix.

All researchers underwent Good Clinical Practice training and procedures conducted in accordance with the 1964 Declaration of Helsinki and later amendments.

7.6 Methods

7.6.1 Participant population and recruitment

An exercise model was chosen for the validation of the *in vivo* performance of the lactate microneedle biosensor. Increased production of L-lactate from skeletal muscle occurs during aerobic exercise – this build-up of lactate is transient, and rapidly cleared at rest. As the aims of the model are to compare the relationships between raised ISF and blood lactate concentrations and assess their change over time, findings from this model will be broadly applicable to those seen in pathological states.

Healthy volunteers were asked to perform aerobic activity on an ergometer bicycle with concurrent measurements of ISF lactate through microdialysis and intermittent blood lactate from venesection.

The clinical evaluation aimed to recruit healthy volunteers in the context of a first in human Phase I medical device study. Participants were selected on the basis of fitness and ability to perform aerobic exercise e.g., cycling or running for at least 30 minutes duration regularly given the nature of the study.

Participants were identified through recruitment posters placed around Imperial College London advertising the study. Those interested were asked to contact the study team directly by email or phone and were provided with a patient information leaflet. They were then invited to a screening visit. All study procedures were done in the NIHR-Imperial Clinical Research Facility at Hammersmith Hospital, London, UK.

Inclusion Criteria

1. Consenting adults ≥ 18 years old
2. Healthy with no previously diagnosed medical condition from a medical practice
3. Able to perform moderately intensive exercise on an exercise bicycle without difficulty for at least 30 minutes continuously, and engages in regular aerobic exercise at least twice a week.

Exclusion Criteria

1. Active inflammatory skin condition such as eczema or dermatitis
2. Active soft tissue infection or infection at any site
3. Known hypersensitivity to any microneedle component or dressings
4. Presence of any implantable electronic devices such as a pacemaker or stimulators

5. Currently pregnant

Ethics approval documentation, study protocol and relevant correspondence is attached in the appendix.

7.6.2 Reporting of findings

Two versions of the lactate microneedle biosensor were tested in the same study. Both versions of these biosensors utilise the same microneedle platform but with different compositions to the sensor layer – these layers utilise either lactate oxidase (LOx) or a direct electron transfer lactate dehydrogenase as the mediator. Participants were recruited into both arms separately– recruitment for the lactate oxidase arm took place between 16th February 2021 and 10th July 2021 and for the direct electron transfer version between 3rd August 2021 and 21st February 2022.

All analyses for the direct electron transfer biosensor were planned to be carried out by the department of Chemistry and are ongoing at the time of writing. Therefore, only data from the lactate oxidase biosensors will be presented in this chapter.

7.6.3 Technical aspects of the Lactate Oxidase Microneedle Biosensor

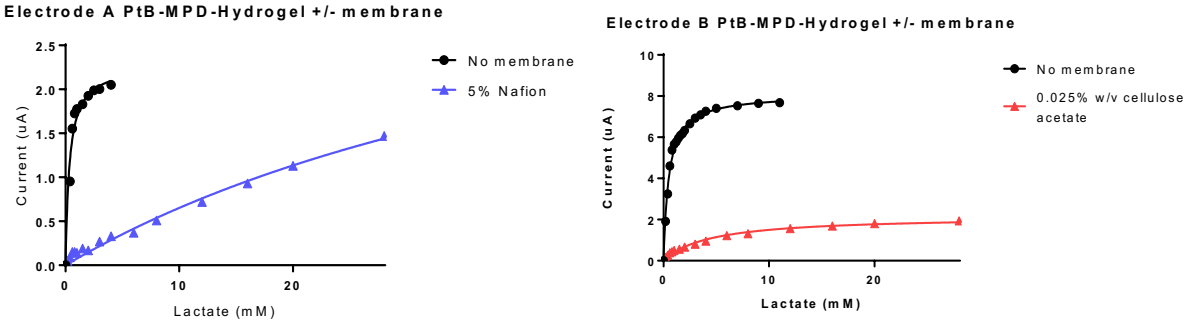
The lactate oxidase microneedle platform utilised in the study has been described in full previously [323]. Briefly, the microneedle array (Torr Scientific, UK) is made from a polycarbonate base which measures 2 × 2 cm and consists of 3 independent working electrode arrays, each consisting of 16 microneedles, and a Ag/AgCl/NaCl reference electrode on a fourth array. A layer of platinum black was applied electrochemically onto the working electrode surfaces followed by deposition of a biocompatible enzyme-hydrogel layer, consisting of lactate oxidase enzyme from *Aerococcus viridans* (Sekisui Diagnostics, Japan), 2% glycerol, 3% w/v bovine serum albumin and 4% polyethylene glycol. Lactate generated in ISF is converted into pyruvate by the embedded enzyme within the hydrogel/skin interface, generating hydrogen peroxide proportionate to lactate concentration.

The resulting hydrogen peroxide change is then amperometrically oxidised leading to a current at the platinum electrode surface measured by a potentiostat applied at 0.7 V at the bedside. An inert membrane consisting of 5% Nafion (Sigma Aldrich, USA) – a biocompatible synthetic polymer – was applied onto the electrode surface in multiple layers by nebulisation in order to extend dynamic range of detection. Single-use

biosensors were fabricated using a clean process up to a week ahead of the clinical study and kept refrigerated at 4°C until the day of use.

Microneedle biosensors were calibrated by the technical team from bioengineering against L-lactate standards before, and after clinical testing to confirm function. This was done using a tissue phantom technique with insertion of the biosensor into agar gel plates made up with specific lactate concentrations. Unpublished results obtained by laboratory work done prior to the PhD project supporting the use of the Nafion membrane are shown in Figure 7.2.

Figure 7.2 Response of lactate oxidase biosensors with and without either: a Nafion (Electrode A), or a cellulose acetate (Electrode B) membrane. These membrane layers restrict diffusion of lactate from the solution into the sensor surface and therefore extend the linear range at the expense of decreasing current. L-lactate was added to a solution containing these electrodes and the current outputs resulting from an increasing concentration are shown. The 5% Nafion showed a linear current response across concentrations up to 20 mM and was taken forward for clinical testing. These graphs are produced from unpublished data collected during my own laboratory work in 2019.



7.6.4 Study procedures

All participants provided written informed consent and were assigned a study number. Procedures were performed to comply with COVID-19 precautions including the use of appropriate personal protective equipment throughout the study.

Single-use lactate microneedle biosensors were placed on inner forearm skin surface cleaned with 2% chlorhexidine gluconate / 70% isopropyl alcohol and applied using firm thumb pressure for 60 seconds. The biosensor was secured in place with an elasticated strap and had a wired connection to the potentiostat (CHI instruments, USA) with real-time microneedle current displayed on a laptop.

The microneedle biosensor was left in-situ on the forearm 60 minutes for stabilisation. The participant was then asked to cycle on an exercise bicycle (Ergoselect 200, Ergoline Germany) at 60 rpm at power increments of 35 W up to a maximum of 210 W, for a total

of 30 minutes according to the protocol. The maximum exercise intensity reached was adjusted throughout the study according to participant ability and preference, but ensuring moderate activity was achieved. Power output values were recorded from the exercise bike. A rest period of 30 minutes with cessation of all exercise followed. Throughout the study period both forearms remained stationary on the handlebars. At the end of the rest period, the biosensor was removed, and participants were given a questionnaire and photographs of the forearm obtained.

Blood sampling

Venous lactate was sampled at regular 5-minute intervals over 60 minutes from an indwelling cannula in the contralateral arm to the biosensor using fluoride/oxalate vacutainer blood bottles (BD, USA). Samples were stored at 4°C and processed within 12 hours of collection at a UKAS-accredited laboratory located at North-West London Pathology, Hammersmith Hospital UK. All samples were processed on the Architect Ci8200 analyser platform (Abbott, USA) through a colorimetric assay. All blood samples were discarded after 72 hours of analysis in line with local policy.

Microdialysis

Microdialysis is an invasive method to sample and measure substrates in the interstitial fluid. A 63 microdialysis catheter (M Dialysis, Sweden) was inserted into the subcutaneous layer of the forearm using sterile technique. Topical anaesthesia (EMLA 5%) was applied on the skin surface prior to insertion. The inserted catheter was perfused with T1 sterile perfusion liquid (M Dialysis, Sweden) at a rate of 2 μL/minute allowing analytes, including lactate, in the subcutaneous ISF to diffuse into the catheter for downstream analysis. The catheter outflow was connected to a microfluidic chip housing which contains an online lactate sensor developed in-house at Imperial College London. This enabled the microdialysis system to detect perfusate lactate in real time. The microdialysis probe was inserted during 60 minutes of stabilisation, and continued until the end of the study period.

Feedback

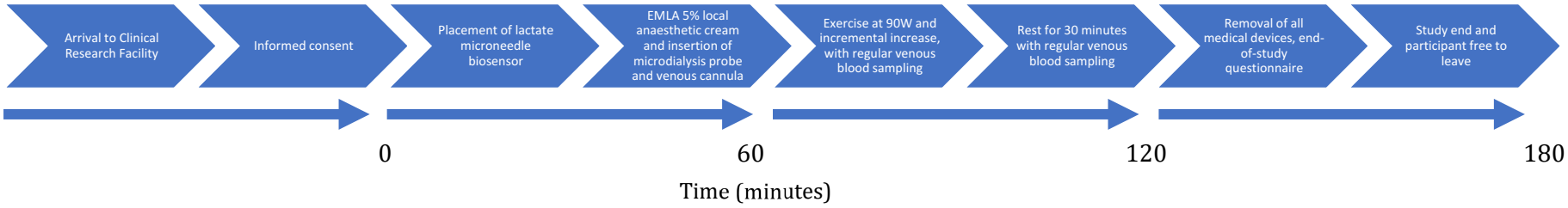
A visual analogue score in a questionnaire was administered to the participants after 2 hours of microneedle placement. A feedback questionnaire (appendix) was given to all participants at the end of the study.

Data management

Participants were assigned unique study numbers, from which all data will be recorded against. Any identifiable information (name, date of birth, contact details) were stored using the electronic health record system (Cerner Corporation, Missouri USA) used at Imperial College Healthcare NHS Trust. Signed consent forms and screening questionnaires were stored in a physical file along with the participant's research notes and kept in a locked research cabinet at Hammersmith Hospital.

An outline for study procedures is provided in Figure 7.3 and the full study protocol is provided in the thesis appendix.

Figure 7.3 Study outline and timings. The study begins after informed consent with placement of the microneedle biosensor for the 60-minute stabilisation period. During this period venous cannulation and microdialysis procedures are also carried out. The participant is then asked to exercise at 90W with incremental increases in power output for 30 minutes, and asked to rest for another 30 minutes during which venous lactate sampling takes place. Removal of all medical devices and end of study questionnaire takes place after which the study concluded.



7.6.5 Sample size, data and statistical analyses

A formal sample size calculation was not performed as this was a descriptive first-in-human study to explore and understand performance of the lactate microneedle biosensor against comparative measurements. Given the possibility of technical and clinical issues arising during the study, the original protocol had specified a conservative sample size of up to 15 participants, in line with most Phase I medical device studies. The decision to stop recruitment was based on the quality and reproducibility of the data obtained.

A robust LOWESS (locally weighted scatterplot smoothing) filter with a 60-second window was applied to raw microneedle and microdialysis data to remove outliers, motion artefact and attenuate noise. In order to calibrate individual biosensors with their corresponding blood lactate concentrations, individual standard for each biosensor were fitted. The absolute differences between values from the standard curve and the actual venous lactate concentrations were then used for Bland-Altman analysis and plotting. Analyses were done independently for the exercise (up to 30/35 minutes) and the rest phase (30 – 60 minutes) as it was assumed that the ISF:blood lactate dynamics and the relationship between these two phases was different. The use of this technique for biosensor calibration is established in other settings such as continuous glucose monitoring [324].

Venous lactate was assumed to change in a linear fashion given that each measurement was obtained 5 minutes apart. Additional venous lactate points were interpolated at 1-minute timepoints between actual measurement points. Microdialysis dialysate results were calibrated to provide concentration values by calibrating the microdialysis sensor against known L-lactate standards to produce a 5-point calibration plot. Microdialysis ISF lactate concentration were plotted against the interpolated venous lactate values and biosensor current.

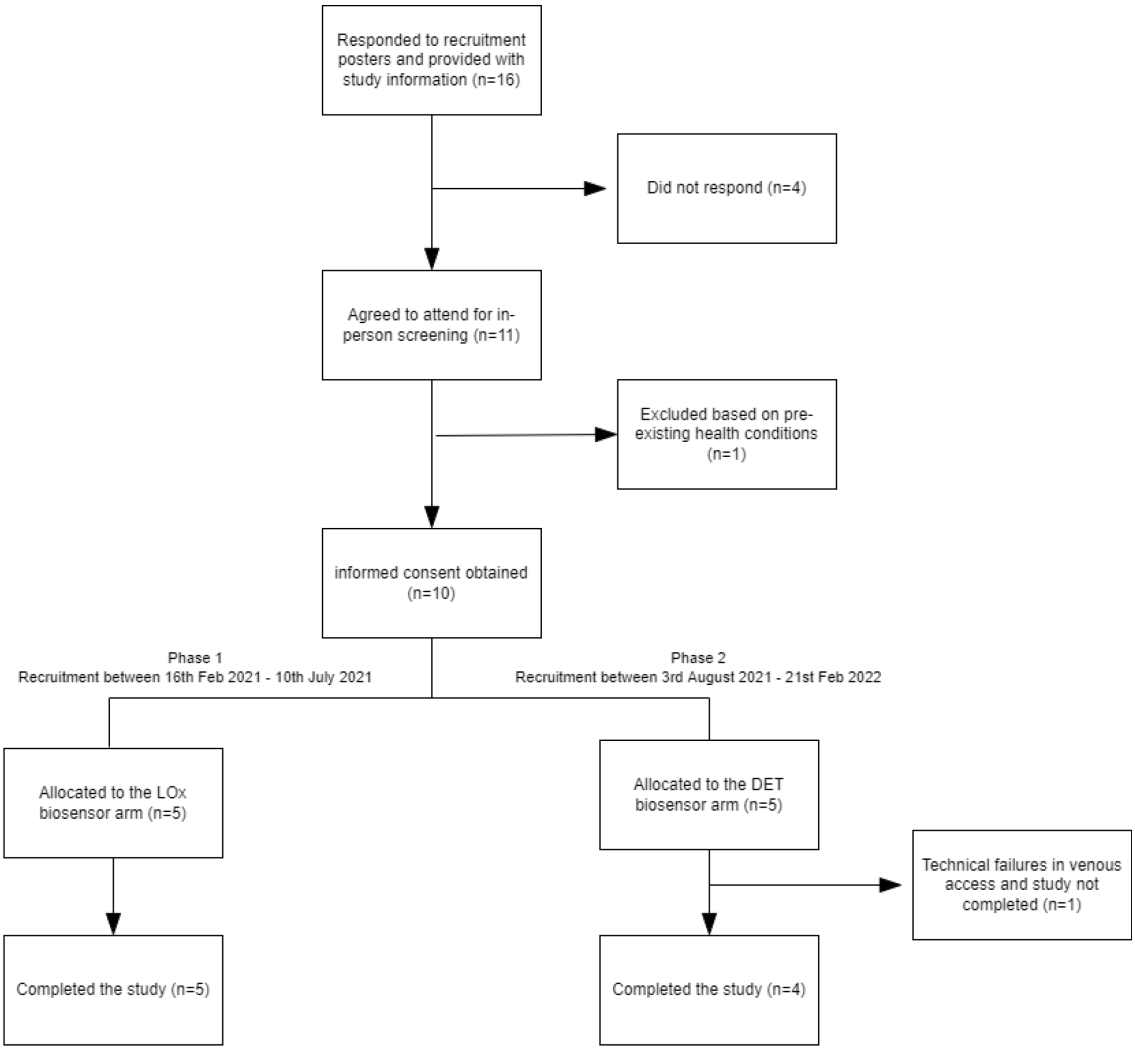
In order to compare the dynamic relationships between biosensor current and venous lactate, all values were normalised to a range between 0 and 1 to allow for comparison. A 5-minute rolling window mean average was applied to calculate the rate of change across the different timepoints. The linear relationship between the rates of change for both measurements were derived using Pearson's coefficient. To estimate the lag time between biosensor and venous lactate values, the point of inflection where there was zero rate of change was identified for both current and venous lactate readings. The difference between these points were then calculated and averaged.

7.7 Results

7.7.1 Baseline description of participants and characteristics

Between 16th February 2021 and 21st February 2022, I identified and screened 16 individuals through recruitment posters placed around Imperial College London who had responded to the study advertisement. Eleven individuals agreed to proceed to in-person screening and one participant was excluded on the basis of pre-existing health conditions. Five participants were enrolled in the LOx biosensor arm of the study and five enrolled in the direct electron transfer arm of the study. One individual from the direct electron transfer arm did not complete the study given technical failures in obtaining venous access. Only results from the LOx arm are presented in this chapter. The180pprox.tment flowchart is shown in Figure 7.4.

Figure 7.4 Study recruitment. Assignment of participants to the LOx or the direct electron transfer biosensors occurred sequentially in two phases over Feb 2021 and 2022.



The median age of participants enrolled in the study was 32 years (Interquartile range, IQR 27-33) and 1 out 5 participants was female. The median baseline pre-exercise venous lactate was 1.40 mmol/L (IQR 1.23-1.52), peak venous lactate was 9.25 mmol/L (IQR 6.73 – 10.71). The lactate concentration at the end of the resting period after 60 minutes was 2.41 mmol/L (IQR 2.06-2.90). All participants completed aerobic exercise and produced a median power output of 113 W (IQR 93-130 W). No adverse events were reported during the study. The characteristics of the participants, their results and power output are shown in Table 7.2

Table 7.2 Baseline characteristics and results of participants enrolled for the LOx microneedle biosensor

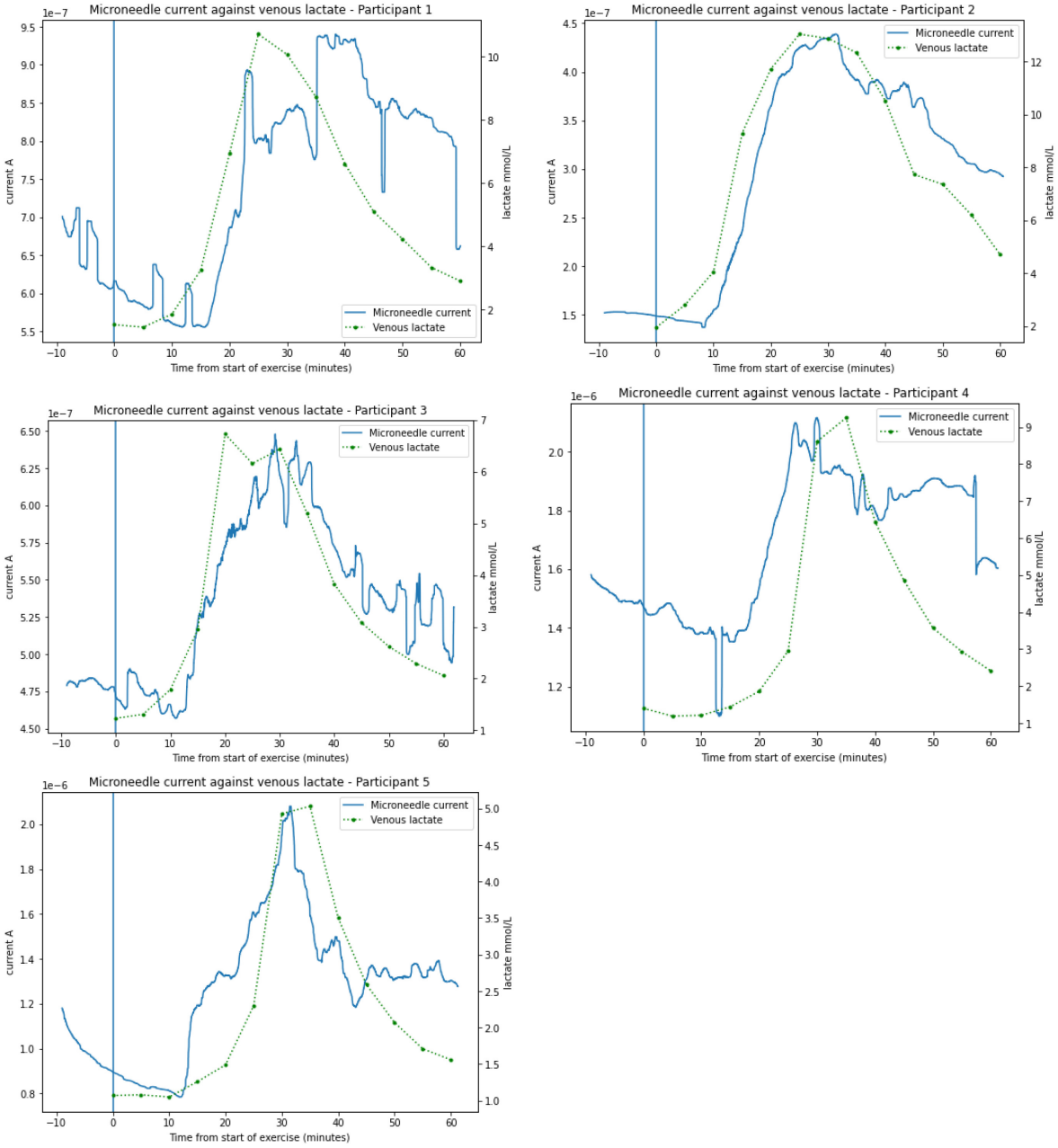
Characteristic	Participant				
	1	**2**	**3**	**4**	**5**
Age (years)	30-35	20-25	26-30	30-35	30-35
Gender	Male	Male	Female	Male	Male
Mean power output (W)	99	74	95	115	112
Peak power output (W)	175	105	140	210	210
Baseline lactate (mmol/L)	1.52	1.96	1.23	1.4	1.07
Lactate max (mmol/L)	10.71	13.03	6.73	9.25	5.03
End lactate (mmol/L)	2.9	4.71	2.06	2.41	1.56

7.7.2 Results of microneedle measurement of lactate

Microneedle biosensor current was filtered for noise and outliers through a robust LOWESS method and aligned with venous lactate results for each participant. No significant changes were observed in the biosensor current output during the initial stabilisation phase where the participant remained at rest and lactate levels were assumed to be stable.

With the onset of exercise (0-30 minutes), there was a rapid rise in venous lactate and biosensor current readings. With rest (30-60 minutes) both venous lactate and biosensor current readings decreased consistently. Continuous microneedle biosensor current followed venous concentrations closely over time. For participants 1-3, significant increases in biosensor current followed a rise in venous lactate but for participants 4 and 5, there was increase in biosensor current *before* measurable increases in venous lactate. In addition, there was also a greater lag in biosensor change with respect to venous lactate during the resting phase compared with in the exercise phase. These results are displayed in Figure 7.5:

Figure 7.5 Plots of individual microneedle current against venous lactate levels for each participant. The biosensor was placed on participants up to 60 minutes before the start of exercise (vertical blue line, time 0). The exercise phase took place between 0 and 30 minutes, and the rest phase between 30 and 60 minutes when the biosensor was removed. The biosensor current (continuous blue) is displayed against venous blood (green dotted) which was obtained from participants at 5 minute intervals throughout the study.

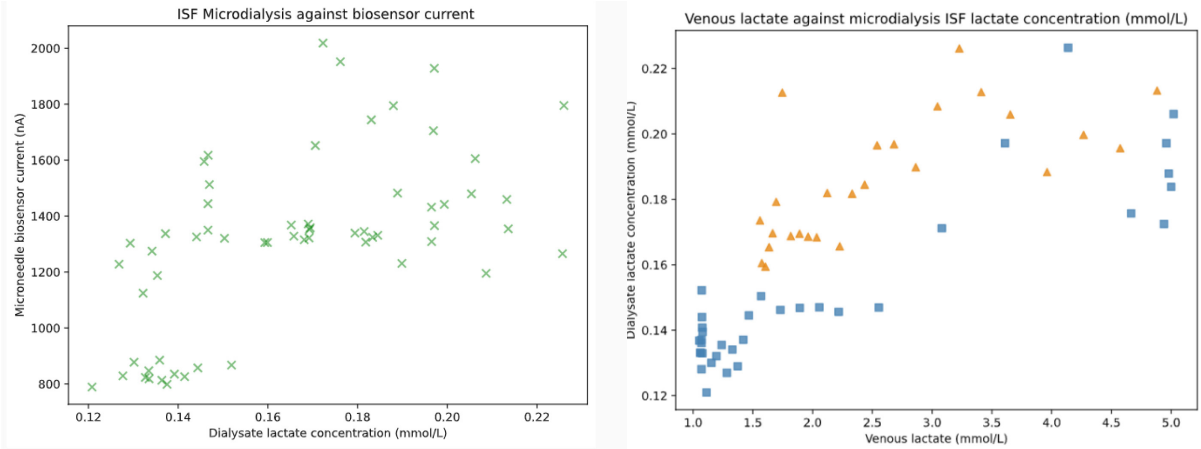


Microdialysis

Microdialysis was performed to understand trends in ISF lactate concentration and provide a reference measurement between biosensor current and blood lactate. Consent for microdialysis was obtained only from one participant (number 5) and therefore performed only for that individual. In general there was a positive association between dialysate and venous lactate concentrations, with suggestion that

these relationships differed depending on whether the measurements were taken in the184pprox.184e or rest phase. In general dialysate lactate levels from ISF were higher at rest compared with venous lactate. There was also positive relationship between dialysate lactate and biosensor current. Of note is that the lactate concentrations measured from the dialysate were significantly lower compared with that of blood for reasons including variable recovery of lactate (as a function of lactate equilibrium between the microdialysis catheter and ISF) which is also modified by the rate of dialysate perfusion through the microdialysis catheter. These results are shown in Figure 7.6:

Figure 7.6 Results of microdialysis for participant 5. Left – calibrated dialysate concentrations from microdialysis were plotted against microneedle current output showing a positive association. Right – interpolated venous lactate results plotted against calibrated dialysate concentrations – the blue squares denote results obtained from the exercise phase (0 to 30 minutes) and the orange triangles from the rest phase (30-60 minutes).



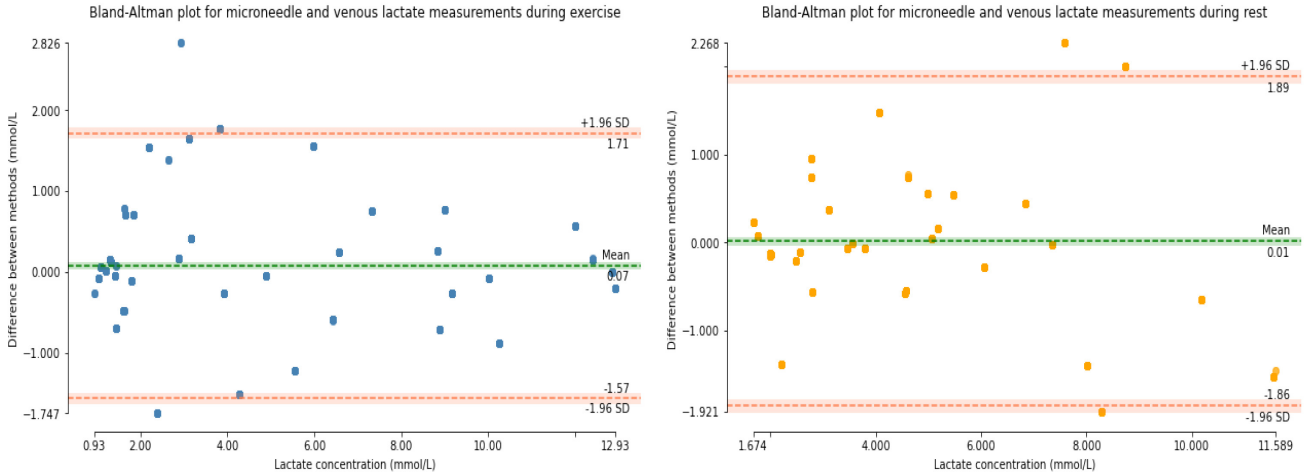
Biosensor and blood lactate concentrations

In order to provide a calibration between ISF biosensor current and the corresponding venous blood lactate concentration, assumptions that there was no significant time-lag between the two compartments were made. The biosensor devices were individually hand-fabricated and exhibited significant variation in output measurements. Therefore individual calibrations were done for each biosensor.

For each biosensor, two separate calibrations curves were made depending on whether the data was obtained during the exercise (0-30 minute) or resting phase (30-60 minutes) as the assumption was that the relationships between ISF and blood differed between these phases. After calibration a Bland-Altman analysis was performed with

these results, and although there was good overall mean agreement between the measurement modalities, the 95% CI difference of up to +/- 1.89 mmol/L was seen (Figure 7.7):

Figure 7.7 Bland-Altman analyses of calibrated biosensor outputs aggregated across all participants and compared with discrete venous blood lactate values taken during the study. These analyses were done separately for the exercise (left) and the rest (right) phases. Overall there was no significant mean difference but individual readings deviated similarly across all lactate ranges – with the 95% confidence interval (CI) around +/- 1.89 mmol/L.



Biosensor and blood lactate rate of change

The ability of the microneedle biosensors to measure changes in ISF lactate concentrations continuously over time was approached through comparing the relative rates of change between modalities. Biosensor currents and venous lactate for each participant were rescaled and normalised into a value between 0 and 1 so they could be comparable. The rate of change (linear gradient) was then derived from current and concentration values contained within a 5 minute rolling window.

Median differences in response time between changes in venous lactate and changes in biosensor current were estimated by the identifying the inflection point where the lactate rate of change went from positive to negative (after cessation of exercise) – on average there was a difference of +5 minutes (IQR -4 to 11) between the modalities. Individual participant and aggregated rates of change are displayed in Figure 7.8. The linear Pearson's correlation between biosensor current and venous lactate concentration for the aggregated results was 0.86, supporting a relationship between the two modalities (Figure 7.9).

Figure 7.8 Rate of change between biosensor and venous lactate for all participants. A 5-minute window rolling average was applied to normalised values and a linear gradient for each window calculated. In general microneedle rates of change followed that of venous lactate for both exercise and rest periods.

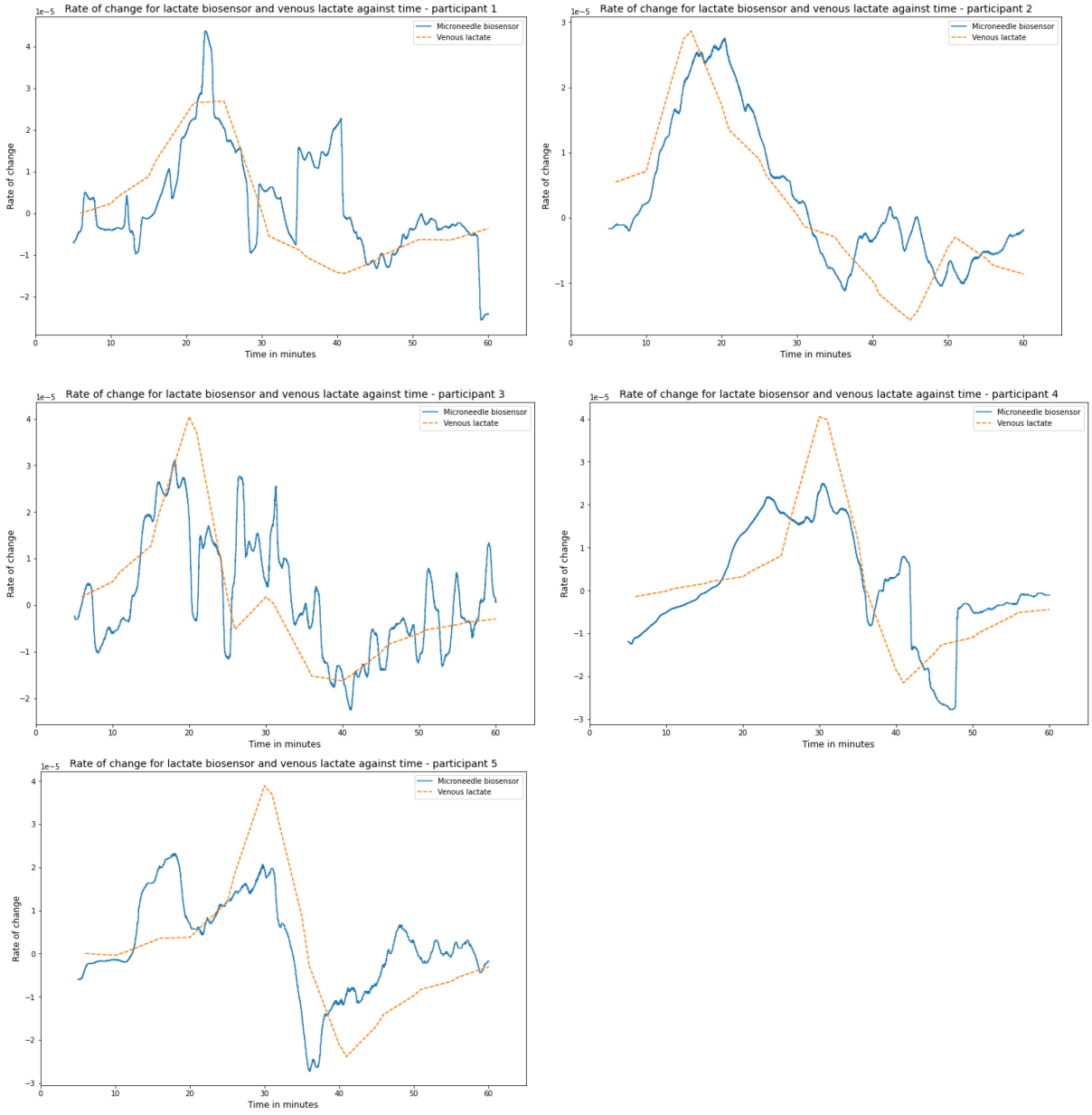
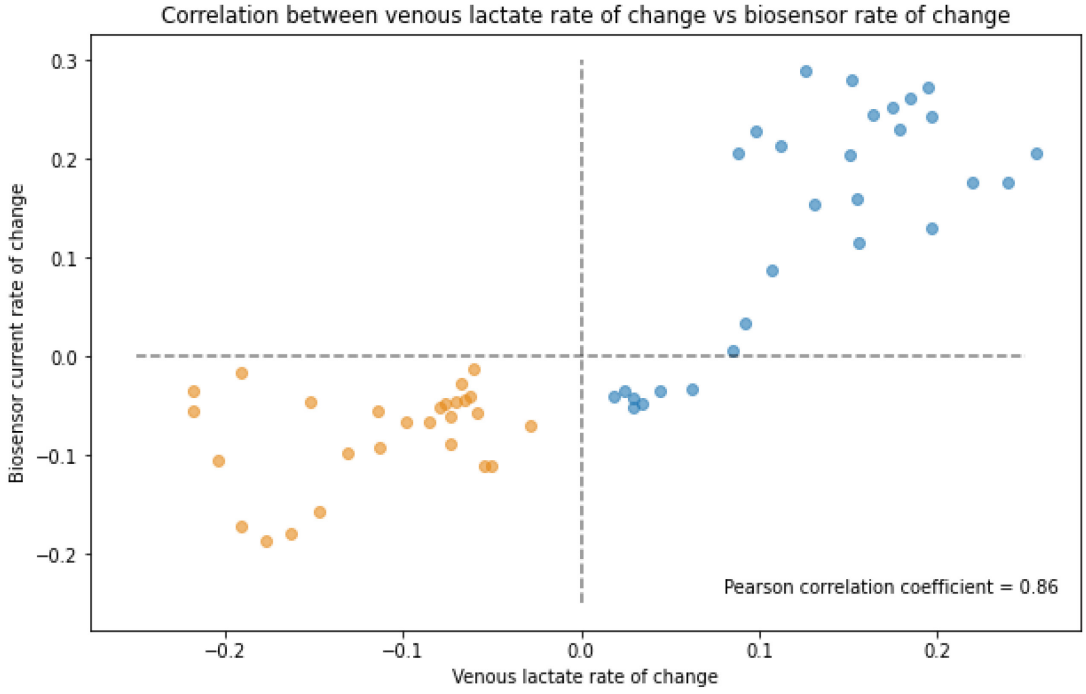


Figure 7.9 Aggregated rate of change for all 5 participants. Top – normalised biomarker and venous lactate rate of change values from a 5-minute rolling window were averaged across all participants and plotted. Bottom – the correlation between the two values across the whole study period is shown at the bottom plot showing a linear relationship between the two (Pearson correlation coefficient – 0.86).





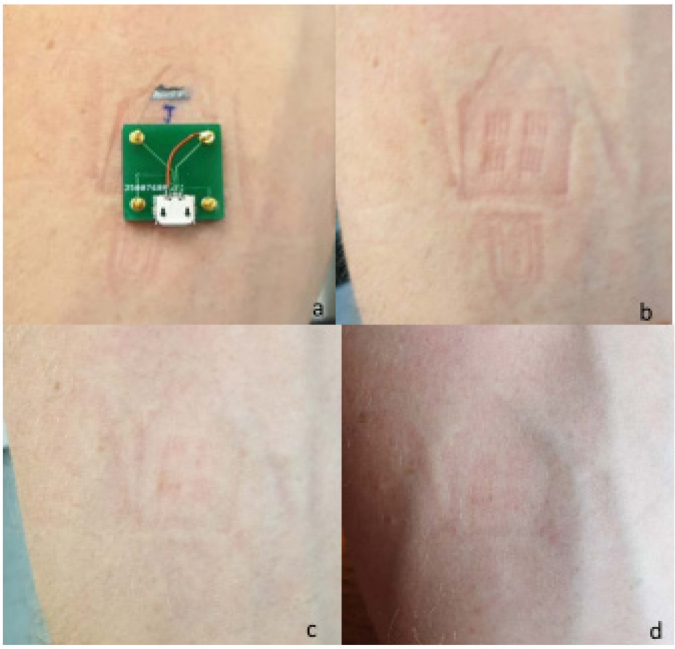
7.7.3 Participant acceptability of the microneedle biosensor

From the post-study questionnaire, the median discomfort rating at the biosensor site using the Visual Analogue Scale was 0.4/10, and 1.0/10 for the cannula in the 5 participants. The degree of restriction imposed by the biosensor was 2.9/10 with

comments relating to the wired nature of the microneedle biosensor. All participants feel that the use of microneedle patch was preferable over frequent blood testing in a clinical context.

A photograph of the skin from one participant at 0, 15, 60 minutes and on removal of the biosensor after 120 minutes of placement is shown in Figure 7.10. There is almost complete resolution of the skin changes by 60 minutes after microneedle removal.

Figure 7.10 Photographs of skin after removal of the microneedle biosensor - a) lactate biosensor in situ on participant forearm without connecting wires; b) underlying skin after immediate removal of biosensor; c) underlying skin after 15 minutes of biosensor removal; d) underlying skin after 60 minutes after biosensor removal. This figure has been published in [311].



7.8 Public acceptability of the microneedle biosensor

Perceptions of the microneedle patch were explored in a public engagement event held at the Great Exhibition Road science festival on 18th June 2022. As part of a People' Research Café – members of the public were invited to attend the café to discuss the microneedle biosensor project. During a 2-hour session, another researcher and I presented the microneedle biosensor device and its potential role to the participants. Attitudes towards use of the biosensor were generally favourable. We spoke to around 20 people individually. Almost all felt that using this biosensor patch was preferable to having blood tests taken in terms of comfort and safety. The findings and main themes summarised from these discussions are displayed in Table 7.3 **Error! Reference**

source not found.and provided insight into the importance of the physical design as much as functional aspects:

Table 7.3 Summarised findings from public acceptability of the microneedle biosensor patch through individual and group conversations in a People's Research Café setting.

Theme	Discussion
The biosensor is preferable if performance compared with blood test is comparable	Clinical trials to evaluate its performance and safety are important. We discussed possibilities of prospective studies to compare the performance of the biosensor with routine investigations.
Ensuring acceptability in small children	Making the design children-friendly (bright colours, cartoon designs) and also safe (cables and wiring cannot be accessed by small children.
Integration into health systems	Ensuring that clinicians can use the biosensor system in parallel with their work. We discussed how the intervention should not just be introduced as a standalone tool but should fit in with the processes.
Privacy and confidentiality	How to ensure that the data is secure and not accessible by 3rd parties unless consent provided. We spoke about how consent and data processing will become much more important and relevant in healthcare.
Using the information for own health	The user being able to use the biosensor to guide and improve their own health would be attractive. Discussion about commercial wearable devices and how they can help with improving individual lifestyles in terms of exercise.
Size of device	Making it smaller and more compact would improve adoption.
Manufacturing waste	Disposable nature of the device generates electronic waste which is becoming an important issue.

7.9 Discussion

Results from the first in-human evaluation of a minimally-invasive lactate biosensor are presented in this chapter. The performance of the biosensor supports its role in the continuous measurement of clinical biomarkers at point of care without need for laboratory testing. Use of such a device would allow for the dynamics and change in lactate to be automatically captured in real-time. These dynamic data in turn support clinical decision-making for a range of conditions, including AFIs. These data can additionally feed into models (alongside other signals) which explicitly make use of dynamic and time-series data. The minimally-invasive nature of the biosensor avoids need for repeated and invasive blood sampling, increasing acceptability in different patient groups and role in settings with limited laboratory access. The material costs for the device is approximately £50 per biosensor but this is likely to decrease with increasing scale of manufacture.

In the clinical evaluation of healthy participants, venous lactate levels between 1.07 and 13.03 mmol/L were generated over an hour. Calibration of the biosensor current against venous lactate levels and subsequent Bland-Altman analyses showed that there were no significant mean differences between the modalities across different blood lactate concentrations seen in the study, with a 95% confidence interval of +/- 1.89 mmol/L. Aggregation of these results through rolling windows over time and comparison of rates of change show that the biosensor was able to quantify lactate change over time during, and after the period of exercise with good concordance (R=0.86). There were no adverse reactions in the study over the 2-hour period of biosensor placement, and participant acceptability was high which supports future implementation.

Understanding the physiology of ISF lactate and its relationship with blood will be important for future clinical application. This aspect has hitherto not been characterised in detail in part due to the lack of measurement modalities for directly sampling ISF. Observations from individual participants in the study show that although biosensor current increases rapidly in exercise alongside venous lactate, there was a slower decrease during rest. This has been described in a microdialysis study previously [325] and may relate to a relative difference in perfusion between phases. With rest and cooling post-exercise, blood diverted away from the skin tissue would mean that lactate clearance in skin ISF is reduced relative to systemic (blood) clearance. This is also supported by the microdialysis data in this study – whereby

relationships between dialysate lactate concentration differed between the exercise and rest phase.

Although an exercise model was used as a proxy for pathology e.g., shock – it would be important to understand how skin perfusion is altered in conditions such as sepsis, or severe dengue which would impact on the biosensor ability to detect physiological derangements. Microdialysis studies from a single centre in critically ill and sepsis patients [318, 320] have supported the notion that ISF provides a sensitive early indicator for changes in tissue perfusion. Measurement of ISF lactate and/or its clearance dynamics could provide relevant information to guide clinical decisions prior to onset of hyperlactatemia. However, the relationships between microdialysis-ISF and blood lactate are not consistently seen across clinical settings [326] nor is it clear how states of impaired microvascular perfusion, such as that seen in *P. falciparum* malaria [327] will affect ISF lactate dynamics. It is also important to note that microdialysis provides indirect quantification of substrates in ISF and will have important differences with a direct microneedle approach.

Limitations and future work

Limitations to this study include small sample size (n=5) although this is in line with most first-in-human volunteer studies of novel medical devices. The duration of placement was also relatively short (2 hours) whereas a device which can remain sensing for days would be ideal for clinical use. It is accepted changes in blood lactate occur over hours rather than minutes in pathology [299]; it is likely that longer periods of biosensor monitoring will be needed to provide meaningful information in the clinical setting. Each biosensor also exhibited significant differences in the amplitude of electrical current output, ranging from 10^{-7} to 10^{-6} A. Biosensors for the purpose of the study were hand-fabricated, emphasising need for standardisation and quality assurance in future manufacture.

The placement of biosensors onto participant skin In this study was done by thumb pressure for one minute and kept in situ by adhesives. It was not possible to confirm that the microneedles were placed within the dermis for the entirety of the study. Although it is possible that lactate in sweat from the skin surface might have contributed to the biosensor signal, the rapid fall in biosensor signal at the end of the exercise phase goes against a major contribution of sweat – it is also likely that lactate concentrations continue to increase as the water component is evaporated. Biosensor

placement, composition of underlying tissue, placement site and depth of optimal insertion are likely to have played a role in individual variability and these need to be better defined.

The relationships between ISF and blood lactate and time-lag were estimated from the peak signal at the point of inflection. It is likely that this relationship is much more complex and not consistent across different levels of venous lactate. Further data points through prolonged monitoring and greater number of participants would be useful to address this in detail.

One major challenge in extrapolating findings from the biosensor pilot to clinical settings is the very different physiology present in shock states. Poor tissue perfusion, derangements in acid-based balance and hypoxia in the local environment could result in both very different relationships between ISF and blood and also affect performance of the biosensor. Cross-reactivity with other compounds found in clinical settings within ISF are known limitations for enzyme-based biosensors [328]. Compounds including paracetamol, uric acid and ascorbic acid undergo redox reactions at potentials utilised for these biosensors. The Nafion membrane of the biosensor acts as a diffusion limiter and can counter against these interferents [329] but other approaches utilising low potentials at the electrode, such as the use of a direct electron transfer enzyme, could be promising [330].

Future perspectives

The microneedle platform has been tested in previous clinical studies and retained efficacy for up to 24 hours of continuous use [331]. Quantifying lactate concentrations and dynamics could be of clinical utility in multiple conditions and contexts. As a self-contained and minimally-invasive wearable, the biosensor could provide actionable information to clinicians in pre-hospital, ambulatory care and other settings where access to laboratory services is challenging.

Continuous measurement allows for early recognition of patient deterioration as it bypasses behavioural barriers to initiate measurement of blood lactate: this can be particularly relevant in caring for special populations such as neonates or small children. Such approaches are being explored in greater detail: continuous real-time measurement of glucose [332] resulted in better glycaemic control in pre-term infants.

The ability for this biosensor approach to measure a number of biomarkers continuously in real time could provide the information needed for personalised approaches to patient medicine. Incorporation of established markers like CRP, inflammatory cytokines with the ability to conduct therapeutic drug monitoring coupled with closed loop antimicrobial administration would provide a route towards an integrated approach to infection management. Coupling of the biosensor with other wearable technologies [333] such as those utilising non-invasive modalities like photoplethysmography could improve performance. Linkage with decision-support systems built on models which can fully utilise these time-series continuous data are crucial.

Chapter 8 - Discussion and conclusions

8.1 Summary of thesis findings

Existing clinical prediction tools for infection management are used to provide an evidence-based approach to decision-making. Derivations of many of these tools are commonly done through regression techniques and direct comparisons between patient cohorts [56]. Resulting findings on risk factors and odds ratios guide the development of clinical scoring systems, which traditionally emphasise a balance between adequate simplicity for bedside implementation, with predictive performance across a range of patients. However, a major limitation is that such existing tools do not adequately make use of the range of existing healthcare data which is available and are unable to account for individual dynamic time-series relationships important in patient management.

In an era of digital health, connectivity and availability of mobile computing processing, there exists significant scope for the role of electronic clinical decision support-systems (CDSS) to augment the quality and efficiency of healthcare. Such CDSS could provide far greater personalisation to the decision support provided and demonstrated their clinical utility for specific purposes in healthcare. Such systems, given increasing digital literacy and ownership of devices such as smartphones, are amenable for scale-up and deployment [254]. Novel approaches to data usage such as machine learning integrates well with this digitalisation and its inclusion in healthcare should be a necessity. Through use of diverse datasets at sufficient granularity, findings provided by novel methods could inform clinical decisions and offer significant benefits.

Much of healthcare datasets include repeated longitudinal data collected from individuals over time. Specifically accounting for the time component in patient health and monitoring for changes in physiology is standard practice in clinical assessment. Understanding how patient state changes with time provides understanding of disease progression and treatment effects. Existing clinical scoring systems and tools in use, or in current research however handle repeated observations poorly and do not adequately account for these time-series relationships.

My research hypothesis was that machine learning models which did account for such time-series dependencies in healthcare data offer significant advantages in terms of performance, or potential clinical utility over conventional approaches. The scope of my work was focused on clinical decision-making in the management of AFI.

I utilised routinely collected and existing real-world clinical datasets with focus on exemplar conditions from two contrasting contexts: dengue in Vietnam and bacterial bloodstream infections in the UK. My findings support the inclusion and use of repeated measurement, dynamic time-series data and non-healthcare data in decision-making for AFI.

My specific research questions were addressed in further detail throughout chapters in the thesis, and these are summarised as follows:

1. *Do machine learning clinical prediction models which explicitly account for dynamic changes in data over time offer better predictive performance, robustness and/or clinical utility over conventional static approaches? What are the comparative strengths and weaknesses of such approaches across different healthcare settings?*

This question was answered through 3 contrasting exemplars utilising routinely collected dynamic healthcare data for clinical management in AFI. These examples compare novel approaches to dynamic time-series data with baseline methods:

Chapter 3 investigated the role and use of existing healthcare biomarker data in models which explicitly account for the dynamic time-series relationships in patient information. Models were developed and evaluated to predict bacterial bloodstream infections. I found that through use of a counterfactual approach and data aggregation methods which progressively reduced the time-series component in data – models which account for change over time in general had higher performance over static, time-invariant models. This was further supported by interpretability analyses which show that the change in biomarker values were particularly important for prediction. Such methods were particularly suited for settings where healthcare data can aggregated electronically.

Chapter 4 examined how machine learning models accounting for dynamic change could play a role in dengue management in a data-limited setting. I trained and evaluated a range of machine learning models for the prediction of dengue shock syndrome in hospitalised patients in Vietnam. These models used routine biomarker (haematocrit and platelet) information collected as part of prospective clinical studies. The dynamics of the biomarkers over 48 hours of admission were used as features. Discrimination performances of these models were higher compared to ones which used time-invariant, static data (i.e. results on hospital admission only). Interpretability analysis again support that the magnitude of change in haematocrit and platelet count contributes towards outcome prediction. I found that the data summaries through

minimum, median and maximum values were suitable approaches where data was limited or difficult to collect.

Chapter 5 examined the concept of data as a diagnostic adjunct. I showed that the inclusion of non-healthcare data, namely temperature and rainfall, into clinical models provided added contextual information important for clinical deployment. I used basic full blood count biomarker features collected on attendance in a cohort of children with acute febrile illness to predict final diagnosis. I found that that performance characteristics and predictive values of the base model changed significantly over time as a function of changing disease prevalence. Through development of dynamic models, I demonstrate that the adjustment of thresholds based on recent observations (as a form of continuous learning), or directly accounting for exogenous factors such as rainfall and temperature resulted in systems which perform with significantly greater consistency and clinical utility.

2. *How can machine learning models be implemented effectively into a healthcare setting, and what important factors related to clinical decision-making need to be addressed?*

I investigated the implementation of decision-support models for clinical settings, using dengue machine learning models developed in Chapters 4 and 5. Introduction of interventions need to be appropriate and relevant to healthcare context. In Chapter 6 I explored and characterised existing priorities and decision-making processes of clinicians who manage dengue at the HTD, Vietnam. Through mapping healthcare processes and understanding clinician decision-making in detail, I was able to hold workshops to better characterise ideal features of a clinical decision support system and role in routine patient care. Such methods used in the implementation work, and the specific findings in this context are likely applicable for other technological interventions which arise in the future.

3. *How can the limitations relating to healthcare data availability be addressed? What innovations can be used to improve the acquisition of dynamic clinical data to support decision-making and model development?*

Chapter 7 directly addressed one of the major challenges identified in this thesis for utilising dynamic time-series data for clinical practice. The limited availability of healthcare data in many settings restrict its application in informing decision-making and further model development. Such difficulties in data acquisition are prominent particularly in low- and middle-income country settings and relate to a multitude of

factors: these include the need for healthcare workers to initiate, and collect clinical information (e.g., vital signs measurements), factors restricting access to laboratory diagnostic services (e.g., cost and transport) as well as requirement for invasive sampling. I investigated a continuous monitoring approach with the aim of minimising missing data: such approaches have conventionally been only feasible within intensive care environments, usually in high resource settings. I present research on a novel non-invasive biosensor for continuous real-time lactate measurement, and demonstrate its effectiveness and acceptability in a cohort of healthy volunteers. It is likely that such biosensor and wearable approaches which provide real-time information at point of care would optimise the role and application of dynamic time-series patient data and improve patient care.

8.2 Findings in a wider context

The proposal that medicine should aim to provide care which is predictive, preventative, personalised and participatory (P4 medicine) has been adopted across different healthcare systems [334–336]. Advances in systems biology and platform sciences field related to -omics developments [337], availability of consumer technologies such as mobile smartphones and wearables [333] and the convergence of data sciences methods for understanding healthcare data are attributed to drive such changes [338]. The ultimate goals are that P4 medicine approaches provide better understanding of health and disease, deliver cost-effective interventions, reduce disease impact through prevention and provide a close linkage between science discovery and healthcare delivery.

The thesis provides new understanding on the clinical role of machine learning for utilising longitudinal healthcare data and these findings support the delivery of P4 medicine. Strength of such data-driven approaches are that they can robustly identify patterns and associations from multiple, diverse data sources. The ability to include a large hypothesis-generating component in research is supportive of P4 medicine in which linkage and identification of latent relationships between clinical, genomic, biomarker and physiological data sources provide far greater personalisation to the individual.

It is crucial that data sharing and data linkage can occur in such a way which promotes research. Personal data collected by institutions, across different sectors and

segregated by national boundaries mean that fair data sharing is often a significant challenge [339]. Tensions between the priorities of individual bodies, the need for confidentiality and privacy with potential (but unproven) benefits can be difficult to reconcile. This is particularly in light of data breaches which undermine public confidence [340].

Diverse models have been proposed to address the issues of privacy and confidentiality through use of incentives [341], generation of synthetic data [342], and novel data-sharing structures in which model development takes places without direct access to individual data [343]. Nonetheless problems of data ownership, downstream governance of shared data, verification of its source (with implications on liability when used to guide healthcare decision-making) and interoperability remain pertinent [344]. These issues are particularly relevant to many LMIC settings [345] where there exists a drive for healthcare digitalisation [17] yet access to quality digital data remains restrictive.

In this thesis, work on developing models for dengue prediction utilised prospective research data as this was the only viable data source, given much of routine data remains digitalised. The work also provides support that non-healthcare data (such as rainfall and temperature data) can augment and support these clinical prediction models through providing context. However, it is acknowledged that this use of secondary data is inherently prone to bias and may not be reflective of data from patients undergoing routine clinical care [15]. It is therefore crucial that such models are examined carefully for bias and validated in real-world settings before deployment.

The healthcare data in my thesis focused on information acquired from hospital settings. These datatypes include information related to patient history, clinical, and laboratory investigations. I analysed these data using time-series methods provided by machine learning and non-linear methods. Comparisons between different approaches such as deep learning versus conventional approaches namely multivariable logistic regression on the whole did not result in significantly different performances. This is consistent with current literature in use of clinical data for formulation of prediction scores [346]. Biomarker time-series data used in the chapters are often initiated by healthcare workers, resulting in limited frequency of data collection and restrictions in dataset granularity. The use of such relatively short sequences of data for this work may contribute towards the lack of significant differences between model approaches.

The development of novel modalities which enhance and support dynamic data collection is therefore important and the use of biosensors and wearables will likely play a significant role in healthcare in the near future. Use of such technologies in environments where frequent in-person assessments are not desired (such as for purposes of infection control) or feasible could provide significant advantages [155]. This thesis supports future investigations through providing methodologies for the clinical translation of such data sources.

