Imperial College London

Assessing and improving methods and software tools for real-time outbreak response

Rebecca Katherine Nash

January 2025

MRC Centre for Global Infectious Disease Analysis

Department of Infectious Disease Epidemiology

Imperial College London

Submitted in part fulfilment of the requirements for the degree of

Doctor of Philosophy at Imperial College London

Abstract

Infectious diseases are responsible for millions of preventable deaths per year globally. The frequency and severity of infectious disease outbreaks are increasing due to factors such as urbanisation, globalisation and climate change. In order to identify ways to improve our ability to respond effectively to infectious threats, we must evaluate the strengths and weaknesses of our current methods and software tools for outbreak analytics.

This thesis uses branching process methods for estimating the time-varying reproduction number (R_t) as a case study to reveal the common challenges encountered when responding to outbreaks in real-time. The findings highlight the need for resources to facilitate rapid and accurate parameterisation of models, the importance of extending tools to directly answer policy-relevant questions, and the necessity of accounting for limitations in disease incidence data.

In response to these issues, open source R packages and resources have been developed. An open source database and R package have been created to rapidly parameterise models for a WHO priority disease, Ebola Virus Disease. Additionally, the R package EpiEstim has been extended to estimate the transmission advantage of new pathogen variants in real-time and an Expectation-Maximisation algorithm has been implemented to reconstruct daily incidence data from any temporal aggregation of incidence. These developments prioritise usability alongside functionality, ensuring that they are widely accessible and applicable in a variety of outbreak contexts.

Overall, this thesis emphasises the importance of continuous innovation and evaluation of outbreak analytics tools to meet the evolving challenges of future epidemics and pandemics.

Statement of originality

I declare that the work presented in this thesis is my own. All secondary sources have been appropriately cited. All research was conducted under the guidance and support of my supervisors, Dr Anne Cori and Professor Pierre Nouvellet.

Research in chapters 2, 3, 4 and 5 were published as first author publications with contributions from my supervisors and co-authors.

Chapter 2: Nash RK, Nouvellet P, Cori A. "Real-time estimation of the epidemic reproduction number: Scoping review of the applications and challenges.", *PLOS Digital Health*, 2022, Vol:1, ISSN:2767-3170, Pages:e0000052-e0000052.

Chapter 3: Nash RK, Bhatia S, Morgenstern C, Doohan P, Jorgensen D, McCain K, et al. "Ebola virus disease mathematical models and epidemiological parameters: a systematic review.", *The Lancet Infectious Diseases*, 2024 Aug 7, `https://doi.org/10.1016/S1473-3099(24)00374-8`.

Chapter 4: Bhatia S*, Wardle J*, Nash RK*, Nouvellet P, Cori A. "Extending EpiEstim to estimate the transmission advantage of pathogen variants in real-time: SARS-CoV-2 as a case-study.", *Epidemics*, 2023, Sep 1;44:100692. (*equal contribution)

Chapter 5: Nash RK, Bhatt S, Cori A, Nouvellet P. "Estimating the epidemic reproduction number from temporally aggregated incidence data: A statistical modelling approach and software tool.", *PLOS Computational Biology*, 2023, Aug 28;19(8):e1011439.

Copyright Statement

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-NonCommercial 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.

Acknowledgements

I would like to begin with a huge thank you to my supervisors, Anne Cori and Pierre Nouvellet. Your mentorship has been truly invaluable, and I have grown so much as a scientist because of you. Thank you for the opportunities you have provided, which have enriched my PhD experience in countless ways. Je suis profondément reconnaissante pour vos conseils, votre patience et votre soutien personnel et professionnel tout au long de mon doctorat. It has been an absolute privilege to learn from you both.

To "Team Cori" and all the friendly faces I have worked with in DIDE (past and present), thank you for making this journey so rewarding. I have learnt so much from so many of you, but you have also made my PhD experience far more enjoyable. Your hard work and resilience, particularly during the COVID-19 response, has been nothing short of inspiring. I feel proud to have worked alongside such a remarkable and dedicated group of individuals.

To my incredible friends, you bring so much joy to my life. After four years of "work hard, play hard", thank you for being a huge part of the latter. I am deeply grateful to each and every one of you for being there for me through the hard times and the good.

To my wonderful family, I cannot thank you enough for your constant support. Mum and Dad, thank you for believing in me and for instilling in me that I can achieve anything I set my mind to. You have taught me that the sky is the limit and I would not be where I am today without you.

Finally, to my husband and best friend, Aria, I am eternally grateful for your unwavering support and encouragement. Thank you for always wanting the best for me, for going above and beyond, and for being the greatest teammate in life I could ever have dreamed of.

Contents

List of Publications 11

1 Introduction **22**

1.1 Background 22

1.1.1 Infectious diseases 22

1.1.2 Outbreak response 24

1.1.3 Understanding transmission dynamics 26

1.1.4 Outbreak forecasts 30

1.2 Motivation 32

1.3 Aims and objectives 36

2 Real-time estimation of the epidemic reproduction number: Scoping review of the applications and challenges **37**

2.1 Introduction 37

2.2 Methods 39

2.2.1 Scoping review: Search strategy and selection criteria 39

2.2.2 Evaluation of R package or tool usability 40

2.2.3 Questionnaire 40

2.2.4 Ethics statement 40

2.3 Results 40

2.3.1 Modifications involving incidence data 40

2.3.2 Other input data or method modification 43

2.3.3 Modifications involving geographical factors 44

2.3.4 Modifications with practical or logistical applications 45

2.3.5 Questionnaire 46

2.4 Discussion 47

2.5 Conclusion 49

3 Ebola virus disease mathematical models and epidemiological parameters: a systematic review 51

3.1 Summary . . . 52
3.2 Introduction . . . 52
3.3 Methods . . . 54
3.3.1 Search strategy and selection criteria . . . 54
3.3.2 Data analysis . . . 55
3.3.3 Data sharing . . . 56
3.4 Results . . . 57
3.5 Discussion . . . 70
3.6 Conclusion . . . 72

4 Estimating the transmission advantage of pathogen variants in real-time: SARS-CoV-2 as a case-study 73

4.1 Introduction . . . 73
4.2 Methods . . . 75
4.3 Results . . . 84
4.4 Discussion . . . 90
4.5 Conclusion . . . 93

5 Estimating the epidemic reproduction number from temporally aggregated incidence data: A statistical modelling approach and software tool 94

5.1 Introduction . . . 95
5.2 Methods . . . 95
5.2.1 EpiEstim . . . 95
5.2.2 Extending EpiEstim for coarsely aggregated data . . . 96
5.2.3 Initialisation . . . 97
5.2.4 Expectation . . . 97
5.2.5 Maximisation . . . 98
5.2.6 Case studies . . . 99
5.2.7 Influenza . . . 99
5.2.8 COVID-19 . . . 99
5.2.9 Simulation study . . . 100
5.3 Results . . . 101
5.3.1 Influenza . . . 101
5.3.2 COVID-19 cases . . . 102
5.3.3 COVID-19 deaths . . . 103

5.3.4 Simulation study . . . 105
5.4 Discussion . . . 108
5.5 Conclusion . . . 111

6 Forecasting COVID-19 cases in US states using reconstructed incidence data 112
6.1 Introduction . . . 112
6.2 Methods . . . 114
6.3 Results . . . 115
6.3.1 Reporting patterns by US state . . . 115
6.3.2 Reporting patterns by outbreak phase . . . 116
6.3.3 Forecasting performance by state . . . 117
6.3.4 Forecasting performance by outbreak time period . . . 119
6.4 Discussion . . . 119
6.5 Conclusion . . . 122

7 Discussion 123
7.1 Incidence data will always be imperfect . . . 124
7.2 Facilitating rapid and accurate parameterisation is essential for outbreak preparedness 129
7.3 Maximising the utility of models in policy-relevant decision making . . . 131
7.4 Increased functionality should not be at the expense of usability . . . 132
7.5 Further Limitations & Future Research . . . 134
7.6 Conclusion . . . 135

A Supplementary Information: Real-time estimation of the epidemic reproduction number: scoping review of the applications and challenges 203
A.1 Overview . . . 203
A.2 Scoping review: additional results . . . 203
A.3 Exploration of R package and tool usability . . . 204
A.3.1 Ease of installation . . . 204
A.3.2 Documentation and tutorials . . . 206
A.3.3 Speed . . . 206
A.4 Questionnaire: additional results . . . 208

B Supplementary Information: Ebola Virus Disease mathematical models and epidemiological parameters: a systematic review and meta-analysis 214
B.1 Overview . . . 214
B.2 Ebola outbreaks . . . 214
B.3 Additional Methods . . . 216

B.3.1 Study Selection . . . 216
B.3.2 Inclusion and exclusion criteria . . . 217
B.3.3 Backwards citation screening . . . 217
B.3.4 Data extraction . . . 218
B.3.5 Quality Assurance . . . 220
B.3.6 Data analysis . . . 222
B.3.7 PRISMA checklist . . . 223
B.4 Results . . . 227
B.4.1 Overview of all EVD parameters . . . 227
B.4.2 Seroprevalence . . . 229
B.4.3 Transmission parameters . . . 232
B.4.4 Temporal dynamics of disease progression . . . 243
B.4.5 Severity . . . 248
B.4.6 Mutation rates . . . 253
B.4.7 Transmission Models . . . 254
B.5 Pathogen Epidemiology Review Group (PERG) membership . . . 263

C Supplementary Information: Extending EpiEstim to estimate the transmission advantage of pathogen variants in real-time: SARS-CoV-2 as a case-study 265
C.1 Overview . . . 265
C.2 SARS-CoV-2 variant-specific incidence data . . . 266
C.2.1 Incidence data from England . . . 266
C.2.2 Incidence data from France . . . 267
C.3 Estimating the effective transmission advantage . . . 268
C.3.1 Serial interval distribution . . . 268
C.3.2 Non-parametric approach . . . 268
C.3.3 Estimation using MV-EpiEstim . . . 268
C.4 Estimates of the effective transmission advantages of SARS-CoV-2 variants . . . 269
C.4.1 Alpha over wildtype . . . 269
C.4.2 Beta and Gamma over wildtype (France) . . . 276
C.4.3 Delta over Alpha (England) . . . 279
C.5 Method performance using simulated data . . . 282
C.5.1 Simulation approach . . . 282
C.5.2 Description of figures . . . 284
C.5.3 Baseline scenario . . . 284
C.5.4 Sensitivity to serial interval mean . . . 287

C.5.5 Misspecification of serial interval mean . . . 289
C.5.6 Sensitivity to serial interval CV . . . 292
C.5.7 Misspecification of serial interval CV . . . 295
C.5.8 Sensitivity to superspreading . . . 298
C.5.9 Sensitivity to under-reporting . . . 300
C.5.10 Time-varying R_t . . . 302
C.5.11 Two locations with time-varying R_t . . . 303
C.5.12 Time-varying transmission advantage . . . 305
C.6 Literature review . . . 306
C.7 Code and Data availability . . . 307

D Supplementary Information: Estimating the epidemic reproduction number from temporally aggregated incidence data 308

D.1 Overview . . . 308
D.2 Additional results . . . 308
D.2.1 Influenza . . . 309
D.2.2 COVID-19 cases . . . 309
D.2.3 COVID-19 deaths . . . 314
D.2.4 Reported incidence patterns by weekday . . . 315
D.2.5 Classification tables . . . 316
D.2.6 Computational time . . . 319
D.3 Simulation study . . . 319
D.3.1 Constant R_t . . . 320
D.3.2 Discontinuities in the reconstructed incidence data . . . 321
D.3.3 Time-varying R_t: sudden change . . . 322
D.3.4 Time-varying R_t: gradual change . . . 326
D.3.5 Influence of R_t plotting time relative to the time window used for estimation . 327
D.3.6 Number of iterations . . . 331
D.3.7 Different temporal aggregations . . . 332
D.4 Alternative approach . . . 335
D.5 Zika virus disease . . . 339
D.6 Alternative analysis pipelines . . . 340

List of Publications

Publications associated with this thesis

First Author

- Nash RK, Bhatia S, Morgenstern C, Doohan P, Jorgensen D, McCain K, et al. "Ebola virus disease mathematical models and epidemiological parameters: a systematic review.", *The Lancet Infectious Diseases*, 2024 Aug 7, https://doi.org/10.1016/S1473-3099(24)00374-8.

- Nash RK, Bhatt S, Cori A, Nouvellet P. "Estimating the epidemic reproduction number from temporally aggregated incidence data: A statistical modelling approach and software tool.", *PLOS Computational Biology*, 2023, Aug 28;19(8):e1011439.

- Bhatia S*, Wardle J*, Nash RK*, Nouvellet P, Cori A. "Extending EpiEstim to estimate the transmission advantage of pathogen variants in real-time: SARS-CoV-2 as a case-study.", *Epidemics*, 2023, Sep 1;44:100692. (*equal contribution)

- Nash RK, Nouvellet P, Cori A. "Real-time estimation of the epidemic reproduction number: Scoping review of the applications and challenges.", *PLOS Digital Health*, 2022, Vol:1, ISSN:2767-3170, Pages:e0000052-e0000052.

Co-author

- Morgenstern C, Rawson T, Routledge I, Kont M, Imai-Eaton N, Skarp J, ..., Nash RK et al. "Severe acute respiratory syndrome (SARS) mathematical models and disease parameters: a systematic review and meta-analysis.", *medRxiv*, 2024. https://www.medrxiv.org/content/10.1101/2024.08.13.24311934v1

- Doohan P, Jorgensen D, Naidoo TM, McCain K, Hicks JT, McCabe R,..., Nash RK et al. "Lassa fever outbreaks, mathematical models, and disease parameters: a systematic review and meta-analysis.", *The Lancet Global Health.*, 2024 Dec 1;12(12):e1962–72.

- Cuomo-Dannenburg G, McCain K, McCabe R, Unwin HJT, Doohan P, Nash RK, et al. "Marburg virus disease outbreaks, mathematical models, and disease parameters: a systematic review.", *The Lancet Infectious Diseases*, 2023, Nov 28. https://doi.org/10.1016/S1473-3099(23)00515-7

Other publications

- Ogi-Gittins I, Hart WS, Song J, Nash RK, Polonsky J, Cori A, et al. "A simulation-based approach for estimating the time-dependent reproduction number from temporally aggregated disease incidence time series data.", *Epidemics*, 2024, Jun. https://doi.org/10.1016/j.epidem.2024.100773

- Bhatia S, Parag KV, Wardle J, Nash RK, Imai N, Elsland SLV, et al. "Retrospective evaluation of real-time estimates of global COVID-19 transmission trends and mortality forecasts.", *PLoS One*, 2023, 18(10):e0286199.

- Challenger JD, Nash RK, Ngufor C, Sanou A, Toé KH, Moore S, et al. "Assessing the variability in experimental hut trials evaluating insecticide-treated nets against malaria vectors.", *Current Research in Parasitology & Vector-Borne Diseases*, 2023, 1;3:100115.

- Sherrard-Smith E, Winskill P, Hamlet A, ..., RK Nash, et al. "Optimising the deployment of vector control tools against malaria: a data-informed modelling study.", *The Lancet Planetary Health*, 2022, Feb 1;6(2):e100–9.

- Whittaker C, Slater H, Nash RK, Bousema T, Drakeley C, Ghani AC, et al. "Global patterns of submicroscopic Plasmodium falciparum malaria infection: insights from a systematic review and meta-analysis of population surveys.", *The Lancet Microbe*, 2021, Aug 1;2(8):e366–74.

- Nash RK, Lambert B, N'Guessan R, Ngufor C, Rowland M, Oxborough R, et al. "Systematic review of the entomological impact of insecticide-treated nets evaluated using experimental hut trials in Africa.", *Current Research in Parasitology & Vector-Borne Diseases*, 2021, Jan 1;1:100047.

List of Algorithms

1 Pseudocode for `estimate_advantage()` applied to two variants, A and B. 81

2 Pseudocode for simulating the epidemic trajectories of two variants. 83

3 Pseudocode for simulating the epidemic trajectories for R_t scenarios. 100

List of Figures

1.1 Schematic of the outbreak analytics process. 33

2.1 Schematic of the forecasting process using EpiEstim and Projections. 38

2.2 Schematic to illustrate the scoping review process. 41

2.3 Summary of the modifications made to the EpiEstim method in 54 papers. 42

3.1 Study selection. 57

3.2 Basic reproduction numbers by outbreak (1/2). 59

3.2 Basic reproduction numbers by outbreak (2/2). 60

3.3 Meta-analysis for the mean serial interval, incubation period, time from symptom onset to death, and time from symptom onset to recovery (1/2). 62

3.3 Meta-analysis for the mean serial interval, incubation period, time from symptom onset to death, and time from symptom onset to recovery (2/2). 63

3.4 Case-fatality ratio estimates across outbreaks (1/3). 66

3.4 Case-fatality ratio estimates across outbreaks (2/3). 67

3.4 Case-fatality ratio estimates across outbreaks (3/3). 68

4.1 Effective transmission advantage of Alpha over wildtype in England 85

4.2 Method performance on simulated data. 88

5.1 Schematic of the EM algorithm approach used to reconstruct daily incidence from temporally aggregated incidence data. 96

5.2 R_t estimates from daily incidence that was either reported or reconstructed from weekly aggregated influenza data. 102

5.3 R_t estimates from daily incidence that was either reported or reconstructed from weekly aggregated COVID-19 case data. 103

5.4 R_t estimates from daily incidence that was either reported or reconstructed from weekly aggregated COVID-19 death data. 104

5.5 Assessing the performance of our method in the presence of weekend effects in the reported data. 106

5.6 Assessing the ability to detect genuine mid-aggregation window variations in transmissibility 107

6.1 Incidence of COVID-19 cases in the US (7-day average) from March 2020 to February 2022. 113

6.2 Relative reporting of COVID-19 cases by day of the week for each US state. 116
6.3 Relative reporting of COVID-19 cases by day of the week for each outbreak phase. 117
6.4 Violin plot showing the distribution of log-transformed (base 10) CRPS values by US state. 118
6.5 Violin plot showing the distribution of log-transformed (base 10) CRPS values by outbreak time period. 119
7.1 Schematic showing how each aim of the thesis addresses aspects of outbreak response . . 124
7.2 Schematic showing an example of the infection process for an individual and the corresponding epidemiological delays. 125
A.1 The disease or pathogen under investigation in A) the papers that used an unmodified version of the EpiEstim package or the method and B) the papers that used a modified version of the approach or package. 204
A.2 Map showing the country in which each questionnaire respondent is based. 208
A.3 The profession of each questionnaire respondent (A) and the purpose of their analysis (B). 209
A.4 Disease(s) investigated by each respondent (A) and the categories of input data (B). . . . 209
A.5 Broad reason for the use of EpiEstim. 210
A.6 Questionnaire responses on package satisfaction and features to be improved 210
A.7 Questionnaire 212
B.1 QA scores (%) across parameters. 222
B.2 Time periods for the infection process 243
B.3 All delays from symptom onset to various endpoints 244
B.4 Delays beyond admission to care. 245
B.5 Meta-analysis for A) incubation period B) onset-to-death and C) onset-to-recovery regardless of QA score, with a species sub-group analysis 246
B.6 Meta-analysis for A) infectious period and B) serial interval regardless of QA score . . . 247
C.1 Daily incidence of SARS-CoV-2 wildtype, Alpha and Delta variants in the 7 NHS regions in England. 266
C.2 Daily incidence of SARS-CoV-2 wildtype, Alpha and Beta/Gamma variants for the 18 ADM2 regions in France. 267
C.3 Effective transmission advantage of Alpha over wildtype in England. 270
C.4 Effective transmission advantage of Alpha over wildtype in England from 1st September 2020 to 14th March 2021. 271
C.5 Effective transmission advantage of Alpha over wildtype for all NHS England regions from 1st September 2020 to 14th March 2021. 272
C.6 Effective transmission advantage of Alpha over wildtype in France 274
C.7 Effective transmission advantage of Alpha over wildtype for all ADM2 regions in France from 18th February to 30th May 2021. 275

C.8 Effective transmission advantage of Beta and Gamma (combined) over wildtype in France 277
C.9 Effective transmission advantage of Beta and Gamma (combined) over wildtype for all ADM2 regions in France from 18th February to 30th May 2021 278
C.10 Effective transmission advantage of Delta over Alpha in England 280
C.11 Effective transmission advantage of Delta over Alpha for all NHS England regions from 15th March to 16th June 2021. 281
C.12 Method performance using simulated data assuming the same natural history for the reference and the variant (using 10 or 20 days of incidence data). 285
C.13 Method performance using simulated data assuming the same natural history for the reference and the variant (using 30 or 50 days of incidence data). 286
C.14 Method performance using simulated data assuming the same natural history for the reference and the variant. 287
C.15 Method performance using simulated data assuming different SI mean for the reference and the variant (using 10 or 20 days of incidence data). 288
C.16 Method performance using simulated data assuming different SI mean for the reference and the variant (using 30 or 50 days of incidence data). 289
C.17 Method performance using simulated incidence data when the mean SI of the variant is different and is misspecified during estimation (using 10 or 20 days of incidence data). . . 290
C.18 Method performance using simulated incidence data when the mean SI of the variant is different and is misspecified during estimation (using 30 or 50 days of incidence data). . . 291
C.19 Method performance using simulated incidence data when the CV of the SI distribution of the variant is different and is correctly specified during estimation (using 10 or 20 days of incidence data). 293
C.20 Method performance using simulated incidence data when the CV of the SI distribution of the variant is different and is correctly specified during estimation (using 30 or 50 days of incidence data). 294
C.21 Method performance using simulated incidence data when the CV of the SI distribution of the variant is different and is misspecified during estimation (using 10 or 20 days of incidence data). 296
C.22 Method performance using simulated incidence data when the CV of the SI distribution of the variant is different and is misspecified during estimation (using 30 or 50 days of incidence data). 297
C.23 Method performance using incidence data simulated with superspreading (using 10 or 20 days of incidence data). 298
C.24 Method performance using incidence data simulated with superspreading (using 30 or 50 days of incidence data). 299

C.25 Method performance using using simulated incidence data with under-reporting (using 10 or 20 days of incidence data). 300
C.26 Method performance using using simulated incidence data with under-reporting (using 30 or 50 days of incidence data). 301
C.27 Method performance using incidence data simulated with time-varying R_t. 302
C.28 Method performance using incidence data simulated with time-varying R_t in two locations (same change). 304
C.29 Method performance using incidence data simulated with time-varying R_t in two locations (different change). 304
C.30 Effective transmission advantage estimated using MV-EpiEstim on simulated data in which the value of the transmission advantage changes over time. 305
C.31 Method performance using incidence data simulated with time-varying ϵ in one location. 305
D.1 R_t estimates for influenza 310
D.2 Comparison of reported versus reconstructed influenza incidence and R_t estimates 311
D.3 R_t estimates for COVID-19 cases 312
D.4 Comparison of reported versus reconstructed COVID-19 incidence and R_t estimates . . . 313
D.5 R_t estimates for COVID-19 deaths 314
D.6 Comparison of reported versus reconstructed COVID-19 deaths and R_t estimates 315
D.7 Reported incidence patterns by weekday 316
D.8 Comparison of the percentage agreement in the classification of R_t estimates 318
D.9 Performance of the method when estimating a constant R_t for 100 simulated epidemics. . 321
D.10 An example of reconstructed daily incidence data (on the log scale) with clear discontinuities on the borders between time periods where the data had been aggregated. 322
D.11 Performance of the method when estimating a time varying R_t for 100 simulated epidemics (sudden decrease in R_t). 324
D.12 Performance of the method when estimating a time varying R_t for 100 simulated epidemics (sudden increase in R_t). 326
D.13 Performance of the method when estimating a time varying R_t for 100 simulated epidemics (gradual decrease in R_t). 328
D.14 Performance of the method when estimating a time varying R_t for 100 simulated epidemics (gradual increase in R_t). 329
D.15 Plotting the gradual change in R_t scenarios for 100 simulated epidemics in the middle of the time window. 330
D.16 Plotting the stepwise change in R_t scenarios for 100 simulated epidemics in the middle of the time window. 331

D.17 The mean weekly sliding R_t estimates generated after each of the 10 iterations from 100 stochastically simulated epidemics 332
D.18 Performance of the method using alternative aggregations of incidence data when the end of aggregation windows align with the step change in R_t. 333
D.19 Performance of the method using alternative aggregations of incidence data when the end of aggregation windows do not align with the step change in R_t. 334
D.20 Comparison between the reconstructed daily incidence generated by the EM algorithm and LOESS smoothing. 336
D.21 R_t estimates made using reconstructed daily incidence generated by the EM algorithm and by LOESS smoothing 338
D.22 The absolute difference between R_t estimates made using reconstructed daily incidence generated by the EM algorithm and by LOESS smoothing 339
D.23 Comparison of R_t estimates obtained using a workaround approach (where a weekly SI is supplied to match the incidence data) and the EM approach. 340
D.24 Supplying reconstructed daily influenza incidence data to EpiLPS. 341

List of Tables

1.1 Examples of diseases, diagnostics and key control measures for each pathogen type . . . 24
1.2 Examples of common disease transmission pathways . . . 25
1.3 Key temporal parameters for transmission . . . 27
2.1 Inclusion and exclusion criteria for the scoping review. . . . 39
2.2 Summary table of the R packages and tool (*) that address some of the issues identified in the scoping review. . . . 46
3.1 Ranges of estimates for basic reproduction numbers by outbreak and estimate type. . . . 60
3.2 Ranges of estimates for epidemiological delays by Ebola virus species. . . . 63
3.3 Ranges of case-fatality ratio estimates by outbreak and country. . . . 69
A.1 Summary table of the papers identified that used an unmodified version of EpiEstim. . . . 205
A.2 Usability of each R package or tool identified in the scoping review. . . . 207
A.3 Summary of the issues and suggestions reported in questionnaire feedback categorised by broad theme. . . . 211
B.1 Outbreaks of Ebola Virus Disease . . . 215
B.2 Inclusion and exclusion criteria . . . 217
B.3 Review papers used in backwards citation screening . . . 218
B.4 Quality assessment questionnaire . . . 220
B.5 PRISMA 2020 Abstracts Checklist . . . 223
B.6 PRISMA 2020 Checklist. . . . 224
B.7 Overview of the total number of parameters extracted within each parameter group . . . 227
B.8 Overview of the total number of parameters extracted for each parameter type . . . 228
B.9 Seroprevalence in countries with history of officially reported EVD cases . . . 229
B.10 Seroprevalence in countries with no history of officially reported EVD cases and multi-country analyses . . . 230
B.11 Risk factors for seropositivity . . . 231
B.12 Attack rates . . . 232
B.13 Basic R estimates . . . 233
B.14 Effective R estimates . . . 236

B.15 Secondary attack rates . . . 238
B.16 Growth rates . . . 239
B.17 Doubling times . . . 240
B.18 Overdispersion table . . . 241
B.19 Risk factors for infection . . . 242
B.20 Risk factors for onward transmission by significance and adjustment. . . . 242
B.21 Case fatality ratios . . . 248
B.22 Risk factors associated with severity . . . 252
B.23 Mutation rates . . . 253
B.24 Summary of EVD model types . . . 254
B.25 Summary of EVD model assumptions . . . 254
B.26 Models overview . . . 255
B.27 Summary of EVD models with code . . . 262
B.28 PERG membership . . . 263
C.1 Estimates of the effective transmission advantage of SARS-CoV-2 Alpha variant over the wildtype for 7 NHS regions in England. . . . 271
C.2 Estimates of the effective transmission advantage of SARS-CoV-2 Alpha variant over the wildtype for 18 ADM2 regions in France. . . . 273
C.3 Estimates of the combined effective transmission advantage of SARS-CoV-2 Beta and Gamma variants over the wildtype in France for 18 ADM2 regions in France. . . . 276
C.4 Estimates of the effective transmission advantage of SARS-CoV-2 Delta variant over Alpha for 7 NHS regions in England. . . . 279
C.5 Parameter values used in the simulations. . . . 283
C.6 Classification of the new variant (as unclear, 'less' or 'more' transmissible) when the true transmission advantage is 1. . . . 302
C.7 Reference R_t values used to simulate incidence in scenarios with time-varying R_t profiles. 303
C.8 Github repositories for packaged code used in 5 of the studies identified in the literature review. . . . 306
D.1 Time taken to estimate R_t in each of the real data scenarios. . . . 319
D.2 Criteria used to assess the performance of our method in the simulation study. . . . 319
D.3 Differences between the re-aggregated reconstructed daily incidence data generated by either the EM algorithm method or LOESS smoothing. . . . 337

Acronyms

AIDS	Acquired Immunodeficiency Syndrome
CDC	Centers for Disease Control and Prevention
CFR	Case Fatality Ratio
CI	Confidence Interval
CRAN	The Comprehensive R Archive Network
CrI	Credible Interval
CRPS	Continuous Ranked Probability Score
CV	Coefficient of Variation
DRC	Democratic Republic of the Congo
ECDC	European Centre for Disease prevention and Control
EM	Expectation Maximisation
EVD	Ebola Virus Disease
FDA	Food and Drug Administration
HIV	Human Immunodeficiency Virus
IFR	Infection Fatality Ratio
IQR	Interquartile Range
JAGS	Just Another Gibbs Sampler
LOESS	Locally Estimated Scatter plot Smoothing
MCMC	Markov Chain Monte Carlo
MERS	Middle East Respiratory Syndrome
MERS-CoV	Middle East Respiratory Syndrome Coronavirus
MRC	Medical Research Council
MRSA	Methicillin-Resistant *Staphylococcus aureus*
NHS	National Health Service
PCR	Polymerase Chain Reaction
PERG	Pathogen Epidemiology Review Group
PRISMA	Preferred Reporting Items for Systematic reviews and Meta-Analyses
RSV	Respiratory Syncytial Virus

SARS	Severe Acute Respiratory Syndrome
SD	Standard Deviation
SE	Standard Error
SGTF	S-gene target failure
SI	Serial Interval
SPI-M-O	Scientific Pandemic Influenza group on Modelling, Operational sub-group
UK	United Kingdom
UKHSA	UK Health Security Agency
US	United States
VOC	Variant Of Concern
VOCs	Variants Of Concern
VOIs	Variants Of Interest
WHO	World Health Organisation

Chapter 1

Introduction

1.1 Background

1.1.1 Infectious diseases

Infectious diseases, also known as communicable diseases, are illnesses caused by pathogens, such as bacteria, viruses, parasites and fungi [1]. According to the World Health Organisation (WHO) Global Health Estimates, infectious diseases and respiratory infections accounted for over 16.2 million deaths globally in 2021 [2]. Of these, COVID-19 accounted for approximately 8.7 million deaths. The burden of infectious diseases is particularly high in low-income countries, where they accounted for eight of the ten leading causes of death in 2021 [3].

An epidemic or 'outbreak' of an infectious disease occurs when there is a sudden increase in cases arising in a population or region compared to what would normally be expected [4]. Outbreaks may result from the emergence of novel pathogens that have not previously circulated, such as SARS-CoV-2, the virus that causes COVID-19. Conversely, they can also involve the re-emergence of known diseases, such as the viral hemorrhagic fever Ebola Virus Disease (EVD), which has caused repeated outbreaks since it was first recognised in 1976, including the 2014-16 West African EVD outbreak which resulted in over 11,000 deaths [5].

Once an infectious disease emerges or re-emerges, it is crucial to rapidly contain the outbreak to prevent the disease becoming endemic, which means that it maintains a relatively stable presence within a population over space and time [4]. For example, malaria is endemic to large parts of Africa, the Americas and Asia, while some diseases, such as tuberculosis, are endemic worldwide [6, 7]. If a disease becomes endemic, it is important to minimise its circulation as much as possible and avoid surges in cases. For example, influenza viruses, primarily types A and B, are endemic throughout the world and cause annual seasonal outbreaks [8]. In temperate regions, these outbreaks typically occur during the winter months, while in tropical regions, they can happen sporadically throughout the year [8].

Outbreaks of infectious diseases are a persistent threat to the health of populations across the world, and the risk of outbreaks continues to increase due to factors such as climate change, rapid urbanisation and globalisation [9]. The scale of these outbreaks can range from a localised surge in cases affecting small groups, such as neighbourhoods, hospitals or schools, to widespread epidemics and global pandemics. Sometimes, pathogens follow predictable patterns, as is the case for seasonal influenza, while others emerge or re-emerge unpredictably [10].

Outbreak control is defined as the implementation of interventions in order to reduce disease incidence, prevalence, morbidity and mortality to a level considered acceptable within a specific area [11]. However, if possible, the ideal outcome of infectious disease response is the elimination or eradication of a disease. Elimination occurs when the incidence of a disease is reduced to zero in a defined geographical area, but interventions will still be required as the disease may be re-introduced and re-establish itself in the population [11]. For example, re-introduction can occur through transmission to humans from zoonotic reservoirs or imported cases from other regions where elimination has not been reached. A well-known example of re-introduction after elimination is measles: after the implementation of successful vaccination programmes, measles had been declared eliminated from 83 countries by the end of 2022 [12]. Elimination in this context meant that there had been no recorded cases of measles in the population for 12 months. Due to low vaccination rates, there has been a widespread resurgence in measles cases, and at the time of writing, outbreaks of measles are occurring in every region of the world [13].

In comparison, eradication is achieved when the incidence of a disease is permanently reduced to zero globally and intervention efforts are no longer needed [11]. For many diseases, eradication is not possible due to a combination of factors, including the lack of effective interventions like vaccines, inadequate surveillance systems to identify cases, insufficient long-term immunity among the population, limited political will or international cooperation, high economic costs and the presence of animal or environmental reservoirs that maintain the pathogen outside of human populations (e.g., the natural reservoir of EVD is believed to be fruit bats) [14, 15]. As eradication is so challenging, only one human disease has been eradicated in history - smallpox [16]. This was achieved through a rigorous control campaign, encompassing targeted surveillance and containment strategies alongside vaccination (universal childhood vaccination or mass vaccination depending on the country) [17].

Since elimination is often temporary and eradication is rarely attainable, many infectious diseases re-emerge periodically or remain endemic, and there is always the risk of new pathogens emerging. Therefore, rapid and effective outbreak response is imperative to mitigate morbidity and mortality in affected populations.

1.1.2 Outbreak response

Outbreaks are not always detected, and even when they are identified, surveillance systems may not be sufficient to monitor their progression. In an ideal scenario, when an outbreak occurs, a systematic investigation would start in order to identify its source, the infectious agent, transmission characteristics and the most effective measures to control the spread of disease [18–20]. The source of an outbreak can either be a 'point source', where there is only a brief period of exposure (e.g. an outbreak of *E.coli* due to contaminated food at a party), 'continuous source', where the exposure is over a longer time period, 'intermittent source', where exposure happens sporadically, or 'propagated', where the pathogen is spread from person to person over a longer period of time [4, 21]. The most common types of pathogen for humans are bacteria, viruses, fungi and parasites (including protozoa and worms) (Table 1.1) [22]. A rarer type of pathogen is prions, which are misfolded proteins that can cause neurodegenerative disease [23]. If a specific pathogen has previously circulated in a population, there may be pre-existing immunity, diagnostics, treatments, or interventions, such as vaccines, which can mitigate the severity of the outbreak. However, novel pathogens may have no established natural history or countermeasures, hindering our ability to rapidly respond.

Table 1.1: **Examples of diseases, diagnostics and key control measures for each pathogen type [4]**

Pathogen Type	**Disease Examples**	**Diagnostics**	**Key Control and Treatment Measures**
Bacteria	Cholera, tuberculosis, chlamydia, Methicillin-Resistant *Staphylococcus aureus* (MRSA)	Polymerase Chain Reaction (PCR), cultures	Antibiotics, isolation, sanitation, vaccination
Viruses	*Caused by DNA viruses:* Acquired Immunodeficiency Syndrome (AIDS), hepatitis B, smallpox, mpox *Caused by RNA viruses:* Influenza, COVID-19, Zika, Ebola Virus Disease (EVD), measles, dengue	PCR, antigen tests	Antivirals, quarantine, isolation, vaccination, vector control
Fungi	Candidiasis, aspergillosis	Microscopy, cultures	Antifungals, limiting exposure, environmental controls
Parasites	Malaria, schistosomiasis, leishmaniasis	Microscopy, antigen tests, PCR	Antiparasitics, vector control, sanitation, mass drug administration
Prions	Creutzfeldt-Jakob disease	Brain biopsy (post-mortem)	Preventing contamination of food products, strict handling, no treatment

The investigation also needs to determine the pathogen's mode of transmission (Table 1.2). The various transmission pathways include direct contact with an infected individual (skin, blood or bodily fluids), sexual transmission, transmission via fomites (i.e., contact with infected surfaces or objects), respiratory pathways (such as airborne aerosol transmission or via droplets), vector-borne transmis-

sion (typically via blood-feeding insects or arthropods, such as mosquitoes, fleas and ticks), zoonotic transmission (from animals to humans), contaminated food or water, and vertical transmission from mother-to-child [4, 24]. Individual pathogens can also have multiple transmission pathways, for example, the SARS-CoV-2 virus is transmitted through direct contact, respiratory droplets, respiratory aerosols, and via fomites [25]. Another example is Ebola virus, which can be transmitted through direct contact with the bodily fluids of infected animals (e.g. during food preparation or consumption) and humans who are currently sick or have died (including blood, saliva and semen). Ebola virus can also be transmitted vertically by infected pregnant women and via fomites contaminated with bodily fluids [26]. Similarly, Zika virus can also be transmitted vertically during pregnancy, via blood or blood product transfusions and sexual contact. However, Zika virus is primarily transmitted via vectors, specifically through the bite of infected *Aedes aegypti* mosquitoes [27].

Table 1.2: **Examples of common disease transmission pathways [4]**

Transmission Pathway	**Definition**	**Disease Examples**
Direct contact	Transmission through direct contact with an infected individual's skin, blood or other bodily fluids.	EVD, mpox, COVID-19
Sexual	Transmission through sexual activity or sexual contact with an infected individual.	Syphilis, chlamydia, AIDS
Droplets	Respiratory transmission via droplet particles at close distance (within 1 metre), typically through coughs and sneezes.	Influenza, COVID-19, common cold.
Airborne	Respiratory transmission via smaller aerosols (generally considered $<5\mu$m in diameter), which can travel longer distances and remain suspended in the air for prolonged periods.	Tuberculosis, measles, influenza, COVID-19.
Fomites	Transmission via contaminated surfaces or objects.	Hand foot and mouth disease, influenza, COVID-19
Vector-borne	Transmission through the bite of an infected vector, typically blood-feeding insects or arthropods (e.g. mosquitoes, fleas, or ticks).	Malaria, Zika, dengue, Lyme disease
Zoonotic	Transmission from animals to humans.	EVD, rabies, plague
Food- or waterborne	Transmission via contaminated food or water.	Cholera, typhoid fever
Vertical	Mother-to-child transmission during pregnancy, childbirth or breastfeeding.	Malaria, AIDS, Zika

Once the pathogen and other clinical characteristics are understood, a case definition can be established [18–20, 28]. The case definition outlines the specific criteria to be met for an individual to be linked to the outbreak, such as symptoms, time period of illness, geographical location and other key characteristics. Using this definition, cases of a disease can be identified through active case detection, using methods like contact tracing and screening programmes, or through passive surveillance of the healthcare system and individuals seeking care. Once identified, cases can be classified as 'suspected',

'probable', or 'confirmed' [29]. Generally, a suspected case would have the typical clinical features of the disease but may be missing an epidemiological link to the outbreak or laboratory confirmation. A probable case meets the same clinical criteria as a suspected case but may have partial laboratory results or an epidemiological link to a confirmed case. If a case is verified through laboratory testing, then they may be considered a confirmed case. These categorisations can be further refined depending on the characteristics of the disease under investigation [4, 30].