Additional considerations to translation and model interpretability

Understanding acceptability of machine learning in healthcare is important for clinical implementation [347], particularly for deep learning models which are intrinsically complex and function through a "black-box" approach [348]. Model interpretability impacts on trust and accountability of the system and their predictions. This in turn have implications for liability when healthcare errors do occur in the system, such as a result of hitherto uncharacterised biases [349]. There are arguments that full patient consent to the use of a system which is intrinsically unknowable is not possible, and opposes individual rights [350].

I used post-hoc approaches through counterfactual examples and SHAP analysis were in the thesis to provide better insight into the models. For future studies, it would be important to map clinician understanding of these explanations, particularly given the significant inter-individual variability of clinical practice, and approach to decision-making. Proposals for placing the onus on model developers and regulatory agencies are ongoing. These aim to ensure such machine learning models possess properties of interpretability as much as performance [88, 217, 351]. Guidance frameworks for use of decision support software as a medical device have undergone consultation, and consensus statements relevant to major Western markets issued (USA, UK, and European Union) [147, 149, 352]. Understanding of whether such statements are applicable to other specific countries, particularly those in LMIC settings where data structure, privacy and healthcare delivery differ significantly will be equally important.

8.3 Future perspectives and recommendations

My recommendations which arise from the findings in this thesis are:

- Machine learning methods for clinical prediction provide robust approaches in the development of clinical prediction scores for dengue and bloodstream infections. These techniques are likely of utility when applied to other acute febrile illnesses. Analyses of clinical studies and datasets should aim to include machine learning and data-driven methods in a complementary fashion to conventional analyses in order to maximise understanding of healthcare data and research findings.

Consideration and inclusion of time-series dynamics, where feasible, in clinical prediction models can provide robustness to predictions when compared with traditional linear approaches. This is particularly likely to apply if the patient data population is large and datasets are diverse (tabular, signals and images). Frameworks for data collection in studies and analysis plans should specifically account for this dimension.

- Ensuring interoperability between datasets from clinical records, laboratory, radiology findings and inclusion of epidemiological or other diverse data strongly supports the data-driven and personalised approach to medicine. Healthcare systems in the era of digitalisation should aim to ensure that such data can be available for research whilst addressing privacy and accountability issues related to data sharing.

- Models and interventions should undergo real-world prospective evaluation to understand performance, despite use of statistical methods of cross-validation or hold out set testing. This is given the intrinsic biases in healthcare data collection which may not be apparent until a formal dedicated post-hoc review.

- Model development and prediction analyses whether through machine learning or other methodologies should explicitly account for clinical implementation at the onset. Domains in terms of feasibility of roll-out for the setting, end user preferences and modality of content delivery should be understood and addressed. Ensuring such models are suited for real-world application will

increase the chances of implementation success. Interpretability and accountability of more complex models should also be essential parts of model development to ensure these are fully aligned with the principles of medical ethics.

References

1. World Health Organization. WHO informal consultation on fever management in peripheral health care settings: A global review of evidence and practice. 2013.

2. Crump JA, Kirk MD. Estimating the Burden of Febrile Illnesses. PLOS Neglected Tropical Diseases. 2015;9:e0004040.

3. Holmes AH, Moore LSP, Sundsfjord A, Steinbakk M, Regmi S, Karkey A, et al. Understanding the mechanisms and drivers of antimicrobial resistance. The Lancet. 2016;387:176–87.

4. Robert R. Hoffman, Laura G. Militello. Perspectives on Cognitive Task Analysis | Taylor & Francis Group. 1st edition. Psychology Press; 2008.

5. Bate L, Hutchinson A, Underhill J, Maskrey N. How clinical decisions are made. Br J Clin Pharmacol. 2012;74:614–20.

6. Challener DW, Prokop LJ, Abu-Saleh O. The Proliferation of Reports on Clinical Scoring Systems: Issues About Uptake and Clinical Utility. JAMA. 2019;321:2405–6.

7. Knight SR, Ho A, Pius R, Buchan I, Carson G, Drake TM, et al. Risk stratification of patients admitted to hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: development and validation of the 4C Mortality Score. BMJ. 2020;370:m3339.

8. Centor RM., Witherspoon JM, Dalton HP, Brody CE, Link K. The Diagnosis of Strep Throat in Adults in the Emergency Room. Med Decis Making. 1981;1:239–46.

9. Lim W, van der Eerden MM, Laing R, Boersma W, Karalus N, Town G, et al. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. Thorax. 2003;58:377–82.

10. Durack DT, Lukes AS, Bright DK, Duke Endocarditis Service. New criteria for diagnosis of infective endocarditis: utilization of specific echocardiographic findings. The American Journal of Medicine. 1994;96:200–9.

11. Organization WHealth. Dengue : Guidelines for Diagnosis Treatment Prevention and Control (New Edition 2009). World Health Organization; 2009.

12. Chow JW, Yu VL. Combination antibiotic therapy versus monotherapy for gram-negative bacteraemia: a commentary. International Journal of Antimicrobial Agents. 1999;11:7–12.

13. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. Intensive Care Med. 1996;22:707–10.

14. Eva Silvestre. How Electronic Health Records Strengthen the Health Systems of Low- and Middle-Income Countries: Learning from Eswatini and Mexico. Chapel Hill, NC: MEASURE Evaluation, University of North Carolina; 2018.

15. Näher A-F, Vorisek CN, Klopfenstein SAI, Lehne M, Thun S, Alsalamah S, et al. Secondary data for global health digitalisation. The Lancet Digital Health. 2023;5:e93–101.

16. Topol EJ. Preparing the healthcare workforce to deliver the digital future: An independent report on behalf of the Secretary of State for Health and Social Care. 2019.

17. World Health Organization. Global strategy on digital health 2020-2025. Geneva: World Health Organization; 2021.

18. A plan for digital health and social care.

19. Chau NVV, Hai HB, Greeff H, Quoc KPN, Trieu HT, Khoa LDV, et al. Wearable remote monitoring for patients with COVID-19 in low-resource settings: case study. BMJ Innovations. 2021;7 Suppl 1.

20. Hulsen T, Jamuar SS, Moody AR, Karnes JH, Varga O, Hedensted S, et al. From Big Data to Precision Medicine. Front Med. 2019;6.

21. Arnold RC, Shapiro NI, Jones AE, Schorr C, Pope J, Casner E, et al. Multicenter study of early lactate clearance as a determinant of survival in patients with presumed sepsis. Shock (Augusta, Ga). 2009;32:35–9.

22. Zhang Z, Xu X. Lactate clearance is a useful biomarker for the prediction of all-cause mortality in critically Ill patients: A systematic review and meta-analysis. Critical Care Medicine. 2014;42:2118–25.

23. Ming DK, Myall AC, Hernandez B, Weiße AY, Peach RL, Barahona M, et al. Informing antimicrobial management in the context of COVID-19: understanding the longitudinal dynamics of C-reactive protein and procalcitonin. BMC Infect Dis. 2021;21:932.

24. Royal College of Physicians. National Early Warning Score (NEWS) 2: Standardising the assessment of acute-illness severity in the NHS. Updated report of a working party. RCP; 2017.

25. Karakike E, Kyriazopoulou E, Tsangaris I, Routsi C, Vincent J-L, Giamarellos-Bourboulis EJ. The early change of SOFA score as a prognostic marker of 28-day sepsis mortality: analysis through a derivation and a validation cohort. Crit Care. 2019;23:387.

26. Scott LJ, Tavaré A, Hill EM, Jordan L, Juniper M, Srivastava S, et al. Prognostic value of National Early Warning Scores (NEWS2) and component physiology in hospitalised patients with COVID-19: a multicentre study. Emerg Med J. 2022;39:589–94.

27. Greenhalgh T, Howick J, Maskrey N. Evidence based medicine: a movement in crisis? BMJ. 2014;348:g3725.

28. Allen D, Harkins KJ. Too much guidance? Lancet. 2005;365:1768.

29. Johnstone IM, Titterington DM. Statistical challenges of high-dimensional data. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2009;367:4237–53.

30. Huber JH, Childs ML, Caldwell JM, Mordecai EA. Seasonal temperature variation influences climate suitability for dengue, chikungunya, and Zika transmission. PLOS Neglected Tropical Diseases. 2018;12:e0006451.

31. Stoddard ST, Forshey BM, Morrison AC, Paz-Soldan VA, Vazquez-Prokopec GM, Astete H, et al. House-to-house human movement drives dengue virus transmission. Proc Natl Acad Sci U S A. 2013;110:994–9.

32. Lee HS, Nguyen-Viet H, Nam VS, Lee M, Won S, Duc PP, et al. Seasonal patterns of dengue fever and associated climate factors in 4 provinces in Vietnam from 1994 to 2013. BMC Infectious Diseases. 2017;17:218.

33. Nguyen V-H, Tuyet-Hanh TT, Mulhall J, Minh HV, Duong TQ, Chien NV, et al. Deep learning models for forecasting dengue fever based on climate data in Vietnam. PLOS Neglected Tropical Diseases. 2022;16:e0010509.

34. Scott I, Cook D, Coiera E. Evidence-based medicine and machine learning: a partnership with a common purpose. BMJ Evidence-Based Medicine. 2021;26:290–4.

35. Mitchell TM. Machine Learning. New York: McGraw-Hill; 1997.

36. KPMG and Oxford University Clinical Research Unit. Digital Health in Vietnam - Market Intelligence Report. 2020.

37. Benjamens S, Dhunnoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. npj Digit Med. 2020;3:1–8.

38. A E, B K, Ra N, J K, Sm S, Hm B, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542.

39. Serag A, Ion-Margineanu A, Qureshi H, McMillan R, Saint Martin M-J, Diamond J, et al. Translational AI and Deep Learning in Diagnostic Pathology. Frontiers in Medicine. 2019;6.

40. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global distribution and burden of dengue. Nature. 2013;496:504–7.

41. Array programming with NumPy | Nature. https://www.nature.com/articles/s41586-020-2649-2. Accessed 10 Jan 2022.

42. Reback J, McKinney W, jbrockmendel, Bossche JV den, Augspurger T, Cloud P, et al. pandas-dev/pandas: Pandas 1.0.3. 2020.

43. Hunter JD. Matplotlib: A 2D Graphics Environment. Computing in Science Engineering. 2007;9:90–5.

44. Waskom ML. seaborn: statistical data visualization. Journal of Open Source Software. 2021;6:3021.

45. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods. 2020;17:261–72.

46. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research. 2011;12:2825–30.

47. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016;:785–94.

48. Chollet F, others. Keras. 2015.

49. Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. 2015.

50. McKinney W. Python for Data Analysis : Data Wrangling with Pandas Numpy and Ipython. Second. O'Reilly Media; 2017.

51. Brownlee J. How to Avoid Data Leakage When Performing Data Preparation. MachineLearningMastery.com. 2020. https://machinelearningmastery.com/data-preparation-without-data-leakage/. Accessed 15 Dec 2022.

52. Hamer RM, Simpson PM. Last Observation Carried Forward Versus Mixed Models in the Analysis of Psychiatric Clinical Trials. AJP. 2009;166:639–41.

53. Masking and padding with Keras | TensorFlow Core. https://www.tensorflow.org/guide/keras/masking_and_padding. Accessed 15 Dec 2022.

54. Mack C, Su Z, Westreich D. Types of Missing Data. Agency for Healthcare Research and Quality (US); 2018.

55. 3.1. Cross-validation: evaluating estimator performance. scikit-learn. https://scikit-learn/stable/modules/cross_validation.html. Accessed 15 Feb 2023.

56. Hastie T, Hastie T, Tibshirani R, JH Friedman. The elements of statistical learning : data mining, inference, and prediction. 6th edition. Springer New York; 2001.

57. Stine R. An Introduction to Bootstrap Methods: Examples and Ideas. Sociological Methods & Research. 1989;18:243–91.

58. Efron B, Tibshirani RJ. An introduction to the bootstrap. CRC press; 1994.

59. 4.2. Permutation feature importance. scikit-learn. https://scikit-learn/stable/modules/permutation_importance.html. Accessed 21 Oct 2022.

60. Kuhn M, Johnson K. Applied Predictive Modeling. Springer New York; 2013.

61. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. BMJ. 1995;310:170.

62. Gareth James DW Trevor Hastie, Robert Tibshirani. An introduction to statistical learning : with applications in R. New York : Springer, [2013] ©2013; 2013.

63. Tibshirani R. Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society Series B (Methodological). 1996;58:267–88.

64. Müller AC, Guido S. Introduction to Machine Learning with Python: A Guide for Data Scientists. O'Reilly Media; 2016.

65. Brownlee J. 14 Different Types of Learning in Machine Learning. MachineLearningMastery.com. 2019. https://machinelearningmastery.com/types-of-learning-in-machine-learning/. Accessed 15 Feb 2023.

66. Ho TK. Random Decision Forests. In: Proceedings of the Third International Conference on Document Analysis and Recognition (Volume 1) - Volume 1. M: IEEE Computer Society; 1995. p. 278–82.

67. Breiman L. Classification and Regression Trees. 1st edition. Routledge.; 1984.

68. sklearn.ensemble.RandomForestClassifier. scikit-learn. https://scikit-learn/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html. Accessed 16 Dec 2022.

69. Schapire RE. The strength of weak learnability. Mach Learn. 1990;5:197–227.

70. Schölkopf B, Smola AJ, Williamson RC, Bartlett PL. New Support Vector Algorithms. Neural Computation. 2000;12:1207–45.

71. Hofmann T, Schölkopf B, Smola AJ. Kernel methods in machine learning. The Annals of Statistics. 2008;36:1171–220.

72. Pearson K. LIII. On lines and planes of closest fit to systems of points in space. 1901. https://doi.org/10.1080/14786440109462720.

73. Chollet F. Deep Learning with Python. 1st edition. USA: Manning Publications Co.; 2017.

74. Good IJ. Rational Decisions. Journal of the Royal Statistical Society Series B (Methodological). 1952;14:107–14.

75. Kiefer J, Wolfowitz J. Stochastic Estimation of the Maximum of a Regression Function. The Annals of Mathematical Statistics. 1952;23:462–6.

76. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. 2017.

77. Team K. Keras documentation: Keras API reference. https://keras.io/api/. Accessed 19 Dec 2022.

78. Sherstinsky A. Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network. Physica D: Nonlinear Phenomena. 2020;404:132306.

79. Bagnall A, Lines J, Bostrom A, Large J, Keogh E. The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. Data Min Knowl Disc. 2017;31:606–60.

80. Ruiz AP, Flynn M, Large J, Middlehurst M, Bagnall A. The great multivariate time series classification bake off: a review and experimental evaluation of recent algorithmic advances. Data Min Knowl Disc. 2021;35:401–49.

81. Hochreiter S, Schmidhuber J. Long Short-Term Memory. Neural Computation. 1997;9:1735–80.

82. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention Is All You Need. 2017.

83. sklearn.model_selection.GridSearchCV. scikit-learn. https://scikit-learn/stable/modules/generated/sklearn.model_selection.GridSearchCV.html. Accessed 12 Jan 2023.

84. Frazier PI. A Tutorial on Bayesian Optimization. 2018.

85. O'Malley T, Bursztein E, Long J, Chollet F, Jin H, Invernizzi L, et al. KerasTuner. 2019.

86. Trevethan R. Sensitivity, Specificity, and Predictive Values: Foundations, Pliabilities, and Pitfalls in Research and Practice. Frontiers in Public Health. 2017;5.

87. Youden WJ. Index for rating diagnostic tests. Cancer. 1950;3:32–5.

88. Kolyshkina I, Simoff S. Interpretability of Machine Learning Solutions in Public Healthcare: The CRISP-ML Approach. Frontiers in Big Data. 2021;4.

89. Hart S. Shapley Value. Game Theory. 1989;:210–6.

90. Lundberg SM, Lee S-I. A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems. 2017;30:4765–74.

91. Everett ED, Hirschmann JV. Transient bacteremia and endocarditis prophylaxis. A review. Medicine (Baltimore). 1977;56:61–77.

92. Timsit J-F, Ruppé E, Barbier F, Tabah A, Bassetti M. Bloodstream infections in critically ill patients: an expert statement. Intensive Care Med. 2020;46:266–84.

93. Bai AD, Lo CKL, Komorowski AS, Suresh M, Guo K, Garg A, et al. Staphylococcus aureus bacteraemia mortality: a systematic review and meta-analysis. Clinical Microbiology and Infection. 2022;28:1076–84.

94. Ikuta KS, Swetschinski LR, Aguilar GR, Sharara F, Mestrovic T, Gray AP, et al. Global mortality associated with 33 bacterial pathogens in 2019: a systematic analysis for the Global Burden of Disease Study 2019. The Lancet. 2022;0.

95. Gupta S, Sakhuja A, Kumar G, McGrath E, Nanchal RS, Kashani KB. Culture-Negative Severe Sepsis: Nationwide Trends and Outcomes. Chest. 2016;150:1251–9.

96. Vincent J-L. Defining sepsis (with or without positive blood cultures). The Lancet Child & Adolescent Health. 2017;1:85–6.

97. Laupland KB, Davies HD, Church DL, Louie TJ, Dool JS, Zygun DA, et al. Bloodstream Infection-AssociatedSepsis and Septic Shock in Critically Ill Adults: APopulation-Based Study. Infection. 2004;32:59–64.

98. Phua J, Ngerng WJ, See KC, Tay CK, Kiong T, Lim HF, et al. Characteristics and outcomes of culture-negative versus culture-positive severe sepsis. Critical Care. 2013;17:R202.

99. Paquette K, Sweet D, Stenstrom R, Stabler SN, Lawandi A, Akhter M, et al. Neither Blood Culture Positivity nor Time to Positivity Is Associated With Mortality Among

Patients Presenting With Severe Manifestations of Sepsis: The FABLED Cohort Study. Open Forum Infect Dis. 2021;8:ofab321.

100. Rudd KE, Johnson SC, Agesa KM, Shackelford KA, Tsoi D, Kievlan DR, et al. Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study. The Lancet. 2020;395:200–11.

101. Cuthbertson BH, Elders A, Hall S, Taylor J, MacLennan G, Mackirdy F, et al. Mortality and quality of life in the five years after severe sepsis. Critical Care. 2013;17:R70.

102. Lodise TP, McKinnon PS, Swiderski L, Rybak MJ. Outcomes analysis of delayed antibiotic treatment for hospital-acquired Staphylococcus aureus bacteremia. Clin Infect Dis. 2003;36:1418–23.

103. Rüddel H, Thomas-Rüddel DO, Reinhart K, Bach F, Gerlach H, Lindner M, et al. Adverse effects of delayed antimicrobial treatment and surgical source control in adults with sepsis: results of a planned secondary analysis of a cluster-randomized controlled trial. Critical Care. 2022;26:51.

104. English surveillance programme for antimicrobial utilisation and resistance (ESPAUR) report 2020 to 2021. 2020;:209.

105. Rodríguez-Baño J, López-Prieto MD, Portillo MM, Retamar P, Natera C, Nuño E, et al. Epidemiology and clinical features of community-acquired, healthcare-associated and nosocomial bloodstream infections in tertiary-care and community hospitals. Clin Microbiol Infect. 2010;16:1408–13.

106. UK SMI B 37: investigation of blood cultures (for organisms other than Mycobacterium species). GOV.UK. https://www.gov.uk/government/publications/smi-b-37-investigation-of-blood-cultures-for-organisms-other-than-mycobacterium-species. Accessed 15 Nov 2022.

107. Vincent J-L, Brealey D, Libert N, Abidi NE, O'Dwyer M, Zacharowski K, et al. Rapid Diagnosis of Infection in the Critically Ill, a Multicenter Study of Molecular Detection in Bloodstream Infections, Pneumonia, and Sterile Site Infections*. Crit Care Med. 2015;43:2283–91.

108. Quick J, Ashton P, Calus S, Chatt C, Gossain S, Hawker J, et al. Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of Salmonella. Genome Biol. 2015;16:114.

109. Jordana-Lluch E, Giménez M, Quesada MD, Rivaya B, Marcó C, Domínguez MJ, et al. Evaluation of the Broad-Range PCR/ESI-MS Technology in Blood Specimens for the Molecular Diagnosis of Bloodstream Infections. PLoS One. 2015;10:e0140865.

110. Seifert H. The Clinical Importance of Microbiological Findings in the Diagnosis and Management of Bloodstream Infections. Clinical Infectious Diseases. 2009;48 Supplement_4:S238–45.

111. Cheng MP, Stenstrom R, Paquette K, Stabler SN, Akhter M, Davidson AC, et al. Blood Culture Results Before and After Antimicrobial Administration in Patients With Severe Manifestations of Sepsis. Ann Intern Med. 2019;171:547–54.

112. Baron EJ, Miller JM, Weinstein MP, Richter SS, Gilligan PH, Thomson RB Jr, et al. A Guide to Utilization of the Microbiology Laboratory for Diagnosis of Infectious Diseases: 2013 Recommendations by the Infectious Diseases Society of America (IDSA) and the American Society for Microbiology (ASM)a. Clinical Infectious Diseases. 2013;57:e22–121.

113. Chassagne P, Perol M-B, Doucet J, Trivalle C, Ménard J-F, Manchon N-D, et al. Is presentation of bacteremia in the elderly the same as in younger patients? The American Journal of Medicine. 1996;100:65–70.

114. Avva U, Mueller M. Fever In A Neonate. In: StatPearls. Treasure Island (FL): StatPearls Publishing; 2022.

115. Nannan Panday RS, Wang S, van de Ven PM, Hekker TAM, Alam N, Nanayakkara PWB. Evaluation of blood culture epidemiology and efficiency in a large European teaching hospital. PLoS One. 2019;14:e0214052.

116. Rhodes A, Evans LE, Alhazzani W, Levy MM, Antonelli M, Ferrer R, et al. Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock: 2016. Intensive Care Med. 2017;43:304–77.

117. Mignot-Evers L, Raaijmakers V, Buunk G, Brouns S, Romano L, Herpt T van, et al. Comparison of SIRS criteria and qSOFA score for identifying culture-positive sepsis in the emergency department: a prospective cross-sectional multicentre study. BMJ Open. 2021;11:e041024.

118. Otani T, Ichiba T, Seo K, Naito H. Clinical prediction rule is more useful than qSOFA and the Sepsis-3 definition of sepsis for screening bacteremia. Am J Emerg Med. 2021;46:84–9.

119. Shapiro NI, Wolfe RE, Wright SB, Moore R, Bates DW. Who Needs a Blood Culture? A Prospectively Derived and Validated Prediction Rule. The Journal of Emergency Medicine. 2008;35:255–64.

120. Bates DW, Sands K, Miller E, Lanken PN, Hibberd PL, Graman PS, et al. Predicting Bacteremia in Patients with Sepsis Syndrome. J INFECT DIS. 1997;176:1538–51.

121. Fontanarosa PB, Kaeberlein FJ, Gerson LW, Thomson RB. Difficulty in predicting bacteremia in elderly emergency patients. Annals of Emergency Medicine. 1992;21:842–8.

122. Pfitzenmeyer P, Decrey H, Auckenthaler R, Michel JP. Predicting Bacteremia in Older Patients. Journal of the American Geriatrics Society. 1995;43:230–5.

123. Nestor D, Andersson H, Kihlberg P, Olson S, Ziegler I, Rasmussen G, et al. Early prediction of blood stream infection in a prospectively collected cohort. BMC Infect Dis. 2021;21:316.

124. Roimi M, Neuberger A, Shrot A, Paul M, Geffen Y, Bar-Lavie Y. Early diagnosis of bloodstream infections in the intensive care unit using machine-learning algorithms. Intensive Care Med. 2020;46:454–62.

125. Bhavani SV, Lonjers Z, Carey KA, Afshar M, Gilbert ER, Shah NS, et al. The Development and Validation of a Machine Learning Model to Predict Bacteremia and

Fungemia in Hospitalized Patients Using Electronic Health Record Data. Crit Care Med. 2020;48:e1020–8.

126. Lee KH, Dong JJ, Kim S, Kim D, Hyun JH, Chae M-H, et al. Prediction of Bacteremia Based on 12-Year Medical Data Using a Machine Learning Approach: Effect of Medical Data by Extraction Time. Diagnostics (Basel). 2022;12:102.

127. Choi DH, Hong KJ, Park JH, Shin SD, Ro YS, Song KJ, et al. Prediction of bacteremia at the emergency department during triage and disposition stages using machine learning models. Am J Emerg Med. 2022;53:86–93.

128. Goh V, Chou Y-J, Lee C-C, Ma M-C, Wang WYC, Lin C-H, et al. Predicting Bacteremia among Septic Patients Based on ED Information by Machine Learning Methods: A Comparative Study. Diagnostics (Basel). 2022;12:2498.

129. Lee KH, Dong JJ, Jeong SJ, Chae M-H, Lee BS, Kim HJ, et al. Early Detection of Bacteraemia Using Ten Clinical Variables with an Artificial Neural Network Approach. J Clin Med. 2019;8:E1592.

130. Mahmoud E, Al Dhoayan M, Bosaeed M, Al Johani S, Arabi YM. Developing Machine-Learning Prediction Algorithm for Bacteremia in Admitted Patients. Infect Drug Resist. 2021;14:757–65.

131. Pai K-C, Wang M-S, Chen Y-F, Tseng C-H, Liu P-Y, Chen L-C, et al. An Artificial Intelligence Approach to Bloodstream Infections Prediction. J Clin Med. 2021;10:2901.

132. Park HJ, Jung DY, Ji W, Choi C-M. Detection of Bacteremia in Surgical In-Patients Using Recurrent Neural Network Based on Time Series Records: Development and Validation Study. J Med Internet Res. 2020;22:e19512.

133. Boner Z, Moore CC, Nguyen NR. Deep Learning Risk Prediction of Bloodstream Infection in the Intensive Care Unit. 2022;:7.

134. Robineau O, Robert J, Rabaud C, Bedos J-P, Varon E, Péan Y, et al. Management and outcome of bloodstream infections: a prospective survey in 121 French hospitals (SPA-BACT survey). Infect Drug Resist. 2018;11:1359–68.

135. The vanishing gradient problem during learning recurrent neural nets and problem solutions | International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems. https://dl.acm.org/doi/10.1142/S0218488598000094. Accessed 13 Jan 2023.

136. Cho K, van Merrienboer B, Bahdanau D, Bengio Y. On the Properties of Neural Machine Translation: Encoder-Decoder Approaches. 2014.

137. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. 2017.

138. Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization: ACM Transactions on Mathematical Software: Vol 23, No 4. https://dl.acm.org/doi/10.1145/279232.279236. Accessed 13 Jan 2023.

139. mateuszbuda. Machine Learning Statistical Utils. 2022.

140. Seale AC, Gordon NC, Islam J, Peacock SJ, Scott JAG. AMR Surveillance in low and middle-income settings - A roadmap for participation in the Global Antimicrobial Surveillance System (GLASS). Wellcome Open Res. 2017;2:92.

141. Fukuda K, Limmathurotsakul D, Okeke I, Shetty N, van Doorn R, Feasey N, et al. Surveillance and Epidemiology of Drug Resistant Infections Consortium (SEDRIC): Supporting the transition from strategy to action [version 2; peer review: 3 approved]. Wellcome Open Research. 2018;3.

142. Deeny SR, van Kleef E, Bou-Antoun S, Hope RJ, Robotham JV. Seasonal changes in the incidence of Escherichia coli bloodstream infection: variation with region and place of onset. Clinical Microbiology and Infection. 2015;21:924–9.

143. Sacchetti B, Travis J, Steed LL, Webb G. Effects of COVID-19 on Blood Culture Contamination at a Tertiary Care Academic Medical Center. Microbiol Spectr. 10:e00277-22.

144. Lee CS, Lee AY. Applications of continual learning machine learning in clinical practice. Lancet Digit Health. 2020;2:e279–81.

145. Samuel G, Derrick G. Defining ethical standards for the application of digital tools to population health research. Bull World Health Organ. 2020;98:239–44.

146. Challen R, Denny J, Pitt M, Gompels L, Edwards T, Tsaneva-Atanasova K. Artificial intelligence, bias and clinical safety. BMJ Qual Saf. 2019;28:231–7.

147. International Medical Device Regulators Forum. Software as a Medical Device (SaMD): Key definitions. 2013; IMDRF/SaMD WG/N10FINAL:2013:9.

148. European Comission. On Artificial Intelligence - A European approach to excellence and trust. 2020. https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf. Accessed 28 Nov 2022.

149. Good Machine Learning Practice for Medical Device Development: Guiding Principles. GOV.UK. https://www.gov.uk/government/publications/good-machine-learning-practice-for-medical-device-development-guiding-principles/good-machine-learning-practice-for-medical-device-development-guiding-principles. Accessed 28 Nov 2022.

150. Adamson AS, Welch HG. Machine Learning and the Cancer-Diagnosis Problem — No Gold Standard. N Engl J Med. 2019;381:2285–7.

151. Chen P-HC, Mermel CH, Liu Y. Evaluation of artificial intelligence on a reference standard based on subjective interpretation. The Lancet Digital Health. 2021;3:e693–5.

152. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nat Med. 2020;26:1364–74.

153. Plana D, Shung DL, Grimshaw AA, Saraf A, Sung JJY, Kann BH. Randomized Clinical Trials of Machine Learning Interventions in Health Care: A Systematic Review. JAMA Network Open. 2022;5:e2233946.

154. Lamy B. Blood culture time-to-positivity: making use of the hidden information. Clinical Microbiology and Infection. 2019;25:268–71.

155. Ming DK, Sangkaew S, Chanh HQ, Nhat PTH, Yacoub S, Georgiou P, et al. Continuous physiological monitoring using wearable technology to inform individual management of infectious diseases, public health and outbreak responses. Int J Infect Dis. 2020;96:648–54.

156. Abebe Tadesse G, Javed H, Thanh NLN, Ha Thai HD, Le Van T, Thwaites L, et al. Multi-modal Diagnosis of Infectious Diseases in the Developing World. IEEE Journal of Biomedical and Health Informatics. 2020;:1–1.

157. Galbo H, Kall L. Circadian variations in clinical symptoms and concentrations of inflammatory cytokines, melatonin, and cortisol in polymyalgia rheumatica before and during prednisolone treatment: a controlled, observational, clinical experimental study. Arthritis Research & Therapy. 2016;18:174.

158. Gwee SXW, St John AL, Gray GC, Pang J. Animals as potential reservoirs for dengue transmission: A systematic review. One Health. 2021;12:100216.

159. Kolimenakis A, Heinz S, Wilson ML, Winkler V, Yakob L, Michaelakis A, et al. The role of urbanisation in the spread of Aedes mosquitoes and the diseases they transmit—A systematic review. PLoS Negl Trop Dis. 2021;15:e0009631.

160. Brady OJ, Gething PW, Bhatt S, Messina JP, Brownstein JS, Hoen AG, et al. Refining the global spatial limits of dengue virus transmission by evidence-based consensus. PLoS Negl Trop Dis. 2012;6:e1760.

161. Global Burden of Diseases. Global Health Metrics - dengue. 2020.

162. Hung TM, Shepard DS, Bettis AA, Nguyen HA, McBride A, Clapham HE, et al. Productivity costs from a dengue episode in Asia: a systematic literature review. BMC Infect Dis. 2020;20:393.

163. Shepard DS, Undurraga EA, Halasa YA. Economic and disease burden of dengue in Southeast Asia. PLoS Negl Trop Dis. 2013;7:e2055.

164. Chan M, Johansson MA. The Incubation Periods of Dengue Viruses. PLoS One. 2012;7:e50972.

165. Wilder-Smith A, Ooi E-E, Horstick O, Wills B. Dengue. The Lancet. 2019;393:350–63.

166. Trung DT, Thao LTT, Hien TT, Hung NT, Vinh NN, Hien PTD, et al. Liver Involvement Associated with Dengue Infection in Adults in Vietnam. Am J Trop Med Hyg. 2010;83:774–80.

167. World Health Organization. Dengue guidelines for diagnosis, treatment, prevention and control : new edition. World Health Organization; 2009.

168. Lam PK, McBride A, Le DHT, Huynh TT, Vink H, Wills B, et al. Visual and Biochemical Evidence of Glycocalyx Disruption in Human Dengue Infection, and Association With Plasma Leakage Severity. Frontiers in Medicine. 2020;7.

169. Lam PK, Tam DTH, Diet TV, Tam CT, Tien NTH, Kieu NTT, et al. Clinical Characteristics of Dengue Shock Syndrome in Vietnamese Children: A 10-Year Prospective Study in a Single Hospital. Clin Infect Dis. 2013;57:1577–86.

170. Trieu HT, Khanh LP, Ming DKY, Quang CH, Phan TQ, Van VCN, et al. The compensatory reserve index predicts recurrent shock in patients with severe dengue. BMC Medicine. 2022;20:109.

171. St. John AL, Rathore APS. Adaptive immune responses to primary and secondary dengue virus infections. Nat Rev Immunol. 2019;19:218–30.

172. Sabin AB. Research on dengue during World War II. Am J Trop Med Hyg. 1952;1:30–50.

173. Montoya M, Gresh L, Mercado JC, Williams KL, Vargas MJ, Gutierrez G, et al. Symptomatic versus inapparent outcome in repeat dengue virus infections is influenced by the time interval between infections and study year. PLoS Negl Trop Dis. 2013;7:e2357.

174. Rothman AL. Immunity to dengue virus: a tale of original antigenic sin and tropical cytokine storms. Nat Rev Immunol. 2011;11:532–43.

175. Katzelnick LC, Gresh L, Halloran ME, Mercado JC, Kuan G, Gordon A, et al. Antibody-dependent enhancement of severe dengue disease in humans. Science. 2017;358:929–32.

176. Tricou V, Minh NN, Farrar J, Tran HT, Simmons CP. Kinetics of Viremia and NS1 Antigenemia Are Shaped by Immune Status and Virus Serotype in Adults with Dengue. PLOS Neglected Tropical Diseases. 2011;5:e1309.

177. Khor CC, Chau TNB, Pang J, Davila S, Long HT, Ong RTH, et al. Genome-wide association study identifies susceptibility loci for dengue shock syndrome at MICB and PLCE1. Nat Genet. 2011;43:1139–41.

178. Whitehorn J, Nguyen CVV, Khanh LP, Kien DTH, Quyen NTH, Tran NTT, et al. Lovastatin for the Treatment of Adult Patients With Dengue: A Randomized, Double-Blind, Placebo-Controlled Trial. Clin Infect Dis. 2016;62:468–76.

179. Tam DTH, Ngoc TV, Tien NTH, Kieu NTT, Thuy TTT, Thanh LTC, et al. Effects of short-course oral corticosteroid therapy in early dengue infection in Vietnamese patients: a randomized, placebo-controlled trial. Clin Infect Dis. 2012;55:1216–24.

180. Nguyen NM, Tran CNB, Phung LK, Duong KTH, Huynh H le A, Farrar J, et al. A randomized, double-blind placebo controlled trial of balapiravir, a polymerase inhibitor, in adult dengue patients. J Infect Dis. 2013;207:1442–50.

181. Low JG, Sung C, Wijaya L, Wei Y, Rathore APS, Watanabe S, et al. Efficacy and safety of celgosivir in patients with dengue fever (CELADEN): a phase 1b, randomised, double-blind, placebo-controlled, proof-of-concept trial. Lancet Infect Dis. 2014;14:706–15.

182. Tricou V, Minh NN, Van TP, Lee SJ, Farrar J, Wills B, et al. A randomized controlled trial of chloroquine for the treatment of dengue in Vietnamese adults. PLoS Negl Trop Dis. 2010;4:e785.

183. Ng TC, Teo CH, Toh JY, Dunn AG, Ng CJ, Ang TF, et al. Factors influencing healthcare seeking in patients with dengue: Systematic review. Tropical Medicine & International Health. 2022;27:13–27.

184. World Health Organization. Global Strategy for Dengue Prevention and Control. 2012.

185. Barniol J, Gaczkowski R, Barbato EV, da Cunha RV, Salgado D, Martínez E, et al. Usefulness and applicability of the revised dengue case classification by disease: multi-centre study in 18 countries. BMC Infectious Diseases. 2011;11:106.

186. Morra ME, Altibi AMA, Iqtadar S, Minh LHN, Elawady SS, Hallab A, et al. Definitions for warning signs and signs of severe dengue according to the WHO 2009 classification: Systematic review of literature. Reviews in Medical Virology. 2018;28:e1979.

187. Srikiatkhachorn A, Rothman AL, Gibbons RV, Sittisombut N, Malasit P, Ennis FA, et al. Dengue—How Best to Classify It. Clinical Infectious Diseases. 2011;53:563–7.

188. Tomashek KM, Wills B, See Lum LC, Thomas L, Durbin A, Leo Y-S, et al. Development of standard clinical endpoints for use in dengue interventional trials. PLoS Negl Trop Dis. 2018;12:e0006497.

189. Sangkaew S, Ming D, Boonyasiri A, Honeyford K, Kalayanarooj S, Yacoub S, et al. Risk predictors of progression to severe disease during the febrile phase of dengue: a systematic review and meta-analysis. The Lancet Infectious Diseases. 2021;21:1014–26.

190. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. European Heart Journal. 2014;35:1925–31.

191. Steyerberg EW. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. Cham: Springer International Publishing; 2019.

192. Nguyen MT, Ho TN, Nguyen VVC, Nguyen TH, Ha MT, Ta VT, et al. An Evidence-Based Algorithm for Early Prognosis of Severe Dengue in the Outpatient Setting. Clin Infect Dis. 2017;64:656–63.

193. Lam PK, Tam DTH, Dung NM, Tien NTH, Kieu NTT, Simmons C, et al. A Prognostic Model for Development of Profound Shock among Children Presenting with Dengue Shock Syndrome. PLOS ONE. 2015;10:e0126134.

194. Lee I-K, Liu J-W, Chen Y-H, Chen Y-C, Tsai C-Y, Huang S-Y, et al. Development of a Simple Clinical Risk Score for Early Prediction of Severe Dengue in Adult Patients. PLoS ONE. 2016;11:e0154772.

195. Pang J, Lindblom A, Tolfvenstam T, Thein T-L, Naim ANM, Ling L, et al. Discovery and Validation of Prognostic Biomarker Models to Guide Triage among Adult Dengue Patients at Early Infection. PLOS ONE. 2016;11:e0155993.

196. Tan KW, Tan B, Thein TL, Leo Y-S, Lye DC, Dickens BL, et al. Dynamic dengue haemorrhagic fever calculators as clinical decision support tools in adult dengue. Trans R Soc Trop Med Hyg. 2020;114:7–15.

197. Goodfellow I, Bengio Y, Courville A. Deep Learning. MIT Press; 2016.

198. Lam PK, Ngoc TV, Thu Thuy TT, Hong Van NT, Nhu Thuy TT, Hoai Tam DT, et al. The value of daily platelet counts for predicting dengue shock syndrome: Results from

a prospective observational study of 2301 Vietnamese children with dengue. PLoS Negl Trop Dis. 2017;11.

199. Tam DTH, Ngoc TV, Tien NTH, Kieu NTT, Thuy TTT, Thanh LTC, et al. Effects of Short-Course Oral Corticosteroid Therapy in Early Dengue Infection in Vietnamese Patients: A Randomized, Placebo-Controlled Trial. Clin Infect Dis. 2012;55:1216–24.

200. Tuan NM, Nhan HT, Chau NVV, Hung NT, Tuan HM, Tram TV, et al. Sensitivity and Specificity of a Novel Classifier for the Early Diagnosis of Dengue. PLoS Negl Trop Dis. 2015;9.

201. Muller DA, Depelsenaire ACI, Young PR. Clinical and Laboratory Diagnosis of Dengue Virus Infection. The Journal of Infectious Diseases. 2017;215 suppl_2:S89–95.

202. National High Blood Pressure Education Program Working Group on High Blood Pressure in Children and Adolescents. The fourth report on the diagnosis, evaluation, and treatment of high blood pressure in children and adolescents. Pediatrics. 2004;114 2 Suppl 4th Report:555–76.

203. Breiman L. Random Forests. Machine Learning. 2001;45:5–32.

204. Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995;20:273–97.

205. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw. 2010;33:1–22.

206. Smits N. A note on Youden's Jand its cost ratio. BMC Medical Research Methodology. 2010;10:89.

207. Rocca E. Probability for the Clinical Encounter. In: Anjum RL, Copeland S, Rocca E, editors. Rethinking Causality, Complexity and Evidence for the Unique Patient: A CauseHealth Resource for Healthcare Professionals and the Clinical Encounter. Cham: Springer International Publishing; 2020. p. 37–54.

208. Cearns M, Hahn T, Clark S, Baune BT. Machine learning probability calibration for high-risk clinical decision-making. Aust N Z J Psychiatry. 2020;54:123–6.

209. Brier GW. VERIFICATION OF FORECASTS EXPRESSED IN TERMS OF PROBABILITY. Monthly Weather Review. 1950;78:1–3.

210. 1.16. Probability calibration. scikit-learn. https://scikit-learn/stable/modules/calibration.html. Accessed 13 Oct 2022.

211. Amann J, Blasimme A, Vayena E, Frey D, Madai VI, the Precise4Q consortium. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak. 2020;20:310.

212. Thai KTD, Nishiura H, Hoang PL, Tran NTT, Phan GT, Le HQ, et al. Age-Specificity of Clinical Dengue during Primary and Secondary Infections. PLoS Negl Trop Dis. 2011;5:e1180.

213. Corada M, Mariotti M, Thurston G, Smith K, Kunkel R, Brockhaus M, et al. Vascular endothelial-cadherin is an important determinant of microvascular integrity in vivo. Proc Natl Acad Sci USA. 1999;96:9815–20.

214. Wang W-K, Chao D-Y, Kao C-L, Wu H-C, Liu Y-C, Li C-M, et al. High Levels of Plasma Dengue Viral Load during Defervescence in Patients with Dengue Hemorrhagic Fever: Implications for Pathogenesis. Virology. 2003;305:330–8.

215. Kleinert S, Horton R. Pathology and laboratory medicine: the Cinderella of health systems. The Lancet. 2018;391:1872–3.

216. Turner HC, Hao NV, Yacoub S, Hoang VMT, Clifton DA, Thwaites GE, et al. Achieving affordable critical care in low-income and middle-income countries. BMJ Global Health. 2019;4:e001675.

217. Yoon CH, Torrance R, Scheinerman N. Machine learning in medicine: should the pursuit of enhanced interpretability be abandoned? Journal of Medical Ethics. 2021. https://doi.org/10.1136/medethics-2020-107102.

218. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. jair. 2002;16:321–57.

219. Crump JA, Newton PN, Baird SJ, Lubell Y. Febrile Illness in Adolescents and Adults. In: Holmes KK, Bertozzi S, Bloom BR, Jha P, editors. Major Infectious Diseases. 3rd edition. Washington (DC): The International Bank for Reconstruction and Development / The World Bank; 2017.

220. Vos T, Lim SS, Abbafati C, Abbas KM, Abbasi M, Abbasifard M, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. The Lancet. 2020;396:1204–22.

221. Drugs I of M (US) C on the E of A, Arrow KJ, Panosian C, Gelband H. A Brief History of Malaria. National Academies Press (US); 2004.

222. World Health Organization. Global Malaria Programme. World malaria report 2017.

223. Choisy M, McBride A, Chambers M, Quang CH, Quang HN, Chau NTX, et al. Climate change and health in Southeast Asia – defining research priorities and the role of the Wellcome Trust Africa Asia Programmes. 2021.

224. Mayxay M, Castonguay-Vanier J, Chansamouth V, Dubot-Pérès A, Paris DH, Phetsouvanh R, et al. Causes of non-malarial fever in Laos: a prospective study. Lancet Glob Health. 2013;1:e46-54.

225. Kasper MR, Blair PJ, Touch S, Sokhal B, Yasuda CY, Williams M, et al. Infectious etiologies of acute febrile illness among patients seeking health care in south-central Cambodia. The American journal of tropical medicine and hygiene. 2012;86:246–53.

226. Messacar K, Parker SK, Todd JK, Dominguez SR. Implementation of Rapid Molecular Infectious Disease Diagnostics: the Role of Diagnostic and Antimicrobial Stewardship. J Clin Microbiol. 2017;55:715–23.

227. Luz PM, Vanni T, Medlock J, Paltiel AD, Galvani AP. Dengue vector control strategies in an urban setting: an economic modelling assessment. Lancet. 2011;377:1673–80.

228. Budd J, Miller BS, Manning EM, Lampos V, Zhuang M, Edelstein M, et al. Digital technologies in the public-health response to COVID-19. Nat Med. 2020;26:1183–92.

229. Fleming KA, Horton S, Wilson ML, Atun R, DeStigter K, Flanigan J, et al. The Lancet Commission on diagnostics: transforming access to diagnostics. The Lancet. 2021;398:1997–2050.

230. Wilson ML, Fleming KA, Kuti MA, Looi LM, Lago N, Ru K. Access to pathology and laboratory medicine services: a crucial gap. The Lancet. 2018;391:1927–38.

231. Horton S, Sullivan R, Flanigan J, Fleming KA, Kuti MA, Looi LM, et al. Delivering modern, high-quality, affordable pathology and laboratory medicine to low-income and middle-income countries: a call to action. The Lancet. 2018;391:1953–64.

232. Hopkins H, Bruxvoort KJ, Cairns ME, Chandler CIR, Leurent B, Ansah EK, et al. Impact of introduction of rapid diagnostic tests for malaria on antibiotic prescribing: analysis of observational and randomised studies in public and private healthcare settings. BMJ. 2017;356:j1054.

233. Teoh B-T, Sam S-S, Tan K-K, Johari J, Abd-Jamil J, Hooi P-S, et al. The Use of NS1 Rapid Diagnostic Test and qRT-PCR to Complement IgM ELISA for Improved Dengue Diagnosis from Single Specimen. Sci Rep. 2016;6:27663.

234. Hang VT, Nguyet NM, Trung DT, Tricou V, Yoksan S, Dung NM, et al. Diagnostic Accuracy of NS1 ELISA and Lateral Flow Rapid Tests for Dengue Sensitivity, Specificity and Relationship to Viraemia and Antibody Responses. PLOS Neglected Tropical Diseases. 2009;3:e360.

235. Yacoub S, Kotit S, Yacoub MH. Disease appearance and evolution against a background of climate change and reduced resources. Philos Trans A Math Phys Eng Sci. 2011;369:1719–29.

236. Whitehorn J, Yacoub S. Global warming and arboviral infections. Clin Med (Lond). 2019;19:149–52.

237. Li C, Wang X, Wu X, Liu J, Ji D, Du J. Modeling and projection of dengue fever cases in Guangzhou based on variation of weather factors. Science of The Total Environment. 2017;605–606:867–73.

238. Colón-González FJ, Bastos LS, Hofmann B, Hopkin A, Harpham Q, Crocker T, et al. Probabilistic seasonal dengue forecasting in Vietnam: A modelling study using superensembles. PLOS Medicine. 2021;18:e1003542.

239. Kiang MV, Santillana M, Chen JT, Onnela J-P, Krieger N, Engø-Monsen K, et al. Incorporating human mobility data improves forecasts of Dengue fever in Thailand. Sci Rep. 2021;11.

240. Rangarajan P, Mody SK, Marathe M. Forecasting dengue and influenza incidences using a sparse representation of Google trends, electronic health records, and time series data. PLoS Comput Biol. 2019;15:e1007518.

241. Bonifay T, Vesin G, Bidaud B, Bonnefoy C, Dueymes M, Nacher M, et al. Clinical characteristics and predictive score of dengue vs. chikungunya virus infections. Médecine et Maladies Infectieuses. 2019;49:250–6.

242. Rambaud-Althaus C, Shao AF, Kahama-Maro J, Genton B, d'Acremont V. Managing the Sick Child in the Era of Declining Malaria Transmission: Development of

ALMANACH, an Electronic Algorithm for Appropriate Use of Antimicrobials. PLOS ONE. 2015;10:e0127674.

243. Blanchet K, Lewis JJ, Pozo-Martin F, Satouro A, Somda S, Ilboudo P, et al. A mixed methods protocol to evaluate the effect and cost-effectiveness of an Integrated electronic Diagnosis Approach (IeDA) for the management of childhood illnesses at primary health facilities in Burkina Faso. Implementation Sci. 2015;11:111.

244. Integrated management of childhood illness. https://www.who.int/teams/maternal-newborn-child-adolescent-health-and-ageing/child-health/integrated-management-of-childhood-illness. Accessed 18 Oct 2022.

245. Pacific WHORO for the W. Dengue Situation Updates 2022. 2022.

246. Ebi KL, Nealon J. Dengue in a changing climate. Environmental Research. 2016;151:115–23.

247. Murray NEA, Quam MB, Wilder-Smith A. Epidemiology of dengue: past, present and future prospects. Clin Epidemiol. 2013;5:299–309.

248. Chen Y, Li N, Lourenço J, Wang L, Cazelles B, Dong L, et al. Measuring the effects of COVID-19-related disruption on dengue transmission in southeast Asia and Latin America: a statistical modelling study. The Lancet Infectious Diseases. 2022;22:657–67.

249. World Health Organization. The selection and use of essential in vitro diagnostics - TRS 1031.

250. World Bank Climate Change Knowledge Portal. https://climateknowledgeportal.worldbank.org/. Accessed 4 Nov 2021.

251. sklearn.model_selection.LeaveOneGroupOut. scikit-learn. https://scikit-learn/stable/modules/generated/sklearn.model_selection.LeaveOneGroupOut.html. Accessed 19 Oct 2022.

252. Berndt DJ, Clifford J. Using dynamic time warping to find patterns in time series. In: Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining. Seattle, WA: AAAI Press; 1994. p. 359–70.

253. Do TTT, Martens P, Luu NH, Wright P, Choisy M. Climatic-driven seasonality of emerging dengue fever in Hanoi, Vietnam. BMC Public Health. 2014;14:1078.

254. Ming D, Rawson T, Sangkaew S, Rodriguez-Manzano J, Georgiou P, Holmes A. Connectivity of rapid-testing diagnostics and surveillance of infectious diseases. Bull World Health Organ. 2019;97:242–4.

255. Pham VH, Nguyet DPH, Mai KNH, Truong KH, Huynh LV, Pham THT, et al. Measles Epidemics Among Children in Vietnam: Genomic Characterization of Virus Responsible for Measles Outbreak in Ho Chi Minh City, 2014. EBioMedicine. 2014;1:133–40.

256. Rocklöv J, Dubrow R. Climate change: an enduring challenge for vector-borne disease prevention and control. Nat Immunol. 2020;21:479–83.

257. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ. 1996;312:71–2.

258. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. npj Digital Medicine. 2020;3:1–10.

259. Bates DW, Leape LL, Cullen DJ, Laird N, Petersen LA, Teich JM, et al. Effect of computerized physician order entry and a team intervention on prevention of serious medication errors. JAMA. 1998;280:1311–6.

260. McMullin ST, Lonergan TP, Rynearson CS. Twelve-month drug cost savings related to use of an electronic prescribing system with integrated decision support in primary care. J Manag Care Pharm. 2005;11:322–32.

261. Kwok R, Dinh M, Dinh D, Chu M. Improving adherence to asthma clinical guidelines and discharge documentation from emergency departments: implementation of a dynamic and integrated electronic decision support system. Emerg Med Australas. 2009;21:31–7.

262. Kwan JL, Lo L, Ferguson J, Goldberg H, Diaz-Martinez JP, Tomlinson G, et al. Computerised clinical decision support systems and absolute improvements in care: meta-analysis of controlled clinical trials. BMJ. 2020;370:m3216.

263. Peiffer-Smadja N, Rawson TM, Ahmad R, Buchard A, Georgiou P, Lescure F-X, et al. Machine learning for clinical decision support in infectious diseases: a narrative review of current applications. Clinical Microbiology and Infection. 2020;26:584–95.

264. Mechelli A, Vieira S. From models to tools: clinical translation of machine learning studies in psychosis. npj Schizophr. 2020;6:1–3.

265. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond Adoption: A New Framework for Theorizing and Evaluating Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies. Journal of Medical Internet Research. 2017;19:e367.

266. Bates DW, Kuperman GJ, Wang S, Gandhi T, Kittler A, Volk L, et al. Ten Commandments for Effective Clinical Decision Support: Making the Practice of Evidence-based Medicine a Reality. J Am Med Inform Assoc. 2003;10:523–30.

267. Medical Research Council. Developing and evaluating complex interventions. 2006.

268. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ. 2021;374:n2061.

269. Sanders D, Haines A. Implementation Research Is Needed to Achieve International Health Goals. PLOS Medicine. 2006;3:e186.

270. Bernasconi A, Crabbé F, Adedeji AM, Bello A, Schmitz T, Landi M, et al. Results from one-year use of an electronic Clinical Decision Support System in a post-conflict context: An implementation research. PLOS ONE. 2019;14:e0225634.

271. International Monetary Fund. Regional Economical Outlook: Sub-Saharan Africa. 2020.

272. ISO. Human-centred design for interactive systems: ISO9241-210.

273. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implementation Science. 2009;4:50.

274. NHS England and NHS Improvement. Mapping the last 10 patients.

275. Walker D, Myrick F. Grounded theory: an exploration of process and procedure. Qual Health Res. 2006;16:547–59.

276. SAGE Research Methods - Encyclopedia of Case Study Research. https://methods.sagepub.com/reference/encyc-of-case-study-research. Accessed 14 Dec 2022.

277. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology. 2006;3:77–101.

278. Experience WL in R-BU. Personas Make Users Memorable for Product Team Members. Nielsen Norman Group. https://www.nngroup.com/articles/persona/. Accessed 8 Dec 2022.

279. Experience WL in R-BU. The Definition of User Experience (UX). Nielsen Norman Group. https://www.nngroup.com/articles/definition-user-experience/. Accessed 5 Dec 2022.

280. Experience WL in R-BU. Tabs, Used Right. Nielsen Norman Group. https://www.nngroup.com/articles/tabs-used-right/. Accessed 8 Dec 2022.

281. Marwaha JS, Landman AB, Brat GA, Dunn T, Gordon WJ. Deploying digital health tools within large, complex health systems: key considerations for adoption and implementation. npj Digit Med. 2022;5:1–7.

282. Nielsen J. Iterative user-interface design. Computer. 1993;26:32–41.

283. Wilson FP, Martin M, Yamamoto Y, Partridge C, Moreira E, Arora T, et al. Electronic health record alerts for acute kidney injury: multicenter, randomized clinical trial. BMJ. 2021;372:m4786.

284. Samuel G, Derrick G. Defining ethical standards for the application of digital tools to population health research. Bull World Health Organ. 2020;98:239–44.

285. National Institude for Health and Care Excellence. QuikRead go for C-reactive protein testing in primary care. Medtech innovation briefing. :16.

286. Ferguson BS, Rogatzki MJ, Goodwin ML, Kane DA, Rightmire Z, Gladden LB. Lactate metabolism: historical context, prior misinterpretations, and current understanding. Eur J Appl Physiol. 2018;118:691–728.

287. Qvisth V, Hagström-Toft E, Enoksson S, Bolinder J. Catecholamine regulation of local lactate production in vivo in skeletal muscle and adipose tissue: role of -adrenoreceptor subtypes. J Clin Endocrinol Metab. 2008;93:240–6.

288. Villar J, Short JH, Lighthall G. Lactate Predicts Both Short- and Long-Term Mortality in Patients With and Without Sepsis. Infect Dis (Auckl). 2019;12:1178633719862776.

289. Kruse O, Grunnet N, Barfod C. Blood lactate as a predictor for in-hospital mortality in patients admitted acutely to hospital: a systematic review. Scand J Trauma Resusc Emerg Med. 2011;19:74.

290. Bakker J, Gris P, Coffernils M, Kahn RJ, Vincent J-L. Serial blood lactate levels can predict the development of multiple organ failure following septic shock. The American Journal of Surgery. 1996;171:221–6.

291. Casserly B, Phillips GS, Schorr C, Dellinger RP, Townsend SR, Osborn TM, et al. Lactate measurements in sepsis-induced tissue hypoperfusion: results from the Surviving Sepsis Campaign database. Crit Care Med. 2015;43:567–73.

292. Yacoub S, Trung TH, Lam PK, Thien VHN, Hai DHT, Phan TQ, et al. Cardio-haemodynamic assessment and venous lactate in severe dengue: Relationship with recurrent shock and respiratory distress. PLoS Negl Trop Dis. 2017;11:e0005740.

293. Krishna S, Waller DW, Kuile F ter, Kwiatkowski D, Crawley J, Craddock CFC, et al. Lactic acidosis and hypoglycaemia in children with severe malaria: pathophysiological and prognostic significance. Transactions of the Royal Society of Tropical Medicine and Hygiene. 1994;88:67–73.

294. Aslar AK, Kuzu MA, Elhan AH, Tanik A, Hengirmen S. Admission lactate level and the APACHE II score are the most useful predictors of prognosis following torso trauma. Injury. 2004;35:746–52.

295. Govender P, Tosh W, Burt C, Falter F. Evaluation of Increase in Intraoperative Lactate Level as a Predictor of Outcome in Adults After Cardiac Surgery. J Cardiothorac Vasc Anesth. 2020;34:877–84.

296. Arnold R, Zhang Z, Patel S, Smola S, Isserman J, Jackson E. Delayed assessment of serum lactate in sepsis is associated with an increased mortality rate. Critical Care. 2014;18 Suppl 1:P174.

297. Aramburo A, Todd J, George EC, Kiguli S, Olupot-Olupot P, Opoka RO, et al. Lactate clearance as a prognostic marker of mortality in severely ill febrile children in East Africa. BMC medicine. 2018;16:37.

298. Marbach JA, Stone S, Schwartz B, Pahuja M, Thayer KL, Faugno AJ, et al. Lactate Clearance Is Associated With Improved Survival in Cardiogenic Shock: A Systematic Review and Meta-Analysis of Prognostic Factor Studies. J Card Fail. 2021;27:1082–9.

299. Vincent J-L, Quintairos E Silva A, Couto L, Taccone FS, Taccone FS. The value of blood lactate kinetics in critically ill patients: a systematic review. Critical care (London, England). 2016;20:257.

300. Jansen TC, van Bommel J, Schoonderbeek FJ, Sleeswijk Visser SJ, van der Klooster JM, Lima AP, et al. Early Lactate-Guided Therapy in Intensive Care Unit Patients. American Journal of Respiratory and Critical Care Medicine. 2010;182:752–61.

301. Graham CA, Leung LY, Lo RS, Lee KH, Yeung CY, Chan SY, et al. Agreement between capillary and venous lactate in emergency department patients: prospective observational study. BMJ Open. 2019;9:e026109.

302. Daurat A, Dick M, Louart B, Lefrant J-Y, Muller L, Roger C. Continuous lactate monitoring in critically ill patients using microdialysis. Anaesthesia Critical Care & Pain Medicine. 2020;39:513–7.

303. Gil B, Ip H, Kassanos P, Lo B, Yang G-Z, Anastasova S. Smart implanted access port catheter for therapy intervention with pH and lactate biosensors. Materials Today Bio. 2022;15:100298.

304. Lin C-E, Hiraka K, Matloff D, Johns J, Deng A, Sode K, et al. Development toward a novel integrated tear lactate sensor using Schirmer test strip and engineered lactate oxidase. Sensors and Actuators B: Chemical. 2018;270:525–9.

305. Schabmueller CGJ, Loppow D, Piechotta G, Schütze B, Albers J, Hintsche R. Micromachined sensor for lactate monitoring in saliva. Biosensors and Bioelectronics. 2006;21:1770–6.

306. Nagamine K, Mano T, Nomura A, Ichimura Y, Izawa R, Furusawa H, et al. Noninvasive Sweat-Lactate Biosensor Emplsoying a Hydrogel-Based Touch Pad. Sci Rep. 2019;9:10102.

307. Chien M-N, Fan S-H, Huang C-H, Wu C-C, Huang J-T. Continuous Lactate Monitoring System Based on Percutaneous Microneedle Array. Sensors (Basel). 2022;22:1468.

308. Tehrani F, Teymourian H, Wuerstle B, Kavner J, Patel R, Furmidge A, et al. An integrated wearable microneedle array for the continuous monitoring of multiple biomarkers in interstitial fluid. Nat Biomed Eng. 2022;:1–11.

309. Dror N, Weidling J, White S, Ortenzio F, Shreim S, Keating MT, et al. Clinical evaluation of a novel subcutaneous lactate monitor. J Clin Monit Comput. 2022;36:537–43.

310. Cass AEG, Sharma S. Microneedle Enzyme Sensor Arrays for Continuous In Vivo Monitoring. Methods Enzymol. 2017;589:413–27.

311. Ming DK, Jangam S, Gowers SA, Wilson R, Freeman DM, Boutelle MG, et al. Real-time Continuous Measurement of Lactate through a Minimally-invasive Microneedle Biosensor: a Phase I Clinical Study. 2021.

312. Enfield J, O'Connell M-L, Lawlor K, Jonathan E, O'Mahony C, Leahy M. In-vivo dynamic characterization of microneedle skin penetration using optical coherence tomography. J Biomed Opt. 2010;15:046001.

313. Sharma S, Huang Z, Rogers M, Boutelle M, Cass AEG. Evaluation of a minimally invasive glucose biosensor for continuous tissue monitoring. Analytical and bioanalytical chemistry. 2016;408:8427–35.

314. Rawson TM, Gowers SAN, Freeman DME, Wilson RC, Sharma S, Gilchrist M, et al. Microneedle biosensors for real-time, minimally invasive drug monitoring of phenoxymethylpenicillin: a first-in-human evaluation in healthy volunteers. The Lancet Digital Health. 2019;1:e335–43.

315. Rawson TM, Ming D, Gowers SA, Freeman DM, Herrero P, Georgiou P, et al. Public acceptability of computer-controlled antibiotic management: An exploration of automated dosing and opportunities for implementation. Journal of Infection. 2019;78:75–86.

316. Fogh-Andersen N, Altura BM, Altura BT, Siggaard-Andersen O. Composition of interstitial fluid. Clinical chemistry. 1995;41:1522–5.

317. Ellmerer M, Haluzik M, Blaha J, Kremen J, Svacina S, Plasnik A, et al. Clinical Evaluation of Subcutaneous Lactate Measurement in Patients after Major Cardiac Surgery. International Journal of Endocrinology. 2009;2009:e390975.

318. Kopterides P, Nikitas N, Vassiliadi D, Orfanos SE, Theodorakopoulou M, Ilias I, et al. Microdialysis-assessed interstitium alterations during sepsis: Relationship to stage, infection, and pathogen. Intensive Care Medicine. 2011;37:1756–64.

319. MacLean DA, Bangsbo J, Saltin B. Muscle interstitial glucose and lactate levels during dynamic exercise in humans determined by microdialysis. Journal of Applied Physiology. 1999;87:1483–90.

320. Kopterides P, Theodorakopoulou M, Ilias I, Nikitas N, Frantzeskaki F, Argyro Vassiliadi D, et al. Interrelationship between blood and tissue lactate in a general intensive care unit: A subcutaneous adipose tissue microdialysis study on 162 critically ill patients. Journal of Critical Care. 2012;27:9–742.

321. Ilias I, Apollonatou S, Vassiliadi D-A, Nikitas N, Theodorakopoulou M, Diamantakis A, et al. Adipose Tissue Lactate Clearance but Not Blood Lactate Clearance Is Associated with Clinical Outcome in Sepsis or Septic Shock during the Post-Resuscitation Period. Metabolites. 2018;8:28.

322. Beck RW, Bergenstal RM, Laffel LM, Pickup JC. Advances in technology for management of type 1 diabetes. Lancet. 2019;394:1265–73.

323. Bollella P, Sharma S, Cass AEG, Antiochia R. Microneedle-based biosensor for minimally-invasive lactate detection. Biosens Bioelectron. 2019;123:152–9.

324. Acciaroli G, Vettoretti M, Facchinetti A, Sparacino G. Calibration of Minimally Invasive Continuous Glucose Monitoring Sensors: State-of-The-Art and Current Perspectives. Biosensors (Basel). 2018;8:24.

325. Gowers SAN, Curto VF, Seneci CA, Wang C, Anastasova S, Vadgama P, et al. 3D Printed Microfluidic Device with Integrated Biosensors for Online Analysis of Subcutaneous Human Microdialysate. Anal Chem. 2015;87:7763–70.

326. Schierenbeck F, Nijsten MWN, Franco-Cereceda A, Liska J. Introducing intravascular microdialysis for continuous lactate monitoring in patients undergoing cardiac surgery: a prospective observational study. Critical Care. 2014;18:R56.

327. Hanson J, Lee SJ, Hossain MA, Anstey NM, Charunwatthana P, Maude RJ, et al. Microvascular obstruction and endothelial activation are independently associated with the clinical manifestations of severe falciparum malaria in adults: an observational study. BMC Medicine. 2015;13:122.

328. Sasso SV, Pierce RJ, Walla Robert, Yacynych AM. Electropolymerized 1,2-diaminobenzene as a means to prevent interferences and fouling and to stabilize immobilized enzyme in electrochemical biosensors. Anal Chem. 1990;62:1111–7.

329. Romero MR, Ahumada F, Garay F, Baruzzi AM. Amperometric Biosensor for Direct Blood Lactate Detection. Anal Chem. 2010;82:5568–72.

330. Freeman D, Ming D, Wilson R, Herzog P, Schulz C, Felice A, et al. Continuous Measurement of Lactate Concentration in Human Subjects through Direct Electron Transfer from Enzyme to Microneedle Electrode. ACS Sens. 2023.

331. Sharma S, El-Laboudi A, Reddy M, Jugnee N, Sivasubramaniyam S, El Sharkawy M, et al. A pilot study in humans of microneedle sensor arrays for continuous glucose monitoring. Anal Methods. 2018;10:2088–95.

332. Beardsall K, Thomson L, Guy C, Iglesias-Platas I, Weissenbruch MM van, Bond S, et al. Real-time continuous glucose monitoring in preterm infants (REACT): an international, open-label, randomised controlled trial. The Lancet Child & Adolescent Health. 2021;5:265–73.

333. Brophy K, Davies S, Olenik S, Cotur Y, Ming D, Van Zalk N, et al. The future of wearable technologies. Report. 2021.

334. Hood L, Balling R, Auffray C. Revolutionizing medicine in the 21st century through systems approaches. Biotechnol J. 2012;7:992–1001.

335. White House Precision Medicine Initiative. The White House. https://obamawhitehouse.archives.gov/node/333101. Accessed 20 Jan 2023.

336. Academy of Medical Science. Stratified, personalised or P4 medicine: a new direction for placing the patient at the centre of healthcare and health education. 2015.

337. Ladner JT, Grubaugh ND, Pybus OG, Andersen KG. Precision epidemiology for infectious disease control. Nat Med. 2019;25:206–11.

338. Flores M, Glusman G, Brogaard K, Price ND, Hood L. P4 medicine: how systems medicine will transform the healthcare sector and society. Per Med. 2013;10:565–76.

339. Hulsen T. Sharing Is Caring—Data Sharing Initiatives in Healthcare. Int J Environ Res Public Health. 2020;17:3046.

340. Department of Health, Data Sharing and Cyber Security Team. Your Data: Better Security, Better Choice, Better Care. 2017.

341. Rowhani-Farid A, Allen M, Barnett AG. What incentives increase data sharing in health and medical research? A systematic review. Research Integrity and Peer Review. 2017;2:4.

342. Han T, Nebelung S, Haarburger C, Horst N, Reinartz S, Merhof D, et al. Breaking medical data sharing boundaries by using synthesized radiographs. Science Advances. 2020;6:eabb7973.

343. Rieke N, Hancox J, Li W, Milletarì F, Roth HR, Albarqouni S, et al. The future of digital health with federated learning. npj Digit Med. 2020;3:1–7.

344. Netherlands AI Coalition. Responsible data sharing in AI. 2020.

345. Population C on, Education D of B and SS and, The National Academies of Sciences E. Exploring the Ethical Imperative for Data Sharing. National Academies Press (US); 2015.

346. Lynam AL, Dennis JM, Owen KR, Oram RA, Jones AG, Shields BM, et al. Logistic regression has similar performance to optimised machine learning algorithms in a clinical setting: application to the discrimination between type 1 and type 2 diabetes in young adults. Diagnostic and Prognostic Research. 2020;4:6.

347. Rawson TM, Ming D, Gowers SA, Freeman DM, Herrero P, Georgiou P, et al. Public acceptability of computer-controlled antibiotic management: An exploration of automated dosing and opportunities for implementation. Journal of Infection. 2018. https://doi.org/10.1016/j.jinf.2018.08.005.

348. Durán JM, Jongsma KR. Who is afraid of black box algorithms? On the epistemological and ethical basis of trust in medical AI. Journal of Medical Ethics. 2021;47:329–35.

349. Vokinger KN, Feuerriegel S, Kesselheim AS. Mitigating bias in machine learning for medicine. Commun Med. 2021;1:1–3.

350. Andreotta AJ, Kirkham N, Rizzi M. AI, big data, and the future of consent. AI & Soc. 2022;37:1715–28.

351. Lipton ZC. The Mythos of Model Interpretability. arXiv:160603490 [cs, stat]. 2017.

352. Medical Device Coordination Group European Union. Guidance on Qualification and Classification of Software in Regulation (EU) 2017/745 – MDR and Regulation (EU) 2017/746 – IVDR.

353. Beane A, De Silva AP, De Silva N, Sujeewa JA, Rathnayake RMD, Sigera PC, et al. Evaluation of the feasibility and performance of early warning scores to identify patients at risk of adverse outcomes in a low-middle income country setting. BMJ Open. 2018;8:e019387.

354. Berner ES, editor. Clinical Decision Support Systems. Cham: Springer International Publishing; 2016.

355. Rawson TM, Hernandez B, Moore LSP, Herrero P, Charani E, Ming D, et al. A Real-world Evaluation of a Case-based Reasoning Algorithm to Support Antimicrobial Prescribing Decisions in Acute Care. Clin Infect Dis. https://doi.org/10.1093/cid/ciaa383.

356. Rawson TM, Hernandez B, Wilson RC, Ming D, Herrero P, Ranganathan N, et al. Supervised machine learning to support the diagnosis of bacterial infection in the context of COVID-19. JAC-Antimicrobial Resistance. 2021;3.

357. Georgiou P, Moore L, Herrero P, Charani E, Rawson T, Hernandez B, et al. Enhanced, personalized and integrated care for infection management at point of care.

358. European Comission. High level expert group on artificial intelligence - Trustworthy AI assessment List. 2018.

359. Islam R, Weir CR, Jones M, Del Fiol G, Samore MH. Understanding complex clinical reasoning in infectious diseases for improving clinical decision support design. BMC Medical Informatics and Decision Making. 2015;15:101.

360. Rawson TM, Charani E, Moore LSP, Hernandez B, Castro-Sánchez E, Herrero P, et al. Mapping the decision pathways of acute infection management in secondary care among UK medical physicians: a qualitative study. BMC Medicine. 2016;14:208.

361. Hinton GE. Connectionist learning procedures. Artificial Intelligence. 1989;40:185–234.

362. Platt JC. Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. In: Advances in Large Margin Classifiers. MIT Press; 1999. p. 61–74.

363. Niculescu-Mizil A, Caruana R. Predicting good probabilities with supervised learning. In: Proceedings of the 22nd international conference on Machine learning - ICML '05. Bonn, Germany: ACM Press; 2005. p. 625–32.

364. A Recipe for Training Neural Networks. http://karpathy.github.io/2019/04/25/recipe/#2-set-up-the-end-to-end-trainingevaluation-skeleton--get-dumb-baselines. Accessed 17 Jan 2023.

I Thesis appendix

"Applied machine learning and the role of dynamic time-series data in optimising clinical management of acute febrile illnesses"

Submitted to:

Imperial College London for the degree of Doctor of Philosophy

Damien Keng Yen Ming

The appendix contains study protocols, ethics committee approvals and other correspondence to relevant bodies related to research carried out for the purposes of the thesis.

a List of peer-reviewed publications arising from the thesis

Chapter 4

Ming DK, Hernandez B, Sangkaew S, Vuong NL, Lam PK, Nguyet NM, et al. Applied machine learning for the risk-stratification and clinical decision support of hospitalised patients with dengue in Vietnam. PLOS Digital Health. 2022;1:e0000005.

Sangkaew S, **Ming D**, Boonyasiri A, Honeyford K, Kalayanarooj S, Yacoub S, et al. Risk predictors of progression to severe disease during the febrile phase of dengue: a systematic review and meta-analysis. The Lancet Infectious Diseases. 2021;21:1014–26.

Chapter 5

Ming DK, Tuan NM, Hernandez B, Sangkaew S, Vuong NL, Chanh HQ, et al. The Diagnosis of Dengue in Patients Presenting With Acute Febrile Illness Using Supervised Machine Learning and Impact of Seasonality. Front Digit Health. 2022;4:849641.

Chapter 6

Huy NQ, **Ming DK,** An LP, Chanh HQ, Tam DTH, Truong NT, Huy VX, Hernandez B, Van Nuil JI, Paton C, Georgiou P, Nguyen NM, Holmes A, Tho PV, Yacoub S Mapping patient pathways and understanding clinical decision-making in dengue management to inform the development of digital health tools BMC informatics and decision making 2023; 23:24

Chapter 7

Ming DK, Jangam S, Gowers SAN, Wilson R, Freeman DME, Boutelle MG, et al. Real-time continuous measurement of lactate through a minimally invasive microneedle patch: a phase I clinical study. BMJ Innov. 2022;:bmjinnov-2021-000864.

Zhang S, Chen Y-C, Riezk A, **Ming D**, Tsvik L, Sützl L, et al. Rapid Measurement of Lactate in the Exhaled Breath Condensate: Biosensor Optimization and In-Human Proof of Concept. ACS Sens. 2022;7:3809–16.

Other publications and presentations related to the machine learning in acute febrile illnesses published during, and related to the thesis:

Ming DK, Myall AC, Hernandez B, Weiße AY, Peach RL, Barahona M, et al. Informing antimicrobial management in the context of COVID-19: understanding the longitudinal dynamics of C-reactive protein and procalcitonin. BMC Infect Dis. 2021;21:932.

Ming DK, Sangkaew S, Chanh HQ, Nhat PTH, Yacoub S, Georgiou P, et al. Continuous physiological monitoring using wearable technology to inform individual management of infectious diseases, public health and outbreak responses. Int J Infect Dis. 2020;96:648–54.

Trieu HT, Khanh LP, **Ming DKY**, Quang CH, Phan TQ, Van VCN, et al. The compensatory reserve index predicts recurrent shock in patients with severe dengue. BMC Medicine. 2022;20:109.

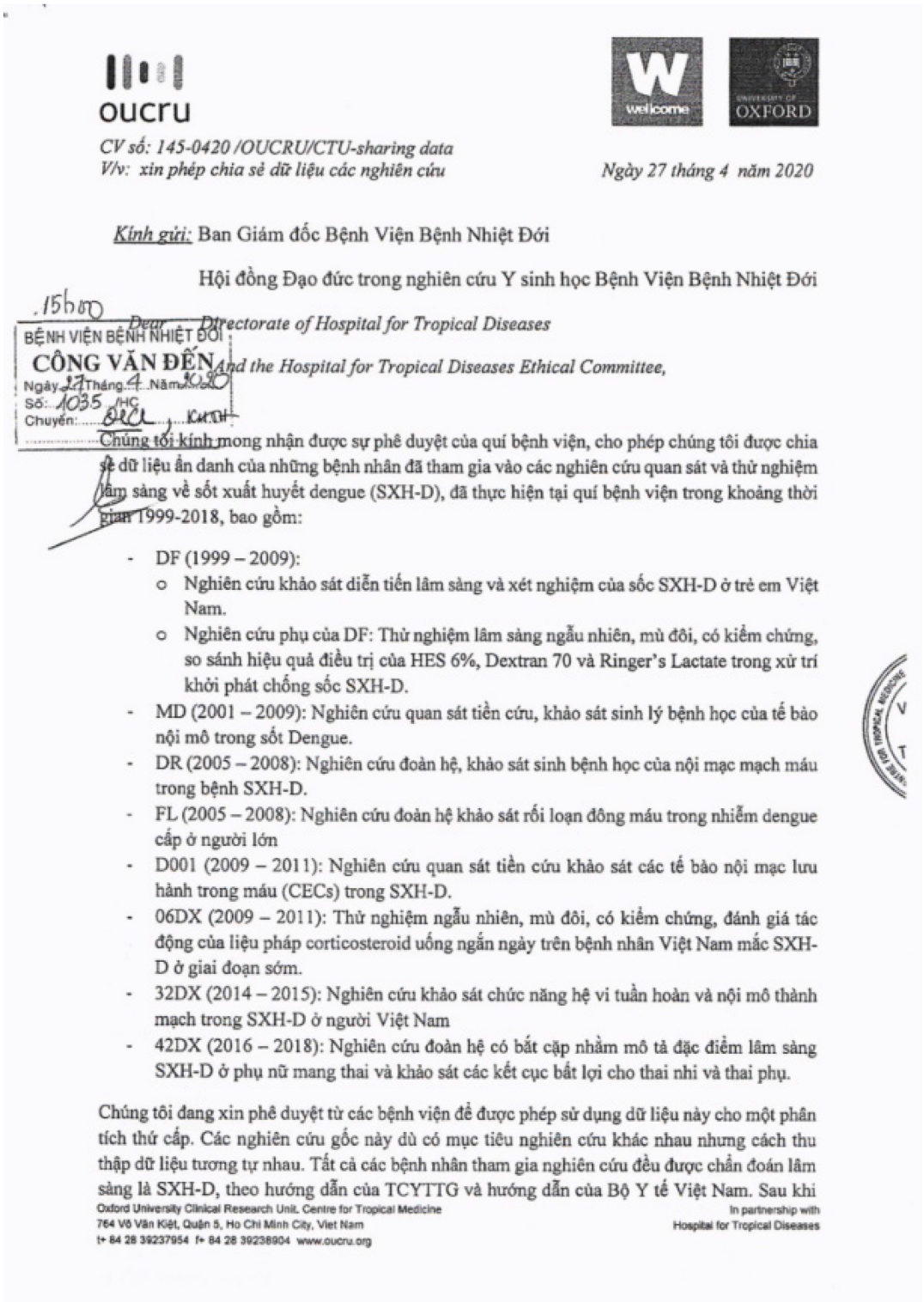
Rawson TM, **Ming D**, Ahmad R, Moore LSP, Holmes AH. Antimicrobial use, drug-resistant infections and COVID-19. Nat Rev Microbiol. 2020.

Rawson TM, Hernandez B, Wilson RC, **Ming D**, Herrero P, Ranganathan N, et al. Supervised machine learning to support the diagnosis of bacterial infection in the context of COVID-19. JAC-Antimicrobial Resistance. 2021;3:dlab002.

Karolcik S, Manginas V, Chanh HQ, Daniels J, Giang NT, Huyen VNT, Tu Van HM, Quoc KPN, Hernandez B, **Ming DK**, Holmes AH, Thwaites L, Yacoub S, Georgiou P, A prospective clinical study on the use of a non-invasive wearable device and neural network models for patients with dengue, oral presentation at International Society of Infectious Diseases congress 2022, Malaysia

Hernandez Perez B, Stiff O, **Ming D,** Ho Quang C, Vuong Nguyen L, Tuan Nguyen M, Chau Nguyen VV, Nguyet Nguyen M, Huy Nguyen Q, Lam Phung K, Tam Dong Thi H, Trung Dinh T, Trieu Huynh T, Wills B, Cameron Paul S, Holmes A, Yacoub S, Georgiou P Learning meaningful latent space representations for patient risk stratification: Model development and validation for dengue and other acute febrile illness *Frontiers in Digital Health*, ISSN: 2673-253X

Data request correspondence and committee approval documentation for dengue datasets



oucru

CV số: 145-0420 /OUCRU/CTU-sharing data
V/v: xin phép chia sẻ dữ liệu các nghiên cứu

Ngày 27 tháng 4 năm 2020

Kính gửi: Ban Giám đốc Bệnh Viện Bệnh Nhiệt Đới

Hội đồng Đạo đức trong nghiên cứu Y sinh học Bệnh Viện Bệnh Nhiệt Đới

15h00
BỆNH VIỆN BỆNH NHIỆT ĐỚI
CÔNG VĂN ĐẾN
Ngày 27 Tháng 4 Năm 2020
Số: 1035 /HC
Chuyển: OUCRU Ký nhận

Dear Directorate of Hospital for Tropical Diseases
And the Hospital for Tropical Diseases Ethical Committee,

Chúng tôi kính mong nhận được sự phê duyệt của quí bệnh viện, cho phép chúng tôi được chia sẻ dữ liệu ẩn danh của những bệnh nhân đã tham gia vào các nghiên cứu quan sát và thử nghiệm lâm sàng về sốt xuất huyết dengue (SXH-D), đã thực hiện tại quí bệnh viện trong khoảng thời gian 1999-2018, bao gồm:

- DF (1999 – 2009):
  - Nghiên cứu khảo sát diễn tiến lâm sàng và xét nghiệm của sốc SXH-D ở trẻ em Việt Nam.
  - Nghiên cứu phụ của DF: Thử nghiệm lâm sàng ngẫu nhiên, mù đôi, có kiểm chứng, so sánh hiệu quả điều trị của HES 6%, Dextran 70 và Ringer's Lactate trong xử trí khởi phát chống sốc SXH-D.
- MD (2001 – 2009): Nghiên cứu quan sát tiền cứu, khảo sát sinh lý bệnh học của tế bào nội mô trong sốt Dengue.
- DR (2005 – 2008): Nghiên cứu đoàn hệ, khảo sát sinh bệnh học của nội mạc mạch máu trong bệnh SXH-D.
- FL (2005 – 2008): Nghiên cứu đoàn hệ khảo sát rối loạn đông máu trong nhiễm dengue cấp ở người lớn
- D001 (2009 – 2011): Nghiên cứu quan sát tiền cứu khảo sát các tế bào nội mạc lưu hành trong máu (CECs) trong SXH-D.
- 06DX (2009 – 2011): Thử nghiệm ngẫu nhiên, mù đôi, có kiểm chứng, đánh giá tác động của liệu pháp corticosteroid uống ngắn ngày trên bệnh nhân Việt Nam mắc SXH-D ở giai đoạn sớm.
- 32DX (2014 – 2015): Nghiên cứu khảo sát chức năng hệ vi tuần hoàn và nội mô thành mạch trong SXH-D ở người Việt Nam
- 42DX (2016 – 2018): Nghiên cứu đoàn hệ có bắt cặp nhằm mô tả đặc điểm lâm sàng SXH-D ở phụ nữ mang thai và khảo sát các kết cục bất lợi cho thai nhi và thai phụ.

Chúng tôi đang xin phê duyệt từ các bệnh viện để được phép sử dụng dữ liệu này cho một phân tích thứ cấp. Các nghiên cứu gốc này dù có mục tiêu nghiên cứu khác nhau nhưng cách thu thập dữ liệu tương tự nhau. Tất cả các bệnh nhân tham gia nghiên cứu đều được chẩn đoán lâm sàng là SXH-D, theo hướng dẫn của TCYTTG và hướng dẫn của Bộ Y tế Việt Nam. Sau khi

Oxford University Clinical Research Unit, Centre for Tropical Medicine
764 Võ Văn Kiệt, Quận 5, Ho Chi Minh City, Viet Nam
t+ 84 28 39237954 f+ 84 28 39238904 www.oucru.org

In partnership with
Hospital for Tropical Diseases



tham gia vào nghiên cứu, các bệnh nhân được thăm khám và theo dõi hàng ngày trong suốt thời gian bệnh và vào ngày tái khám. Kết quả của các nghiên cứu này đã được công bố theo danh sách đính kèm. Các nghiên cứu gốc với dữ liệu được ẩn danh hoàn toàn đều được thông qua Hội đồng đạo đức các bệnh viện và Hội đồng đạo đức nghiên cứu bệnh nhiệt đới Đại học Oxford (OxTREC). Việc lấy chấp thuận của đối tượng tham gia cho phép chia sẻ dữ liệu cho nghiên cứu này sẽ không được thực hiện. Thay vào đó, chúng tôi sẽ xin phê duyệt của Ban Giám đốc bệnh viện, nơi thực hiện các nghiên cứu này, cho phép chia sẻ dữ liệu như là một hình thức thay thế cho việc lấy chấp thuận nghiên cứu.

We would be very grateful for your consideration of granting approval for sharing of anonymous patient data of observational studies and clinical trials of Dengue conducted at the Hospital for Tropical Diseases (HTD) during the 1999-2018 period, which includes:

- DF (1999-2009):
  - A prospective observational study investigating clinical and laboratory features of dengue shock syndrome in Vietnamese children with Dengue
  - Ancillary study: A double-blinded, randomised, controlled clinical trial investigating the effects HES 6%, Dextran 70 and Ringer's Lactate as an initial resuscitation solution for Dengue shock.
- MD (2001 – 2009): A prospective observational study investigating pathophysiology of endothelium in Dengue fever.
- DR (2005 – 2008): A cohort study investigating pathophysiology of vascular endothelial cells in dengue infection
- FL (2006 – 2008): A prospective observational study describing coagulopathy features in adults with acute dengue infection.
- D001 (2009 – 2011): A prospective observational study investigate about circulating endothelial cells in dengue
- 06DX (2009 – 2011): A double-blinded randomized controlled trial investigating effects of short-course oral corticosteroid therapy in early dengue infection in Vietnamese patients.
- 32DX (2014 – 2015): An observation study of microvascular and endothelial function in Vietnamese patients with dengue infection
- 42DX (2016 – 2018): A matched cohort study investigating clinical features of dengue infection in pregnancy and adverse outcomes for the mother and newborn.

We are in the process of asking the hospitals that took part in these studies for their approvals of secondary use of these data. Albeit these studies had different study aims, data collection was done similarly. All patients enrolled to studies were diagnosed Dengue clinically in accordance to the WHO guideline and the guideline of Vietnamese Ministry of Health. After participating in a study, patients were examined and monitored daily throughout the course of illness and at the follow-up. The results of these studies were reported in the list of publication attached. All the studies that will be included in pooled anonymous dataset have had full OxTREC and local hospital ethical approvals. No further individual patient consent will be

performed. Instead, the project will seek the proxy consents from IRBs, where studies were conducted, in order to have these data shared.

Các bộ dữ liệu được chia sẻ sẽ được sử dụng để phát triển một hệ thống hỗ trợ quyết định lâm sàng nhằm phát hiện sớm sốc SXH-D ở bệnh nhân người lớn và trẻ em nhập nội trú. Dự án này được thực hiện với sự hợp tác giữa trường Imperial College, Luân Đôn và Đơn vị nghiên cứu lâm sàng của Đại học Oxford, Việt nam. Trong dự án này, chúng tôi muốn phát triển một hệ thống hỗ trợ quyết định lâm sàng, dựa trên phân loại độ nặng và kết quả lâm sàng của những ca bệnh tương tự trước đây, bằng cách sử dụng các kỹ thuật phân tích tự động – đào tạo cho máy móc (machine learning), nhằm cung cấp những dự đoán trong tương lai để hỗ trợ bác sĩ lâm sàng ra quyết định và cải thiện hiệu quả điều trị. Chúng tôi hy vọng các mô hình đào tạo máy nâng cao sẽ đưa ra được những tiên lượng tốt hơn. Hệ thống được thiết kế để chạy trên điện thoại thông minh và máy tính bảng, và định hướng chuyển giao cho những cơ sở y tế bên ngoài Bệnh viện Bệnh Nhiệt đới, những nơi không có bác sĩ chuyên khoa nhiễm. Tất cả các quy trình trong dự án này đều được thực hiện tuân theo các điều khoản nêu trong hợp đồng chia sẻ dữ liệu. Xin vui lòng xem thêm thông tin trong đề cương rút gọn đính kèm.

The shared datasets will be used in developing a clinical decision support system for the recognition of dengue shock for children and adults admitted to secondary care. This project will be collaboration between Imperial College London and Oxford University Clinical Research Unit in Vietnam. In this project, we aim to develop a clinical decision support system with the classification of previous, similar cases of dengue and their outcomes through machine learning methods in order to provide information in the future for clinicians to make informed decisions and improve treatment. The models with the advance of machine learning are expected to gain better predictive performance. The system will be designed to run on basic smartphones and tablets, and the technology transferable to settings outside HTD, where specialist infectious diseases services are not available. All works will be conducted in accordance with the sharing agreement. Please see the brief protocol of the project for more details.

Trân trọng cảm ơn và rất mong nhận được sự chấp thuận của Ban Giám đốc bệnh viện.

Đơn Vị Nghiên Cứu Lâm Sàng Đại Học Oxford

VĂN PHÒNG ĐẠI DIỆN TẠI VIỆT NAM

Evelyne Kestelyn

Trưởng Phòng Thử Nghiệm Lâm sàng, OUCRU-VN

Đại diện nghiên cứu viên

TS. BS. Sophie Yacoub

Trưởng Nhóm Nghiên cứu SXH-D, OUCRU-VN

3

[redacted]

Đại diện nghiên cứu viên chính của các nghiên cứu

- Nghiên cứu DF, MD, DR, FL, D001, 06DX, 42DX:

Prof. Bridget Wills

- Nghiên cứu 32DX:

TS. BS. Sophie Yacoub

Ý KIẾN CỦA BỆNH VIỆN

.....

.....

.....

.....

.....

.....

.....

Tp. HCM, ngày 14 tháng 5 năm 2020

Thủ trưởng cơ quan

(Họ, tên, chữ ký và đóng dấu)

SỞ Y TẾ THÀNH PHỐ HỒ CHÍ MINH
BỆNH VIỆN BỆNH NHIỆT ĐỚI

TS.BS. NGUYỄN VĂN VĨNH CHÂU

4

oucru

wellcome

UNIVERSITY OF OXFORD

DANH SÁCH CÁC NGHIÊN CỨU ĐÃ ĐĂNG BÁO

Wills BA, Nguyen MD, Ha TL, Dong TH, Tran TN, Le TT, et al. Comparison of three fluid solutions for resuscitation in dengue shock syndrome. N Engl J Med. 2005;353(9):877-89.

Dinh The T, Le Thi Thu T, Nguyen Minh D, Tran Van N, Tran Tinh H, Nguyen Van Vinh C, et al. Clinical Features of Dengue in a Large Vietnamese Cohort: Intrinsically Lower Platelet Counts and Greater Risk for Bleeding in Adults than Children. PLOS Neglected Tropical Diseases. 2012;6(6):e1679.

Lam PK, Tam DTH, Diet TV, Tam CT, Tien NTH, Kieu NTT, et al. Clinical Characteristics of Dengue Shock Syndrome in Vietnamese Children: A 10-Year Prospective Study in a Single Hospital. Clinical Infectious Diseases. 2013;57(11):1577-86.

Lam PK, Ngoc TV, Thu Thuy TT, Hong Van NT, Nhu Thuy TT, Hoai Tam DT, et al. The value of daily platelet counts for predicting dengue shock syndrome: Results from a prospective observational study of 2301 Vietnamese children with dengue. PLOS Neglected Tropical Diseases. 2017;11(4):e0005498.

Tam DTH, Ngoc TV, Tien NTH, Kieu NTT, Thuy TTT, Thanh LTC, et al. Effects of Short-Course Oral Corticosteroid Therapy in Early Dengue Infection in Vietnamese Patients: A Randomized, Placebo-Controlled Trial. Clinical Infectious Diseases. 2012;55(9):1216-24.

Yacoub S, Lam PK, Vu LH, Le TL, Ha NT, Toan TT, Van NT, Quyen NT, Le Duyen HT, Van Kinh N, Fox A. Association of microvascular function and endothelial biomarkers with clinical outcome in dengue: an observational study. The Journal of infectious diseases. 2016 Sep 1;214(5):697-706.

CLINICAL RESEARCH UNIT

Oxford University Clinical Research Unit. Centre for Tropical Medicine
764 Võ Văn Kiệt, Quận 5, Ho Chi Minh City, Viet Nam
t+ 84 28 39237954 f+ 84 28 39238904 www.oucru.org

In partnership with
Hospital for Tropical Diseases

oucru

CV số: 146-0420 /OUCRU/CTU-sharing data
V/v: xin phép chia sẻ dữ liệu các nghiên cứu *Ngày 27 tháng 4 năm 2020*

Kính gửi: Ban Giám đốc Bệnh Viện Bệnh Nhiệt Đới Trung Ương

Hội đồng Đạo đức trong nghiên cứu Y sinh học Bệnh Viện Bệnh Nhiệt Đới Trung Ương

Dear Directorate of National Hospital for Tropical Diseases

And the National Hospital for Tropical Diseases Ethical Committee,

Chúng tôi kính mong nhận được sự phê duyệt của quí bệnh viện, cho phép chúng tôi được chia sẻ dữ liệu ẩn danh của những bệnh nhân đã tham gia vào nghiên cứu quan sát "32DX - Nghiên cứu khảo sát chức năng hệ vi tuần hoàn và nội mô thành mạch trong sốt xuất huyết Dengue (SXH-D) ở người Việt Nam", đã thực hiện tại quí bệnh viện trong khoảng thời gian 2014-2015.

Chúng tôi đang xin phê duyệt từ các bệnh viện để được phép sử dụng dữ liệu này cho một phân tích thứ cấp. Các nghiên cứu gốc này dù có mục tiêu nghiên cứu khác nhau nhưng cách thu thập dữ liệu tương tự nhau. Tất cả các bệnh nhân tham gia nghiên cứu đều được chẩn đoán lâm sàng là SXH-D, theo hướng dẫn của TCYTTG và hướng dẫn của Bộ Y tế Việt Nam. Sau khi tham gia vào nghiên cứu, các bệnh nhân được thăm khám và theo dõi hàng ngày trong suốt thời gian bệnh và vào ngày tái khám. Trong quy trình chấp thuận, nghiên cứu gốc 32DX đã lấy được sự đồng thuận của đối tượng tham gia cho phép chia sẻ dữ liệu cho nghiên cứu trong tương lai (xin xem các mẫu phiếu chấp thuận của nghiên cứu đính kèm). Kết quả nghiên cứu gốc đã được công bố theo danh sách đính kèm.

We would be very grateful for your consideration of granting approval for sharing of anonymous patient data of an observational study 'An observation study of microvascular and endothelial function in Vietnamese patients with dengue infection' conducted at the National Hospital for Tropical Diseases (NHTD) during the 2014 – 2015 period.

We are in the process of asking the hospitals that took part in these studies for their approvals of secondary use of these data. Albeit these studies had different study aims, data collection was done similarly. All patients enrolled to studies were diagnosed Dengue clinically in accordance to the WHO guideline and the guideline of Vietnamese Ministry of Health. After participating in a study, patients were examined and monitored daily throughout the course of illness and at the follow-up. The results of these studies were reported in the list of publication attached. As part of the consent process for these studies, consent for future data sharing was given for these studies. The results of these studies were reported in the list of publication attached.

Oxford University Clinical Research Unit. Centre for Tropical Medicine
764 Võ Văn Kiệt, Quận 5, Ho Chi Minh City, Viet Nam
t+ 84 28 39237954 f+ 84 28 39238904 www.oucru.org

In partnership with
Hospital for Tropical Diseases

Các bộ dữ liệu được chia sẻ sẽ được sử dụng để phát triển một hệ thống hỗ trợ quyết định lâm sàng nhằm phát hiện sớm sốc SXH-D ở bệnh nhân người lớn và trẻ em nhập nội trú. Dự án này được thực hiện với sự hợp tác giữa trường Imperial College, Luân Đôn và Đơn vị nghiên cứu lâm sàng của Đại học Oxford, Việt nam. Trong dự án này, chúng tôi muốn phát triển một hệ thống hỗ trợ quyết định lâm sàng, dựa trên phân loại độ nặng và kết quả lâm sàng của những ca bệnh tương tự trước đây, bằng cách sử dụng các kỹ thuật phân tích tự động – đào tạo cho máy móc (machine learning), nhằm cung cấp những dự đoán trong tương lai để hỗ trợ bác sĩ lâm sàng ra quyết định và cải thiện hiệu quả điều trị. Chúng tôi hy vọng các mô hình đào tạo máy nâng cao sẽ đưa ra được những tiên lượng tốt hơn. Hệ thống được thiết kế để chạy trên điện thoại thông minh và máy tính bảng, và định hướng chuyển giao cho những cơ sở y tế bên ngoài Bệnh viện Bệnh Nhiệt đới, những nơi không có bác sĩ chuyên khoa nhiễm. Tất cả các quy trình trong dự án này đều được thực hiện tuân theo các điều khoản nêu trong hợp đồng chia sẻ dữ liệu. Xin vui lòng xem thêm thông tin trong đề cương rút gọn đính kèm.

The shared datasets will be used in developing a clinical decision support system for the recognition of dengue shock for children and adults admitted to secondary care. This project will be collaboration between Imperial College London and Oxford University Clinical Research Unit in Vietnam. In this project, we aim to develop a clinical decision support system with the classification of previous, similar cases of dengue and their outcomes through machine learning methods in order to provide information in the future for clinicians to make informed decisions and improve treatment. The models with the advance of machine learning are expected to gain better predictive performance. The system will be designed to run on basic smartphones and tablets, and the technology transferable to settings outside HTD, where specialist infectious diseases services are not available. All works will be conducted in accordance with the sharing agreement. Please see the brief protocol of the project for more details.

Trân trọng cảm ơn và rất mong nhận được sự chấp thuận của Ban Giám đốc bệnh viện.

Đơn Vị Nghiên Cứu Lâm Sàng Đại Học Oxford | **Đại diện nghiên cứu viên**

VĂN PHÒNG ĐẠI DIỆN TẠI VIỆT NAM

Evelyne Kestelyn | TS. BS. Sophie Yacoub

Trưởng Phòng Thử Nghiệm Lâm sàng, OUCRU-VN | Trưởng Nhóm Nghiên cứu SXH-D, OUCRU-VN

2

Đại diện nghiên cứu viên chính của nghiên cứu 32DX

TS. BS. Sophie Yacoub

Ý KIẾN CỦA BỆNH VIỆN

Bệnh viện Bệnh Nhiệt đới Trung ương và Hội đồng đạo đức Bệnh viện đồng ý cho nghiên cứu 32DX chia sẻ dữ liệu ẩn danh của nghiên cứu để phân tích thứ cấp.

Tp. HCM, ngày 2 tháng 6 năm 2020

Thủ trưởng cơ quan

(Họ, tên, chữ ký và đóng dấu)

BỆNH VIỆN BỆNH NHIỆT ĐỚI TRUNG ƯƠNG

GIÁM ĐỐC

Phạm Ngọc Thạch

3

oucru

wellcome

UNIVERSITY OF OXFORD

24 May 2019

Dear Hospital for Tropical Diseases Ethical Committee,

We would be very grateful for your consideration of granting ethical approval for sharing of anonymous patient data from a study conducted at the Hospital for Tropical Diseases, (as well as 6 other hospitals; Children's Hospital No. 1, Children's Hospital No. 2, Tien Giang Provincial Hospital, Dong Nai Children's Hospital, Binh Duong Provincial Hospital, and Long An Provincial Hospital). We are in the process of asking all the hospital ethical committees that took part in the study, for their approval of secondary use of this data. The study recruited 7,544 children presenting at outpatient departments within 72 hours of fever. Of those, 2060 (27.3%) had laboratory-confirmed dengue and 117 (1.5%) developed severe dengue. As part of the consent process for these studies, consent for future data sharing *was* given for this study. The results of the study were reported in a publication by Nguyen et al. in 2017 (1).

The shared data set will be used in developing and evaluating prognostic models for dengue deterioration in the febrile phase. This project will be a collaboration between Imperial College London and Oxford University Clinical Research Unit in Vietnam. In this project, we aim to develop prognostic models using machine learning techniques incorporating both baseline and dynamic clinical parameters. The models with the advance of machine learning are expected to gain better predictive performance. Final models will be provided to clinicians in a primary care setting to improve triage patients and dengue case management.

Yours sincerely,

Dr Sophie Yacoub (Head of Dengue Group, OUCRU-VN)

Oxford University Clinical Research Unit. Centre for Tropical Medicine
764 Võ Văn Kiệt, Quận 5, Ho Chi Minh City, Viet Nam
t+ 84 28 39237954 f+ 84 28 39238904 www.oucru.org

In partnership with
Hospital for Tropical Diseases

b Ethics approval for use of BSI dataset from Imperial College Healthcare NHS Trust

Enhanced, Personalized, and Integrated Care for Infection Management

Research type

Research Study

Full title

Enhanced, Personalized and Integrated Care for Infection Management at the Point-Of-Care (EPIC IMPOC)

IRAS ID

204949

Contact name

Alison H Holmes

Contact email

alison.holmes@imperial.ac.uk

Duration of Study in the UK

1 years, 6 months, 10 days

Shape the future of the HRA website. Take part in testing our new website or feedback on our current site.

time as bacteria develop resistance to them. Antimicrobial usage itself can lead to development and spread of antimicrobial resistance. Antimicrobial resistance is now a major threat to patient safety. To conserve the effectiveness of antimicrobials we need to develop ways to use them more sensibly healthcare professionals who diagnose and treat infections must be able to access antimicrobial guidelines and test results at the patient bedside. This needs to be provided rapidly and with support to make sure that the decisions on prescribing antimicrobials are the best that can be made.

Prototype software to achieve this has been developed through collaboration between healthcare professionals and biomedical engineers. This prototype software (run on a mobile device) retrieves patient results from various laboratory and clinical databases (securely within the Trust firewall) and displays this to the clinician making the prescribing decision. Furthermore a machine learning algorithm is applied to the data, and similar anonymised historical cases (and the antimicrobials prescribed and the clinical outcomes) are also displayed to the clinician to further inform their decision making. The prototype has been designed for use in intensive care, where the risk of infection is high, but through the research project detailed here, the software will be developed and tested as a proof-of-concept across other areas of hospital patient care. Furthermore there is a key need to engage patients with how decisions are made around antimicrobial prescribing. We propose to adapt the prototype to meet these needs. This system should improve patient safety and help preserve the effectiveness of existing antimicrobials.

REC name

London - Chelsea Research Ethics Committee

REC reference

17/LO/0047

Date of REC Opinion

20 Feb 2017

REC opinion

Further Information Favourable Opinion

c Study protocol for MIRACLE (Minimally Invasive Realtime Assessment of Continuous Lactate in Exercise)

Imperial College London

Minimally Invasive Realtime Assessment of Continuous Lactate in Exercise (MIRACLE)

Study Management Group

Principal Investigator:

Professor Alison Holmes

Co-investigators:

Dr Danny O'Hare
Professor Anthony Cass
Dr Damien Ming
Dr David Freeman

Sponsor

Imperial College London is the main research Sponsor for this study. For further information regarding the sponsorship conditions, please contact the Head of Research Governance and Integrity at r.nicholson@imperial.ac.uk .

Funders

Department of Health and Social Care

This protocol describes the MIRACLE study and provides information about procedures for entering participants. Every care was taken in its drafting, but corrections or amendments may be necessary. These will be circulated to investigators in the study. Problems relating to this study should be referred, in the first instance, to the Principal Investigator.

This study will adhere to the principles outlined in the UK Policy Framework for Health and Social Care Research It will be conducted in compliance with the protocol, Data Protection Act 2018 and General Data Protection Regulations (Europe) and other regulatory requirements as appropriate.

TITLE	Minimally Invasive Realtime Assessment of Continuous Lactate in Exercise
DESIGN	In-house Phase I validation study for a microneedle biosensor for continuous lactate measurement
AIMS	(i) Examine the relationship between microneedle biosensor output with reference to venous blood lactate during exercise (ii) Evaluate performance of microneedle biosensor with reference to interstitial fluid lactate from microdialysis during exercise (iii) Assess the comfort and acceptability of biosensor during exercise and rest
OUTCOME MEASURES	(i) Precision and accuracy of the lactate microneedle with reference to venous lactate as a gold standard (ii) Participant acceptability of the biosensor through an end-of-study questionnaire
POPULATION	15 healthy volunteers
DURATION OF STUDY	6 months (1/9/2020 – 1/04/2021)

1. INTRODUCTION

1.1 Background

Lactate is a compound produced as a by-product of glycolysis. The increase in lactate concentration can result from a variety of causes, such as during anaerobic metabolism in exercise. The measurement of blood lactate during exercise has been routinely used to determine individual anaerobic thresholds, and optimise training in athletes (1). However, an abnormal increase in lactate can also occur in haemodynamic shock, and this is an important physiological end-point for a number of diseases. In this situation, lactate monitoring is useful to guide clinical management. Blood lactate concentrations correlate with all-cause mortality (2), and its clearance over time an independent predictor of outcomes in conditions such as sepsis (3).

The measurement of the lactate trend requires multiple blood tests, and this can be uncomfortable. The need for repeated procedures may therefore restrict the full utility of the biomarker. Although point-of-care lactate measurements are possible, the barriers to initiating testing may also limit its role in clinical decision-making.

The interstitial fluid (ISF) is large body of extracellular fluid with a similar composition to blood plasma, with compounds freely exchanged between the compartments. The measurement of ISF is already extensively utilised in continuous glucose monitoring, but has applications in other areas of medicine such as surgery. Lactate concentrations in the ISF is closely related to serum lactate in animal models and healthy volunteers. From microdialysis studies, the increase in ISF lactate from patients with sepsis also preceded a rise in blood lactate by 4 hours, suggesting that this could be a valuable biomarker (4).

A promising avenue for ISF lactate measurement is the use of a microneedle array as a biosensor (5). This small, low-cost biosensor is applied painlessly to the skin where it is able to measure lactate in real-time through continuous contact with the ISF. There is extensive clinical safety data in the use of microneedles, and the core utility of ISF measurements well-established. Testing in clinical studies have shown good tolerability of the microneedle in patients for up to 24 hours (6). In vitro testing of the microneedle using human serum and lactate have also demonstrated excellent performance (unpublished data). From public engagement work, there is also strong support for the use of microneedles as an acceptable

alternative to traditional modes of testing (7). Although studies have been published on real-time detection of lactate through other methods, such as through intravenous catheters, there have been none to date which measure ISF lactate in real-time in humans.

1.2 Study Rationale

The aim of the study is the validation of the microneedle-based lactate biosensor for real-time continuous lactate measurement in healthy volunteers. As the biosensor is placed on the skin surface in a minimally-invasive fashion (see section 3), it has significant advantages over traditional venesection / arterial blood sampling particularly in the paediatric population, where acceptability by patient and carers is important. The use of a microneedle biosensor would result in reduced training requirements, infrastructure costs and risk of needlestick injuries.

For exercise physiology, the validation of the lactate biosensor would lead to development of a "wearable" device which can be used in conjunction with endurance training methods. This would provide information regarding individualised anaerobic thresholds to a much greater resolution compared with current methods. Using an exercise model, participants routinely generate transient lactate concentrations of up to 14 mM in their blood. Unlike in sepsis, this build-up of lactate is physiological and normalises rapidly on cessation of exercise.

II. This project will therefore provide the first-in-human validation of the microneedle platform for continuous lactate measurement in healthy volunteers engaged in exercise. The results of this study will directly inform a larger clinical trial to evaluate the clinical performance of physiological monitoring in exercise, as well as subsequently for healthcare applications such as sepsis recognition, trauma resuscitation or post-operative fluid resuscitation. The measurement system is self-contained and there is scope for near-patient use in resource-limited settings where laboratory services are not available–- for example childhood malaria. As ISF exists in equilibrium with blood, small molecules diffuse freely between compartments. There also exists scope for the use of the microneedle as a platform for the measurement of novel biomarkers, and contribute towards implementation of individualised, precision medicine.

STUDY OBJECTIVES

Primary objectives

II) Validation of the microneedle biosensor in vivo within exercising volunteers for the continuous measurement of interstitial fluid lactate: Evaluate the relationship of the microneedle biosensor output with reference to regular venous lactate during exercise.

ii) Evaluate biosensor performance with reference to interstitial fluid lactate from microdialysis during exercise.

Secondary objective

i) Comfort and acceptability of microneedle during exercise and rest.

3. STUDY DESIGN

Study outline

This is an in-house Phase I observational device validation study for a novel lactate microneedle-based biosensor to measure continuous interstitial fluid lactate.

Fifteen healthy participants will be recruited to perform aerobic exercise on an exercise bicycle for 30 minutes with the microneedle sensor, with concurrent interstitial and blood lactate measurements using standard analysis.

The duration of the study is anticipated to take 3 hours in total from consent to the end of study, and will take place in the NIHR- Imperial Clinical Research Facility located at Hammersmith Hospital, London.

Pre-study

An e-mail will be sent advertising the project to the volunteer database at Imperial College London. Posters will also be placed up within the campus to advertise the study. Volunteers expressing an interest will then be contacted directly by the research team by email or phone, and invited to the screening meeting. Written participant information leaflet and study outline will be sent to the volunteer by email or post, and any questions can be answered before the screening visit. All screening and recruitment information will be recorded on an electronic master study list which is password protected and accessible only to the research staff.

The screening meeting will take place at least 24 hours after receipt of the study material.

Sample Size Calculation

The target recruitment number of this study is 15 participants. As the novel lactate biosensor has not previously undergone clinical testing, there are no preliminary data to perform a formal sample size calculation. The sample size has derived from previous similar microneedle studies on glucose and antibiotic measurement, as well as unpublished preliminary in vitro data.

Screening and informed consent

The screening visit will take place at the NIHR- Imperial Clinical Research Facility based at Hammersmith Hospital. The screening will involve the use of a questionnaire covering basic medical history, exercise habits, and will identify any criteria for exclusion. This will be performed in person with the participant to confirm eligibility. Physical examination will be carried out as necessary by the study physician – for example to ensure suitability of sites for microneedle biosensor and catheter placement. The study will be discussed with reference to the participant information leaflet and any questions answered. Informed consent will be obtained, with completed paper consent forms and screening questionnaires securely stored at the NIHR-Imperial Clinical Research Facility, Hammersmith Hospital.

The participant will then be assigned a study number from a master list and this identifier will be used for the duration of the study. The study can take place immediately after the screening process.

Study Method

The study will take place in a research room within the NIHR-Imperial Clinical Research Facility. After screening and informed consent, the study will begin:

1) A sterilised lactate microneedle biosensor will be placed on the non-dominant forearm and connected to a potentiostat, which records data. There will be a run-in stabilisation time of approximately 60 minutes to normalise readings.

2) EMLA 5% topical anaesthetic cream will be applied to the site of microdialysis insertion.

3) A peripheral cannula will be inserted in the arm for blood lactate testing.

4) A microdialysis catheter will be inserted on the non-dominant forearm for the measurement of interstitial fluid lactate (see section below for description).

5) At 60 minutes into the study, baseline blood lactate and ISF lactate will be taken from the peripheral cannula and microdialysis catheter respectively.

6) The participant will be asked to begin exercise on an exercise bicycle starting at 90W, with increments in power every 5 minutes up to a maximum of 210W.

7) During the study, 2 ml of blood will be drawn from the peripheral cannula every 5 minutes for venous lactate determination, and will continue until the end of the rest period.

8) Microdialysis samples will also be drawn from the microdialysis catheter every 5 minutes for ISF lactate determination and will continue until the end of the study.

9) The exercise period will stop 90 minutes into the study (after 30 minutes of total exercise) and the rest period will begin for 30 minutes.

10) At the end of the rest period, all monitoring devices (microneedle, peripheral cannula, microdialysis catheter) will be removed from the participant.