Investigators can collect case data into a line list - a dataset where each line corresponds to a known case of a disease, and each column contains specific information such as their date of illness onset or date reported, and their respective characteristics, such as age and sex [28, 31]. Line list data enables epidemiologists to look more closely at the characteristics of individual cases and identify risk factors, for instance, to determine whether a disease disproportionally affects individuals of a specific age group, sex, or socioeconomic group [32, 33]. To identify temporal trends in the data and potential changes in disease spread over time, epidemiologists and public health professionals aggregate case data into counts of daily incidence to perform time series analysis [34]. Incidence data, containing the number of new cases per unit time, is key for 'outbreak analytics', a term which encompasses the methods used to collate, curate, visualise, analyse, model and report outbreak data [35]. During prolonged outbreaks or when surveillance systems are under pressure, the level of detail collected often decreases, leading to more aggregated incidence over time (e.g. weekly or monthly case counts) and space (e.g. district-level instead of village-level) [36–39]. As a result, the resolution of the data diminishes, making it more challenging to capture granular trends [39, 40]. It is crucial to note that, particularly in large or widespread outbreaks, it is common for some cases to go unreported [41]. When analysing incidence data, it is important to consider the likely proportion of missed cases and how this may change over time to ensure a more accurate interpretation of the data.

1.1.3 Understanding transmission dynamics

Once incidence data becomes available, one of the primary tasks for public health professionals is to estimate the pathogen's transmissibility, which refers to how easily the pathogen spreads within the population. Transmissibility is influenced by various factors such as the pathogenicity of the infectious agent, the population's current immunity levels, how quickly infected individuals can infect others, demographics, and social connectivity [42].

Temporal characteristics of transmission

To better understand the transmission timeline, investigators need to estimate the incubation period, latent period, infectious period, generation time and serial interval (see Table 1.3 for definitions) [43].

The relationship between the incubation period and latent period of a disease is an important consideration (Table 1.3). For some diseases, the latent period is the same or longer than the incubation

Table 1.3: **Key temporal parameters for transmission [43]**

Temporal parameter	**Definition**
Incubation period	Time between infection and symptom onset.
Latent period	Time between infection and becoming infectious.
Infectious period	Time between onset of infectiousness and an individual no longer being capable of transmitting the pathogen to others.
Generation time	Time between an individual becoming infected and subsequently infecting someone else.
Serial interval	Time between a case becoming symptomatic and the individual they infect becoming symptomatic.

period, meaning individuals only become infectious after symptoms appear. For example, in diseases like EVD or Marburg virus disease, transmission begins after symptom onset, making it possible to quickly identify and isolate cases in order to minimise onward transmission [15, 44]. In contrast, for some pathogens the latent period is shorter than the incubation period, allowing for asymptomatic transmission. In these cases, individuals may never develop symptoms or spread the infection before starting to show symptoms. For example, the transmission risk of Human Immunodeficiency Virus (HIV) is highest in the first weeks of infection, but some individuals remain asymptomatic until the infection advances to Acquired Immunodeficiency Syndrome (AIDS), a progression that can take months or even years [45]. Asymptomatic transmission can contribute substantially to the spread of disease, as individuals are often unaware that they are infected, less likely to be identified through passive surveillance and left untreated [46–48]. In these instances, active testing is required to detect and manage cases effectively.

For each infectious disease, individuals are capable of transmitting the infection to others for the duration of the infectious period (Table 1.3). Characteristics of interventions such as isolation (separating infected individuals from others) and quarantine (separating exposed individuals from others in case they have been infected) aim to coincide as much as possible with this time period to ensure that infectious individuals are separated from the general population until they are no longer contagious [49]. Pathogens with a long duration of infectiousness, such as HIV and *Mycobacterium tuberculosis* (causative agent of Tuberculosis), can remain transmissible for months to years if left untreated [50, 51]. This extended period of ongoing transmission, particularly in the absence of symptoms, complicates containment efforts.

The generation time and serial interval give an indication of how quickly a disease spreads in a population (Table 1.3). The generation time is influenced by several factors, including the characteristics of the pathogen itself (e.g. how quickly a virus replicates), the host (e.g. pre-existing immunity or behaviour changes), the environment (e.g. population density or the season) and current control measures (e.g. vaccination or quarantine) [52]. Diseases with shorter generation times transmit more

rapidly, such as influenza and COVID-19 [53, 54]. Since the time of infection itself can be difficult to observe directly, the serial interval often serves as a practical proxy [55]. This is because the generation time and serial interval share a similar mean value, although their distributions may differ (i.e., in terms of their variance or skewness), especially when the incubation period is not aligned with the latent period [56].

Understanding these temporal characteristics of transmission is essential for developing and implementing effective control measures. Additionally, they are important parameters in models used to estimate transmissibility (Section 1.1.3) and forecast outbreaks (Section 1.1.4).

Quantifying transmissibility

Two of the most common measures of transmissibility are the growth rate and the reproduction number. The growth rate is the change in the number of new cases per unit time, providing an insight into how quickly an outbreak is growing or on the decline [57]. It can be estimated using simple models (e.g. log-linear models) fitted to incidence data [58].

The reproduction number is the average number of secondary cases caused by an infected individual and has several variations, including the basic reproduction number (R_0) and the effective (or time-varying) reproduction number (R_t) [59, 60]. R_0 is the most simplified form of reproduction number, quantifying the average number of people a primary case would infect in a large, entirely susceptible population [59]. R_0 is most useful at the initial stages of an outbreak to give an indication of a disease's baseline transmission potential prior to the implementation of interventions.

In contrast, R_t represents the average number of people a case would be expected to infect at a specified time point during an outbreak, where the population is not fully susceptible to the disease [60]. This is a more accurate reflection of infectious disease dynamics due to factors such as pre-existing immunity, acquired immunity, changes in behaviour, and interventions.

There are two versions of R_t, either the instantaneous reproduction number or the case or cohort reproduction number (R_c) [60, 61]. R_c measures the average number of people that a cohort of individuals infected at a specific time would go on to infect [61]. R_c therefore requires data after the time for which the reproduction number is to be estimated and is less useful when quantifying transmissibility to inform real-time response. In this thesis, R_t refers to the instantaneous reproduction number, which reflects transmission at a specific point in time based on previous data [60]. As described by Cori et al., the difference between R_t and R_c is analogous to the distinction between the estimated life expectancy for individuals born this year and their actual lifespan [60]. While the actual lifespan of individuals can only be determined retrospectively, once all individuals born this year have passed away, life expectancy can be estimated in the present by assuming that future death rates will remain similar to those observed today. Common methods used to infer R_t in real-time include renewal equation based methods (equation 1.1), which rely on a branching process model. The model assumes that the

incidence of new cases on day t (I_t) can be represented by a Poisson process:

$$I_t \sim Pois\left(R_t \sum_{s=1}^{t} I_{t-s}\omega_s\right) \tag{1.1}$$

where R_t is the instantaneous reproduction number at time t, and the past incidence (I_{t-s}) is weighted by ω_s, the probability mass function of the generation time (the time between infection in a case and their infector) [62, 63]. As stated in Section 1.1.3, the generation time can be difficult to observe directly, and the serial interval often serves as a substitute [55]. Therefore, in addition to incidence data, these methods require reliable information on the generation time (when I_t reflects the number of cases by date of infection) or the serial interval (when I_t reflects cases by date of symptom onset).

Transmissibility is a crucial element in understanding the current state of an outbreak, as it indicates whether case numbers are rising or falling and by how much. An outbreak is considered to be growing when the growth rate > 0, equivalent to $R_t > 1$, and declining if the growth rate < 0, equivalent to $R_t < 1$. In this thesis, I focus on using the reproduction number to quantify transmissibility instead of the growth rate, as it provides greater practical value in outbreak response. The growth rate only reflects changes in the observed outcome of the infection process (i.e., incidence), whereas the reproduction number allows us to better understand the individual level generative process of infection, i.e., the relationship between transmissibility and the timescale of transmission (e.g., generation time) [64]. This is important to explore as different combinations of R_t and the generation time of a disease (e.g., high R_t with a long generation time, versus low R_t with a short generation time) can lead to the same observed incidence initially (and therefore the same growth rate), but have different implications for control. For instance, in scenarios with a high R_t and longer generation time, new generations of cases would appear more slowly, but due to the high transmissibility the epidemic would ultimately result in a higher predicted burden. This suggests that an intervention that targets a larger proportion of the population would be most suitable (e.g., reducing the number of susceptible individuals through a high coverage mass vaccination programme). Whereas, in an equivalent scenario (from the perspective of the growth rate) with a lower R_t and shorter generation time, new generations of cases occur in rapid succession and would require a response capable of interrupting transmission over shorter timescales (e.g., contact tracing to identify and isolate cases more quickly) [64]. The reproduction number also provides important context for designing vaccination strategies. For example, the herd immunity threshold (the minimum proportion of the population that needs to be immune to ultimately stop transmission) is based on bringing the reproduction number below 1, and is not directly related to the growth rate.

Transmissibility estimates are therefore essential for situational awareness and can help with intervention planning and understanding whether current disease control measures are effective or need to

be intensified. Furthermore, these estimates can feed into real-time incidence forecasts, which provide important insights for logistical and resource planning. For example, they can help predict whether hospital bed capacity is likely to be exceeded or identify areas where the allocation of resources should be prioritised [65–68]. Given the numerous applications, accurate and precise transmissibility estimates are essential to ensure timely response, effective interventions and optimal resource allocation.

1.1.4 Outbreak forecasts

Outbreak forecasting has gained increasing prominence, especially during the COVID-19 pandemic, as the need for timely and accurate predictions of the number of future cases, hospitalisations and deaths became critical for managing public health responses globally. Various organisations established collaborative forecasting hubs to model epidemic trajectories, which combine the findings from multiple epidemiological models into ensemble predictions that contributed to shaping policy decisions [69]. Hubs include the United Kingdom (UK)'s Scientific Pandemic Influenza group on Modelling, Operational sub-group (SPI-M-O) and UK Health Security Agency (UKHSA) Epidemiological Ensemble Group (UKHSA Epi-Ensemble) modelling hub, the United States (US) COVID-19 Forecast Hub (a partnership between the US Centers for Disease Control and Prevention (CDC) with the University of Massachusetts Amherst), and the European Centre for Disease prevention and Control (ECDC) respiratory diseases forecasting hub (RespiCast) [70–72]. The ensemble forecasts produced by these hubs combine the strengths of different epidemiological models to enhance the overall reliability of forecasts [73].

Two common types of forecasting models used in epidemiology include branching process models and compartmental models. Branching process models, such as renewal equation based models (Section 1.1.3 & Equation (1.1)) focus on the probabilistic spread of infections and are useful for capturing the stochastic nature of the transmission process. The renewal equation relies on relatively little data for parameterisation: the past incidence, the latest estimate for R_t and either the serial interval or generation time (depending on whether past incidence is by date of infection or date of symptom onset). Using these inputs, future incidence can be projected over a specified time window assuming that R_t remains constant [74]. These models are particularly useful in the early stages of outbreaks when data to reliably parameterise more complex models may be limited. However, the accuracy of these forecasts is highly sensitive to the precision of R_t estimates and the specified serial interval (or generation time), as even small errors can lead to significant discrepancies in projected outcomes [75]. Moreover, given that R_t is assumed to remain the same throughout the forecasting period, they are best suited to short-term (i.e., week ahead) forecasts. Over longer time frames, transmissibility is likely to vary due to factors such as increasing population immunity, behaviour change or the implementation or relaxation of interventions, which cannot be captured simply in branching process models.

In contrast, compartmental models, like the well-known SIR or SEIR model, are often used to

describe long-term transmission dynamics of disease [59, 76, 77]. In the SIR model, the population is divided into three compartments: Susceptible (S), those who can be infected by the pathogen; Infected (I), those who are currently infected and can transmit the disease to others; and Recovered (R), those who have recovered from the disease. The SEIR model introduces an additional Exposed (E) compartment, representing individuals who have been infected but are not yet infectious (due to the latent period). This basic structure can be extended with additional compartments to capture more complex disease dynamics. For instance, compartments can be added to represent asymptomatic cases, hospitalised individuals, vaccinated or unvaccinated individuals, or those who die. Each compartment represents the number of individuals in the population assigned to each group, with the transition between groups driven by either deterministic (no randomness or uncertainty) or stochastic processes (which incorporate randomness and uncertainty). The deterministic form of the SEIR model can be written as the system of ordinary differential equations:

$$\frac{dS}{dt} = \frac{-\beta SI}{N} \tag{1.2}$$

$$\frac{dE}{dt} = \frac{\beta SI}{N} - \sigma E \tag{1.3}$$

$$\frac{dI}{dt} = \sigma E - \gamma I \tag{1.4}$$

$$\frac{dR}{dt} = \gamma I \tag{1.5}$$

Where β represents the infection rate (i.e., the rate of infection per person per time unit), γ is the recovery rate (1 divided by the average duration of infectiousness), σ is the incubation rate (1 divided by the average duration of the latent period) and the total population size, N, is equal to S + E + I + R. The stochastic version draws the number of people transitioning between each pair of compartments from a probability distribution at each time step that the model is run. Therefore, to parameterise and intialise compartmental models, knowledge of the infection rate, recovery rate, incubation rate, the total population size and the initial number of susceptible, exposed, infectious, and recovered individuals are needed. Once parameterised, the models can then be calibrated or fitted to incidence data. Unlike branching process models, R_0 or R_t are not inputs but can be directly inferred from a combination of parameters from the model. In this example, R_0 can be estimated from the ratio:

$$R_0 = \frac{\beta}{\gamma} \tag{1.6}$$

Although compartmental models require more data and involve greater complexity, they can ac-

count for population saturation as individuals become immune, directly influencing the trajectory of disease cases over time. They also allow for detailed scenario modeling, which enables public health professionals to explore the potential impacts of various interventions, such as vaccination campaigns, social distancing or quarantine measures [78]. Compartmental models can simulate how changes in disease characteristics, like increased transmissibility or the emergence of a new variant, might alter the course of an epidemic. This makes compartmental models valuable tools for forecasting, planning resource allocation, and assessing the likely outcomes of different public health strategies in various hypothetical situations. Despite their strengths, the robustness of compartmental models depends heavily on the quality of input data, making reliable parameterisation essential.

1.2 Motivation

Epidemic and pandemic preparedness is a key priority of national and international health organisations [79–81]. Effectively preparing for the next outbreak of a newly emerging or re-emerging pathogen requires strengthening our capacity for rapid response [82, 83]. Central to this effort is the active development of shareable analytical tools that are broadly applicable in a range of disease contexts. Many real-time analyses are repeated multiple times within the same outbreak (e.g., to monitor variations in parameters such as R_t over time) and between outbreaks. To make these analyses more efficient and reusable, methods are typically implemented as code in widely-used programming languages such as R, with code packaged (e.g. in R packages), documented, and distributed as open source in public repositories such as The Comprehensive R Archive Network (CRAN) or GitHub [84–86]. This ensures that other researchers can access and apply these tools for consistent analyses, but only if they are robust, intuitive, user-friendly and well-documented. In this thesis, I will focus on outbreak analytics packages in the R programming language, as it is currently the most widely used in the field of infectious disease epidemiology [35, 87]. I will explore the use of tools designed to address some of the key questions that arise during outbreak response (Figure 1.1), evaluating their strengths and limitations to identify opportunities to develop robust solutions to common challenges.

Once an outbreak has been detected in a population and incidence data has been made available, one of the first questions to answer is *"how transmissible is this pathogen?"* using measures such as R_t (Figure 1.1 A & B). As discussed, R_t is important for understanding the current state of an outbreak (Section 1.1.3) and for parameterising forecasting models to predict future epidemic trajectories (Section 1.1.4). Software tools to estimate R_t are most useful if they can be applied in diverse settings and contexts, however specific data format requirements (e.g., daily incidence data) or the unavailability of necessary input data for parameterisation (e.g., the serial interval) can limit their applicability.

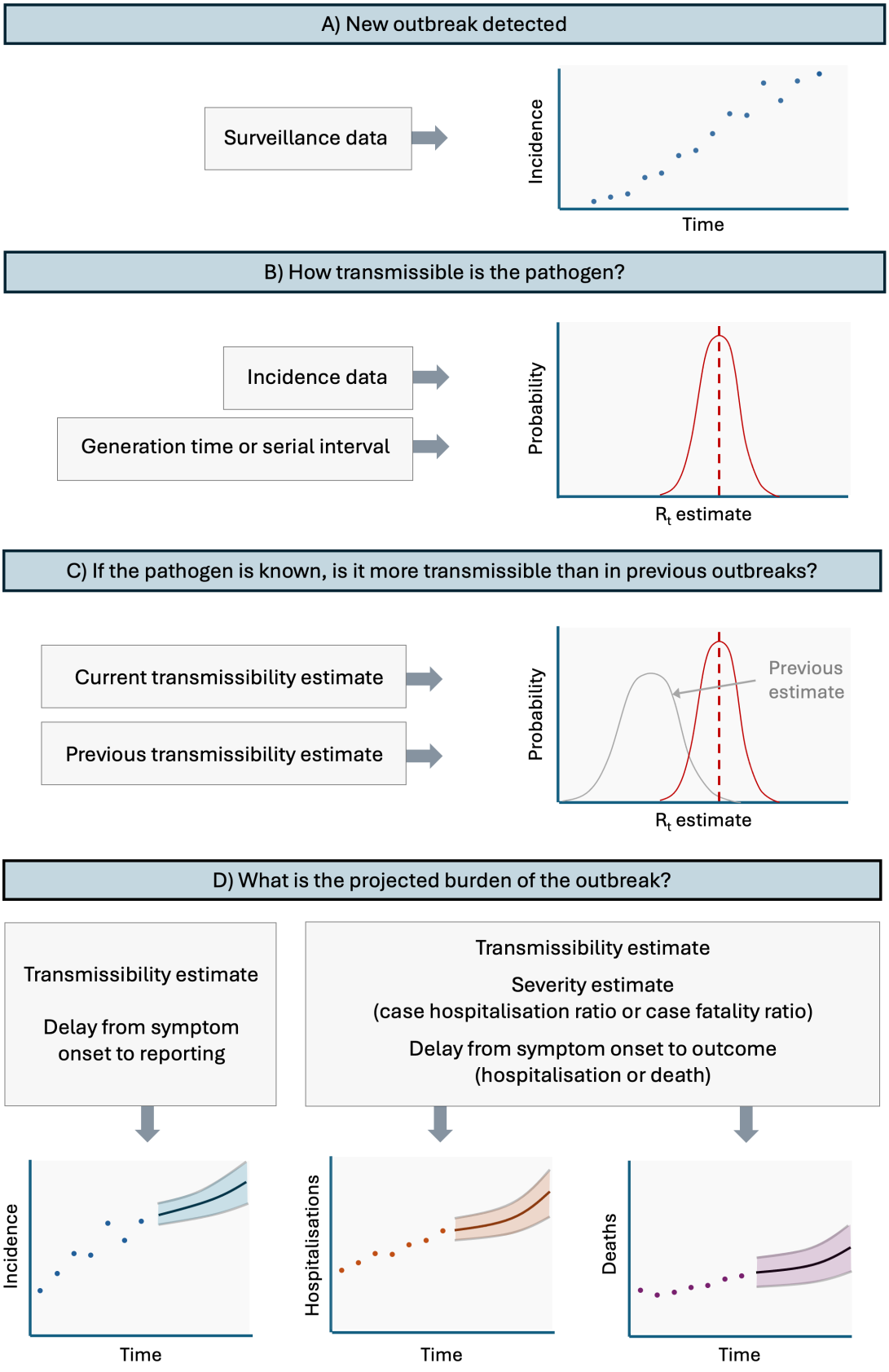


Figure 1.1: **Schematic of the outbreak analytics process.** Each box with an arrow corresponds to the necessary inputs required at each stage.

The R package EpiEstim is one of the first and most widely used R packages developed to estimate R_t in real-time [62]. The COVID-19 pandemic led to a significant increase in the development and applications of new and existing methods, such as EpiEstim, for estimating R_t. Despite some advancements being made, gaps remain in fully optimising the use of these tools, and new challenges have arisen due to the sheer number of tools being developed without much coordination, making it difficult for researchers to identify the most appropriate methods for their specific needs. To improve their practical utility, it is crucial to investigate the current landscape of R_t estimation methods and identify gaps in development that, once addressed, would enhance their applicability in future outbreaks.

When a pathogen re-emerges (e.g., EVD), another key question is *"how does this outbreak compare to previous outbreaks?"*. This may involve comparing the current R_t estimate to those from past outbreaks to determine whether the pathogen is more transmissible (Figure 1.1 C), or evaluating whether severity estimates and other characteristics of the disease, such as the incubation period and the serial interval, are different to what we have seen before. Such comparisons can also be made between outbreaks of a new pathogen and other known pathogens belonging to the same family. For example, SARS-CoV-2 has been compared to SARS-CoV-1 (the coronavirus causing Severe Acute Respiratory Syndrome (SARS)) and Middle East Respiratory Syndrome Coronavirus (MERS-CoV), since all three belong to the same family and genus of viruses [88, 89]. Rapidly assessing whether the current outbreak is worse than usual requires that data from previous outbreaks are readily accessible and kept up-to-date (e.g., in open source databases). In the early stages of the 2014-16 EVD outbreak, analyses were conducted to understand why this outbreak far exceeded the scale of all previous EVD outbreaks combined [65]. These analyses suggested that transmissibility estimates and the clinical timeline of infection were consistent with previous outbreaks, indicating that the scale of the outbreak may be due to inadequate control measures and specific characteristics of the affected populations, including the high degree of cross-border mobility in Guinea, Liberia and Sierra Leone [65]. Therefore, timely access to historical estimates of transmissibility and other key transmission characteristics is crucial to guide effective responses by identifying the factors driving new outbreaks.

If new pathogen variants emerge within the same outbreak, this question changes to *"how does this new variant compare to currently or previously circulating variants?"*. For COVID-19, numerous new variants and sub-variants of SARS-CoV-2 emerged sporadically throughout the pandemic [90, 91]. Some variants showed early signs of increased transmissibility, more severe disease or immune evasion and were classified as Variants Of Concern (VOCs) by the WHO, including Alpha, Delta, Gamma and Omicron [91, 92]. Rapidly characterising the transmission advantage of new variants is critical for informing effective real-time response. Many studies developed methods to estimate the transmission advantage by integrating evidence from multiple data sources, including genomic data [93–101]. These approaches require substantial time and expertise to implement and are specific to each dataset and context. There is a need to develop simple tools that incorporate less data to estimate

the transmission advantage of pathogen variants in real-time, yet still arrive at the same conclusion as these more complex methods. By removing the requirement for extensive resources and specialised knowledge, such tools would be more accessible and widely applicable across different contexts.

Once transmissibility has been characterised, the next step is to determine *"what is the projected burden of the outbreak?"* using forecasting models (Figure 1.1 D). As discussed in Section 1.1.3 and Section 1.1.4, it is critical that models are reliably parameterised in order to ensure accurate and precise estimates of transmissibility and the future burden of the disease, including the incidence of cases, or alternative measures, such as hospitalisations and deaths. R_t estimates depend heavily on the correct specification of the serial interval or generation time, as errors in these parameters can result in biased R_t estimates, ultimately leading to bias in forecasts [102]. In addition to R_t, branching process forecasting models require reliable data on delays, including the time between symptom onset and reporting and symptom onset to hospitalisation or death, depending on the outcome of interest.

Given their complexity, compartmental models require further data and can be parameterised to align with the best available estimates of key natural history parameters, such as the duration of infection and the latent period (Section 1.1.4). Scenario modelling, which explores hypothetical situations such as the impact of changes in intervention policies, requires further assumptions. For example, modelling the rollout of a vaccine that decreases the transmission rate requires reliable assumptions about the relative difference in infectiousness of vaccinated and unvaccinated individuals.

For re-emerging pathogens, systematic reviews of pathogen transmission characteristics from previous outbreaks, when compiled into open source databases and analysed through meta-analyses, provide a valuable resource for parameterising models. This is particularly important for diseases that are deemed by the WHO to be the greatest risk to public health due to their epidemic potential, such as EVD [103]. Accurate parameterisation is especially problematic for emerging pathogens, where historical data is lacking and key parameters are unknown. During the early stages of an outbreak caused by a novel pathogen, researchers often rely on the estimated transmission characteristics of closely related pathogens. For example, when SARS-CoV-2 emerged, initial estimates for parameters such as the incubation period were approximated using estimates from SARS-CoV and MERS-CoV [89]. These resources allow for rapid and informed responses when new outbreaks occur, drawing upon previous experience and data.

In summary, the motivation for the projects presented in this thesis (see Section 1.3) is the critical need to evaluate and improve the tools and methods developed to answer the key questions of outbreak response. Overall, this thesis intends to contribute to outbreak preparedness by developing solutions that support rapid, data-driven decision-making in real-time.

1.3 Aims and objectives

1. **Assess the strengths and limitations of current software tools for estimating the time-varying reproduction number (R_t)**

   (a) Conduct a scoping review using a widely used R package for R_t estimation, EpiEstim, [104] as a case study to evaluate the contexts in which R_t estimation methods have been applied during real-time outbreak response (Chapter 2).

   (b) Survey EpiEstim users to identify the key challenges faced when estimating R_t in real-time (Chapter 2).

   (c) Synthesise the findings from the scoping review and user survey to identify unmet needs and priority areas for development, which would enable broader applicability of these methods in real-time outbreak response (Chapter 2).

2. **Address issues in model parameterisation to facilitate preparedness and rapid response**

   (a) For EVD, a WHO priority disease, systematically review the available literature on R_t, the serial interval, generation time and other key transmission parameters (Chapter 3) [103].

   (b) Establish a database of key EVD transmission parameters and store them in an open source R package to facilitate their use in R_t estimation and outbreak modelling (Chapter 3).

   (c) Perform a meta-analysis of suitable parameters to obtain optimal estimates for use in outbreak models, enhancing preparedness for future EVD outbreaks (Chapter 3).

3. **Address data-related challenges that impact the applicability of outbreak analytics tools in certain contexts**

   (a) Develop a software tool to estimate the transmission advantage of pathogen variants in real-time without the need for genomic sequencing data (Chapter 4).

   (b) Assess the performance of this tool using simulated data and data for COVID-19 (Chapter 4).

   (c) Develop an algorithm to reconstruct daily incidence data from any temporal aggregation of reported incidence data (Chapter 5).

   (d) Using simulated data and real-world data for COVID-19, influenza, and Zika virus disease, assess whether reconstructed incidence data provides comparable or superior accuracy to reported data in estimating the reproduction number (R_t) (Chapter 5) and generating reliable forecasts (Chapter 6).

Chapter 2

Real-time estimation of the epidemic reproduction number: Scoping review of the applications and challenges

The contents of this chapter have been adapted from a manuscript previously published in PLOS Digital Health [105]. I conducted all work presented in this chapter, with input and feedback from my supervisors.

In this chapter, I investigate the current landscape of R_t estimation methods using the popular R package, EpiEstim, as a case study. A scoping review, complemented by insights from a small EpiEstim user survey, highlight issues with the current approaches, including the quality of input incidence data, the inability to account for geographical factors, and other methodological issues. I summarise the methods and software developed to tackle the problems identified, but conclude that significant gaps remain which should be addressed to enable easier, more robust and applicable estimation of R_t during epidemics. The findings from this chapter provide the foundation for the subsequent research projects detailed in this thesis.

2.1 Introduction

As discussed in the main introduction (Section 1.1.3), transmissibility quantifies how easily a pathogen can spread through a population and can depend on numerous factors, such as the pathogenicity of the infectious agent but also the current level of immunity, demographics, and connectivity in the population. Transmissibility is typically measured by the time-varying (or effective) reproduction number (R_t), which represents the average number of secondary infections generated by a case at time t of an outbreak [62, 74, 106, 107]. R_t is an important element of outbreak analysis as it indicates whether case numbers are rising ($R_t > 1$) or falling ($R_t < 1$) and by how much. This can guide

intervention planning and help to determine whether current control measures are effective, and if not, to what extent they need to be intensified. R_t estimates can also feed into real-time incidence forecasts, which can assist logistics and resource planning, for instance, by identifying whether hospital bed capacity is likely to be exceeded or if resources need to be allocated to specific areas [68, 108–110].

As stated in Section 1.1.3, many methods have been developed to estimate R_t in real-time; one of the most commonly used is through the renewal equation (Equation (1.1)). Shown again here for clarity, the renewal equation relies on a branching process model. The model assumes that the incidence of new cases on day t (I_t) can be represented by a Poisson process:

$$I_t \sim Pois\left(R_t \sum_{s=1}^{t} I_{t-s}\omega_s\right) \tag{1.1}$$

where R_t is the time-varying reproduction number (i.e. the average number of cases caused by a primary case infected at time t, assuming that conditions remain the same after time t), and the past incidence (I_{t-s}) is weighted by ω_s, the probability mass function of the generation time (the time between infection in a case and their infector) [62, 63]. In practice, as infection itself is difficult to observe, the incidence of symptomatic cases can be used instead and ω_s can be approximated by the serial interval (SI, the time between symptom onset in a case and their infector. See Britton and Scalia Tomba for the implications of this approximation) [111]. Cori *et al.* developed a method to estimate R_t from the renewal equation that is suitable for real-time application [104]. The method is implemented in the R package EpiEstim, where estimation is performed over user-defined time windows within which R_t is assumed constant. Longer windows typically lead to smoother estimates, but may artificially hide some of the temporal variability [112].

R_t estimates can then be used to project forward in time and forecast future epidemic trajectories (using Equation (1.1), see Figure 2.1), as implemented in the R package projections [113].

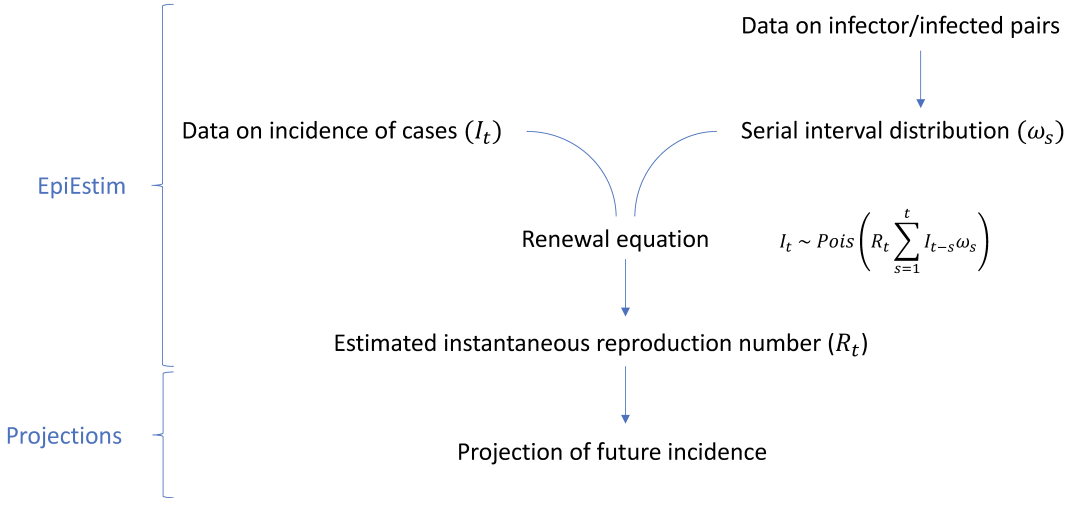


Figure 2.1: **Schematic of the forecasting process using EpiEstim and Projections.**

The COVID-19 pandemic has generated a surge in methodological and tool developments to estimate R_t, as well as numerous applications of these methods. We aimed to describe the landscape of methods and real-world applications of R_t estimation, and to identify potential gaps that could be addressed to make those methods more easily applicable and more useful in practice in the future.

EpiEstim has been widely used for R_t estimation during the COVID-19 pandemic and has been shown to perform better than many other methods in terms of estimation accuracy [112]. Therefore, we use EpiEstim as a starting point to identify ways in which renewal equation methods have been used and modified [112]. We reviewed research articles that either used EpiEstim (the method or the software) or described a modified approach to address unmet needs. Alongside this scoping literature review, a questionnaire was distributed to known users of EpiEstim and advertised on social media to gather information on issues that were unlikely to be described in publications, e.g. computational speed or usability. By collating the findings, we aim to reveal some of the key challenges when estimating R_t in real-time and those that are yet to be addressed.

2.2 Methods

2.2.1 Scoping review: Search strategy and selection criteria

Google scholar was used to identify all articles or reports up to 10th December 2020 that cited one of the two papers describing EpiEstim (Cori et al. 2013 or Thompson et al. 2019) [62, 114]. After full text screening, two databases were compiled using the inclusion and exclusion criteria outlined in Table 2.1, collating papers or reports that used: i) an unmodified version of the EpiEstim method or software and ii) a modified version of the EpiEstim method or software.

Table 2.1: **Inclusion and exclusion criteria for the scoping review.**

Inclusion	**Exclusion**
• Cited Cori et al. 2013 [62] or Thompson et al. 2019 [114]. **Unmodified papers** • Did not mention any alterations to the original method. **Modified papers** • Described altering the approach e.g., used a method that is "similar to..." or "modified from..." EpiEstim.	• Did not implement the method or a modified version of the method. • Updated pre-print (by 31st March 2021) that no longer cites the paper(s). • Duplicate (e.g. updated pre-print versions with different names). • Withdrawn pre-print.

2.2.2 Evaluation of R package or tool usability

Each R package or tool identified in the scoping review was appraised in terms of its ease of installation, available documentation (e.g. vignette or tutorials) and speed of estimation of the reproduction number (see Appendix A.3 and Table A.2 for more detail).

2.2.3 Questionnaire

We designed an online questionnaire including 16 questions aiming to reveal the key challenges users encountered when using EpiEstim to estimate R_t. It was shared via Twitter and distributed to a list of 41 known EpiEstim users [115]. The questionnaire was live between 11th and 31st January 2021.

2.2.4 Ethics statement

Ethical approval was not required as the questionnaire was evaluative and solely to understand the current use and limitations of the EpiEstim R package. Respondents were informed that their survey responses were being collected for academic purposes.

2.3 Results

281 papers were included, which cited the original Cori et al. 2013 paper or the paper by Thompson et al. 2019 describing the latest version of EpiEstim (Figure 2.2) [62, 114]. Papers investigated a variety of diseases; the most common was COVID-19, followed by Ebola and influenza (Figure A.1 in Appendix A). The majority (n = 242) had used an unmodified version of EpiEstim (Table A.1 in Appendix A), and 54 papers described at least one modification [110, 116–166]. Fifteen papers fell into both categories i.e., used the original EpiEstim method as a baseline to compare a new approach to [116–130].

We classified the 106 issues addressed in the 54 papers proposing modifications into 13 categories (Figure 2.3). The most common included delays in the reporting of cases or missing data (n = 20), weekly administrative noise (n = 11), choice of prior for R_t (n = 15), and modelling different regions at the same time (n = 13). Each modification type was further categorised into four themes: 1) incidence data, 2) other input data or method modification 3) extension to account for geographical or spatial factors, and 4) practical or logistical issues.

2.3.1 Modifications involving incidence data

The accuracy of R_t estimates directly depend on the quality of the input data, particularly the incidence data, which can be influenced by several factors. Incidence by date of infection is best suited to measure real-time transmission, but it is rarely available. A common workaround is to use incidence by date of symptom onset, resulting in "delayed" R_t estimates, representing transmission one incubation period

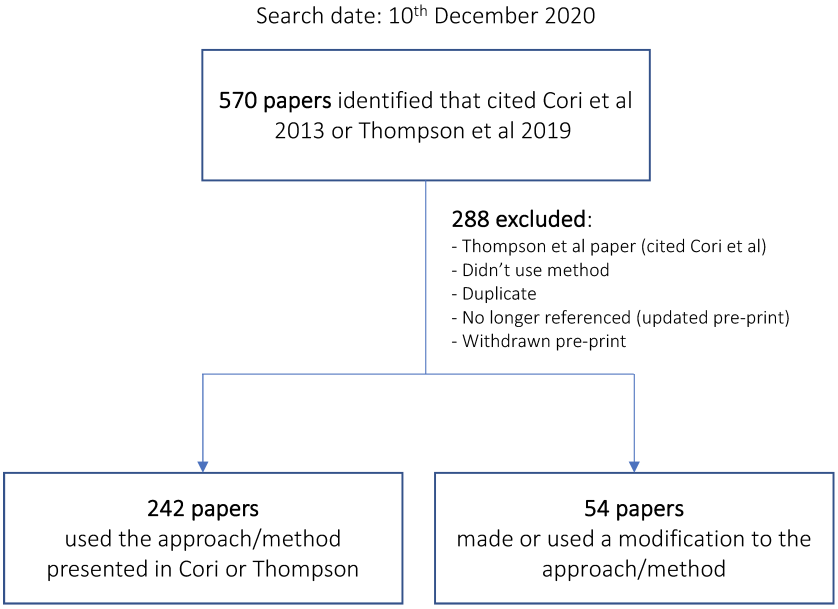


Figure 2.2: **Schematic to illustrate the scoping review process.** Note: fifteen papers are counted as both unmodified and modified (i.e., EpiEstim was used as a baseline comparator to a modified approach).

before time t [62]. There are often additional delays between symptom onset and subsequent reporting of cases, leading to right censoring when observing cases in real time [62]. Recent incidence observed in real-time is therefore often underestimated, which can bias R_t estimates if not accounted for. This is further complicated when reporting delays vary over time, e.g. at the start of an outbreak when robust reporting systems are not yet in place. Beyond issues related to delays in reporting, some cases may not be reported at all; however this should not bias R_t estimates, unless reporting varies over short time scales e.g. if testing capacity fluctuates [62].

Thirty percent of modifications (n = 32) addressed issues with incidence data. The most common was to account for reporting delays or missing data (i.e. underreporting or cases with missing date information, n = 20).

The R packages bayEStim, EpiNow, and EpiNow2 (which is also used to generate R_t estimates for routine forecasting on the website EpiForecasts [151, 153]) all allow back-calculation of the incidence of symptom onset or infection from the incidence of reported cases using a reporting delay or incubation period distribution [118, 125–127, 131, 147, 151, 153, 154, 167, 168]. Such delays can be informed by individual case timelines or by characterising the delay after which reported incidence stabilises. Huisman et al. deconvolve over the observed incidence time series with a delay distribution that is specific to the type of case detection (e.g. positive test, hospital admission, death) [120]. Li et al. use region and time dependent delay distributions [123].