11) The participant will be given a questionnaire to complete. The site of the microneedle placement on the skin will also be photographed by the study camera – this will not contain any participant identifiable information and will be consented for explicitly. The photograph will be stored electronically on servers maintained by Imperial College Healthcare NHS Trust within the NIHR-Imperial Clinical Research Facility.

12) End of study, and participant free to leave.

Laboratory sampling

It is expected that a total of 13 blood samples will be drawn from each participant during the study into oxalate vacutainer bottles. This will amount to around 2.0 ml × 13 = 26.0 ml blood taken in total. For microdialysis samples, a continuous infusion of perfusion fluid into the probe will occur and 13 sampling points will result in removal of 200 µl x 13 = 2.6 ml of the infused perfusion fluid in total. Blood and interstitial fluid will be frozen immediate at -80 oC freezer immediately after sampling at the NIHR-Imperial Clinical Research Facility. Analysis will be performed by North West London pathology laboratories at Hammersmith Hospital, Imperial College Healthcare NHS Trust. All samples will be destroyed after analysis in line with local policy after 72 hours.

Post study

There is no planned follow-up visit. However, participants will be sent a copy of their exercise and lactate profile if they wish. They will also be given information on contacting the research team if any issues arise as a result of the study.

Pseudonymised data will be downloaded and stored on secure computers hosted by Imperial College London. Blood and ISF sample analyses are described in the "Laboratory sample testing" section below.

Data collection and analysis

Participant information will be kept on case record forms in the study folder. Blood and ISF lactate results will be downloaded from the Imperial Healthcare NHS Trust pathology information system and stored in password-protected databases on secure Imperial College computers against participant number only, thus data is pseudonymised. Data from the microneedle reading will be transferred from the potentiostat and stored in the same database against participant number. The data will be subsequently analysed for concordance using Bland-Altman plots and correlation coefficients using packages written in Python. The photographs of the skin will not contain any participant identifiable information and will be stored on secure Imperial College Healthcare NHS Trust server computers.

Data management

All research data will be stored within Imperial College data storage infrastructure and will comply with the College research data management policy https://www.imperial.ac.uk/media/imperial-college/research-and-innovation/research-office/public/Imperial-College-RDM-Policy.pdf

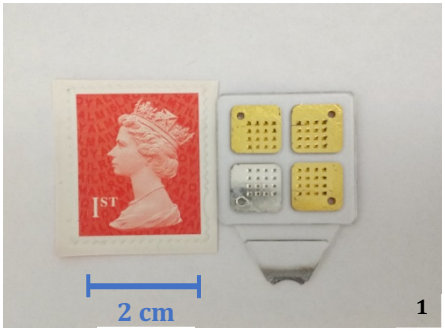
Case report forms will be stored in a locked cabinet within a keycard protected room on Imperial College estates. All collected research data is pseudonymised and will be backed up on secure, encrypted servers managed through Imperial College London. All data will be password protected and accessible by the research team only. Data which will be used for subsequent analysis will undergo full anonymisation by decoupling of the participant number.

On completion of the study, all study data including consent forms and other participant identifiable information will be stored by Imperial College London archiving facility for a minimum of 10 years.

1.6 Study Equipment and Processes

Description of the Microneedle biosensor

The microneedle biosensor is a single-use, sterilised biosensor developed at Imperial College London. The device is fabricated through injection moulding of polycarbonate and measures 2 × 2 cm. It has 1.0 mm protrusions which are applied to the skin surface, in direct contact with the interstitial fluid. As the depth of penetration is superficial 247pprox.x. 0.8 mm), the biosensor resides within the top layer within the viable epidermis and does not come into contact with nerve fibres or blood vessels. There is extensive literature to support safety and tolerability of the microneedle (8). The microneedle platform has also been used in previous studies at Imperial College (6,9) with good results.



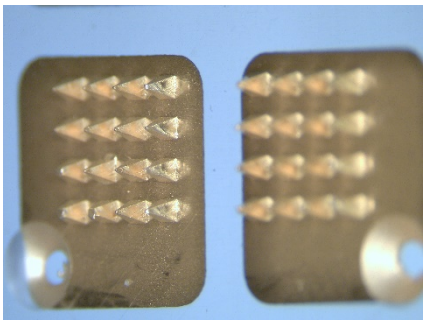


Figure 1 (left) – microneedle biosensor adjacent to a postage stamp measuring 2 cm in width showing the three gold electrodes and a silver reference electrode (bottom left).

Figure 2 (right)– microneedle biosensor under 10x magnification showing 1 mm protrusions which are positioned in the skin

The lactate microneedle is quality-checked and undergoes a process of metallization through a magnetron sputter coater which is provided for, and quality-checked by Torr Scientific Ltd (Bexhill-on-Sea, UK). The surface of the microneedle consists of a layer of lactate oxidase enzyme (Sekisui diagnostics, MA, USA) immobilised to the electrode surface with a polyethylene glycol-based hydrogel. The superficial layer of the biosensor is then coated with a Nafion membrane. All components used to fabricate the lactate microneedle biosensor are non-toxic. Safety information, along with fabrication details are provided in the microneedle appendix.

Sterilisation of the sensors has been organised through TCM associates Ltd, who have a standard protocol in place for sterilisation of the biosensors as part of previous studies using the same microneedles. This includes completion of bioburden assays and quality assurance procedures.

Following sterilisation, the needles will be returned and stored before use. In vitro assessment has been undertaken to demonstrate that these sensors are still functional within expected limits after 2 months of storage. In the case of this study, it is expected that sensors will be used within 2 weeks of fabrication. The microneedle biosensor is connected to a potentiostat device which will continuously record data. A datasheet on the microneedle platform is provided in the appendix.

Description of the Microdialysis catheter

Microdialysis will be performed using the single-use 63 Microdialysis Catheter (M Dialysis AB, Johanneshov, Sweden) according to manufacturer instructions. The device conforms to the ISO 13485:2016 standard for the design and manufacture of medical devices. Briefly, the device is inserted into the subcutaneous space with topical anaesthesia as a sterile procedure. A slow infusion of perfusion fluid is used to obtain a flow of interstitial fluid, which is then sampled and sent for analysis. The microdialysis catheter is withdrawn at the end of the study and pressure applied to the puncture site for 5 minutes.

Venepuncture

All blood samples will be drawn from a peripheral cannula inserted at the start of the study.

Exercise

Participants will be asked to cycle on a dedicated exercise bicycle at increasing power outputs for 30 minutes in total. During this period blood and ISF sampling will occur in the upper limb. The participant is free to pause, stop or adjust the intensity of the exercise at any point during the study by informing the research team. The exercise may also be stopped at the discretion of the research physician.

1.7 Precautions and Adverse Reactions

Associated with Microneedle use

The risks associated with microneedle use on intact skin are low, with the main adverse reactions being skin irritation or discomfort at the site. The materials used for fabrication are safe to use externally (see technical appendix). Placement of the microneedle is well-tolerated and skin irritation occurs only in prolonged applications (> 12 hours). From previous studies microneedles had a low pain score of 0.4 out of 10 using a pain visual analogue scale (6). Any evidence of pain or inflammation on the skin will result in removal of the microneedle and the participant observed for resolution of symptoms. The risk of subsequent infection is also likely to be low as the epidermis is not breached with the microneedle biosensor, and the microneedle is sterile on application. The participant will be asked to inform the study staff after the end of the study if any erythema develops.

Associated with microdialysis catheter insertion

The microdialysis catheter will be inserted into the subcutaneous layer of the forearm using an aseptic technique. Topical anaesthetic (EMLA Cream 5% containing lidocaine/prilocaine) will be applied to the participant's forearm for up to 30 minutes prior to insertion to ensure anaesthesia. Pain and bruising can nonetheless occur as a result of insertion, although from previous studies this process was self-limiting and tolerated well. Given the anatomical location of the microdialysis catheter it is unlikely to cause any damage to deeper structures of the arm, such as blood vessels or nerve fibres. Insertion of the catheter will be done by the study physician experienced in percutaneous procedures. The research team is to be contacted again for any redness which develops after completion of the study.

Associated with venous cannula insertion

Cannulation will be done using aseptic non-touch technique and in accordance with Imperial College Healthcare NHS Trust guidelines. Failure to cannulate and site discomfort are the main adverse effects of this procedure in which case another site on the arm will be considered.

Associated with exercise

II. Participants will be screened for their ability to perform regular moderate intensity aerobic exercise. All equipment (microneedle, microdialysis and cannula) is secured on the upper arm to reduce any effects resulting from cycling exercise in the lower limb. Blood collection and microdialysis will be able to proceed without requiring cessation of cycling, although the participant may be asked to momentarily suspend activity if a need arises. PARTICIPANT RECRUITMENT

An e-mail will be sent advertising the project to the dedicated volunteer database curated by Imperial College London, and posters placed in the College. Volunteers expressing an

interest will be contacted directly by the research team by email or phone, and invited to the screening meeting. Written participant information leaflet and study outline will be sent to the volunteer by email or post.

4.1 Pre-recruitment evaluations

Participants will be selected on the basis of fitness and ability to perform aerobic exercise e.g., cycling or running for at least 30 minutes duration regularly given the nature of the study.

4.2 Inclusion Criteria

1. Consenting adults ≥ 18 years old

2. Healthy with no previously diagnosed medical condition from a medical practice

3. Able to perform moderately intensive exercise on an exercise bicycle without difficulty for at least 30 minutes continuously, and engages in regular aerobic exercise at least twice a week

4.3 Exclusion Criteria

1. Active inflammatory skin condition such as eczema or dermatitis

2. Active soft tissue infection or infection at any site

3. Known hypersensitivity to any microneedle component or dressings

4. Presence of any implantable electronic devices such as a pacemaker or stimulators

5. Currently pregnant

4.4 Withdrawal Criteria

Participants can withdraw at any stage during the study by informing the research team verbally or by writing without having to provide a reason. Data collected up to the point of withdrawal will be included in the final analysis.

5. ADVERSE EVENTS

5.1 Definitions

Adverse Event (AE): any untoward occurrence in a participant.

Serious Adverse Event (SAE): any untoward and unexpected occurrence or effect that:

- Results in death
- Is life-threatening – refers to an event in which the subject was at risk of death at the time of the event; it does not refer to an event which hypothetically might have caused death if it were more severe
- Requires hospitalisation

II.1. • Results in persistent or significant disability or incapacity **Reporting Procedures**

All adverse events should be reported, and any questions concerning adverse event reporting should be directed to the Principal Investigator in the first instance.

II.1.2 Non serious aEs

II.1.2 All such events, whether expected or not, should be recorded. Serious aEs

An SAE form should be completed and emailed to the Principal Investigator within 24 hours.

All SAEs should be reported to the Bloomsbury Research Ethics Committee where in the opinion of the Principal Investigator, the event was:

- 'related', i.e. resulted from the administration of any of the research procedures; and
- 'unexpected', i.e. an event that is not listed in the protocol as an expected occurrence

Local investigators should report any SAEs as required by their Local Research Ethics Committee, Sponsor and/or Research & Development Office.

Contact details for reporting SAEs

Fax: 0208383394

Urgent attention Prof Alison Holmes (alison.holmes@imperial.ac.uk)

Tel: 02033132732 (Mon to Fri 09.00 – 17.00)

Sponsor office

II. Email: JRCO@imperial.ac.uk ASSESMENT AND FOLLOW UP

There is no planned follow up of participants after the end of the study. Participants will be given the contact information of the research team if any questions arise in the future.

7. REGULATORY ISSUES

7.1 Ethics approval

The Study Coordination Centre has obtained approval from the Bloomsbury Research Ethics Committee (REC) and Health Research Authority (HRA). The study must also receive confirmation of capacity and capability from each participating NHS Trust before accepting participants into the study or any research activity is carried out. The study will be conducted in accordance with the recommendations for physicians involved in research on human subjects adopted by the 1^{8t}h World Medical Assembly, Helsinki 1964 and later revisions.

7.2 Consent

Consent to enter the study must be sought from each participant only after a full explanation has been given, an information leaflet offered and time allowed for consideration. Signed participant consent should be obtained. The right of the participant to refuse to participate without giving reasons must be respected. All participants are free to withdraw at any time without having to provide a reason.

7.3 Confidentiality

The Principal Investigator will preserve the confidentiality of participants taking part in the study and fulfil transparency requirements under the General Data Protection Regulation for health and care research. Data and all appropriate documentation will be stored for a minimum of 10 years after the completion of the study, including the follow-up period.

Participants will be given unique study numbers, from which all data will be recorded against. A study number is used to replace identifiable information. The identifying unique study number, and all identifiable data (such as name, date of birth, e-mail address), will be stored in a password protected NHS computer at Hammersmith Hospital, Imperial College Healthcare NHS Trust. The signed consent form and screening questionnaire will be stored in a physical file along with the participant's research notes and kept at the NIHR-Imperial Clinical Research Facility, Centre for Translational and Experimental Medicine, Hammersmith Hospital. Only pseudonymised data will be stored in Imperial College London, in an encrypted database within the Imperial College London Firewall.

For the purposes of laboratory sample testing of blood and interstitial fluid, the participant will be allocated a patient identifiable number to facilitate analysis through the laboratory at Hammersmith Hospital, Imperial College Healthcare NHS Trust. Access to these results will be through a NHS smart-card-protected electronic records system (Cerner Corporation, Missouri USA) restricted to staff with the appropriate license to attend. Results will be downloaded to a database against the participant study number.

7.4 Indemnity

Imperial College London holds negligent harm insurance policies which apply to this study.

7.5 Sponsor

Imperial College London will act as the main Sponsor for this study. Delegated responsibilities will be assigned to the NHS trusts taking part in this study.

7.6 Funding

Department of Health and Social Care are funding this study

7.7 Audits

The study may be subject to inspection and audit by Imperial College London under their remit as sponsor and other regulatory bodies to ensure adherence to GCP and the UK Policy Framework for Health and Social Care Research.

8. STUDY MANAGEMENT

The day-to-day management of the study will be co-ordinated through the research physician (DM). The project will benefit from management by the Head of Operations within the research group. Interim safety and study data will be reviewed daily by the research physician to ensure completeness and this will be further discussed at a working group meeting held fortnightly. A steering committee consisting of the study PI, research team, external representatives and a patient representative will meet quarterly to review study progress and approve progression of the study.

9. PUBLICATION POLICY

The methods and results from the evaluation of the continuous lactate assessment through a microneedle will be presented at national and international conferences to groups involved in exercise research, as well as to infectious diseases specialists and decision makers. Peer-reviewed publications will arise from this study which will allow for a wide range of researchers to engage.

10. REFERENCES

1. Foster C, Rodriguez-Marroyo JA, de Koning JJ. Monitoring Training Loads: The Past, the Present, and the Future. Int J Sports Physiol Perform. 2017 Apr;12(Suppl 2):S22–8.

2. Zhang Z, Xu X. Lactate clearance is a useful biomarker for the prediction of all-cause mortality in critically Ill patients: A systematic review and meta-analysis. Crit Care Med. 2014;42(9):2118–2125.

3. Arnold R, Zhang Z, Patel S, Smola S, Isserman J, Jackson E. Delayed assessment of serum lactate in sepsis is associated with an increased mortality rate. Crit Care. 2014;18(Suppl 1):P174.

4. Kopterides P, Theodorakopoulou M, Ilias I, Nikitas N, Frantzeskaki F, Argyro Vassiliadi D, et al. Interrelationship between blood and tissue lactate in a general intensive care unit: A subcutaneous adipose tissue microdialysis study on 162 critically ill patients. J Crit Care. 2012;27(742):9–742.

5. Sharma S, Saeed A, Johnson C, Gadegaard N, Cass AE. Rapid, low cost prototyping of transdermal devices for personal healthcare monitoring. Sens Bio-Sens Res. 2017 Apr;13:104–8.

6. Sharma S, Huang Z, Rogers M, Boutelle M, Cass AEG. Evaluation of a minimally invasive glucose biosensor for continuous tissue monitoring. Anal Bioanal Chem. 2016 Nov;408(29):8427–8435.

7. Rawson TM, Ming D, Gowers SA, Freeman DM, Herrero P, Georgiou P, et al. Public acceptability of computer-controlled antibiotic management: An exploration of automated dosing and opportunities for implementation. J Infect [Internet]. 2018 Aug; Available from: http://www.ncbi.nlm.nih.gov/pubmed/30099085 https://linkinghub.elsevier.com/retrieve/pii/S01634453183025000

8. Donnelly RF, Raj Singh TR, Woolfson AD. Microneedle-based drug delivery systems: Microfabrication, drug delivery, and safety. Drug Deliv. 2010 May;17(4):187–207.

9. Rawson TM, Gowers SAN, Freeman DME, Wilson RC, Sharma S, Gilchrist M, et al. Microneedle biosensors for real-time, minimally invasive drug monitoring of phenoxymethylpenicillin: a first-in-human evaluation in healthy volunteers. Lancet Digit Health. 2019 Nov 1;1(7):e335–43.

Lactate Microneedle Biosensor – technical and safety summary

Design

II) The lactate microneedle biosensor measures 2 × 2 cm in size and is composed of three layers on its functional surface: A polycarbonate plastic base metallised with a thin layer of gold (Torr Scientific, Bexhill, UK) and platinum.

II) ii) A middle layer of lactate oxidase enzyme (Sekisui diagnostics, MA, USA) immobilised in a polyethylene glycol (PEG)-hydrogel. An outer membrane of Nafion (Sigma-Aldrich, MO, USA).

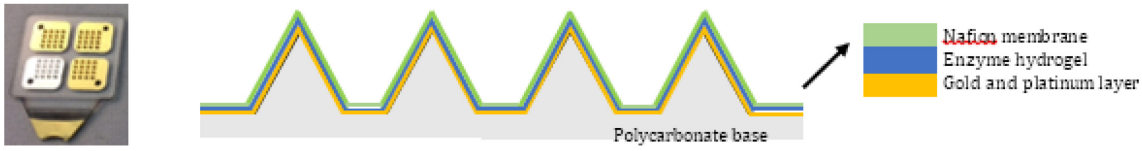
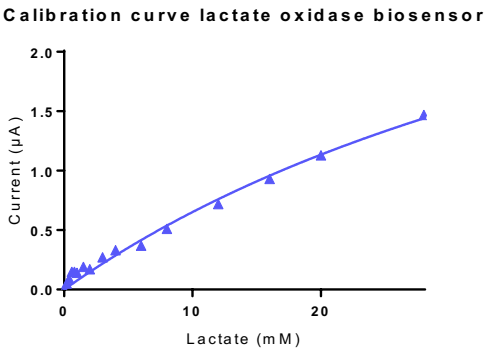


Figure 1. Photograph of the microneedle biosensor and a cross sectional diagram of layers

The biosensor is inserted into the viable epidermis of the skin at a depth of approximately 0.8 mm where it is in direct contact with interstitial fluid. When interstitial fluid lactate comes into contact with the surface of the biosensor it undergoes enzymatic oxidation, producing an electrical charge proportionate to the lactate concentration. This continuous process is detected by a potentiostat device connected to the biosensor.

Efficacy

In vitro calibrations have shown that the biosensor performs well across a range of lactate concentrations (0 – 20 mM):



Safety information

All components used in the fabrication of the lactate microneedle biosensor are non-toxic and safe for use in the skin. The polycarbonate base and metals (gold, silver and platinum) are inert. The PEG used for the hydrogel construction have a good safety profile, and are used in many existing biomedical applications [1]. The membrane of the sensor is composed of Nafion, an inert polymer and which is non-toxic [2].

Clinical safety information

The microneedle platform has been tested in human studies on glucose and antibiotic sensing, and have shown good performance and tolerability over periods of 8-24 hours of continuous use [3], [4]. Consensus from patient and public involvement through science festivals and events have also shown increased acceptability of the microneedle biosensor over blood testing [5].



Figure 2. (Left) Photograph of the skin after microneedle removal at different timepoints after placement for 8 hours. Note lack of surrounding redness in all photographs. (Right) visual analogue scale of microneedle (MN) discomfort against a peripheral cannula in 8 volunteers showing significantly better tolerability.

References

[1] "Polyethylene Glycol (PEG) Selection Guide," Sigma-Aldrich. [Online]. Available: https://www.sigmaaldrich.com/technical-documents/articles/materials-science/polyethylene-glycol-selection-guide.html.

[2] "Nafion Sigma-Aldrich Safety Data Sheet (SDS)," Sigma-Aldrich. [Online]. Available: https://www.sigmaaldrich.com/catalog/product/aldrich/309389.

[3] S. Sharma et al., "A pilot study in humans of microneedle sensor arrays for continuous glucose monitoring," Anal. Methods, vol. 10, no. 18, pp. 2088–2095, 2018.

[4] T. M. Rawson et al., "Microneedle biosensors for real-time, minimally invasive drug monitoring of phenoxymethylpenicillin: a first-in-human evaluation in healthy volunteers," Lancet Digit. Health, vol. 1, no. 7, pp. e335–e343, Nov. 2019.

[5] T. M. Rawson et al., "Public acceptability of computer-controlled antibiotic management: An exploration of automated dosing and opportunities for implementation," J. Infect., Aug. 2018.

d Patient information sheet for MIRACLE study

Department of Infectious Disease
Imperial College London
8th Floor Commonwealth Building
Imperial College London at Hammersmith Campus
Du Cane Road, London
W12 ONN, UK
Tel: +44 (0)20 3313 2732

INFORMATION SHEET FOR RESEARCH PARTICIPANTS – HEALTHY VOLUNTEER STUDY

You will be given a copy of this information sheet and a signed copy of your consent form to keep, should you decide to participate in the study.

Study title: Minimally Invasive Realtime Assessment of Continuous Lactate in Exercise (MIRACLE)

Chief Investigator: Professor Alison Holmes

You are invited to take part in a research study exploring the use of a new microneedle biosensor to measure a compound called lactate in exercise.
It is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with friends, relatives and your doctor if you wish. Contact us (details at the end) if anything is unclear and we can aim to provide more information.
The study is conducted for the Centre for Antimicrobial Optimisation which is funded by the Department of Health and Social Care, as well as part of a PhD project.
You are free to withdraw at any time without explanation and thank you for reading this.

1. WHAT IS THE PURPOSE OF THE STUDY?

Lactate is an important compound which is produced naturally by the body in exercise. However, its level can also increase in illness (in conditions like sepsis). Currently

lactate can only be measured through blood tests, and this can lead to difficulties in repeated measurements are needed.
Imperial College London have developed a new low-cost lactate microneedle biosensor. This is a device which has small (<1 mm) blunt spikes which are placed on the surface of your skin. It is minimally-invasive, painless and does not involve any blood testing. The biosensor is able to continuously measure the lactate profile in your skin, and transmit the information to a computer. It works through detecting the lactate in interstitial fluid, which is a thin layer of clear fluid between the cells in your skin which helps bring nutrients and oxygen from blood to the individual cells. The microneedle biosensor has been shown to be effective, and well-tolerated in our previous studies on diabetes and antibiotic measurements, also conducted at Imperial College.
The aim is to use the microneedle to measure the lactate produced during, and after a period of moderate intensity exercise and compare this with the normal ways of measuring lactate (i.e. blood tests).
The purpose for this study is to see how well the microneedle performs for lactate measurement, and to see how participants feel about wearing it during exercise. The results will allow us to improve the device and potentially test it on patients in the future.

2. WHY HAVE I BEEN INVITED?

You have previously given permission for your contact details to be included on the Imperial Clinical Research Facility Healthy Volunteer database, or have contacted us through posters placed around Imperial. Healthy adult volunteers who are able to do moderate intensity exercise continuously for up to 30 minutes are being invited to participate in this study. Moderate intensity exercises are activities which increase your heart rate, and may include brisk walking, jogging, cycling, hiking or dancing continuously without stopping for 30 minutes.

3. DO I HAVE TO TAKE PART?

No, there is no obligation to take part in the study. Even if you do change your mind after agreeing you can withdraw from the study at any stage without giving a reason. Doing this will not affect your medical care in the future and it will not have any consequences on your health.

4. WHAT WILL HAPPEN TO ME IF I DECIDE TO TAKE PART?

You will be invited to a screening visit at the NIHR-Imperial Clinical Research Facility, Hammersmith Hospital where you will meet a researcher face to face. The researcher will explain the study to you and ensure you are suitable for the study. You will have the opportunity to ask any questions related to the study. We will then ask you to complete and sign a consent form if you would like to participate.
The study can proceed straight after the screening step. We will place a microneedle biosensor on your arm by pressing it gently on the skin and fixing it with a dressing. We will also place a small tube called a microdialysis catheter under the skin in your arm– this will be done with local anaesthetic cream to reduce any discomfort. The microdialysis catheter will remove small volumes of interstitial fluid for lactate testing. In addition, another small intravenous cannula will be placed in the vein in the arm, in order to draw blood for lactate during the study.
Once set up, we ask you to perform some gentle exercise on an exercise bike. The intensity of exercise will increase every few minutes and will be equivalent to cycling uphill at a moderate pace at its most intense. During this time, blood and a small amount of interstitial fluid will be sampled every 5 minutes from the catheters. The exercise period will stop after 30 minutes and we ask you to sit down and rest for the

remainder of the study. Blood and interstitial fluid sampling will continue during this period.
The study will end an hour after the start of exercising. All equipment will be removed and you will be asked to fill out a short questionnaire. You will be asked for a photograph of the skin where the biosensor has been placed – this is for our records. We will ensure that you will not be able to be identified from the photograph alone. After this you are free to leave.
You are free to withdraw from the study at any point during the study. If the exercise intensity is too high this can also be adjusted or the exercise can stop early.

5. WHAT DO I HAVE TO DO?

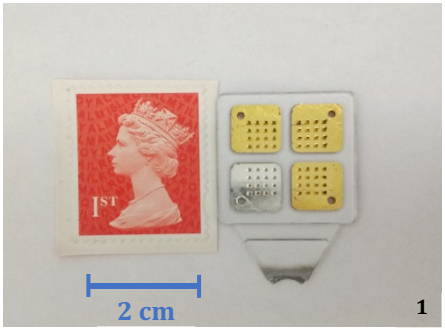
You will be asked to cycle on an exercise bike for 30 minutes, at an increasing intensity. We also ask you to bring some clothing and footwear suitable for the exercise.

6. WHAT ARE WE TESTING FOR?

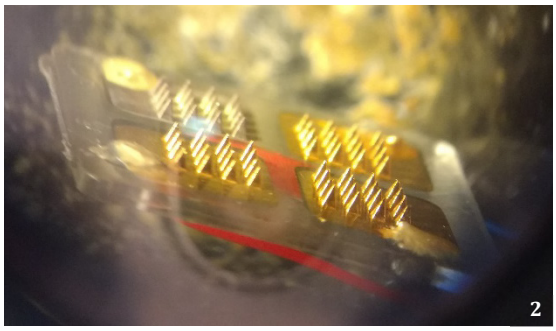
We are testing how well the lactate microneedle biosensor works in human volunteers during exercise. To do this we will also measure blood, and interstitial fluid lactate at the same time as a comparison. We want to know how you find wearing the biosensor during the study, and will ask you to complete a questionnaire at the end.

7. WHAT DO THE BIOSENSOR AND MICRODIALYSIS DEVICES LOOK LIKE?

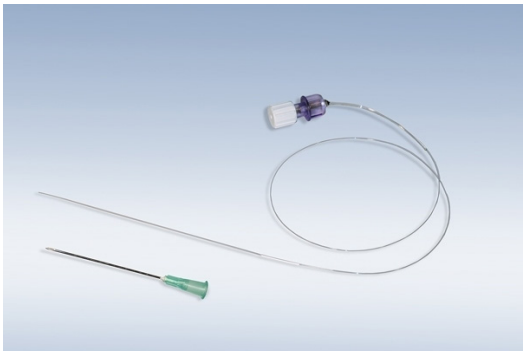
The lactate microneedle biosensor is a small flat device measuring 2 × 2 cm with protrusions of 1 mm. This is shown below in figure 1, and under magnification in figure 2. The device is placed onto the skin painlessly and held in place with a dressing:



1


2

The microdialysis catheter is a small plastic tube measuring less than 1 mm in width and is placed under the skin with anaesthetic (shown with a green needle for comparison below)



8. WHAT ARE THE SIDE EFFECTS, DISADVANTAGES AND RISKS OF TAKING PART?

All procedures and equipment used in this study have been shown to be safe in previous human studies. We therefore do not expect any significant side effects during, or after the study.

For the lactate microneedle biosensor, the main adverse reactions are possible skin irritation or discomfort. The microneedle has been used for 24 hours in human studies without any issues with comfort. The materials used to construct the biosensor are safe for external use in humans. If there are any signs of skin issues during the study, the biosensor will be removed and the skin observed. Risk for this is minimalised as the device will only be used for a short duration in this study (1 hour). As the sterile device is placed on the top part of your skin (less than 1 mm in depth), any damage to the arm or infection is unlikely.

You will have a cannula placed in a vein in your arm in order to allow us to draw blood easily from you during the study. The cannula is a sterile plastic tube which is commonly used in hospital care and is well-tolerated and should not cause pain after placement. However, if the placement is unsuccessful as we are unable to locate your vein, we might need to repeat its placement in another vein. This might result in bruising in your skin although it will not cause any other problems – the bruising typically resolves in less than 1 week. We will also be taking 13 blood samples from your cannula – the total amount of blood drawn (26 mL) is small and will not result in any problems with your health.

The microdialysis is also well-tolerated, and pain is minimised through use of a local anaesthetic cream which numbs the skin. Any damage to the arm is low as it is placed near the top of the skin, although the procedure itself might still cause pain and minor bruising. There is a low risk of infection although this is minimised through use of a sterile technique.

The exercise will be done on an exercise bicycle – the regimen is designed to make you feel tired in order to build up lactate. The total duration of exercise is 30 minutes but if you feel you need to stop or slow down then please tell the researcher.

9. WHAT ARE THE POSSIBLE BENEFITS OF TAKING PART?

There is no direct benefit to you if you take part – the information however will be useful in improving the design of the lactate microneedle biosensor for use in exercise as well as in patients in the future.

We can provide you with your lactate results in exercise on request – please contact us on the email address at the end.

10. WHAT IF NEW INFORMATION BECOMES AVAILABLE?

If new information becomes available during a research project regarding the question being studied, your research team will tell you about it and discuss whether you wish to continue participating in the study.

If you do decide to continue in the study, you will be asked to sign an updated consent form. Alternatively, on receiving new information your research team might consider it to be in your best interests to withdraw you from the study.

11. WHAT HAPPENS WHEN THE RESEARCH STUDY STOPS?

Once the study has finished, the results of the study can be made available to you should you wish. If you have any problems immediately following the study, then you should contact one of the research doctors on the numbers provided below.

12. WHAT IF I CHANGE MY MIND?

If you decide that you do not wish to take part in the study anymore and wish to stop, please inform the researcher in person, by phone or email. You do not have to provide a reason for this.

You can do this at any stage of the study and this will not affect your care or any future treatment. Please note that the information which have been collected so far will still be used for the purposes of the study unless specified.

13. WHAT IF SOMETHING GOES WRONG?

Imperial College London holds insurance policies which apply to this study. If you experience harm or injury as a result of taking part in this study, you will be eligible to claim compensation without having to prove that Imperial College is at fault. This does not affect your legal rights to seek compensation.

If you are harmed due to someon''s negligence, then you may have grounds for a legal action. Regardless of this, if you wish to complain, or have any concerns about any aspect of the way you have been treated during the course of this study then you should immediately inform the investigator (Professor Alison Holmes, 0203 313 2732, alison.holmes@imperial.ac.uk).

The normal National Health Service mechanisms are also available to you. If you are still not satisfied with the response, you may contact the Imperial College, Joint Research Compliance Office (jrco@ic.ac.uk).

14. WILL MY TAKING PART IN THIS STUDY BE KEPT CONFIDENTIAL?

Yes–- all information that is collected about you during this research will be kept strictly confidential. On entering the study, you will be provided a participant number under which all your results will be pseudonymised and not immediately identifiable. Basic identifiable information such as name, contact details and date of birth will be stored on an Imperial College Healthcare NHS Trust computer.

For blood testing, you will be assigned an Imperial College Healthcare NHS Trust patient number – the results will be stored against your name. However, this information is also password protected and will only be accessed by the research team. Your signed consent form will be kept with your medical notes at the NIHR-Imperial Clinical Research Facility. All other information will be pseudonymised and stored on an Imperial College London computer.

Imperial College London is the sponsor for this study based in the United Kingdom. We will be using information from you in order to undertake this study and will act as the data controller for this study. This means that we are responsible for looking after your information and using it properly. Imperial College London. will keep identifiable information about you.

- 10 years after the study has finished in relation to data subject consent forms.
- 10 years after the study has completed in relation to primary research data.

Further information on Imperial College London's retention periods may be found at https://www.imperial.ac.uk/media/imperial-college/administration-and-support-services/records-and-archives/public/RetentionSchedule.pdf

Your rights to access, change or move your information are limited, as we need to manage your information in specific ways in order for the research to be reliable and accurate. If you withdraw from the study, we will keep the information about you that we have already obtained. To safeguard your rights, we will use the minimum personally-identifiable information possible. You can find out more about how we use your information through contacting the principal investigator:

Prof Alison Holmes (alison.holmes@imperial.ac.uk)
Tel: 02033132732 (Mon to Fri 09.00 – 17.00)

Imperial College London will use your name, age, gender, NHS number and contact details to contact you about the research study, and make sure that relevant information about the study is recorded for your care, and to oversee the quality of the study. If you do not have a set of medical records at Imperial College Healthcare NHS Trust, this will be created for you for the purposes of the study, and only accessed where it is directly relevant to your taking part in the study, such as to record your participation or to document other findings. Individuals from Imperial College London and regulatory organisations may also look at your medical and research records to check the accuracy of the research study only where it is necessary. Imperial College Healthcare NHS Trust will pass these details to Imperial College London along with the information collected from you. The only people in Imperial College London who will have access to information that identifies you will be people who need to contact you f1or the research study or to follow up in event of unexpected findings or audit the data collection process. The people who analyse the information will not be able to identify you and will not be able to find out your name, NHS number or contact details.

LEGAL BASIS
As a university we use personally-identifiable information to conduct research to improve health, care and services. As a publicly-funded organisation, we have to ensure that it is in the public interest when we use personally-identifiable information from people who have agreed to take part in research. This means that when you agree to take part in a research study, we will use your data in the ways needed to conduct and analyse the research study.
Health and care research should serve the public interest, which means that we have to demonstrate that our research serves the interests of society as a whole. We do this by following the UK Policy Framework for Health and Social Care Research.

INTERNATIONAL TRANSFERS
There may be a requirement to transfer information to countries outside the European Economic Area (for example, to a research partner). Where this information contains your personal data, Imperial College London will ensure that it is transferred in accordance with data protection legislation. If the data is transferred to a country which is not subject to a European Commission (EC) adequacy decision in respect of its data protection standards, Imperial College London will enter into a data sharing agreement with the recipient organisation that incorporates EC approved standard contractual clauses that safeguard how your personal data is processed.

CONTACT US
If you wish to raise a complaint on how we have handled your personal data or if you want to find out more about how we use your information, please contact Imperial College London's Data Protection Officer via email at dpo@imperial.ac.uk, via telephone on 020 7594 3502 and via post at Imperial College London, Data Protection Officer, Faculty Building Level 4, London SW7 2AZ.

If you are not satisfied with our response or believe we are processing your personal data in a way that is not lawful you can complain to the Information Commissioner's Office (ICO). The ICO does recommend that you seek to resolve matters with the data controller (us) first before involving the regulator.

When you agree to take part in a research study, the information about your health and care may be provided to researchers running other research studies in this organisation and in other organisations. These organisations may be universities, NHS organisations or companies involved in health and care research in this country or abroad. Your information will only be used by organisations and researchers to conduct research in accordance with the UK Policy Framework for Health and Social Care Research.

This information will not identify you and will not be combined with other information in a way that could identify you. The information will only be used for the purpose of health and care research, and cannot be used to contact you or to affect your care. It will not be used to make decisions about future services available to you, such as insurance. We will also inform your general practitioner by letter that you have taken part in the study unless you do not wish for us to.

15. WHO WILL ACCESS MY DATA?

Only the study team will have direct access to your personal, research data and photographs. The information will be securely stored at Imperial College Healthcare NHS Trust with access restrictions.

16. WHAT WILL HAPPEN TO THE RESULTS OF THE RESEARCH STUDY?

The results are likely to be published in the six months following the study. We might also include the photographs in the publication but you will not be able to be identified from these images. Your confidentiality will always be ensured, and you will not be identified in any publication. At the end of the study, the results can be made available to you on request.

17. WHO IS ORGANISING AND FUNDING THE RESEARCH?

This study is being organised by the Centre for Antimicrobial Optimisation at Imperial College London, and funded by the Department of Health and Social Care.

18. WHO HAS REVIEWED THE STUDY?

This study has been reviewed by Joint Research Compliance Office at Imperial College London, Bloomsbury Research Ethics Committee and Health Research Authority

19. CONTACT FOR FURTHER INFORMATION

If you experience any problems during the study, you may withdraw at any stage. Additionally, you may contact us after completion of the study. We will provide you with a telephone number providing direct access, to one of the doctors (Dr Ming or Prof Holmes) involved in the study. The doctors may also be contacted through Professor Holmes' Head of Operations (0203 313 2072) or contact details provided below. The hospital switchboard (020 8383 1000) holds the home and mobile phone numbers for all the doctors involved in the study and can contact them at any time outside normal working hours. Please note that the Clinical Research Facility is not open 24-hours a day.

20. REIMBURSEMENT

The research team can offer reimbursement for your time and travel expenses of up to £15 per participant, plus travel expenses.

21. CONTACT DETAILS FOR THE STUDY

Dr Damien Ming (research physician)
E: antibiotic.research@imperial.ac.uk
T: 020 3313 2732

Thank you for being a part of this study!

e Patient questionnaire

Minimally Invasive Realtime Assessment of Continuous Lactate in Exercise

Post-study questionnaire

Participant Number _____

Date ___/____/___

Please take a moment to complete the follow questions by marking a vertical line at the point the corresponds to your current situation relating to the questions.

1) I have pain where the biosensor is placed:

No pain |——————————————————————| Worst possible pain

I have discomfort where the biosensor has been placed:

2)

No discomfort |——————————————————————| Worst possible discomfort

3) I feel restricted by the biosensor:

No restriction |——————————————————————| Worst possible restriction

4) My skin feels hot / burning:

No heat / burning |——————————————————————| Worst possible heat / burning

5) I have pain where the cannula is placed:

No pain |——————————————————————| Worst possible pain

6) I have pain where the microdialysis catheter is placed:

No pain |——————————————————————| Worst possible pain

Other comments?

f Patient consent form for MIRACLE study

Minimally Invasive Realtime Assessment of Continuous Lactate in Exercise (MIRACLE)

PARTICIPANT CONSENT FORM (VERSION 1.1)

CHIEF INVESTIGATOR: PROF. ALISON HOLMES (ANTIBIOTIC.RESEARCH@IMPERIAL.AC.UK)

The participant should complete the whole of this sheet (please initial each statement if it applies to you)

	Initials
I confirm that I have read and understand the Patient Information Sheet dated 10/2/2020 (version 1.1) for the above study and I have had the opportunity to ask questions which have	☐
I understand that my participation is voluntary and I am free to withdraw at any time, without giving any reason and without my medical care or legal rights being affected.	☐
I understand that sections of any of my medical notes may be looked at by responsible individuals from Imperial College London, from Imperial College Healthcare NHS Trust or from	☐
I give permission for these individuals to access my records that are relevant to this research.	☐
I **give/do not give** (delete as applicable) consent for information collected about me to be used to support other research in the future, including those outside of the EEA.	☐
I consent to taking part in the whole of the MIRACLE study and undergo testing of the medical device.	☐
I consent to having photographs of my skin taken for research purposes and understand that these will not contain any personal identifiable information.	☐
I consent to these photographs to be used in further research development as well as sharing.	☐
I consent to my general practitioner being informed regarding participating in this study.	☐

_____	_____	_____
Name of Participant	Signature	Date
_____	_____	_____
Name of Person taking consent	Signature	Date
_____	_____	_____
Principal Investigator	Signature	Date

1 copy for patient/participant; 1 copy for Principal Investigator; 1 copy to be kept with hospital notes

g Sponsorship letter from Imperial College London for MIRACLE study

Imperial College London

Imperial College Healthcare NHS Trust

Joint Research Compliance Office
Academic Health Science Centre
Imperial College London and
Imperial College Healthcare NHS Trust
Room 215, Medical School Building
St Mary's Hospital
Pread Street
London W2 1NY
Tel: +44 (0)20 759 49459
Becky.ward@imperial.ac.uk
www.ic.ac.uk/clinicalresearchgovernanceoffice

Becky Ward
Research Governance Manager

14/02/2020

Professor Alison Holmes
Commonwealth Building
Hammersmith Hospital
DuCane Road
London
W12 0NN

Dear Professor Holmes

RE: Minimally Invasive Realtime Assessment of Continuous Lactate in Exercise (MIRACLE)

Joint Research Compliance Office Reference number: 19HH5646

This is to confirm that the above named research project utilises human participants, their organs, tissue and/or data as defined under the sponsorship requirements of the UK policy framework for health and social care, incorporating the Medicines for Human Use (Clinical Trials) Regulations 2004.

On behalf of Imperial College of Science, Technology and Medicine, we undertake to act as the identified Research Sponsor for this project.

This letter confirms:

- The research proposal has been discussed, assessed and registered with the Joint Research Compliance Office, Imperial College Academic Health Science Centre, Imperial College London and provisional sponsor approval granted.
- The Chief Investigator has undergone a process of scientific critique commensurate with the scale of the project.
- Indemnity and insurance arrangements have been put in place to cover the project.
- Resources and support are available to the research team to aid delivery of the research as proposed.
- Management, monitoring and reporting responsibilities for the research have been approved.
- Imperial College will undertake and enforce those sponsor duties set out in the UK policy Framework for Health and Social Care.

Imperial College Sponsorship is conditional on the project receiving applicable ethical and regulatory approval for all research related aspects of its conduct. It is also conditional on successful contract and agreement negotiations and sign off via the Joint Research Office, where relevant, and before the study commences.

A copy of the ethics approval letter **must** be sent to the Research Governance Manager prior to the study commencing. Sponsorship is dependant on obtaining R&D Office approval for all NHS sites where the research is being conducted.

Yours sincerely

[redacted]

Becky Ward
Research Governance Manager

h Correspondence with MHRA and notice of no objection for device validation study

From: Smolenska, Daniella
To: Ming, Damien Keng Yen
Cc: O'Hare, Danny; Holmes, Alison H
Subject: RE: E/2019/2081 Advice for in-house testing of a novel device and healthcare institution exemption E/2019/2081
Date: 15 January 2020 15:45:34
Attachments: image001.png

Dear Damien

This enquiry has been passed to me.

I note that the proposed study is intended to assess feasibility of a microneedle biosensor intended to measure lactate which has been developed within Imperial College London. The microneedle device will undergo optimisation and possible further development depending on the results of the study and further clinical studies in sepsis are proposed.

It is clear from the correspondence below and the protocol that a medical purpose is foreseen for the microneedle biosensor. The study appears to be being conducted within Imperial college london without any third party involvement.

The main consideration at this time is therefore whether Imperial intend to develop this device for use outside of their entity.

We base our advice on the information presented to us and consider that where there is a possible future commercial intent, that studies being conducted on humans fall within the remit of the regulations.

In earlier correspondence it was stated that 'Commercialisation might come afterwards but there are no immediate plans currently'. You have since clarified this to confirm that there is no intention of commercialisation. As long as this is the case and there is no intention to develop the device for use outside of Imperial college London for a medical purpose, then MHRA would not consider these activities to fall under the current Medical Devices Regulations 2002 (as amended). Notification to MHRA will therefore not be required.

Please note that new medical devices regulations (EU regulation 2017/745) will be fully applicable from 26th May 2020. Article 5 covers the Health Institution Exemption and does not include clinical investigations. MHRA are currently considering the requirements for clinical investigations that are being conducted within a health institution and will be consulting on this very soon. Please keep an eye on our website from next Monday for further details on this.

I hope this email is helpful

Please note that whilst we are willing to give any help and advice we can, any views given by us on the interpretation of the Medical Device Regulations represent our best judgement at the time, based on the information available. Such views are not meant to be a definitive statement of law, which may only be given by the Courts. Accordingly we would always advise you to seek the views of your own professional advisors.

Please note, we would advise you to familiarise yourself with the new EU Medical Devices Regulation 2017/745 (MDR) and the *in vitro* Diagnostic Medical Devices Regulation 2017/746 (IVDR) as they will fully apply in the UK from the 26 May 2020 and 26 May 2022 respectively.

The Regulations introduce extensive changes, which you can read about in our guidance. These changes will affect manufacturers, importers, distributors, authorised representatives, as well as health institutions.

Kind regards
Daniella

Working pattern: My working days are Tuesdays and Wednesdays

Mrs Daniella Smolenska
Regulatory Affairs Manager
Devices Regulatory Group
MHRA
10 South Colonnade
Canary Wharf
London
E14 4PU

Tel: +44 (0) 20 3080 7363
Email: Daniella.smolenska@mhra.gov.uk

From: Ming, Damien Keng Yen <d.ming@imperial.ac.uk>
Sent: 15 January 2020 05:11
To: Dhillon, Annie <Annie.Dhillon@mhra.gov.uk>
Cc: Fleetcroft, Camilla <Camilla.Fleetcroft@mhra.gov.uk>; Holmes, Alison H <alison.holmes@imperial.ac.uk>; O'Hare, Danny <d.ohare@imperial.ac.uk>
Subject: FW: E/2019/2081 Advice for in-house testing of a novel device and healthcare institution exemption E/2019/2081

Dear Annie and Camilla,

Re: E/2019/2081 Minimally Invasive Realtime Assessment of Continuous Lactate in Exercise study

I am a clinical researcher at Imperial College London and we are proposing the testing of a novel microneedle device developed within Imperial, in order to measure lactate. There is no intention of commercialisation and we have been seeking a MHRA confirmation that the study falls under the healthcare institution exemption. This study will take place in-house within Imperial College London.

Study designs based on the same microneedle platform for measuring glucose (IRAS 190530) and antibiotic levels (IRAS 236047) have been approved in the past under the healthcare institution exemption.

Our original submission was dated 1st Nov 2019.

Leo Withers has been handling our case however I understand he has left the MHRA, with outstanding enquires. His autoreply has suggested that we contact you for

ongoing help. The issue revolves around future commercialisation – however I would like to clarify that this is for proof-of-concept purposes and again, there is no intention to commercialise and the device will only be used within Imperial College.

We are obviously keen to resolve this issue soon given the time it has taken, and would be very grateful for your help.

I attach the protocol for review and would be happy to speak over the phone next week if this is helpful too.

Regards,
Damien

Clinical Research Fellow
Department of Infectious Disease
Imperial College London
d.ming@imperial.ac.uk
+44(0)7786462238

From: Ming, Damien Keng Yen
Sent: 15 January 2020 11:48
To: Withers, Leo <leo.withers@mhra.gov.uk>; Holmes, Alison H <alison.holmes@imperial.ac.uk>; O'Hare, Danny <d.ohare@imperial.ac.uk>
Subject: RE: E/2019/2081 Advice for in-house testing of a novel device and healthcare institution exemption E/2019/2081

Hi Leo,

I'm following up on this and wonder if I can provide any more information regarding the proposal.

If there is any additional uncertainty would it be possible to arrange a call to discuss starting next week (20th Jan 2020) onwards?

Regards,
Damien

Clinical Research Fellow
Department of Infectious Disease
Imperial College London
d.ming@imperial.ac.uk
+44(0)20331132732

From: Ming, Damien Keng Yen <d.ming@imperial.ac.uk>
Sent: 02 January 2020 17:08
To: Withers, Leo <leo.withers@mhra.gov.uk>; Holmes, Alison H <alison.holmes@imperial.ac.uk>; O'Hare, Danny <d.ohare@imperial.ac.uk>
Subject: Re: E/2019/2081 Advice for in-house testing of a novel device and healthcare institution exemption E/2019/2081

Dear Leo,

Thank you for your email.

Re: E/2019/2081 Minimally Invasive Realtime Assessment of Continuous Lactate in Exercise (MIRACLE)

I would like to make it clear that the primary purpose of the proposed research is wholly academic, and the aim is proof of concept for minimally-invasive lactate sensing.

I apologise for the ambiguity in my previous email to you dated 25/11/19. There is no intention of commercialisation.

This study is of a similar design to two previous published research projects at imperial which utilises the same microneedle platform (IRAS 236047 and 190530).

I would be extremely grateful if you could tell me if this will qualify for the MHRA in-house exemption.

I have copied in the head of development Dr Danny O'Hare and the principal investigator Professor Alison Holmes who will be able to provide more detailed information if required.

Likewise I would be very happy to talk over the phone with you at your convenience - please tell me your contact number and a good time.

Thank you,
Damien

From: Withers, Leo <Leo.Withers@mhra.gov.uk>
Sent: Monday, 30 December 2019, 14:40
To: Ming, Damien Keng Yen
Subject: RE: E/2019/2081 Advice for in-house testing of a novel device and healthcare institution exemption E/2019/2081

This email from Leo.Withers@mhra.gov.uk originates from outside Imperial. Do not click on links and attachments unless you recognise the sender. If you trust the sender, add them to your safe senders list to disable email stamping for this address.

Dear Damien,

To clarify, you are asking whether your study would be considered in-house.

Your device appears to be a microneedle biosensor developed within Imperial College London intended to continuous measure lactate in human subjects.

You state you have no intention to commercialise your microneedle device, however also state you may commercialise the device in the future following the completion of this planned study.

Based on the information you have provided, your study does not fall under the in-house exemption as you appear to intend to commercialise your device or a future version of the device in the future based on the results of your study.

As you intend to test your device on human subjects, it is non-CE marked and your study does not fall within the in-house exemption you will be required to notify MHRA of your planned study.

Here you will find a link to the approval process for new clinical investigations:
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/780231/Guidance_for_mfrs_on_clinical_trials_February_2019.pdf

Here you will also find information on the required documents for your submission and how to apply:
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/780236/Guidance_for_mfrs_-_compiling_a_submission_to_MHRA.pdf

Note all applications are now submitted electronically via the IRAS portal.

As set out in the medical device directive, MHRA will review your submission and provide a response of 'objection' or 'no objection' within 60 days of the clinical trial being validated and any additional information being supplied to MHRA by the manufacturer. You cannot commence the trial without a letter of 'no objection' from MHRA.

You will also need approval of an approved Research Ethics Committee (REC) along that of the Health Research Authority (HRA). We would advise you to seek further information regarding their approval processes.

Please note, we would advise you to familiarise yourself with the new EU Medical Devices Regulation 2017/745 (MDR) and the in vitro Diagnostic Medical Devices Regulation 2017/746 (IVDR) as they will fully apply in the UK from the 26 May 2020 and 26 May 2022 respectively.

The Regulations introduce extensive changes, which you can read about in our guidance. These changes will affect manufacturers, importers, distributors, authorised representatives, as well as health institutions.

Please note that whilst we are willing to give any help and advice we can, any views given by us on the interpretation of the Medical Device Regulations represent our best judgement at the time, based on the information available. Such views are not meant to be definitive statements of the law, which may only be given by the Courts. Accordingly, we would always advise enquirers to seek the views of their own professional advisors.

Kind regards,
Leo

Leo Withers
Regulatory Affairs Specialist (for Clinical Investigations of medical devices)
Devices Regulatory Group
MHRA
10 South Colonnade
Canary Wharf
London
E14 4PU

Email: leo.withers@mhra.gov.uk

From: Ming, Damien Keng Yen <d.ming@imperial.ac.uk>
Sent: 11 December 2019 11:20
To: Withers, Leo <Leo.Withers@mhra.gov.uk>
Subject: RE: E/2019/2081 Advice for in-house testing of a novel device and healthcare institution exemption E/2019/2081

Hi Leo,

RE: E/2019/2081 (Minimally Invasive Realtime Assessment of Continuous Lactate in Exercise (MIRACLE) study)

I wonder if you had an update on this enquiry submitted on 1st Nov?

It'd be good to confirm that the planned clinical investigation will not require MHRA notification as it is used in-house and will not be commercialised.

i IRAS application form for MIRACLE study for Health Research Authority approval

Full Set of Project Data IRAS Version 5.13

Welcome to the Integrated Research Application System

IRAS Project Filter

The integrated dataset required for your project will be created from the answers you give to the following questions. The system will generate only those questions and sections which (a) apply to your study type and (b) are required by the bodies reviewing your study. Please ensure you answer all the questions before proceeding with your applications.

Please complete the questions in order. If you change the response to a question, please select 'Save' and review all the questions as your change may have affected subsequent questions.

Please enter a short title for this project (maximum 70 characters)
Minimally-Invasive Lactate in Exercise

1. Is your project research?

(•) Yes ( ) No

2. Select one category from the list below:

( ) Clinical trial of an investigational medicinal product
(•) Clinical investigation or other study of a medical device
( ) Combined trial of an investigational medicinal product and an investigational medical device
( ) Other clinical trial to study a novel intervention or randomised clinical trial to compare interventions in clinical practice
( ) Basic science study involving procedures with human participants
( ) Study administering questionnaires/interviews for quantitative analysis, or using mixed quantitative/qualitative methodology
( ) Study involving qualitative methods only
( ) Study limited to working with human tissue samples (or other human biological samples) and data (specific project only)
( ) Study limited to working with data (specific project only)
( ) Research tissue bank
( ) Research database

If your work does not fit any of these categories, select the option below:

( ) Other study

2a. Is the study sponsored or funded by a device manufacturer or other commercial company?

( ) Yes (•) No

Please select one of the following:

( ) Clinical study of a non-CE marked device where commercialisation of the product is intended
(•) Clinical study of a non-CE marked device for use within the institution, where commercialisation is not intended
( ) Clinical study of one or more CE marked devices for an off-label indication
( ) Clinical study of one or more CE marked devices for a labelled indication, involving a change to standard care or randomisation between groups

DRAFT

○ Clinical study of one or more CE marked devices for a labelled indication, involving *no* change to standard care or randomisation between groups

○ Pre-clinical device development or performance testing

2b. Please answer the following question(s):

a) Does the study involve the use of any ionising radiation? ○ Yes ◉ No

b) Will you be taking new human tissue samples (or other human biological samples)? ◉ Yes ○ No

c) Will you be using existing human tissue samples (or other human biological samples)? ○ Yes ◉ No

3. In which countries of the UK will the research sites be located? *(Tick all that apply)*

☑ England
☐ Scotland
☐ Wales
☐ Northern Ireland

3a. In which country of the UK will the lead NHS R&D office be located:

◉ England
○ Scotland
○ Wales
○ Northern Ireland
○ This study does not involve the NHS

4. Which applications do you require?

☑ IRAS Form
☐ Confidentiality Advisory Group (CAG)
☐ Her Majesty's Prison and Probation Service (HMPPS)

5. Will any research sites in this study be NHS organisations?

◉ Yes ○ No

5a. Are all the research costs and infrastructure costs (funding for the support and facilities needed to carry out research e.g. NHS Support costs) for this study provided by a NIHR Biomedical Research Centre, NIHR Collaboration for Leadership in Health Research and Care (CLAHRC), NIHR Patient Safety Translational Research Centre or Medtech and In Vitro Diagnostic Cooperative in all study sites?

Please see information button for further details.

○ Yes ◉ No

Please see information button for further details.

5b. Do you wish to make an application for the study to be considered for NIHR Clinical Research Network (CRN) Support and inclusion in the NIHR Clinical Research Network Portfolio?

DRAFT

Please see information button for further details.

(•) Yes ( ) No

The NIHR Clinical Research Network provides researchers with the practical support they need to make clinical studies happen in the NHS e.g. by providing access to the people and facilities needed to carry out research "on the ground".