Some papers adjusted their approach to explicitly account for underreporting [120, 123, 130, 141, 142]. The R package Epidemia does so through a time-varying reporting rate [137, 146, 169]. A commonly adopted approach to circumvent time-varying case reporting is to use the incidence of deaths,

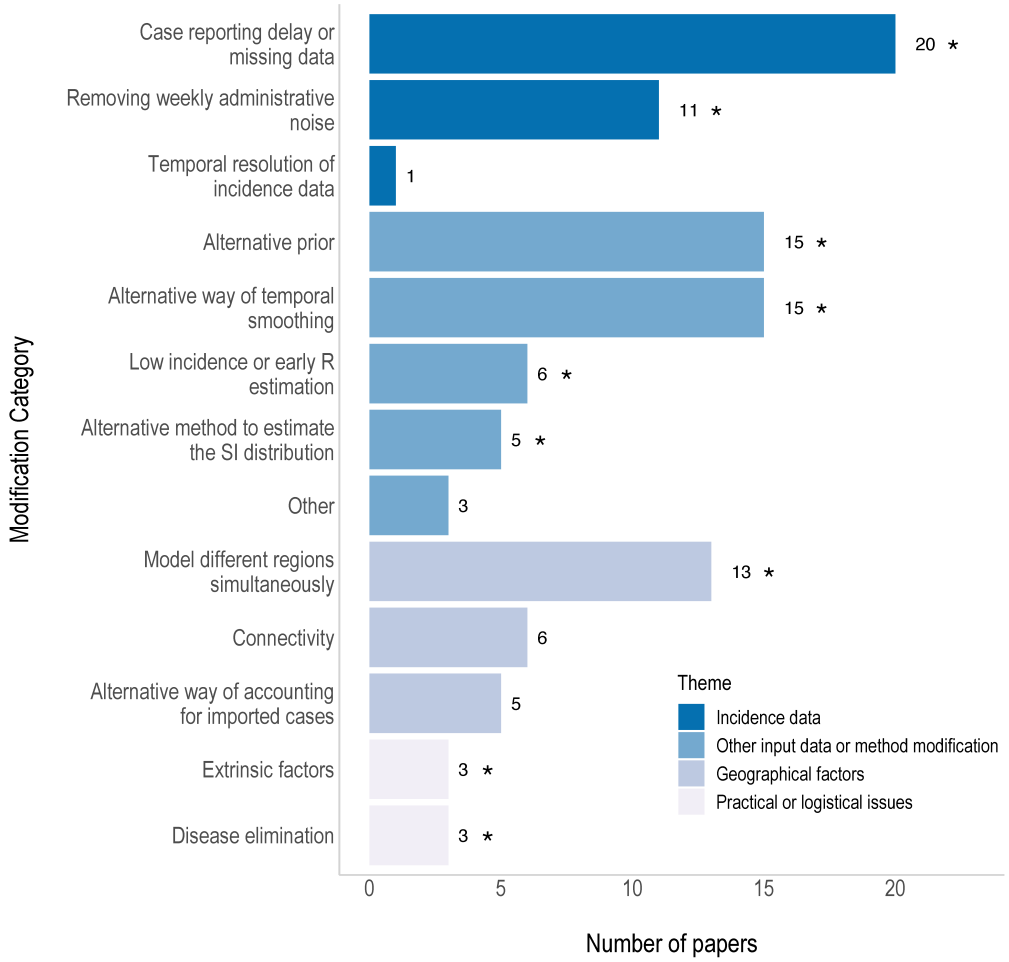


Figure 2.3: **Summary of the modifications made to the EpiEstim method in 54 papers.** Papers which made multiple modifications are counted more than once. Asterisks are present next to the modification category if an R package or opensource tool has been identified that addresses an issue within that category. See text for a description of all modifications and Table 2.2 for a summary of the R packages identified.

as reporting of deaths is usually less prone to time fluctuations than reporting of cases. Nouvellet et al. and Djafaara et al. used COVID-19 death data to estimate R_t directly using the renewal equation [117, 158]. Others have estimated R_t from the incidence of infection or symptom onset, back-calculated from the observed incidence of death [122, 144]. Epidemia and EpiNow2 use a similar approach, and estimates can be based on incidence by date of infection, hospital admission or death [118, 125–127, 131, 137, 146, 147, 151, 153].

The second most common modification was adjusting for weekly administrative noise (n = 11), where reporting can vary depending on the day of the week e.g. due to fewer tests being performed or longer processing times on the weekends [170]. As the default, EpiEstim estimates R_t over weekly time windows to smooth out these temporal trends. Modifications (including in EpiNow2) have been made to account for weekly noise in the incidence data by explicitly including a day of the week effect [118, 125–127, 131, 134, 135, 147, 151, 153, 155].

One paper adjusted the temporal resolution of incidence data. Currently, EpiEstim requires the incidence data and the SI distribution to be on the same time scale. For acute infections, where the SI is short, this means weekly incidence is not supported and daily incidence is required as an input. Yamauchi et al. use EpiEstim on daily incidence inferred from weekly observations using smoothing splines [165].

2.3.2 Other input data or method modification

As the default, EpiEstim uses a gamma distributed prior for R_t, with a mean and standard deviation of 5. Such a high mean value avoids estimating a subcritical reproduction number ($R_t < 1$) unless the data supports it, but can lead to overestimation of R_t in periods with low incidence. Priors for R_t are fixed across all time windows, but the length of the time window, which can be varied, determines the smoothness of the resulting estimates. Many methods have addressed the problem of temporal smoothness using alternative, often autoregressive, prior specifications [118, 124–128, 130, 131, 137, 146, 147, 151, 153, 162, 163]. Turbé et al. updated the gamma distributed prior for each time window with the previous posterior, so that estimates benefit from the information that becomes available over time [128]. Other methods, including those implemented in Epidemia, EpiNow2, and another opensource tool, EpiFilter, can control the temporal variation in R_t using random walks, such as gaussian processes [118, 124–127, 130, 131, 137, 146, 147, 151, 153, 162, 163, 171]. In Epidemia, in addition to being smoothed using a random walk, R_t can also be constrained by covariates such as mobility, and these covariates themselves are given priors [137, 146].

Overall, fifteen papers described modifications related to the temporal smoothing of successive R_t estimates, which in EpiEstim involves a subjective time window choice over which each R_t is estimated. Choosing a time window length is a trade-off between identifying important temporal trends and smoothing out random noise in the data. Pathogens with short SIs and/or in the context of fast-changing transmission (e.g. due to control measures) will require shorter time windows to capture the former, while in low incidence settings, longer time windows may be needed to encompass more data. Four papers use metrics (either the Accumulated Prediction Error, implemented in the R package called APEestim, or the Fisher Information Approximation) to identify the optimal time window [119, 124, 159, 160, 172]. Instead of relying on a time window, methods which can control temporal variation using random walks can incorporate parameters which control the autocorrelation of successive R_t estimates [118, 124–127, 130, 131, 147, 151, 153, 162, 163].

Six papers addressed issues with R_t estimation when incidence is low, where EpiEstim would typically return the prior [121, 124, 143, 148, 149, 162]. The R package earlyR is the frequentist equivalent of EpiEstim, and therefore returns estimates that are only data-driven, but it is not set up to estimate time-varying reproduction numbers [121, 143, 148, 149, 173]. EpiFilter provides more robust estimates of R_t in low-incidence settings by using a retrospective smoothing approach that

combines the strengths of the Cori et al. and Wallinga and Teunis methods [61] (see Section 1.1.3). EpiFilter treats R_t as a latent variable and uses Bayesian inference to condition on both past and future incidence data, producing stable, low-variance estimates. The conditioning on future incidence means that this approach can only be applied retrospectively. However, EpiFilter can also support real-time estimation through a filtering approach that uses only past data. Compared to EpiEstim and APEestim, it does not destabilise as easily in periods of low incidence and does not over-rely on the prior [124, 162].

In EpiEstim, the SI distribution can be specified as parametric (where only the mean and standard deviation are required), non-parametric (where the whole discrete distribution is provided), or "uncertain", where a range of potential SIs are considered. In 2019, Thompson et al. extended the framework so that the SI distribution can be directly estimated within EpiEstim from data on known pairs of primary and secondary cases [114]. Five papers propose alternative approaches to derive the SI that they subsequently use in EpiEstim or another R package [133, 143, 145, 150, 166]. Al-Wahaibi et al. used an alternative R package, Epicontacts, to estimate the SI distribution directly from contact tracing data [133, 174]. Donnat and Holmes effectively assume a linearly decreasing probability mass function for the SI with parameters optimised based on past incidence data [143]. Knight et al. explored a method to permit distributions that allow negative SI values, which occur when transmission can take place before symptom onset as e.g. for COVID-19 [150]. For applicability to vector-borne disease, some approaches modelled the SI as a distribution for both the human and vector [145, 166].

Less common methodological modifications included aiming to improve the quantification of uncertainty using a Bayesian filtering method [139] or jointly estimating the reproduction number and the early incidence to account for highly variable reporting over time [74, 110].

2.3.3 Modifications involving geographical factors

The simple renewal equation detailed in Equation (1.1) describes transmission in a closed population in a single location. Thirteen papers developed or used methods that allow R_t to be estimated simultaneously for different locations. EpiNow2 simply adds the functionality to input incidence for multiple locations at once in a single function call and runs the same model to estimate R_t independently for each region, producing a single output containing the region-specific estimates and facilitating comparisons between locations [118, 125–127, 131, 147, 151, 153]. Abry et al.'s approach enables simultaneous estimation of R_t from incidence time series in different locations, either independently or jointly using spatial smoothing [132]. Some approaches model transmission in different regions using linear predictors (which can include environmental, demographic, or intervention related factors), as in Epidemia, where parameters can be fixed or varying by location [137, 146, 155, 163].

Six papers further modified the approach to explicitly account for connectivity [136, 140, 152, 157, 158, 165]. To account for connectivity between individuals or locations, Ng et al. weighted

the probability of transmission between two cases depending on the distance between them, [157] while other studies inferred connectivity between locations using a gravity model approach [136, 140]. Chirombo et al. developed an estimator for a spatiotemporal R that accounts for susceptibility and infectiousness levels in each location and connectivity between them [140]. Backer et al. identified areas with a key role in disease spread by combining estimates of location-specific R_t with estimated movements out of each location to other areas [136].

Transmissibility may also be constrained by explicitly using mobility data [152, 158, 165]. Yamauchi et al. used a similar approach to Backer et al. but used census data to inform movement between jurisdictions [136, 165]. In a less specific way, several studies examined the extent to which reduction in the level of mobility due to social distancing led to reductions in R_t [152, 158].

A simpler approach to estimate R_t for a location whilst accounting for cases arising elsewhere is to distinguish between imported and locally acquired cases, as proposed in the EpiEstim extension by Thompson et al [114]. Five other papers described similar modifications to account for imported cases or for an exogenous force of infection [138, 147, 156, 162, 164]. Bourhy et al. extended the renewal equation to incorporate the rate of introduction of rabies into a city [138]. Two papers estimated R_t separately for locally acquired versus overseas acquired cases in order to model the effects of different interventions targeted at each group [147, 164].

2.3.4 Modifications with practical or logistical applications

R_t estimation is often used as the basis for practical decision making in response to changing transmissibility in real-time. For example, the implementation of interventions, such as social distancing measures, estimating the adequacy of current control measures, or deciding when to declare that an outbreak has ended [116, 175–177]. Despite the clear link between R_t estimation and policymaking, very few modifications (let alone R packages) have been made to directly answer important practical or logistical questions.

Modifications were made to estimate the impact of potential drivers of transmission on R_t estimates (n = 3). Wang et al. modelled the effects of interventions by incorporating additional parameters to explicitly account for the impact of interventions on the infectious period or the overall level of infectiousness of infected cases [129]. Epidemia accounts for extrinsic factors, such as interventions and changes in mobility, by creating a linear predictor which includes the factors influencing changes in transmission [137, 146].

Three papers sought to develop probabilistic methods to predict local disease elimination or the risk of epidemic resurgence, one of which is implemented in EpiFilter [116, 161, 162].

A summary of the R packages and software tool that incorporate some of the modifications identified through the literature review is available in Table 2.2. We also performed additional exploration to assess their usability and the speed of estimation of the reproduction number (when relevant examples

were provided). This is a non-comprehensive assessment based on the R packages/tools available at the time of writing (see Table A.2, Appendix A.3 and the documentation of each package for more details). Two of the seven were very easy to install, and the majority (n = 5) provided documentation and vignettes. Of the R packages/tool that estimate the reproduction number, three were able to perform the estimation (using provided examples) in under a minute, however, only earlyR scored full marks in the usability category.

Table 2.2: **Summary table of the R packages and tool (*) that address some of the issues identified in the scoping review.** The ticks (✓) indicate whether an R package includes that modification type and rows that are grey show the modification types that have not been incorporated into a known R package or tool. The final three rows summarise additional exploration by the authors to assess how easily each package/tool can be installed and used. For full details of how each classification (very good = ✓✓, good = ✓, poor = X) was defined, see Table A.2 in Appendix A and the associated text (Appendix A.3).

Theme	Modification type	APEestim *(v 0.0.1)*	bayEStim *(v 0.0.1)*	earlyR *(v 0.0.5)*	epicontacts *(v 1.1.2)*	Epidemia *(v 1.0.0)*	EpiFilter* *(*Tool)*	EpiNow2 *(v 1.3.2)*
Incidence	Case reporting delay or missing data		✓			✓		✓
	Removing weekly administrative noise							✓
	Temporal resolution of incidence data							
Other input data or method modification	Low incidence or early R estimation			✓			✓	
	Alternative prior					✓	✓	✓
	Alternative way of temporal smoothing	✓				✓	✓	✓
	Alternative method to estimate the SI distribution				✓			
Geographical factors	Connectivity							
	Model different regions simultaneously					✓		✓
	Alternative way of accounting for imported cases							
Practical/ logistical	Extrinsic factors					✓		
	Disease elimination						✓	
Additional exploration (✓✓ = very good, ✓ = good, ✗ = poor)								
Usability	Ease of installation	✓	✗	✓✓	✓✓	✗	✗	✓
	Documentation and tutorials (e.g. vignette)	✗	✗	✓✓	✓✓	✓✓	✓✓	✓✓
	Speed of R estimation◇ (◇except for epicontacts, which estimates the SI)	✓✓	NA	✓✓	✓✓◇	✓	✓✓	✗

2.3.5 Questionnaire

The questionnaire had 17 responses from 11 countries (Figure A.2). The majority of respondents were academic researchers, public health consultants or health professionals (Figure A.3). The small sample size makes it difficult to statistically analyse the responses, but overall, the questionnaire respondents echoed the broad themes identified during the scoping review (Figure 2.3 and Table A.3). In addition, three other themes were identified: usability, speed and compatibility (Table A.3).

The most common issue affecting 53% of questionnaire respondents was usability, e.g. difficulties in using the package (at least initially - see Appendix A.4). Respondents suggested links to more resources, such as R code examples, to demonstrate how to apply and interpret outputs of EpiEstim in multiple scenarios. Speed issues were reported by 18% of questionnaire respondents, all using the 'uncertain_si' feature, where R_t estimates are integrated over a range of serial interval distributions. Compatibility issues were reported by users wanting to use alternative programmes or programming languages, such as JAGS or Julia, or wanting to use EpiEstim in a workflow with other R packages and suggested the addition of information on how to do so (see Appendix A.4 for further details).

2.4 Discussion

To improve methods of R_t estimation, it is crucial to first gain an understanding of the current key challenges preventing, or restricting, practical applications of these methods to quantify pathogen transmissibility in real-time. To get an overview of such issues, we chose EpiEstim, one of the most popular and best-performing methods for estimating R_t in real-time, as a case study [112]. We analysed the literature citing EpiEstim as well as direct feedback from EpiEstim users collected through a questionnaire. We have also summarised the R packages identified that use alternative methodologies.

The most common challenge identified was to deal with factors influencing the quality of incidence data. Addressing frequently encountered issues, such as delays in reporting, underreporting, and administrative noise, appeared to be a priority in the field. Delays in reporting of cases or time-varying reporting rates have been a particular issue throughout the COVID-19 pandemic, linked to a high proportion of asymptomatic cases, fluctuations in testing, and health systems being under strain [178–180]. To counter some of these issues, R packages such as bayEStim and EpiNow2 explicitly account for delay distributions enabling "nowcasting", which aims to eliminate any time-lag when estimating the impact of control measures or factors such as changes in behaviour [167, 168].

When incidence by date of infection or symptom onset was unavailable or unreliable, some approaches used incidence by date of death. For instance, COVID-19 case data was initially very sporadic, whilst death data was more reliably reported. The R packages Epidemia and EpiNow2 allow the use of death data to maximise the use of available information in real-time [168, 169]. In many epidemics, including COVID-19, the reliability of case data improves over time and can provide more up-to-date insights into changing transmission dynamics. The ability to use either dataset means they can complement each other during different phases of the outbreak. However, the suitability of the type of data used will depend on the pathogen under investigation and the local context.

To our knowledge, no existing R package fully addresses the methodological challenges associated with estimating R_t using weekly-aggregated incidence data, which was highlighted in questionnaire responses and is a frequent query in correspondence with users of EpiEstim. While it is possible to

supply EpiEstim and EpiNow2 with weekly-aggregated data if the SI is provided on the same time scale, aggregating a short SI to a weekly distribution may affect the quality of the R_t estimates and is not possible when the SI is shorter than a week [145]. This is a key ongoing issue, particularly for diseases such as influenza and Zika, where incidence is typically reported on a weekly basis, [145, 181] but also for COVID-19, with several US states having recently moved from reporting daily to weekly cases only [182].

The scoping review identified additional methodological or data-related issues. These led to the development of methods that are more appropriate for use during periods of low incidence, [121, 124, 143, 148, 149, 162] or offer alternatives to allow less subjective and more flexible inputs into EpiEstim, e.g. the prior for R_t, the SI distribution, or an alternative way of temporally smoothing R_t estimates [118, 119, 121, 124–128, 130, 131, 133, 137, 143, 145–151, 153, 159, 160, 162, 163, 166].

Modifications concerning geographical factors were also common in the literature, although only mentioned by one questionnaire respondent. Beyond accounting for imported cases [114, 162] and estimating R_t across different regions at the same time, [118, 125, 127, 131, 137, 146, 147, 151, 153] we found no R packages that explicitly addressed spatial interactions. This means that, aside from bespoke studies, [136, 140, 152, 155, 157, 158, 163, 165] important aspects of the spatial structure in transmission may have been overlooked. For instance, to our knowledge, there is no ready-to-use tool to analyse how easing restrictions in an area may influence transmission in neighbouring locations.

Another theme highlighted in both the scoping review and questionnaire responses was practical or logistical issues, such as evaluating or accounting for the impact of interventions, assessing when elimination has been reached, or extending the framework for logistical planning in hospitals. Transmissibility estimates ultimately inform intervention design, e.g. to select interventions that appear most effective and identify locations which should be targeted for implementation. The ability to estimate the impact of interventions on R_t across different countries is a useful feature of Epidemia, which uses a logistic regression framework to infer the impact of non-pharmaceutical interventions (e.g. social distancing) and reductions in mobility on transmission [169]. Additionally, an important practical consideration is when one can confidently estimate that elimination has been reached, which has been addressed by EpiFilter [124, 162, 171]. More R packages that directly address a variety of recurring logistical issues is a promising avenue for future research in the field.

Despite the small sample size (n = 17) and the imperfect nature of the questionnaire (see Appendix A.4), it reinforced many of the findings from the scoping review and revealed additional challenges with computational speed, usability, and compatibility. We evaluated each of the identified R packages and tool based on these criteria and found that, of the packages that estimate the reproduction number, only earlyR scored full marks in the usability category (Table 2.2 and Table A.2). These themes emphasise that package developers (of both new and existing methods) should ensure that increased functionality does not come at the expense of speed, and that additional resources are

available to aid understanding of the package and how it can be used in conjunction with others in an outbreak analysis workflow. In response, we created a new EpiEstim vignette, which intends to provide a greater range of examples of using EpiEstim in practice [183]. A suggestion to improve usability by providing a library of pre-built configurations for different pathogens has already begun to be addressed to some extent by the R package EpiNow2 [168].

This is a fast-moving field and novel extensions or applications of EpiEstim-like methods are continuously emerging. It is therefore likely that we have missed valuable modifications not covered by articles available at the time of our literature search, such as an extension to the method by Johnson et al. to account for superspreading [184] and by Brizzi et al. to account for unobserved generations of infections at the start of an outbreak [185]. It is also possible that through using EpiEstim as a starting point for our literature search, we may have missed relevant papers that did not cite the method. However, given that EpiEstim has been recognised as the most reliable currently existing approach for real-time R_t estimation, [112] we believe it is unlikely that we would have missed important contributions that would alter our conclusions. To our knowledge, there are still numerous issues regarding real-time estimation of the reproduction number that remain unaddressed by R packages or opensource software. For instance, we found no readily available tools that consider non-spatial population heterogeneities or the ability to include time-varying generation times [186, 187].

The intention of this review was to provide a broad overview of the current landscape of renewal-equation based R_t estimation methods and tools. Therefore, beyond providing an appraisal of the usability of each identified R package/tool in terms of ease of installation, available documentation, and speed, we have not tested or compared their performance. Future work should focus on a more systematic and critical evaluation of the estimation accuracy of R packages intending to tackle the same issues. Moreover, it would be interesting to characterise the added value of additional features in relation to their cost, such as longer computational time.

2.5 Conclusion

The quality of incidence data continues to pose significant challenges to real-time estimation of R_t. Numerous methods and R packages have been developed to address issues such as reporting delays [118, 125–127, 131, 147, 151, 153, 154, 167, 168], underreporting [120, 123, 130, 137, 141, 142, 146, 169] and administrative noise [118, 125–127, 131, 134, 135, 147, 151, 153, 155], but significant gaps remain, such as the inability to directly use temporally aggregated incidence data. Despite the importance of spatial factors in transmission, no R package has been identified to account for movement or interactions between locations when estimating R_t. It is also clear that extensions to these methods could allow for rapid translation of R_t estimates into logistically relevant outputs, such as predicting how the demand for hospital resources may change in real-time. Addressing these recurring issues and extending

the methodology to directly answer important practical and logistical questions are key priorities for widening the applicability of R_t estimation methods during epidemics. However, package developers should keep in mind that speed, ease of use, and access to sufficient resources, will be key to the uptake of these new or improved tools.

The remaining chapters of this thesis address issues related to three of the broad themes identified in this study: "incidence data", "other input data" and "practical or logistical issues". First, Chapter 3 focuses on the need for a robust resource to accurately parameterise model inputs, such as the serial interval, using historical data. Next, Chapter 4 addresses a key policy-relevant issue by applying the renewal equation to determine, in real-time, whether newly emerging pathogen variants have a transmission advantage over currently circulating variants. Finally, Chapter 5 and Chapter 6 address the inability to use temporally aggregated incidence data in branching process models.

Chapter 3

Ebola virus disease mathematical models and epidemiological parameters: a systematic review

The contents of this chapter have been adapted from a manuscript previously published in The Lancet Infectious Diseases [188]. Dr Juliette Unwin performed the literature search as described in Section 3.3.1. The screening and data extraction process, as well as the assessment of bias and quality of the included papers, was a collaborative effort involving contributions from 18 other co-authors (see Section 3.3.1). My involvement in the initial title and abstract screening was minimal, with 10 papers reviewed. I reviewed the full text of 304 eligible papers, representing the fourth-largest contribution to full-text screening, and extracted data from 59 studies – the third-largest contribution to data extraction. Alongside myself, Dr Sangeeta Bhatia, Dr Anne Cori and Dr Juliette Unwin contributed to additional data extraction and analysis of modelling results, and further analysis of severity estimates in response to peer reviewer comments. Apart from these analyses, I was responsible for all data cleaning, code and all other data analyses presented in this chapter. I also initiated the core functionalities of the epireview R package. The manuscript was written with input from all co-authors, with primary contributions from myself, Dr Sangeeta Bhatia, Dr Anne Cori and Dr Juliette Unwin.

As discussed in Chapter 1 and highlighted in the findings of Chapter 2, reliable parameterisation of models is critical for robust estimation of R_t and infectious disease modeling. In order to determine whether an outbreak is worse than usual (Figure 1.1) and to enhance preparedness and rapid response efforts, access to comprehensive databases of parameter estimates from historical outbreaks serves as an invaluable resource.

The WHO have identified "priority diseases" deemed to have the most epidemic potential [103]. These diseases include COVID-19, Ebola Virus Disease (EVD), Marburg virus disease, Crimean-Congo haemorrhagic fever, Lassa fever, Middle East Respiratory Syndrome (MERS), Severe Acute Respiratory

Syndrome (SARS), Nipah and henipaviral diseases, Rift Valley fever, Zika and 'Disease X', which represents a disease caused by a currently unknown pathogen capable of causing a serious epidemic or pandemic [103]. This work focuses on one of the priority diseases, EVD, to systematically review all available literature on key epidemiological and modelling parameters. This includes estimates of R_t, the Serial Interval (SI), generation time, other epidemiological delays, seroprevalence, attack rates, severity estimates (e.g. Case Fatality Ratio (CFR)), mutation rates and risk factors for infection or death. A database of key EVD parameters is established and stored in an open-source R package, epireview, to facilitate their use in R_t estimation and outbreak modelling [189]. Additionally, a meta-analysis of suitable parameters is conducted to obtain the optimal estimates for use in outbreak models.

This work contributes to a wider initiative by the Pathogen Epidemiology Review Group (PERG) at Imperial College London to systematically review all WHO priority pathogens and was critical in initiating the functionalities of the epireview R package where the data is stored [190].

3.1 Summary

EVD poses a recurring risk to human health. We conducted a systematic review (PROSPERO CRD42023393345) of EVD transmission models and parameters published from database inception to July 7, 2023, from PubMed and Web of Science. Two people screened each abstract and full text. Papers were extracted with a bespoke Access database, 10% were double extracted. We extracted 1280 parameters and 295 models from 522 papers. Basic reproduction number estimates were highly variable, as were effective reproduction numbers, likely reflecting spatiotemporal variability in interventions. Estimates obtained from a random-effects model were 15.4 days (95% Confidence Interval (CI) 13.2–17.5) for the SI, 8.5 days (7.7–9.2) for the incubation period, 9.3 days (8.5–10.1) for the symptom-onset-to-death delay, and 13.0 days (10.4–15.7) for symptom-onset-to-recovery. Estimates obtained from a common-effect model were similar, albeit with narrower CIs. Case Fatality Ratio (CFR) estimates were generally high but highly variable, which could reflect heterogeneity in underlying risk factors. Although a substantial body of literature exists on EVD models and epidemiological parameter estimates, many of these studies focus on the west African Ebola epidemic and are primarily associated with Zaire Ebola virus, which leaves a key gap in our knowledge regarding other Ebola virus species and outbreak contexts.

3.2 Introduction

The COVID-19 pandemic, and multiple outbreaks of re-emerging pathogens over the past 10 years, have highlighted the tremendous threat of infectious pathogens to the human population. The Democratic Republic of the Congo (DRC) experienced an outbreak of EVD from 2018 to 2020. Seven additional EVD outbreaks in the DRC, Guinea, and Uganda have been declared since. In May, 2022, an outbreak

of mpox affected several countries beyond the known endemic region [191]. In February, 2023, to March, 2023, Equatorial Guinea and Tanzania faced a Marburg virus disease outbreak, the first since 2014 [192]. These recurring events reinforce the need for pandemic preparedness. WHO has listed Ebola virus as one of nine pathogens posing the greatest threat to public health due to its high epidemic potential and insufficient countermeasures [193].

Ebola virus is a deadly filovirus, [194] transmitted through close contact and bodily fluids, especially during traditional burials and caregiving, which has caused 38 known outbreaks since its discovery in 1976 (Appendix B.2) [195]. Most of these outbreaks have occurred in central and western Africa, and the largest epidemic, the west African EVD epidemic, caused more than 11,000 reported deaths between 2013 and 2016, mainly across Guinea, Liberia, and Sierra Leone [196, 197]. Four species of Ebola virus are known to affect humans: Zaire, Bundibugyo, Sudan, and Taï Forest. One species, Reston, is only known to cause disease in non-human primates [194] and a sixth species, Bombali, was identified in samples taken from bats in Sierra Leone [198]. The symptoms of Ebola virus infection can be sudden and include fever, fatigue, muscle pain, headache, and sore throat, followed by vomiting, diarrhoea, rash, and internal and external bleeding.

Treatment of EVD involves supportive care, such as rehydration (e.g. intravenous fluids or oral rehydration) and the stabilisation of oxygen saturation and blood pressure. There are two monoclonal antibody treatments recommended for confirmed cases of infection with the Zaire species, REGN-EB3 and mAb114, but access is limited by uncertainties surrounding pricing and future supply [199]. There are no licensed treatments for other EVD species. A vaccine (rVSV-ZEBOV [Ervebo; Merck, Rahway, NJ, USA]) trialled against the Zaire species during the west African epidemic [200] is now used as part of outbreak response activities to prevent disease with a ring vaccination strategy. An alternative vaccine combination consists of two heterologous doses (Ad26.ZEBOV [Zabdeno] and MVA-BN-Filo [Mvabea; Janssen-Cilag International, Beerse, Belgium]) given 8 weeks apart and is therefore not suitable for use in an emergency context, in which immediate protection is necessary [201]. Three vaccine candidates for the Sudan species are in various trial phases [202]. In the absence of widely available vaccines and therapeutics, mitigation of EVD outbreaks relies on a suite of public health and social measures, such as case identification and isolation, contact tracing and quarantine, personal protective equipment for health-care workers, safe and dignified burials, and community engagement.

Mathematical modelling and outbreak analytics are a component of monitoring and responding to epidemics that have been used effectively during previous outbreaks. Examples include estimating the severity, epidemic delays, and projections during the west African epidemic; [65, 66] estimating bed capacity in Sierra Leone during the west African epidemic; [203] producing estimates of severity, epidemic delays, and reproduction number in the DRC in 2018; [204] and making disease forecasts during the 2018–19 DRC outbreak [205]. Mathematical epidemic models typically use various parameters that characterise the pathogen as inputs (e.g. the incubation and infectious periods), with the robustness

of modelling outputs depending on the reliability of parameter values. Furthermore, these parameters have direct implications for outbreak control; for example, the upper bound of the incubation period determines the necessary duration of follow-up for contacts of cases [65]. Therefore, centralising evidence around these parameters to inform the rapid design of mathematical models that could support the response to future EVD outbreaks is important. We aim to carry out a systematic review with a meta-analytic component to build a database of EVD models and related parameters.

3.3 Methods

We used Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines for this systematic review. Checklists are included in the appendix: Table B.5 and Table B.6.

3.3.1 Search strategy and selection criteria

This systematic review was prospectively registered with the international prospective register for systematic reviews, known as PROSPERO (CRD42023393345). We searched PubMed and Web of Science databases for English language peer-reviewed articles including EVD transmission models, parameters characterising EVD transmission, evolution, natural history, severity, seroprevalence, and risk factors published from database inception to July 7, 2023 (search terms used provided in Appendix B.3.1). Each title and abstract and then full text were screened by two reviewers from a group of 15 (RKN, SB, CM, PD, DJ, KM, RM, AF, GC-D, JH, TN, IR, SvE, AC, and HJTU) using Covidence and inclusion and exclusion criteria (Appendix B.3.2). Disagreements were resolved by consensus between reviewers. We used backward citation chaining from 12 of the 179 review papers [206–217] identified during screening to add in missing papers (Appendix B.3.3).

19 authors (RKN, SB, CM, PD, KM, RM, DN, GC-D, JH, RS, TN, SvE, CG, TR, SIL, JW, KF, AC, and HJTU) extracted data about the articles, models, and parameters from our included papers using Microsoft Access (Appendix B.3.4). Four authors (RKN, SB, AC, and HJTU) extracted further data about the models from the papers that mentioned publicly available code to identify models that could in theory be re-used in future outbreaks (Appendix B.3.4).

To ensure consistency of the data extraction process, data were double extracted from a randomly selected sample of 10% (n=55) of papers, with disagreements resolved by consensus. We used a customised questionnaire to assess the quality and risk of bias of each paper (Appendix B.3.5). Data extractors responded to each question with yes, no, or not applicable if the question was not applicable to the article. For each article, the quality score was determined as the ratio of yes answers to the total number of applicable questions answered for that article.

3.3.2 Data analysis

All analysis was done in R with the orderly2 R package [218] for workflow management (Appendix B.3.6). Curated and annotated data are made available in our R package epireview [189]. We conducted meta-analyses for incubation periods, SIs, and time from symptom onset to death or recovery, in which the number of estimates exceeded at least two, with the metamean function from the meta R package [219]. This package returns two estimates, one from a common-effect model and one from a random-effects model.

The common-effect model assumes that all effect sizes come from a single, homogeneous population, meaning that any differences between studies are solely due to sampling error. More formally, each study's estimate θ_i is assumed to reflect a single true underlying effect θ, with normally distributed error scaled by its standard error s_i (typically calculated as the standard deviation of the outcome measure divided by the square root of the study sample size):

$$\theta_i = \theta + s_i\epsilon_i, \quad \epsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1). \tag{3.1}$$

The error terms ϵ_i are independent and identically distributed (i.i.d) meaning that they are drawn independently from the same standard normal distribution. Under this assumption, differences in the effect size estimates across studies are attributed solely to random sampling error, not from underlying heterogeneity, such as differences in the study population or methods [220].

The pooled mean estimate $\hat{\theta}_{CE}$ is calculated as:

$$\hat{\theta}_{CE} = \frac{\sum_{i=1}^{I} \theta_i w_i}{\sum_{i=1}^{I} w_i} \tag{3.2}$$

where $w_i = 1/s_i^2$ is the inverse-variance weight, with s_i^2 being the variance of θ_i [221]. Larger studies, which tend to have smaller standard errors due to their increased sample size, have more weight on the pooled estimate [220].

The variance of the pooled estimate $\hat{\theta}_{CE}$ is given by:

$$Var(\hat{\theta}_{CE}) = \frac{1}{\sum_{i=1}^{I} w_i}. \tag{3.3}$$

The random-effects model assumes that the estimates from different studies θ_i arise from a distribution centred around a true mean θ, incorporating both within-study (s_i^2) and between-study variance (τ^2) [221]. The model can be expressed as:

$$\theta_i = \theta + u_i + s_i\epsilon_i, \quad \epsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1), \quad u_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\tau^2), \tag{3.4}$$

where u_i represents the random effect for study i, and is assumed to be independent of the error term

ϵ_i. Therefore, the model assumes that the observed differences between studies arise from sampling variability and true heterogeneity across studies, such as differences in the study design and population. The weight of each study is adjusted using the formula:

$$w_i^* = \frac{1}{s_i^2 + \tau^2} \tag{3.5}$$

and using the adjusted weights, the pooled mean effect size is calculated as:

$$\hat{\theta}_{RE} = \frac{\sum_{i=1}^{I} \theta_i w_i^*}{\sum_{i=1}^{I} w_i^*}. \tag{3.6}$$

The variance of $\hat{\theta}_{RE}$ is given by:

$$Var(\hat{\theta}_{RE}) = \frac{1}{\sum_{i=1}^{I} w_i^*} \tag{3.7}$$

We use the I^2 statistic to estimate the percentage of total variability in effect sizes that is due to between-study heterogeneity as opposed to within-study sampling error [222]. It is calculated as:

$$I^2 = \frac{Q - (I-1)}{Q} \times 100\% \tag{3.8}$$

where I is the total number of studies and Q is Cochran's Q statistic, which estimates the total observed variability around the pooled estimate from the common-effect model:

$$Q = \sum_{i=1}^{I} w_i(\theta_i - \hat{\theta}_{CE})^2. \tag{3.9}$$

A higher value of Q indicates greater heterogeneity than would be expected from sampling error alone, and I^2 expresses this excess as a percentage of the total variability.

The metamean function was also used to conduct sub-group meta-analyses to explore whether these delays varied by Ebola virus species. We did not do meta-analyses for other parameters due to too few data in the required format (Appendix B.3.6) or high variability between the study settings. We did perform meta analyses in which we did not filter on quality assessment.

We only include parameters from articles with a high quality assessment score of at least 50% (Appendix B.3.5; all parameters are included in Appendix B.4). We did sensitivity analyses in which we did not filter for high quality papers.

3.3.3 Data sharing

Data are available from `https://github.com/mrc-ide/epireview/tree/main/inst/extdata`. Code is available from `https://github.com/mrc-ide/epireview` and our full analysis can be reproduced using `https://github.com/mrc-ide/priority-pathogens`.

3.4 Results

Our search returned 24,338 articles, which was reduced to 14,690 after deduplication and the addition of two papers identified through other systematic reviews (Figure 3.1). Following title and abstract screening, 1,674 articles were retained for full text screening, with 522 meeting our inclusion criteria.

We extracted 1,280 parameters from 354 articles (Appendix B.4.1) and 295 models from 280 articles. We could link 1,213 of the parameters to a specific EVD outbreak; of these, 858 (70.7%) reported on the west African Ebola epidemic. 1,229 parameters could be linked to a specific Ebola virus species, of which 1,136 (92.4%) were associated with the Zaire species.

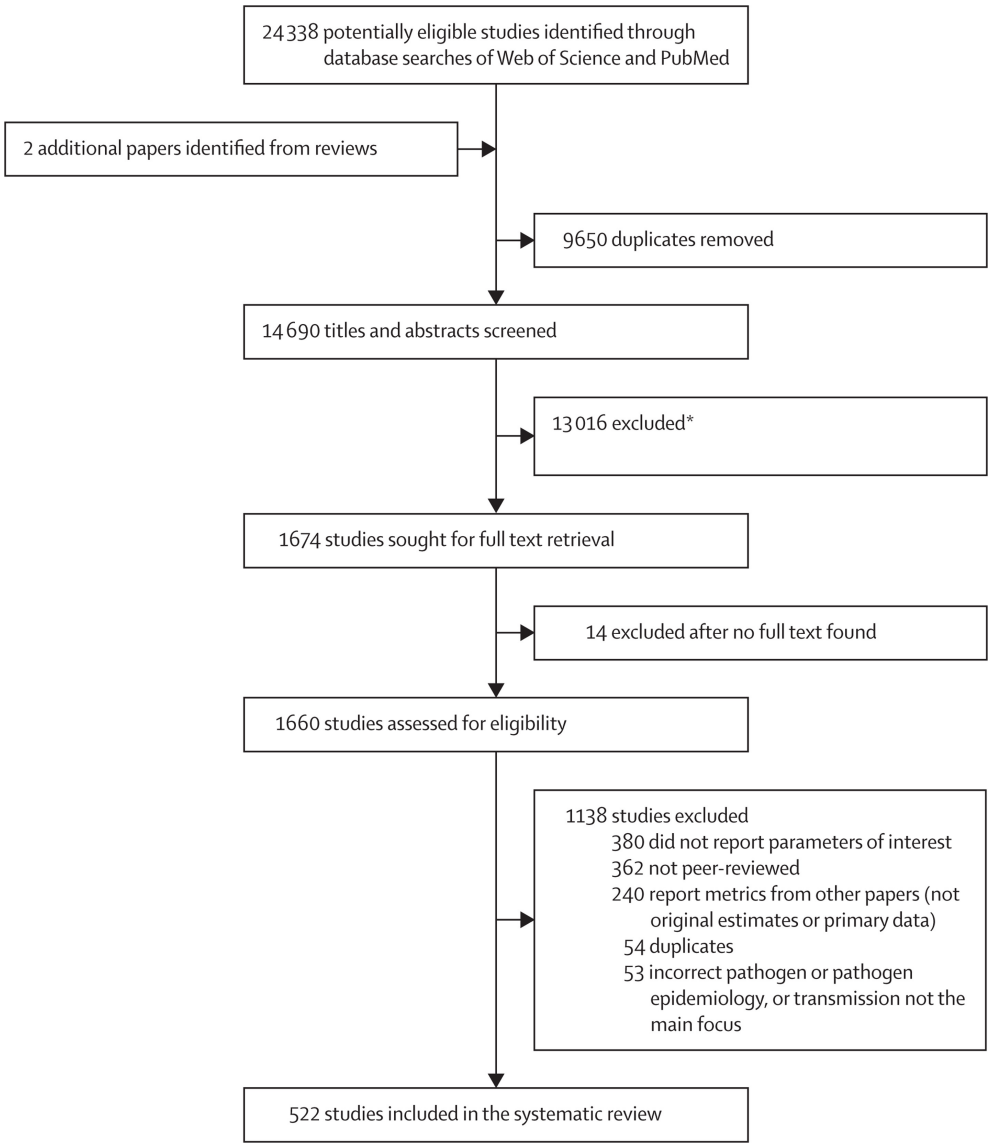


Figure 3.1: **Study selection.** Reasons for exclusion given in Appendix B.3.2. Figure shown as presented in [188].