If you select yes to this question, you must complete a NIHR Clinical Research Network (CRN) Portfolio Application Form (PAF) immediately after completing this project filter question and before submitting other applications. Failing to complete the PAF ahead of other applications e.g. HRA Approval, may mean that you will be unable to access NIHR CRN Support for your study.

6. Do you plan to include any participants who are children?

( ) Yes (•) No

7. Do you plan at any stage of the project to undertake intrusive research involving adults lacking capacity to consent for themselves?

( ) Yes (•) No

Answer Yes if you plan to recruit living participants aged 16 or over who lack capacity, or to retain them in the study following loss of capacity. Intrusive research means any research with the living requiring consent in law. This includes use of identifiable tissue samples or personal information, except where application is being made to the Confidentiality Advisory Group to set aside the common law duty of confidentiality in England and Wales. Please consult the guidance notes for further information on the legal frameworks for research involving adults lacking capacity in the UK.

8. Do you plan to include any participants who are prisoners or young offenders in the custody of HM Prison Service or who are offenders supervised by the probation service in England or Wales?

( ) Yes (•) No

9. Is the study or any part of it being undertaken as an educational project?

(•) Yes ( ) No

Please describe briefly the involvement of the student(s):
The project will be a part of a PhD project enrolled at Imperial College London - the student will be responsible for study design, recruitment, day-to-day project mangement and analysis of the study.

The academic supervision will be provided by the principal investigator.

9a. Is the project being undertaken in part fulfilment of a PhD or other doctorate?

(•) Yes ( ) No

10. Will this research be financially supported by the United States Department of Health and Human Services or any of its divisions, agencies or programs?

( ) Yes (•) No

11. Will identifiable patient data be accessed outside the care team without prior consent at any stage of the project (including identification of potential participants)?

( ) Yes (•) No

DRAFT

DRAFT

Integrated Research Application System
Application Form for Medical Device Study

The Chief Investigator should complete this form. Guidance on the questions is available wherever you see this symbol displayed. We recommend reading the guidance first. The complete guidance and a glossary are available by selecting Help.

Please define any terms or acronyms that might not be familar to lay reviewers of the application.

Short title and version number: (maximum 70 characters - this will be inserted as header on all forms)
Minimally-Invasive Lactate in Exercise

PART A: Core study information

1. ADMINISTRATIVE DETAILS

A1. Full title of the research:

Minimally Invasive Realtime Assessment of Continuous Lactate in Exercise (MIRACLE)

A2-1. Educational projects

Name and contact details of student(s):

Student 1

	Title	Forename/Initials	Surname
	Dr	Damien	Ming
Address	Commonwealth Building Hammersmith Hospital Du Cane Rd, London		
Post Code	W12 0NN		
E-mail	d.ming@ic.ac.uk		
Telephone	020 3313 2732		
Fax			

Give details of the educational course or degree for which this research is being undertaken:
Name and level of course/ degree:
PhD in Medicine

Name of educational establishment:
Imperial College London

Name and contact details of academic supervisor(s):

Academic supervisor 1

	Title	Forename/Initials	Surname
	Professor	Alison	Holmes
Address	Commonwealth Building		

DRAFT

	Hammersmith Hospital
	Du Cane Rd, London
Post Code	W12 0NN
E-mail	alison.holmes@imperial.ac.uk
Telephone	
Fax	

Please state which academic supervisor(s) has responsibility for which student(s):
Please click "Save now" before completing this table. This will ensure that all of the student and academic supervisor details are shown correctly.

Student(s)	Academic supervisor(s)
Student 1 Dr Damien Ming	☑ Professor Alison Holmes

A copy of a current CV for the student and the academic supervisor (maximum 2 pages of A4) must be submitted with the application.

A2-2. Who will act as Chief Investigator for this study?

- ○ Student
- ◉ Academic supervisor
- ○ Other

A3-1. Chief Investigator:

	Title	Forename/Initials	Surname
	Professor	Alison	Holmes
Post	Professor in Infectious Diseases		
Qualifications	MBBS MA MD MPH FMedSCi		
ORCID ID			
Employer			
Work Address	Commonwealth Building Hammersmith Hospital Du Cane Rd, London		
Post Code	W12 0NN		
Work E-mail	alison.holmes@imperial.ac.uk		
* Personal E-mail	alison.holmes@imperial.ac.uk		
Work Telephone			
* Personal Telephone/Mobile	020 3313 2732		
Fax			

** This information is optional. It will not be placed in the public domain or disclosed to any other third party without prior consent.*
A copy of a current CV (maximum 2 pages of A4) for the Chief Investigator must be submitted with the application.

A4. Who is the contact on behalf of the sponsor for all correspondence relating to applications for this project?
This contact will receive copies of all correspondence from REC and HRA/R&D reviewers that is sent to the CI.

DRAFT

	Title	Forename/Initials	Surname
	Ms	Becky	Ward
Address	Joint Research Compliance Office		
	Room 221 Medical School Building		
	St Marys Campus London		
Post Code	W2 1PG		
E-mail	becky.ward@imperial.ac.uk		
Telephone	020 7594 9459		
Fax			

A5-1. Research reference numbers. *Please give any relevant references for your study:*

Applicant's/organisation's own reference number, e.g. R & D (if available):	19HH5646
Sponsor's/protocol number:	19HH5646
Protocol Version:	1.2
Protocol Date:	10/02/2020
Funder's reference number (enter the reference number or state not applicable):	NIHR200646
Project website:	

Registry reference number(s):
The UK Policy Framework for Health and Social Care Research sets out the principle of making information about research publicly available. Furthermore: Article 19 of the World Medical Association Declaration of Helsinki adopted in 2008 states that "every clinical trial must be registered on a publicly accessible database before recruitment of the first subject"; and the International Committee of Medical Journal Editors (ICMJE) will consider a clinical trial for publication only if it has been registered in an appropriate registry. Please see guidance for more information.

International Standard Randomised Controlled Trial Number (ISRCTN):
ClinicalTrials.gov Identifier (NCT number): NCT04238611

Additional reference number(s):

Ref.Number	Description	Reference Number

A5-2. Is this application linked to a previous study or another current application?

(•) Yes ( ) No

Please give brief details and reference numbers.
Use of microneedle for antibiotic measurement: IRAS 236047 and use of microneedle in glucose monitoring: IRAS 190530

Members of these previous studies are also involved in the current study proposal. The microneedle biosensor platform used is similar in design.

2. OVERVIEW OF THE RESEARCH

To provide all the information required by review bodies and research information systems, we ask a number of specific questions. This section invites you to give an overview using language comprehensible to lay reviewers and members of the public. Please read the guidance notes for advice on this section.

A6-1. Summary of the study. *Please provide a brief summary of the research (maximum 300 words) using language easily understood by lay reviewers and members of the public. Where the research is reviewed by a REC within the UK*

Health Departments' Research Ethics Service, this summary will be published on the Health Research Authority (HRA) website following the ethical review. Please refer to the question specific guidance for this question.

Lactate is a naturally-occurring compound produced by the body, and measuring its level is useful for clinicians. A normal increase in lactate can come from exercise and the levels can help with designing training regimens for athletes. However, an abnormal increase in lactate can also result from serious conditions like sepsis. Measuring lactate in this situation is useful to assess the severity of the condition, and help with treatment. At the moment, lactate can only be measured by a blood test. This can be uncomfortable, and if you needed to know how the lactate levels change over time, repeated blood tests are required.

We would like to test a new device called a microneedle biosensor to measure lactate. This is a small postage-stamped device placed on the top of your skin, and can measure lactate continuously. The device has been tested in people with diabetes and for measuring antibiotic levels – it is safe, painless and has shown to work well. We have laboratory data showing that the microneedle can also measure lactate accurately, but this needs to be tested in real life.

We plan to recruit 15 healthy participants to cycle on an exercise bike for 30 minutes. During this time the microneedle will be used to measure lactate in the arm. We will also measure lactate normally using blood tests and a technique called microdialysis which measures lactate from the skin. We will compare how well the microneedle performs compared with the other ways of measurement, and get feedback from the participants. The study will last around 3 hours overall.

This research will be important in improving the microneedle design. The results will also allow us to explore if we can use the microneedle for patients who are ill in hospitals, and improve their care in the future.

A6-2. Summary of main issues. *Please summarise the main ethical, legal, or management issues arising from your study and say how you have addressed them.*

Not all studies raise significant issues. Some studies may have straightforward ethical or other issues that can be identified and managed routinely. Others may present significant issues requiring further consideration by a REC, R&D office or other review body (as appropriate to the issue). Studies that present a minimal risk to participants may raise complex organisational or legal issues. You should try to consider all the types of issues that the different reviewers may need to consider.

Use of the microneedle platform

The microneedle is a small, sterile device which is placed on the skin surface. Small (< 1 mm) protrusions are in direct contact with the interstitial fluid between cells and it is able to continuously measure compounds in this fluid. The device sits above all nerve fibres and blood vessels and therefore does not hurt. The same platform has been demonstrated to be safe, well-tolerated and effective in previous Phase I and Phase II clinical studies for antibiotic (IRAS 236047) and glucose measurement (IRAS 190530) respectively. In the glucose study, the microneedle device was kept on the skin surface of patients with diabetes for up to 24 hours and resulted in a consistently low pain score (median of 1 out of 10 on a visual analogue scale). There were no complications reported from the either of the studies.

The risk of infection and skin irritation is low given it does not breach the epidermis, and the device will be kept in place on the skin only for a short period of time (less than 2 hours). The microneedles are functionalised to detect lactate using a lactate oxidase enzyme immobilised on the surface. All compounds in the microneedle are safe to use, and are inert and non-irritant/non-toxic to the skin.
During the study any pain, or observed redness will result in the removal of the microneedle device. Participants will be advised to contact the research team if any issues develop after the study.

Safety of microdialysis

Microdialysis is an established clinical and research technique for the measurement of interstitial fluid. The technique involves insertion of a subcutaneous probe using aseptic non-touch technique. Pain or discomfort is minimised through the use of a topical anaesthesia cream (EMLA 5% containing lidocaine/prilocaine). There is a small risk of skin bruising which is self-limiting, but any other injury as a result of microdialysis probe insertion is low as it is inserted near the top of the skin and avoids any deeper parts of the body.

Exercise

Participants will be pre-screened for ability to perform aerobic exercise, with the requirement that they perform regular moderate intensity activity twice weekly. The exercise protocol involves cycling on an exercise bicycle for 30 minutes at increasing intensity. All equipment for the study will be secured safely in the upper arm and any sampling will be able

to proceed without cessation of cycling. The participant is free to pause, stop or adjust the intensity of the exercise at any point during the study by informing the research team. The exercise can also be stopped at the discretion of the research physician.

Blood-taking

All blood samples will be taken from an intravenous cannula, which is placed in the arm. Insertion of the cannula will follow sterile technique according to local Imperial Healthcare NHS Trust guidelines. There is a risk of bruising and discomfort, but this method of taking blood samples will reduce the number of times venepuncture occurs and reduces discomfort.

There is no risk to pregnant women being included in the study.

Confidentiality and data management

All data collected is pseudo-anonymised in the first instance through assigning the participants a study number. Clinical records and results are stored securely, and only the research team have access to the information. The data used for analysis, and publication will be fully anonymised.

3. PURPOSE AND DESIGN OF THE RESEARCH

A7. Select the appropriate methodology description for this research. *Please tick all that apply:*

- ☐ Case series/ case note review
- ☐ Case control
- ☐ Cohort observation
- ☐ Controlled trial without randomisation
- ☐ Cross-sectional study
- ☐ Database analysis
- ☐ Epidemiology
- ☑ Feasibility/ pilot study
- ☐ Laboratory study
- ☐ Metanalysis
- ☐ Qualitative research
- ☐ Questionnaire, interview or observation study
- ☐ Randomised controlled trial
- ☑ Other (please specify)

Device validation study

A10. What is the principal research question/objective? *Please put this in language comprehensible to a lay person.*

The main objective is to assess the performance of the microneedle device for measuring lactate, and compare this with blood test (gold standard) and microdialysis.

A11. What are the secondary research questions/objectives if applicable? *Please put this in language comprehensible to a lay person.*

The secondary objective is to assess how well the microneedle is tolerated for exercise using a participant questionnaire.

A12. What is the scientific justification for the research? *Please put this in language comprehensible to a lay person.*

DRAFT

Lactate is an important compound which is routinely measured in hospital care and for sports medicine. It is produced in a variety of ways by the body but in general, a higher blood lactate level in illness is associated with increased risk of death, or poorer outcomes. In exercise the lactate build-up also relates to the intensity of exercise a person can do – however this usually clears quickly with resting and does not cause harm to the person. The use of lactate has been extensively validated in many major international studies. Its measurement is also included in guidelines as a valuable marker and helps clinicians decide on treatment for conditions such as sepsis. The change of lactate over time (trend) is also important - for example, a lactate level which falls over a few hours suggests that the current treatment is improving the patient.

Currently lactate measurement is only possible through a blood test. This means that it can be uncomfortable to measure repeatedly. There are also disadvantages to its use in special populations such as children or the elderly, where blood taking can be difficult or distressing. Blood-taking is associated with risks to staff, such as needlestick injuries. The need to take blood may also influence clinicians' practice and affect how readily this compound is measured – for example if lactate is measured late in illness, then the appropriate treatment might also be delayed. There is also a time-cost to blood tests and the delay in measuring this important marker may also affect patient outcome.

The microneedle is a new platform which can continuously sense compounds in the fluid in the skin (the interstitial fluid, ISF). It has been shown that this microneedle design is a safe, and effective alternative to blood tests from two recent human studies in diabetes antibiotic level measurement, both performed at Imperial College by the same team. Lactate from the ISF is known to relate closely to lactate in the blood from studies in intensive care. Therefore, measuring ISF lactate using a microneedle could provide advantages if used in healthcare, as it is minimally-invasive and also gives a continuous measurement of lactate over time. Although the device works well in the laboratory it would be important to validate its performance in real life – an exercise model in healthy participants to generate lactate has been chosen for this study. As lactate applies to many other areas of healthcare – such as in emergency trauma, cardiac surgery, heart failure – if the platform is validated then it could potentially be applied to many different areas of healthcare.

A13. Please summarise your design and methodology. *It should be clear exactly what will happen to the research participant, how many times and in what order. Please complete this section in language comprehensible to the lay person. Do not simply reproduce or refer to the protocol. Further guidance is available in the guidance notes.*

The study will aim to recruit 15 healthy participants who regularly perform exercise. Participants will perform a 30-minute period of moderate-intensity exercise, whilst having their lactate levels measured through the microneedle, through repeated blood tests, and through microdialysis. The study will be conducted with a research physician within the NIHR-Clinical Research Facility located at the Hammersmith Hospital and will last approximately 4 hours from participants arriving to leaving.

Participants will be recruited from a combination of posters and from the healthy volunteer database kept by Imperial College London. Interested volunteers will be asked to contact the research team, and written participant information will be sent via email with a discussion over the phone if requested.

The screening will take place at least 24 hours after providing the participant information and an informed consent process will take place, with the study physician going through the participant information again.

The screening will involve the use of a questionnaire covering basic medical history, exercise habits, and will identify any criteria for exclusion. This will be performed in person with the participant to confirm eligibility. Physical examination will be carried out as necessary by the study physician – for example to ensure suitability of sites for microneedle biosensor and catheter placement.

The participant is informed that he/she can withdraw from the study at any point without providing a reason, and also request the exercise intensity to be decreased as required.

The study can take place immediately after screening:

1) A lactate microneedle biosensor is placed on the left arm and a stabilisation period of 30-60 minutes will take place. The microneedle will be connected to a potentiostat device which continuously records the ISF signals from the skin.

2) EMLA cream, which is a local anaesthetic, will also be applied on the skin of the participant on the left arm.

3) A peripheral cannula will be placed in the right arm to assist in taking blood.

4) A microdialysis catheter will be placed at the site of the local anaesthetic cream after cleaning.

5) Baseline blood and microdialysis samples will be taken from the participant – this is painless as the cannula / catheters are already in place.

6) The participant will be asked to start cycling on the specialised exercise bicycle, and the power output increased by 30 Watts every 5 minutes up to a maximum of 210 Watts (Similar to moderate uphill cycling).

7) During the exercise every 5 minutes, blood will be drawn from the cannula until the end of the study for lactate analysis

8) Microdialysis samples will also be removed every 5 minutes until the end of the study

9) The exercise duration will stop at 30 minutes, and the participant asked to rest.

10) The blood and microdialysis sampling will continue into the rest period for another 30 minutes

11) After this point, all cannulas / catheters will be removed, and the staff will ensure there are no complications which develop. The site of the device will also be photographed ensuring the participant cannot be identified.

12) The participant will be asked to complete a questionnaire and the study will end afterwards. At this point they are free to leave, and asked to contact the research team if there are any problems arising from the study.

The blood and microdialysis samples will be immediately refrigerated after the study for analysis, and the microneedle readings stored on a computer for analysis.

A14-1. In which aspects of the research process have you actively involved, or will you involve, patients, service users, and/or their carers, or members of the public?

- [x] Design of the research
- [ ] Management of the research
- [ ] Undertaking the research
- [ ] Analysis of results
- [x] Dissemination of findings
- [ ] None of the above

Give details of involvement, or if none please justify the absence of involvement.
The microneedle platform design and testing has been explored in numerous previous patient public involvement events - notably at Imperial Science Festival 2017 and 2018, as well as Imperial Lates events at South Kensington. Demonstrations and public opinion on the use of the devices, as an alternative to blood testing have been conducted and results published in peer reviewed journals (https://doi.org/10.1016/j.jinf.2018.08.005). The regular events have also been used as opportunities to showcase the technology and disseminate results.

4. RISKS AND ETHICAL ISSUES

RESEARCH PARTICIPANTS

A15. What is the sample group or cohort to be studied in this research?

Select all that apply:

- [ ] Blood
- [ ] Cancer
- [ ] Cardiovascular
- [ ] Congenital Disorders

- [ ] Dementias and Neurodegenerative Diseases
- [ ] Diabetes
- [ ] Ear
- [ ] Eye
- [x] Generic Health Relevance
- [ ] Infection
- [ ] Inflammatory and Immune System
- [ ] Injuries and Accidents
- [ ] Mental Health
- [x] Metabolic and Endocrine
- [ ] Musculoskeletal
- [ ] Neurological
- [ ] Oral and Gastrointestinal
- [ ] Paediatrics
- [ ] Renal and Urogenital
- [ ] Reproductive Health and Childbirth
- [ ] Respiratory
- [ ] Skin
- [ ] Stroke

Gender:	Male and female participants
Lower age limit: 18	Years
Upper age limit:	Years

A17-1. Please list the principal inclusion criteria (list the most important, max 5000 characters).

1. Consenting adults ≥ 18 years old

2. Healthy with no other previous medical history

3. Able to perform moderately intensive exercise on an exercise bicycle without difficulty for at least 30 minutes continuously, and engages in regular aerobic exercise at least twice a week

A17-2. Please list the principal exclusion criteria (list the most important, max 5000 characters).

1. Active inflammatory skin condition such as eczema or dermatitis

2. Active soft tissue infection or infection at any site

3. Known hypersensitivity to any microneedle component or dressings

4. Presence of any implantable electronic devices such as a pacemaker or stimulators

RESEARCH PROCEDURES, RISKS AND BENEFITS

A18. Give details of all non-clinical intervention(s) or procedure(s) that will be received by participants as part of the research protocol. *These include seeking consent, interviews, non-clinical observations and use of questionnaires.*

Please complete the columns for each intervention/procedure as follows:

DRAFT

1. Total number of interventions/procedures to be received by each participant as part of the research protocol.
2. If this intervention/procedure would be routinely given to participants as part of their care outside the research, how many of the total would be routine?
3. Average time taken per intervention/procedure (minutes, hours or days)
4. Details of who will conduct the intervention/procedure, and where it will take place.

Intervention or procedure	1	2	3	4
Recruitment Process – correspondence by email / phone	2	0	5 minutes	Research physician, in the Clinical research facility
Screening and consent in person	1	0	5 minutes	Research physician, in the Clinical research facility
Exercise on exercise bicycle	1	0	30 minutes	Participant, in the Clinical Research Facility
Study questionnaire	1	0	15 minutes	Participant, in the Clinical Research Facility

A19. Give details of any clinical intervention(s) or procedure(s) to be received by participants as part of the research protocol. *These include uses of medicinal products or devices, other medical treatments or assessments, mental health interventions, imaging investigations and taking samples of human biological material. Include procedures which might be received as routine clinical care outside of the research.*

Please complete the columns for each intervention/procedure as follows:

1. Total number of interventions/procedures to be received by each participant as part of the research protocol.
2. If this intervention/procedure would be routinely given to participants as part of their care outside the research, how many of the total would be routine?
3. Average time taken per intervention/procedure (minutes, hours or days).
4. Details of who will conduct the intervention/procedure, and where it will take place.

Intervention or procedure	1	2	3	4
Physical examination limited to skin and sites of catheter placement	1	0	2 minutes	Research physician, in the Clinical Research Facility
Application of microneedle biosensor	1	0	5 minutes	Research physician, in the Clinical Research Facility
Cannulation for blood taking	1	0	5 minutes	Research physician, in the Clinical Research Facility
Blood sampling from cannula	13	0	3 minutes	Research physician, in the Clinical Research Facility
Insertion of microdialysis catheter	1	0	10 minutes	Research physician, in the Clinical Research Facility
Interstitial fluid sampling from microdialysis	13	0	3 minutes	Research physician, in the Clinical Research Facility
Removal of microneedle device, cannula and microdialysis catheter	1	0	5 minutes	Research physician, in the Clinical Research Facility

A20. Will you withhold an intervention or procedure, which would normally be considered a part of routine care?

○ Yes ◉ No

A21. How long do you expect each participant to be in the study in total?

DRAFT

3 hours

A22. What are the potential risks and burdens for research participants and how will you minimise them?

For all studies, describe any potential adverse effects, pain, discomfort, distress, intrusion, inconvenience or changes to lifestyle. Only describe risks or burdens that could occur as a result of participation in the research. Say what steps would be taken to minimise risks and burdens as far as possible.

Microneedle biosensor

Irritation of the skin – the components used to construct the lactate microneedle biosensor are non-irritant as certified by the Safety Data Sheet from manufacturers. The device has been proven to be well-tolerated from other studies referenced. As the device will be in place for approximately 1 hour (previous studies have used this for up to 24 hours) it is anticipated that the risk of adverse reaction is low. The device will be removed at any signs of irritation of pain during the study. As the device is placed in the epidermis layer superficially – it is not expected to be associated with side effects associated with a breach of skin e.g. pain, bleeding or infection.

Exercise

The exercise period is limited to 30 minutes' of cycling on an exercise bicycle. The intensity of cycling will be increased incrementally and is equivalent to an uphill cycle at a moderate speed at its highest level. The participants will all have been screened for their ability to perform aerobic exercise beforehand. They will be asked to stop at any signs of abnormal discomfort and can withdraw from the exercise protocol at any point.

Microdialysis

Microdialysis insertion can be uncomfortable although this is minimised through the use of a local anaesthetic cream. Previous experiences have shown that this is well-tolerated. The microdialysis process itself may be uncomfortable and the flow rates of dialysis can be reduced if this occurs. The participants will be provided information on this during the consent process. The insertion will be undertaken as an aseptic procedure to minimise risk of infection. As the device is in place for approximately 1 hour, these risks are low.

Cannulation and blood taking

Cannulation of a large peripheral vein is a standard clinical procedure and not commonly associated with distress. Aseptic non touch techniques will be used to minimise risk of infection. Venepuncture may occasionally fail and require placement of the cannula in another vein, resulting in minor bruising. Blood taking will remove around 2 mL of blood for 13 blood draws = 26 mL: this is a small volume of blood to remove and will not result in any clinical consequences for the participant

A23. Will interviews/ questionnaires or group discussions include topics that might be sensitive, embarrassing or upsetting, or is it possible that criminal or other disclosures requiring action could occur during the study?

○ Yes ◉ No

A24. What is the potential for benefit to research participants?

There is no direct benefit for the participant in this study. However individual lactate results during exercise can be provided on request, which might be of interest for athlete training.

A25. What arrangements are being made for continued provision of the intervention for participants, if appropriate, once the research has finished? *May apply to any clinical intervention, including a drug, medical device, mental health intervention, complementary therapy, physiotherapy, dietary manipulation, lifestyle change, etc.*

Not applicable

A26. What are the potential risks for the researchers themselves? *(if any)*

The risks to participants are minimal but relate to insertion of cannulation and handling of unscreened blood samples. All staff on the research project have been trained in terms of procedure and safe sharps disposal.

DRAFT

RECRUITMENT AND INFORMED CONSENT

In this section we ask you to describe the recruitment procedures for the study. Please give separate details for different study groups where appropriate.

A27-1. How will potential participants, records or samples be identified? Who will carry this out and what resources will be used?*For example, identification may involve a disease register, computerised search of social care or GP records, or review of medical records. Indicate whether this will be done by the direct care team or by researchers acting under arrangements with the responsible care organisation(s).*

Participants will be identified through a combination of recruitment posters at Imperial College London sites and the use of the healthy volunteer database kept by Imperial College London. Study participants will be invited to contact the research team directly for further discussion based on the information provided.

A27-2. Will the identification of potential participants involve reviewing or screening the identifiable personal information of patients, service users or any other person?

(●) Yes ( ) No

Please give details below:
The identification of participants will use a healthy volunteer database which is maintained by the NIHR-Imperial Clinical Research Facility. Participants who have previously expressed an interest in research studies will have signed up to this mailing list which contains information regarding name, age, gender, height, weight, ethnicity, smoking status and contact details.

A27-3. Describe what measures will be taken to ensure there is no breach of any duty of confidentiality owed to patients, service users or any other person in the process of identifying potential participants.*Indicate what steps have been or will be taken to inform patients and service users of the potential use of their records for this purpose. Describe the arrangements to ensure that the wishes of patients and service users regarding access to their records are respected. Please consult the guidance notes on this topic.*

The healthy volunteer database is securely maintained according to the Data Protection Act 1998 until 24 May 2018 and the General Data Protection Regulations (the "GDPR") from 25 May 2018 onwards. Personal information is used only for specified purposes which includes suitability for clinical research studies, and access restricted to authorised parties specified.

Participants which have signed up for this database will have had information regarding how the data is stored and used explained to them in a privacy notice.

This is explicitly documented in the NIHR-Imperial Clinical Research Facility notice: http://imperial.crf.nihr.ac.uk/wp-content/uploads/2018/08/ICRF-GD PR-Privacy-Notice.pdf

A27-4. Will researchers or individuals other than the direct care team have access to identifiable personal information of any potential participants?

( ) Yes (●) No

A28. Will any participants be recruited by publicity through posters, leaflets, adverts or websites?

(●) Yes ( ) No

If Yes, please give details of how and where publicity will be conducted, and enclose copy of all advertising material (with version numbers and dates).
Posters will be placed on site at estates of Imperial College London where students and staff will be able to view the material.

A29. How and by whom will potential participants first be approached?

Participants will not be approached but rather they will respond to email and posters.

A30-1. Will you obtain informed consent from or on behalf of research participants?

(•) Yes ( ) No

If you will be obtaining consent from adult participants, please give details of who will take consent and how it will be done, with details of any steps to provide information (a written information sheet, videos, or interactive material). Arrangements for adults unable to consent for themselves should be described separately in Part B Section 6, and for children in Part B Section 7.

If you plan to seek informed consent from vulnerable groups, say how you will ensure that consent is voluntary and fully informed.

Not applicable

If you are not obtaining consent, please explain why not.

Not applicable

Please enclose a copy of the information sheet(s) and consent form(s).

A30-2. Will you record informed consent (or advice from consultees) in writing?

(•) Yes ( ) No

A31. How long will you allow potential participants to decide whether or not to take part?

At least 24 hours

A32. Will you recruit any participants who are involved in current research or have recently been involved in any research prior to recruitment?

(•) Yes
( ) No
( ) Not Known

If Yes, please give details and justify their inclusion. If Not Known, what steps will you take to find out?

The proposed study will not be affected by concurrent enrollment in another research study nor will it impact their welfare as a result.

A33-1. What arrangements have been made for persons who might not adequately understand verbal explanations or written information given in English, or who have special communication needs? *(e.g. translation, use of interpreters)*

A telephone translation service will be used where possible in cases where the persons are not able to understand English adequately. As this is a healthy volunteer study it is anticipated that this will not be a significant issue.

A34. What arrangements will you make to ensure participants receive any information that becomes available during the course of the research that may be relevant to their continued participation?

They will be contacted directly if there is new information relevant to their participation. However as the total study period is 3 hours it is unlikely that this will be an issue.

A35. What steps would you take if a participant, who has given informed consent, loses capacity to consent during the study? *Tick one option only.*

DRAFT

○ The participant and all identifiable data or tissue collected would be withdrawn from the study. Data or tissue which is not identifiable to the research team may be retained.

○ The participant would be withdrawn from the study. Identifiable data or tissue already collected with consent would be retained and used in the study. No further data or tissue would be collected or any other research procedures carried out on or in relation to the participant.

○ The participant would continue to be included in the study.

○ Not applicable – informed consent will not be sought from any participants in this research.

◉ Not applicable – it is not practicable for the research team to monitor capacity and continued capacity will be assumed.

Further details:
The study will take place over 3 hours only in healthy volunteers and it is not anticipated that the participant will lose capacity.

CONFIDENTIALITY

In this section, personal data means any data relating to a participant who could potentially be identified. It includes pseudonymised data capable of being linked to a participant through a unique code number.

Storage and use of personal data during the study

A36. Will you be undertaking any of the following activities at any stage (including in the identification of potential participants)? *(Tick as appropriate)*

- ☐ Access to medical records by those outside the direct healthcare team
- ☐ Access to social care records by those outside the direct social care team
- ☑ Electronic transfer by magnetic or optical media, email or computer networks
- ☐ Sharing of personal data with other organisations
- ☐ Export of personal data outside the EEA
- ☑ Use of personal addresses, postcodes, faxes, emails or telephone numbers
- ☐ Publication of direct quotations from respondents
- ☐ Publication of data that might allow identification of individuals
- ☑ Use of audio/visual recording devices
- ☑ Storage of personal data on any of the following:
  - ☑ Manual files (includes paper or film)
  - ☑ NHS computers
  - ☐ Social Care Service computers
  - ☐ Home or other personal computers
  - ☐ University computers
  - ☐ Private company computers
  - ☐ Laptop computers

Further details:
Use of personal addresses, postcodes, faxes, emails or telephone numbers: If participants would like to receive information of the outcome of the study on completion they will be able to provide contact details for information to be communicated to them upon publication of data. This information will be kept on an NHS computer within the Trust firewall on a secure network.

DRAFT

Digital photographs will be taken of the participant's skin at the end of the study and stored on NHS computers.

Manual files include signed consent forms and questionnaires will be stored in the NIHR-Clinical Research Facility.

Participant laboratory results and study data will be stored on server space on NHS computers - this data will be pseudo-anonymised. This will include data recorded by the potentiostat from the lactate microneedle device, which will be a time-series data file recorded against the study number of the participant.

A37. Please describe the physical security arrangements for storage of personal data during the study?

All personal identifiable information namely consent forms, master list of patients, screening questionnaires and assigned participant number will be stored in the NIHR-Clinical Research Facility. Access to the main building is keycard protected and the files are stored in a key locked cabinet in a locked room.

The trial master file will be kept similarly in a locked cabinet within a key-locked room at Imperial College London, access to the building will be through keycard access as well.

A38. How will you ensure the confidentiality of personal data? *Please provide a general statement of the policy and procedures for ensuring confidentiality, e.g. anonymisation or pseudonymisation of data.*

All data from the trial device will be identifiable through participant number only thus data is pseudo-anonymised. There will not be any means of linking participant number to personal information beyond the master file stored in NHS servers. Data for the purposes of subsequent analysis will be fully anonymised through decoupling participant number.

All patient-identifiable data will be kept on NHS computers which belong to Imperial College Healthcare NHS Trust. Access to personal data is restricted only to study staff only who will use their NHS smartcard to access data through Lightweight Directory Access Protocol (LDAP) authentication. Only members of the research team, the NHS, Imperial College London as sponsor, and regulatory authorities may review records when appropriate

The master list of patient details and participant number will also be kept on a NHS computer server which is password protected. Again access is restricted to study staff only on NHS computer terminals - the physical access of these computers require keycard access to Imperial College Healthcare NHS Trust estates.

A dedicated digital camera will be used to obtain photographs of the participant's device placement site on the skin at the end of the study - this will not contain any personal identifiable information (e.g. the face or visible markings). Images will be transferred to the server space on NHS computers and deleted from the camera device.

All study blood results will be discarded after analysis as per local policy at Imperial College Healthcare NHS Trust.

At the end of the study, data will be stored in Imperial College London archiving facility for a minimum of 10 years.

A40. Who will have access to participants' personal data during the study? *Where access is by individuals outside the direct care team, please justify and say whether consent will be sought.*

Only the study staff of the MIRACLE study will have access to participants' personal data. This will be analysed on computers within the Imperial College London firewall.

Storage and use of data after the end of the study

A41. Where will the data generated by the study be analysed and by whom?

Data will be analysed by the MIRACLE study staff, all of whom are employees of Imperial College London. The data will be fully anonymised at the analysis stage through removal of the participant number, and will be analysed at Imperial College London.

A42. Who will have control of and act as the custodian for the data generated by the study?

DRAFT

	Title	Forename/Initials	Surname
	Professor	Alison	Holmes
Post Qualifications	MBBS MA MD MPH FMedSCi		
Work Address	Commonwealth Building Hammersmith Hospital Du Cane Rd, London		
Post Code	W12 0NN		
Work Email	alison.holmes@imperial.ac.uk		
Work Telephone			
Fax			

A43. How long will personal data be stored or accessed after the study has ended?

- ○ Less than 3 months
- ○ 3 – 6 months
- ○ 6 – 12 months
- ○ 12 months – 3 years
- ◉ Over 3 years

If longer than 12 months, please justify:
10 years as per Imperial College London retention schedule. This requires that all study data, including consent forms will be securely stored at Imperial College London for 10 years following completion of the study.

A44. For how long will you store research data generated by the study?

Years: 10
Months: 0

A45. Please give details of the long term arrangements for storage of research data after the study has ended. *Say where data will be stored, who will have access and the arrangements to ensure security.*

All research data will be stored within Imperial College data storage infrastructure and will comply with the College research data management policy https://www.imperial.ac.uk/media/imperial-college/research-and-in novation/research-office/public/Imperial-College-RDM-Policy.pdf

All physical materials relating to the study including the master log, trial master file, case report and consent forms will be archived securely by Imperial College London for 10 years in a secure location and authorised access only. All electronic research data will be pseudo-anonymised and will be backed up on secure, encrypted servers managed through Imperial College London.

Consent forms and other participant identifiable information will be stored by Imperial College London for a minimum of 10 years.

Data will be stored in Imperial College London archiving facility at the end of the study. This will only be able to be accessed by lead researchers, the study PI and the Joint Research Compliance Office team at Imperial College London.

INCENTIVES AND PAYMENTS

A46. Will research participants receive any payments, reimbursement of expenses or any other benefits or incentives for taking part in this research?

- ◉ Yes ○ No

DRAFT

If Yes, please give details. For monetary payments, indicate how much and on what basis this has been determined.
£15 will be provided for time and travel expenses and calculated based on INVOLVE rates

A47. Will individual researchers receive any personal payment over and above normal salary, or any other benefits or incentives, for taking part in this research?

○ Yes ◉ No

A48. Does the Chief Investigator or any other investigator/collaborator have any direct personal involvement (e.g. financial, share holding, personal relationship etc.) in the organisations sponsoring or funding the research that may give rise to a possible conflict of interest?

○ Yes ◉ No

NOTIFICATION OF OTHER PROFESSIONALS

A49-1. Will you inform the participants' General Practitioners (and/or any other health or care professional responsible for their care) that they are taking part in the study?

◉ Yes ○ No

If Yes, please enclose a copy of the information sheet/letter for the GP/health professional with a version number and date.

A49-2. Will you seek permission from the research participants to inform their GP or other health/ care professional?

◉ Yes ○ No

It should be made clear in the participant's information sheet if the GP/health professional will be informed.

PUBLICATION AND DISSEMINATION

A50-1. Will the research be registered on a public database?

The UK Policy Framework for Health and Social Care Research sets out the principle of making information about research publicly available. Furthermore: Article 19 of the World Medical Association Declaration of Helsinki adopted in 2008 states that "every clinical trial must be registered on a publicly accessible database before recruitment of the first subject"; and the International Committee of Medical Journal Editors (ICMJE) will consider a clinical trial for publication only if it has been registered in an appropriate registry. Please see guidance for more information.

◉ Yes ○ No

Please give details, or justify if not registering the research.
This has been registered on the clinicaltrials.gov website (NCT04238611)

Please ensure that you have entered registry reference number(s) in question A5-1.

A51. How do you intend to report and disseminate the results of the study? *Tick as appropriate:*

☑ Peer reviewed scientific journals
☑ Internal report
☑ Conference presentation
☑ Publication on website

☐ Other publication

☐ Submission to regulatory authorities

☐ Access to raw data and right to publish freely by all investigators in study or by Independent Steering Committee on behalf of all investigators

☐ No plans to report or disseminate the results

☐ Other (please specify)

A52. If you will be using identifiable personal data, how will you ensure that anonymity will be maintained when publishing the results?

Not applicable

A53. Will you inform participants of the results?

◉ Yes ○ No

Please give details of how you will inform participants or justify if not doing so.
They will be able to request a copy of their results, as well as published article through contacting the research team after the end of the study.

5. Scientific and Statistical Review

A54-1. How has the scientific quality of the research been assessed? *Tick as appropriate:*

☑ Independent external review

☐ Review within a company

☑ Review within a multi−centre research group

☐ Review within the Chief Investigator's institution or host organisation

☑ Review within the research team

☑ Review by educational supervisor

☐ Other

Justify and describe the review process and outcome. If the review has been undertaken but not seen by the researcher, give details of the body which has undertaken the review:
The study is a multi-disciplinary endeavor and has been reviewed by members of Imperial College departments of medicine, chemistry and bioengineering.

An external peer review organised by Imperial College London has also been conducted and completed on 22/11/2019.

For all studies except non-doctoral student research, please enclose a copy of any available scientific critique reports, together with any related correspondence.

For non-doctoral student research, please enclose a copy of the assessment from your educational supervisor/ institution.

A56. How have the statistical aspects of the research been reviewed? *Tick as appropriate:*

☐ Review by independent statistician commissioned by funder or sponsor

☐ Other review by independent statistician

☐ Review by company statistician

☐ Review by a statistician within the Chief Investigator's institution

☐ Review by a statistician within the research team or multi−centre group

☐ Review by educational supervisor

☐ Other review by individual with relevant statistical expertise

☑ No review necessary as only frequencies and associations will be assessed – details of statistical input not required

In all cases please give details below of the individual responsible for reviewing the statistical aspects. If advice has been provided in confidence, give details of the department and institution concerned.

Title Forename/Initials Surname

Department
Institution
Work Address

Post Code
Telephone
Fax
Mobile
E-mail

Please enclose a copy of any available comments or reports from a statistician.

A57. What is the primary outcome measure for the study?

The primary outcome is the concordance between the lactate microneedle biosensor with venous and interstitial fluid lactate.

A58. What are the secondary outcome measures? *(if any)*

Participant acceptability of the biosensor through an end-of-study questionnaire

A59. What is the sample size for the research? *How many participants/samples/data records do you plan to study in total? If there is more than one group, please give further details below.*

Total UK sample size:	15
Total international sample size (including UK):	15
Total in European Economic Area:	0

Further details:
All participants will be recruited from within the UK.

A60. How was the sample size decided upon? *If a formal sample size calculation was used, indicate how this was done, giving sufficient information to justify and reproduce the calculation.*

The target recruitment number of this study is 15 participants. As the novel lactate biosensor has not previously undergone clinical testing, there are no preliminary data to perform a formal sample size calculation. The sample size has derived from previous similar microneedle studies on glucose and antibiotic measurement, as well as unpublished preliminary in vitro data.

A61-1. Will participants be allocated to groups at random?

○ Yes ◉ No

DRAFT

A62. Please describe the methods of analysis (statistical or other appropriate methods, e.g. for qualitative research) by which the data will be evaluated to meet the study objectives.

The data from the biosensor will be analysed for concordance with blood and interstitial fluid lactate using Bland-Altman plots and correlation coefficients using packages written in Python. Descriptive analyses of the change in lactate during and after exercise will also be performed.

6. MANAGEMENT OF THE RESEARCH

A63. Other key investigators/collaborators. *Please include all grant co-applicants, protocol co-authors and other key members of the Chief Investigator's team, including non-doctoral student researchers.*

	Title	Forename/Initials	Surname
	Professor	Tony	Cass
Post	Professor in Chemistry		
Qualifications	PhD		
Employer	Imperial College London		
Work Address	Molecular Sciences Research Hub White City Campus London		
Post Code	W12 0BZ		
Telephone	020 7594 7222		
Fax			
Mobile			
Work Email	t.cass@imperial.ac.uk		

	Title	Forename/Initials	Surname
	Dr	Danny	O'Hare
Post	Reader in Sensor Research		
Qualifications	PhD		
Employer	Imperial College London		
Work Address	Royal School of Mines South Kensington Campus Imperial College London		
Post Code	SW7 2AZ		
Telephone			
Fax			
Mobile			
Work Email	d.ohare@imperial.ac.uk		

	Title	Forename/Initials	Surname
	Dr	Damien	Ming
Post	Clinical Research Fellow		
Qualifications	BMBCh MA MRCP DTM&H		
Employer	Imperial College London		
Work Address	Department of Infectious Diseases Hammersmith Hospital London		

DRAFT

Post Code	W12 0HS
Telephone	02075942023
Fax	
Mobile	
Work Email	d.ming@ic.ac.uk

Title	Forename/Initials	Surname
Dr	David	Freeman

Post	Postdoctoral Research Associate
Qualifications	PhD
Employer	Imperial College London
Work Address	Molecular Sciences Research Hub White City Campus London
Post Code	W12 0BZ
Telephone	020 7594 7222
Fax	
Mobile	
Work Email	d.freeman@imperial.ac.uk

A64. Details of research sponsor(s)

A64-1. Sponsor

Lead Sponsor

Status:
- ○ NHS or HSC care organisation
- ◉ Academic
- ○ Pharmaceutical industry
- ○ Medical device industry
- ○ Local Authority
- ○ Other social care provider (including voluntary sector or private organisation)
- ○ Other

Commercial status: Non-Commercial

If Other, please specify:

Contact person

Name of organisation	Imperial College London JRCO
Given name	Becky
Family name	Ward
Address	Room 221
Town/city	Medical School Building, St Marys Campus
Post code	W2 1PG
Country	UNITED KINGDOM
Telephone	02075949459

Fax
E-mail becky.ward@imperial.ac.uk

A65. Has external funding for the research been secured?

Please tick at least one check box.

- [x] Funding secured from one or more funders
- [ ] External funding application to one or more funders in progress
- [ ] No application for external funding will be made

What type of research project is this?

- ○ Standalone project
- ○ Project that is part of a programme grant
- ◉ Project that is part of a Centre grant
- ○ Project that is part of a fellowship/ personal award/ research training award
- ○ Other

Other – please state:

Please give details of funding applications.

Organisation	Department of Health & Social Care
Address	CENTRAL COMMISSIONING FACILITY Grange House 15 Church Street
Post Code	TW1 3NL
Telephone	020 8843 8000
Fax	020 8843 8001
Mobile	
Email	ccf@nihr.ac.uk

Funding Application Status: ◉ Secured ○ In progress

Amount: £4,047,193

Duration
Years: 2
Months: 0

If applicable, please specify the programme/ funding stream:

What is the funding stream/ programme for this research project?
Seed project for optimising antimicrobial use

A66. Has responsibility for any specific research activities or procedures been delegated to a subcontractor (other than a co-sponsor listed in A64-1) ? *Please give details of subcontractors if applicable.*

○ Yes ◉ No

DRAFT

A67. Has this or a similar application been previously rejected by a Research Ethics Committee in the UK or another country?

○ Yes ◉ No

Please provide a copy of the unfavourable opinion letter(s). You should explain in your answer to question A6-2 how the reasons for the unfavourable opinion have been addressed in this application.

A68-1. Give details of the lead NHS R&D contact for this research:

	Title	Forename/Initials	Surname
	Ms	Becky	Ward
Organisation	Imperial College London		
Address	Imperial College London and Imperial College Healthcare NHS Trust Room 215, Level 2, Medical School Building, Norfolk Place, London		
Post Code	W2 1PG		
Work Email	becky.ward@imperial.ac.uk		
Telephone	0207 594 9459		
Fax			
Mobile			

Details can be obtained from the NHS R&D Forum website: http://www.rdforum.nhs.uk

A68-2. Select Local Clinical Research Network for NHS Organisation identified in A68-1:

North West London

For more information, please refer to the question specific guidance.

A69-1. How long do you expect the study to last in the UK?

Planned start date: 01/05/2020
Planned end date: 01/11/2020
Total duration:
Years: 0 Months: 6 Days: 1

A70.

Definition of the end of trial, and justification in the case where it is not the last visit of the last subject undergoing the trial

Last visit of last subject in study (participant number 15).

A71-1. Is this study?

◉ Single centre
○ Multicentre

A71-2. Where will the research take place? *(Tick as appropriate)*

- [x] England
- [ ] Scotland
- [ ] Wales
- [ ] Northern Ireland
- [ ] Other countries in European Economic Area

Total UK sites in study 1

Does this trial involve countries outside the EU?

( ) Yes (•) No

A72. Which organisations in the UK will host the research?*Please indicate the type of organisation by ticking the box and give approximate numbers if known:*

[x] NHS organisations in England	1
[ ] NHS organisations in Wales	
[ ] NHS organisations in Scotland	
[ ] HSC organisations in Northern Ireland	
[ ] GP practices in England	
[ ] GP practices in Wales	
[ ] GP practices in Scotland	
[ ] GP practices in Northern Ireland	
[ ] Joint health and social care agencies (eg community mental health teams)	
[ ] Local authorities	
[ ] Phase 1 trial units	
[ ] Prison establishments	
[ ] Probation areas	
[ ] Independent (private or voluntary sector) organisations	
[x] Educational establishments	1
[ ] Independent research units	
[ ] Other (give details)	
Total UK sites in study:	2

A73-1. Will potential participants be identified through any organisations other than the research sites listed above?

( ) Yes (•) No

A74. What arrangements are in place for monitoring and auditing the conduct of the research?

The Joint Research Compliance Office at Imperial College London will ensure the research fulfills the legal, ethical and scientific obligations of the healthcare research process. The team may audit this this study in line with research governance arrangements.

DRAFT

The project will have a working group that will meet every 2 weeks to review data and discuss issues that arise during the recruitment process.

A steering committee consisting of the study PI, research team, external representatives, and a patient representative will meet quarterly to review study progress and approve progression of the study.

The day to day running of the study will be managed by the lead researcher.

A75-1. What arrangements will be made to review interim safety and efficacy data from the trial? Will a formal data monitoring committee or equivalent body be convened?

Given the small patient sample size (n=15) safety will be monitored through adverse event reporting during the study. Safety will also be discussed in the project and steering committee which will meet every 2 weeks and quarter.

If a formal DMC is to be convened, please forward details of the membership and standard operating procedures to the Research Ethics Committee when available. The REC should also be notified of DMC recommendations and receive summary reports of interim analyses.

A75-2. What are the criteria for electively stopping the trial or other research prematurely?

The study will aim to recruit the full number of participants 15 in order to increase the generalisability of the results.

Failure of the lactate microneedle biosensor in measuring lactate, or any significant concerns about safety will also be criteria for electively stopping the study prematurely.

A76. Insurance/ indemnity to meet potential legal liabilities

Note: in this question to NHS indemnity schemes include equivalent schemes provided by Health and Social Care (HSC) in Northern Ireland

A76-1. What arrangements will be made for insurance and/or indemnity to meet the potential legal liability of the sponsor(s) for harm to participants arising from the management of the research? *Please tick box(es) as applicable.*

Note: Where a NHS organisation has agreed to act as sponsor or co-sponsor, indemnity is provided through NHS schemes. Indicate if this applies (there is no need to provide documentary evidence). For all other sponsors, please describe the arrangements and provide evidence.

- ☐ NHS indemnity scheme will apply (NHS sponsors only)
- ☑ Other insurance or indemnity arrangements will apply (give details below)

Imperial College London negligent and non-negligent harm insurance cover applies

Please enclose a copy of relevant documents.

A76-2. What arrangements will be made for insurance and/ or indemnity to meet the potential legal liability of the sponsor(s) or employer(s) for harm to participants arising from the design of the research? *Please tick box(es) as applicable.*

Note: Where researchers with substantive NHS employment contracts have designed the research, indemnity is provided through NHS schemes. Indicate if this applies (there is no need to provide documentary evidence). For other protocol authors (e.g. company employees, university members), please describe the arrangements and provide evidence.

- ☐ NHS indemnity scheme will apply (protocol authors with NHS contracts only)
- ☑ Other insurance or indemnity arrangements will apply (give details below)

Imperial College London negligent and non-negligent harm insurance cover applies

Please enclose a copy of relevant documents.

A76-3. What arrangements will be made for insurance and/ or indemnity to meet the potential legal liability of investigators/collaborators arising from harm to participants in the conduct of the research?

Note: Where the participants are NHS patients, indemnity is provided through the NHS schemes or through professional indemnity. Indicate if this applies to the whole study (there is no need to provide documentary evidence). Where non-NHS sites are to be included in the research, including private practices, please describe the arrangements which will be made at these sites and provide evidence.

- [x] NHS indemnity scheme or professional indemnity will apply (participants recruited at NHS sites only)
- [ ] Research includes non-NHS sites (give details of insurance/ indemnity arrangements for these sites below)

Imperial College London negligent and non-negligent harm insurance cover applies

Please enclose a copy of relevant documents.

A77. Has the sponsor(s) made arrangements for payment of compensation in the event of harm to the research participants where no legal liability arises?

- (•) Yes
- ( ) No

If Yes, please give details of the compensation policy:
Imperial College London negligent and non-negligent harm insurance cover applies

Please enclose a copy of relevant documents.

A78. Could the research lead to the development of a new product/process or the generation of intellectual property?

- (•) Yes
- ( ) No
- ( ) Not sure

A79. Please select the level of commercial participation in this project.

- (•) None
- ( ) Industry funding, but not industry sponsored
- ( ) Industry funding and industry sponsored
- ( ) Industry sponsored, but not industry funded

A80. Please select the main subject area of research. Additional sub-topics may be selected, if required

- [ ] Age and Ageing
- [ ] Anaesthetics
- [ ] Cancer (includes malignant haematology
- [ ] Cardiovascular
- [ ] Clinical
- [ ] Critical Care
- [ ] Dementias and Neurodegenerative Diseases
- [ ] Dermatology
- [ ] Diabetes
- [ ] Ear, Nose and Throat
- [ ] Gastrointestinal

DRAFT

- [ ] Genetics
- [ ] Health Services Research
- [ ] Hepatology
- [ ] Immunology and Inflammation
- [ ] Infectious Disease and Microbiology
- [ ] Injuries and Accidents
- [ ] Medicines for Children (does not include Paediatrics)
- [ ] Mental Health
- [x] Metabolic and Endocrine
- [ ] Musculoskeletal (Rheumatoid Arthritis is a separate category)
- [ ] Nervous System Disorders
- [ ] Non-malignant Haematology
- [ ] Ophthalmology
- [ ] Oral and Dental
- [ ] Paediatrics (does not include Medicines for Children)
- [ ] Primary Care
- [ ] Public Health Research
- [ ] Renal
- [ ] Reproductive Health and Childbirth
- [ ] Respiratory
- [ ] Rheumatoid Arthritis
- [ ] Stroke
- [ ] Surgery
- [ ] Urogenital

Part B: Section 2

A. General information

Information in this sub-section will be included in applications to the Research Ethics Committee and NHS R & D offices at the research sites.

1. Is the manufacturer (or other organisation responsible for developing the device) the same organisation named as lead sponsor for this study?

(•) Yes ( ) No

Organisation	Imperial College London JRCO
Address	Room 221 Medical School Building, St Marys Campus
Post Code	W2 1PG
Country	
Telephone	02075949459

DRAFT

Fax	
Mobile	
E-mail	becky.ward@imperial.ac.uk

2. Details of the medical devices to be used in the study

Name of the manufacturer:	Imperial College London JRCO
Manufacturer's trade name for the device:	Lactate Microneedle Biosensor
Device identification name and/or number:	
Name:	Lactate Microneedle Biosensor
Number:	
Generic name of device and principal intended use(s):	Lactate Microneedle Biosensor
Length of time since device came into use:	

3-1. Further details of the purpose of the study

Does the study involve:

- (●) Investigation of a new medical device
- ( ) Investigation of new implantable material
- ( ) Use of an existing product outside the terms of its CE market intended purpose
- ( ) Use of a modified product
- ( ) Use of an existing product within its CE market intended purpose

3-2. Please give further details below including the following:

Description of any new device, materials, method of use or operation with a summary of the intended purpose.

The study aims to examine the efficacy of the lactate microneedle biosensor in exercising participants. The same microneedle platform has been successfully trialed in healthy volunteers for antibiotic measurement and glucose sensing with good acceptability and performance (see references at bottom).

The lactate microneedle biosensor measures 2 × 2 cm in size and is composed of three layers on its functional surface:

i)A polycarbonate plastic base metallised with a thin layer of gold (Torr Scientific, Bexhill, UK) and platinum.

ii) A middle layer of lactate oxidase enzyme (Sekisui diagnostics, MA, USA) immobilised in a polyethylene glycol (PEG)-hydrogel.

iii) An outer membrane of Nafion (Sigma-Aldrich, MO, USA).

The device has 1.0 mm protrusions which are applied to the skin surface, in direct contact with the interstitial fluid. As the depth of penetration is superficial (approx. 0.8 mm), the biosensor resides within the top layer within the viable epidermis and does not come into contact with nerve fibres or blood vessels.

Sterilisation of the sensors has been organised through TCM associates Ltd, who have a standard protocol in place for sterilisation of the biosensors as part of previous studies using the same microneedles. This includes completion of bioburden assay's and quality assurance procedures

Method of use - the device will be connected to a potentiostat which is a device which is able to sense the electrical signals originating from the microneedle. The continuous readout will be recorded, and displayed on a laptop

DRAFT

computer.

References

T. M. Rawson et al., "Microneedle biosensors for real-time, minimally invasive drug monitoring of phenoxymethylpenicillin: a first-in-human evaluation in healthy volunteers," Lancet Digit. Health, vol. 1, no. 7, pp. e335–e343, Nov. 2019.

S. Sharma et al., "A pilot study in humans of microneedle sensor arrays for continuous glucose monitoring," Anal. Methods, vol. 10, no. 18, pp. 2088–2095, 2018.

Composition of any new implantable materials, including summary of biocompatibility findings from studies to date.

Not applicable

A summary of any modifications to CE marked devices.

Not applicable

A summary of any proposed changes to the CE market intended purpose.

Not applicable

For all products with CE mark please attach instructions for use.

4. Please describe the arrangements for manufacture of the investigational device. *Include details of the quality assurance system in place within the legal entity. Give details of any collaboration with a commercial manufacturer or other sub- contractor and enclose a copy of the contract.*

The lactate microneedle biosensor is manufactured in-house at Imperial College London by the departments of Chemistry and Bioengineering:

1. Base needle fabrication process:
A metal inlay is used for the injection moulding of polycarbonate pellets. The polycarbonate pellets were dried at 110°C for 24 hours under vacuum prior to use before injection moulding process at Tm = 270 °C (PC melt temperature), Tw = 80 °C (tool temperature) at injection speed of 20 cm3 s−1 and shot volume of 4.4 cm3 and a cooling time of 5s. Each polycarbonate microneedle structure (25 × 25 × 2 mm) comprised of four 4 × 4 microneedle arrays.