Published EVD seroprevalence estimates were highest in countries with reported outbreaks (Appendix B.2 and Appendix B.4.2), but varied depending on the population sampled. In the Democratic Republic of the Congo (DRC), community-based seropositivity was between 0% and 18.7% (five studies,

five estimates) across the different assays, whereas in hospitals it was between 0% and 37.0% (five studies, eight estimates). Similarly, in Guinea, a single population-level estimate (one study, one estimate) was 0.07% but ranged from 59.4% to 99.8% in hospitals. Despite no official outbreaks, seropositivity has been reported among population groups in Cameroon, Central African Republic, Kenya, Madagascar, Mali, Tanzania, and two multi-country studies (ten studies, 12 estimates; appendix Table B.10).

Across eight articles reporting attack rates (Appendix B.4.3), nine central estimates were all less than 10%, except for one paper [223] that suggested that up to 31% of physicians had been infected in the 1995 outbreak in Kikwit, DRC. Seven (70%) of ten estimates focused on the general population with generally low uncertainty.

Published estimates of the basic reproduction number (R_0) were highly variable (appendix Table B.13). After removing papers with low quality assessment scores (<50%), 71 R_0 central estimates across 52 studies ranged from 0.05 to 12 (Figure 3.2, Table 3.1). 58 (82%) of 71 R_0 estimates, including the most extreme central estimates, were for the 2013–16 west African epidemic. Across all other outbreaks, R_0 central estimates ranged from 1.1 to 7.7 (both estimates were for the 2018–20 DRC outbreak). Uncertainty around west African central estimates of R_0 was also highly variable, with lower bounds of 0 and upper bounds of 18.5. 69 (97%) of 71 R_0 estimates were for the Zaire species and two were for the Sudan species (from outbreaks in Uganda in 2000–01 and 2022–23). Sudan R_0 central estimates were less variable than Zaire estimates (range 2.0 to 2.7).

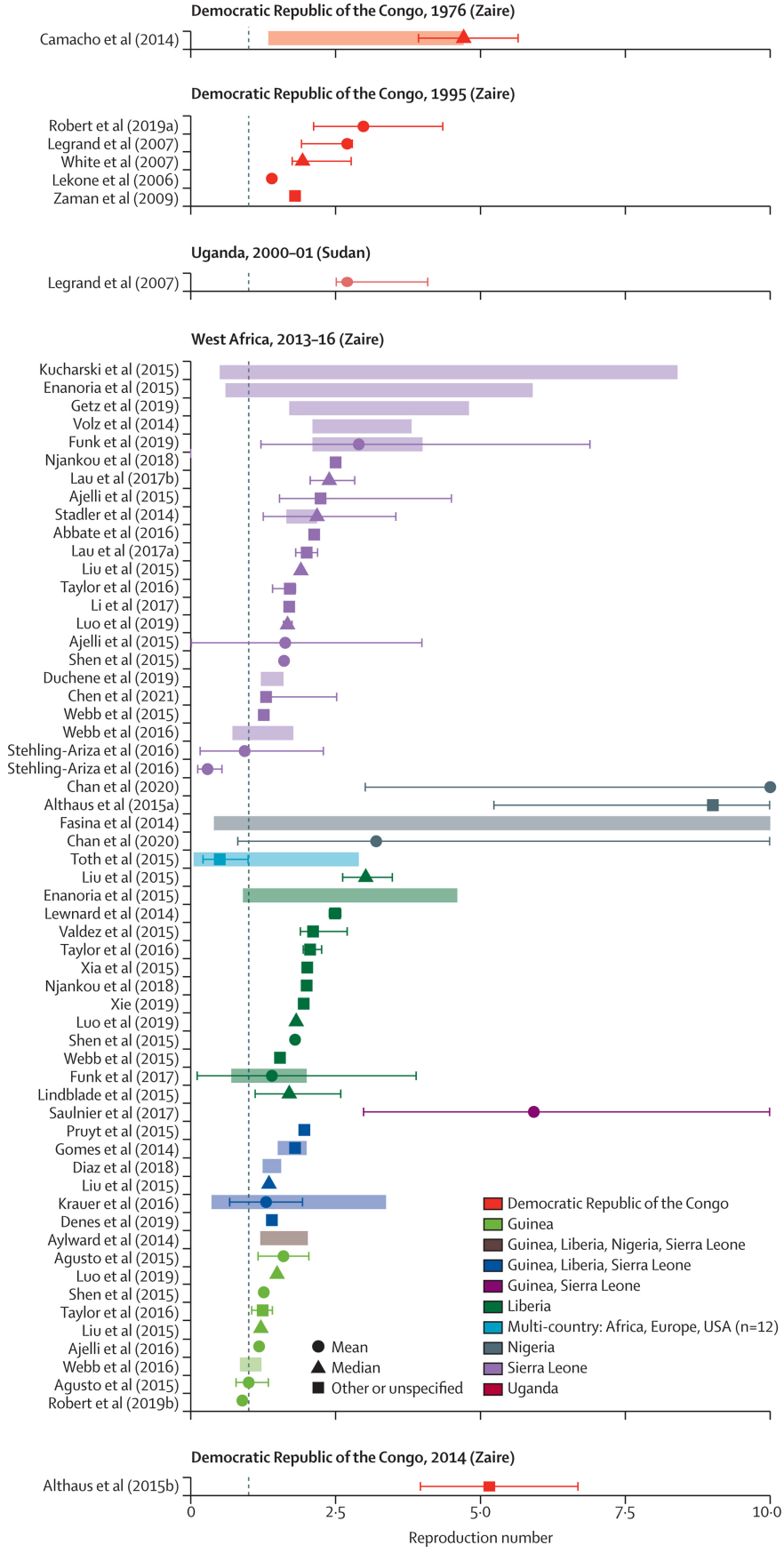


Figure 3.2: **Basic reproduction numbers by outbreak (1/2).** Figure 3.2 continues on the next page.

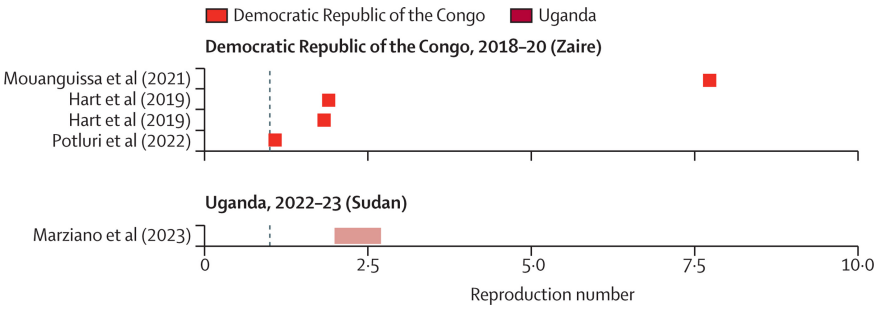


Figure 3.2: **Basic reproduction numbers by outbreak (2/2).** Reasons for exclusion given in Appendix B.3.2. Each section corresponds to a different outbreak of EVD with the associated Ebola virus species in brackets. Points represent central estimates, with symbol shapes corresponding to value type. Thick coloured shaded lines represent the range of central estimates in which the reproduction number was estimated (e.g. across regions or over time). Solid coloured bars represent the uncertainty around the central estimate, which was reported in different formats including Standard Deviation (SD) (in which case the bar represents one SD either side), 95% highest posterior density interval, range, Interquartile Range (IQR), 95% CrI, and 95% CI. The x-axis has been restricted to a maximum of 10 for clarity. All parameters are from articles with a quality assessment score of at least 50% (appendix Table B.13). Some studies reported multiple estimates for different countries or situations; different publications with the same lead author and year of publication are indicated with lowercase letters (e.g. 2019a and 2019b). Reference list for cited publications is available in the supplementary appendix in [188]. Figure shown as presented in [188].

Table 3.1: **Ranges of estimates for basic reproduction numbers by outbreak and estimate type.** Total number of parameters specifies the number of parameters (quality assessment filtered ≥50%) included in the summary range. Some parameter entries provide aggregated ranges of central estimates (e.g. across time and countries; Appendix B.3.4) if more than three values were provided. Additionally, not all central estimates were reported with an associated uncertainty interval.

	Central estimate range	Uncertainty range	Total number of parameters
Democratic Republic of the Congo, 1976			
Median	1.34-4.71	3.92-5.66	1
Democratic Republic of the Congo, 1995			
Mean	1.4-2.98	1.9-4.36	3
Median	1.93	1.74-2.78	1
Other or unspecified	1.8	..	1
Uganda, 2000-01			
Mean	2.7	2.5-4.1	1
West Africa 2013-16			
Mean	0.29-10	0-18.5	17
Median	0.6-5.9	1.1-3.55	14
Other or unspecified	0.05-12	0.2-15.55	27
Democratic Republic of the Congo, 2014			
Other or unspecified	5.15	3.95-6.69	1
Democratic Republic of the Congo, 2018-20			
Other or unspecified	1.08-7.73	..	4
Uganda, 2022-23			
Other or unspecified	1.99-2.7	..	1

Estimates of the effective reproduction number, which measures transmissibility in the presence of potential interventions and population immunity, were also highly variable (appendix Table B.14). In 23 high quality assessment studies (≥50% quality assessment scores), central estimates ranged from 0 to 9 across 32 parameter estimates, of which all were for Ebola virus Zaire. 23 (72%) of 32 parameters were for the west African epidemic, and the rest were all from the DRC, with only one estimate before the west African epidemic (0.73 in the 1995 DRC outbreak).

Estimates of secondary attack rates (12 studies, 15 estimates; appendix Table B.15), growth rates (six studies, 13 estimates; appendix Table B.16), and doubling times (four studies, six estimates; appendix Table B.17) were similarly heterogeneous across studies. Central estimates of secondary attack rates were in the range of 0.1% to 89% (eight estimates for household-based studies, three estimates for community-based studies, and three estimates for contact-based studies), daily growth rates were between 0.0 and 1.4, and doubling times were up to decades.

We extracted 14 estimates of overdispersion from 11 studies, which all used a negative binomial distribution to characterise the offspring distribution. Central overdispersion estimates ranged from 0.02 to 2.20, with lower values indicating more overdispersion (appendix Table B.18). Other than one entry with an unspecified Ebola virus species, all were for Zaire (13 estimates), and 12 (86%) were for the west African Ebola epidemic.

We identified 31 studies that examined risk factors for being infected with Ebola virus (Appendix B.4.3). Conflicting findings were found across studies, with most risk factors found to be both non-significant and significant. Risk factors such as contact with an infected individual (close contact with the individual or their bodily fluids, household contact, and non-household contact), participation in funerals (either attendance or involvement in unsafe burial practices), age, occupation, and admission to hospital for an unrelated risk factor for EVD were most frequently significant. However, sex was more commonly non-significant (12 analyses) than significant (three analyses).

Risk factors for seropositivity were similar to those investigated for infection and showed similar conflicting findings across studies (appendix Table B.11 and Table B.19). Contact with an animal was most commonly significant. Age, sex, occupation, close contact, and household and non-household contact were most frequently non-significant. Studies were evenly divided on the significance of funeral participation, with some finding it to be a significant risk factor and others finding no significant association. Two studies analysed risk factors for onwards transmission; significant risk factors included age, sex, socioeconomic status, survival, funeral, and being part of the first generation of a transmission chain (appendix Table B.20).

We extracted 21 estimates of the SI that were reported from 17 studies (19 estimates for Zaire species, two estimates for Sudan), and a single generation time estimate was reported for Zaire (Table 3.2 and appendix Figure B.2). The mean SI estimate for high quality assessment studies (six studies and six estimates, all Zaire species) was 16.5 days (95% CI 16.1–16.9) for common effect and

15.4 days (13.2–17.5) for random effect (Figure 3.3). The high I^2 value (94%) is indicative of high amounts of heterogeneity in the SI, suggesting that the estimates from the common effect might be overprecise and the random effect estimates more reliable. Estimates including three additional low quality assessment studies were similar (appendix Figure B.6).

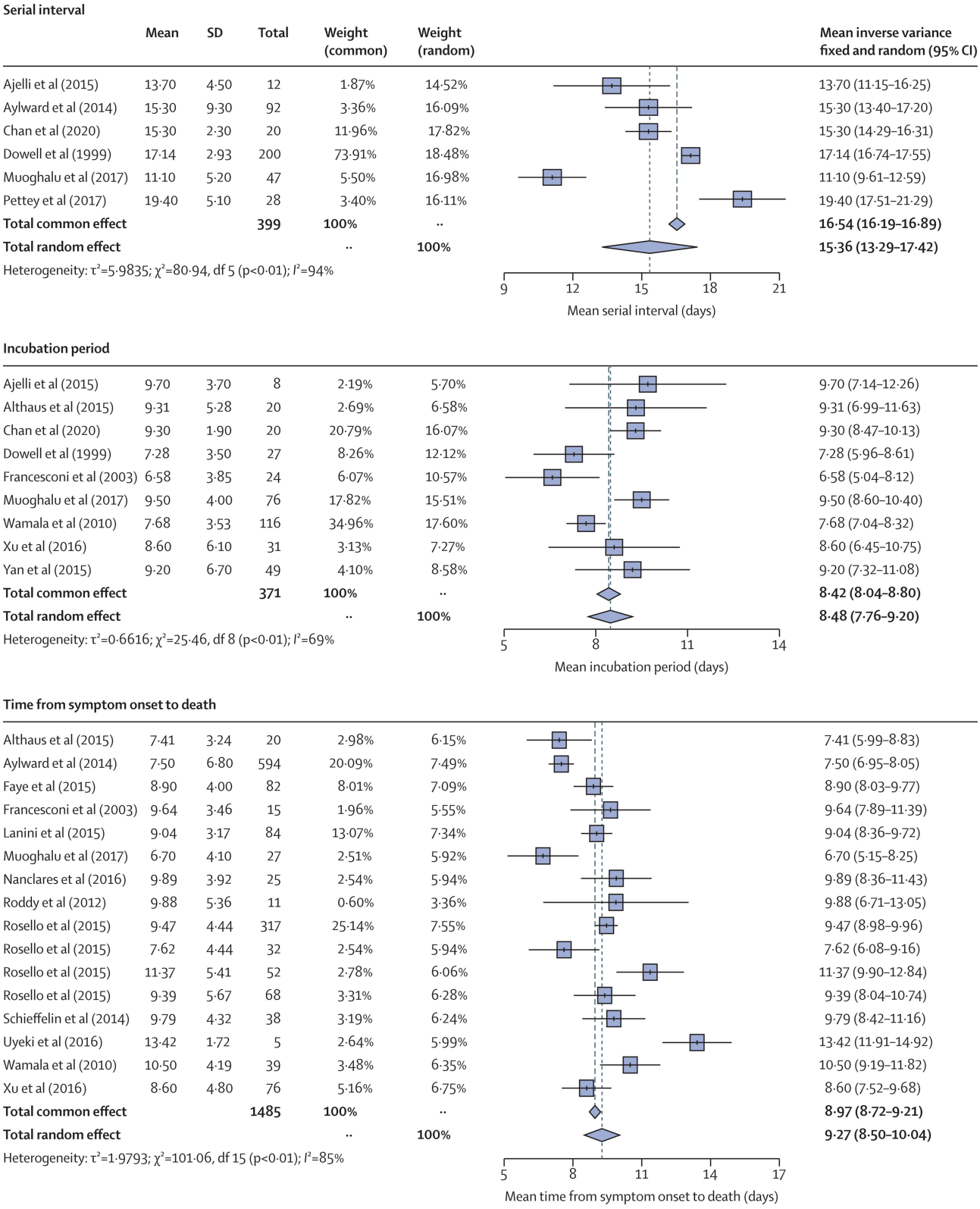


Figure 3.3: **Meta-analysis for the mean serial interval, incubation period, time from symptom onset to death, and time from symptom onset to recovery (1/2).** Figure 3.3 continues on the next page.

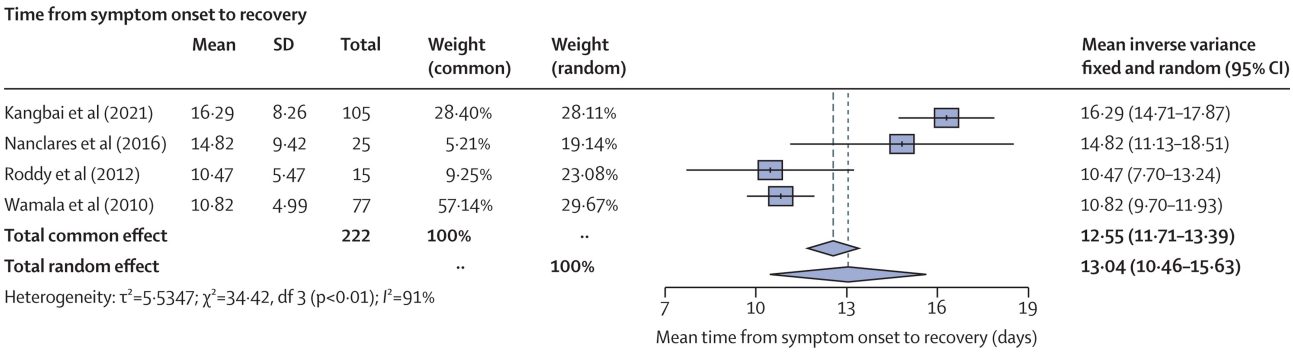


Figure 3.3: **Meta-analysis for the mean serial interval, incubation period, time from symptom onset to death, and time from symptom onset to recovery (2/2).** Some studies reported multiple estimates for different countries or situations. All included studies have a quality assessment score of 50% or more. Parameters used in the meta-analyses are paired mean and SD of the sample or were converted into mean and SD of the sample from the following combinations: mean and Standard Error (SE); median and IQR; or median and range (Appendix B.3.6). Blue squares are mean estimates from each study with 95% CIs. Diamonds represent the overall mean across studies for the common and random-effect models. The random-effect model accounts for within-study and between-study variance. Reference list for cited publications is available in the supplementary appendix in [188]. Figure shown as presented in [188].

Table 3.2: **Ranges of estimates for epidemiological delays by Ebola virus species.** Total number of parameters specifies the number of parameters (quality assessment filtered ≥50%) included in the summary range. Some parameter entries provide aggregated ranges of central estimates (e.g. across time and countries; Appendix B.3.4) if more than three values were provided.

	Central estimate range	Uncertainty range	Total number of parameters
Generation time (days)			
Zaire	10.9	9.2-13	1
Incubation period (days)			
Bundibugyo	5.7-11.3	2-27	3
Sudan	6-14	1-16	3
Zaire	0.1-18	0-33	30
Infectious period (days)			
Zaire	1.7-29.6	1.2-10.8	12
Latent period (days)			
Zaire	0.1-31.2	0.1-71	7
Serial interval (days)			
Sudan	11.7-12	8.2-15.8	2
Zaire	7-19.4	2-25	15
Symptom onset to death (days)			
Bundibugyo	9-11.4	3-21	3
Sudan	3-11	3-15	2
Zaire	1-14	0-27	28
Symptom onset to recovery or non-infectiousness (days)			
Bundibugyo	9.9-10	2-26	2
Zaire	8.4-14	7.5-20	5

We identified 52 estimates for the incubation period from 43 studies, 11 estimates of the latent period from ten studies, and 27 estimates of the infectious period from 23 studies (appendix Figure B.2).

Excluding papers with low quality assessment scores (<50%), all estimates of the incubation period were for Ebola Zaire species (30 estimates), except for six estimates (three Bundibugyo and three Sudan), with central estimates for all species ranging between 0.1 and 18. Central estimates ranged from 0.1 to 31.2 days for the latent period and from 1.7 to 29.6 days for the infectious period (Table 3.2). There were too few estimates with high quality assessment scores for the latent and infectious periods (one estimate for each) in the required format to conduct meta-analyses (Appendix B.3.6). The pooled mean incubation period estimate based on nine studies was 8.4 days (95% CI 8.0–8.8) for total common effect and 8.5 days (7.7–9.2) for total random effect (Figure 3.3). Despite sparse estimates for species other than Zaire in the analysis (three estimates), there are statistically significant differences (p=0.02) in mean incubation periods between Ebola virus species (with mean incubation period ranging between 4.77 and 8.36 days for the random-effect model; appendix Figure B.5). After including all infectious period estimates (regardless of quality assessment score; three estimates), the pooled mean was 5.4 days (95% CI 5.3–5.5) for common effect and 5.0 days (3.7–6.3) for random effect (appendix Figure B.6). Again, the high I^2 value (100%) suggests the random-effects model is more reliable.

Clinical progression (e.g. symptom onset to death or recovery) and case management (e.g. admission to hospital) are also key modelling inputs. We extracted estimates for delays from symptom onset to test, test result, diagnosis, reporting (or WHO notification), seeking care, admission to care, quarantine, recovery, negative test or undetectable viral load, discharge from care, and death (appendix Figure B.3). We also extracted estimates for delays from admission to care and from admission to discharge, recovery, or death (appendix Figure B.4). We found delays in the clinical timeline to be highly variable across contexts. Across all studies (including those with low quality assessment, seven studies), the central estimate of symptom onset to reporting varied between 0 and 25.7 days (15 estimates), although most (13 estimates) were around half a week to 2 weeks (appendix Figure B.3). Similarly, the central symptom onset to discharge from care delay varied between 6.3 and 28 days (11 studies, 14 estimates; appendix Figure B.3). Time in care was also highly variable irrespective of the outcome but tended to be shorter for those dying: the central delay from admission to care to death was 0–11 days (26 studies, 33 estimates) versus 2.6–17 days for admission to recovery (ten studies, 12 estimates; appendix Figure B.4).

By contrast, reported delays from symptom onset to admission to care were remarkably consistent. Central estimates varied between 0 and 23 days for high quality assessment studies (39 studies, 47 estimates; appendix Figure B.3), but 44 (94%) of 47 central estimates fell between 3 and 6 days. We extracted 14 estimates from 12 studies for the delay from symptom onset to recovery, with central estimates of those with a high quality assessment (seven estimates) ranging from 8.4 to 14.0 days (Table 3.2 and appendix Figure B.3). Based on four estimates from four high quality assessment studies, the pooled mean estimate for the time from symptom onset to recovery was 12.6 days (95% CI 11.7–13.4) for total common effect and 13.0 days (10.4–15.7) for total random effect (Figure 3.3). Again

I^2 was high (91%), suggesting the random-effects model is more reliable. We extracted 48 estimates from 39 studies for the delay from symptom onset to death, with a pooled mean estimate across 16 suitable estimates (which provided sufficient information to derive standard errors (s_i) required for the meta-analysis - see Equations (3.1) and (3.4)) from 13 high quality assessment studies of 9.0 days (8.7–9.2) for total common effect and 9.3 days (8.5–10.1) for total random effect (Figure 3.3). Species subgroup meta-analyses with random-effects models indicated that the time from symptom onset to death might be longer (p=0.01) and symptom onset to recovery might be shorter (p<0.01) for those infected with the Bundibugyo species compared with Zaire (appendix Figure B.5). Similar results were found in the meta-analyses in which we did not filter on quality assessment (appendix Figure B.5 and Figure B.6).

We extracted 166 estimates of the Case Fatality Ratio (CFR) from 130 papers. Early estimates from the DRC in 1976 and 1995 suggested a CFR greater than 69% (Table 3.3, Figure 3.4; appendix Table B.21). However, more recent estimates from the west African epidemic (2013–16) and DRC (2018–20) had lower central values ranging from 18.5% to 93.5%, but with wide uncertainties in some settings. In general, estimates of CFR are highly context-specific and factors such as study design, sampling strategy, type of surveillance, location of outbreak, and whether estimates were generated post-outbreak had a strong influence. We determined that a meta-analysis was not a feasible approach for synthesis because the settings and methods that were used to generate the extracted estimates were variable. We therefore used naive measures (crude and weighted mean) to summarise CFR estimates. The mean CFR across all estimates was 57.8%. For 112 (67%) of 166 estimates in which we also had information on the sample size, the mean CFR weighted by the study population size was 64.6%. Age, sex, and occupation were frequently investigated as potential risk factors for death in both multivariate and univariate analyses (appendix Table B.22). Age was significantly associated with death in 41 (60%) of 68 parameter entries, whereas most analyses did not find a significant association between death and sex or occupation.

We extracted 24 estimates of mutation, substitution, and evolutionary rates from 20 studies (appendix Table B.23). Central estimates ranged from 0.36×10^{-4} to 0.0036 substitutions per site per year. The upper bound was an estimate from a single outbreak in the DRC [224]. By contrast, the same study estimated a much lower substitution rate of 6.9×10^{-4} across multiple countries and outbreaks from 1976 to 2018. Apart from three estimates for all species (Bundibugyo, Sudan, Zaire, and Taï forest) [225, 226] and a single estimate for Bundibugyo alone, [227] all estimates were for Zaire (one study did not specify the species).

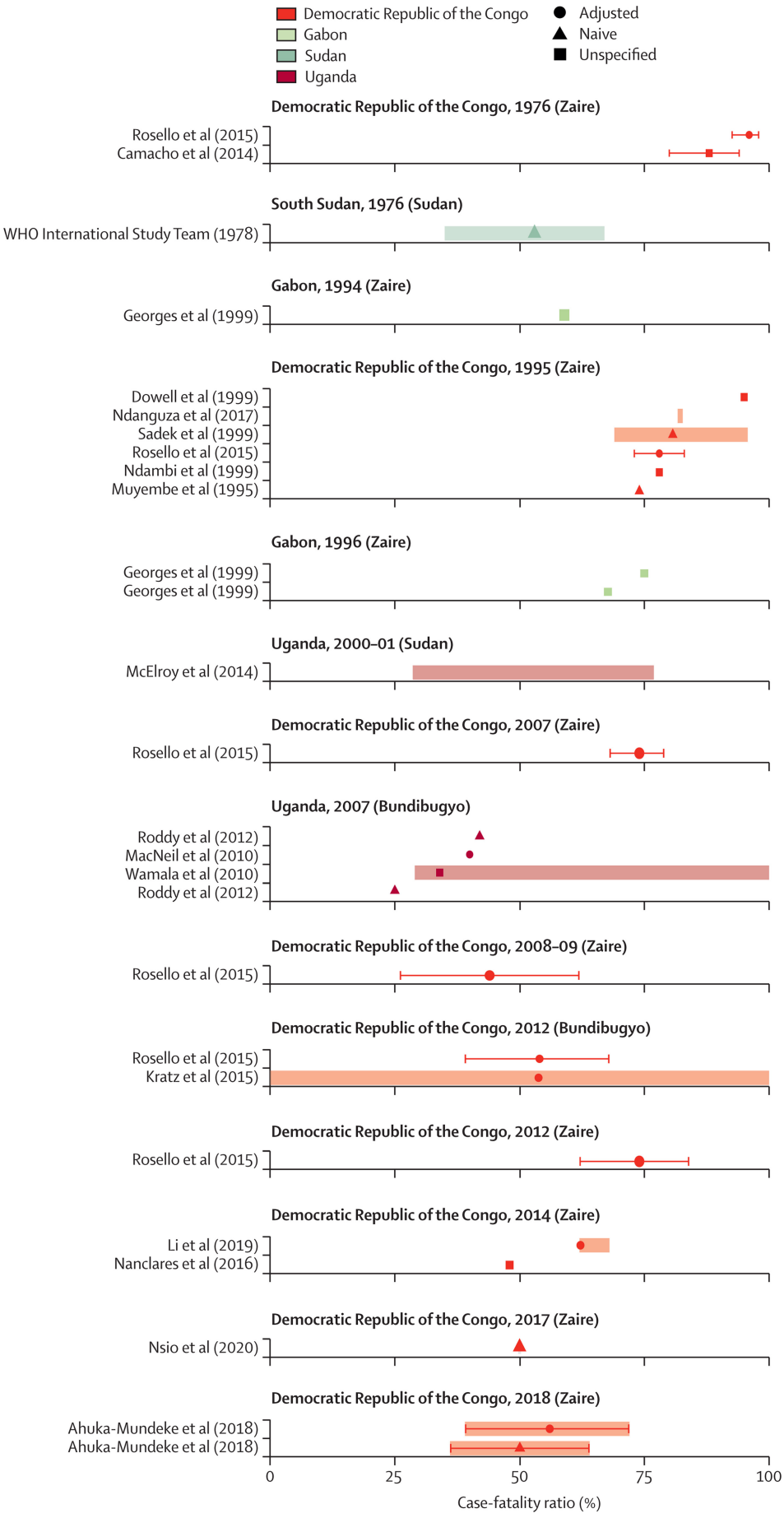


Figure 3.4: **Case-fatality ratio estimates across outbreaks (1/3).** Figure 3.4 continues on the next page.

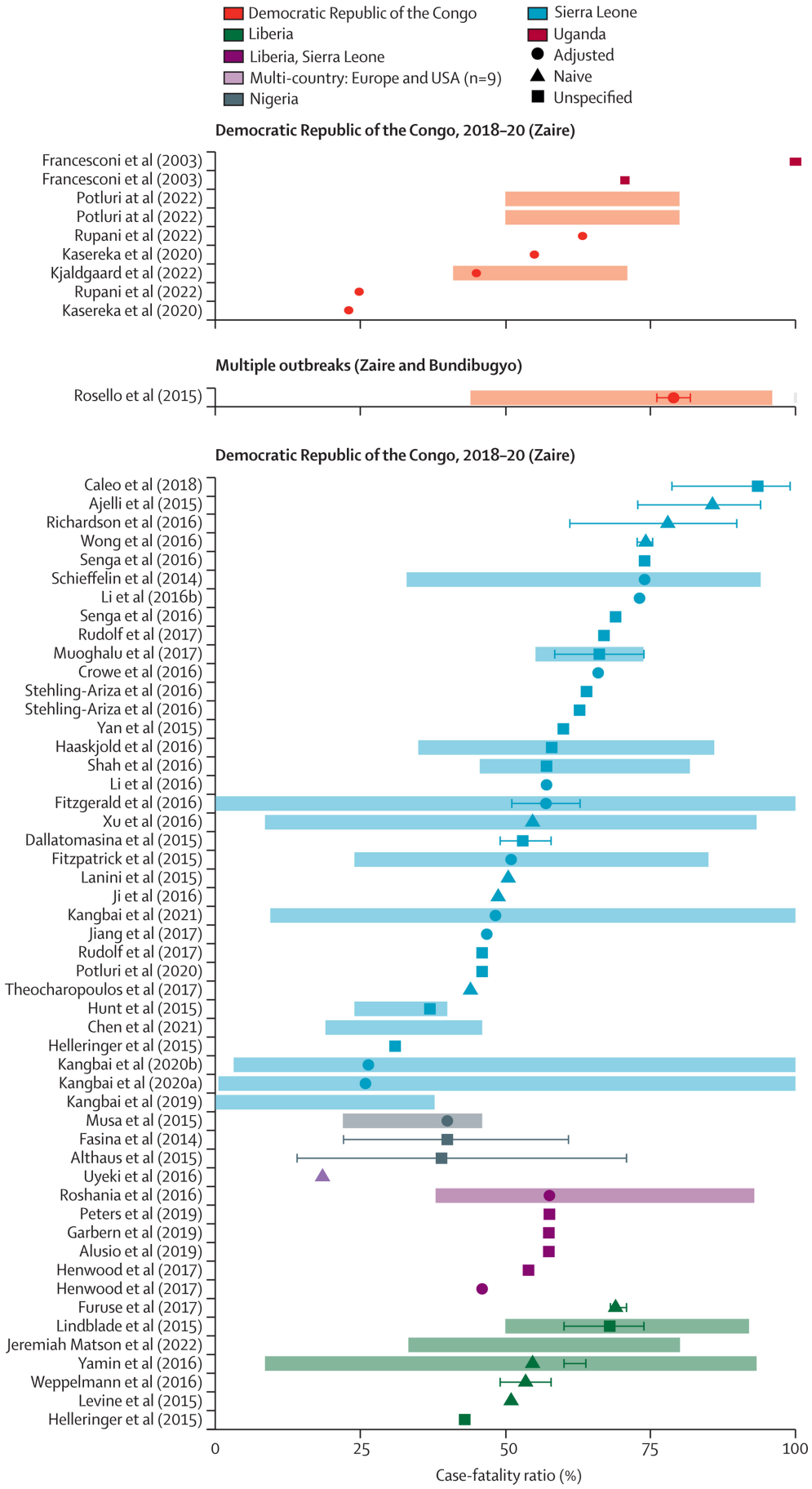


Figure 3.4: **Case-fatality ratio estimates across outbreaks (2/3).** Figure 3.4 continues on the next page.

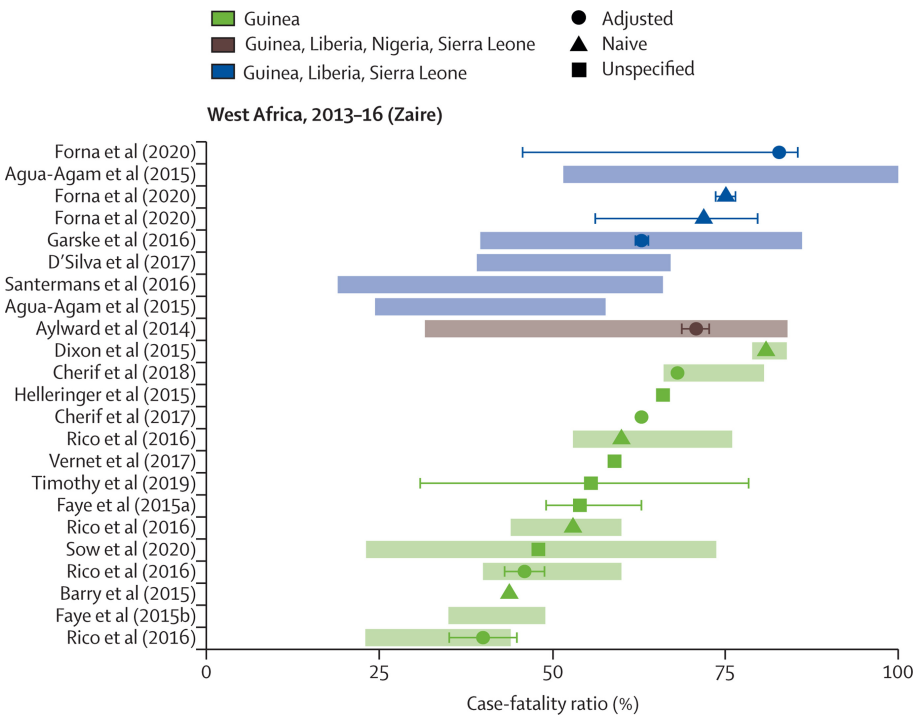


Figure 3.4: **Case-fatality ratio estimates across outbreaks (3/3).** Each section corresponds to a different EVD outbreak, with the associated Ebola virus species in parentheses. Points represent central estimates, with symbol shapes corresponding to analysis type: adjusted, naive, or unspecified. Thick coloured shaded lines represent the range of central estimates when the CFR was estimated across regions or over time. Solid coloured bars represent the uncertainty around the central estimates, reported in different formats including 95% CrIs and 95% CIs. All parameters are from articles with quality assessment scores of 50% or more (appendix Table B.21). Some studies reported multiple estimates for different countries or situations; different publications with the same lead author and year of publication are indicated with lowercase letters (e.g. 2019a and 2019b). Reference list for cited publications is available in the supplementary appendix in [188]. Figure shown as presented in [188].

Table 3.3: **Ranges of case-fatality ratio estimates by outbreak and country.** Total number of parameters specifies the number of parameters (quality assessment filtered ≥50%) included in the summary range. Some parameter entries provide aggregated ranges of central estimates (e.g. across time and countries; Appendix B.3.4) if more than three values were provided. Additionally, not all central estimates were reported with an associated uncertainty interval.

	Central estimate range	Uncertainty range	Total number of parameters
Democratic Republic of the Congo, 1976			
Democratic Republic of the Congo	88–96%	80–98%	2
South Sudan, 1976			
Sudan	35–67%	..	1
Gabon, 1994			
Gabon	59%	..	1
Democratic Republic of the Congo, 1995			
Democratic Republic of the Congo	69–96%	73–83%	6
Gabon, 1996			
Gabon	68–75%	..	2
Uganda, 2000–01			
Uganda	29–77%	..	1
Democratic Republic of the Congo, 2007			
Democratic Republic of the Congo	74%	68–79%	1
Uganda, 2007			
Uganda	25–100%	..	4
Democratic Republic of the Congo, 2008–09			
Democratic Republic of the Congo	44%	26–62%	1
Democratic Republic of the Congo, 2012			
Democratic Republic of the Congo	0–100%	39–84%	3
West Africa 2013–16			
Guinea	23–84%	31–78%	14
Guinea, Liberia, Nigeria, Sierra Leone	32–84%	69–73%	1
Guinea, Liberia, Sierra Leone	19–100%	46–86%	8
Liberia	9–93%	49–74%	7
Liberia, Sierra Leone	38–93%	..	6
Multi-country: Europe and USA (n=9)	18%	..	1
Nigeria	22–46%	14–71%	3
Sierra Leone	0–100%	49–99%	34
Democratic Republic of the Congo, 2014			
Democratic Republic of the Congo	48–68%	..	2
Democratic Republic of the Congo, 2017			
Democratic Republic of the Congo	50%	..	1
Democratic Republic of the Congo, 2018			
Democratic Republic of the Congo	36–72%	36–72%	2
Democratic Republic of the Congo, 2018–20			
Democratic Republic of the Congo	23–80%	..	7
Uganda	71–100%	..	2
Multiple outbreaks			
Democratic Republic of the Congo	44–96%	76–82%	1

Finally, we extracted 295 existing models from 280 studies (appendix Table B.26). 210 (71%) of the 295 published EVD models were compartmental, 19 (6%) used a branching process, and 17 (6%) were agent-based models (appendix Table B.24). There were also 49 other model types or combinations of models. Various assumptions, including homogenous mixing, heterogeneity in transmission rates between groups or over time, and the latent period being the same as the incubation period, were made in the models (appendix Table B.25). Despite this wealth of knowledge, only 37 (13%) of 295 models mentioned publicly available code and we could access the code for 23 (8%) of 295 models, limiting re-usability. Even if we could access the associated code, only 14 (61%) of 23 models with accessible code had accompanying instructions or a README (a text file that introduces the project and explains how to use the code). Of the 23 models with available code (appendix Table B.27), 13 (57%) addressed research questions about transmission, four (17%) focused on method developments, and three (13%) generated forecasts. In addition, seven (30%) were spatial and none considered spillover. 20 (87%) of the 23 models with available code were fitted to observed data, with ten (43%) using Markov chain Monte Carlo methods and five (22%) using maximum likelihood estimation. 15 (65%) of the models were coded in R and three (13%) were coded in MATLAB.

3.5 Discussion

This systematic review presents a comprehensive set of epidemiological parameters and mathematical models for EVD. Most estimates of epidemiological delays were highly variable across studies, which agrees with previous reviews [208, 209]. The heterogeneity is probably driven by differences in epidemic context, with time to hospitalisation often being used as a marker of response performance [65, 66]. We found significant differences between species for incubation periods and time from symptom onset to recovery and death, although there was more evidence for Zaire than for other species. The frequent inconsistency in endpoint definitions (e.g. reporting vs WHO notification or recovery vs testing negative) made comparison of epidemiological delays across studies challenging.

Despite the plethora of published evidence, we chose not to summarise all parameters through meta-analysis because there were too few estimates or the definitions and contexts (across which the parameters were estimated) varied. For example, seroprevalence estimates varied considerably depending on the population groups being sampled. Seropositivity was high when looking at serology in people who had past infections [228] compared with the general population [229]. Differences in R_0 values might be driven by more than just the epidemic context (e.g. some of the early epidemics with high R_0 estimates were driven mostly by nosocomial transmission [230, 231]) or, for instance, differences in the methodological approach taken (e.g. branching processes [65] or using phylogenies [232]). Uncertainty in R_0 is especially high in the Nigeria west African epidemic context, likely because it was a small outbreak [233]. Estimates of seroprevalence [212, 214, 217] secondary attack rates [207]

and R_0 [208] are in line with previous reviews, although we did not focus on specific settings so there is greater variability in previous reviews. In particular, we note the setting for the secondary attack rate is important to explain heterogeneities, with estimates from studies with contact and household settings being higher than community-based settings.

CFR estimates also varied greatly between and within outbreaks, which could be due to differences in the resilience of health-care systems, patient demographics, and the presence of conflict [65, 234, 235]. The range in our estimates is consistent with previous reviews [208, 214–217]. However, there is insufficient evidence to distinguish whether one or more species have an inherently higher severity, or whether the observed differences are driven by contextual factors such as improved case management over time. Summarising parameter estimates that have been generated with widely different methods and relate to varied contexts presents an important statistical problem. Nevertheless, a reasonable synthesis of the plethora of information available from previous outbreaks can provide a useful starting point during a new outbreak. We presented both an average CFR and an average weighted by the sample size informing the estimate. In line with previous findings from a preprint study, [236] we found CFR estimates to be higher in studies with small sample size. We therefore suggest the use of weighted average CFR in modelling studies with sensitivity analyses over a broad range for robustness. Future research could investigate improved methods to synthesise highly heterogenous estimates such as those presented here.

We also extracted parameters not considered in previous systematic reviews. We found substantial but uncertain amounts of overdispersion, ranging from 0.02 to 2.2. Evidence from Lloyd-Smith and colleagues [237] suggests that this overdispersion corresponds to between 30–90% of transmission being attributed to the most infectious 20% of individuals, which is less overdispersive than severe acute respiratory syndrome but highly uncertain. The high evolutionary rate of the Ebola virus supports that genomic data might be an important asset to characterise the transmission dynamics in future outbreaks (e.g. by reconstructing who infected whom [238]). However, similar to most parameters, there is high variability and uncertainty, and most evolutionary estimates are for the Zaire species and the west African epidemic context. Some of the variability in evolutionary rates are likely due to differences in substitution rates between outbreaks compared with within outbreaks [239].

Our study has several limitations. First, due to the extensive literature published on EVD, we restricted our review to published peer-reviewed studies in English. Second, although we attempted to ensure consistency in data extraction by double extracting 55 (11%) of the 522 papers, inconsistencies across extractors or incompleteness of data extraction from studies with multiple parameters are possible. Additionally, due to the subjectivity of quality assessment and scoring of papers as a whole rather than by individual parameters, consensus among reviewers was difficult to achieve. Third, substantial heterogeneity in the way estimates were reported sometimes made direct comparisons between studies challenging. Often, studies did not distinguish whether uncertainty pertained to the sample

or the sample mean, which affected our ability to include them in meta-analyses. Fourth, specificity of the different seroprevalence tests was often not mentioned and historic papers did not always specify the assay used, making comparisons challenging. Fifth, the large range of reported evolutionary rates could reflect our data extraction method; in particular we did not differentiate between studies using samples solely from humans and those including some samples from animals, nor the method used for estimation or sampling. Sixth, we did not extract odds ratios characterising risk factors or the direction of protection or risk, which limited our ability to assess the strength and consistency of associations across studies. For example, studies appeared evenly divided on the significance of funeral participation as a risk factor for seropositivity, but without directionality it is difficult to interpret these findings or compare results across studies - particularly given variation in the representativeness of data sources and other methodological differences, such as how exposures were defined (e.g., funeral participation involving direct contact with the deceased versus attendance). We also acknowledge that significance classification is somewhat arbitrary and dependent on study design. We encourage readers specifically interested in risk factors to investigate these papers further, for example by accessing the data from this review through the epireview R package [189].

3.6 Conclusion

Being prepared at the onset of an infectious disease outbreak is imperative to mount a rapid and effective response to combat the spread of disease. Here we present, synthesise, and analyse the breadth of evidence on the transmissibility, severity, delays, risk factors, mutation rates, and seroprevalence of Ebola virus, as well as identify transmission models for EVD, expanding on previous modelling reviews [210, 213]. We curated 1280 parameter estimates and 295 model descriptions from 522 studies and made our data available through an easy-to-use R package, epireview [189]. We expect that this comprehensive repository will serve as an important resource for modellers and the public health community, who can add to this dynamic database as and when more evidence becomes available, ensuring that this database provides a live picture of the latest evidence on EVD. Much is already known about the Ebola Zaire species; however, our review highlights a crucial shortage of evidence for other species, such as Sudan, Bundibugyo, and Taï Forest, which is important given that the 2022 EVD outbreak was of the Sudan species. Initial analyses suggest statistically significant differences in key parameters between species, such as the incubation period and delays from onset to death or recovery. We suggest analysing any data collected during the 2022 Ugandan outbreak along with preparing countries to collect and analyse data from future outbreaks for wider research could help fill this knowledge gap. Finally, we note the paucity of publicly available source code for EVD models (8%). Publicly releasing the source code and instructions for future models will increase usability of existing models in real-time settings.

Chapter 4

Estimating the transmission advantage of pathogen variants in real-time: SARS-CoV-2 as a case-study

The contents of this chapter have been adapted from a manuscript previously published in the journal Epidemics [240]. I was a joint first author of this manuscript and it was written with input from all co-authors. I worked on the simulation study, the development of the software and the literature review. I also created the vignette for the software and the supplementary database of other methods used to estimate the transmission advantage of pathogen variants (available in [240]).

A key theme identified in Chapter 2 was the need for R_t estimation tools to provide practical and logistically relevant outputs. In this chapter I describe the development of a software tool to estimate the transmission advantage of pathogen variants in real-time, which was ultimately used to inform the UK government response to COVID-19 [241].

4.1 Introduction

The SARS-CoV-2 pandemic has highlighted the potentially dramatic influence that emerging novel pathogen variants can have on transmission dynamics and on the control measures needed to mitigate the epidemic burden. The emergence of the Alpha variant of SARS-CoV-2 in September 2020, and of the Delta variant in December 2020 drastically altered the trajectory of the COVID-19 epidemic in several countries leading to renewed imposition of public health measures such as lockdowns [242, 243]. The continued high level of transmission of SARS-CoV-2 globally makes the emergence of new variants very likely. As of April 2023, the World Health Organization has classified five variants of SARS-CoV-2 as Variants Of Concern (VOCs) (i.e. Alpha, Beta, Gamma, Delta, and Omicron), because of their increased transmissibility, severity, and/or immune escape properties compared to the

circulating SARS-CoV-2 variants [244]. Rapidly quantifying characteristics of such emerging variants is critical to anticipate their potential impact and adjust interventions accordingly. Estimates of the transmission advantage of a new variant over previously circulating variants can help re-evaluate key metrics which depend on the reproduction number, e.g. the herd-immunity threshold, and adjust short-term projections of the epidemic trajectory. They can also help to disentangle the impact on transmission levels of the newly introduced variant from other factors, such as concurrent changes in control measures [245]. Shortly after the emergence of the Alpha variant in England in September 2020 [246], a number of studies aimed to estimate its transmission potential, compared to the previously circulating non-VOC lineages [93–101]. More recently, several papers have evaluated the transmissibility of the Delta variant compared to Alpha [245, 247–251], and of one or more VOCs [96, 250, 252–257]. All of these studies have developed new approaches to estimate the transmission advantages of the Alpha and the Delta variants, often synthesising evidence from multiple data sources including genomic data. The time and expertise required to design and implement such approaches, with methods tailored to the specificity of each dataset and context, often limit their widescale and real-time use.

In this study, we present a new Bayesian inference method, MV-EpiEstim (for Multi-Variant EpiEstim), to estimate in real-time the transmission advantage of a new variant of a pathogen compared to a reference variant, using simple data consisting of the time series of incidence of cases of each variant in one or more locations. The aims of this work are to: 1) develop a statistically robust method and accompanying software tool to estimate the transmission advantage; 2) assess how well it can estimate the transmission advantage across a range of simulated and real-life epidemic scenarios.

In the rest of the manuscript, we refer to different "variants" but the method can be equally applied to different strains. We present the method for one reference and one new variant, but the method naturally extends to more than one new variant. Our work builds on a previously published methodology [62, 114] to estimate the instantaneous reproduction number R_t.

We assume that locally, the transmissibility of all variants follows the same temporal pattern, i.e. the reproduction number of the new variant is the same as that of the reference variant, albeit with a multiplicative factor. We refer to this multiplicative factor as the "effective transmission advantage" of the new variant, compared to the reference variant. We further assume that the effective transmission advantage remains constant over a user-defined time-window and across all locations under consideration. Note that both the time-window and the set of locations over which the effective transmission advantage is assumed to be constant can be varied by the user.

We provide an open source implementation of our method in the R package EpiEstim [258]. The approach, which we validate on an extensive simulation study, is computationally efficient as it takes advantage of an analytical formulation of marginal posterior densities of both the instantaneous reproduction number for the reference variant, and the transmission advantage of the new variant.

We illustrate the use of our tool by retrospectively estimating the effective transmission advantage

of SARS-CoV-2 VOCs (Alpha, Beta/Gamma, and Delta) over the previously circulating variants using data from England and France. In addition, we perform a literature review to summarise other existing approaches and tools available for estimating the transmission advantage of new variants from incidence data. We show that the estimates from our method are consistent with those from other studies, and that our fast ready-to-use tool allows timely estimation and easy exploration of changes in the transmission advantage over time and space. Our inference framework and open source software should allow rapid quantification and monitoring of the effective transmission advantage of future new variants in real-time.

4.2 Methods

We extend the methodology from Cori et al. [62] and Thompson et al. [114] to develop an inference framework for jointly estimating the transmissibility (instantaneous reproduction number R_t) of a reference variant and the effective transmission advantage of novel variants, compared to the reference. For simplicity, we present the method for two variants only (a reference and a new variant). The method is applicable to, and has been implemented for, estimating the transmission advantages of multiple variants over a single reference.

Assumptions. Our method relies on daily incidence data of the reference and the variant. Where data from more than one location are used, we assume that the epidemics in each location are independent and closed except for cases of each variant on the first day who are assumed to be imported. The effective reproduction number is defined as the ratio of newly infected cases to the total infectiousness (due to past cases) in a location. For more details, see [62, 114].

Notation. We use the following notations:

- Indexes t for time, l for location and v for variant, with $v = 0$ denoting the reference variant and $v = 1$ the new variant,
- n_l the number of locations considered
- T the number of days of observation
- $I_t^{l,v}$ denotes the number of incident cases of variant v at time t in location l,
- $R_t^{l,v}$ denotes the instantaneous reproduction number for variant v at time t in location l. For simplicity we use R_t^l to denote the instantaneous reproduction number for the reference variant in location l i.e. $R_t^l = R_t^{l,0}$.
- w^v is the probability mass function of the discrete Serial Interval (SI) for variant v, assumed the same across all locations, but potentially different between variants (w_s^v is the probability that the SI lasts s days, $s = 1, \ldots, SI_{max}$; and we assume $w_0^v = 0$).
- For simplicity we introduce the generic notation $\mathbf{X}_t^{L,V} = \left\{X_t^{l,v}\right\}_{l=1,\ldots,n_l;v=0,1}$ for the variable X at time t across all locations and both variants.

Following the notation above, we estimate the overall infectiousness for variant v at time t and in location l due to past incident cases of that variant in that location as

$$\Lambda_t^{l,v} = \sum_{s=1}^{t} I_{t-s}^{l,v} w_s^v. \tag{4.1}$$

We assume two independent branching processes for the incidence of cases caused by the reference variant and the new variant. We assume that the daily incidence of cases caused by the reference variant at time t in location l (I_t^l) follows a Poisson distribution:

$$I_t^l \sim Poisson(R_t^l \Lambda_t^l) \tag{4.2}$$

Similarly, the daily incidence of cases caused by the new variant at time t in location l ($I_t^{l,1}$) also follows a Poisson distribution:

$$I_t^{l,1} \sim Poisson(R_t^{l,1} \Lambda_t^{l,1}), \tag{4.3}$$

where

$$R_t^{l,1} = \epsilon R_t^l \tag{4.4}$$

i.e. the reproduction number of the new variant is proportional to that for the reference variant; the proportional factor ϵ is the effective transmission advantage (if $\epsilon > 1$, or disadvantage if $\epsilon < 1$) of the new variant compared to the reference variant, assumed constant over a time-window and across a set of locations defined by the user.

In the software, we temporarily assign the most transmissible variant as the 'reference' during estimation, irrespective of which variant was originally supplied as the reference variant (see Algorithm 1). This reordering stabilises inference by ensuring that ϵ is bounded between 0 and 1, avoiding an unbounded posterior space (1 to $+\infty$) when $\epsilon > 1$. After inference is complete, if reordering occurred, this reordering is reversed by reporting the reciprocal of ϵ (i.e. $1/\epsilon$), so that the output reflects ϵ for the user-specified new variant relative to the user-specified reference variant.

In the simulation study, all scenarios explored values of $\epsilon > 1$ for simplicity and because they align with typical use cases (e.g., evaluating variants of concern). Given the internal reordering of variants within the software (Algorithm 1), the estimation process itself is agnostic to which variant has the advantage - values of $\epsilon < 1$ effectively correspond to swapping the reference and new variant.

The likelihood of the time series of incident cases of the reference and the new variants can be written as

$$\mathcal{L}_t = P\left(\mathbf{I}_t^{L,V} | \mathbf{I}_0^{L,V}, \ldots, \mathbf{I}_{t-1}^{L,V}, \mathbf{w}^V, \mathbf{R}_t^{L,V}, \epsilon\right) \tag{4.5a}$$

$$= \prod_{l=1}^{n_l} \prod_{v=0}^{1} P(I_t^{l,v} | \mathbf{I}_0^{L,V}, \ldots, \mathbf{I}_{t-1}^{L,V}, \mathbf{w}^V, R_t^{l,v}) \tag{4.5b}$$

$$= \prod_{l=1}^{n_l} \prod_{v=0}^{1} \frac{\left(R_t^{l,v}\right)^{I_t^{l,v}} e^{-R_t^{l,v}\Lambda_t^{l,v}}}{I_t^{l,v}!} \tag{4.5c}$$

$$\propto \prod_{l=1}^{n_l} (R_t^l)^{I_t^{l,0}} (\epsilon R_t^l)^{I_t^{l,1}} e^{-R_t^l(\Lambda_t^{l,0} + \epsilon\Lambda_t^{l,1})} \tag{4.5d}$$

$$\propto \prod_{l=1}^{n_l} \left(\left(R_t^l\right)^{\sum_{v=0}^{1} I_t^{l,v}} \epsilon^{I_t^{l,1}} e^{-R_t^l\left(\Lambda_t^{l,0} + \epsilon\Lambda_t^{l,1}\right)} \right) \tag{4.5e}$$

$$\propto \epsilon^{\sum_{l=1}^{n_l} I_t^{l,1}} e^{-\sum_{l=1}^{n_l} R_t^l\left(\Lambda_t^{l,0} + \epsilon\Lambda_t^{l,1}\right)} \prod_{l=1}^{n_l} \left(R_t^l\right)^{\sum_{v=0}^{1} I_t^{l,v}} \tag{4.5f}$$

Equation (4.5a) defines $\mathcal{L}_t$ as the conditional probability of observing the incidence of cases across all locations and variants at time t ($\mathbf{I}_t^{L,V}$), given the observed past incidence ($\mathbf{I}_0^{L,V}, \ldots, \mathbf{I}_{t-1}^{L,V}$), the variant-specific serial interval distributions ($\mathbf{w}^V$) and model parameters ($\mathbf{R}_t^{L,V}$ and ϵ).

To evaluate this, we assume conditional independence, meaning that given the past incidence and parameters, the current incidence in each location and for each variant ($I_t^{l,v}$) are independent of each other (Equation (4.5b)). This allows us to factor the likelihood across all locations and variants, as shown in Equations (4.5b) and (4.5c), where we model incidence using Poisson distributions following the renewal equations defined in Equation (4.2) and Equation (4.3).

Next, we simplify Equation (4.5c) by removing the factorials in the denominator ($I_t^{l,v}!$), which are constant with respect to the parameters being estimated. Equation (4.5d) shows a likelihood expression that is proportional to ($\propto$) the original, where we introduce the $R_t^{l,1} = \epsilon R_t^l$ relationship (Equation (4.4)) to express the variant-specific transmission in terms of a shared baseline and relative advantage (or disadvantage). In Equation (4.5e) the variant-specific powers of R_t^l are combined into a single power using $R^a R^b = R^{a+b}$ and more clearly shows the dependence of the likelihood on the shared baseline reproduction number R_t^l.

Finally, Equation (4.5f) rearranges Equation (4.5e) to clearly separate the overall contribution of ϵ from R_t^l.

We assume Gamma priors for each R_t^l, with the same shape a and scale b across times and locations, and for ϵ, with shape c and scale d. These priors reflect our initial beliefs about the distributions of the parameters before observing any data. In the software, the priors are relatively uninformative by default: ϵ has shape 1 and scale 1, corresponding to a mean of 1 and standard deviation of 1. Each R_t^l has shape 0.04 and scale 25, corresponding to a mean of 1 and standard deviation of 5. These priors can be altered by the user to better reflect their own assumptions. The joint posterior distribution

represents our updated beliefs about the parameters after observing the data $(\mathbf{I}_0^{L,V}, \ldots, \mathbf{I}_T^{L,V})$ assuming that the SI distributions for both variants $\mathbf{w}^V$ are known:

$$P\left(\epsilon, \mathbf{R}_1^L \ldots, \mathbf{R}_T^L \mid \mathbf{I}_0^{L,V}, \ldots, \mathbf{I}_T^{L,V}, \mathbf{w}^V\right) \tag{4.6a}$$

$$\propto \prod_{t=1}^{T} \mathcal{L}_t \times \; P(\epsilon) \prod_{t=1}^{T} \prod_{l=1}^{n_l} P(R_t^l) \tag{4.6b}$$

$$\propto \prod_{t=1}^{T} \left(\mathcal{L}_t \prod_{l=1}^{n_l} \left(R_t^l\right)^{a-1} e^{-\frac{R_t^l}{b}} \right) \epsilon^{c-1} e^{-\frac{\epsilon}{d}} \tag{4.6c}$$

$$\propto \left(\prod_{t=1}^{T} \prod_{l=1}^{n_l} \left(R_t^l\right)^{a-1+\sum_{v=0}^{1} I_t^{l,v}} \right) \epsilon^{c-1+\sum_{t=1}^{T}\sum_{l=1}^{n_l} I_t^{l,1}} \tag{4.6d}$$

$$\times \; e^{-\frac{\epsilon}{d} - \sum_{t=1}^{T}\sum_{l=1}^{n_l} R_t^l \left(\frac{1}{b} + \Lambda_t^{l,0} + \epsilon \Lambda_t^{l,1}\right)}$$

Equation (4.6a) defines the posterior as the conditional probability of ϵ and reproduction numbers across all times and locations $(\mathbf{R}_1^L \ldots, \mathbf{R}_T^L)$ given the observed incidence $(\mathbf{I}_0^{L,V}, \ldots, \mathbf{I}_T^{L,V})$ across all variants, locations and times, and given the serial interval distributions $(\mathbf{w}^V)$.

Using Bayes' theorem, we can express the posterior as being proportional to the product of the likelihoods ($\mathcal{L}_t$ - the probability of the observed data at each time step, as defined in Equation (4.5)) and the prior distributions for ϵ and for each R_t^l (Equation (4.6b)). In Equation (4.6c) we express the priors explicitly as Gamma distributions: $R_t^l \sim Gamma(a, b)$, $\epsilon \sim Gamma(c, d)$ so that $P(R_t^l) \propto (R_t^l)^{a-1} e^{-\frac{R_t^l}{b}}$ and $P(\epsilon) \propto \epsilon^{c-1} e^{-\epsilon/d}$.

Equation (4.6d) then rewrites the joint posterior by substituting the simplified form of the likelihood (from Equation (4.5f)) and incorporating the prior terms into a single expression. The exponents on R_t^l and ϵ are increased by the observed case counts, reflecting how the data updates the Gamma priors (shape parameters a and c). Similarly, the final exponential term combines contributions from both priors and the likelihood. It penalises parameter combinations that are unlikely under the priors (given the scale parameters b and d) and that, through the overall infectiousness implied by past case history, would poorly explain the observed incidence.

The joint posterior is not analytically tractable because Equation (4.6d) does not correspond to the probability density function of any standard multivariate distribution. The presence of terms involving the product of ϵ and R_t^l introduces dependencies that make joint sampling challenging. Instead, we exploit that the conditional posterior distributions for ϵ and R_t^l are Gamma distributions, which are analytically tractable. This allows us to use a Gibbs sampler, a Markov Chain Monte Carlo (MCMC) algorithm, in an iterative process in which we alternate between updating ϵ given the current estimate for R_t^l, and updating R_t^l given the current value of ϵ, until the estimates converge. Compared to Metropolis-Hastings, the Gibbs sampler is more computationally efficient as it draws directly from the

marginal posterior distributions without an acceptance step. The full implementation is described in the MCMC inference section below.

The conditional posterior distribution for ϵ describes our updated belief about the value of ϵ given the observed incidence data (for all variants, locations, and time steps) and the reproduction numbers for the reference variant across all locations and time steps. By conditioning on R_t^l and isolating the terms that involve ϵ from Equation (4.6d) we obtain:

$$\begin{aligned} &P\left(\epsilon \mid \mathbf{I}_0^{L,V}, \ldots, \mathbf{I}_T^{L,V}, \mathbf{R}_1^L \ldots, \mathbf{R}_T^L, \mathbf{w}^V\right) \\ &\propto \epsilon^{c-1+\sum_{t=1}^{T}\sum_{l=1}^{n_l} I_t^{l,1}} e^{-\epsilon\left(\frac{1}{d}+\sum_{t=1}^{T}\sum_{l=1}^{n_l} R_t^l \Lambda_t^{l,1}\right)} \end{aligned} \tag{4.7}$$

Equation (4.7) shows a Gamma distribution with shape $c + \sum_{t=1}^{T}\sum_{l=1}^{n_l} I_t^{l,1}$ and scale $\frac{1}{\frac{1}{d}+\sum_{t=1}^{T}\sum_{l=1}^{n_l} R_t^l \Lambda_t^{l,1}}$.

Similarly, by conditioning on ϵ and isolating the terms involving R_t^l from Equation (4.6d), we obtain the conditional posterior:

$$\begin{aligned} &P(R_t^l \mid \mathbf{I}_0^{L,V}, \ldots, \mathbf{I}_T^{L,V}, \epsilon, \\ &\quad \mathbf{R}_1^L \ldots, \mathbf{R}_{t-1}^L, \mathbf{R}_{t+1}^L, \ldots, \mathbf{R}_T^L, \mathbf{R}_t^1 \ldots, \mathbf{R}_t^{l-1}, \mathbf{R}_t^{l+1} \ldots, \mathbf{R}_t^{n_l}, \mathbf{w}^V) \\ &\propto \left(R_t^l\right)^{a-1+\sum_{v=0}^{1} I_t^{l,v}} e^{-R_t^l\left(\frac{1}{b}+\Lambda_t^{l,0}+\epsilon\Lambda_t^{l,1}\right)} \end{aligned} \tag{4.8}$$

Equation (4.8) also shows a Gamma distribution, with shape $a + \sum_{v=0}^{1} I_t^{l,v}$ and scale $\frac{1}{\frac{1}{b}+\Lambda_t^{l,0}+\epsilon\Lambda_t^{l,1}}$.

Precision in ϵ estimates. An analytical formulation of the marginal posterior distribution of ϵ allows us to quantify the expected precision in ϵ estimates. Precision refers to how narrowly concentrated a posterior distribution is around its mean and can be expressed through the Coefficient of Variation (CV). Since the CV of a gamma distributed random variable is $1/\sqrt{shape}$, the CV of the marginal posterior distribution of ϵ is

$$\begin{aligned} &CV\left(\epsilon \mid \mathbf{I}_0^{L,V}, \ldots, \mathbf{I}_T^{L,V}, \mathbf{R}_1^L \ldots, \mathbf{R}_T^L, \mathbf{w}^V\right) \\ &= \frac{1}{\sqrt{c + \sum_{t=1}^{T}\sum_{l=1}^{n_l} I_t^{l,1}}} \\ &\leq \frac{1}{\sqrt{\sum_{t=1}^{T}\sum_{l=1}^{n_l} I_t^{l,1}}} \end{aligned} \tag{4.9}$$

That is, the CV of ϵ scales inversely with the square root of the total incidence of the new variant across all locations in the time-window used for estimation. Therefore, as the observed incidence data for the new variant increases, the uncertainty in our estimate of ϵ decreases. When incidence is low, estimates of ϵ will be more uncertain and strongly influenced by the prior.

Markov Chain Monte Carlo (MCMC) inference. The analytical formulation of the marginal posterior distributions for R_t^l and ϵ allow us to use a multi-stage Gibbs sampler for the MCMC inference.

To initialise R_t^l, we use EpiEstim to estimate a single reproduction number for the reference variant over the entire time period of observations, and using incidence aggregated across all locations. The posterior mean is then used as the initial value for R_t^l. We independently use the same approach to estimate a single reproduction number for the new variant; ϵ is then initialised to the median of the ratio of the reproduction numbers for the new variant and the reference.

We first sample from the marginal distribution of R_t^l, conditional on ϵ, and then we sample from the marginal distribution of ϵ, conditional on the newly sampled value of R_t^l. We repeat this procedure for a fixed number of iterations or until convergence is achieved. Convergence is assessed using Gelman-Rubin convergence diagnostic [259] using 1.1 as a cut-off value.

Implementation. The inference method is implemented in a new function "estimate_advantage" of the development version of the R package EpiEstim available at `https://github.com/mrc-ide/EpiEstim`. The pseudocode for the new function is provided in Algorithm 1.

Algorithm 1: Pseudocode for `estimate_advantage()` applied to two variants, A and B.

Input:

`incid` ← 3D array [time, location, variant] of incidence data. Variant A (index 1) is the reference and Variant B (index 2) is the "new" variant
`si_distr` ← matrix [time, variant] of discrete serial interval probability mass functions
`priors` ← list of Gamma prior parameters (shape and scale) for R and ϵ
`mcmc_control` ← list with MCMC settings:
  `n_iter` ← integer, total number of MCMC iterations
  `burnin` ← integer, number of initial iterations to discard (burn-in period)
  `thin` ← integer, thinning interval - keep every ith sample after burn-in
`t_min`, `t_max` ← integers, start and end time steps of estimation
`reorder_incid` ← boolean. If TRUE, variants are internally reordered temporarily so that the most transmissible becomes the reference (index 1) during estimation. Results are later mapped back to the original order

1) Initialisation:

Create empty `epsilon` [new variant, iteration] and `R` [time, location, iteration] for outputs
Compute overall infectivity `lambda` [time, location, variant] from `incid` and `si_distr` (Equation (4.1))
Estimate single initial R (`R_init`) for each variant using incidence aggregated across all locations and estimated over all time steps (`t_min` : `t_max`) (Equation (1.1))

Re-order variants
if *`reorder_incid` = TRUE* ***and*** *`R_init` for Variant B > `R_init` for Variant A* **then**
  Reorder variants so that Variant B is at index 1 and treated as the reference
  Reorder `incid`, `lambda`, and `R_init` accordingly
  Estimate initial ϵ (`epsilon`$[,1]$) ← `R_init` for Variant A / `R_init` for Variant B
else
  `epsilon`$[,1]$ ← `R_init` for Variant B / `R_init` for Variant A
end

2) MCMC using Gibbs sampling - alternate conditional draws of R and ϵ:

Sample `R`$[\ ,\ ,1]$ from its gamma conditional posterior given `epsilon`$[,1]$, `incid`, `lambda`, and `priors` (Equation (4.8))

for $i = 1$ *to* *`n_iter`* **do**
  Sample `R`$[,,i+1]$ given `epsilon`$[,i]$, `incid`, `lambda`, and `priors` (Equation (4.8))
  Sample `epsilon`$[,i+1]$ given `R`$[,,i+1]$, `incid`, `lambda`, and `priors` (Equation (4.7))
end

Remove `burnin` and apply `thin` to `epsilon` and `R` chains
Compute convergence diagnostics (Gelman-Rubin) for each ϵ chain

3) Output:

Reverse variant re-ordering (if applicable)
Invert `epsilon` to express the advantage of Variant B relative to Variant A (original reference): `epsilon` ← 1/`epsilon`
Rescale `R` to obtain estimates for Variant A again: `R` ← `R` / `epsilon`

return list:

`epsilon` ← matrix of relative transmissibility compared to the reference variant. Each row is a "new" variant (in this example only 1 - Variant B)
`R` ← array of R estimates for the reference variant (Variant A)
`convergence` ← logical vector indicating whether the MCMC chain for epsilon has converged
`diag` ← full Gelman-Rubin diagnostics for epsilon

Choosing a time-period for estimation of ϵ. Users can set the time period over which estimation will be carried out. We recommend that the estimation is started after at least one generation of cases has been observed. The default starting point in the software is set to the first day of non-zero incidence across all locations plus the 95^{th} percentile of the SI distribution.

Classification of a variant. We used the posterior distribution of the effective transmission advantage to classify a new variant (in relation to the reference variant) as:

- 'More transmissible' if the 2.5^{th} quantile of the posterior distribution was greater than 1;
- 'Less transmissible' if the 97.5^{th} quantile of the posterior distribution was less than 1; and,
- 'Unclear' if the 95% Credible Interval (CrI) contained 1.

We note that here we used the 2.5 and 97.5 posterior percentiles for variant classification, which provides an easy metric to summarise across simulations. However our EpiEstim implementation of the approach provides the entire posterior distribution of the transmission advantage, and therefore the user could use different thresholds for classification, balancing the desired sensitivity and specificity. The user could also quantify the posterior probability that a novel variant is more transmissible.

Method validation. We assessed the validity of our method using a large simulation study, where we considered several scenarios with different values for the transmissibility of each variant, allowing for superspreading, under-reporting or time-varying transmission advantages, as well as differences in natural history between variants (Appendices C.5 and C.5.1). Case incidence for the reference and new variants were simulated independently following Equations (4.2) and (4.3), where at each time step the incidence of new cases is drawn from a Poisson distribution with a mean equal to the sum of past infections weighted by the serial interval and scaled by either the reproduction number (for the reference variant), or the reproduction number multiplied by ϵ (for the new variant). We seeded the epidemic with 20 cases of the reference variant per day for the first 10 days, and a single case of the new variant on the 10^{th} day. We then simulated forward from day 11 for 100 additional days, generating 100 stochastic epidemic trajectories for each simulation scenario and for each combination of parameter values considered within that scenario (Table C.5). For scenarios with multiple locations, each location is assumed independent, with a closed population. The effective transmission advantage was then estimated using 10, 20, 30, or 50 days of data starting from day 11 (see Appendix C.3.2). Pseudocode to illustrate the simulation process is provided in Algorithm 2.

Algorithm 2: Pseudocode for simulating the epidemic trajectories of two variants.

Input:

`nlocations` ← integer, number of locations to simulate
`nvariants` ← integer, number of variants to simulate (2 in this case)
`ndays` ← integer, number of days to run the simulation for
`incid_init` ← list of length `nvariants`, each containing a vector of daily incidence used to seed the epidemic:
    `incid_init[[1]]` ← 20 cases per day for 10 days (reference variant)
    `incid_init[[2]]` ← 1 case on day 10 (new variant)
`rt_ref` ← numeric, simulated reproduction number for the reference variant
`epsilon` ← numeric, simulated transmission advantage of the "new" variant
`simatrix` ← matrix [time, variant] of discrete serial interval probability mass functions
`nsims` ← integer, number of simulations to run
`p_report` ← numeric, probability of cases being reported (default = 1)

1) Initialisation:

Compute R for the new variant: `rt_variant` ← `epsilon` × `rt_ref`
Initialise R array [time, location, variant]:
    `R`[, , 1] ← `rt_ref` across time and locations
    `R`[, , 2] ← `rt_variant` across time and locations
`init_length` ← length of `incid_init[[1]]`
Initialise empty list `sim_results` (for the output of `nsims` simulations)

2) Simulate incidence:

while *`length(sim_results)` < `nsims`* **do**
    Initialise array `I`[time, location, variant]
    Set first `init_length` rows using `incid_init`
    for *each day* $t = 1$ *to* ***`ndays`****, location* l *and variant* v **do**
        Estimate overall infectiousness and simulate cases:
        Equation (4.1):
        `overall_infectiousness`[t, l, v] ← $\sum_{s=1}^{t}$ `I`$[t - s +$ `init_length`$, l, v] \times$ `si_matrix`$[s, v]$
        Equation (4.2) and Equation (4.3):
        `expected_cases`[t, l, v] ← `R`[t, l, v] × `overall_infectiousness`[t, l, v]
        `I`[t + `init_length`, l, v] ∼ Poisson(`expected_cases`[t, l, v])
    end
    Apply probability of cases being reported
    for *each time, location, and variant* **do**
        `reported`[t, l, v] ∼ Binomial(`I`[t, l, v], `p_report`)
    end
    Retain only informative simulations
    if *cumulative reported cases for 'new' variant in first 20 days* < 5 **then**
        Discard simulation
    else
        Append `reported` to `sim_results`
    end
end

3) Output:

`sim_results`: list of length `nsims`, each an array [time, location, variant] of simulated reported incidence

We measured method performance on the simulations using several metrics (detailed in Appendix C.5.1). First, we measured the bias as the difference between the mean posterior estimate of the effective transmission advantage and its true value. Second, we measured uncertainty in the posterior distribution and considered the coverage probability, which measures whether uncertainty in estimates is adequate. Finally, we considered the ability of the model to adequately classify the variant as "more transmissible", "less transmissible" than the reference or "unclear".

Literature review. On 21st February 2023, we searched for all studies published from 2020 onward which estimated the transmission advantage of SARS-CoV-2 variants. The following search terms were entered into Web of Science: AB=((Transmission OR Transmissibility) & (Variant OR VOC) & (estimat*) & (SARS-CoV-2 OR COVID)). In total, 336 studies were identified in the search and uploaded to Covidence for screening. 244 studies were excluded after title and abstract screening, and an additional 31 studies were excluded during full text screening. For completeness, 5 additional papers that were found prior to the search were also included. Of these 66 (61 + 5) studies, 53 explicitly provided one or more estimates of the transmission advantage of one variant over another. For these, we extracted the value and type of transmission advantage estimated, as well as the method used and its availability (or not) as packaged software. See Appendix C.6 and the supplementary database (available in [240]) for further details.

4.3 Results

Transmission advantage of SARS-CoV-2 variants

We used MV-EpiEstim to retrospectively estimate the transmission advantage of SARS-CoV-2 variants using data from England and France. The Alpha variant originated in late summer to early Autumn 2020 in England (before vaccination was initiated), where it became dominant in early 2021 (Figure 4.1A). The Delta variant, first detected in India, emerged in England around March 2021 and accounted for most SARS-CoV-2 cases in all regions by late Spring 2021 (appendix Figure C.1). England never experienced substantial transmission of the Beta and Gamma variants, first detected in South Africa and Brazil respectively [244].

In France, the Alpha variant emerged in early 2021, rapidly dominating cases in metropolitan France and the French West Indies [260, 261]. The Beta and Gamma variants were also circulating from January 2021 in most regions, and accounted for the majority of cases in French Guyana and la Réunion from Spring 2021 [262] (appendix Figure C.2). The Delta variant emerged in France in early June 2021 after the period covered in this study.

We considered daily variant-specific incidence data from 7 National Health Service (NHS) regions in England between 1st September 2020 and 14th March 2021 (appendix Figure C.1), and from 18 ADM2 regions in France between 18th February and 30th May 2021 (appendix Figures C.2 and C.6).

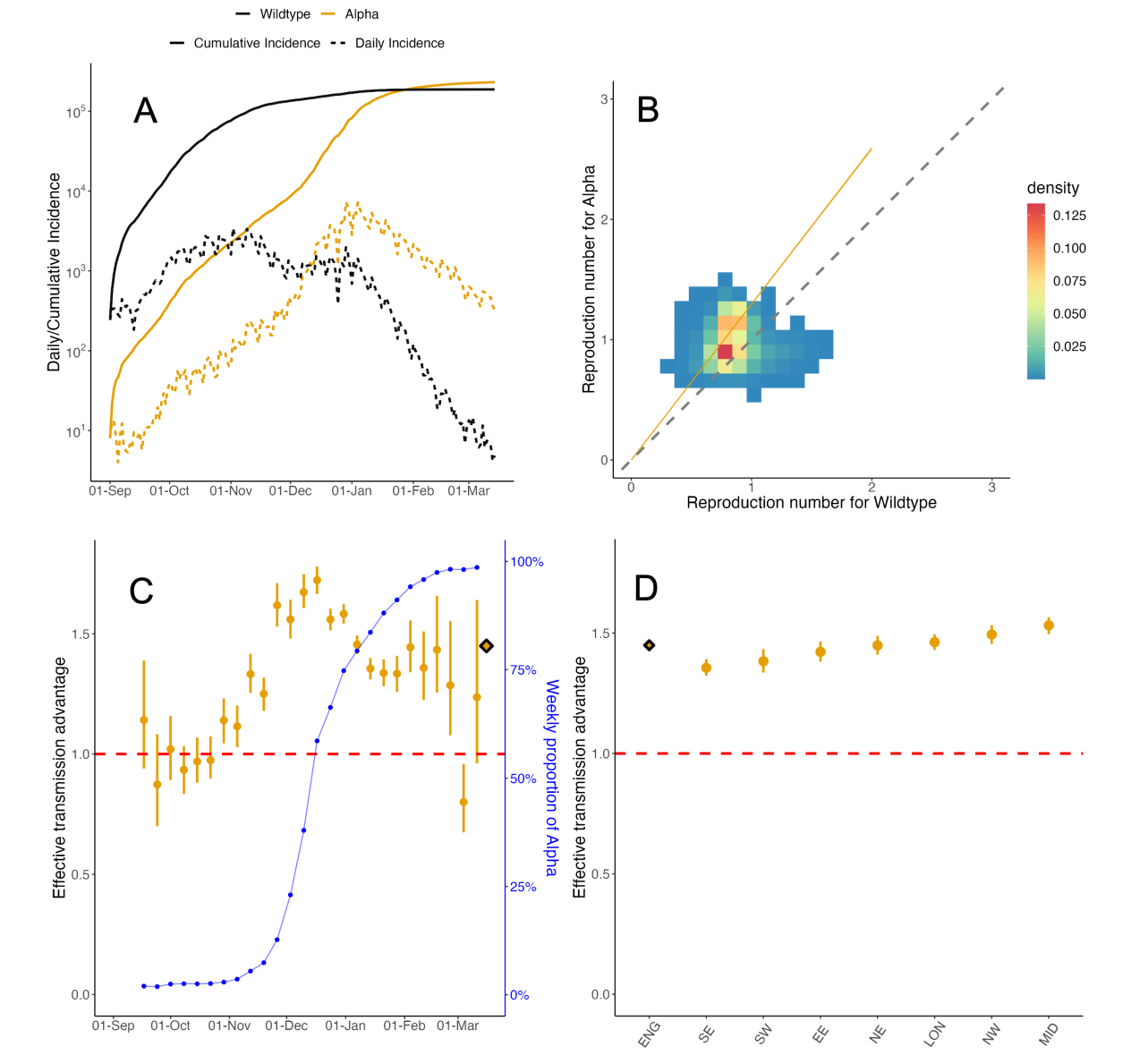


Figure 4.1: **Effective transmission advantage of Alpha over wildtype in England** (A) The daily reported incidence of cases of the wildtype (black) and Alpha (yellow) in England from September 2020 to March 2021. (B) The effective reproduction number R_t estimated independently for the wildtype (x-axis) and Alpha (y-axis) on sliding weekly windows. The colour of the cells indicates the density of the draws from the respective posterior distributions of R_t. The dashed diagonal line indicates the $x = y$ threshold. Coloured cells lying above the diagonal line suggest that Alpha is more transmissible. The yellow line denotes the median effective transmission advantage estimated using MV-EpiEstim, assuming no temporal or spatial heterogeneity. 95% Credible Interval (CrI) were so narrow that they could not be distinguished from the line. (C) Effective transmission advantage estimated using MV-EpiEstim using data in the week ending on the date specified on the x-axis (yellow) and the entire time series (diamond). The dark blue circles and the vertical bars denote respectively the mean and 95% binomial Confidence Interval (CI) of the proportion of incidence of Alpha (right y-axis) in the week of estimation. Because of the high incidence of both wildtype and Alpha in the study period, the 95% CI are small and hence difficult to distinguish. (D) Effective transmission advantage estimated using MV-EpiEstim for all NHS England regions together (diamond) and separately (solid circles), using data from 1st September 2020 to 14th March 2021. The NHS England regions are - East of England (EE), London (LON), Midlands (MID), North-East (NE), North-West (NW), South-East (SE), South-West (SW). In panels (C) and (D), the solid yellow circles denote the median estimate, the vertical lines indicate the 95% CrI, and the red dashed line denotes the $\epsilon = 1$ threshold.