2. The base structures are then quality checked using microscopy and undergo metalisation using a custom magnetron sputter coater, which is provided by Torr Scientific Ltd.

3. Upon completion of sputtering, platinum is deposited on the surface of the microneedle through either electrochemical methods, or metalisation.

4. A layer of lactate oxidase suspended in a PEG-hydrogel is applied to the functional surface of the electrode.

5. Sterilisation of the sensors has been organised through TCM associates Ltd, who have a standard protocol in place for sterilisation of the devices as part of previous studies using the same microneedles. This includes completion of bioburden assay's and quality assurance procedures.

6. Following sterilisation, the needles will be returned and stored before use. In-vitro assessment has been undertaken to demonstrate that these sensors are still functional within expected limits after 2 months of storage. In the case of this study, it is expected that sensors will be used within 4 weeks of fabrication.

5. What safety and performance testing has been undertaken on the investigational device and its constituents? *Please give summarised details of appropriate tests (including outcome i.e. pass/fail), e.g. mechanical, electrical, biological, toxicological, sterilisation.*

All components used in the fabrication of the lactate microneedle biosensor are non-toxic and safe for use in the skin.

The polycarbonate base and metals (gold, silver and platinum) are inert. The PEG used for the hydrogel construction have a good safety profile, and are used in many existing biomedical applications. The membrane of the sensor is composed of Nafion, an inert polymer and which is non-toxic.

The specific array device used within our study has been demonstrated to have demonstrated biocompatability in previously published studies by members of our and other research groups, with a good safety profile - please refer to Phase I clinical studies referenced below by our group.

The microneedle platform has been tested in human studies on glucose and antibiotic sensing, and have shown good performance and tolerability over periods of 8-24 hours of continuous use (referenced in section 3-2). Consensus from patient and public involvement through science festivals and events have also shown increased acceptability of the microneedle biosensor over blood testing.

6. Please describe the sponsor's plans for further development and use of the device. *Indicate whether the plans include making it available (whether for a fee or not) to other legal entities or working with a device manufacturer or other company to commercialise the product.*

This is a pilot, in house feasibility study that aims to evaluate the performance of the device and estimate potential impact. Following this study, data will be reviewed and a follow on application will be made to request permission to evaluate the sensors and controller systems in patients as part of a larger clinical study in clinical settings, such as in sepsis. There are no immediate plans to commercialise the device at this time.

7. Declaration

This declaration should be authorised by the head of clinical engineering or equivalent at the institution developing the device.

I confirm that the information provided in this section is accurate to the best of my knowledge. I take full responsibility for ensuring that the device has been manufactured to the standards expected of an equivalent CE marked device and that all relevant testing to demonstrate compliance with these standards has been undertaken.

Name: .....

Post: .....

Organisation: .....

Date: .....

DRAFT

Part B: Section 5 – Use of newly obtained human tissue(or other human biological materials) for research purposes

1. What types of human tissue or other biological material will be included in the study?

1. Blood test for serum lactate
2. Interstitial fluid for lactate

2. Who will collect the samples?

The study clinician will collect the samples

3. Who will the samples be removed from?

- [x] Living donors
- [ ] The deceased

4. Will informed consent be obtained from living donors for use of the samples? *Please tick as appropriate*

In this research?

(•) Yes ( ) No

In future research?

( ) Yes ( ) No (•) Not applicable

6. Will any tissues or cells be used for human application or to carry out testing for human application in this research?

( ) Yes (•) No

8. Will the samples be stored: *[Tick as appropriate]*

In fully anonymised form? *(link to donor broken)*

( ) Yes (•) No

In linked anonymised form? *(linked to stored tissue but donor not identifiable to researchers)*

(•) Yes ( ) No

If Yes, say who will have access to the code and personal information about the donor.

All samples will analysed by North West London NHS pathology laboratory located at Hammersmith Hospital - the samples will be stored linked to the participant's hospital number which will be generated by the NIHR-Imperial Clinical Research Facility for the purposes of this study.

There will be a period of sample storage prior to analysis (less than 8 hours) and only the research team will have access to the study number and personal information of the study participant.

The sample will be stored, and discarded after analysis in line with local policy after 72 hours in the laboratory.

In a form in which the donor could be identifiable to researchers?

( ) Yes (•) No

DRAFT

9. What types of test or analysis will be carried out on the samples?

Lactate concentration only will be tested on the blood and interstitial fluid samples.

10. Will the research involve the analysis or use of human DNA in the samples?

○ Yes ◉ No

11. Is it possible that the research could produce findings of clinical significance for donors or their relatives?

○ Yes ◉ No

12. If so, will arrangements be made to notify the individuals concerned?

○ Yes ○ No ◉ Not applicable

13. Give details of where the samples will be stored, who will have access and the custodial arrangements.

All samples will be discarded after analysis, and this is estimated to take place within 72 hours of the samples being obtained. In the interim samples will be stored in the North West London Pathology laboratory located at Hammersmith Hospital, Imperial College Healthcare NHS Trust.

14. What will happen to the samples at the end of the research? *Please tick all that apply and give further details.*

☐ Transfer to research tissue bank

(If the bank is in England, Wales or Northern Ireland the institution will require a licence from the Human Tissue Authority to store relevant material for possible further research.)

☐ Storage by research team pending ethical approval for use in another project

(Unless the researcher's institution holds a storage licence from the Human Tissue Authority, or the tissue is stored in Scotland, or it is not relevant material, a further application for ethical review should be submitted before the end of this project.)

☐ Storage by research team as part of a new research tissue bank

(The institution will require a licence from the Human Tissue Authority if the bank will be storing relevant material in England, Wales or Northern Ireland. A separate application for ethical review of the tissue bank may also be submitted.)

☐ Storage by research team of biological material which is not "relevant material" for the purposes of the Human Tissue Act

☑ Disposal in accordance with the Human Tissue Authority's Code of Practice

☐ Other

☐ Not yet known

Please give further details of the proposed arrangements:

All samples will be discarded as per local Imperial College Healthcare NHS Trust pathology policies within 72 hours of processing.

PART C: Overview of research sites

Please enter details of the host organisations (Local Authority, NHS or other) in the UK that will be responsible for the research sites. *For further information please refer to guidance.*

Investigator identifier	Research site		Investigator Name	
IN1	◉ NHS/HSC Site ○ Non-NHS/HSC Site		Forename	Alison
			Middle name	
			Family name	Holmes
	Organisation name	IMPERIAL COLLEGE HEALTHCARE NHS TRUST	Email	alison.holmes@imperial.ac.uk
	Address	ST. MARYS HOSPITAL PRAED STREET LONDON GREATER LONDON	Qualification (MD...)	MBBS MA MD MPH FMedSCi
	Post Code	W2 1NY	Country	UNITED KINGDOM
	Country	ENGLAND		

DRAFT

j HRA ethics approval letter for MIRACLE study and related correspondence with NHS Research Ethics Committee

Health Research Authority

London - Bloomsbury Research Ethics Committee
HRA RES Centre Manchester
Barlow House 3rd Floor
4 Minshull Street
Manchester
M1 3DZ
Telephone: 0207 104 8196

Please note: This is the favourable opinion of the REC only and does not allow you to start your study at NHS sites in England until you receive HRA Approval

03 April 2020

Professor Alison Holmes
Commonwealth Building
Hammersmith Hospital
Du Cane Rd, London
W12 0NN

Dear Professor Holmes

Study title:	**Minimally Invasive Realtime Assessment of Continuous Lactate in Exercise (MIRACLE)**
REC reference:	**20/LO/0364**
Protocol number:	**19HH5646**
IRAS project ID:	**274161**

The Research Ethics Committee reviewed the above application at the meeting held on 01 April 2020. Thank you to Dr Damien Ming and Dr Danny O'Hare for attending via teleconference to discuss the application.

Ethical opinion

The members of the Committee present gave a favourable ethical opinion of the above research on the basis described in the application form, protocol and supporting documentation, subject to the conditions specified below.

Conditions of the favourable opinion

The REC favourable opinion is subject to the following conditions being met prior to the start of the study.

Number	Action Required
1	Clearly define the meaning of "healthy with no other previous medical history" in the Protocol, explaining that this means participants will not have any diagnosed medical condition from a medical practice.
2	Clearly define the term "moderate intensity exercise" in the Participant Information Sheet (PIS). The Committee suggest using the following definition "This accelerates heartrate and can include things like brisk walking, dancing, cycling slowly, gardening"
3	Make it clear in the protocol that pregnant women can participate.
4	Provide a clear explanation of the statistical analyses and any reviews undertaken to support section A56 the IRAS form. Please note that you do not need to update the IRAS form, supporting text within this response table will suffice.
5	Ensure that the payment of £15 plus travel expenses is consistently described across the relevant documents, as currently the reimbursement of additional travel expenses is not clear.
6	Clarify the types of medical notes that will be accessed, as statement 3 of the Informed Consent Form (ICF) currently states "any of my medical notes may be looked at". Make this explicit in the ICF, PIS and GP letter.
7	Explain the term 'interstitial fluid' in lay language in the PIS.
8	Correct the typo 'ovewr' in the GP letter.
9	Confirm whether Good Clinical Practice training has been undertaken and provide evidence of this by submitting the relevant certifications.
10	Please comment on the intentions of the study start date or how this will be otherwise handled, given the current Covid-19 pandemic. The following guidance on the HRA is provided for reference: https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/covid-19-guidance-sponsors-sites-and-researchers/.

You should notify the REC once all conditions have been met (except for site approvals from host organisations) and provide copies of any revised documentation with updated version numbers. Revised documents should be submitted to the REC electronically from IRAS. The REC will acknowledge receipt and provide a final list of the approved documentation for the study, which you can make available to host organisations to facilitate their permission for the study. Failure to provide the final versions to the REC may cause delay in obtaining permissions.

Confirmation of Capacity and Capability (in England, Northern Ireland and Wales) or NHS management permission (in Scotland) should be sought from all NHS organisations involved in the study in accordance with NHS research governance arrangements. Each NHS organisation must confirm through the signing of agreements and/or other documents that it has given permission for the research to proceed (except where explicitly specified otherwise).

Guidance on applying for HRA and HCRW Approval (England and Wales)/ NHS permission for research is available in the Integrated Research Application System.

For non-NHS sites, site management permission should be obtained in accordance with the procedures of the relevant host organisation.

Sponsors are not required to notify the Committee of management permissions from host organisations.

Registration of Clinical Trials

It is a condition of the REC favourable opinion that **all clinical trials are registered** on a publicly accessible database. For this purpose, 'clinical trials' are defined as the first four project categories in IRAS project filter question 2. Registration is a legal requirement for clinical trials of investigational medicinal products (CTIMPs), except for phase I trials in healthy volunteers (these must still register as a condition of the REC favourable opinion).

Registration should take place as early as possible and within six weeks of recruiting the first research participant at the latest. Failure to register is a breach of these approval conditions, unless a deferral has been agreed by or on behalf of the Research Ethics Committee (see here for more information on requesting a deferral: https://www.hra.nhs.uk/planning-and-improving-research/research-planning/research-registration-research-project-identifiers/

As set out in the UK Policy Framework, research sponsors are responsible for making information about research publicly available before it starts e.g. by registering the research project on a publicly accessible register. Further guidance on registration is available at: https://www.hra.nhs.uk/planning-and-improving-research/research-planning/transparency-responsibilities/

You should notify the REC of the registration details. We routinely audit applications for compliance with these conditions.

Publication of Your Research Summary

We will publish your research summary for the above study on the research summaries section of our website, together with your contact details, no earlier than three months from the date of this favourable opinion letter. Should you wish to provide a substitute contact point, make a request to defer, or require further information, please visit: https://www.hra.nhs.uk/planning-and-improving-research/application-summaries/research-summaries/

Notice of no objection must be obtained from the Medicines and Healthcare products Regulatory Agency (MHRA).

The sponsor is asked to provide the Committee with a copy of the notice from the MHRA, either confirming no objection or giving grounds for objection, as soon as this is available.

It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

After ethical review: Reporting requirements

The attached document "After ethical review – guidance for researchers" gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators

- Notification of serious breaches of the protocol
- Progress and safety reports
- Notifying the end of the study, including early termination of the study
- Final report

The latest guidance on these topics can be found at https://www.hra.nhs.uk/approvals-amendments/managing-your-approval/.

Ethical review of research sites

NHS/HSC Sites

The favourable opinion applies to all NHS/HSC sites taking part in the study taking part in the study, subject to confirmation of Capacity and Capability (in England, Northern Ireland and Wales) or NHS management permission (in Scotland)being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion" below).

Non-NHS/HSC sites

I am pleased to confirm that the favourable opinion applies to any non NHS/HSC sites listed in the application, subject to site management permission being obtained prior to the start of the study at the site.

Approved documents

The documents reviewed and approved at the meeting were:

Document	*Version*	*Date*
Copies of advertisement materials for research participants [Lactate in exercise - recruitment poster]	1.0	10 February 2020
Evidence of Sponsor insurance or indemnity (non NHS Sponsors only) [Evidence of indemnity - Imperial College London]		23 July 2019
GP/consultant information sheets or letters [Lactate in exercise - GP notification]	1.0	12 November 2019
IRAS Application Form [IRAS_Form_18022020]		18 February 2020
Letter from funder [Letter from funder - CAMO]		10 April 2019
Letter from sponsor [Imperial College London sponsor letter]		14 February 2020
Letters of invitation to participant [Lactate in exercise - letters of invitation]	1.0	10 February 2020
MHRA Notice of No Objection Letter (Medical Devices) and relevant correspondence [Email from MHRA - notice of no objection]		15 January 2020
Non-validated questionnaire [Lactate in exercise study questionnaire]	1.1	10 February 2020
Participant consent form [Lactate in Exercise participant consent form]	1.1	10 February 2020
Participant information sheet (PIS) [Lactate in exercise - participant information sheet]	1.1	10 February 2020
Referee's report or other scientific critique report [Lactate in exercise - external peer review certificate]		22 November 2019
Research protocol or project proposal [Lactate in exercise - protocol]	1.2	10 February 2020
Summary CV for Chief Investigator (CI) [Alison Holmes CV]		
Summary CV for student [Damien Ming CV]		16 February 2020

Summary, synopsis or diagram (flowchart) of protocol in non technical language [Lactate in exercise - flow diagram]	1.1	10 February 2020

Membership of the Committee

The members of the Ethics Committee who were present at the meeting are listed on the attached sheet.

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

User Feedback

The Health Research Authority is continually striving to provide a high quality service to all applicants and sponsors. You are invited to give your view of the service you have received and the application procedure. If you wish to make your views known please use the feedback form available on the HRA website: http://www.hra.nhs.uk/about-the-hra/governance/quality-assurance/

HRA Learning

We are pleased to welcome researchers and research staff to our HRA Learning Events and online learning opportunities– see details at: https://www.hra.nhs.uk/planning-and-improving-research/learning/

IRAS project ID: 274161 **Please quote this number on all correspondence**

With the Committee's best wishes for the success of this project.

Yours sincerely

[redacted]

Signed on behalf of;
Reverend Jim Linthicum
Chair

E-mail: bloomsbury.rec@hra.nhs.uk

Enclosures: List of names and professions of members who were present at the meeting and those who submitted written comments.

Copy to: Ms Becky Ward

London - Bloomsbury Research Ethics Committee

Attendance at Committee meeting on 01 April 2020

Committee Members:

Name	*Profession*	*Present*	*Notes*
Mr Christopher Exeter	Data Scientist	Yes	
Ms Kalvinder Gahir	Practice Support Pharmacist	No	
Dr Paul Gorczynski	Chartered Psychologist	Yes	Vice Chair
Miss Allyson Gray	Paediatric Clinical Research Nurse	No	
Dr Michael Jacobs	Retired Academic Pharmacologist	Yes	
Monica King	Named Nurse for safeguarding children	Yes	
Mr Enrico Leipzig-Pagani	Medical Distributor/Responsible Person	Yes	
Reverend Jim Linthicum	Hospital Chaplain	No	
Dr Mary Lowth	General Practitioner / Medicolegal doctor	No	
Ms Cathy MacLean	Clinical Project Manager	No	
Ms Clare Madin	Semi-Retired Clinical Data Management Manager	No	
Mrs Michelle McPhail	Senior Lecturer in Management Studies	Yes	
Ms Pippa Sipanoun	Research Associate, Great Ormond Street Hospital	Yes	
Mr David Wagstaff	Organisational Consultant	Yes	
Dr Ruth Williams	Consultant Paediatric Neurologist	No	

Also in attendance:

Name	*Position (or reason for attending)*
Miss Anna Martin	Approvals Administrator
Damilola Odunlami	Approvals Specialist

Ymchwil Iechyd a Gofal Cymru
Health and Care Research Wales

NHS
Health Research Authority

Professor Alison Holmes
Commonwealth Building
Hammersmith Hospital
Du Cane Rd, London
W12 0NN

Email: approvals@hra.nhs.uk
HCRW.approvals@wales.nhs.uk

07 May 2020

Dear Professor Holmes

HRA and Health and Care Research Wales (HCRW) Approval Letter

Study title:	**Minimally Invasive Realtime Assessment of Continuous Lactate in Exercise (MIRACLE)**
IRAS project ID:	**274161**
Protocol number:	**19HH5646**
REC reference:	**20/LO/0364**
Sponsor	**Imperial College London JRCO**

I am pleased to confirm that **HRA and Health and Care Research Wales (HCRW) Approval** has been given for the above referenced study, on the basis described in the application form, protocol, supporting documentation and any clarifications received. You should not expect to receive anything further relating to this application.

Please now work with participating NHS organisations to confirm capacity and capability, in line with the instructions provided in the "Information to support study set up" section towards the end of this letter.

How should I work with participating NHS/HSC organisations in Northern Ireland and Scotland?
HRA and HCRW Approval does not apply to NHS/HSC organisations within Northern Ireland and Scotland.

If you indicated in your IRAS form that you do have participating organisations in either of these devolved administrations, the final document set and the study wide governance report (including this letter) have been sent to the coordinating centre of each participating nation. The relevant national coordinating function/s will contact you as appropriate.

Please see IRAS Help for information on working with NHS/HSC organisations in Northern Ireland and Scotland.

How should I work with participating non-NHS organisations?
HRA and HCRW Approval does not apply to non-NHS organisations. You should work with your non-NHS organisations to obtain local agreement in accordance with their procedures.

What are my notification responsibilities during the study?

The standard conditions document "*After Ethical Review – guidance for sponsors and investigators*", issued with your REC favourable opinion, gives detailed guidance on reporting expectations for studies, including:

- Registration of research
- Notifying amendments
- Notifying the end of the study

The HRA website also provides guidance on these topics, and is updated in the light of changes in reporting expectations or procedures.

Who should I contact for further information?
Please do not hesitate to contact me for assistance with this application. My contact details are below.

Your IRAS project ID is **274161**. Please quote this on all correspondence.

Yours sincerely,
Damilola Odunlami

Approvals Specialist

Email: approvals@hra.nhs.uk

Copy to: Dr Damien Ming
Ms Becky Ward

List of Documents

The final document set assessed and approved by HRA and HCRW Approval is listed below.

Document	*Version*	*Date*
Copies of advertisement materials for research participants [Lactate in exercise - recruitment poster]	1.0	10 February 2020
Evidence of Sponsor insurance or indemnity (non NHS Sponsors only) [Evidence of indemnity - Imperial College London]		23 July 2019
GP/consultant information sheets or letters [Lactate in exercise - GP notification]	1.1	
IRAS Application Form [IRAS_Form_18022020]		18 February 2020
Letter from funder [Letter from funder - CAMO]		10 April 2019
Letter from sponsor [Imperial College London sponsor letter]		14 February 2020
Letters of invitation to participant [Lactate in exercise - letters of invitation]	1.1	
MHRA Notice of No Objection Letter (Medical Devices) and relevant correspondence [Email from MHRA - notice of no objection]		15 January 2020
Non-validated questionnaire [Lactate in exercise study questionnaire]	1.1	10 February 2020
Other [Response letter to REC conditions]		
Other [GCP training certificate for Damien Ming]		
Other [GCP training certificate for Alison Holmes]		
Other [GCP training certificate for David Freeman]		
Participant consent form [Lactate in Exercise participant consent form]	1.2	
Participant information sheet (PIS) [Lactate in exercise - participant information sheet]	1.2	
Referee's report or other scientific critique report [Lactate in exercise - external peer review certificate]		22 November 2019
Research protocol or project proposal [Lactate in exercise - protocol]	1.3	
Summary CV for Chief Investigator (CI) [Alison Holmes CV]		
Summary CV for student [Damien Ming CV]		16 February 2020
Summary, synopsis or diagram (flowchart) of protocol in non technical language [Lactate in exercise - flow diagram]	1.1	10 February 2020

IRAS project ID	274161

Information to support study set up

The below provides all parties with information to support the arranging and confirming of capacity and capability with participating NHS organisations in England and Wales. This is intended to be an accurate reflection of the study at the time of issue of this letter.

Types of participating NHS organisation	Expectations related to confirmation of capacity and capability	Agreement to be used	Funding arrangements	Oversight expectations	HR Good Practice Resource Pack expectations
This is a single site study sponsored by the participating NHS organisation's partner University. Therefore there is only one site type.	This is a single site study sponsored by the participating NHS organisations partner University. You should work with the participating sites R&D office to make arrangements to set up the study and confirm local capacity and capability. This R&D office will confirm to you when the study can start following issue of HRA and HCRW Approval.	This is a non-commercial single site study taking place in the NHS where that single NHS organisation's partner University is the study sponsor. Therefore no study agreements are required.	External study funding has been secured.	A Principal Investigator should be appointed at study sites of this type.	No Honorary Research Contracts, Letters of Access or pre-engagement checks are expected for local staff employed by the participating NHS organisations. Where arrangements are not already in place, network staff (or similar) undertaking any of the research activities listed in the IRAS form (except for administration of questionnaires or surveys), would be expected to obtain an honorary research contract from one NHS organisation (if university employed), followed by Letters of Access for subsequent organisations. This would be on the basis of a Research Passport (if university employed) or an

k Correspondence with Imperial College Healthcare NHS Trust for study set up post-HRA approval

Ming, Damien Keng Yen

From: MULLANEY, Scott (IMPERIAL COLLEGE HEALTHCARE NHS TRUST) <scott.mullaney@nhs.net>
Sent: 26 October 2020 13:36
To: Holmes, Alison H; Lee, Jacob; Ming, Damien Keng Yen
Cc: NGUYEN, Vinh (IMPERIAL COLLEGE HEALTHCARE NHS TRUST); TAMANG, Tara (IMPERIAL COLLEGE HEALTHCARE NHS TRUST); Nicholson, Ruth M
Subject: Restarting Clinical Research Activity at ICHT (Stage 2a): Study 19HH5646 - Minimally-Invasive Lactate in Exercise

This email from scott.mullaney@nhs.net originates from outside Imperial. Do not click on links and attachments unless you recognise the sender. If you trust the sender, add them to your safe senders list to disable email stamping for this address.

Dear Professor Holmes:

Thank you for your Stage 2a application to restart the feasibility review for: **Study 19HH5646 - Minimally-Invasive Lactate in Exercise.**

I have reviewed the application and, as you have approval from the CRF, **I am happy to give permission for this study to resume research activity pending Sponsor restart greenlight approval**.

Thank you for your application. I will document on DOCUMAS that this was granted today.

Please do come to me if there are any further queries.

Best Wishes,

Scott Mullaney



Scott Mullaney RN BSN MSc | Divisional Research Manager - Medicine & Integrated Care | Imperial College Healthcare NHS Trust
Email: scott.mullaney@nhs.net | Phone: 020 3313 3554 | www.imperial.nhs.uk | @ImperialNHS
Divisional Research Management Team | 1st Floor Admin Block South | Hammersmith Hospital | Du Cane Road | London | W12 0HS

Please note: I do not expect an answer to this email outside of your normal working hours.

****
This message may contain confidential information. If you are not the intended recipient please inform the

1

From: Gardner, Naomi C A
To: Ming, Damien Keng Yen
Cc: Holmes, Alison H; Lee, Jacob
Subject: FW: Restarting Clinical Research Activity at ICHT (Stage 2a): Study 19HH5646 - Minimally-Invasive Lactate in Exercise
Date: 26 October 2020 16:00:49
Attachments: image002.png
image003.png

Dear Damien,

Please take this e-mail as sponsor acknowledgement that we are happy for you to restart.

Please ensure that you add this email and add a file note in the TMF/file for your records.

Please let us know if there are any changes to the documentation moving forward or if you have any further queries.

Kind regards

Naomi Gardner

Naomi Gardner |Research Regulatory Facilitator | Imperial College London | Imperial College Healthcare NHS Trust
Research Governance and Integrity Office, Medical School Building , Norfolk Place, W2 1PG
Phone: 0207 594 2589 (direct line) | Email: n.gardner@imperial.ac.uk

From: Nicholson, Ruth M <r.nicholson@imperial.ac.uk>
Sent: 26 October 2020 14:40
To: Gardner, Naomi C A <n.gardner@imperial.ac.uk>
Subject: Fw: Restarting Clinical Research Activity at ICHT (Stage 2a): Study 19HH5646 - Minimally-Invasive Lactate in Exercise

Hi Naomi,

This is allocated to you.

BW

Ruth Nicholson

Head of Research Governance and Integrity

Room 215

Level 2, Medical School Building

Norfolk Place,
London W2 1PG

020759 41862

Gallagher

Insurance | Risk Management | Consulting

23rd July 2019

TO WHOM IT MAY CONCERN

We, the undersigned Insurance Brokers, hereby certify that the following described insurance:

VERIFICATION OF INSURANCE

Unique Market Reference:	B1262FI0114919
Type:	Clinical Trials Insurance
Insured:	Imperial College London
Period:	From: 1st August 2019 To: 31st July 2020 Both days inclusive at Local Standard Time.
Interest:	This Policy will indemnify/Cover the Insured for No Fault Compensation and Legal Liability for Human Clinical Trials and as more fully described within the Policy Wording
Limit of Indemnity:	GBP 15,000,000 any one Claim and in the aggregate, Legal Costs in addition
Excess:	GBP 2,500 each and every Claim other than United States of America jurisdiction where USD 25,000 each and every claim shall apply
Underwriter:	100.0000% Newline Syndicate 1218

This document is for information only and does not make the person or organisation to whom it is issued an additional Insured, nor does it modify in any manner the Contract of Insurance between the Insured and the Insurers. Any amendment, change or extension to such Contract can only be affected by specific endorsement attached thereto.

Should the above mentioned Contract of Insurance be cancelled, assigned or changed during the above policy period in such manners as to affect this document, no obligation to inform the holder of this document is accepted by the undersigned or by the Insurers. The information provided is correct at the date of signature.

[redacted]

Authorised Signatory
Arthur J Gallagher

l Ethics protocol for process mapping study at Hospital for Tropical Diseases, Vietnam

Understanding the role of decision making and clinical decision support systems in the management of acute dengue infection

Principal Investigators

Dr Sophie Yacoub, Oxford University Clinical Research Unit, Vietnam

Dr Nguyen Thanh Truong, Hospital for Tropical Diseases, Vietnam

Investigators

Dr Louise Thwaites, Oxford University Clinical Research Unit, Vietnam

Dr Ho Quang Chanh, Oxford University Clinical Research Unit, Vietnam

Dr Nguyen Quang Huy, Oxford University Clinical Research Unit, Vietnam

Dr Dong Thi Hoai Tam, Oxford University Clinical Research Unit, Vietnam

Dr Nguyen Minh Nguyet, Oxford University Clinical Research Unit, Vietnam

Dr Nguyen Thi Xuan Chau, Oxford University Clinical Research Unit, Vietnam

Dr Jennifer Van Nuil, Oxford University Clinical Research Unit, Vietnam

Luu Phuoc An, Oxford University Clinical Research Unit, Vietnam

Dr Pantelis Georgiou, Imperial College London, UK

Dr Bernard Hernandez, Imperial College London, UK

Professor Alison Holmes, Imperial College London, UK

Dr Damien Ming, Imperial College London, UK

Dr Chris Paton, University of Oxford, UK

Sponsor

The University of Oxford

CRF: Case report form

CDSS: Clinical decision support system

EPIC-IMPOC: Enhanced, personal and integrated care for infection management at point of care

EPIC-IMPOC Dengue: Enhanced, personal and integrated care for infection management at point of care in dengue

HTD: Hospital for Tropical Diseases

OUCRU: Oxford University Clinical Research Unit

Synopsis

Full title	Understanding the role of decision making and clinical decision support systems (CDSS) in the management of acute dengue infection	
Short title	Decision making and CDSS in dengue	
Study design	Qualitative observational study	
Study participants	Healthcare workers and relevant stakeholders	
Study site	Hospital for Tropical Diseases	
Sample size	Stage 1: 12-30 participants for semi-structured interviews Stage 2: 6-12 participants for each workshop session Stage 3: 50 individual usage instances of the CDSS	
Planned study period	June 2021 – June 2022	
Outcomes	**Objectives**	**Outcome measures**
1	Understand the workflow and decision-making needs of healthcare staff involved in dengue management at the HTD.	Cognitive task analysis and description of major themes from semi-structured interviews.
2	Optimise interface and functionality of the CDSS to suit local needs through interviews and feedback.	Human centred design approaches to describe needs and feedback during workshop activities.
3	Obtain feedback from healthcare workers on the usability of the final CDSS prototype.	Verbal or questionnaire feedback of usability, utility and other aspects regarding the CDSS prototype after end-user testing.

Introduction

Within low and middle-income countries (LMICs) such as Vietnam, dengue imposes significant healthcare burdens on the population. Although most symptomatic patients recover without complications, a small proportion of individuals subsequently develop severe disease. This severe form of dengue is characterized by profound haemodynamic shock, bleeding and/or organ dysfunction, and hospitalization and supportive treatment is often necessary. Effective decision-making in dengue during early illness is particularly important to optimize use of healthcare resources although this can be challenging because of variations in clinical practice and availability of specialist support in different healthcare settings [353].

Clinical decision support systems (CDSS) are tools designed to improve decision-making at the point of use. These systems exist in different formats such as paper proformas, aide memoirs but more recently include medical applications displayed through computers, tablets or smart phones [354]. Such systems have been shown to been useful in clinical care such as through improving consistency in prescribing, and clinical management. For example a CDSS tool for managing bacterial infections resulted in a higher proportion of appropriate and safe antibiotic recommendations compared with standard care in clinical evaluation [355] and another aimed at appropriate prescribing resulted in cost savings over a year of implementation [260]. Furthermore, these digital platforms also allow for and support use of data methodologies such as machine learning to be used in clinical practice. Machine learning refers to a set of algorithms and information processing techniques which are particularly suited to the analysis of large datasets such as that found in healthcare. These interventions could support decision-making by aggregating clinical information with the goal of improving risk stratification for severe disease, reducing unwanted variation in practice or improving time management of the end-user.

Within our research group we have developed a flexible digital CDSS platform and interface which can support such interventions. The project entitled, EPIC-IMPOC, “Enhanced, personal and integrated care for infection management at point of care” was developed for optimal management of bacterial infections and has been validated in prospective real-world settings in UK hospitals [355, 356]. This system is web-based and can be run with minimal hardware requirements on mobile phones, computer tablets and laptop computers through a local server. However the successful implementation of CDSS needs to be designed with the end-users in mind, integrate with existing workflows and match the context and needs for a healthcare system in Viet Nam. In this proposed study we therefore plan to understand and map the workflow of clinicians involved in the management of dengue and use this information to design a CDSS, provisionally termed: EPIC-IMPOC Dengue (Enhanced, personal and integrated care for infection management at point of care in dengue) suitable for local use at the Hospital for Tropical Diseases (HTD), Ho Chi Minh City.

This study forms part of the Flagship innovation programme at HTD/OUCRU Ho Chi Minh City aiming to provide proof-of-principle that new technologies and analytical techniques can improve the clinical care of critically ill patients. The work is done in collaboration with the new data science at OUCRU Vietnam, and the Centre for BioInspired Technology at Imperial College London.

Aims and Objectives

The aims of this research are to understand the clinical decision-making processes involved in dengue management at the HTD by healthcare workers (focusing on doctors and nurses). These findings will be used to guide and optimise design of EPIC-IMPOC Dengue, a digital CDSS system based on EPIC-IMPOC [357], with the goal of supporting clinical management of patients hospitalised with acute dengue (Appendix I).

The work will be done through a series of semi-structured interviews, workshops and feedback sessions aimed at capturing information from relevant healthcare workers and stakeholders directly involved in dengue care. In parallel there will be an iterative cycle of design improvements to EPIC-IMPOC Dengue, followed by end-user testing, analysis of these results in order to inform further design improvements and testing. This is done in order to maximise usability and impact of the final prototype of EPIC-IMPOC Dengue. Furthermore, the workshops will address issues of accountability in the use of CDSS and machine learning such as end-user understanding of the implications of risk and decisions provided by the system [358].

This research is an observational study using qualitative and human-centred design [272] methods in healthcare workers and relevant stakeholders. There is no direct patient involvement and the research will not affect clinical patient management.

This is a research study and none of the processes and outcomes will affect clinical care. The final CDSS product resulting from this research will then proceed through national regulatory approvals to ensure compliance with medical device guidelines, as well as compliance with ministry of health guidelines on the management of dengue.

Research questions:

1) What are the main sources of information used by healthcare workers and processes involved in the clinical management of dengue patients at the HTD?
2) What are the optimal designs, function and roles of a CDSS for dengue management?
3) What are the optimal designs, function and roles of a CDSS to maximise utility and uptake?
4) How could we implement EPIC-IMPOC Dengue to improve existing workflow and decision-making when managing patients with dengue attending the HTD?

***Objectives*:**

1) Understand the workflow and decision-making needs of healthcare staff involved in management of dengue patients hospitalised at the HTD.
2) Optimise interface and functionality of the CDSS to suit local needs through interviews and workshops.
3) Obtain feedback from healthcare workers on the usability of the all EPIC-IMPOC Dengue prototypes.

Study design

This is a qualitative study aimed at understanding decision-making processes in dengue management in order to optimise a CDSS, “EPIC-IMPOC Dengue”. As the study focuses on the design and implementation of EPIC-IMPOC Dengue only there is no patient involvement and no impact on clinical care in this study.

The study is divided into 3 main stages and will proceed in the following order:

Stage 1: healthcare workers and relevant stakeholders at the HTD involved in the management of patients with dengue will be invited to participate in semi-structured interviews to map out clinical decision-making. These personnel will be identified in consultation with the department heads and relevant stakeholders in the outpatient, emergency, inpatient adult and paediatric wards and intensive care departments. The aim is to enrol participants with a variety of specialties and clinical experience levels.

The interview process will be explained, and informed consent sought. Participants will be allocated a sequential study number. All work will be done outside working hours. Basic information including gender, role, and place of work will be captured on the interview. A semi-structured interview conducted individually lasting 45-60 minutes and will be conducted by the researchers. The research interview will be conducted by a social scientist and a medically-qualified doctor to ensure that the interview proceeds appropriately. The list of suggested starting questions is displayed in appendix II and has been developed and piloted by clinical members of the dengue research group at OUCRU. Similar additional questions based on responses may be added throughout the study.

Audio recordings of the interview will be captured, transcribed and translated into English by the researchers. The transcripts will be reviewed and analysed by the study team consisting of clinicians, data scientists and social science researchers. The timeline for this stage is for 1-2 interviews to be conducted each week.

All responses of the interviews will be formally reviewed by a research committee in regular debrief sessions held every 2 weeks. The committee is formed from:

- A junior doctor with less than 5 years' clinical experience
- A senior doctor with 5 or more years of clinical experience
- A member of the OUCRU social sciences research team
- A member of the dengue research group

A cognitive task analysis [4] and grounded theory approach [275] will be used to capture emerging themes which arise from the interviews. Interviews will continue until thematic saturation or up to 30 interviews, whichever occurs earlier.

Stage 2: A series of workshops will be held after completion of stage 1 in order to implement human-centred design principles and further explore the roles of decision-making and use of CDSS. These workshops will be held in Oxford University Clinical Research Unit (OUCRU) premises and will last approximately 2 hours each. Participants will be invited from a range of backgrounds including clinicians, nurses and dengue researchers in order to obtain a range of views. Contents of the workshops will include:

- Development of personas and group work to characterise typical end users of the CDSS
- Wireframe designs of the digital CDSS by the participants
- Evaluation of CDSS design and prototypes and feedback
- Discussions of clinical scenarios and implementation of the CDSS

Stage 3: Testing of EPIC-IMPOC Dengue on clinician end users. This will be done outside working hours and will not involve patients or use any patient clinical information. The aim of this stage is to evaluate the ease of navigation, utility of the system and obtain feedback for improvements. These participants will be identified by the head of departments of different specialties at the HTD.

This work will take place in clinical settings at the HTD to simulate an encounter with a patient admitted with dengue. Study participants will be asked to input artificial data generated for the purposes of this study into EPIC-IMPOC Dengue displayed on a computer tablet or laptop and asked to provide immediate feedback through an electronic questionnaire after its use. Metrics in-built to EPIC-IMPOC Dengue will record user interactions. The researcher will also observe the interaction between the participant and EPIC-IMPOC Dengue during this time using a standardised observation tool (see appendix III) and record verbal feedback. No individual identifiable data will be captured from the participant.

All of the above work is done in a manner and time, which does not interrupt or influence the workflow of the participants and their clinical responsibilities. This is also to ensure there is no impact on routine patient care at the hospital.

Inclusion/Exclusion criteria:

a) Inclusion criteria:

Participants must healthcare workers at the HTD and be involved in the management of patients with dengue.

b) **Exclusion criteria:**

Informed consent not obtained.

Endpoints and Planned Analyses:

The endpoints of this research are to characterise the main pathways and themes which underline decision-making in dengue management, use these data for CDSS development and obtain feedback of the final CDSS prototype. This will be done by compiling interviews, outcomes from workshop activities and feedback from healthcare workers. The results will also inform and enable future research into examining the need and impacts of such a system in healthcare settings.

As this is a qualitative descriptive research study, we plan to analyse interview transcripts and workshop results using a grounded theory, and interviews will continue until thematic saturation. The typically number of participants will be expected to be around 12-30 individuals for the interviews. We aim also to hold around 3 workshops with between 6-12 individuals in each. Cognitive task analysis frameworks will be applied and discussions within the research group will allow for specific tasks and priorities to be characterised.

A human-centred design approach in the workshops and testing of the digital CDSS will be utilised in order to understand the local needs. This will involve qualitative feedback on designs and ideas presented to the participants as well as multiple choice responses to clinical scenarios.

Statistical approaches will be limited to medians, interquartile ranges for responses and scoring during feedback.

Sample size

This is a qualitative study aimed at exploring the range of experiences and feedback and therefore a formal sample size calculation has not been done. For stage 1 we aim to continue interviews until thematic saturation after review by the research committee – this is typically 12-30 participants from experience. We aim to recruit 2-4 interviewees each from the outpatient department, inpatient wards, emergency department, adult intensive care and paediatric intensive care units and further

targeted recruitment will take place depending on whether thematic saturation has been reached. The interviewees will consist also of a mix of junior and senior clinicians to ensure a balanced view.

For stage 2 workshops are aimed to engage in collaborative tasks and at least 6 participants will be required for each. For observations of CDSS testing we plan to obtain at least 50 observations of CDSS use (the same participant can contribute more than one observation at different times: as the CDSS prototype undergoes constant improvement, this will provide different information at each observation).

Safety and risk assessment

This research is unlikely to pose any risks to the participants or patients.

Risks and Benefits

There is minimal risk to the participants as this is an observational qualitative study. An informed consent process will be undertaken for the potential participants listing the risk and benefits of this study. There will also be able to discuss participation with the research team who can answer questions not covered in the participant information sheet. There will be no risk to the patients admitted to the HTD given that study processes will not interfere with the workflow of the participant.

Participant Confidentiality

Confidentiality will be maintained throughout the study. All participants will be allocated a study number, but no other identifiable information will be recorded.

Ethics considerations

The Investigator will ensure that this study is conducted in accordance with the principles of the Declaration of Helsinki (Seoul 2008) and the terms of approval of the appropriate ethics committees.

The Investigator will ensure that this study is conducted in full conformity with relevant regulations and with the ICH Guidelines for Good Clinical Practice July 1996.

This protocol and all associated informed consent forms will be submitted to the ethics committee of the Hospital for Tropical Diseases and Oxford Tropical Research Ethics Committee. The Investigators will submit and, where necessary, obtain approval from the above parties for all substantial amendments to the original approved documents.

Data Management

Data collected in this research will be in the form of transcripts and their English translations, notes, responses from questionnaires and digital audio recordings. All of these data will be stored securely in a password protected server accessible only by the research team. All primary audio recordings will be deleted at the completion of the study. Participants involved in Stages 1 and 3 of the research will be identifiable

only by the allocated study number. Analysis of the results will not be identifiable by person.

Transcripts from audio recordings will be uploaded to the OUCRU-VN data server which is secure, backed up and password protected. These files will be stored on the server for 3 years from the end of the study. All other collected information relating to the study including consent form, recruitment logs and results will be stored for a minimum of 10 years in OUCRU.

Expense and benefits

Study funding will cover costs associated with workshops and group discussions. Participants will be compensated for their time at standard OUCRU rates. Those who participate in the interview will be reimbursed 200.000 VNĐ after completing the interview. Each observed use of CDSS will also be reimbursed 50.000 VNĐ.

Finance and Insurance

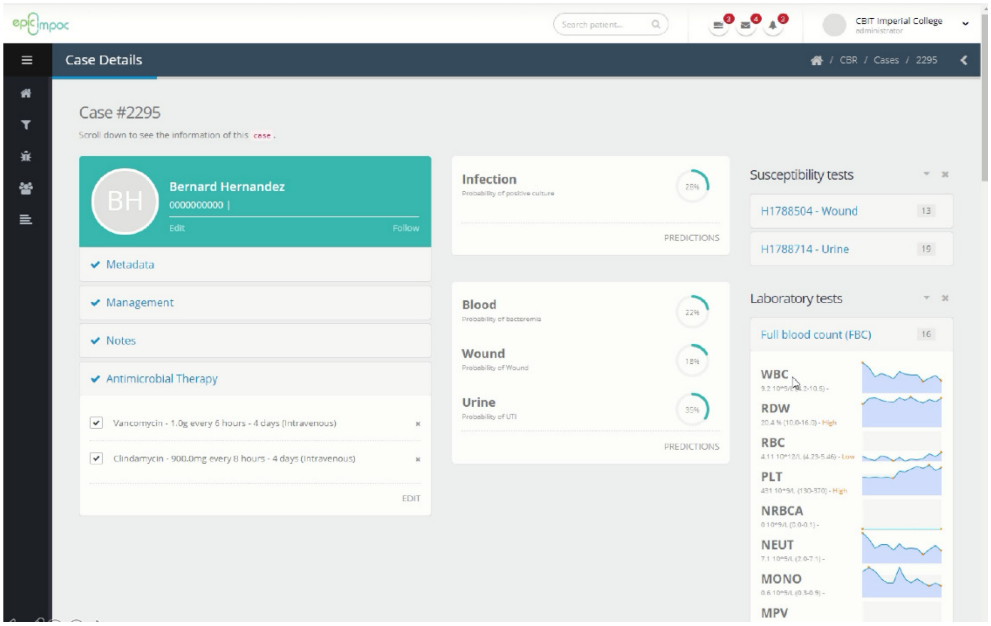
The conduct of this study is funded by the Wellcome Trust and sponsored by the University of Oxford. This research will be appropriately covered through the University of Oxford's legal liability insurances

Publication Policy

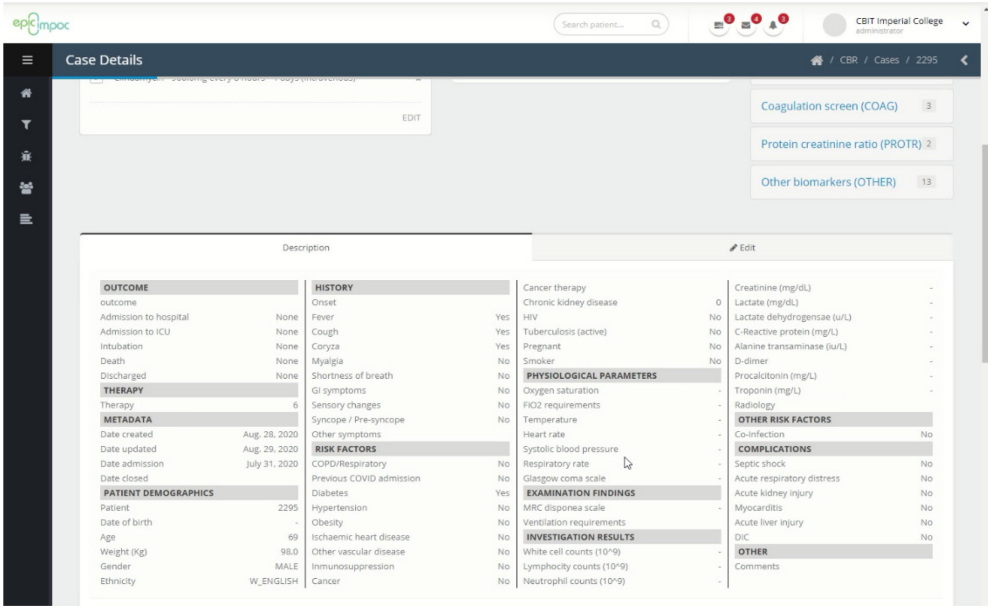
The primary outcome data will be analysed and reported in a peer-reviewed publication. The authors (and their respective positions in the author list) will be agreed prior to the start of the study in accordance with the guidelines of the International Committee of Medical Journal Editors.

Appendix I – Outline of example EPIC-IMPOC Dengue interface for testing

Interface examples of the digital CDSS system used for evaluation by the end-user – this is html web based and can be displayed on computers, mobile phones and tablets. The designs are based on existing versions of EPIC-IMPOC designed for management of bacterial infections. The final appearance of the CDSS will be subject to change to suit dengue management and take into account findings from this research study.



Screenshot 1: Initial page showing patient details, relevant information and display of result trends at a glance



Screenshot 2: Backend page showing details of previously inputted clinical data for reference and use in the machine learning inference module

Appendix II – Proposed interview guide

Interviews over 45-60 minutes

1) Recall a case of dengue you have seen recently – what do you consider as the most important clinical decisions to make in your role?

2) How do you tell dengue apart from other infectious diseases which cause fever? Is this difficult, what tools do you find useful?

- **II)** Follow up: Ask about the NS1 RDT assay and when they might disregard the results. Other blood tests – just FBC? **Describe how you would assess the severity of a patient with dengue.**
- **II)** Follow up: What sources of information do you find useful to make this decision, how does each of them affect your decision? If any of these are missed off then ask how important they are and rank them in order: [clinical symptoms, underlying diseases, laboratory investigations vital signs, POCT, history, guardian/parents, assessment, notes, response to therapy] **How do you know or decide when a patient should be admitted, discharged or referred to another service (ICU or ward)?**
- **II)** Follow up: What factors might affect how you make this decision especially with triage? Ask about: [disease severity, seasonality, time of day, staffing, time-pressures, hospital environment, uncertainty about the case, WHO warning signs, co-morbidities, age of patient, distance from the hospital] and how these affect their decision-making. **How do you define a complex dengue case? [*wait for response and then follow-up with next question*] How do you approach this?**
- **II)** Follow up: Expand on any of the following: use of guidelines, risk prediction scores/calculators, discussion with colleagues or senior staff. **How do you assess if a patient needs intravenous fluids or blood therapy? What information do you base this decision on?**
- **II)** Follow up: Expand on use of vital signs or monitoring devices to assess fluid response. **How do final decisions about management in a patient with dengue get made?**

Follow up: Expand on disagreements about optimal management e.g., between team members, and how these are resolved, role of asking for advice or permission from senior colleagues. What is the role of the patient and their representative in decision-making?

8) Is there a situation where patients with dengue will need antibiotics? Can you describe examples?

9) Can you tell me what personal or hospital factors allow for an ideal approach in dengue management?

Follow up: What other services are needed? Are there any information you would use if available for you? How do you make use of published guidelines?

10) If a computer program could take in clinical and laboratory information and accurately provide a risk of severe dengue for a patient, would it be used, and how?

Follow up: explore any uncertainties or if another setting might be more suited.

11) Is there anything else you would like to add?

Appendix III – Observation Guide

- Describe the purpose of this observation is to evaluate the performance of the clinical decision support tool (CDSS) in a clinical setting.
- Explain that by using the CDSS, the participant will be helping us improve its performance.
- Reassure the participant that this is not a test, and they are free to stop anytime.
- Explain that during the observation the researcher will not be able to provide any help or answer questions.
- Finally, ask the participant to think aloud with any impressions or feedback during the process.

Step 1

Present the clinical scenario* to the participant and ask them to read through it carefully

Step 2

Present the CDSS on the tablet device and ask them to follow the on-screen instructions

Describe the first interaction** with the CDSS _____

Record any verbal feedback during this stage _____

Step 3

Observe the data entry using the clinical scenario

Describe their interaction with the CDSS _____

Record any verbal feedback during this stage _____

Step 4

Observe how the participant interacts with the CDSS output screen

Describe their interaction with the CDSS _____

Record any verbal feedback during this stage _____

Step 5

Ask if the participant has any further feedback and/or suggestions _____

* Example of clinical scenario: You are working in the Emergency department and a 26 year-old man arrives with fever. He is alert but feels weak – his pulse is 110 bpm, blood pressure 90/60 mmHg and feels cold to touch. His rapid NS1 test is positive.

** Describe: ease-of-navigation, if the system is engaging, if there are any errors made, if the system was used efficiently and any other observations

m Ethics approval and correspondence from HTD and OXTREC for process mapping study

SỞ Y TẾ TP. HỒ CHÍ MINH
BỆNH VIỆN BỆNH NHIỆT ĐỚI

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập – Tự do – Hạnh phúc

Số : 1024 /QĐ-BVBNĐ

Thành phố Hồ Chí Minh, ngày 14 tháng 5 năm 2021

QUYẾT ĐỊNH
Về việc phê duyệt đề tài khoa học công nghệ cấp cơ sở

GIÁM ĐỐC BỆNH VIỆN BỆNH NHIỆT ĐỚI

Căn cứ Quyết định số 5178/QĐ-SYT ngày 29 tháng 7 năm 2020 của Sở Y tế thành phố Hồ Chí Minh ban hành Quy chế tổ chức và hoạt động của Bệnh viện Bệnh nhiệt đới trực thuộc Sở Y tế;

Căn cứ nội dung phiên họp ngày 28/4/2021 và ngày 13/5/2021 của Hội đồng Đạo đức trong nghiên cứu y sinh học Bệnh viện Bệnh nhiệt đới;

Xét đề nghị của Chủ tịch Hội đồng Đạo đức trong nghiên cứu y sinh học Bệnh viện Bệnh nhiệt đới.

QUYẾT ĐỊNH:

Điều 1: Nay phê duyệt đề tài nghiên cứu khoa học cấp cơ sở:

1. Tên nghiên cứu: Tìm hiểu vai trò của việc ra quyết định và hệ thống hỗ trợ ra quyết định lâm sàng trong điều trị bệnh sốt xuất huyết dengue (05NV).
   - Thuyết minh đề tài, phiên bản 1.0 ngày 06/5/2021,
   - Đề cương nghiên cứu, phiên bản 1.0 ngày 06/5/2021.
2. Chủ nhiệm đề tài: BSCKII. Nguyễn Thanh Trường, TS.BS Sophie Yacoub.
   Thành viên tham gia chính: PGS.TS Louise Thwaites, BS. Hồ Quang Chánh, PGS.TS Đông Thị Hoài Tâm, TS.BS Nguyễn Minh Nguyệt, BS. Nguyễn Thị Xuân Châu, BS. Nguyễn Quang Huy, TS. Jennifer Van Nuil, ThS. Lưu Phước An, BS. Damien Ming
3. Mã số nghiên cứu: CS/BND/21/13.
4. Địa điểm thực hiện: Bệnh viện Bệnh nhiệt đới.
5. Thời gian nghiên cứu: từ tháng 06/2021 đến tháng 06/2022.
6. Nguồn kinh phí: Dự án Hỗ trợ kỹ thuật nghiên cứu những bệnh nhiễm trùng quan trọng đối với sức khỏe cộng đồng tại Việt Nam (giai đoạn 2020-2030).

Điều 2: Các Chủ nhiệm đề tài; các Khoa/ Phòng; bộ phận có liên quan theo nội dung đề cương nghiên cứu được phê duyệt chịu trách nhiệm phối hợp thực hiện nghiên cứu đúng quy định của nhà nước về Nghiên cứu khoa học.

Điều 3: Quyết định có hiệu lực thi hành kể từ ngày ký./.

Nơi nhận:
- Chủ nhiệm đề tài;
- Lưu: VT, P.QLCL.

GIÁM ĐỐC

SỞ Y TẾ THÀNH PHỐ HỒ CHÍ MINH
BỆNH VIỆN BỆNH NHIỆT ĐỚI

TS.BS. NGUYỄN VĂN VĨNH CHÂU

Oxford Tropical Research Ethics Committee

University of Oxford
Research Services, University Offices
Wellington Square, Oxford OX1 2JD
Tel. +44 (0)1865 (2)82106
E-mail: oxtrec@admin.ox.ac.uk

Dr. Sophie Yacoub
Oxford University Clinical Research Unit
764 Vo Van Kiet, District 5
Ho Chi Minh City
Vietnam

17 June 2021

Dear Dr. Yacoub,

Full Title of Study: Understanding the role of decision making and clinical decision support systems in the management of acute dengue infection

OxTREC Reference: 531-21

Thank you for your email of the 16 June 2021, and for your minimal risk application form.

I am pleased to confirm that approval has now been granted for this study. This is valid for the first five years and is subject to receiving the local ethical approval (if this approval has not yet been received).

The documents approved for this study are as follows:

Documents:	Version:	Date:
Minimal risk application form	V2.0	14/06/21
PIS/ICF – stage 1	V1.0	14/06/21
PIS/ICF – stage 3	V1.0	14/06/21
Interview questions		

Any subsequent changes to the application must be submitted to the Committee as an Amendment. This should include a letter to give the reasons for the proposed modifications and all revised documents with changes tracked.

Please ensure that you submit a completed Annual Report form on every anniversary of this approval and a final End of Study Report. The relevant forms can be found on the OxTREC website: https://researchsupport.admin.ox.ac.uk/governance/ethics/apply/oxtrec.

Finally, please note the following **important information**:

Data safety—all studies
It is the responsibility of the PI to ensure that all data collected during the course of the study is stored and transferred safely and securely. Further guidance and advice is available from the Research Data Team.

Tel: +44 (0)1865 (2)82106
Email: oxtrec@admin.ox.ac.uk
Web: https://researchsupport.admin.ox.ac.uk/governance/ethics

Interview guide

Understanding the role of decision making and clinical decision support systems in the management of acute dengue infection:

Stage 1 – semi-structured interviews

Overview of the project

Full title	Understanding the role of decision making and clinical decision support systems (CDSS) in the management of acute dengue infection	
Short title	Decision making and CDSS in dengue	
Study design	Qualitative observational study	
Study participants	Healthcare workers and relevant stakeholders	
Study site	Hospital for Tropical Diseases	
Sample size	Stage 1: 12-30 participants for semi-structured interviews Stage 2: 6-12 participants for each workshop session Stage 3: 50 individual usage instances of the CDSS	
Planned study period	June 2021 – June 2022	
Outcomes	**Objectives**	**Outcome measures**
1	Understand the workflow and decision-making needs of healthcare staff involved in dengue management at the HTD.	Cognitive task analysis and description of major themes from semi-structured interviews.
2	Optimise interface and functionality of the CDSS to suit local needs through interviews and feedback.	Human centred design approaches to describe needs and feedback during workshop activities.
3	Obtain feedback from healthcare workers on the usability of the final CDSS prototype.	Verbal or questionnaire feedback of usability, utility and other aspects regarding the CDSS prototype after end-user testing.