For simplicity, we refer to all lineages of SARS-CoV-2 other than the VOCs that were circulating at the time as 'wildtype'. R_t estimates obtained independently for the wildtype and for Alpha indicated that Alpha was more transmissible (Figure 4.1B). However, the magnitude of the transmission advantage (naively estimated as the ratio between the two R_ts, see Appendix C.3 for details) varied over time and across regions. Pooling these non-parametric estimates over time and regions yielded a highly uncertain and non-significant transmission advantage of 1.41 (95% CrI 0.86-2.01) for Alpha compared to the wildtype across all times and regions in England.

MV-EpiEstim allows further exploring how the effective transmission advantage, which we denote as ϵ, varies over time and space, by estimating ϵ over various temporal and spatial units, within which it is assumed constant.

We first assumed that ϵ was constant across all regions but potentially varied over time. Using weekly data aggregated across regions, ϵ estimates from MV-EpiEstim showed a strong temporal variation with the central estimate initially increasing from 1.01 (95% CrI 0.89-1.15) in October to 1.72 (95% CrI 1.67-1.78) in December, and then declining again to 1.24 (95% CrI 0.95-1.62) in March 2021 (Figure 4.1C). We found a similar trend over time when ϵ was estimated independently for each region (appendix Figure C.5). We also used MV-EpiEstim to estimate ϵ separately for each NHS region assuming that it remains constant over time. This highlighted minor regional differences with ϵ estimated across the whole time period ranging from 1.36 (95% CrI 1.33-1.39) in the South-East to 1.54 (95% CrI 1.50-1.58) in the Midlands (Figure 4.1D, appendix Table C.1).

The consistent temporal variability in ϵ estimates when using data from all or individual regions suggests an underlying change in transmission dynamics. Across the entire time period and all regions we found strong evidence that Alpha was more transmissible than the wildtype, with an overall central estimate of ϵ at 1.45 (95% CrI 1.43-1.46) when assuming a single epsilon across time and space (see also appendix Table C.1). However these results mask substantial temporal heterogeneity and small levels of spatial heterogeneity (appendix Figure C.3).

We estimated a similar, albeit slightly lower, overall effective transmission advantage for Alpha using data from the 18 ADM2 regions in France (central ϵ estimate assuming no temporal or spatial heterogeneity at 1.29 (95% CrI 1.29-1.31), see appendix Figure C.6). As for Alpha in England, these estimates masked changes in the transmission advantage over time, declining from 1.46 (95% CrI 1.44-1.48) in March to 0.90 (95% CrI 0.86-0.95) in May 2021 (appendix Figure C.7), and to a lesser extent between regions (from 1.21 (95% CrI 1.20-1.23) in Île-de-France to 1.41 (95% CrI 1.37-1.46) in Bourgogne-Franche-Comté, excluding regions where Beta/Gamma were dominant, see appendix Figure C.6 and Table C.2).

Following the same approach, and using data from France, we demonstrated that the Beta and Gamma variants (combined) are also more transmissible than the wildtype, with ϵ estimated to be 1.25 (95% CrI 1.25-1.27). However behind this overall central estimate we identified a decline in ϵ

over time, and heterogeneities between regions (appendix Figure C.8 and Table C.3). Finally, using data from England, we estimated that Delta is 1.77 (95% CrI 1.69-1.85) times more transmissible than Alpha (appendix Figure C.10 and Table C.4). The spatial and temporal trends in ϵ estimates for Delta, while present, were less marked (appendix Figure C.11).

Method validation

The method performed well across most scenarios considered, with a small bias (defined as the difference between the mean posterior estimate and the true ϵ value, Figure 4.2). The coverage probability, which measures whether uncertainty in estimates is adequate (Appendix C.5.1), was also good across most scenarios (appendix Figures C.12 and C.31).

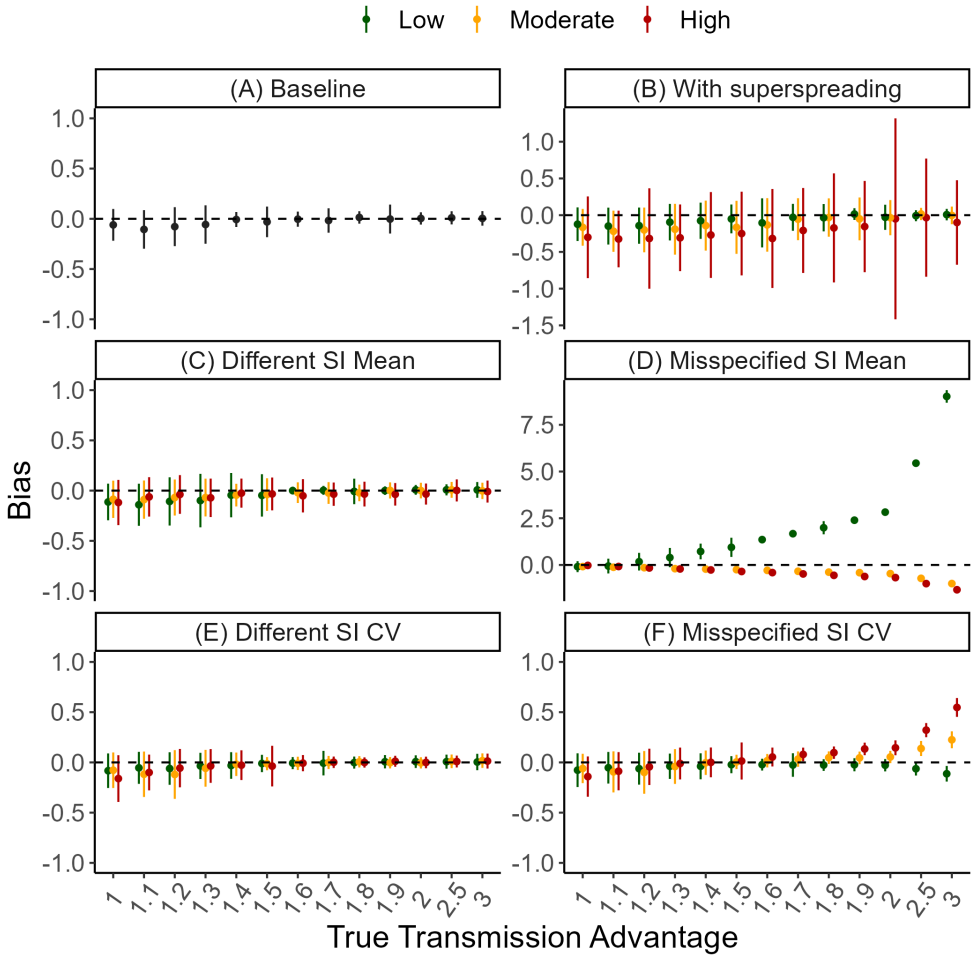


Figure 4.2: **Method performance on simulated data.** We assessed the performance of MV-EpiEstim on a range of scenarios. In each panel, the x-axis shows the true value of the effective transmission advantage (on categorical scale). The y-axis shows the bias i.e., the difference between the posterior mean estimate of the transmission advantage and the true value. The solid dots represent the mean bias (across 100 simulations) and the vertical bars show the Standard Deviation (SD) of the bias. Each panel corresponds to a different simulation scenario. In all scenarios, the R_t for the reference variant was 1.1 and the R_t for the new variant was ϵ times the reference R_t (see Appendix C.5 for details). (A) In the baseline scenario, we assumed no superspreading and the same natural history for both variants. (B) As (A), but with low (overdispersion parameter $\kappa = 1$), moderate ($\kappa = 0.5$) and high ($\kappa = 0.1$) levels of superspreading. (C) As (A), but the mean SI of the new variant is 0.5 (low), 1.5 (moderate) or 2 (high) times that of the reference and is correctly specified during estimation. (D) As (C), but the mean SI of both variants are assumed to be the same during estimation. (E) As (A), but the Coefficient of Variation (CV) (ratio of SD to mean) of the SI distribution of the new variant is 0.5 (low), 1.5 (moderate) or 2 (high) times that of the reference and is correctly specified during estimation. (F) As (E), but the CV of the SI distribution of the reference and new variants are assumed to be the same during estimation. Note that the y-axis range is different for panel D. Results using $R_t = 1.6$ for the reference and using fewer days of data are presented in Appendices C.5.3 to C.5.9. Results using time-varying reference R_t in one or two locations are shown in Appendices C.5.10 and C.5.11.

MV-EpiEstim was able to accurately estimate the transmission advantage when variants were known to differ in their natural history (characterised by the SI distribution, i.e., the delay between onset of symptoms in a case and their infector, Figure 4.2C and E, Appendices C.5.4 and C.5.6). We also explored a scenario typical of real-time outbreak analysis where the natural history of the new variant is different, but in the absence of information, is assumed to be the same as that of the reference (Appendices C.5.5 and C.5.7). Misspecifying the mean SI led to substantial bias (median bias ranged

from -1.8 to 16.7) and poor coverage, especially when the transmission advantage was moderate (more than 1.5) and the mean SI of the new variant was much shorter (0.5 times) than that of the reference (Figure 4.2D). Misspecifying the CV of the SI had little impact on the quality of the estimates (range of median bias: -0.4 to 1.0), unless the transmission advantage was very high (more than 2, Figure 4.2F).

Even in the presence of substantial superspreading (i.e., equivalent to that of SARS-CoV-1, Figure 4.2B) or poor case-reporting (i.e., up to 80% cases not reported, appendix Figure C.26), neither of which is explicitly accounted for by MV-EpiEstim, the transmission advantage remained unbiased (range of median bias with overdispersion parameter 0.1 (-0.3, 0.1); range of median bias with probability of reporting 0.2 (-0.5, 0.0)). However, coverage tended to be low in the presence of high superspreading or high under-reporting, indicating that the credible intervals were too narrow in these scenarios (Appendices C.5.8 and C.5.9).

In all scenarios, using more days of data reduced both the bias and the uncertainty (defined as the posterior standard deviation) in the estimated effective transmission advantage (appendix Figure C.14 and Appendices C.5.3 to C.5.7).

We used the uncertainty in the estimates of ϵ (i.e., the width of the 95% CrI) to determine if the effective transmission advantage was significant and classify the variant as more or less transmissible than the reference (see Section 4.2). Crucially, in many scenarios including some where the bias was substantial, MV-EpiEstim was able to correctly characterise a variant as being more transmissible than the reference. For instance, when the mean SI of the new variant was shorter but misspecified, the variant was correctly classified as more transmissible since ϵ was over-estimated (appendix Figure C.18, scenario type low). Conversely, when the mean SI of the new variant was longer but was misspecified, classification performance was generally poor and correct classification was only feasible with sufficient days of data and a large transmission advantage (appendix Figure C.18, scenario type high).

We also tested the performance of our method in a scenario in which the transmission advantage is changing over time (appendix Figure C.30). We show that, when applied to short time windows, our method is able to detect the changes in transmission advantage, but it can give misleading estimates if applied to a longer time window where the transmission advantage varies (appendix Figure C.31).

More results demonstrating the performance of the method when using fewer days of data, two locations, time-varying R_t and accounting for under-reporting are shown in Appendix C.5.

Literature review results

105 estimates for the transmission advantage of SARS-CoV-2 variants were found across 53 studies. Of the literature which provided estimates for the transmission advantage in the reproduction number R, the estimated advantage for Alpha was in the range of 1.35-1.75, with associated uncertainty estimates (95% CrIs/CIs) ranging from 1.02-2.30. Similarly, the range of the central estimates of the transmission advantage for the Delta variant in the literature was 1.5-2.4.

Out of the 53 studies, only 5 provided packaged code, with a single package requiring only incidence data for the estimation. All literature review results, including extracted estimates of the transmission advantages and hyperlinks to the available code and R packages can be found in the supplementary database (available in [240]).

4.4 Discussion

In this study we present a novel method, MV-EpiEstim, to estimate the transmission advantage of a new variant of a pathogen over a reference variant. MV-EpiEstim builds on the EpiEstim method [62, 114]. As such, MV-EpiEstim offers the same functionalities as EpiEstim. Because it is based on analytical formulations of the marginal posterior densities, the run time of a typical analysis using MV-EpiEstim is less than a few minutes on a standard laptop. MV-EpiEstim is implemented as a new function in the R package EpiEstim [263].

To illustrate the use of MV-EpiEstim, we retrospectively estimated the effective transmission advantage of Alpha, Delta and Beta/Gamma combined over the wildtype variants in England and France. Our analyses showed substantial changes in the effective transmission advantage of Alpha over time, particularly in England, where our analysis covers the full period from early emergence to dominance. The consistency of these temporal trends across regions provides stronger evidence of underlying changes in transmission trends. Volz et al. found a broadly similar temporal trend in the transmission advantage in the UK [93], which Kraemer et al. suggested may be in part explained by spatial patterns of spread [264].

Our analysis also identified temporal changes in the transmission advantage of Beta and Gamma over the wildtype, and Delta over Alpha. Such temporal changes may be due to a combination of factors. First, the detection of a VOC can trigger interventions (such as increased testing and contact tracing) targeted at sub-populations in which the VOC is circulating (e.g. travellers). This can lead to a lower rate of spread of the VOC early on, before its spread is generalised, and therefore an apparent increase in the transmission advantage over time. Changes in population immunity over time, such as increasing immunity to Alpha and waning immunity to previous variants can also contribute to the observed temporal trends in its transmission advantage. Future work could explore extending our approach to incorporate explanatory variables such as the proportion of population susceptible to specific variants over time and across different locations.

The fast run-time and reliance solely on variant-specific incidence time-series and SI distributions make it easy to explore several hypotheses about spatio-temporal trends. We recommend that future users of MV-EpiEstim run multiple sensitivity analyses exploring the spatio-temporal heterogeneities in effective transmission advantage. Consistent signals of a significant transmission advantage independently estimated across these analyses can help raise early warnings about emergence of VOCs.

Our method works well across a range of simulated scenarios, designed to mimic a variety of real-time epidemic contexts, including in the presence of superspreading and when the natural history of the new variant is imperfectly characterised. As expected, the performance deteriorated with larger errors and lower coverage probability in scenarios with high superspreading or under-reporting, and with large misspecification of natural history. Of the packaged tools identified in our literature review, one did explicitly account for overdispersion, but did not demonstrate method performance on simulated data [265]. Moreover, other than a couple of studies, which did not provide a packaged tool [266, 267], all other approaches to estimate the transmission advantage suffer from a similar identifiability issue between a change in generation time versus a change in transmission (see supplementary database in [240] for details).

In the absence of precise information on natural history, MV-EpiEstim's fast run time offers the possibility of exploring various assumptions and, in turn, estimating a range of plausible transmission advantages. Our method is robust to moderate levels of under-reporting and temporal changes in reporting if these affect both the reference and the variant equally. However, if changes in reporting vary between variants over time this can bias estimates of ϵ. For example, changes in diagnostic test protocols that progressively increase sensitivity for a new variant could cause higher apparent growth relative to the reference variant, artificially inflating ϵ.

As discussed in relation to the temporal changes in the transmission advantage of Beta and Gamma over the wildtype, and Delta over Alpha, apparent temporal trends in ϵ may be more reflective of interventions targetted at a particular variant or changes in population immunity, rather than biological changes in transmissibility alone. Other external factors, such as surge testing in population groups where a VOC is known to be circulating, may also lead to increased relative detection of that variant at that time, distorting ϵ estimates. Interpreting temporal changes requires careful consideration of these types of external influences.

Importantly, we show that our method can accurately characterise a variant as being 'more' or 'less' transmissible than a reference variant across many scenarios, including some where the performance at estimating the transmission advantage was only modest. This simple but robust characterisation could be as important as estimating the exact value of the transmission advantage, especially in informing public health response during the early phase of a new variant emerging. For example, in scenarios where the emerging variant has a shorter SI, as was the case for Omicron for example [268], our method will generally be able to detect an increase in transmission, albeit with an overestimated transmission advantage due to assuming that the SI has not changed. Where the new variant has a longer SI though, our approach will have poor ability to detect the increased transmissibility. However, again, these caveats are common to many approaches.

Classification performance depends on multiple factors including the characteristics of the reference and new variants and the amount of data available for estimation. Across all scenarios, the probability

of correctly classifying a more transmissible variant (i.e., true positive rate) increases with increasing baseline transmissibility, higher transmission advantage, and as more data are used for estimation. Conversely, in scenarios with high levels of superspreading or large misspecification of the natural history of the variant, the sensitivity of the classification is reduced (Figures C.17 and C.23) but improves when either more data are used (Figures C.18 and C.24) or with increasing transmission advantage. It is worth noting that in all scenarios, including scenarios where method sensitivity is low, the probability of misclassifying a variant as "more transmissible" when it is not (i.e., false positive rate) remains low (Table C.6) i.e., the method specificity is high. Critically, we note that in such cases, the variant is classified as "unclear", with very low probability of incorrectly classifying it as less transmissible. Further, both the true and false positive rate of classification also depend on the threshold (quantile of the posterior distribution of ϵ) used for classifying a variant as more transmissible, which can easily be modified by the user depending on the relative costs of false negatives and false positives.

We emphasise that our method estimates the *effective* transmission advantage, which will often reflect a combination of several factors such as a true increase in underlying transmissibility and the ability of a new variant to escape immunity. Disentangling these effects is particularly challenging in the context of changing population immunity e.g., due to vaccination roll-out, and may require additional data (see supplementary database in [240] for details). However, regardless of its drivers, early identification of a transmission advantage is a critical first step to a timely response.

We note that in the extreme case where estimates are made independently at all time steps and locations, our method reduces to what we have called the "non-parametric" approach, as it provides a non-parametric estimate of the transmission advantage at each time and location as the ratio between the reproduction numbers for each variant. While such an approach can help in initial exploration, the assumption of independence across time and space can lead to highly uncertain estimates. Pooled non-parametric estimates are simply an average of time and location specific estimates, giving equal weight to each estimate regardless of the underlying incidence.

MV-EpiEstim allows combining information across time and/or locations, assuming that the effective transmission advantage is constant across these. Given that the resulting estimate is a single value of ϵ fitted jointly across all data, the time points and locations with more observed cases contribute more to the estimate. This effectively provides an incidence-weighted estimate and leads to reduced uncertainty compared to the non-parametric approach. Temporal or spatial heterogeneity in the transmission advantage (e.g. reflecting heterogeneity in population immunity) can also be characterised by applying the method separately by location or time period, which is easy to implement in our software.

Our estimated transmission advantage of the SARS-CoV-2 Alpha variant (over the wildtype) is consistent with those from other analyses identified in our literature review, with estimates in the range of 1.02 to 2.30 (See supplementary database in [240]). Similarly, our estimates of the transmission

advantage of the Delta variant over the Alpha variant are broadly consistent with the literature, with central estimates of the advantage in the range of 1.5 to 2.4.

In addition, our results highlight temporal changes in the transmission advantage which were overlooked by some of these studies. The agreement of our findings with those from other studies employing a diversity of modelling approaches including renewal equations, semi-mechanistic models, and phylodynamic models suggests that MV-EpiEstim can be a useful tool for early characterisation of new variants. Importantly, where specific bio-markers are sufficient to distinguish variants (e.g. S-gene), MV-EpiEstim does not need any whole-genome sequencing data. Therefore, it could be used in near real-time, relying only on routinely collected incidence data and not necessarily suffering from potential delays in the sequencing pipeline.

4.5 Conclusion

Given the continued transmission of SARS-CoV-2 and low vaccination coverage globally [269], new variants are likely to continue emerging. Our tool can be used to monitor their transmissibility and rapidly identify variants of concern. Our estimates of the transmission advantage of Delta have been used to inform UK national policy in real-time.

Applications of our work are not limited to SARS-CoV-2; our generic method could easily be used to monitor other pathogens with multiple co-circulating strains such as influenza or *streptococcus pneumoniae*.

Chapter 5

Estimating the epidemic reproduction number from temporally aggregated incidence data: A statistical modelling approach and software tool

The contents of this chapter have been adapted from a manuscript previously published in PLOS Computational Biology [270]. After I had implemented the Expectation Maximisation (EM) algorithm in code, Professor Samir Bhatt helped to formalise the mathematical expressions for publication. I conducted all other work presented in this chapter, with input and feedback from my supervisors.

Depending on the nature of a disease, how protracted an outbreak is, and the resources available in the affected community or country, case incidence data may be reported on a daily basis, at longer intervals (e.g., weekly or monthly), or irregularly (e.g., three times per week) [271–274]. The most common challenge encountered by users of EpiEstim was the inability to provide temporally aggregated incidence data as an input for R_t estimation without using a workaround, such as aggregating the serial interval distribution to match the aggregated incidence data, which is not possible in some cases (Chapter 2). Additionally, when daily incidence data was available, administrative noise often negatively impacted the reliability of R_t estimates. In this chapter, I introduce an Expectation Maximisation (EM) algorithm that reconstructs daily incidence data from temporally aggregated data, whilst preserving the integrity of the observed data by ensuring that, if reaggregated, it matches the original aggregated counts. I implement this novel method in the EpiEstim R package and demonstrate its performance using both simulated and real-world data.

5.1 Introduction

As infectious disease outbreaks become more common, it is increasingly important to rapidly characterise the threat of emerging and re-emerging pathogens [275]. Transmissibility, i.e. a pathogen's ability to spread through a population, can be quantified using the time-varying reproduction number, R_t, defined as the average number of infections that are caused by a primary case at time t of an outbreak. R_t signals whether an outbreak is growing ($R_t > 1$) or declining ($R_t < 1$), and whether current interventions are sufficient to control the spread of the disease.

One of the most popular tools for real-time R_t estimation, the R package EpiEstim, relies on observing the incidence data and supplying an estimated Serial Interval (SI) distribution—the time between symptom onset in a case and their infector. EpiEstim requires that the SI distribution and incidence data are supplied using the same time units. This can be problematic when daily incidence data are not reported, which is common for many diseases, such as influenza, Zika virus disease, and most notifiable diseases in countries such as the United Kingdom (UK) and the United States (US) [271–274]. Additionally, several studies intentionally aggregate data to reduce the impact of daily reporting variability; administrative noise, such as "weekend effects", are characterised by a drop in reported cases over weekends, due to reduced care seeking and longer delays in reporting, followed by a peak on Mondays [276, 277]. A commonly used workaround is to aggregate the SI distribution to match the frequency of incidence reporting, [145, 278] however this is not possible if the SI is shorter than the aggregation of data. For example, influenza-like illness is typically reported on a weekly basis, but influenza has an estimated mean SI of 2–4 days [279, 280]. Similarly, reporting of COVID-19, which has an estimated SI of 3–7 days, has typically moved from daily to weekly [281, 282]. Therefore, enabling estimation of R_t from temporally aggregated data is critical to ensure methods such as EpiEstim are widely applicable [105].

In this study, we combine an EM algorithm with the renewal equation approach implemented in EpiEstim to reconstruct daily incidence from aggregated data and estimate R_t. We assess the performance of the method using influenza and COVID-19 data, in addition to an extensive simulation study.

5.2 Methods

5.2.1 EpiEstim

As described in previous chapters, EpiEstim uses the renewal equation (Equation (5.1)), a form of branching process model [62]. The incidence of new symptomatic cases at time t (I_t) is approximated by a Poisson process, where I_{t-s} is the past incidence, and g_s is the probability mass function of the SI (note that the notation for the SI is different in this chapter).

$$I_t \sim Pois\left(R_t \sum_{s=1}^{t} I_{t-s} g_s\right) \tag{5.1}$$

With EpiEstim, R_t can be assumed to remain constant within user defined time windows, which smooth out estimates.

5.2.2 Extending EpiEstim for coarsely aggregated data

We extended EpiEstim to estimate R_t from aggregated incidence data, where each aggregation window (ω) is >1 day, whilst still conditioning on an assumed serial interval distribution (g_s). We use an EM algorithm to iteratively reconstruct daily incidence from aggregated data, and in turn estimate R_t. We present the method with weekly data in mind, but the method and software can be applied to any temporal aggregation (Figure 5.1 and Appendix D.3.7).

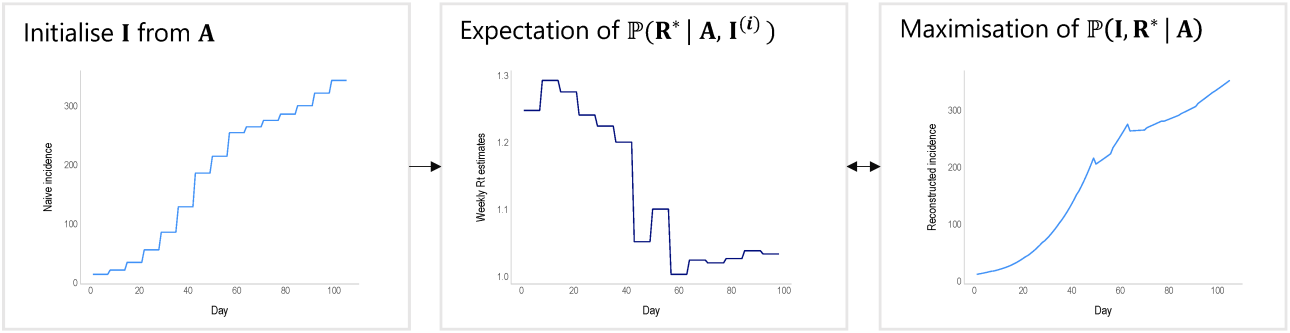


Figure 5.1: **Schematic of the EM algorithm approach used to reconstruct daily incidence (I) from temporally aggregated incidence data (in this case weekly, A).** The algorithm is initialised with a naive disaggregation of the weekly incidence (assuming constant daily incidence throughout the aggregation window, left panel). The resulting daily incidence is then used to estimate the reproduction number for each aggregation window, in this case for each week, $\mathbf{R}^*$ (expectation step, central panel). $\mathbf{R}^*$ is converted into a growth rate (see Equation (5.7)), which is in turn used to reconstruct daily incidence data, whilst ensuring that if $\mathbf{I}$ were to be reaggregated it would still sum to the original weekly totals (maximisation step, right panel). The process cycles between the expectation and maximisation steps until convergence.

We define:

- $\mathbf{I} = \{I_t\}_{t=1,\ldots,T}$ the vector of unobserved daily incidence,
- $\mathbf{A} = \{A_w\}_{w=1,\ldots,W}$ the vector of observed aggregated incidence, so that for each aggregation window w $A_w = \sum_{t=t_{w-1}+1}^{t_w} I_t$
- $\mathbf{R}^* = \{R^*_w\}_{w=1,\ldots,W}$ the vector of reproduction numbers corresponding to each incidence aggregation window. The * indicates that this is only used in the EM algorithm to reconstruct the daily incidence I, and is distinct from the final estimated $\mathbf{R}$.

We use the following indexes:

- t for days ($t = 1, \ldots, T$),
- w for aggregation windows ($w = 1, \ldots, W$),
- i for iterations of the EM algorithm ($i = 0, \ldots, 10$).

In the following, the bold notation signifies vectors. Our overall goal is to maximise the marginal likelihood function

$$\mathbb{P}(\mathbf{I} \mid \mathbf{A}) = \int \mathbb{P}(\mathbf{I}, \mathbf{R}^* \mid \mathbf{A})\, d\mathbf{R}^* = \int \mathbb{P}(\mathbf{I} \mid \mathbf{R}^*, \mathbf{A}) \mathbb{P}(\mathbf{R}^* \mid A)\, d\mathbf{R}^*. \tag{5.2}$$

This marginal likelihood expresses the probability of the unobserved daily incidence $\mathbf{I}$, conditional on the observed aggregated incidence $\mathbf{A}$, by integrating over all values of the latent reproduction number vector $\mathbf{R}^*$. As mentioned, $\mathbf{R}^*$ is used solely within the EM algorithm to reconstruct plausible daily incidence trajectories over each aggregation window and does not correspond to the final estimated reproduction number. The integrand $\mathbb{P}(\mathbf{I} \mid \mathbf{R}^*, \mathbf{A})\mathbb{P}(\mathbf{R}^* \mid A)$ represents the joint distribution of $\mathbf{I}$ and $\mathbf{R}^*$, given the observed data.

In short, the statistical goal is to produce a series of reproduction numbers that can reconstruct daily incidence while still being consistent with the observed aggregated incidence. We use an expectation maximisation scheme to approximate this marginal likelihood at low computational cost. The algorithm involves three steps: initialisation, expectation, and maximisation.

5.2.3 Initialisation

The algorithm is initialised (step $i = 0$) by disaggregating the aggregated incidence by piecewise constant functions (constant daily incidences across each aggregation window). That is, for aggregation window w covering days $t = t_{w-1} + 1, \ldots, t_w$:

$$I_t^{(i=0)} = \frac{1}{t_w - t_{w-1}} A_w. \tag{5.3}$$

Note that this allows non-integer incidence counts.

For iteration $i \geq 1$ of the algorithm, we iterate over two steps: expectation and maximisation.

5.2.4 Expectation

First, we compute the expectation

$$\mathbb{E}_{\mathbb{P}(\mathbf{R}^* \mid \mathbf{A}, \hat{\mathbf{I}}^{(i)})} \left[\mathbb{P}(\mathbf{I}, \mathbf{R}^* \mid \mathbf{A}) \right]. \tag{5.4}$$

This expectation computes the average reproduction number over each data aggregation window, given the original observed aggregated incidence and the previous (i^{th}) iteration of the estimate of daily incidence. To compute this, we use the renewal approach from EpiEstim where the posterior distribution of reproduction numbers is found analytically as $\mathbb{P}(\mathbf{R}^* | \mathbf{I}) = \prod_{w=1}^{W} \mathbb{P}(R_w^* | \mathbf{I})$, with $\mathbb{P}(R_w^* | \mathbf{I}) \sim$ Gamma(shape $= \alpha_w$, scale $= \beta_w$) and where $\alpha_w = a + \sum_{s=t_{w-1}+1}^{t_w} I_s$ and $\beta_w = 1/\left(\frac{1}{b} + \sum_{s=t_{w-1}+1}^{t_w} \sum_{u=1}^{s} I_{s-ug_u}\right)$, where a and b are the shape and scale of the Gamma prior distri-

bution for R_w [62]. The expected value for the reproduction number is therefore calculated as:

$$\hat{R}_w^{*(i)} = \frac{a + \sum_{s=t_{w-1}+1}^{t_w} \hat{I}_s^{(i-1)}}{\frac{1}{b} + \sum_{s=t_{w-1}+1}^{t_w} \sum_{u=1}^{s} \hat{I}_{s-u}^{(i-1)} g_u} \tag{5.5}$$

5.2.5 Maximisation

The maximisation step consists of recovering the most likely daily incidence from the expected $\mathbf{R}^*$ i.e. maximising $\mathbb{P}(\mathbf{I}, \mathbf{R}^* \mid \mathbf{A})$, or maximising $\mathbb{P}(\mathbf{I}, \mathbf{R}^*) \propto \prod_{t \geq 1} \mathbb{P}(I_t | I_0, \ldots, I_{t-1}, \mathbf{R}^*)$, subject to the constraint that daily incidence sums to the aggregated incidence i.e. $A_w = \sum_{t=t_{w-1}+1}^{t_w} I_t$.

In our renewal equation context, $I_t | I_0, \ldots, I_{t-1}, \mathbf{R}$ follows a Poisson distribution with mean $R_w \sum_s I_{t-s} g_s$ (where w is such that $t_{w-1} < t \leq t_w$), and therefore has mode $\hat{I}_t^{(i)} = \lfloor R_w \sum_{s=1}^{\infty} \hat{I}_{t-s}^{(i)} g_s \rfloor$ which we approximate as:

$$\hat{I}_t^{(i)} = \hat{R}_w^{*(i)} \sum_{s=1}^{t} \hat{I}_{t-s}^{(i)} g_s. \tag{5.6}$$

Wallinga and Lipsitch [283] demonstrate, conditional on the generation time distribution, analytical correspondence between reproduction number R and growth rate r through the link function:

$$R = \frac{1}{\sum_{s=1}^{\infty} \exp(-rs) g_s}. \tag{5.7}$$

We therefore assume local exponential growth so that Equation (5.6) is equivalent to:

$$\hat{I}_t^{(i)} = k_w e^{\hat{r}_w^{*(i)}(t - t_{w-1})}, \tag{5.8}$$

where $\hat{r}_w^{*(i)}$ is the exponential growth rate over aggregation window w, obtained from $\hat{R}_w^{*(i)}$ using the link function shown in Equation (5.7).

k_w is calculated to ensure the sum of daily incidence values adds up to the observed weekly totals:

$$k_w = \frac{A_w}{\sum_{t=t_{w-1}+1}^{t_w} exp\left(\hat{r}_w^{*(i)} (t - t_{w-1})\right)}. \tag{5.9}$$

We then use the estimate of $\mathbf{I}$ from Equation (5.8) in the maximisation step and iterate, thus completing the algorithm (Equation (5.5) for the expectation step and Equation (5.8) for the maximisation step). At this point, $\mathbf{I}$ can be used to estimate the full posterior distribution of $\mathbf{R}$ over any time window using EpiEstim (hereafter, this final $\mathbf{R}$ is referred to as R_t).

Given the rapid computational time and convergence (see Appendix D.2.6 and Appendix D.3.6), the default number of iterations was set to 10 in the R package. However, a convergence check ensures that the final iteration of the reconstructed daily incidence does not differ from the previous iteration beyond a tolerance of 10^{-6}, and the number of iterations can be modified by the user.

5.2.6 Case studies

We chose datasets where incidence data was available daily, and then artificially aggregated them to weekly counts. R_t was estimated from daily incidence that was reconstructed from weekly aggregated data using our new approach, and compared to R_t estimates obtained from the reported daily incidence using the original EpiEstim R package. All R_t estimates were made using both daily and weekly sliding time windows. We refer to those estimates as daily R_t estimates and weekly R_t estimates, respectively.

We considered three characteristics: 1) mean R_t estimates, 2) uncertainty in the R_t estimates, and 3) the classification of R_t as increasing, uncertain or declining (Appendix D.2.5). To compare the performance of this approach to the original method, we assessed the correlations between each of the three characteristics when using the reported and reconstructed incidence. For the mean R_t estimates and uncertainty in R_t estimates, we assessed the linear relationships using the Pearson correlation coefficient (where values closer to +1 are indicative of a strong positive correlation).

The gamma distributed priors for $\mathbf{R}^*$ and R_t were set to a mean and standard deviation of 5 (shape = 1, scale = 5), which is the default prior parameterisation used in EpiEstim. The rationale behind this choice is that it ensures that one will not conclude R < 1 unless the data strongly supports that. The user can set the prior themselves.

5.2.7 Influenza

We obtained a five-week subset of a dataset (11^{th} December 2009—14^{th} January 2010) on US active component military personnel (employed by the military as their full-time occupation) that made an outpatient visit to a permanent military treatment facility describing a respiratory-related illness. This daily incidence by date of presentation at a clinic was originally obtained by Riley et al. from the Armed Forces Health Surveillance Center and were digitally extracted for use here [284]. We used a mean SI of 3.6 days and SD of 1.6 days [279].

5.2.8 COVID-19

Incidence of UK COVID-19 cases and deaths were taken from the UK government website [285]. For COVID-19 cases, we obtained ninety-seven weeks of data (21^{st} February 2020 to 30^{th} December 2021) for incidence by date of specimen, which is the date that a sample was taken from an individual which later tested positive. For COVID-19 deaths, we used ninety-six weeks of data (2^{nd} March 2020 to 2^{nd} January 2022) for incidence by date of death within twenty-eight days of a positive test. We assumed a mean SI of 6.3 days and SD of 4.2 days [281].

In Appendix D.5, we also apply the EM algorithm to weekly incidence data for Zika virus disease to assess the performance of the method on a non-respiratory pathogen.

5.2.9 Simulation study

We considered scenarios where R_t either remained constant or varied over time, with a stepwise or gradual change. For each scenario, one hundred seventy-day epidemic trajectories were simulated using a renewal equation-based branching process model, as implemented in the R package projections [113]. Each epidemic was seeded with seven days of ten daily cases and then simulated over seventy days (ten weeks) starting on day eight with daily incidence modelled as a Poisson process (Algorithm 3).

Algorithm 3: Pseudocode for simulating the epidemic trajectories for R_t scenarios.

Input:

`ndays` ← integer, number of days to run the simulation for - 70
`incid_init` ← vector of daily incidence used to seed the epidemic - 10 cases per day for 7 days
`mean_si`, `sd_si` ← parameters for gamma-distributed serial interval
`nsims` ← integer, number of simulations to run (e.g. 100)
`rt` ← either:
  `rt_const` ← vector of constant R_t values (e.g. 1, 1.25, 1.5, 1.75)
  `rt_steps` ← list of two-element vectors for sudden change in R_t (e.g. $1 \rightarrow 1.25$)
  `rt_grad` ← list of vectors representing gradual change in R_t over `ndays`

1) Initialisation:

Generate discrete gamma serial interval distribution `si` using `mean_si` and `sd_si`
Set `init_length` ← length of `incid_init`
Initialise empty list `sim_results`

2) Simulate incidence:

Constant R_t scenarios
for *each constant R in* `rt_const` **do**
  for *sim = 1 to* `nsims` **do**
    Initialise `I`, a numeric vector of length `init_length + ndays`
    Set first `init_length` entries using `incid_init`
      for *each day* $t = 1$ *to* `n_days` **do**
        `overall_infectiousness`[t] ← $\sum_{s=1}^{t}$ `I`$[t - s +$ `init_length`$] \times$ `si`$[s]$
        `expected_cases`[t] ← $R \times$ `overall_infectiousness`$[t]$
        `I`[t + `init_length`] ∼ Poisson(`expected_cases`[t]) (Equation (5.1))
      end
    Append `I[(init_length + 1) : (init_length + ndays)]` to `sim_results`
  end
end

Stepwise change in R_t scenarios
for *each pair* $R = (R_t^1, R_t^2)$ *in* `rt_steps` **do**
  Same as above except with a step change at day 35, using:
  R_t^1 for days $1 \rightarrow 35$
  R_t^2 for days $36 \rightarrow 70$
end

Gradual change in R_t scenarios
for *each vector* R_t *in* `rt_grad` **do**
  Same as above except with a gradual change in R_t over `n_days`
  Apply the same simulation model with $R[t]$ updated each day
end

3) Output:

`sim_results`: for each scenario, a list of length `nsims` where each element contains simulated incidence over `ndays`

Daily datasets were aggregated weekly and used to estimate R_t using the proposed method; these values were compared to R_t estimates obtained from simulated daily data using the original EpiEstim R package. We explored the impact of weekend effects on R_t estimates, the ability to supply alternative temporal aggregations of data e.g., three-day, ten-day, or two-weekly aggregations, the ability to detect mid-aggregation variations in transmissibility, and finally, the number of iterations required to reach convergence when reconstructing daily incidence data. The full simulation study description and details can be found in Appendix D.3.

5.3 Results

Hereafter, we refer to reported and reconstructed incidence data, these are the reported daily incidence and the daily incidence that has been reconstructed from weekly aggregated data, respectively.

5.3.1 Influenza

The reconstructed incidence of influenza was much smoother than the reported incidence, which showed clear weekend effects and lower reported cases on two public holidays, both occurring on Fridays (Figure 5.2A and Appendix D.2.4). Considering weekly sliding R_t first, there was a high correlation in both the mean R_t estimates derived from each dataset ($R^2 = 0.91$, Figure 5.2C and Appendix D.2.1) and their associated uncertainty ($R^2 = 0.93$, Figure 5.2D). The overall agreement in the classification of R_t reached 81.8% (see methods Section 5.2.6 and Appendix D.2.5).

In contrast, mean daily R_t estimates differed markedly depending on whether the reported or reconstructed data were used, with an R^2 of 0.13 and much higher mean R_t and uncertainty in estimates obtained from reported data (Figure 5.2E and Figure 5.2F). Higher mean R_t estimates coincided with large peaks in the reported daily incidence (typically on Mondays), as daily R_t estimates were not smoothed and therefore more affected by intra-weekly variability (Appendix D.2.1). The overall agreement in the classification of daily R_t estimates was much lower, with only 44.4% agreement (Appendix D.2.5).

In this case study, the greatest differences in R_t estimates tended to correspond to time periods when the reported and reconstructed incidence data were most dissimilar (Figure 5.2B and Appendix D.2.1). There was no apparent pattern in the estimates with regard to the outbreak phase, i.e. early, mid or late-phase, but this is likely due to this dataset being a snapshot of incidence taken from within an established epidemic (Figure 5.2).

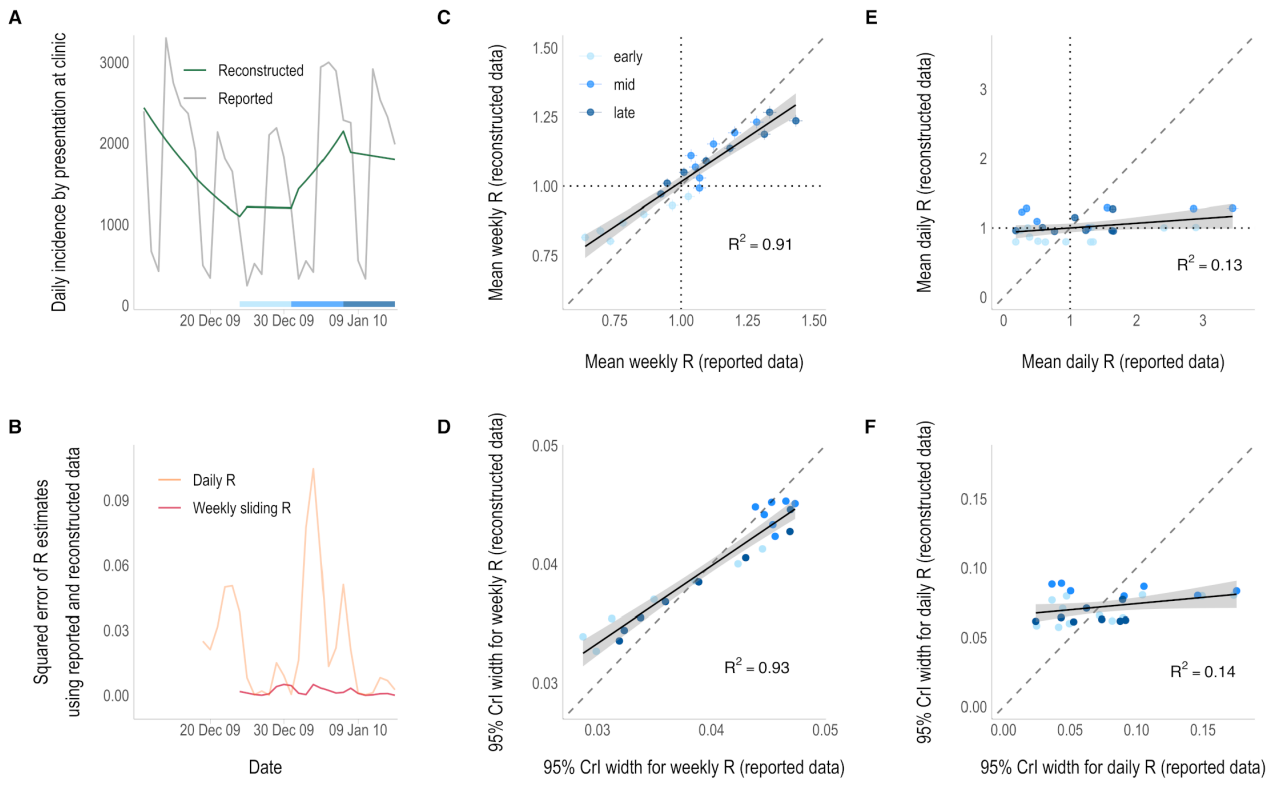


Figure 5.2: R_t **estimates from daily incidence that was either reported or reconstructed from weekly aggregated influenza data.** A) The reported (grey) and reconstructed (green) daily incidence of influenza by date of presentation at a military clinic. B) Squared error of the daily (orange) and weekly sliding (pink) R_t estimates that were made from reconstructed daily data compared to those obtained from the reported daily data. R_t estimation starts on the first day of the second aggregation window (day 8 — 18^{th} December 2009) and is plotted on the last day of the time window used for estimation (i.e., starting on day 9 (19^{th} December) for daily estimates and day 14 (24^{th} December) for weekly estimates). *Note: the x-axis is shared with the incidence plot above.* C & E) Correlation between the weekly sliding (C) and daily (E) mean R_t estimates using reconstructed data (y-axis) and reported daily data (x-axis). Vertical and horizontal lines depict the 95% credible intervals (95% CrIs) and dotted lines show the threshold of $R_t = 1$. D & F) Correlation between the uncertainty in the weekly sliding (D) and daily (F) R_t estimates, defined as the width of the 95% credible intervals, using the reconstructed (y-axis) and reported (x-axis) daily data. The colour of the points in panels C-F correspond to the epidemic phase, i.e. the early (19^{th}–30^{th} December for daily estimates, or 24^{th}–30^{th} December for weekly sliding estimates), middle (31^{st} December – 6^{th} January) or late (7^{th} –14^{th} January) phase of the data, shown by the strip in panel A. Solid lines show the linear model fit with 95% CIs (grey shading). Dashed lines represent the x = y line.

5.3.2 COVID-19 cases

The reconstructed incidence of COVID-19 smoothed out intra-weekly variability, caused by factors such as weekend effects (Figure 5.3A and Appendix D.2.4). Weekly sliding R_t estimates obtained from reconstructed and reported incidence were similar, both in their means ($R^2 = 0.98$) and their level of uncertainty ($R^2 = 0.99$, Figure 5.3C and Figure 5.3D and Appendix D.2.2). Mean daily R_t estimates were less well correlated ($R^2 = 0.67$), although the difference is less marked than in the influenza case study (Figure 5.3E), and the uncertainty in the estimates was similar across both approaches ($R^2 = 0.97$, Figure 5.3F). Most of the discrepant R_t estimates and higher levels of uncertainty coincide with the early phase of the outbreak when incidence was lower (Figure 5.3E and Figure 5.3F). Out-

side of periods of low incidence, the largest differences in R_t estimates tended to correspond to time periods with greater disparities between the reported and reconstructed incidence data (Figure 5.3B and Appendix D.2.2). The overall agreement in the classification of R_t estimates was higher than for influenza, with 74.4% and 94.9% agreement for daily and weekly sliding R_t estimates respectively (Appendix D.2.5).

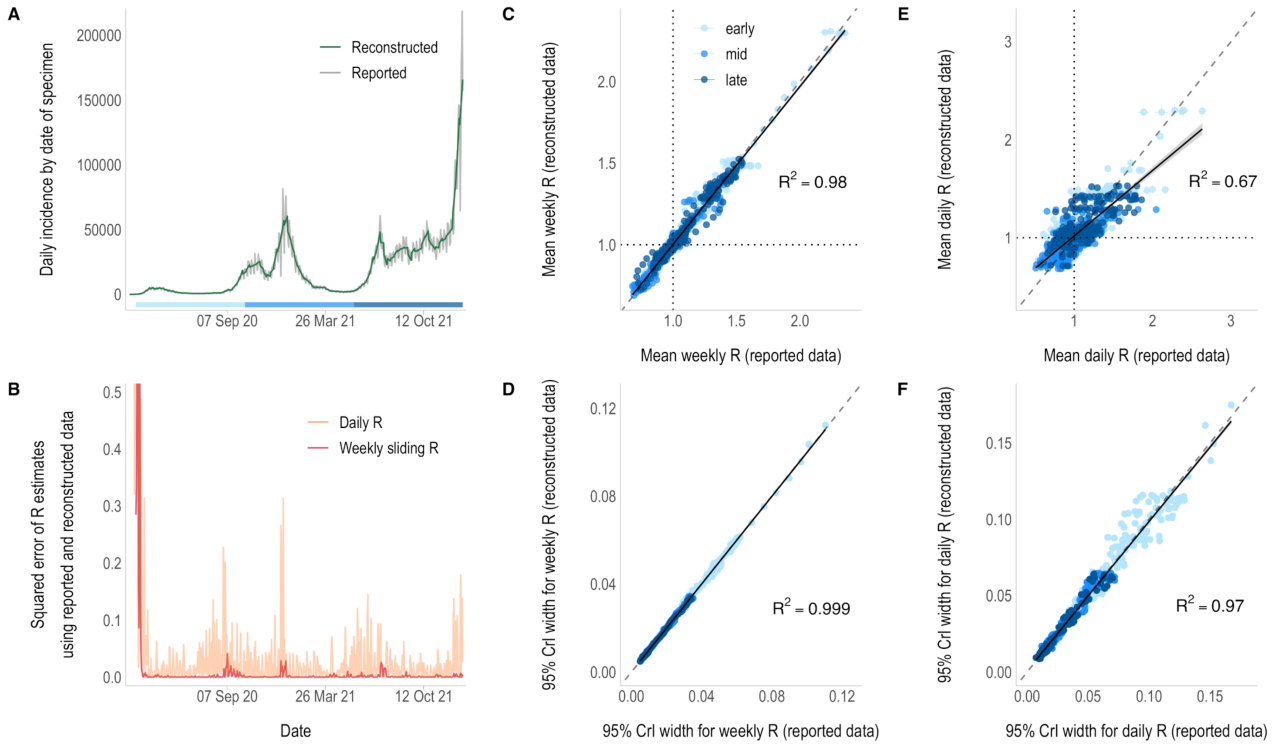


Figure 5.3: R_t **estimates from daily incidence that was either reported or reconstructed from weekly aggregated COVID-19 case data.** A) The reported (grey) and reconstructed (green) daily incidence of COVID-19 by date of specimen. B) Squared error of the daily (orange) and weekly sliding (pink) R_t estimates made from reconstructed data compared to those obtained from the reported daily data. R_t estimation starts on the first day of the second aggregation window (day 8 — 28^{th} February 2020) and is plotted on the last day of the time window used for estimation (i.e., starting on day 9 (29^{th} February) for daily estimates and day 14 (5^{th} March) for weekly estimates). *Note: the x-axis is shared with the incidence plot above and the y-axis has been limited to 0.5 for clarity.* C & E) Correlation between the weekly sliding (C) and daily (E) mean R_t estimates using reconstructed (y-axis) and reported (x-axis) daily data, starting on day 30 due to low incidence. Vertical and horizontal lines depict the 95% credible intervals (95% CrIs) and dotted lines show the threshold of R_t = 1. D & F) Correlation between the uncertainty in the weekly sliding (D) and daily (F) R_t estimates, defined as the width of the 95% credible intervals, using the reconstructed (y-axis) and reported (x-axis) daily data. The colour of the points in panels C-F correspond to the epidemic phase, i.e. the early (21^{st} March – 12^{th} October 2020), middle (13^{th} October 2020 — 22^{nd} May 2021) or late (23^{rd} May – 30^{th} December 2021) phase of the data, shown by the strip in panel A. Solid lines show the linear model fit with 95% CIs (grey shading). Dashed lines represent the x = y line.

5.3.3 COVID-19 deaths

The reported incidence of COVID-19 deaths was much less influenced by day-to-day variation. The reconstructed daily incidence was more similar to the observed daily data than in the previous case studies (Figure 5.4A). Both weekly and daily R_t estimates obtained from weekly data were highly

consistent with those obtained from daily observations ($R^2 = 0.98$ and $R^2 = 0.80$ respectively, Figure 5.4C and Figure 5.4E). The overall agreement in R_t classifications for daily estimates was the highest of all case studies at 85.8%, and 93.3% for weekly R_t estimates (Appendix D.2.5). Discrepancies between the two mostly coincide with periods of particularly low incidence of deaths (Figure 5.4B and Appendix D.2.3). The overall lower incidence of COVID-19 deaths compared to COVID-19 cases means there is greater uncertainty in R_t estimates in this case study (Figure 5.4D and Figure 5.4F and Appendix D.2.3). However, there was minimal difference in the uncertainty of estimates obtained from daily and weekly data (Figure 5.4D and Figure 5.4F).

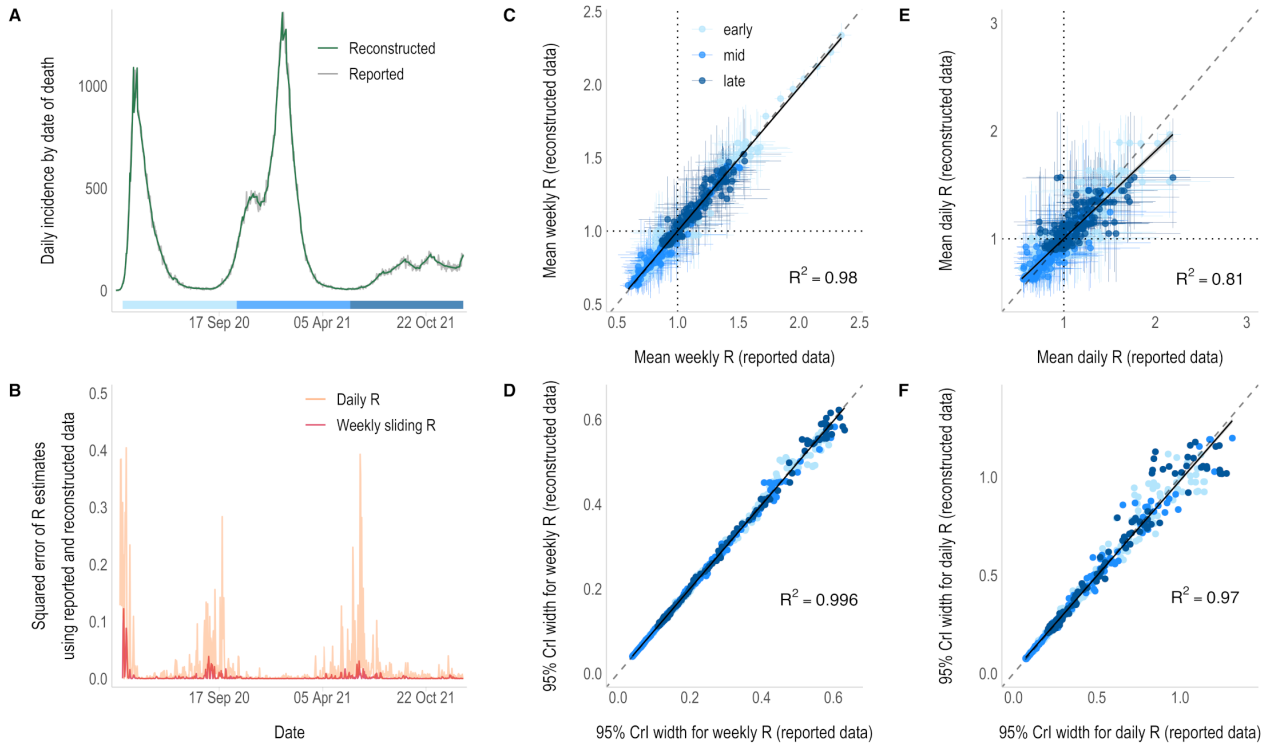


Figure 5.4: R_t **estimates from daily incidence that was either reported or reconstructed from weekly aggregated COVID-19 death data.** A) The reported (grey) and reconstructed (green) daily incidence of COVID-19 by date of death within 28 days of a positive test. B) Squared error of the daily (orange) and weekly sliding (pink) R_t estimates that were made from reconstructed data compared to those obtained from the reported daily data. R_t estimation starts on the first day of the second aggregation window (day 8 — 9^{th} March 2020) and is plotted on the last day of the time window used for estimation (i.e., starting on day 9 (10^{th} March) for daily estimates and day 14 (15^{th} March) for weekly estimates). *Note: the x-axis is shared with the incidence plot above and the y-axis has been limited to 0.5 for clarity.* C & E) Correlation between the weekly sliding (C) and daily (E) mean R_t estimates using reconstructed (y-axis) and reported daily data (x-axis), starting on day 30 due to low incidence. Vertical and horizontal lines depict the 95% credible intervals (95% CrIs) and dotted lines show the threshold of $R_t = 1$. D & F) Correlation between the uncertainty in the weekly sliding (D) and daily (F) R_t estimates, defined as the width of the 95% credible intervals, using the reconstructed (y-axis) and reported daily (x-axis) data. The colour of the points in panels C-F correspond to the epidemic phase, i.e. the early (31^{st} March – 20^{th} October 2020), middle (21^{st} October 2020 — 28^{th} May 2021) or late (29^{th} May 2021— 2^{nd} January 2022) phase of the data, shown by the strip in panel A. Solid lines show the linear model fit with 95% CIs (grey shading). Dashed lines represent the x = y line.

In all case-studies, incidence reconstructions converged within 10 iterations of the EM algorithm. The overall process of Rt estimation from weekly aggregated data took three seconds or less to run on

MacOS (2 GHz Quad-Core Intel Core i5) 16GB RAM (Appendix D.2.6); the influenza scenario, with over 57,000 cases, took two seconds to run, whilst the COVID-19 cases and deaths scenarios, with an overall incidence over 149,000 and 13 million cases respectively, took three seconds to run.

5.3.4 Simulation study

We considered scenarios where R_t either remained constant (Figure D.9) or varied over time, with a stepwise (Figures D.11 and D.12) or gradual (Figures D.13 and D.14) change. For each scenario, one hundred seventy-day epidemic trajectories were simulated using a Poisson branching process as implemented in the R package projections [113]. Daily datasets were aggregated weekly and used to estimate R_t using the proposed method; these values were compared to R_t estimates obtained from simulated daily data using the original EpiEstim R package. The full simulation study description can be found in Appendix D.3.

Weekend effects

To explore the impact of weekend effects on R_t estimates, the simulated incidence for scenarios where R_t remained constant at 1.5, suddenly decreased from 1.5 to 1.25, and gradually increased from 1.25 to 1.5, were modified so that 80% of cases were taken off the final two days of each aggregation window and redistributed uniformly over the first two days of the aggregation window (Figure 5.5A-C). This appears roughly similar to the pattern observed in the influenza case study (Figure D.7). R_t was then estimated using daily (Figure 5.5D-F), weekly (Figure 5.5G-I), and two-weekly (Figure 5.5J-L) sliding windows, using the simulated data (with weekend effects) and the reconstructed daily data.

For each scenario, the R_t estimates were considerably smoother and more accurate when the incidence had been reconstructed from weekly data. This is simply because the reconstruction smoothed out the intra-weekly variability and removed the effect of the noise from R_t estimates.

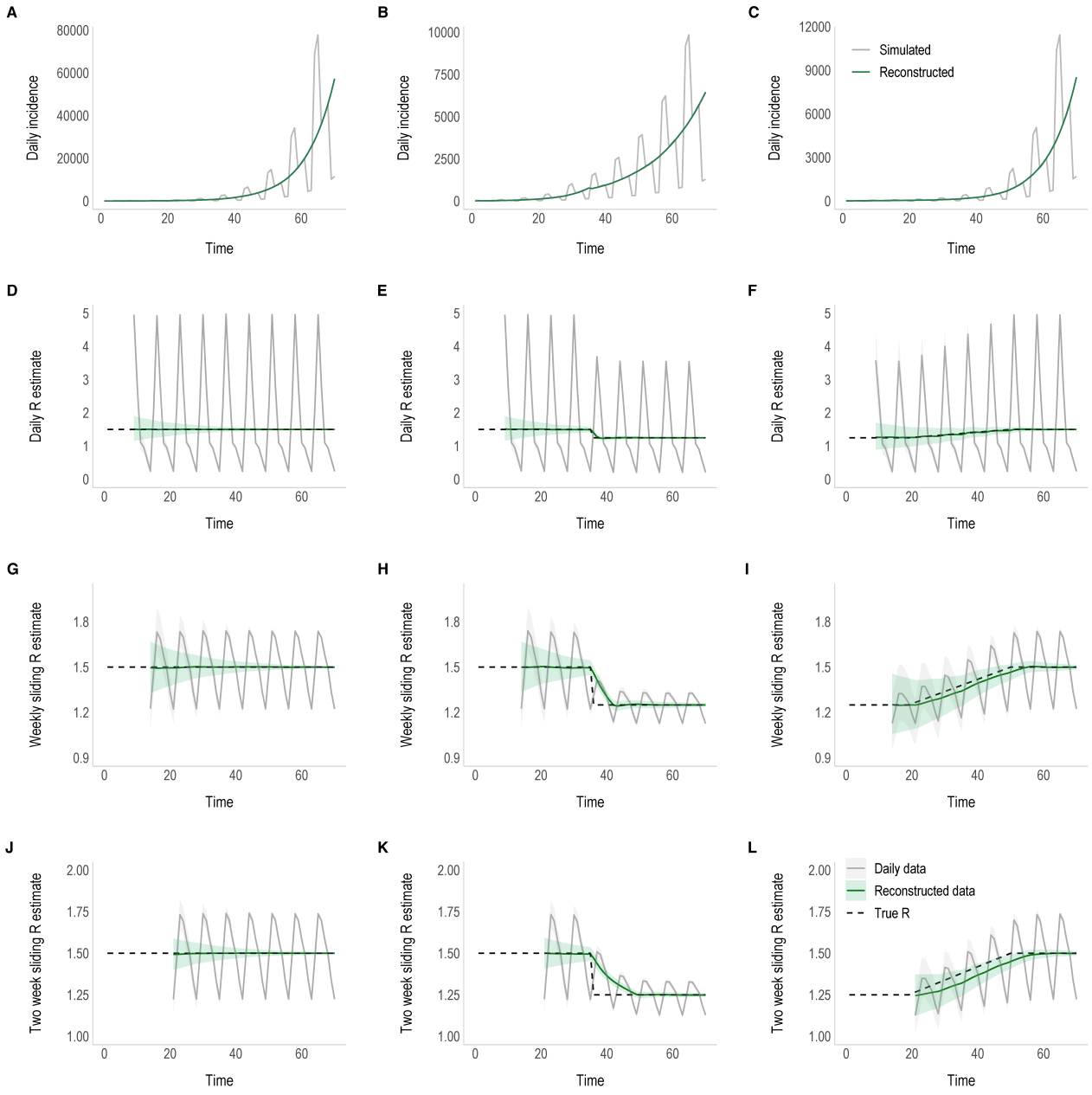


Figure 5.5: **Assessing the performance of our method in the presence of weekend effects in the reported data.** A-C) Example of reported (grey) and reconstructed (green) daily incidence data taken from one of the 100 simulated epidemics (selected at random) for a (A) constant R_t of 1.5, (B) stepwise decrease in R_t from 1.5 to 1.25, and (C) gradual increase in R_t from 1.25 to 1.5. R_t was estimated over D-F) daily, G-I) weekly, and J-L) two-weekly sliding time windows. The black dashed line is the true value of R_t that the simulated data was based on. *Note: y-axis limits for each incidence plot varies and the y-axis scale is different for the daily R_t estimate plots (second row).*

Mid-aggregation variations in transmissibility

While our method is effective at successfully smoothing out noise in the data due to phenomena such as weekend effects in reporting, this smoothing may also disguise genuine variations in transmissibility that occur within aggregation windows. To assess this, we simulated incidence data where transmissibility changed during weekends, either increasing from 1.25 to 1.5 or decreasing from 1.5 to 1.25 (Figure 5.6). Such patterns could occur due to behavioural changes, e.g. transmission may increase during weekends

due to larger group gatherings and events, or decrease if transmission is limited to households rather than workplaces.

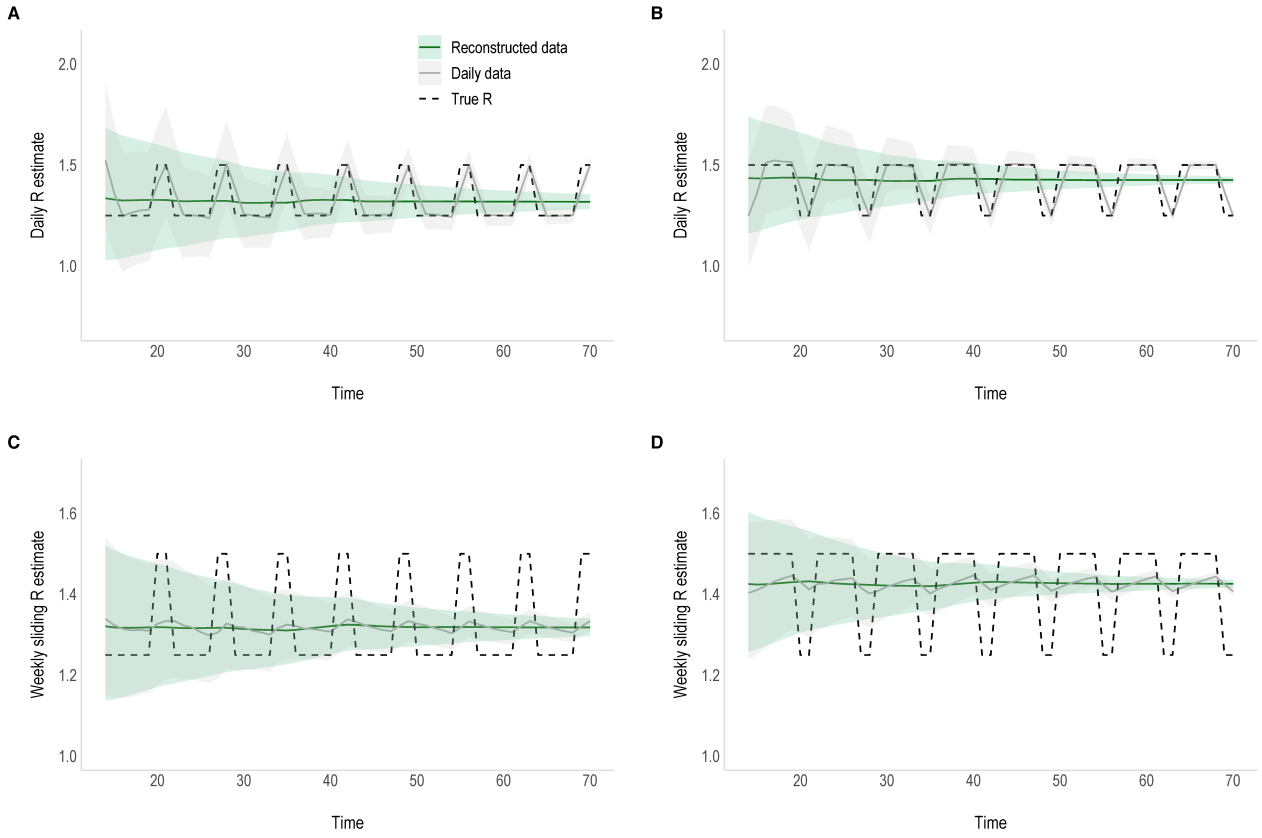


Figure 5.6: **Assessing the ability to detect genuine mid-aggregation window variations in transmissibility** (either increasing (left), or decreasing (right) on weekends) from reconstructed daily data (green) and the true simulated daily data (grey), with R_t estimated over daily (A-B) or weekly-sliding (C-D) time windows.

When R_t is estimated over daily time windows (Figure 5.6A-B) variations in transmissibility are well captured by R_t estimates obtained from the simulated daily data, but completely smoothed out when using the reconstructed daily data. When R_t is estimated over weekly-sliding time windows (Figure 5.6C-D), R_t estimates from the simulated daily data are considerably smoothed out, but small temporal variations are still apparent, whilst they disappear in R_t estimates using the reconstructed daily data.

In a real-world context, it can be difficult to disentangle administrative noise in reporting, such as weekend effects, from these genuine patterns in transmission. Therefore, there is a trade off in temporal smoothing between removing noise and detecting important variations in transmissibility that have epidemiological significance.

We also explored the ability to supply alternative temporal aggregations of data e.g., three-day, ten-day, or two-weekly aggregations (Appendix D.3.7) and the number of iterations required to reach convergence when reconstructing daily incidence data (Figure D.17).

5.4 Discussion

Estimates of the time-varying reproduction number (R_t) have frequently been used to inform and guide policymaking during outbreaks, and a commonly used approach to estimate R_t is EpiEstim, which relies on daily incidence data. However, maintaining daily incidence databases requires substantial time and investment in resources, which is not always feasible, particularly for less acute or routinely reported diseases. Therefore, in practice, many diseases are not reported on a daily basis, including influenza and other notifiable diseases in the UK and US [271–274]. As the COVID-19 pandemic persists, daily reporting is also becoming less common [182]. Coarsely aggregated data can be challenging to deal with in the context of R_t estimation methods, restricting their applications in certain contexts. In this study, we develop a statistical framework and tool that allows R_t estimation from aggregated incidence without introducing bias. Using influenza and COVID-19 data, alongside a simulation study, we demonstrate how a simple expectation-maximisation algorithm approach can rapidly reconstruct daily incidence data and accurately estimate R_t.

In all case studies, direct comparisons between weekly sliding R_t estimates show that very similar estimates can be made from the reported daily incidence and the reconstructed daily incidence from weekly aggregated data. However, daily R_t estimates are more influenced by noise, such as intra-weekly variability, leading to greater disparities in estimates between datasets. There are clear weekend effects exhibited in the influenza and COVID-19 case data (Appendix D.2.4), leading to peaks and troughs in the reported incidence and the resulting daily R_t estimates (Figure 5.2, Figure 5.3, Figure D.1 and Figure D.3). Using reconstructed incidence considerably smoothed the daily R_t estimates, removing the impact of weekend-effects. The overall agreement in the classification of R_t as increasing, uncertain, or declining between estimates made from each dataset rose substantially when some of the variability in the reported data was smoothed by estimating R_t using weekly sliding windows (Appendix D.2.5).

Despite both being affected by weekly periodicity in reporting, concordance of R_t estimates obtained from COVID-19 case data is considerably better than for influenza, perhaps due to the greater quantity of data, with a very strong positive correlation between daily and weekly R_t estimates (Figure 5.3). This is reflected in the high overall agreement in the classification of R_t estimates obtained from the reported and reconstructed datasets. It is important to note that outlying and much larger R_t estimates obtained from both datasets coincide with the early phase of the epidemic, when incidence was lower and the prior for R_t ($\mu = 5$, $\sigma = 5$) had more weight on estimates.

During the early stages of epidemics, despite there being far fewer deaths than cases, death data can sometimes be considered more reliable [146, 158]. For example, case reporting is affected by surveillance system quality and the robustness of testing practices, which can vary considerably over the course of an epidemic, especially early on. COVID-19 incidence by date of death is much less influenced by administrative noise in the data (Appendix D.2.4), and the reconstructed incidence is most similar to

the reported daily incidence of any case study. Therefore, the greatest differences in R_t estimates from death data coincide with periods of low incidence (appendix Figure D.6) when uncertainty increases. Weekly sliding R_t estimates are equally as correlated as those from COVID-19 case data, but daily R_t estimates are the most strongly correlated of any dataset (Figure 5.4). Additionally, there is very high overall agreement in the classification of daily and weekly R_t (Appendix D.2.5). This provides further support that differences between daily R_t estimates for influenza and COVID-19 cases is likely due to the reconstructed incidence smoothing out weekly periodicity in reporting.

To investigate further, weekend effects were artificially introduced to data in the simulation study (Figure 5.5). We have shown that, when using reported incidence, R_t estimates are all strongly influenced by weekend effects (regardless of the smoothing time-window). Reconstructing daily incidence from weekly data completely removes the effect of noise from resulting R_t values, greatly improving the accuracy of estimates. This demonstrates that it may be beneficial to artificially aggregate daily data, as has been done in previous studies [276, 277]. However, we did assume quite an extreme level of administrative noise, so in instances where the pattern is less prominent, it may have less of an impact on estimates. Furthermore, this smoothing effect could disguise genuine variations in transmissibility that occur mid-aggregation window, for instance, increased/decreased transmission over weekends (Figure 5.6). Disentangling important temporal trends in R_t from noise in the data can be difficult, and if aggregated data are used it will be at the cost of reduced temporal resolution in R_t estimates.

This can be seen when the method is applied to data aggregated over longer timescales, such as ten- to fourteen-days (Appendix D.3.7). This approach requires two layers of smoothing: 1) the incidence is smoothed over each aggregation window during the reconstruction process and 2) R_t estimates are smoothed by the sliding window chosen by the user. If a change in R_t occurs at the end of an aggregation window (i.e. on the last day), such as a sudden decrease in R_t due to a strict lockdown, that change is detected with a lag, corresponding to the length of the sliding window used for R_t estimation (appendix Figure D.18). However, if the event occurs mid-aggregation window, then in addition to the usual lag caused by the sliding window, estimates will be affected by the smoothing of the incidence within the aggregation window during reconstruction (appendix Figure D.19). The change in R_t will seem more gradual over the period that data are aggregated over and will appear to start earlier (corresponding to the first day of the aggregation window). It is important for users to be aware of this, particularly when using longer aggregations of data.

Another consideration is that the reconstructed incidence can have discontinuities in the borders between aggregation windows (Appendix D.3.2). This occurs because in reconstructing daily incidence we impose that, if it were to be re-aggregated, it would match the original data. Methods that simply fit smoothing splines to weekly data, inferring daily case counts from the daily difference in cumulative counts, are not affected by this [165, 286]. To circumvent this problem, we recommend that sliding windows used to estimate R_t are at least equal to or longer than the length of aggregation windows to

reduce the impact of discontinuities on estimates (Appendix D.3.7).

Alternative approaches include simple smoothing splines or Locally Estimated Scatter plot Smoothing (LOESS) to reconstruct daily incidence from aggregated data (see Appendix D.4), and modelling frameworks implemented in the Epidemia and EpiNow2 R packages [131, 146, 276]. Daily infections are modelled as a latent process, back-calculated from observed data on cases or deaths, depending on an appropriate infection to observation distribution. In addition, Epidemia integrates further information, such as the infection ascertainment rate (for cases) or the infection fatality rate (for deaths) [146]. This facilitates a 'nowcasting' approach, allowing users to estimate R_t directly from the unobserved infections, but they typically require more data (e.g. incidence of deaths and cases), more assumptions (e.g. delay distributions and ascertainment rates), and are much more computationally intensive, which can be a barrier to the adoption of such methods by users [105].

Here, R_t estimates are based on a single daily incidence reconstruction, meaning R_t can be estimated very rapidly from aggregated data, which is particularly desirable during real-time outbreak analysis [105]. A potential downside is that uncertainty in R_t estimates could be underestimated. However, the simulation study showed that the 95% credible interval of estimates encompassed the correct value of R_t the majority of the time, and we found no substantial indication that this approach detrimentally affected our characterisation of the uncertainty.

Given that this method is directly derived from EpiEstim, it relies on similar assumptions and caveats [62, 102]. As time of infection is more difficult to observe than symptom onset, the SI is typically used as an approximation of the generation time in the renewal equation, which may introduce bias [111]. The SI, the level of undetected cases, and the reporting rate are assumed to remain constant, which is often not the case in practice. Factors such as changes in population immunity, and the introduction of interventions, can alter the SI throughout an epidemic [287]. Whilst changing case definitions, new testing practices, and increased healthcare-seeking behaviour, can all affect case ascertainment [62]. Parameters chosen by users can also influence estimation accuracy, for instance, the time window length for temporal smoothing and the prior for R_t [102]. Finally, EpiEstim's assumption of a Poisson likelihood may be a limitation in instances when data are substantially overdispersed [288, 289].

To make the method simple to implement for current and future users of EpiEstim, this extension has been fully integrated with the 'estimate_R()' function in the original R package on GitHub [290]. Just one additional parameter is required - the number of days data are aggregated over (with some other optional parameters). The reconstructed daily incidence is also generated as an output, so it is possible to use it in other analysis pipelines involving alternative R estimation methods, which may perform better than EpiEstim in certain contexts, e.g. in retrospective analysis (Appendix D.6), [289] or in the presence of delays in reporting [131]. More details regarding the applications of this method can be found in the package vignette and associated examples [290].

5.5 Conclusion

We extended the widely used R_t estimation approach proposed by Cori et al., [62] and implemented in the R package EpiEstim, to incorporate a new feature which allows R_t to be easily estimated from any temporal aggregation of incidence data. We have demonstrated that the method performs well using both simulated and real-world data, recovering or even improving upon the estimates that would have been made from reported daily data. This extension is easy to use and computationally efficient, which will enable epidemiologists and other public health professionals to apply EpiEstim to a wider range of diseases and epidemic contexts.

Chapter 6

Forecasting COVID-19 cases in US states using reconstructed incidence data

Chapter 5 demonstrates that R_t can be accurately estimated using incidence data reconstructed via the Expectation Maximisation (EM) algorithm. This chapter focuses on assessing whether comparable forecasting performance can be achieved with reconstructed incidence data, using COVID-19 cases in the United States (US) as a case study. I conducted all work presented in this chapter, with input and feedback from my supervisors.

6.1 Introduction

Branching process forecasting models for COVID-19

As discussed in Section 1.1.4, branching process forecasting models, such as renewal equation based methods, rely on past daily incidence data, the latest estimate for R_t and either the Serial Interval (SI) or generation time (depending on whether past incidence is by date of infection or date of symptom onset). Using these inputs, future incidence can be projected over a specified time window assuming that R_t remains constant.

Throughout the COVID-19 pandemic, branching process models have been used to forecast future cases, hospitalisations and deaths to inform real-time response globally [291–293]. However, due to the protracted nature of the pandemic and the demand on resources, many countries, regions and public health agencies did not maintain daily reporting of COVID-19 incidence [294–297], and therefore the applications of these models were limited.