Introduction

The aim of this project is to understand how dengue is managed at the HTD, in order to design a clinical decision support system for testing and future implementation.

The semi-structured interviews will describe in detail the decision-making processes (knowledge, attitudes and behaviour) adopted by clinicians when they manage a patient with acute dengue. This work will complement the background process maps and task analysis which have been completed for PICU and AICU.

Methodology

Interviews will be conducted by a member of the social sciences team with clinical support.

The estimated sample size is 12-30 participants, and recruitment to continue until thematic saturation.

To obtain a mix of different views – we aim to recruit 2-4 interviewees, with a mix of experiences from:

1) Outpatient department
2) Inpatient ward
3) Emergency department
4) Adult intensive care
5) Paediatric intensive care

Each interview should take around 45 minutes with audio responses recorded on a voice recorder, and transcribed into English by the interviewer.

The research committee will then review result transcripts every 2 weeks, or as appropriate to extract main themes in decision-making, using a cognitive task analysis approach, examining the combination of skills, strategies and knowledges applied to complex tasks [359, 360].

The committee consists of:

- Sophie Yacoub
- Chanh Ho Quang
- An Luu Phuoc
- Huy Nguyen Quang
- Bernard Hernandez
- Damien Ming

Interviews over 45-60 minutes

1) Recall a case of dengue you have seen recently – what do you consider as the most important clinical decisions to make in your role?

2) How do you tell dengue apart from other infectious diseases which cause fever? Is this difficult, what tools do you find useful?

- **II)** Follow up: Ask about the NS1 RDT assay and when they might disregard the results. Other blood tests – just FBC? **Describe how you would assess the severity of a patient with dengue.**
- **II)** Follow up: What sources of information do you find useful to make this decision, how does each of them affect your decision? If any of these are missed off then ask how important they are and rank them in order: [clinical symptoms, underlying diseases, laboratory investigations vital signs, POCT, history, guardian/parents, assessment, notes, response to therapy] **How do you know or decide when a patient should be admitted, discharged or referred to another service (ICU or ward)?**
- **II)** Follow up: What factors might affect how you make this decision especially with triage? Ask about: [disease severity, seasonality, time of day, staffing, time-pressures, hospital environment, uncertainty about the case, WHO warning signs, co-morbidities, age of patient, distance from the hospital] and how these affect their decision-making. **How do you define a complex dengue case? [*wait for response and then follow-up with next question*] How do you approach this?**
- **II)** Follow up: Expand on any of the following: use of guidelines, risk prediction scores/calculators, discussion with colleagues or senior staff. **How do you assess if a patient needs intravenous fluids or blood therapy? What information do you base this decision on?**
- **II)** Follow up: Expand on use of vital signs or monitoring devices to assess fluid response. **How do final decisions about management in a patient with dengue get made?**

Follow up: Expand on disagreements about optimal management e.g., between team members, and how these are resolved, role of asking for advice or permission from senior colleagues. What is the role of the patient and their representative in decision-making?

8) Is there a situation where patients with dengue will need antibiotics? Can you describe examples?

9) Can you tell me what personal or hospital factors allow for an ideal approach in dengue management?

Follow up: What other services are needed? Are there any information you would use if available for you? How do you make use of published guidelines?

10) If a computer program could take in clinical and laboratory information and accurately provide a risk of severe dengue for a patient, would it be used, and how?

Follow up: explore any uncertainties or if another setting might be more suited.

11) Is there anything else you would like to add?

n Workshop guides

Workshop guide for the development of a clinical decision support system in dengue

Goals of the workshops:

Workshop I

1) Facilitate discussion on clinical decision-making in dengue, based on the main themes identified from interviews.
2) Prioritise the tasks and steps which are taken during dengue clinical care in order of overall importance/impact.
3) Create a wish list of digital interventions which will be of likely utility in order to support above activities and organise them in terms of feasibility.
4) To map possible roles for these interventions by user according to dengue patient flow at the HTD.

Workshop II

1) To present wireframe designs of digital clinical decision support system (CDSS) interventions for evaluation
2) Explore different user interface approaches and understand suitability and usability
3) …. further activities to be determined after first workshop

Outline of workshop I

Timing and outline: After lunch (provided): 2 hour workshop, with a tea break in between

Venue: Conference room – social distancing with table and chair (HTD conference room)

Staff: 1-2 facilitators in addition to Imperial College team via laptop (Main: Huy)

Audience: 6-8 clinicians from HTD: participants who have taken part in the interviews previously but also head / senior nurses from wards.

Materials needed: Marker pens, post-it notes, large sheets of paper, 1-2 boards with pins, laptop computer(s) with internet access and Microsoft Teams installed, phone camera for photographing boards, lunch and refreshments.

Structure:

Part I

1) Introduction and explain purpose of the workshop: as an activity to inform design of a CDSS for dengue management – outline of activities (5 minutes) – Damien/Sophie

   Present findings from the interviews so far (10 minutes) – Damien/An/Huy

2)

   Each participant is asked to create a persona of a possible clinician user: these can be similar to people they know (or themselves), or completely a different person. Try and encourage a range of personas in the group (10 minutes total)

3)

   For each persona, have the name, picture, experience / past, their current role and personality E.g.,

Portrait and name Dr Cookie 	**Experience** Newly qualified doctor Trained in the Never seen a case of dengue!
Current role In charge of ED overnight First to see any new medical patients	**Likes / dislikes** Wants to be a surgeon Likes new technology Has a cat at home

Explain that the aim is to have a range of personalities and for each participant to imagine how their character would then view and interact with a CDSS system.

Imagine they are real people but with very fixed views and needs (maybe Dr Cookie would like a CDSS which lets him access very basic dengue information quickly).

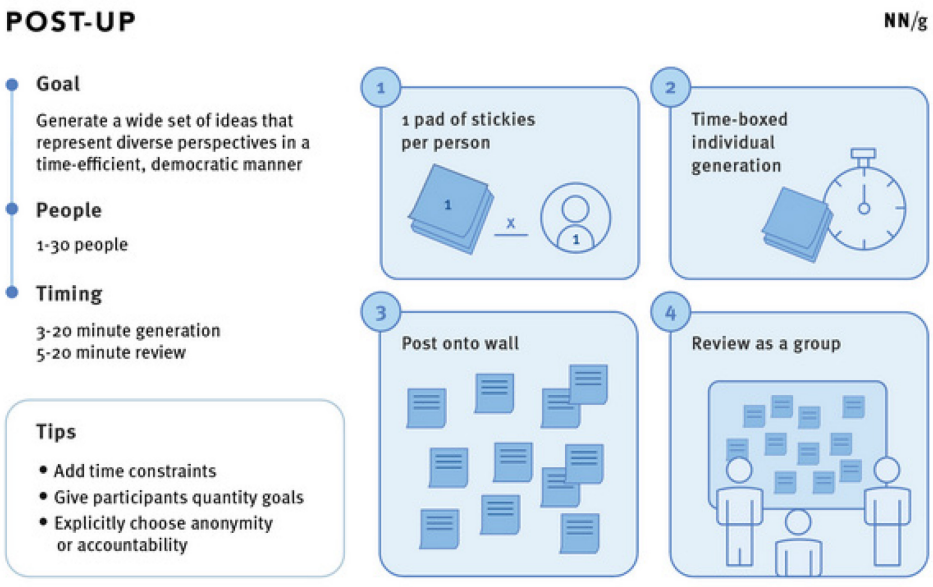
Go around and ask everyone to introduce their persona.

For all activities that follow, try and encourage responses from the participants themselves as well as ask them how their persona might respond.

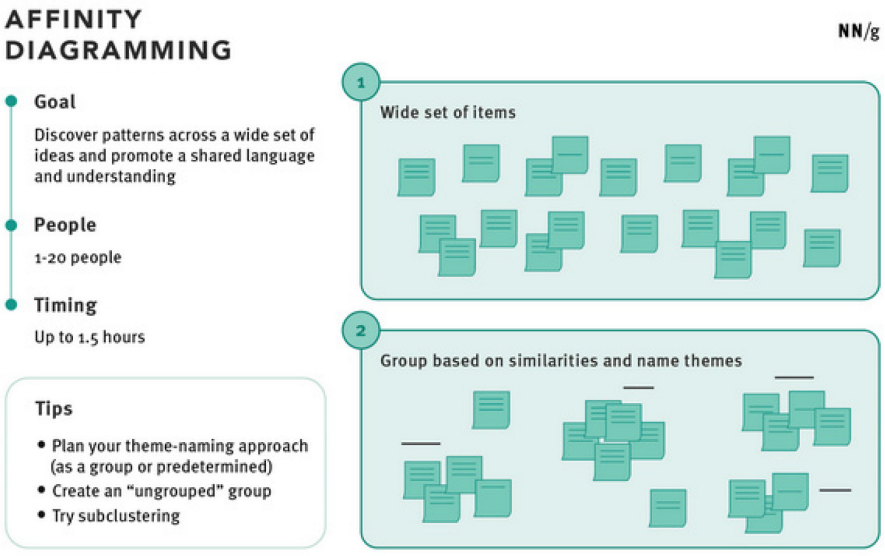
4) For each persona each participant write as many post-it notes (at least 3) on how they might find their current role – what is easy or difficult about their current role? What do they like or dislike? The goal is to generate range of ideas and find common themes specific to their persona.

There will be some overlap with responses with interviews but this is ok.

Ask for volunteer to put these notes up in the wall (5 minutes).



5) Group discussion (15 minutes total)
    a. Ask for volunteer to help:
    Can the participants cluster the post it notes to categories? Are there any common themes which emerge (e.g., workload, lack of information, difficulty accessing information, diagnostics, inexperience, lack of resources, inefficient systems, access to medical records, unpredictability of the disease….)
    b. Can all participants agree on the top 3 most important categories?



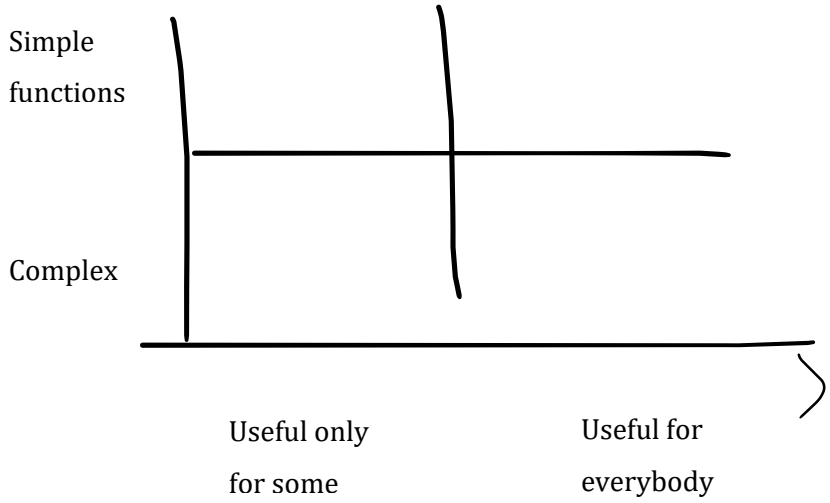
Tea break

Part II

6) Discuss personal experiences of electronic or digital interventions and technologies. Has anyone ever used a medical app on the phone, electronic medical records, online risk score calculators, electronic textbook references, pubmed searching in the hospital to help them with deciding how to manage their patients?
    a. What have they found easy or difficult about it? (10 minutes)

7) Using post-it notes describe features (make a wish-list) on what would make an electronic intervention for dengue useful for their personas or for themselves (or what would **not** be useful!)
    a. Put all responses up on the wall (5 minutes)

8) Ask a volunteer to arrange the post it notes into the diagram below – what features would be easy and useful for everybody for a CDSS (top right)? Discuss and see if everyone would agree (15 minutes)- Huy/Chanh/Co Tam

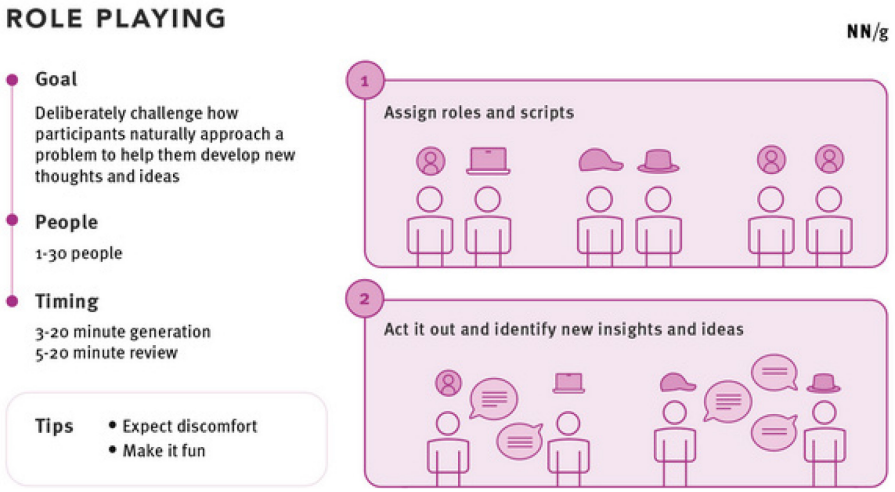


9) Ask the participants to vote and choose 3 functions to take forward for the CDSS for dengue – tell them they have now designed a CDSS!

For each persona created, choose 2-3 and role play how the persona would use the CDSS if they were back in their job – note down what their roles are, and what their likes/dislikes are.

10)

a. One participant plays the persona, and another facilitator can pretend to be the CDSS system (make it fun!). The participant is given a setting by the facilitator e.g., weekend work on ward C before Tet, or during rainy season.
b. If the system functioned exactly as the role play, would this be useful? Who would want to use it?



11) Summarise
    a. Range of personas
    b. Challenges in their roles
    c. Wish list of CDSS functions
    d. Any other thoughts to help design
    e. Explain the next step is to take these functions from the workshop and we will try and build a template functioning CDSS

End of workshop

Workshop 2 for Dengue Clinical Assessment Tool (D-CAT)

Hosted on: https://cdss.nguyentd.com/

Username: test@test.com

Password: 123456

Set up: Board, sticky notes, each participant with laptop/tablet

Functions to evaluate:

1. New patient function – data entry form
2. Patient screen – basic info, medical history, lab results, management, writing new notes
3. Machine learning risk model
4. Guidelines section
5. Calculator functionality
6. Discharge function
7. General comments on interface

Workshop

1. Outline of the CDSS (presentation) – 10 mins
2. Clinical scenario / patient data provided to each participant
    a. Ask them to enter the patient information using a unique name they made up
    b. They can use the templates or make up their own data
3. Go through each section (see feedback form) and have a discussion using patient data as a reference
    a. Feedback form to be completed during session:
        i. Appearance
        ii. Ease-of-use
        iii. Functionality
        iv. Usefulness in their clinical practice
        v. Other suggestions / changes / improvements
4. General discussion on utility and implementation
5. Discuss on next steps
6. End

Patient 1

Name	XXXXX
Date of birth	5th January 1990
Gender	Male
Admission date	Today
Height	150 cm
Weight	60 kg
Hospital No.	123456
Admitted via	Emergency department
Date of symptom onset	4 days ago
Symptoms	Fever, mucosal bleed, nausea, rash, vomiting and itching
Medical history	TB treated, diabetes
Temperature	39
Pulse	120
Blood pressure	130/80 mmHg
Examination	Cool peripheries, some oedema, no bleeding, but tender abdomen
Diagnosis	Dengue with warning signs
Tests	Positive NS1 from 2 days ago; platelet count 100; haematocrit 50%
Management	Admit to Ward Start fluids at 5 ml/kg/hour – what is the infusion rate needed?

Patient 2

Name	XXXXXXX
Date of birth	23rd April 2015
Gender	Female
Admission date	Today
Height	140 cm
Weight	39 kg
Hospital No.	2323433
Admitted via	OPD
Date of symptom onset	3 days ago
Symptoms	Nausea, skin bleeding, headache, fever
Medical history	Previous dengue
Temperature	38.9
Pulse	110
Blood pressure	100/80 mmHg
Examination	Pleural effusion on right side only
Diagnosis	Dengue with warning signs but high risk
Tests	Positive NS1 from 1 day ago; platelet count 20; haematocrit 55%
Management	Admit to Ward Start fluids for this patient at 10 ml/kg/hour What is the infusion rate needed?

Feedback form for DCAT workshop

Please fill this out during the workshop session

1. Patient entry section

Task: Enter new patient information to the system-

Do you like how it looks?

Is it easy to use?

Does everything work?

Are these functions useful to you?

Other comments:

2. Patient information screens – home page and patient information

Tasks: Look at the patient information entered

Look at the medical history section

Enter a new blood result in the lab results section – haematocrit

Enter a new note onto the notes section -

Do you like how it looks?

Is it easy to use?

Does everything work?

Are these functions useful to you?

Other comments:

3. Machine learning icon

Tasks: Look at the predicted risk of dengue shock over the next few days -

Do you like how it looks?

Is it easy to use?

Does everything work?

Are these functions useful to you?

Other comments on how to display risk of shock:

5. Guidelines

Tasks: Look at the different guidelines flowchart displayed -

Do you like how it looks?

Is it easy to use?

Does everything work?

Are these functions useful to you?

Other comments on how to display guidelines:

6. Fluid calculator

Tasks: Provides a infusion rate for adults and paediatrics based on ideal body weight or centile range (CDC)-

Do you like how it looks?

Is it easy to use?

Does everything work?

Are these functions useful to you?

Other comments on how to display information – what other calculators would you like to see?

7. Discharge function
Tasks: Discharge your patient and create a summary – save as pdf or print-

Do you like how it looks? Is it easy to use? Does everything work? Are these functions useful to you? Other comments:

Any other feedback

Workshop 2 for Dengue Clinical Assessment Tool (D-CAT)

Hosted on: https://cdss.nguyentd.com/

Username: test@test.com

Password: 123456

Set up: Board, sticky notes, each participant with laptop/tablet

Functions to evaluate:

8. New patient function – data entry form
9. Patient screen – basic info, medical history, lab results, management, writing new notes
10. Machine learning risk model
11. Guidelines section
12. Calculator functionality
13. Discharge function
14. General comments on interface

Workshop

7. Outline of the CDSS (presentation) – 10 mins
8. Clinical scenario / patient data provided to each participant
    a. Ask them to enter the patient information using a unique name they made up
    b. They can use the templates or make up their own data
9. Go through each section (see feedback form) and have a discussion using patient data as a reference
    a. Feedback form to be completed during session:
        i. Appearance
        ii. Ease-of-use
        iii. Functionality
        iv. Usefulness in their clinical practice
        v. Other suggestions / changes / improvements
10. General discussion on utility and implementation
11. Discuss on next steps
12. End

Patient 1

Name	XXXXX
Date of birth	5th January 1990
Gender	Male
Admission date	Today
Height	150 cm
Weight	60 kg
Hospital No.	123456
Admitted via	Emergency department
Date of symptom onset	4 days ago
Symptoms	Fever, mucosal bleed, nausea, rash, vomiting and itching
Medical history	TB treated, diabetes
Temperature	39
Pulse	120
Blood pressure	130/80 mmHg
Examination	Cool peripheries, some oedema, no bleeding, but tender abdomen
Diagnosis	Dengue with warning signs
Tests	Positive NS1 from 2 days ago; platelet count 100; haematocrit 50%
Management	Admit to Ward Start fluids at 5 ml/kg/hour – what is the infusion rate needed?

Patient 2

Name	XXXXXXX
Date of birth	23rd April 2015
Gender	Female
Admission date	Today
Height	140 cm
Weight	39 kg
Hospital No.	2323433
Admitted via	OPD
Date of symptom onset	3 days ago
Symptoms	Nausea, skin bleeding, headache, fever
Medical history	Previous dengue
Temperature	38.9
Pulse	110
Blood pressure	100/80 mmHg
Examination	Pleural effusion on right side only
Diagnosis	Dengue with warning signs but high risk
Tests	Positive NS1 from 1 day ago; platelet count 20; haematocrit 55%
Management	Admit to Ward Start fluids for this patient at 10 ml/kg/hour What is the infusion rate needed?

Feedback form for DCAT workshop

Please fill this out during the workshop session

4. Patient entry section

Task: Enter new patient information to the system-

Do you like how it looks?

Is it easy to use?

Does everything work?

Are these functions useful to you?

Other comments:

5. Patient information screens – home page and patient information

Tasks: Look at the patient information entered

Look at the medical history section

Enter a new blood result in the lab results section – haematocrit

Enter a new note onto the notes section -

Do you like how it looks?

Is it easy to use?

Does everything work?

Are these functions useful to you?

Other comments:

6. Machine learning icon

Tasks: Look at the predicted risk of dengue shock over the next few days -

Do you like how it looks?

Is it easy to use?

Does everything work?

Are these functions useful to you?

Other comments on how to display risk of shock:

8. Guidelines

Tasks: Look at the different guidelines flowchart displayed -

Do you like how it looks?

Is it easy to use?

Does everything work?

Are these functions useful to you?

Other comments on how to display guidelines:

9. Fluid calculator

Tasks: Provides a infusion rate for adults and paediatrics based on ideal body weight or centile range (CDC)-

Do you like how it looks?

Is it easy to use?

Does everything work?

Are these functions useful to you?

Other comments on how to display information – what other calculators would you like to see?

10. Discharge function

Tasks: Discharge your patient and create a summary – save as pdf or print-

Do you like how it looks?

Is it easy to use?

Does everything work?

Are these functions useful to you?

Other comments:

Any other feedback

II Supplementary data and results

a Prediction of bacterial bloodstream infections

LSTM models for the rolling window approach were shortened slightly from a 7-day to 3-day window. This increases the class imbalance by increasing the number of negative windows (positive class between 7-day vs 3-day windows around 6.5% vs 4.7%). For the temporally ordered window approach, the performances on hold out test set compared with the unordered windows were AUROC 0.84 vs 0.88, and AUPRC 0.21 vs 0.46 respectively for the 3-day window approach.

b Dengue risk for Hospitalised Patients

Prospective clinical studies conducted in Vietnam by Oxford University Clinical Research Unit conducted between 1^{2t}h April 2001 and 3^{0t}h January 2018 were used to derive the final dataset used for analyses. A summary of included studies are included in table 1:

No.	Study identifier code	Recruitment dates	Enrolled patients (n)	Title of study
1	Study 1[198]	12/4/2001- 24/7/2009	3,042	Inpatient-based study examining prognostic factors during the febrile phase
3	Study 2[199]	3/8/2009 – 8/12/2010	225	Randomised control trial comparing use of steroids versus placebo in acute dengue
4	Study 3	08/12/2010 – 16/6/2011	88	Prospective study in acute dengue
5	Study 4[192, 200]	19/10/2010 – 4/12/2014	8,100	Outpatient-based study examining prognostic factors during the febrile phase
6	Study 5	20/10/2016 – 30/1/2018	665	Study of acute dengue in pregnancy during the febrile phase

i Details of individual studies

The following clinical studies and their data were aggregated and used for machine learning development and evaluation. Five studies were included based on their similarities in study design, dengue definitions and recorded outcomes.

1. **"Inpatient-based study examining prognostic factors during the febrile phase" (Study 1) – https://doi.org/10.1371%2Fjournal.pntd.0005498**

Recruitment details

This was a prospective inpatient observational study recruiting children between 5 and 15 years admitted to the paediatric dengue ward at the HTD, Ho Chi Minh City between 12/4/2001- 24/7/2009. A total of 3,042 patients were recruited to the study with confirmed dengue and 2,615 patients were included in the analyses after filtering. Admitted patients were followed up daily until discharge.

Participants who are admitted to the paediatric ward at the HTD are those with uncomplicated illness only and not suitable for outpatient care. Patients with severe dengue or require frequent (4-6 hourly) monitoring are admitted or transferred to the paediatric intensive care unit.

Dengue diagnosis

Diagnosis of dengue is confirmed by detection of dengue virus (DENV) RNA in plasma by reverse transcriptase polymerase chain reaction (RT-PCR) at enrolment, or by seroconversion on IgM and IgG capture ELISA on paired enrolment and early convalescent specimens (Dengue Duo IgM and IgG Capture ELISA, PanBio, Australia) or in-house methods.

Outcomes

The development of shock– the most common manifestation of severe dengue in this population was the primary endpoint. This was defined as narrow pulse pressure (≤ 20 mmHg) or hypotension for age with evidence of impaired peripheral perfusion.

2. **"Randomised control trial comparing use of steroids versus placebo in acute dengue" (Study 2)- https://doi.org/10.1093/cid/cis655**

Recruitment details

This was a randomised placebo-controlled clinical trial examining the effects of low or high dose oral prednisolone in patients aged 5-20 years admitted to the HTD. In total, 225 participants were enrolled – these patients had a fever onset of ≤ 72 hours, did not present with signs of dengue-related complications, did not have past medical history or regular medications and had a confirmed diagnosis of dengue. Of this group, the 74 patients in the placebo group were included in the analyses.

Admission criteria for adults is based on HTD hospital guidelines but are in general consistent with warning signs defined by the World Health Organization 2009 dengue guidelines [11] suggestive of an elevated risk of progression to severe dengue. These include clinical signs such as persistent vomiting, abdominal pain or tenderness, skin or mucosal bleeding or signs of fluid leakage but also includes patients who have comorbidities or would benefit from closer clinical monitoring, at the discretion of the clinical team.

Dengue diagnosis

Diagnosis was supported through a positive rapid test for dengue non-structural protein 1 (NS1 Ag-STRIP, Bio-Rad) and confirmed through either a DENV RT-PCR at enrolment or a capture IgM and IgM using paired specimens (Venture Technologies, Sarawak, Malaysia) within 72 hours of illness onset and at day 7.

Outcomes

Development of dengue shock syndrome according to WHO 2009 dengue guidelines, and severe clinical bleeding as evaluated by clinician were used as outcome definitions.

3. "Prospective study in acute dengue" (Study 3) – unpublished data

Recruitment details

This was a prospective inpatient observational study recruiting patients aged 12-25 admitted to the HTD with early acute dengue (within 72 hours of illness onset) between 20108/12/2010 – 16/6/2011. Admitted patients were followed up daily until discharge. In total 88 patients were enrolled of which 47 were included in the analyses after filtering. Admission criteria to the HTD is mentioned above.

Dengue diagnosis

Diagnosis of dengue is confirmed by detection of DENV-RNA in plasma by RT-PCR at enrolment or by IgM or IgG seroconversion using methods described above

Outcomes

Vascular leak and severe bleeding according to WHO 2009 dengue guidelines.

4. "Outpatient-based study examining prognostic factors during the febrile phase" (Study 4)-- https://doi.org/10.1093/cid/ciw863

Recruitment details

This was a prospective community-based observational study recruiting patients presenting to outpatient departments at three large hospitals in Ho Chi Minh City, Vietnam – these include the HTD, Children's Hospital 1 and 2. Children between 1-15 years of age attending the outpatient with a fever 3 days or less in whom dengue is a possible diagnosis was eligible for enrolment. The exclusion criteria included cases where an alternative diagnosis is more likely or if the participant would be unlikely to be able to attend follow up. In total 8,100 participants were enrolled in the study of which 2,245 had confirmed dengue and 1,010 patients who were hospitalised were included in the analyses.

Diagnosis

A diagnosis of dengue was confirmed through one of the following: positive RT- PCR for DENV, a positive NS1 assay (Platelia NS1 antigen or NS1 Ag STRIP, Bio-Rad, France) at enrolment or detection of IgM seroconversion.

Outcomes

Dengue shock syndrome, dengue with severe bleeding, or dengue with end-organ involvement (central nervous dysfunction, hepatic dysfunction or severe respiratory dysfunction or other major organ involvement) constituted study endpoints.

5. "Study of acute dengue in pregnancy during the febrile phase" (Study 5) – unpublished data

Recruitment details

This was a prospective observational study recruiting women with acute dengue during pregnancy matched with non-pregnant women controls with acute dengue at the HTD and Tu Du Hospital in Ho Chi Minh City. Women between 18 and 45 years old referred to the HTD with a fever and clinical symptoms consistent with dengue and who were either pregnant or not during the recruitment period were eligible. In total 665 patients with dengue were recruited of which 456 were included in the analyses after filtering.

Diagnosis

Diagnosis of dengue is confirmed by detection of DENV-RNA in plasma by RT-PCR, NS1 antigen detection at enrolment or by IgM or IgG seroconversion, or a positive using methods described above.

Outcomes

Patient outcomes are classified according to WHO 2009 dengue guidelines which include development of DSS, significant clinical bleeding or organ involvement.

Grid search parameters for machine learning models used for hyperparameter optimisation

Model development and validation

Five models were chosen for evaluation, these include xGboost[47] , random forest classifier[203], logistic regression, artificial neural networks[361] and support vector machines[362]. A linear implementation of logistic regression only as a baseline model, acknowledging non-linear relationships exist with regards to haematocrit and platelet count with outcome (see below).

A repeat stratified 10-fold cross validation strategy, repeated 3 times was adopted, and hyperparameter tuning done through grid search.

Models from each algorithm were tested through cross-validation with the area under the receiver operating curve (AUROC) as the primary scoring method. Isotonic calibration was subsequently applied to the models[363].

The final models and their respective hyperparameter variables trained on the development and their hyperparameters are as follows:

Extreme gradient boosting (XGBoost) ‘'et'': [0.05,0.1,0.2],‘'min_child_weigh'': [0.05,0.1,0.2],‘'max_dept'': [4,5,6],‘'gamm'': [0.1,0.2,0.3],‘'n_estimator'':[25, 50,75, 150, 300]

Optimal parameters: (min_child_weigh'': 0.001‘'et'': 0.01,‘'gamm'': 0.2,‘'max_dept'': 4,‘'n_estimator'': 300)

Random forest classifier:‘'criterio'': ‘'gin'’'entrop'’],‘'n_estimator'': [250, 500, 1000, 1500, 2000],‘'min_samples_spli'': [8, 12, 14, 16],‘'max_dept'': [5, 8, 10, 20, 40, None]

Optimal parameters: (criterion''entrop'', max_depth=8, min_samples_split=8, n_estimators=500, random_state=42)

a) Logistic regression

‘solve'': ‘'liblinea''],
‘penalt'': ‘'LASS'',‘'Ridge regularisatio''],‘'': [0.01, 0.1, 1, 10, 100]

Optimal parameters: (C=0.01, penalty''LASS'', random_state=42, solver''liblinea'')

b) Artificial neural networks

‘solve'': ‘'sg'’'ada''],
‘activatio'': ‘'rel''],‘'hidden_layer_size'': [(100,),(150,),(200,100)],‘'alph'': [0.0001,0.001,0.05],
‘'max_ite'': [200,400,600]

Optimal parameters: (hidden_layer_sizes=(200, 100), max_iter=50, solver''adam)

c) Support vector machines

kernel: [rbf],
‘'':[0.001,0.01,0.1],‘'gamm'':[0.01,0.1,1],‘'probabilit'':[True]

Optimal parameters: (kernel:rbf, C=0.1, gamma=0.1, probability=True)

As a probabilistic classification was given, the optimal cut-off was determined using the maximum J-statistic, in order to maximise sensitivity and specificity and the default cut off of >0.50 not implemented. The cut-off values for each of the models are as follows:

Model	**Optimal cut-off**
XGBoost	0.090229
Random forest classifier:	0.086768
Logistic regression	0.123531
Artificial neural networks	0.083191
Support vector machines	0.081361

Table 2. Cut-off thresholds for probabilistic classifier. A value above the threshold would be classified as positive (experiencing complication) and below as negative (no complication).

The performances of the models for each of the algorithms from cross validation of the development set as well as against the hold out test set is shown in the main text.

Bootstrapping

In order to derive the confidence intervals for the performance metrics including AUROC, sensitivity, specificity, negative predictive value and positive predictive value, we performed bootstrapping[58] of the development set in order to generate, 1,000 different datasets. Data were resampled with replacement at random through the bootstrapping procedure and underwent repeated testing with 90% of the bootstrapped dataset used to train the final model, and tested against the remaining 10%. The distribution of each of the performance metrics was plotted and assumed to be gaussian in distribution (figure 3), and the 95% interval was taken as the cut offs values at the 2.5 and 97.5 percentile of these values.

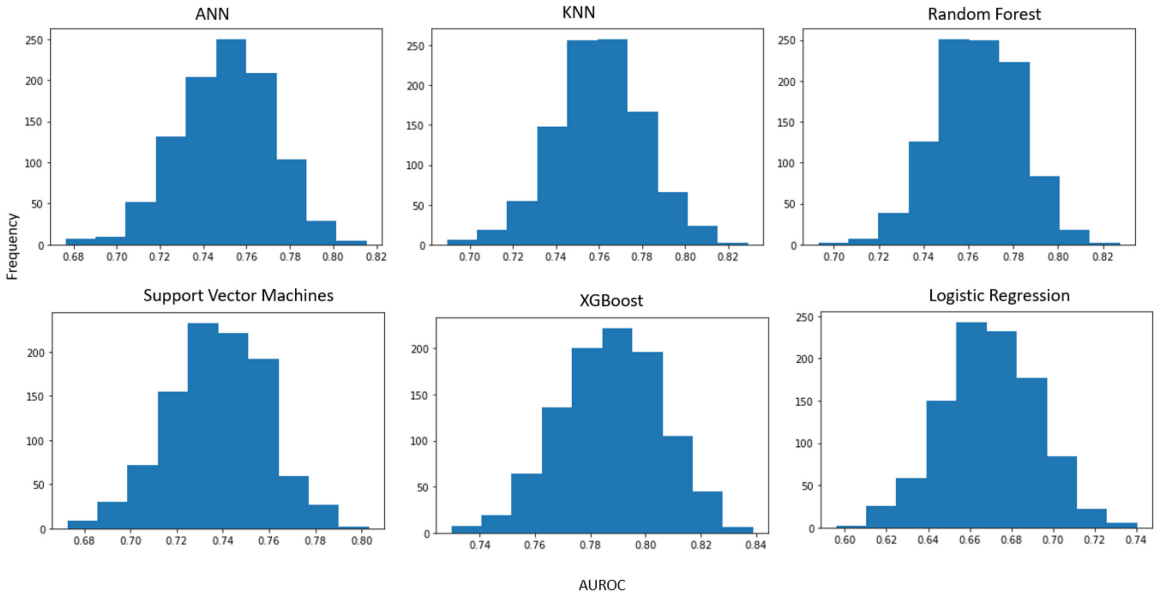


Figure 3. Histograms of AUROC distributions obtained from bootstrapping

ii Full results for model evaluation in prediction of severe dengue outcomes

Composite dengue outcomes were extracted from individual studies to provide classification which better captured the range of severe outcomes seen in the disease, beyond dengue shock.

The following outcomes extracted from each of the studies. These were based on definitions of moderate and severe dengue definition, defined as one of:

- Dengue shock syndrome: rapid and weak pulse and narrow pulse pressure (< 20 mmHg) or hypotension for age, and clinical signs of hypotension.
- Plasma leakage: dengue shock syndrome, clinical fluid accumulation in chest or abdomen or signs of respiratory compromise requiring treatment.
- Bleeding: bleeding deemed severe by clinical assessment and/or that requiring medical treatment such as nasal packing, blood product transfusion.

Limitations to these outcomes are that variables are not timestamped in their original dataset. Timings of onset cannot be inferred and the model examines association rather than being a predictive model per se (see main text in thesis).

The following specific outcomes were recorded from each of the studies:

Study	**Criteria**	**Outcomes captured**

Study 1	Shock	Dengue shock syndrome
	Plasma leakage	Fluid overload by clinical assessment
	Bleeding	Dengue haemorrhagic fever grades III and IV
Study 2	Shock	Dengue shock syndrome
	Plasma leakage	One of: presence of ascites, need for ventilation, oxygen therapy, diuresis or inotropic support
	Bleeding	One of: Severe mucosal or skin bleeding, melaena or haematemesis
Study 3	Shock	Dengue shock syndrome
	Plasma leakage	One of: Fluid overload, need for ventilation, oxygen therapy or diuresis
	Bleeding	Severe bleeding on clinical assessment
Study 4	Shock	Dengue shock syndrome
	Plasma leakage	One of: Respiratory distress, pleural effusion, ascites or pulmonary oedema
	Bleeding	Severe bleeding on clinical assessment
Study 5	Shock	Dengue shock syndrome
	Plasma leakage	One of: Respiratory distress, pleural effusion, ascites or pulmonary oedema
	Bleeding	One of: Severe mucosal or skin bleeding, melaena, haematemesis or haematuria

The performance of models trained using a training set and 10-fold cross validation, and evaluated against hold out test set are shown below. Overall performance was weaker than the main model against dengue shock syndrome, and one possibility is that this is reflective of the heterogeneity in the outcome classification.

Model	**Mean model AUROC**	**Model Specificity**	**Model Sensitivity**	**Model Positive Predictive value**	**Model Negative Predictive value**	**Brier score**	**Hold out test set (AUROC)**

XGBoost	0.759 (0.731-0.805)	0.75 (0.632-0.873)	0.668 (0.54-0.782)	0.198 (0.167-0.308)	0.961 (0.946-0.97)	0.068	0.773
Random forest	0.766 (0.712-0.787)	0.668 (0.606-0.857)	0.745 (0.523-0.789)	0.172 (0.144-0.266)	0.966 (0.947-0.972)	0.068	0.782
Logistic regression	0.701 (0.663-0.74)	0.627 (0.485-0.838)	0.689 (0.445-0.815)	0.145 (0.119-0.208)	0.956 (0.94-0.97)	0.075	0.714
Artificial neural networks	0.758 (0.714-0.791)	0.744 (0.592-0.856)	0.65 (0.512-0.805)	0.19 (0.143-0.259)	0.958 (0.948-0.973)	0.068	0.782
Support vector machines	0.66 (0.69-0.73)	0.654 (0.537-0.854)	0.45 (0.775-0.782)	0.15 (0.127-0.233)	0.954 (0.939-0.966)	0.070	0.738

iii Static models

Static, time-invariant models were developed in order to evaluate the role of including summarised feature data taken over the 48-hour period. These models included only the first feature value for each patient.

The development of dengue shock syndrome was kept as the outcome for prediction. All procedures including grid search for each of the algorithms were unchanged with the use of 10-fold cross validation and AUROC as the metric to select for the best performing model configurations.

Results are summarised as follows, and discussed in the main text:

Algorithm	**Model AUROC in 10-fold cross validation**	**Model specificity**	**Model sensitivity**	**AUROC evaluation against hold out test set**
XGBoost	0.675	0.540	0.818	0.733
Random forest	0.676	0.757	0.600	0.733
Logistic regression	0.687	0.518	0764	0.764
Artificial neural networks	0.675	0.639	0.727	0.727

Support vector machines	0.593	0.487	0.655	0.655

c Seasonality and time on clinical management of acute febrile illnesses

i Optimal model configuration

The optimal XGBoost model configuration after hyperparameter grid search process used was:

XGBClassifier(eta= 0.1, gamma= 0.1, max_depth= 3,min_child_weight= 0.2, n_estimators= 50, eval_metric''erro'',enable_categorical=True,verbosity = 0)

ii Window size

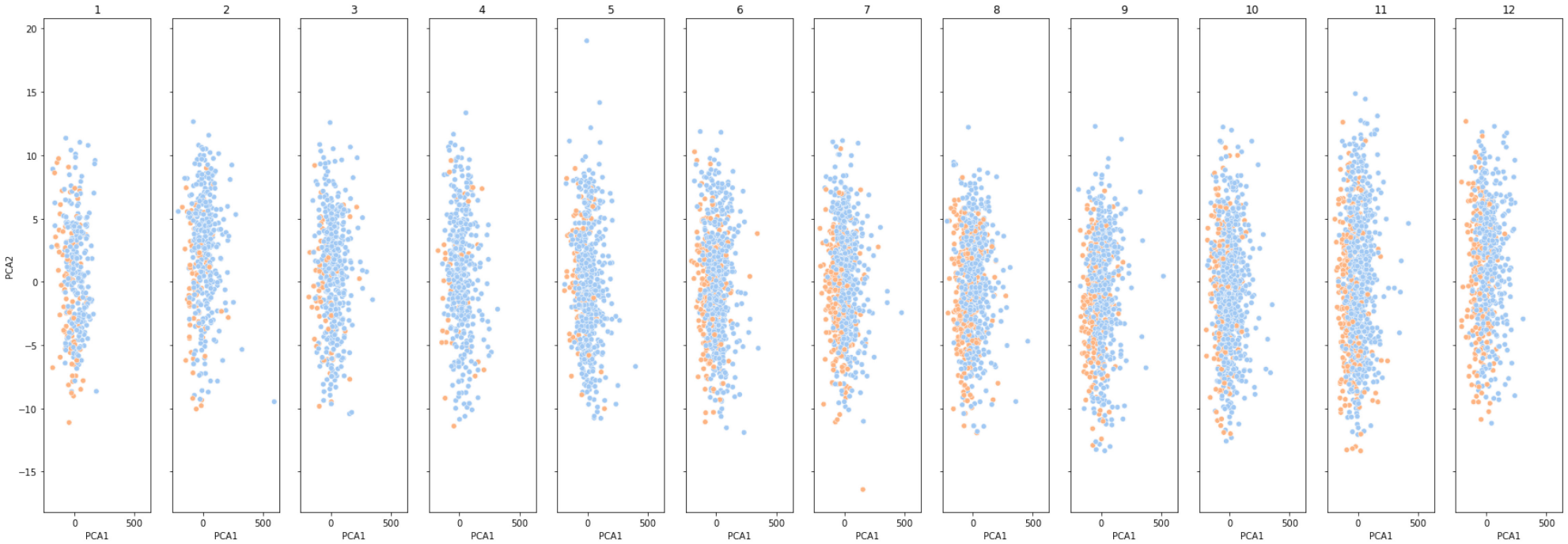
For the empirical threshold tuning approach, sensitivity analyses were performed to evaluate the effect of using different window sizes on performance.

Predictions thresholds were tuned according to the optimal threshold achieved for the patients included in a particular window of either 7, 14, 21 and 28 days. Similar results were achieved:

Window size	Specificity (IQR)	Sensitivity	PPV	NPV	Threshold
7	0.87 (0.65-1.00)	0.80 (0.60-0.91)	0.67 (0.50-0.83)	0.91 (0.90-0.93)	0.36 (0.16-0.62)
14	0.85 (0.68-0.97)	0.77 (0.57-0.88)	0.64 (0.53-0.79)	0.91 (0.90-0.92)	0.35 (0.17-0.57)
21	0.85 (0.68-0.96)	0.76 (0.56-0.87)	0.63 (0.54-0.76)	0.90 (0.90-0.91)	0.35 (0.17-0.54)
28	0.85 (0.68-0.95)	0.75 (0.57-0.86)	0.62 (0.55-0.75)	0.90 (0.90-0.91)	0.34 (0.18-0.53)

iii Clustering approaches for other febrile illnesses

An unsupervised approach was used to investigate if any patterns in the OFI group would emerge when stratified according to month of year. Principal component analysis was used to reduce feature variables (age, sex, haematocrit, platelet, white cell, and lymphocyte count on clinical presentation) into 2-dimensional values.

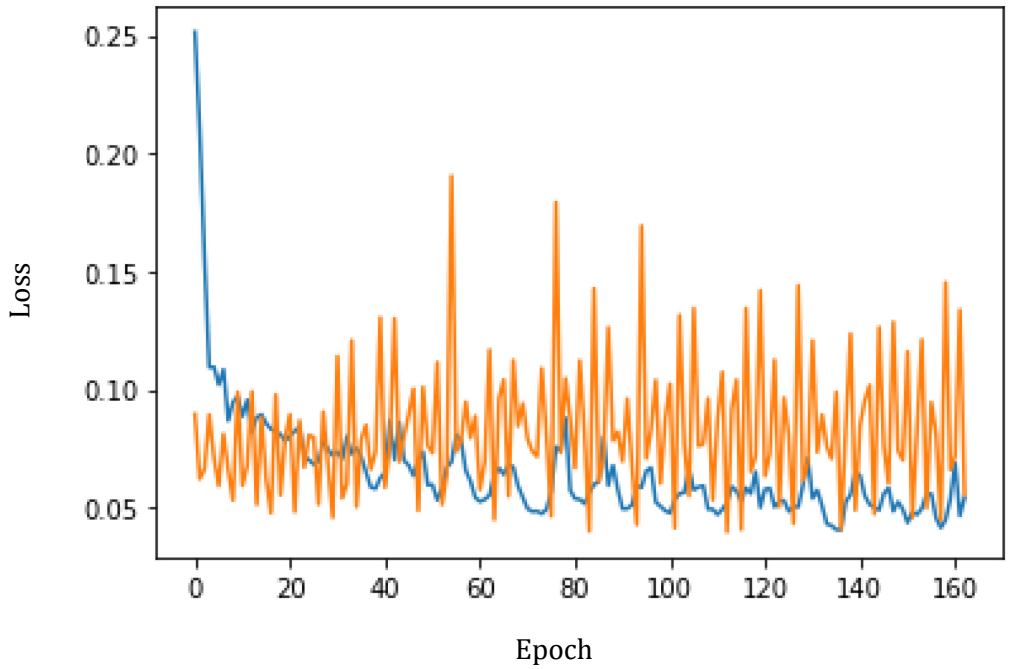


Two dimensional PCA plots for dengue (orange) and OFI (blue) across months 1-12 shown as the title.

iv Deep learning parameters

A stacked long short-term memory model was used for the prediction of optimal thresholds to achieve a constant negative predictive value. The model comprised of two layers of LSTM (size 20 neurons), a dropout of 0.2 in between and a output neuron to predict threshold. Mean absolute error was used as the loss metric and adam as the optimiser.

The learning curves until epoch 160 are as follows, with an early stopping strategy implemented to derive the final model used.



d Deep learning for BSI

i Correlation between parameters

A Pearson correlation heatmap was plotted using feature variables at day 0 (day of blood culture acquisition) to examine for collinearity. In general, there were no overt relationships between features apart for haemoglobin-related features (Figure II.1).

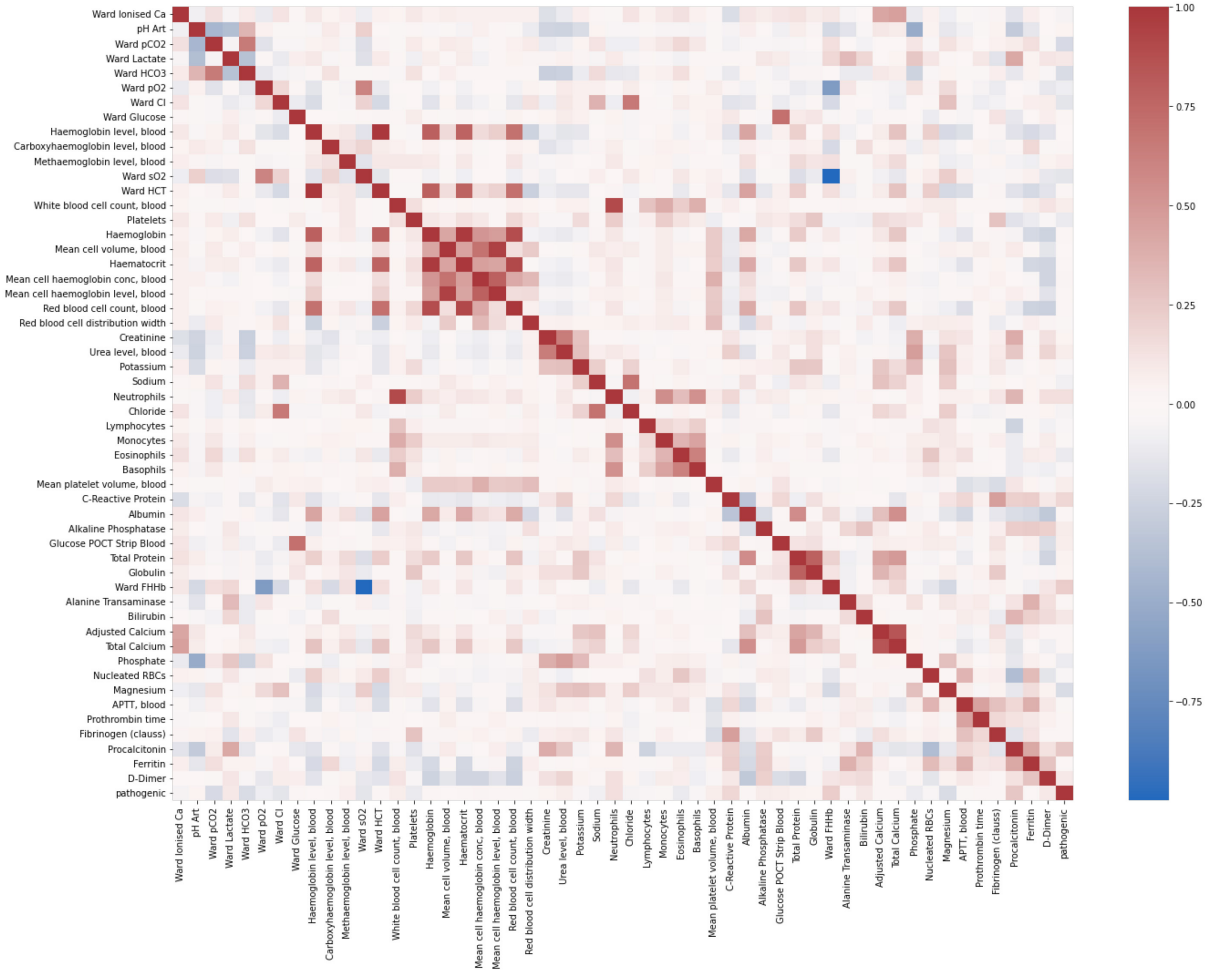


Figure II.1 Pearson correlation heatmap with colour intensity signifying strength of correlation of values measured at the timepoint of blood culture acquisition for all patients. Red values show positive correlation and blue negative correlation.

ii Data completeness and differences

For final feature variables, the amount of missing data on the date of blood culture acquisition was plotted against different outcomes. Heatmaps denote the relative availability of features for inclusion in models. Red colours signify high availability of the feature and blue colours low availability. Overall, data availability was similar across the two different outcomes (pathogenic BSI vs non-BSI) but there were some differences

when classed according to species. Note that for ward point of care tests including blood gas analyses, these are frequently performed more than once a day which could explain their representation in the heatmaps.

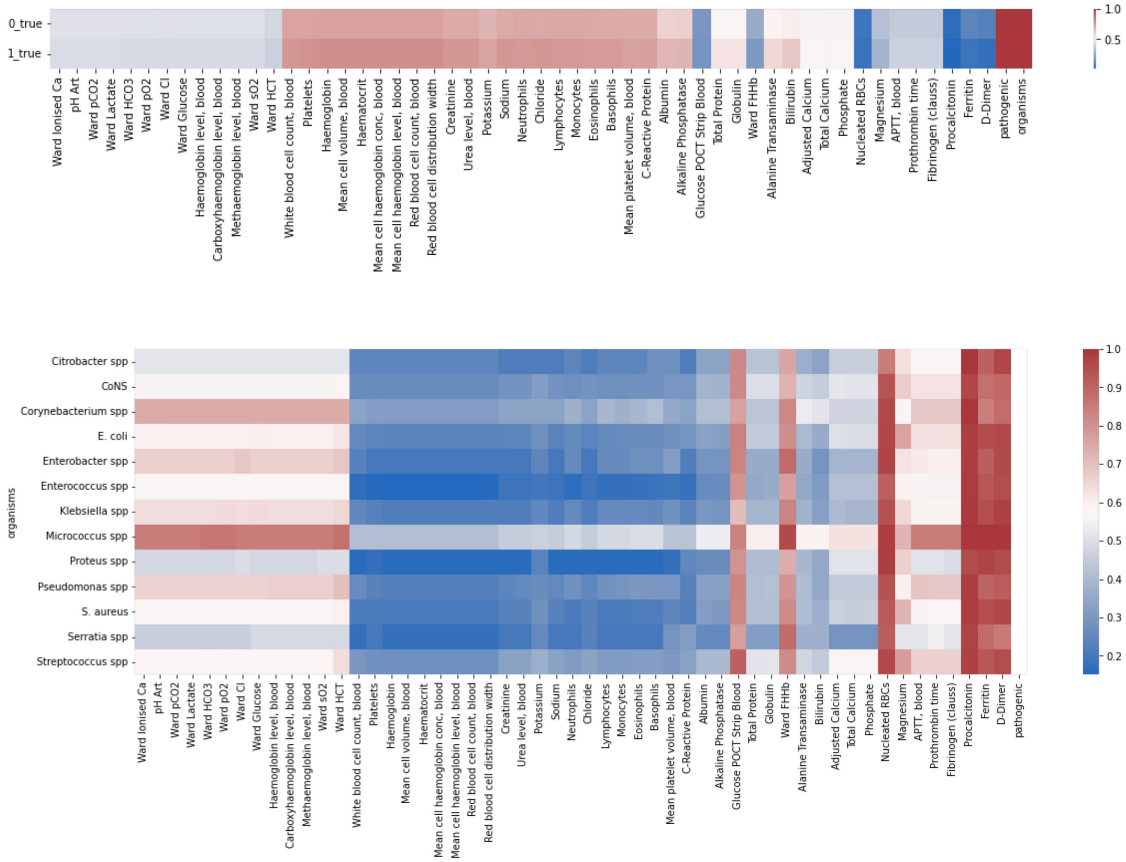


Figure II.2 Heatmaps showing data availability with red representing higher availability of data on the day of blood culture acquisition. The top heatmap is stratified between 1: pathogenic BSI and 0: non-BSI. Bottom heatmap is stratified according to final bacterial species isolated from blood cultures.

iii Models used in the chapter

Initial biases for model output layers were set using the log proportion of pathogenic BSI/ non-BSI class to account for class imbalance [364].

For optimisation of the model architecture, an iterative approach was used through KerasTuner. The library iteratively examines different hyperparameters in stages through Bayesian optimisation process.

The configurable parameters were as follows:

Parameter	Configuration
Layer 1	LSTM, size 8 to 512
Dropout 1	Rate 0.1 to 0.5
BatchNormalisation 1	Add or omit
Layer 2	LSTM, size 8 to 512
Dropout 2	Rate 0.1 to 0.5
BatchNormalisation 2	Add or omit
Layer 3	Dense 8 to 128
Output	Initial bias or not

For the rolling window models the final neural network configuration obtained by KerasTuner was as follows: stacked LSTM layers of size 10, with a 0.2 dropout and BatchNormalisation layer in between. Binary Crossentropy was used as the loss function and the final layer was used a sigmoid activation function for binary classification. This is shown in figure 3.3:

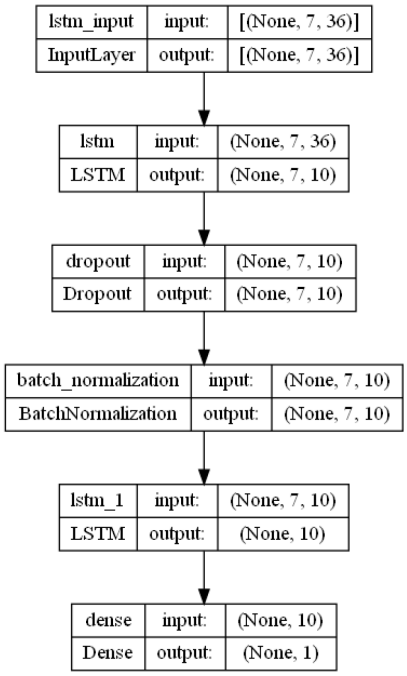


Figure II.3 LSTM model optimised for rolling window predictions consisting of two LSTM layers, with a Dropout and BatchNormalisation layer.

iv A regression approach to lactate prediction

Feature importance and interpretability analysis for the model supports the role of lactate in supporting predictions of BSI.

This sub-analysis examines whether lactate concentrations can be predicted through a data-driven approach using routine blood biomarker values alone. The null hypothesis being that if lactate concentrations can be accounted for through existing data then it is unlikely that its measurement provides significant added information.

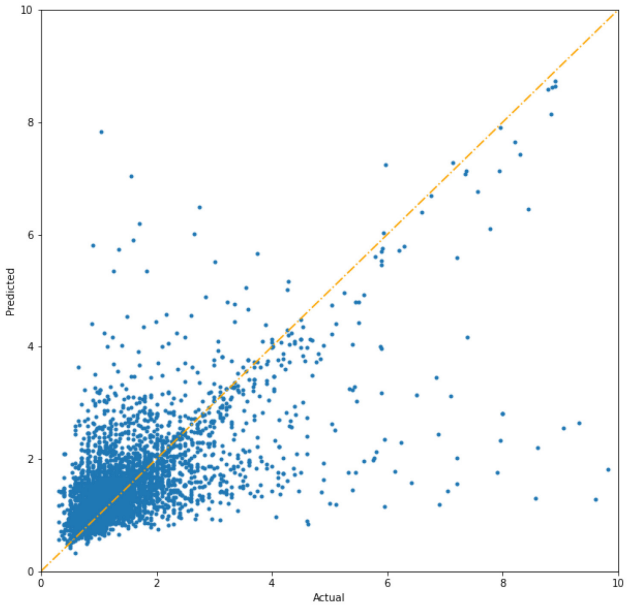
An XGBoost (see chapters 4 and 5) algorithm was used using the BSI dataset as described. Features included were routine haematological, biochemical and clotting biomarkers listed below:

> ['Haemoglobin level, blood', 'White blood cell count, blood', 'Platelets', 'Haemoglobin', 'Mean cell volume, blood', 'Haematocrit', 'Mean cell haemoglobin conc, blood', 'Mean cell haemoglobin level, blood', 'Red blood cell count, blood', 'Red blood cell distribution width', 'Creatinine', 'Urea level, blood', 'Potassium', 'Sodium', 'Neutrophils', 'Chloride', 'Lymphocytes', 'Monocytes', 'Eosinophils', 'Basophils', 'Mean platelet volume, blood', 'C-Reactive Protein', 'Albumin', 'Alkaline Phosphatase', 'Glucose POCT Strip Blood', 'Total Protein', 'Globulin', 'Alanine Transaminase', 'Bilirubin', 'Adjusted Calcium', 'Total Calcium',

'Phosphate', 'Nucleated RBCs', 'Magnesium', 'APTT, blood', 'Prothrombin time', 'Fibrinogen (clauss)', 'Procalcitonin', 'Ferritin', 'D-Dimer']

A total of 20,072 lactate values were available for analysis (median 1.23, IQR 0.91 to 1.76 mmol/L) with a regression approach used with hyperparameter grid search. The mean squared error (MSE) metric was used as loss function.

The mean MSE score for the tuned model to predict lactate levels was 1.09 mmol/L in 5-fold cross validation. A depiction of actual vs predicted values in this model, and Bland-



Altman analysis shown as follows:

Figure II.4 Actual (x-axis) vs predicted (y-axis) lactate values in the optimised XGBoost model for all predictions, R (correlation) = 0.64

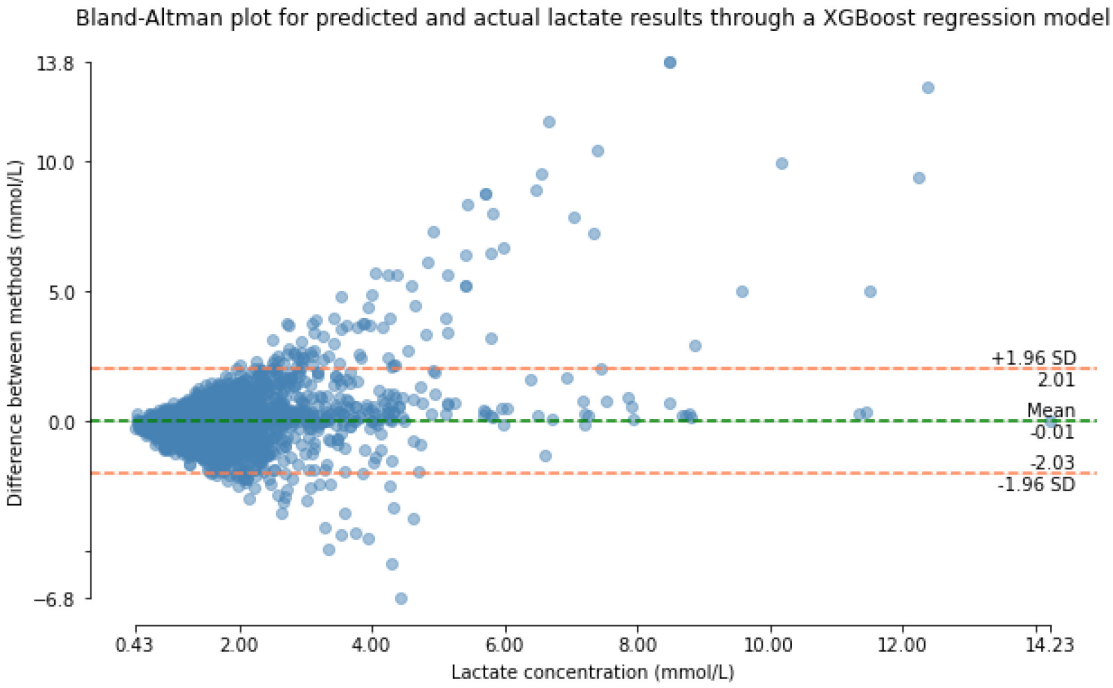


Figure II.5 Bland-Altman plot showing differences between predicted and actual lactate values across different lactate concentrations. The 95% confidence intervals were +2.01 to -2.03 mmol/L.

1. Implementation of models and pathway mapping

v Coding results of interviews

The following table represents coding process and relevant examples extracted from the interview transcripts. Subsequent axial coding processes link up individual codes to create themes on the dengue decision-making process.

Category	Topic	Code	Example(s)	Interview number
Assessment	Classification of severity	Importance of Ministry of Health dengue protocol	...about severity on a dengue fever patient, we use the Ministry of Health's protocol, and the Hospital for tropical Disease also has a protocol which is based on the Ministry of Health's general guide.	1
	Role of vital signs	Importance of specific clinical signs	...if a patient has a critical dengue fever, we evaluate basing on the patient's pulse rate, heart rate; blood pressure is the most important parameter. Besides, we have an evaluation based on one thing which is called… professionally called perfusion of peripheral organs	1
		Blood pressure and pulse rate used in decision-making	the most helpful value, firstly, is pulse rate. . If we have continuous pulse rate monitoring, it should be very good. I think one thing that is more helpful is continuous blood pressure monitoring.	1
	Assessment of symptoms	Screening of clinical conditions is based on MOH guidelines	It's made clearly in the protocol based on which symptoms that a patient is accepted for admission or hospitalization, or he/she may be taken care of at home. It's done mainly through screening in the first place which is mainly based on the patient's clinical conditions.	1

	History	Timing and day of illness is important to consider	For Dengue fever, severity is not seen critical during the first three days. Rare cases of critical conditions, for example, it's an encephalitis case. But in most cases, the condition becomes critical during day Three to Seven. But when the patient's condition becomes critical and we're even not aware if it's a Dengue fever case, we're likely involved in an unidentified treatment not knowing what disease the patient suffers	3
	Co-morbidities	Co-morbidities increase the complexity of case management and are not covered in MOH protocols	if the patient has many underlying diseases. For example, a patient has leukaemia, and has Dengue fever at the same time, then it'd be so challenging as the patient has severe blood clotting disorder, or another patient who has diseases such as critical asthma, and at the time he/she's suffering an asthma attack, so it's complicated with two things: the patient has his/her underlying disease. Generally speaking, all of the Ministry of Health's disease guidelines are for a certain disease only, therefore, there is a principle in medical field that it's said it's to treat a patient rather than to treat a disease. To treat a disease is so easy as who can read can read the guideline and follow it,	3
	Admission to hospital	There is a list of indications for hospitalisation and these are based on MOH protocol	it's to follow the Ministry's protocol, we can see clearly defined conditions with which a patient is to be hospitalized, generally speaking, it's for critical cases. For light cases, for those patients with normal haemorrhage they will be admitted for monitoring such as kids, patients with obesity, in pregnancy, or in old age	2

		WHO 2009 warning signs are used in conjunction with MOH guidelines	...are included in the Ministry of Health's guidelines. It means who among the patients is for discharge, it means if the patient doesn't have any warning symptoms and the patient may go home, and for a patient with warning signs or with certain conditions, or a patient who's from somewhere far away and doesn't have conditions for daily monitoring and has to be hospitalized, so there are regulations for those patients in critical conditions to be admitted into the ICU	3
		There is some ambiguity to the MOH guidelines relating to change in Hct and platelets	There is only a criterion that is rather inconsistent, that is, there hasn't been a particularly regulated threshold in the protocol. It's the threshold of platelet decrease and of HCT increase. There is just a general sentence "sharp decrease and increase of HCT" while it depends on a doctor's identification. Currently, the doctors' identification is a general one which is agreed among the doctors that it's below a threshold, for example, below a hundred, sharp decrease, the word "sharp" here, for example, for a decrease of more than fifty percent in comparison with the previous value, for example, and then we agree for a hospitalization. It's an agreement among the doctors. I've been involved for a long time and seen so but the protocol doesn't include a particular regulated value.	9
	Age	There is a difference in managing adults and paediatric patients	...conditions are varied in adults, children, adolescence, young adults, the aged. Therefore, that if one is not well trained, that one with his specialization works in another, which may be against each other, may make people die.	3

Diagnosis	Differentiating dengue from OFI	Presenting symptoms - absence of specific symptoms	Dengue fever is like a viral infection disease, normally, if, I mean patients' symptoms like having a fever, muscle pain or muscle fatigue, lose appetite.. **only no respiratory symptoms**	1
	Age	Children can be difficult as they provide vague symptoms	clinical symptoms are not so distinctive as those on adults. The kids are too young to make complaints. Second, they can only describe so ambiguous clinical symptoms, so mostly, regarding the fact that if a patient is actually going to have a critical complication, mostly, for the paediatrics, we rely mainly on subclinical matters	6
	Differentiating dengue shock from other forms of shock	Specific clinical signs are present in dengue	Dengue fever patients may have rather cooler arms and legs in comparison with other patients'. However, the overall look is very fresh, the complexion is also rosy but not pale but the patient is in a septic shock… for the patients having a septic shock, their blood pressure would normally drop more dramatically than.... If we're observant, we could figure it out.	2
	Differentiating dengue shock from other forms of shock	Alternative forms of shock can be difficult to exclude	a case having a shock, Dengue fever shock or septic shock. It makes doctors so confused	5
		Experience is important to differentiate between shock types	And especially, with inexperienced people in Dengue fever disease, for example, they may think of sepsis, shock or a blood disease and treatment would be going far and therefore, the treatment for the patient isn't appropriate, and the condition may become worse	3

	Undifferentiated fever	Uncertainty in diagnosis is one of the main challenges in dengue management	...and I mainly meet Dengue fever cases when I work in the examination department so as for me, a complex patient is when the patient's symptoms aren't clearly confirmed. It may be an un clear symptom of having fever, the patient may not remember when he/she has a fever, or he/she may just feel chills or something. Symptoms are unclearly reported and confirmed then it's a complex case. All kinds of tests are requested to be done for the patient. And for those patients with comorbidities, it may challenge us for providing treatment. For example, obesity or else, or chronic diseases that go along…	4
	Epidemiology in clinical practice	Knowledge of seasonality and epidemiology is useful to know for managing patients	… and it depends on epidemic areas, and an important factor: seasons. For example, in Ho Chi Minh City, during the rainy season, patients are admitted with very high temperature.	1
Investigations	NS1 assay	NS1 tests can take a long time to come back	But it takes so long for an NS1 result, not until some hours later for it to be returned while we need clinical tests immediately…	2
	CRP	CRP could be useful but not consistently measured for patients	Because by then if NS1 is negative, it'll mean nothing so…but by the days, we'll have CRP done for the patient. As for other doctors' patients, they may not have CRP done	4
	Role of investigations	Clinical signs and symptoms guide diagnosis	No, we don't rely on… Normally, it's based on clinical, clinical symptoms of Dengue fever, and other very similar viral infections. It's a matter that if we cannot do early diagnosis and until it gives clear symptoms, normally, the patient's condition has been critical,	3
	Use of liver enzymes	ALT/AST to guide assessment	firstly, on symptoms, symptoms as a patient reports, and then rely on what's	4

			from my examination. Next are the tests, blood count test or liver enzyme…	
	Adults vs paediatric cases	Blood sampling can be difficult in children	it's very hard to take blood samples from kids. When it's to do it, as for those challenging kids, I must try two or three times, and sometimes, the results may be incorrect. But mainly, mostly the relatives don't agree for us to take blood samples many times	
Directed therapy	Specialist therapy	Plasma exchange for liver failure is a specialist treatment	Recently, my department has decided to implement an intervention that a plasma exchange for the patients who suffer severe bleeding and severe blood clotting disorder	1
	Deviation from MOH guidelines	Experienced doctors may attempt practices not covered in the protocol but only after no initial response	for example, on the patients, the protocol is applied and no responsiveness on the patients is seen, and then some experienced doctors who are considered experts may try a small customization, so this is done based on experience and not part of the hospital's protocol	1
	Limit of MOH guidelines	Treatment for recurrent shock beyond two episodes is not covered in the MOH protocols	As for those cases under the shock management, the patients suffer repeated shocks three or four times… Two times is still included in the Ministry of Health's protocol but three times and more would be out of the protocol and we must rely on clinical evaluation and make decisions on how/what infusion the patient should have.	2
	Individualised therapy	Fluid volume and rate assessed by vital signs (pulse, pulse pressure) and haematocrit	that the patient has such signs then we decide on an infusion for the patient. And after the infusion, the patient is responsive, the patient's HCT value is expected to decrease, the patient's blood pressure is likely to be wider, for example, at first, the patient has a blood pressure of 10/8, after infusion, the patient is	2

			responsive and it may be 11-8 or eleven over seven, it means	
		Unclear if adjusting fluids by weight or BMI will always lead to improvements	adjust infusion based on the patient's weight. As I don't know if to adjust that may work… if to adjust it based on weight, you know, may work or not. I've seen to adjust the fluid infusion based on weight correctly, sometimes, it didn't help to avoid a shock	4
	Blood transfusion	Indication for transfusion is a difficult decision and needs to be balanced with danger of fluid overload	The case that gave me the impression was the one on which I made a most memorable decision: blood transfusion or no transfusion for the patient. If we could make a correct decision, the patient's life would be saved. If we made a wrong decision, we might have made the patient suffer an infusion overload, which would have made the condition more critical.	5
	Role of continuous or frequent vital signs	Frequency of vital signs measurements is important	as every time we give an infusion we count the hours, for example, we count fifteen millilitres per hour, and we take the blood pressure again after an hour, measuring in between doesn't having much meaning	2
		Hourly vital signs is performed ideally	For a newly admitted patient, if it's necessary for monitoring, it's done within thirty minutes, and after that it's done every hour. For those patients in critical conditions, it's downstairs in the ICU…	4
	Treatment priorities	Adult vs paediatric manifestations of dengue are different and treatment is different too.	as I mentioned a moment ago, it's mainly Dengue fever shocks only. There are rarely complications with multi-organ failure or haemorrhage, while on the adult side it's the opposite. Regarding the complication management, the complication frequencies are different. Adults have more underlying diseases	6

Classification	Formal risk assessment of patients	No formal risk prediction score / scoring sheet / risk calculators exist in routine use	At present, we don't have a reference… does it mean something like SOFA score in sepsis, right? not yet, at present.	1
	Scoring and risk prediction	Systems exist for severe/non-severe classification initially but no scores for exist prognosis over time	Here we don't have any scoring scale at all. In general, for Dengue fever disease, there is not any scale for evaluation to check a patient's severity. Only doctors provide examinations and synthesize and they give expert comments, and there is no scale at all. And to use a scoring scale to evaluate appropriately, so many tests are to be done, or as for septicaemia, so many tests are to be done to generate a scale,	3
		Decision-making can be basic due to time constraints. But more precise ways of aggregating information would be welcome.	we don't have time to work on a score sheet. We think of, first, the child's illness day, if the child has a history of Dengue fever or not, we consider the current test results we have in hand, we can predict the illness development… If I have a predictor, a sheet for prediction, I'd use it.	7
Referral	Departmental referral	Referral between ED and ICU are based on specific agreed indications in the hospital	It doesn't mean an agreement but we have indications to admit to ICU and we have indications not to admit.	2
	Discharge from ICU care	Clinical stability (e.g., 24 hours) is used to support decisions to discharge patients	If we see a patient after being given treatment and getting stable, getting over shocks and staying stable within twenty-four hours, we can confirm conclusively the patient hemodynamic conditions stable without other abnormal signs or other manifestations, we may let the patient to regular ward.	2

	Discharge from hospital	Discharge is based on specific hospital criteria	The discharge standards are also included in the protocol: platelets: over fifty thousand, patients have suffered the disease seven days with no more clinical symptoms, and then they're to be discharged	5
Hospital logistics	Workload	Re-organisation of workflow happens during periods of high workloads	During the high season of dengue fever, yes, there have been times that the ICU has been overloaded. The hospital has decided those departments with dengue fever treatment function upstairs would be responsible for dealing with shocks instead of transferring patients down into the ICU	1
	Overloaded hospitals	There are hard indications for hospital admission	One must be hospitalized if it's a critical case even when there is a severe or non-severe pandemic. those cases, in general, as it was in pandemic and patients came in so continuously, don't know if you've seen this, we had to manage, many had to lie in canvas folding chairs	2
		Admissions can be flexible and compromises in care made	When we see that a patient is having warning symptoms but as it's too crowded, the patient cannot be hospitalized, and if hospitalized, the patient would just be merely monitored only, so we advise the relatives to give further monitoring and make an appointment for check-up in half a day. One may have examination in the morning, and is present for check-up in the afternoon	3

		Dengue management involves a balance between needs of an individual with that of a system (in context of epidemic), and this is done through resource allocation	the more attentively the patients are monitored the better. It's not just for Dengue fever but also for other diseases. It means the more attentively we monitor a patient, the more appropriate our decisions would be. But this disease is not for a person, for this disease, so many people are infected at a time as it's an epidemic, so we must have an optimal plan to make sure monitoring is appropriately done in reference to our healthcare conditions to avoid a waste of the human resource for those non-critical cases... It's not necessary to monitor every patient attentively as it would waste so much our resources.	3
		There exists flexibility in managing cases, especially as outpatient	For example, when an epidemic comes, it may become too occupied and there may be no beds available anymore, so first, too many patients are admitted into fully occupied so no beds for all, and many have to lie on the floor, which is not comfortable; and the staff get overloaded, too, they'll not be able to cover everything… so at those times, I must consider, for example, to request daily check-ups, or advise them to go to other lower hospitals that can follow and provide Dengue fever treatment	4
	Iatrogenic complications	Healthcare workers can cause iatrogenic complications if inexperienced	Nurses need special training separately. The reason is that bleeding easily occurs in Dengue fever patients. If a nurse is not so attentive, gives injections where it's easy, and then bleeding is seen at the spots where the needle has injected. Because when a patient is already in a critical condition, haemorrhage condition then bleeding easily happens. Therefore, those inexperienced nurses may damage the veins they when inject	3

	Capacity	The hospital system is flexible and can cope	In general, there is enough human resource for all situations, so hospitalization or departmental transfers are requested based on patients' condition, and they don't influence the staff at all, the staff cannot make change to the treatment decision.	5
Team working	Hierarchy of command	Department head or deputy gives approval for treatment decisions.	Approval from the department head or head deputy doctor must be given. We cannot do something for the patient on our own.	1
	Team decision making	Consultation is needed for specific decisions.	Consultation is for a case that needs multi-departmental specializations or is involved in antibiotics use, combination, or as for complicated issues, each hospital has its own regulations to authorize, for	3
Trouble shooting	Complicated case management	MOH protocol are tried initially before alternative approaches taken	when we provide treatment to the complicated cases, we still follow strictly the protocol first. Normally, there are situations when a creative idea is raised and to check if there is any better solution for a patient based on scientific evidence	1
Knowledge and learning	Sharing knowledge	Sharing knowledge through cases, or contribution to research is desired	We work but we don't have what is called… There are good cases and we just close the files with reports only and not a real research study, so I hope, it's to help everyone improve, develop, I hope scientific study is done. We work and we find scientific evidence, definitely, and it would help to develop the department. From those things, experience would be passed on to other doctors or next generations.	1
	Medical education	Dissemination of clinical information to unspecialised facilities, medical	You know the information we've collected from the hospitals specialized in treatment of Dengue fever disease is actually necessary for those work in unspecialized level facilities; general hospitals are not	3

		education is needed	specialized in Dengue fever, or district level treatment facilities, or private clinics need the information synthesized from the specialized hospitals' database	
Uncertainty	Recurrent shock treatment	MOH guidelines are not explicit for recurrent shock beyond two episodes	decisions for opinions such as need extended treatment, or severe shock patient treatment. There are some instances we need to ask for clinical opinion such as re-shock, blood infusion for such haemorrhage. When patient re-shock, what are the dosage, do we need to change dosage? Because once patient re-shock, re-shock the 2nd time, there is no dosage recommendation from WHO or MoH, so, we have to base on clinical scenario.	8
	Completeness of guidance	There are a few scenarios not covered by the MOH guidelines but in general they are useful	...when a patient is getting into a shock, two to three times is described and included in the Ministry of Health's protocols but that a patient has repeated shocks three, four, or five times isn't included in the protocols...For other matters, for me, they're all included in the protocols. Nothing is missing.	2
	Up to date knowledge of patient status	Role of monitoring equipment for contemporaneous knowledge of patient state will be helpful	Therefore, I think, for monitoring the patients, if more equipment items are provided, central monitors are supplied, it would help monitoring the patients and early detection on patients' shock periods, or as a patient is going to have a shock, we may be aware early, and it should not be a situation in which a patient has not been found suffering a shock until one or two hours later..	1

	Urgent decision-making	Certain decisions have to be made immediately and consultation is not always possible	for a Dengue fever patient, normally, a doctor must make decision himself/herself, except challenging cases who must be in the ICU and consultations are needed. A patient with many underlying diseases is needed to managed immediately. Therefore, when it's for you to make decision, it'd be late if you go around and ask this person that person. Therefore, a doctor must be flexible in whatever way he/she can to make an immediate decision.	3
	Access to clinical parameters	There is a need for medical equipment to be provided in the examination rooms	Sometimes, the doctors don't have enough time to take vital signs. Some nurses working in the paediatrics department would not be familiar with working on adults and vice versa, so when they take vital signs, they may think it's a normal vital sign but it's not normal actually at the age, or, this is abnormal but in fact, it's normal at this age, you see.... no SpO2 oximeters for patients in the examination room? D: No, there isn't. therefore, it's a shortcoming. Sometimes, I feel the pulse, and like a habit, and also personal experience	4
		Vital signs can be missed because of healthcare capacity and workload	patients deteriorating without noticing, especially, it often happens in regular wards, for example, at night nurses take vital signs of patients with warning signs every 3 hours but there are cases without warning signs and still develop shock, those are the cases that we miss. We can't manage all, 2 nurses during the night cannot measure 50-60 patients of entire department every 1-2 hours, it's overwhelming...	5

Patient factors	Request for hospitalisation	Patient requests affect indication for hospitalisation	it's seen with socio-economic issues, and economic or other benefits, or services that they want to satisfy patients' need so they'll give patients admission for hospitalization	3
		Patient sometimes prefer not to be hospitalised and discussion is often required.	but we must persuade them. Because, sometimes, a patient may have alert signs but the patient may keep say, for example, he's/she's fine, and his/her home is nearby so he/she doesn't want to be hospitalized for management, or sometimes, a patient may feel well, he/she doesn't see any symptoms but in fact, as his/her pulse rate or blood pressure is taken, it shows that he/she is going to suffer a shock so we must persuade him/her he/she is actually not well, not well at all and needs to be hospitalized. Therefore, it's for us to persuade them.	4
	Age specific challenges	Paediatric patients are challenging	Also, the kids are so active, automatic continuous vital sign monitoring is quite difficultly done, they usually take the device stuffs off and play, for example, so they're not likely to have them on. Therefore, normally, to take vital signs, we take advantage during the day and night like thieves, we must wait for them to sleep,	6
	Financial factors	Money is not a large issue in most cases of directing care	think dengue fever treatment cost is not much. Because the number of days of their treatment only counts from 3-7 days, the number of days in the hospital is only from 3-7 days. The amount is really not much.	8
	Financial factors	Money is important	..., they need medicines, treatment methods that require high expenses, especially using ventilator, dialysis machine, it seems that the cost is so high. If the treatment, on average, if patients having both dialysis and ventilator use, it	10

			costs about two to three million dongs a day. If the treatment is to go on, within one to two weeks, half a month, it may reach fifty to sixty million. For those patients without the medical insurance, being manual laborers, with economic constraints, it'll be a very big number.	
Political factors	Dengue as a priority disease	Top down approach for dengue care guided by the ministry of health	Dengue fever is a rather sensitive disease that the Ministry of Health is concerned so when it becomes a big issue, it's so big so it needs approval from the leadership.	2
		MOH guidelines is a form of legal entity	Currently, I think everything complies to the Ministry of Health's protocol, because the Ministry of Health's protocol is legal and it's also based on experts' recommendations.	6
CDSS attitudes	Deskilling	Concerns on gaining experience if a computer program is used	in general, the experience is from one's own learning on patients, one must look at a patient to recognise signs, if it's just to rely on software products, sometimes, one may lose his normal sensitiveness, and they keep looking in such a way	2
		Availability of information is the norm and useful for reference	Skills can be learnt from training, and software products may generate recommendations for management and they're just for reference. At present, we have so many applications on phones or from google. It just needs searching but it's human to make decision, therefore, skills can be learnt from training or coaching, and they are there just to help us make decisions... but the matter of learning by heart is the old days' story, and now a doctor sometimes must be multi-functionable, so the issue wouldn't influence too much, which means the matter would be not its flip side,	3

		Availability of information is the norm and useful for reference	Therefore, I don't think if one relies on that and would lose his/her predicting skill as the first thing a doctor must do is that he/she must listen to the patient about their symptoms, and it's not the fact that he/she would completely rely on something so I think it's feasible and it wouldn't influence judgement matter at all	4
	Matching expectations and context	Expectations for such a programme is varied and should suit the context as well as be flexible	We should meet the expectations in place where a program is used. It would be more appropriate. For example, if it's for district level only, it's to collect inputs on their expected items, needs, goals as they may not expect something deeper… it depends on particular demands at particular units, you should get preferences from many places so that there should be software flexibility...	2
	Domain of decision-making	CDSS cannot make decisions - legally and ethically	… It's impossible to rely on consultation, this software, that software, for example, to make decision. It's never been allowed by the government, or by the law.	3
	Wide domains of information	Patient related information which includes social history is useful to include for reference and decision making	And the application needs to use information about other sources such as symptoms, or clinical information, patients may take vital signs themselves or subclinical information must be sufficient, medical history must be sufficient, that's it, even history, it should be included… I think a category must be created for patient symptoms, or if it's not so, I think of a factor like if a patient lives alone or with a caregiver or a helper. As part of Dengue fever treatment, resting is an important matter. If a patient lives alone, the patient must do everything himself/herself.	4

	Patient/public interface	CDSS systems could feature patient reported symptoms or patient data entry	for Dengue fever, besides clinical symptoms, the symptoms and the signs that patients can report themselves are of higher significance so I think such an application may be feasibly applied. But besides to introduce such an application, patients at home should still have gauges to take their vital signs as it'd be better	4
	Intended users	Software is only useful for junior staff and those in district facilities	I think the software, probably, is only, or it'd be helpful for new doctors. Newly joined doctors are not so experienced then using it'd be more supportive, and the staff at the lower level hospitals, it'd help more. For the senior doctors with much experience then they won't need.	2
	Conflict with outputs	Implicit trust in own judgement processes is more important	We would not be able to accept reluctantly that it may produce what it could and we may rely on ourselves, we trust ourselves, we don't trust it.	1
	Trust	External approval would be importance in validating and trust. A formal oversight and regulatory processes	And data, the way a decision would be made must be based on scientific basis, scientific evidence so that it would be easy, and would have been approved, like a vaccine which has been approved and is made by well-known producer then we are more likely to get it rather than other types. It's the same with a software product. It must be scientific, it must be approved and it must be easy to use	1
	Updated information	The system should not display outdated information	but patients' condition must be updated every day so that we can monitor them attentively and as we're not always directly available there with patients so I think it must be updated every day.	4
	Time for decision-making	A smooth, stable and quick interface is important	Also, it'd be able to produce a decision quickly. It means whenever we need to use it, it could produce something immediately. We don't expect that we	1

			would have to sit and wait and see nothing…	
	Accessibility	Access through a phone would be beneficial	..for example, it would be able to use on phones or it would be popular and it should not become frozen while running, then people would use it easily…	1
	Additive roles	The system will not be used if it does not save time	If it's in a high season, there are too many patients, and if we have to do too many mechanical procedures, we'll not make it, it's actually so because we just read the test results quickly and manage the cases immediately as we don't have time to input and save the data to follow up, because in a department, there are only three doctors in a shift. In one night, if we receive ten Dengue fever cases, we'll not be able to make it.	5
	Objection to double data entry	Will not be used if data entry required	If the program requires the doctors and nurses to input data one more time, I think, in the hospital environment, it'd be impossible as they have too many things to do. If they have more things to do, first, it'd slow down the work, second, about the fact that it's, in fact, kind of prediction, personally, each doctor who monitors a patient from the beginning to the end can predict the patient's condition development so the fact that it requires inputting data from the beginning and then activate the program to calculate, it'd be unnecessary.	6
	CDSS output	The system should be easy to understand (usability)	I wish a software that is clear, easy to understand, to apply it. In general, it should be an easy-to-understand, simple software. I wish so, so that sometimes we apply it, in a particular situation that is so urgent. That's it. In the situation, if we start to use it, we need it to be simple so that we can use it quickly. And, if the	7

			software is available soon, we can check it before use, for each case admitted, we can use it for prediction right there.	
	Interpretation of risk	Risk percentages are inherently difficult to interpret	percentage how could a the doctor assess! The doctor can't evaluate, so I said a software is needed to evaluate if the doctor is questioning, if the doctor has doubts. As for me, I can only evaluate it as dengue fever or not, I only write whether the patient has dengue fever or not, but I don't…, or don't have dengue fever, but I can't consider percentage as part of the diagnosis	8
	Suggested management templates	Recommendation from pre-set monitoring options could be useful	For instance, that patient has warning sign only, no shock but obese... Should I consider transfer patient to ICU or monitor closely hourly or monitor pulse, BP every 3 hours. This is because monitoring patient every hour increase workloads for nurses and doctors, so if there is a software that recommend how to monitor patient, or transfer patient to any ward, then I will use it as its reference is more precise doctor's, it is more comprehensive	8
	Intermittent monitoring can add to workload	Rechecking observations at frequent intervals can be time-consuming and unproductive	For example, when I feel that patient may enter shock state but decide to monitor not so closely, it makes no sense and makes me anxious. Besides monitoring hourly again by myself, there's nothing else I could do. That is all I could try to do. Nevertheless, if I monitor closely mild patients, it will add up to the workload for nurses and also for patients, keep monitoring and taking blood from mild patients continuously is not a right thing to do.	8

	CDSS outputs	Produce an explicit recommendation for approaching the problem e.g., frequency of observations	We enter all kinds of information into the software, it must know how often I should monitor that patient, I should take tests and how often should I monitor clinical.	8
		Concept of ML as a diagnostic - when NS1 unreliable (secondary, late infections) or IgM negative and PCR not available (late shocks)	And another problem is that there are cases where NS1 is considered as a test that accurately diagnoses dengue fever. There are cases when I see NS1 is negative or IgM is negative, we only have platelets and HCT, then only clinical thing is my examination but if NS1 or IgM it is negative, then what is the percentage that a patient will actually have dengue fever and not another disease. That's it, there are cases when it's difficult to diagnose.	8
	Interpretability of risk scoring	Why has it produced this recommendation or score? What depth of explanation is satisfactory? Accountability and interpretability of machine learning models.	if it is good and everyone knows, then I will give the software it calculates like that, I feel something, and the software that no one knows about that software, but give it to the seniors, they will ask "why should you believe it?" I can't explain it	8

vi Next steps of work for decision-support development

Using findings from interviews and both workshops, wireframe designs of a digital clinical decision support system (CDSS), a web-based implementation was designed through a third-party developer.

Work was done as part of the Vietnam ICU Translational Applications Laboratory (VITAL) project in collaboration with colleagues from Oxford University and Hospital for Tropical diseases.

Description: Interface viewed on a computer screen or mobile device in clinical settings to provide functionality in clinical decision-making for dengue patients.

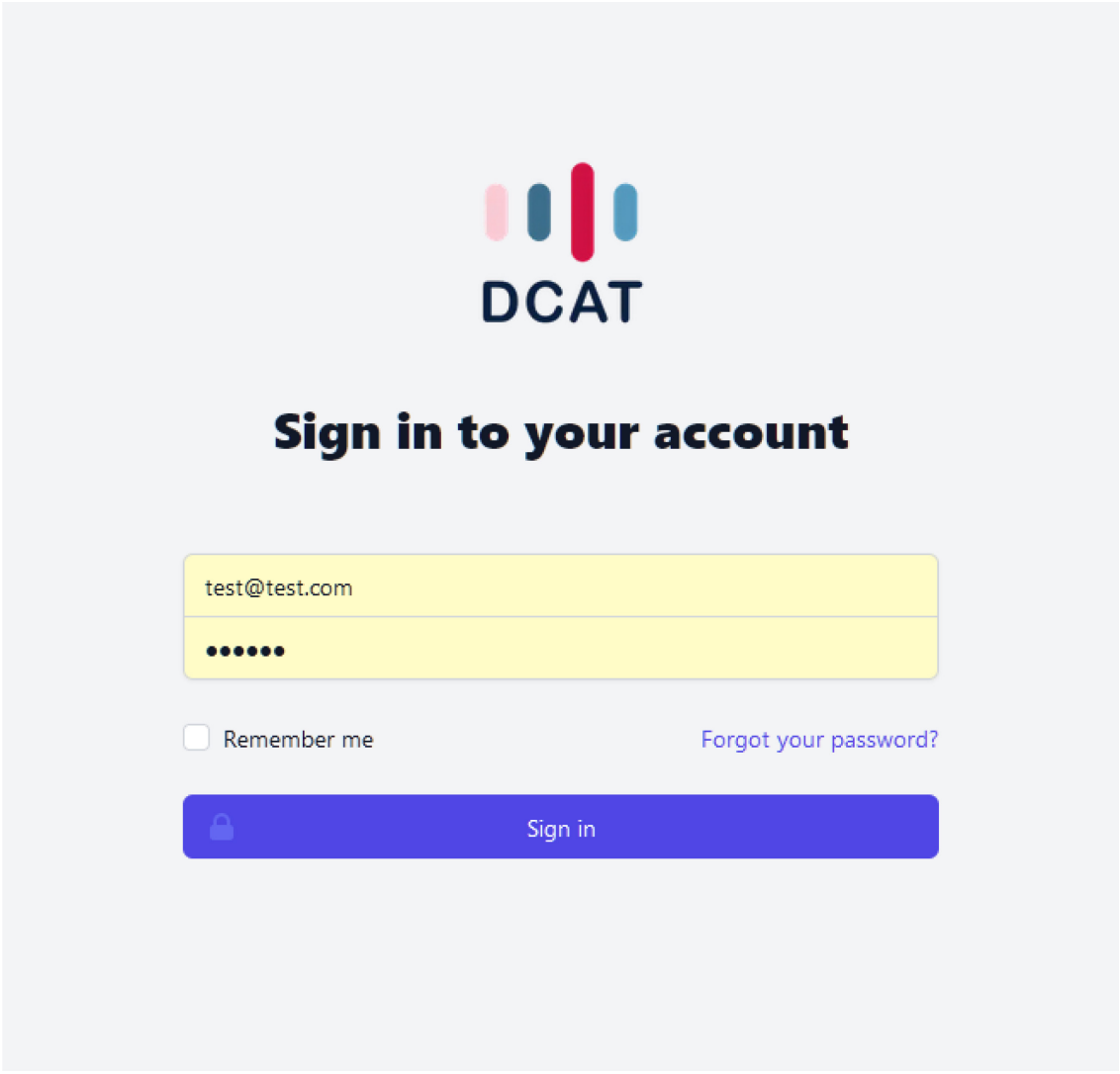
Format: Web pages hosted on OUCRU intranet, communicating to database back end provided by Imperial College London

Functions:

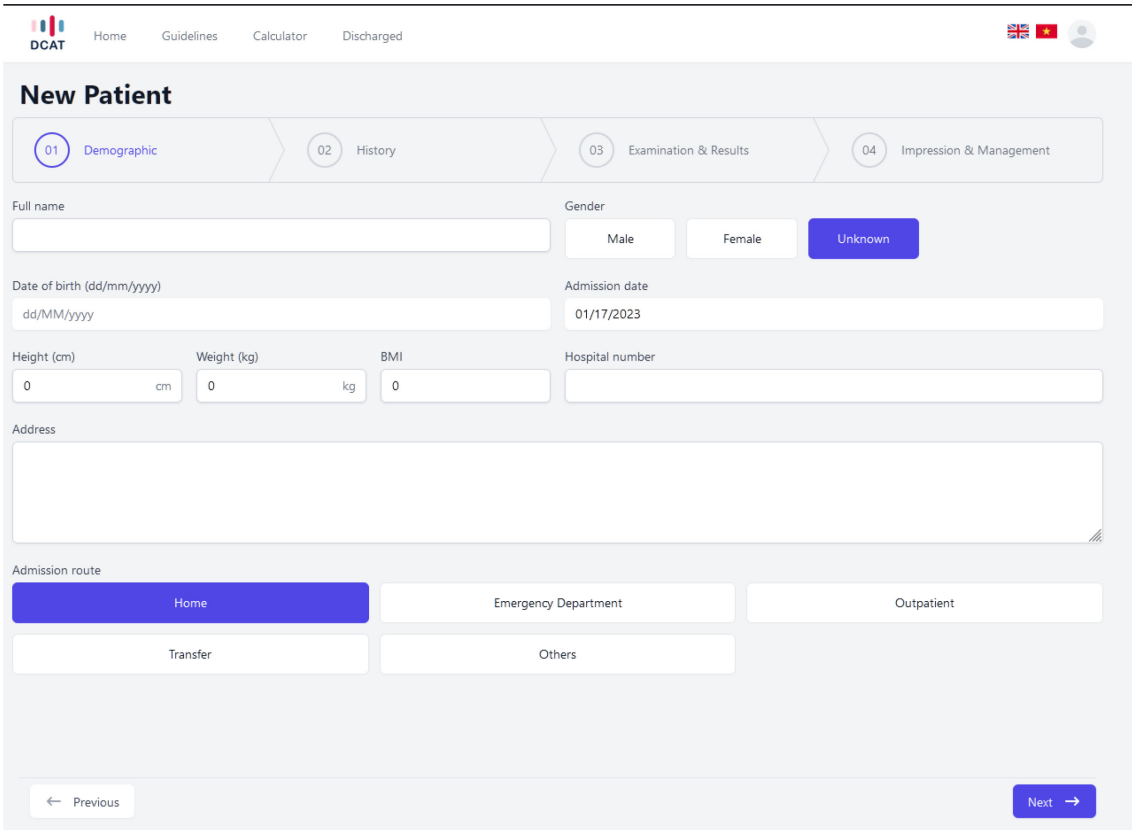
1) A display of patient names, age, gender, dates of admission and blood test results
2) Data entry by user to input clinical information
3) Quick access links to information resources (guidelines, calculators) hosted locally on the intranet
4) Access to patient risk prediction modules
5) A report function which generates a pdf of patient information which can be printed out or emailed
6) Individual user profiles which record the patients which have been accessed most recently
7) Patient message board / chat function for posting information

Screenshots of the D-CAT (Dengue Clinical Assessment Tool) are shown below. The system is web-based and intended for hosting on a secure intranet.

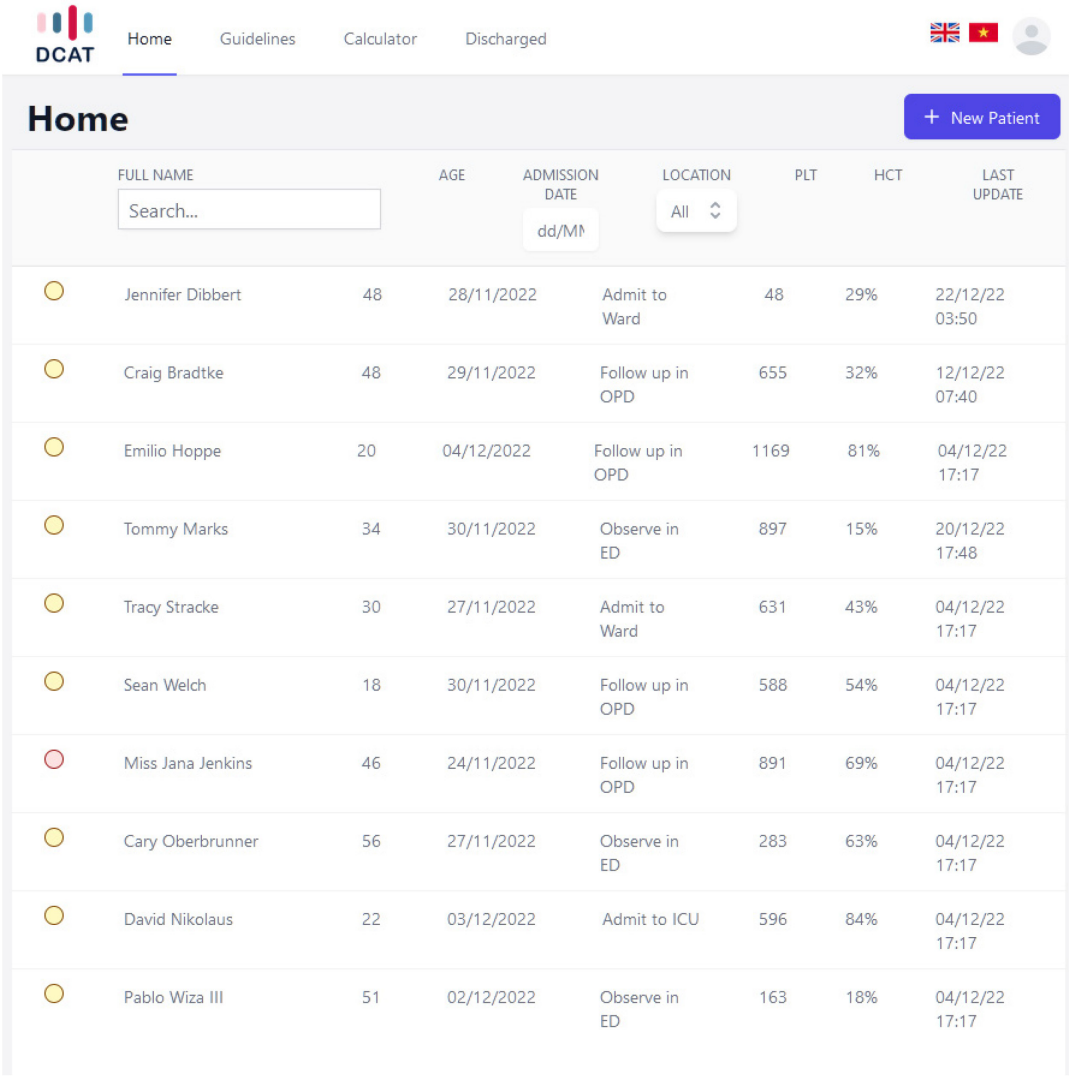
Next steps for implementation will involve usability testing in end-users: clinicians will be invited to use the CDSS in the clinical setting using generated patient details to evaluate user experience and inform iterative changes to improve workflow. Introduction, use and evaluation of CDSS within a clinical setting beyond an observational capacity is discussed in the thesis discussion section.



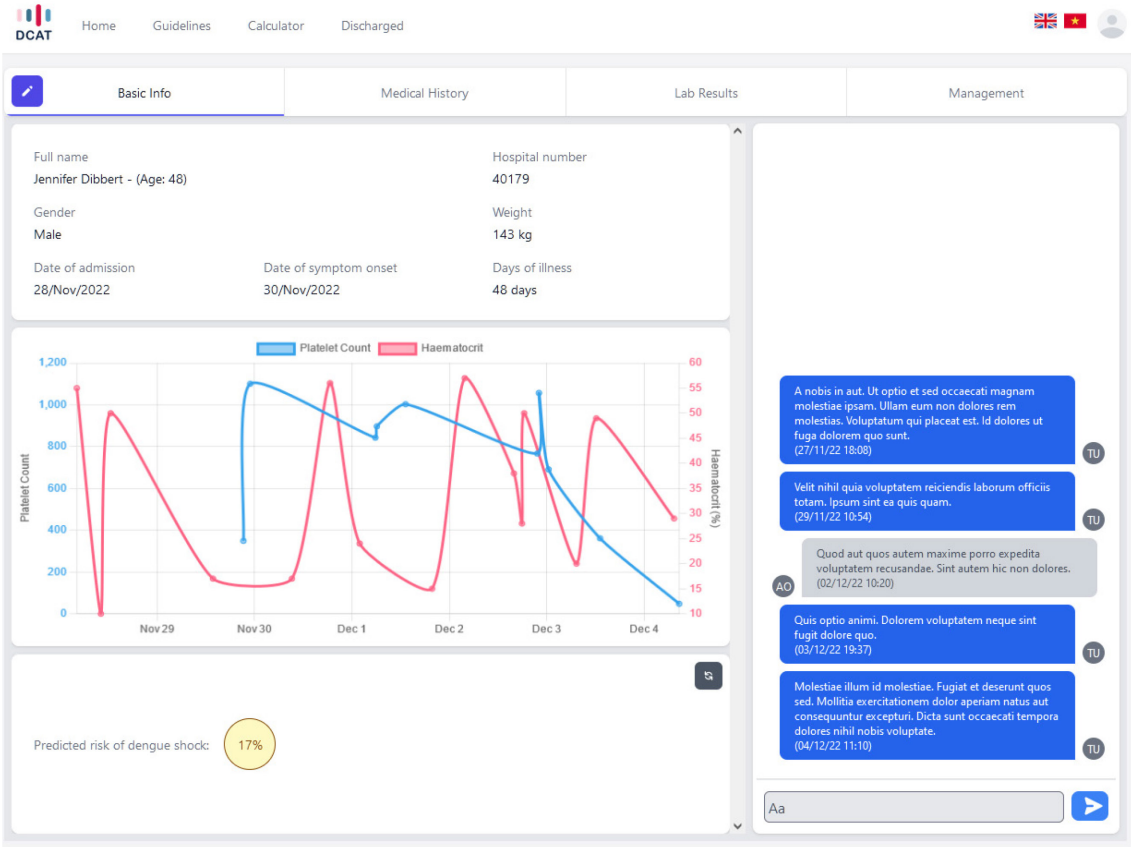
A login screen with individual username and password security to enable different profile users. Unique logins can also facilitate usage audits.



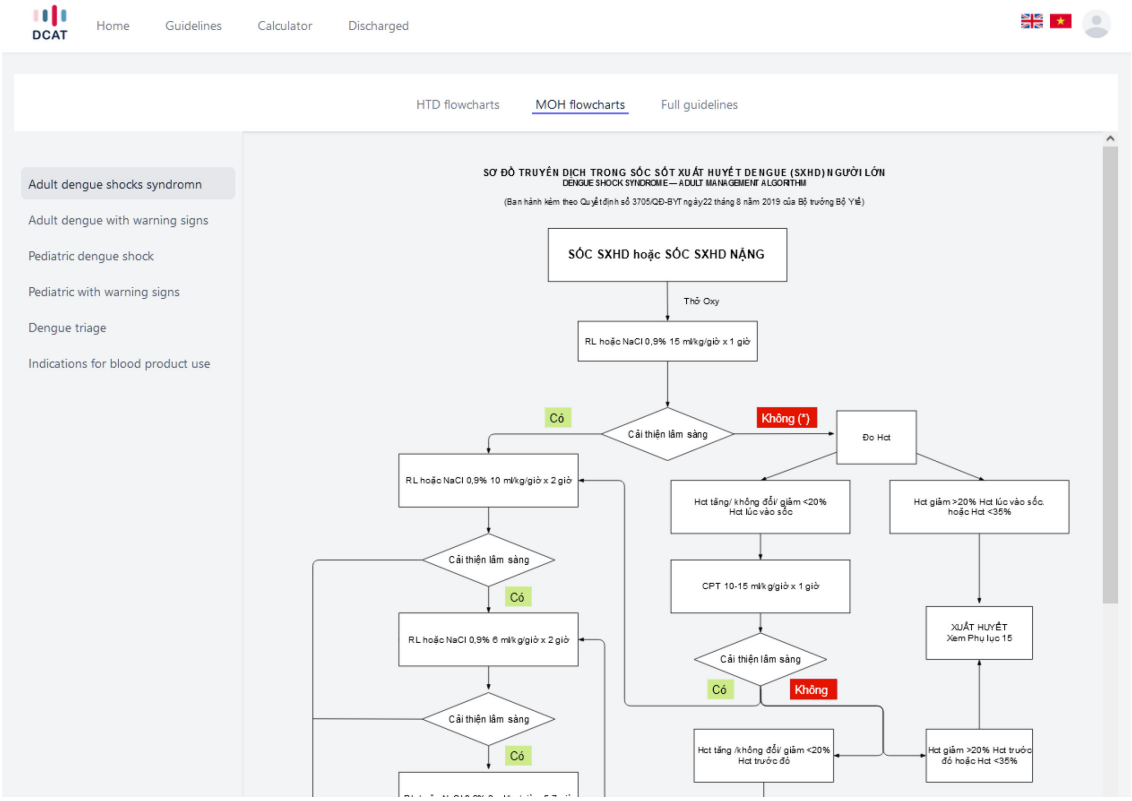
New patient entry screen for entering clinical, examination and laboratory information. This function is intended to be superseded by interoperability with other electronic systems which are to be introduced at the hospital in the future.



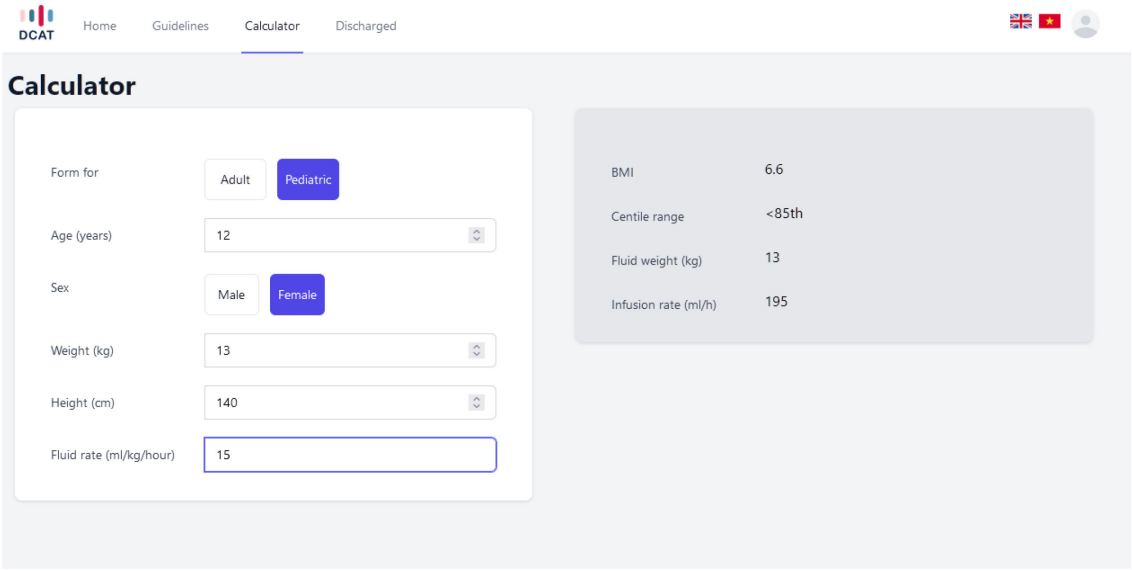
Main screen of patients, details at a glance – this is customised to the last patients accessed by a particular user profile. Updated results and risk score (circle on the left) changes according to most recent results.



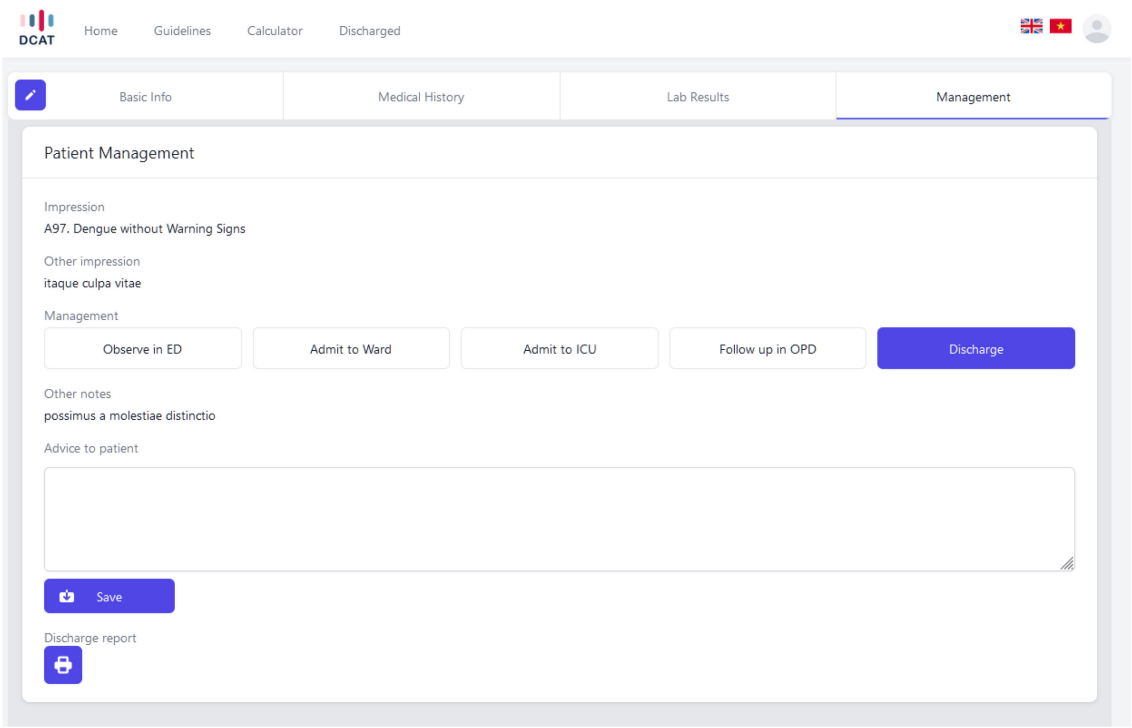
Patient information screen – platelet and haematocrit results graphically displayed, and a chat function on the right shows brief communications between users or allows for note-form documentation. At the bottom a traffic light system displays the predicted risk scores generated by the machine learning models for dengue shock syndrome.



Ministry of Health guidelines in flow chart format which can be browsed, or printed for each condition.



Fluid infusion rate calculator for adults and paediatric patients to provide decision-support



Discharge summary function – aggregation of clinical, examination and laboratory findings to produce a summary for patient transfer of care or own records