COVID-19 cases in the US

The first laboratory-confirmed case of COVID-19 in the US was reported in Washington state on the 20th January 2020 following recent travel to Wuhan, China [298]. On the 11th March 2020, the WHO officially declared COVID-19 a pandemic [299]. In the two years following this declaration (March

2020 - February 2022), the US experienced six distinct outbreak phases, which I have defined as the: 'Early Pandemic', 'Summer Surge', 'Winter Surge', 'Vaccine Rollout', 'Delta Wave' and 'Omicron Wave' (Figure 6.1).

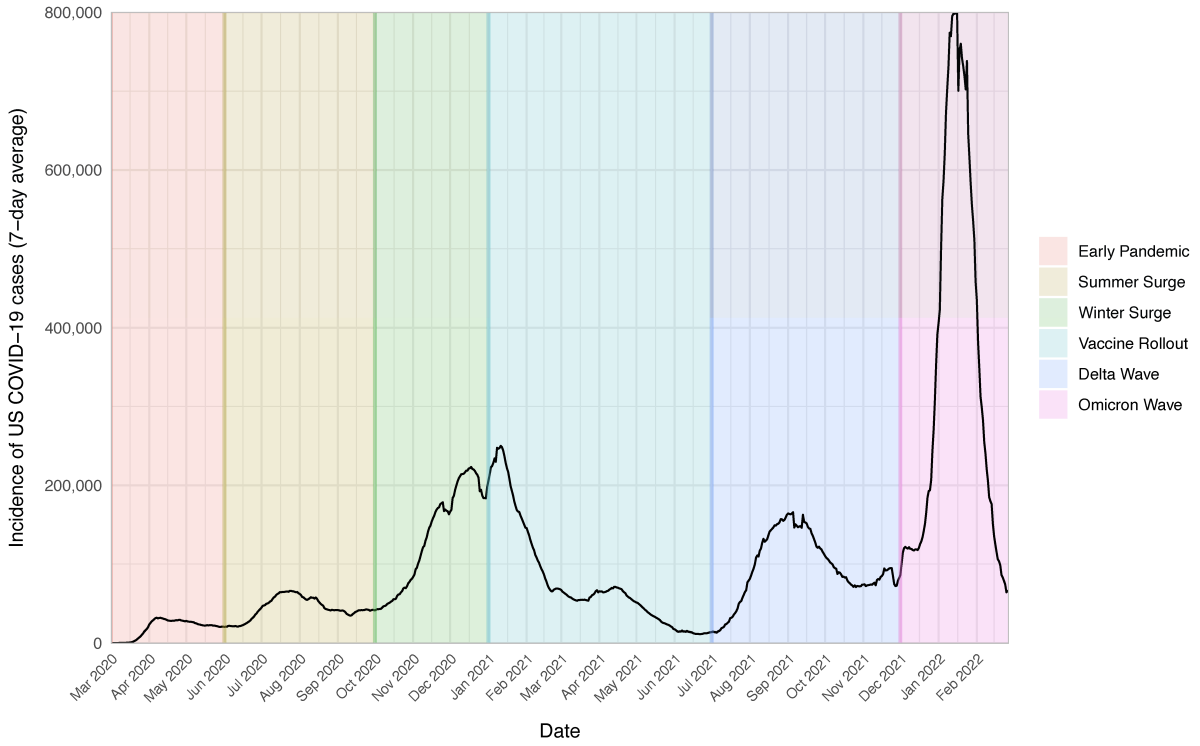


Figure 6.1: **Incidence of COVID-19 cases in the US (7-day rolling average) from March 2020 to February 2022.** Highlighted periods represent key phases of the pandemic, including: 'Early Pandemic' (March - May 2020), 'Summer Surge' (June - September 2020), 'Winter Surge' (October 2020 - December 2020), 'Vaccine Rollout' (January - June 2021), 'Delta Wave' (July - November 2021) and 'Omicron Wave' (December 2021 - February 2022). See text for a description of each phase.

During the 'Early Pandemic' (March - May 2020), containment measures such as social distancing, school and workplace closures and stay-at-home orders began to be implemented to curb the virus's spread [300]. Each US state had the autonomy to impose their own countermeasures and each had their own approach to reporting incidence data [301]. Across states there was considerable variation in the interventions implemented, when they were implemented and their duration [302].

The 'Summer Surge' (June - September 2020) followed the relaxation of initial lockdowns, resulting in a rise in case numbers in many states. A more substantial 'Winter Surge' (October - December 2020) then coincided with public holiday gatherings and colder weather, which led to increased indoor activities and travel [303]. In mid-December 2020, the first vaccine approved by the Food and Drug Administration (FDA) was made available to healthcare workers and older residents of long-term care facilities [300]. In early 2021, COVID-19 cases steadily declined during the phased 'Vaccine Rollout' to the population (January - June 2021) [304]. During this time, many US states reduced the reporting frequency of cases and deaths, typically shifting to reporting 4-5 days per week [296].

In July 2021, the Delta variant became the dominant variant in the US and caused another wave

of cases, thought to have been partly driven by uneven vaccination coverage [304]. The largest wave then occurred in December 2021 with the emergence of the highly transmissible Omicron variant. Fortunately, Omicron infections were associated with a lower risk of hospitalisation and death compared to those caused by the Delta variant [305]. From March 2022 onward, the policy approach to COVID-19 shifted toward endemicity across the US, focusing on long-term management strategies for COVID-19 [306]. At this stage, the vast majority of US states had ceased daily reporting of cases and deaths.

In this study, the forecasting performance of a branching process model will be compared using both daily reported and daily reconstructed COVID-19 case incidence data. This analysis aims to determine whether daily reconstructed data can reliably support accurate forecasting and response planning in instances where daily reporting has ceased. Additionally, it seeks to assess whether maintaining daily reporting offers any significant advantages over the use of reconstructed data.

6.2 Methods

COVID-19 data

Incidence of COVID-19 cases in the US were obtained from the COVID-19 Data Repository by the Center for Systems Science and Engineering at Johns Hopkins University [307, 308]. The time series chosen encompasses two years of COVID-19 circulation in the US from March 2020 to February 2022. Case incidence data for thirteen states were chosen for the analysis as they maintained mostly daily reporting throughout this time period. These states include California, Colorado, Delaware, Hawaii, Maryland, Missouri, New Jersey, New York, North Dakota, Ohio, Pennsylvania, Tennessee and Texas (accounting for around 45% of the US population [309]). Day of the week reporting patterns in the case incidence data were analysed by state and outbreak phase (Figure 6.1) by calculating the percentage of weekly cases reported on each day. Daily percentages were summarised across weeks using boxplots, which display the median, interquartile range, and overall spread of percentages (excluding outliers).

Reported incidence data were artificially aggregated to weekly counts and daily incidence data were reconstructed using the EM algorithm method (see Section 5.2). Both the reported and reconstructed daily incidence data were supplied as inputs to the forecasting model.

Model

The renewal equation is used as described in Chapter 1, Chapter 5 and Equation (5.1) to forecast the future incidence of COVID-19 cases [74]. The Serial Interval (SI) of COVID-19 is believed to have varied throughout the pandemic, becoming shorter with the emergence of new variants [310]. To provide a middle ground, the SI of mean 4.8 days and standard deviation 2.7 days was chosen for this analysis [311]. To calibrate the model, R_t is estimated using case incidence within a fixed calibration window prior to the projection date, where R_t is assumed to remain constant. To smooth out 'weekend

effects' in the data (as discussed in Chapter 5) a 10-day calibration window was chosen, following the approach of Bhatia et al [312]. For the period before the calibration window, combinations of R_t and case incidence consistent with the observed incidence in the calibration window were jointly estimated using Markov Chain Monte Carlo (MCMC) implemented within the 'jointlyr' R package [313]. The MCMC was run for 10,000 iterations, from which 1,000 sets of R_t and back-calculated incidence were sampled. For each sampled set, 10 stochastic realisations of the projected daily incidence of cases were drawn using a Poisson branching process (following the approach shown in Algorithm 3) [113]. The resulting 10,000 trajectories (1,000 parameter sets x 10 stochastic realisations) were summarised by computing quantiles for each forecast day. See Bhatia et al for further details [312].

Using this model, 28-day forecasts were produced (specified using 'ndays' in Algorithm 3) for each week of the dataset between 9th March 2020 to the 27th February 2022. Projections were initiated on each day of the week to explore whether the starting day influenced forecast performance. The 28-day forecasts using reported and reconstructed incidence data were then aggregated to weeks (forecast weeks 1-4) and their ability to capture the true weekly cases in each forecast week was compared. Projection weeks for certain states were excluded if zero cases were reported in the last week or if non-daily reporting was suspected (a conservative threshold was used: more than 30 cases reported in a week with at least one day of zero incidence) within the calibration window.

Forecasting performance was assessed using the Continuous Ranked Probability Score (CRPS) metric, which is a measure of the absolute difference between the observed weekly cases in the forecast period and the forecast distribution. Therefore, the lower the CRPS value, the closer the forecast is to the true observed cases for that forecast week. In order to mitigate the effect of the varying magnitude of case counts across the included US states, a transformation of $log(x + 1)$ was applied to the true cases and the predictions, as recommended by Bosse et al [314]. The CRPS was then calculated using the 'scoringutils' R package, which for this metric acts as a wrapper for the 'scoringRules' R package [315, 316]. CRPS values were then transformed using the base 10 logarithm and compared between the included US states and the outbreak time periods as defined in Figure 6.1.

6.3 Results

As in Chapter 5, hereafter I refer to reported and reconstructed incidence data, these are the reported daily incidence and the daily incidence that has been reconstructed from weekly aggregated data, respectively.

6.3.1 Reporting patterns by US state

The relative reporting of US COVID-19 cases on each day of the week compared to the weekly total varied according to the US state (Figure 6.2). Aggregated across all included states, the median

percentage of weekly reported cases was lower on Sundays and Mondays, with a peak midweek on Thursdays and Fridays. All 13 included states generally reported fewer cases on Sundays, Mondays or Tuesdays. Colorado and North Dakota displayed a typical 'weekend-effect' with lower reporting on weekends and high reporting at the start of the week. Most states, excluding Colorado, Delaware, North Dakota and Texas, had highest case counts on Thursdays or Fridays. The degree of week-to-week variability in daily reporting patterns differed between states. States such as Hawaii and Delaware showed more inconsistency in the percentage of weekly cases reported on different days of the week, whereas states such as Colorado and New York exhibited more consistent and stable reporting trends.

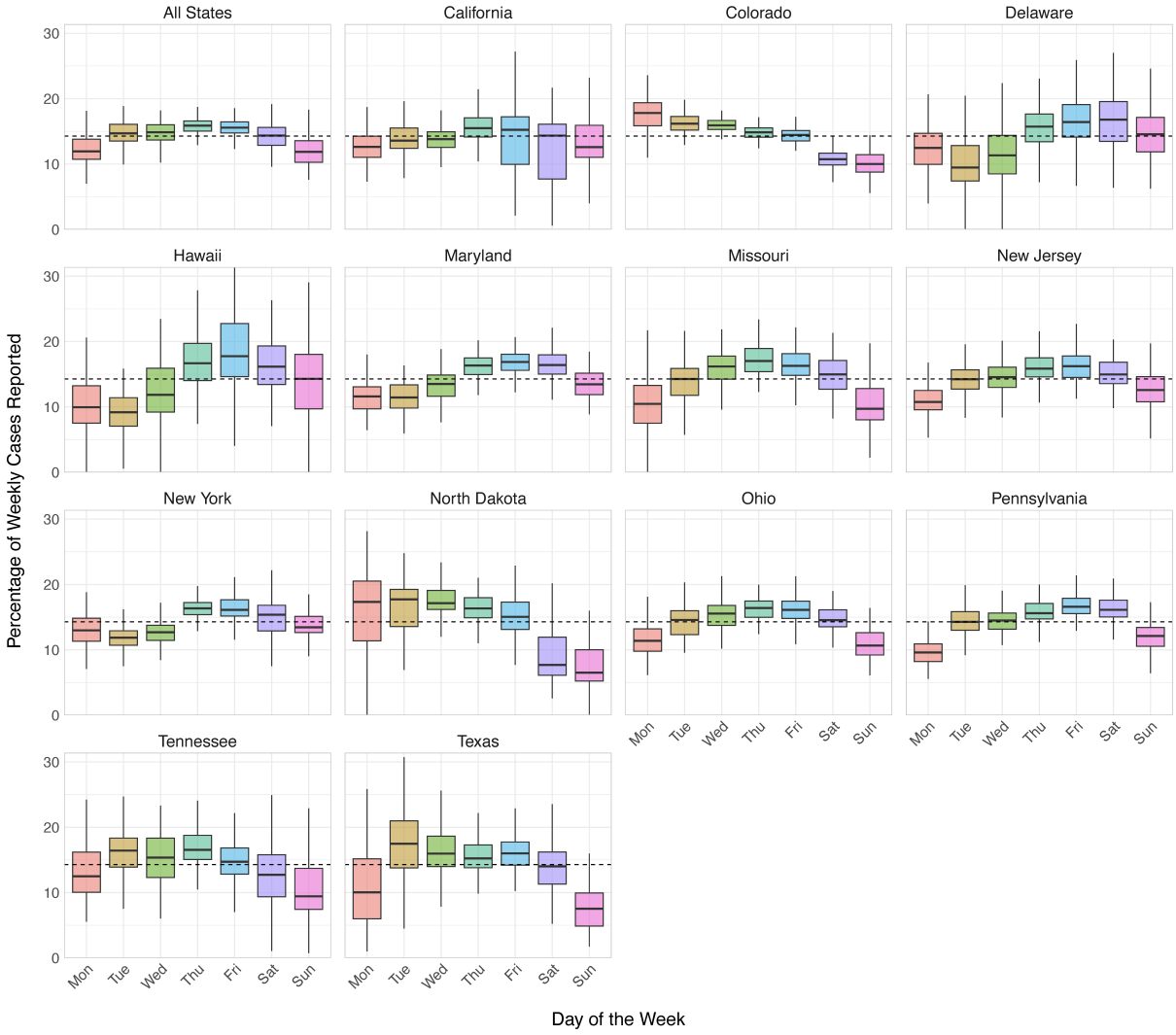


Figure 6.2: **Relative reporting of COVID-19 cases by day of the week for each US state.** Boxplots show the distribution of the percentage of weekly cases reported by day of the week, across weeks. Boxes represent the interquartile range (IQR), lines show medians, and whiskers extend to the most extreme values within 1.5 × IQR. Dashed lines represent the expected percentage if cases were reported equally on each day.

6.3.2 Reporting patterns by outbreak phase

The day of the week pattern of COVID-19 case reporting also varies by outbreak phase (Figure 6.3). The Summer Surge, Winter Surge, Vaccine Rollout and Delta Wave phases exhibited a similar overall

trend to that observed when aggregating across all phases, with a smaller median percentage of weekly cases reported on Sundays and Mondays, with peaks in reporting typically occurring midweek. The median percentage of cases reported peaked on Thursdays for the Vaccine Rollout and Delta Wave, while reported cases typically peaked on Tuesdays and Fridays for the summer and winter surges, respectively. The Early Pandemic exhibited a different trend, with the most cases reported from Thursdays to Saturdays, with lowest reporting on Tuesdays. The Omicron wave showed the most irregular pattern, with peaks on Thursdays and lowest reporting on Saturdays. Both the Delta Wave and Omicron Wave showed considerable week-to-week variability in daily reporting patterns, particularly on Sundays and Mondays.



Figure 6.3: **Relative reporting of COVID-19 cases by day of the week for each outbreak phase.** Each phase corresponds to those shown in Figure 6.1. Boxplots show the distribution of the percentage of weekly cases reported by day of the week, across weeks. Boxes represent the interquartile range (IQR), lines show medians, and whiskers extend to the most extreme values within 1.5 × IQR. Dashed lines represent the expected percentage if cases were reported equally on each day.

6.3.3 Forecasting performance by state

Averaged across all thirteen states included in the analysis, CRPS values showed only marginal differences within each forecast week (Figure 6.4), irrespective of whether reported or reconstructed incidence were used or the day of the week chosen for projections. As expected, CRPS values progressively increased over the forecast weeks, reflecting reduced performance over longer forecasting horizons.

When considered separately, each state follows the same general trend of increasing CRPS over the forecast weeks (Figure 6.4). Within each forecast week CRPS values are typically similar, but for some states there are small differences in performance depending on the projection day. For example, Friday and Sunday forecasts based on reported data appear to perform consistently worse (higher median CRPS values) in North Dakota.

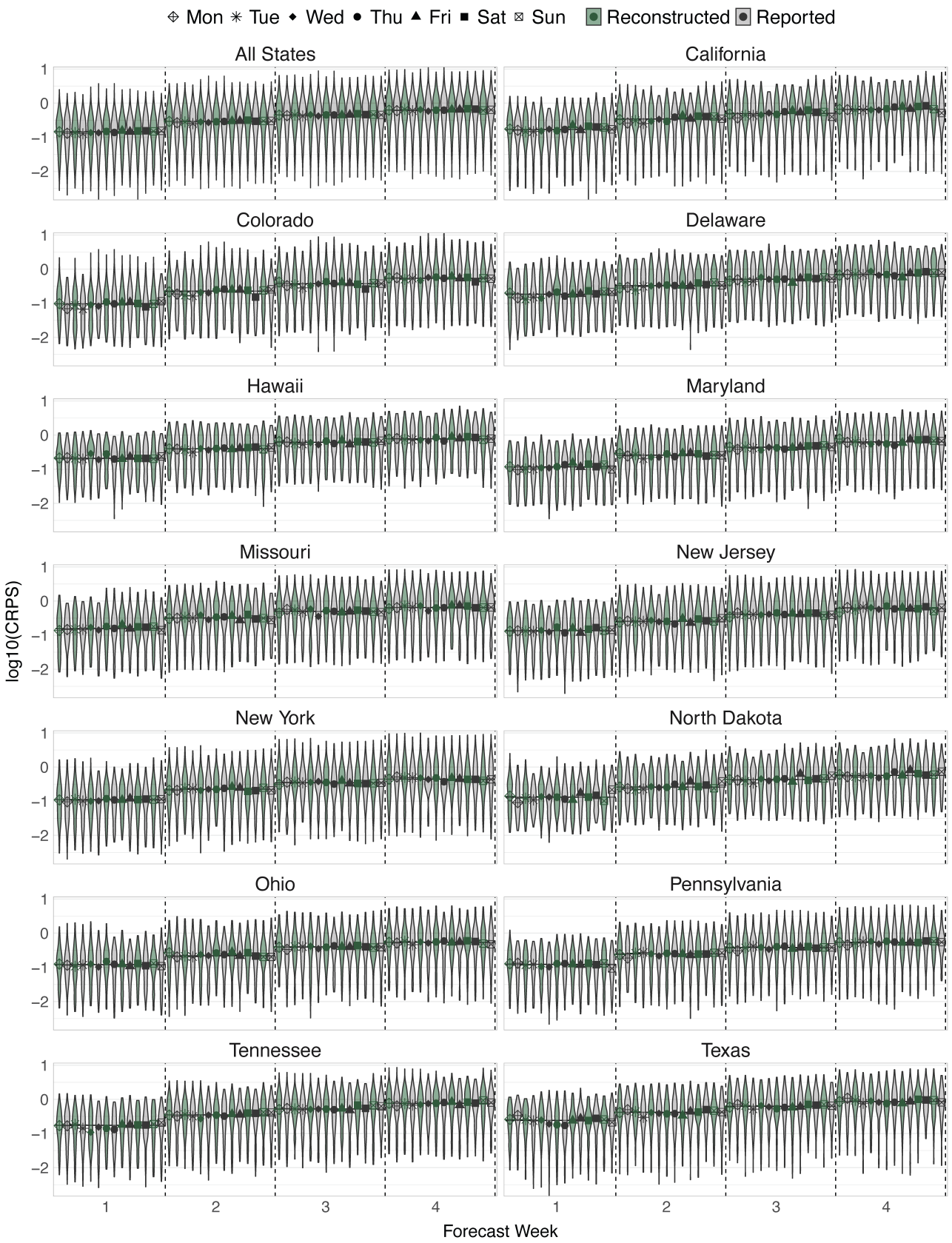


Figure 6.4: **Violin plot showing the distribution of log-transformed (base 10) CRPS values by US state, based on the difference in true observed weekly cases and weekly forecasts produced from reconstructed (green) and reported (grey) incidence data.** Points denote the median CRPS values, with symbols representing the projection day (last day of past incidence data supplied to the model): Mondays (diamonds with cross) Tuesdays (stars), Wednesdays (solid diamonds), Thursdays (circles), Fridays (triangles), Saturdays (solid squares) and Sundays (squares with cross). The solid black line represents the overall median CRPS score for the forecast week. The "All States" panel shows the distribution of CRPS values across all weeks and all individual state forecasts.

6.3.4 Forecasting performance by outbreak time period

When considering outbreak time periods separately (Figure 6.5), forecasting performance was worst overall (higher combined median CRPS in each forecast week) during the early pandemic (March - May 2020) and the large Omicron wave (December 2021 - February 2022). During the early pandemic, forecasts using reported incidence data perform slightly better, other than Saturdays in the final forecast week. During the Omicron wave, Wednesday forecasts using reported data perform best. In the summer surge (June - September 2020), winter surge (October - December 2020), vaccine rollout (January - June 2021) and Delta wave (July - November 2021), differences between forecasts appear negligible.

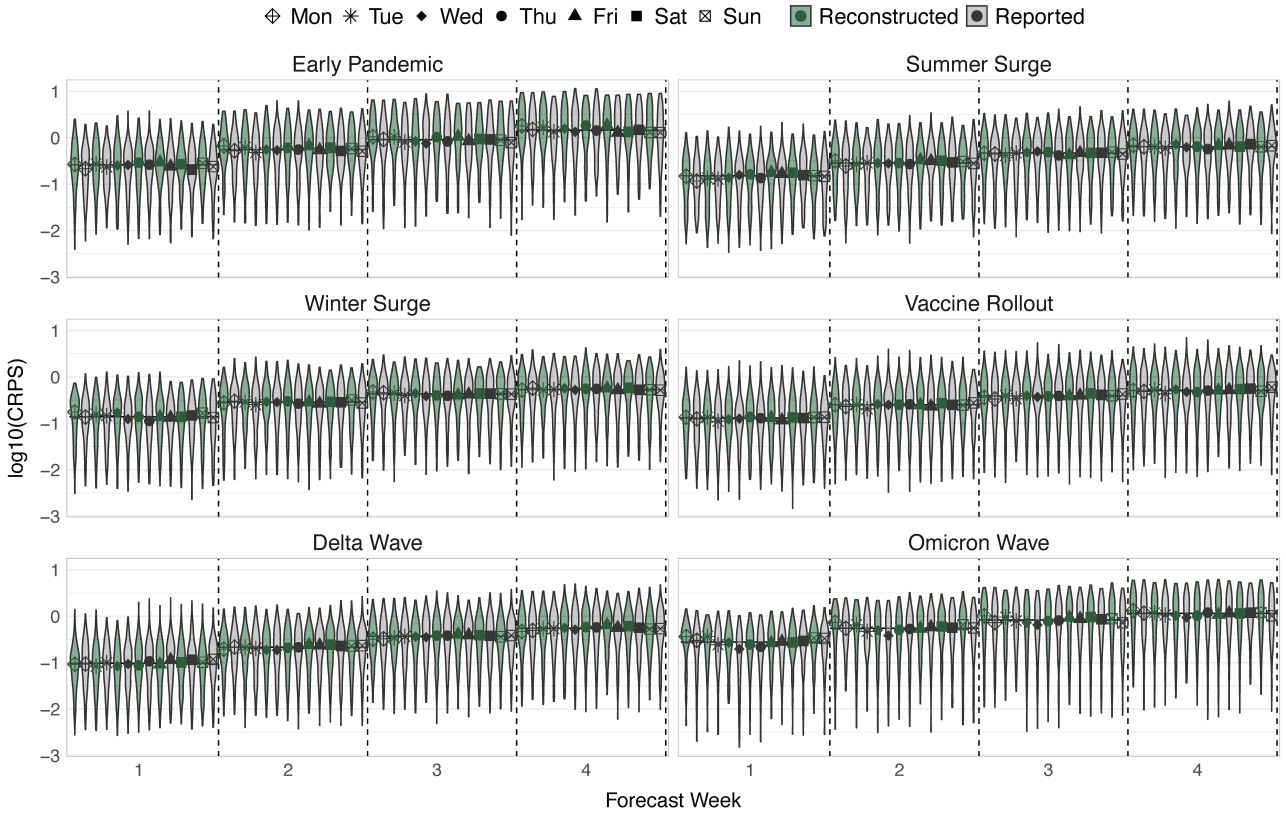


Figure 6.5: **Violin plot showing the distribution of log-transformed (base 10) CRPS values for all included states by outbreak time period, based on the difference in true observed weekly cases and weekly forecasts produced from reconstructed (green) and reported (grey) incidence data.** Outbreak phases correspond to those shown in Figure 6.1. Points denote the median CRPS values, with symbols representing the projection day (last day of past incidence data supplied to the model): Mondays (diamonds with cross) Tuesdays (stars), Wednesdays (solid diamonds), Thursdays (circles), Fridays (triangles), Saturdays (solid squares) and Sundays (squares with cross). The solid black line represents the overall median CRPS score for the forecast week.

6.4 Discussion

The applications of branching process models can be restricted if daily incidence data are unavailable or if there is a transition to aggregated incidence data, such as weekly counts, during prolonged outbreaks (see Chapter 2). Many US states stopped reporting daily incidence for COVID-19 in summer

2021 [296], and the European Centre for Disease prevention and Control (ECDC) switched to reporting weekly incidence as early as 14th December 2020 [297]. As mentioned in Section 1.1.4, branching process models are especially valuable during the early stages of outbreaks when limited data make it challenging to reliably parameterise more complex models. In this study, the Expectation Maximisation (EM) algorithm has been applied to US COVID-19 case incidence data to demonstrate that reconstructed incidence data can achieve comparable forecasting performance to reported incidence data regardless of the projection day (i.e., the last day of past incidence data supplied to the model), US state, or outbreak time period. The findings suggest that while daily reporting may offer some advantages under specific circumstances, such as during the early pandemic, the differences in forecasting accuracy are generally minimal. This highlights the viability of reconstructed data as a practical alternative for supporting accurate forecasting efforts in contexts where setting up and maintaining daily reporting is logistically challenging or resource-intensive.

Forecasting performance across US states

For all included US states combined, model performance was similar regardless of the projection day or whether daily incidence data were reported or reconstructed (Figure 6.4). Increasing median CRPS values across forecast weeks (indicating worse performance) highlights the inherent uncertainty in predictions with longer forecasting horizons. This pattern is expected, as projections assume that transmissibility remains the same as in the calibration window of past incidence data throughout the forecasting period. Over longer time frames this assumption becomes more unlikely due to factors such as increasing behaviour change and the implementation or relaxation of interventions [291].

When US states were considered independently, there was no clear advantage for the use of either reported or reconstructed incidence data. However, the choice of projection day influenced forecasting performance for some individual states (Figure 6.4). US states recorded cases by the date they were reported, which is delayed from the actual test date, and were lowest on Sundays and Mondays with a relatively even distribution around a peak on Thursdays (Figure 6.2), which is similar to the trend observed across all US states by the COVID Tracking Project between March 2020 and February 2021 [317]. This is different to patterns observed in the UK (Chapter 5, Figure D.7 & Figure 6.2), where case counts were recorded by the date of specimen collection for testing and typically exhibited a 'weekend-effect' with lower case counts from Friday to Sunday and relatively stable reporting over the rest of the week (Figure D.7).

Each US state has its own health department and reporting practices, which lead to state-specific differences in reporting patterns (Figure 6.2). Most US states reported data with a delay of one or two days, causing a corresponding shift in the 'weekend-effect' to the beginning of the following week [317]. While incorporating state-specific reporting delay distributions into the model might improve forecast accuracy, detailed data on these delays were not available beyond general patterns

inferred from reporting trends. Moreover, the primary aim of this chapter was not to optimise model performance, but to compare model performance using different types of input data (reported vs reconstructed incidence). There was no obvious correlation between state-specific reporting patterns (Figure 6.2) and the best performing projection days in forecasts. The model calibration window (see methods Section 6.2) was chosen to smooth out reporting patterns in the data, however this 10-day window would have encompassed either two peaks or two dips in the relative reporting of cases, which could have caused projection day variations in the results. As shown in Chapter 5, when using reconstructed incidence, calibration windows should always be at least the length of the original data aggregation and longer windows should help smooth out discontinuities in the reconstructed incidence (see Appendix D.3.2). Future research could consider increasing the calibration window to 14-days, however if it is too long it may be less representative of current transmission.

Given the differences observed in forecasting performance, multiple projection days could be considered for real-time analysis. It is important to also consider the practical utility of forecasts depending on the specific day of the week they are generated. For example, projections made using incidence data up to Sunday are performed on Mondays, which can guide the week's response [291]. In contrast, projections based on incidence data up to Friday or Saturday would need to be performed over the weekend, which may be impractical given that they are not typically working days. The timing of incidence data publication in each state will also play a role, as some datasets may be updated daily, while others may only be released once a week, limiting flexibility in when forecasts can be performed for real-time response. Therefore, the choice of projection day should balance practical constraints with potential gains in forecasting performance.

Forecasting performance across outbreak phases

When considering distinct outbreak phases (Figure 6.1), forecasts made during the summer surge to the delta wave (June 2020 to November 2021) showed minimal differences in forecasting performance regardless of the dataset and projection day. In contrast, forecasting performance was notably worse in the early phase and the Omicron wave (Figure 6.5). Several factors likely contributed to this disparity. Reporting patterns differed most noticeably in the early pandemic and Omicron wave (Figure 6.3). During the early phase, US states had never collected data at the scale the pandemic demanded before, contending with considerable test shortages and longer reporting delays as testing facilities and health departments adjusted to the initial surge in cases [318]. Despite these limitations, forecasts based on daily reported incidence consistently outperformed those using reconstructed incidence, emphasising the importance of prioritising reported daily incidence for forecasting in the early stages of outbreaks.

During the Omicron wave, healthcare systems were also under immense strain, which could have led to inconsistencies in reporting and therefore worse forecasting performance. Taking into account different projection days and datasets, there were only small differences between them during the

Omicron wave, but reported incidence data with a Wednesday projection day performed best.

Limitations & Future Research

While this study provides valuable insights, there are several limitations. Forecasting performance and the reconstruction of incidence data itself can be influenced by high sensitivity to the specified serial interval (or generation time), as even small errors can lead to significant discrepancies in projected outcomes [75]. Given that the method used does not allow for a time-varying SI, the SI chosen for this analysis aimed to provide a reasonable midpoint for the SI throughout the pandemic. However, this may partly account for the worse performance observed in the early pandemic and Omicron wave when the SI was estimated to be slightly longer and shorter respectively (Figure 6.5) [310]. It may also explain the improved performance of forecasts using reported data in the early phase. Future sensitivity analyses could explore whether performance improves with varying SI distributions.

Additionally, weekly smoothing occurs when daily incidence data are reconstructed from weekly case counts (see Chapter 5 Section 5.4). This weekly smoothing, combined with the model's 10-day calibration window, could influence forecasting accuracy during periods of rapid changes in transmission or reporting frequency, such as the early pandemic and Omicron wave. It is also important to consider how the reporting rate varies over the course of an outbreak [319]. For instance, if the healthcare system is overwhelmed and does not have the capacity to handle an influx of cases, the proportion of reported cases would likely be lower than during time periods where the healthcare system is under less pressure. If the proportion of cases reported during the forecast period differs to the calibration window, this will lead to bias in comparisons between forecasts and observed weekly counts. Future research could include simulation studies that introduce a time-varying reporting rate to the incidence data to see how this influences performance using each dataset.

Future research could also explore the forecasting performance of reconstructed incidence in other outbreak contexts. While this study focuses on weekly aggregated incidence, shorter reporting intervals, such as 4-5 times per week (as adopted by some US states [296]), could be considered. Furthermore, this analysis focused solely on COVID-19 cases in the US, and future studies could assess the applicability of this method to other pathogens and countries.

6.5 Conclusion

Reconstructed incidence data can support reliable and accurate forecasting when daily reported incidence is not available. While daily reported incidence remains critical during the early stages of outbreaks, the findings emphasise the viability and practicality of using reconstructed incidence for diseases where cases are typically reported over longer timescales or when reporting changes arise during prolonged outbreaks. Balancing the benefits of daily reporting with the flexibility offered by reconstructed data can ensure more resilient forecasting methods in real-time outbreak response.

Chapter 7

Discussion

Infectious disease outbreaks remain a substantial threat to global public health, with their frequency escalating due to factors such as rapid urbanisation, globalisation and climate change [9]. Recent studies have shown that land-use and climate change have contributed to an increase in the frequency of zoonotic spillover events, in which a pathogen crosses from animal hosts to humans [320]. These trends emphasise the importance of strengthening our preparedness and response strategies to mitigate the impact of future outbreaks.

During an outbreak, several key questions arise (Figure 7.1), including: *How transmissible is this pathogen? How does this outbreak compare to previous outbreaks? How does a new variant compare to current or past variants? What is the projected impact of the outbreak?* Many methods and associated software tools have been developed to answer these questions. However, it is critical to continuously evaluate their strengths and limitations to ensure that they are applicable across a wide variety of outbreak contexts.

In this thesis, I began by assessing the strengths and limitations of tools used to estimate transmissibility in real-time using the time-varying reproduction number, R_t, with a focus on branching-process models (Aim 1 - Section 1.3 and Chapter 2). The insights gained from this study formed the foundation for subsequent projects in this thesis and were applicable to outbreak tools beyond those used to estimate R_t, such as those used to produce epidemic forecasts.

I explored data-related challenges that impact the applicability of tools in certain contexts (Aim 3 - Section 1.3). First, I presented a software tool that is capable of estimating the transmission advantage of pathogen variants in real-time without the need for genomic sequencing data (Chapter 4). Then, to address issues regarding the temporal scale of reported incidence data, I developed an Expectation Maximisation (EM) algorithm to reconstruct daily incidence data from any temporal aggregation of incidence data (Chapter 5). I demonstrated that this reconstructed daily incidence can perform as effectively as, if not better than, reported daily incidence when applied to both R_t estimation (Chapter 5) and short- to medium-term forecasting (Chapter 6).

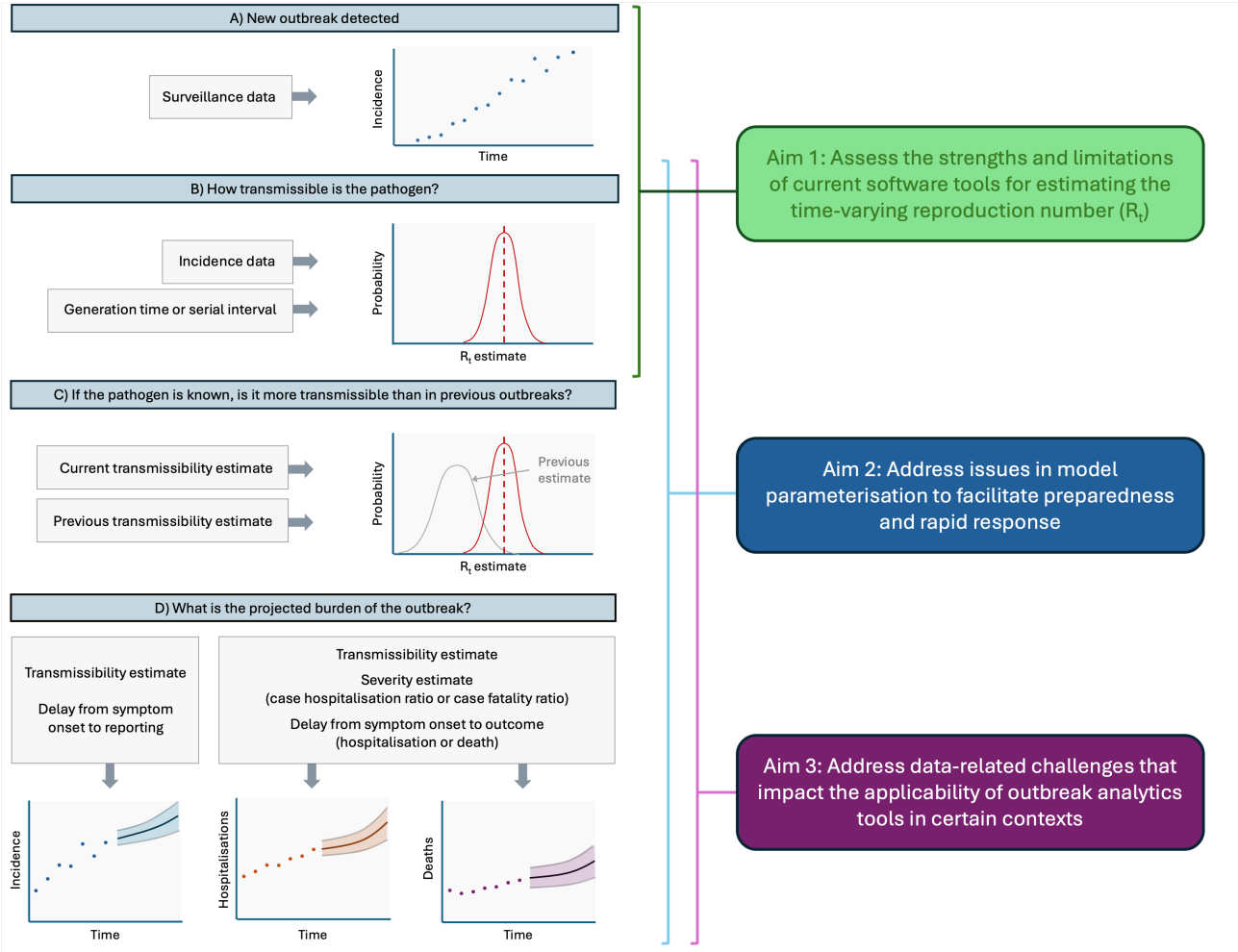


Figure 7.1: **Schematic showing how each aim of the thesis addresses aspects of outbreak response.** The left hand side is Figure 1.1, as shown in Section 1.2.

Even if data were perfect, there are also issues with reliable model parameterisation. To address the challenges in model parameterisation to facilitate rapid response and preparedness for future outbreaks (Aim 2 - Section 1.3), I summarised the key transmission and epidemiological parameters for Ebola Virus Disease (EVD) (Chapter 3). The resulting open source database is dynamic and can be updated in line with new evidence as it becomes available.

By addressing gaps in our current approaches, these improvements ensure that we can better meet the demands of future outbreaks.

7.1 Incidence data will always be imperfect

Accurate and reliable disease incidence data is critical for real-time outbreak response; however, there are numerous issues that influence the quality of incidence data, many of which are predictable yet largely unavoidable during the complexities of outbreak response. When creating outbreak analytics tools and interpreting their findings, it is imperative that these limitations in incidence data are taken into account.

This thesis focuses primarily on the use of the renewal equation for estimating R_t and generating forecasts, which in theory models disease dynamics based on the incidence of cases by date of infection [60, 61, 63]. In practice, incidence by date of infection is difficult to observe and alternative incidence data such as by date of symptom onset, report, hospitalisation or death, are often used instead (Figure 7.2). Each of these proxies comes with its own limitations. For example, incidence by date of symptom onset may be more susceptible to factors such as recall bias, whereas incidence by date of death inherently lags further behind the true transmission dynamics. In Chapter 2, factors such as under-reporting, administrative noise (such as 'weekend-effects') and delays in reporting, were all found to be pervasive issues influencing the reliability of incidence data, limiting R_t estimation accuracy.

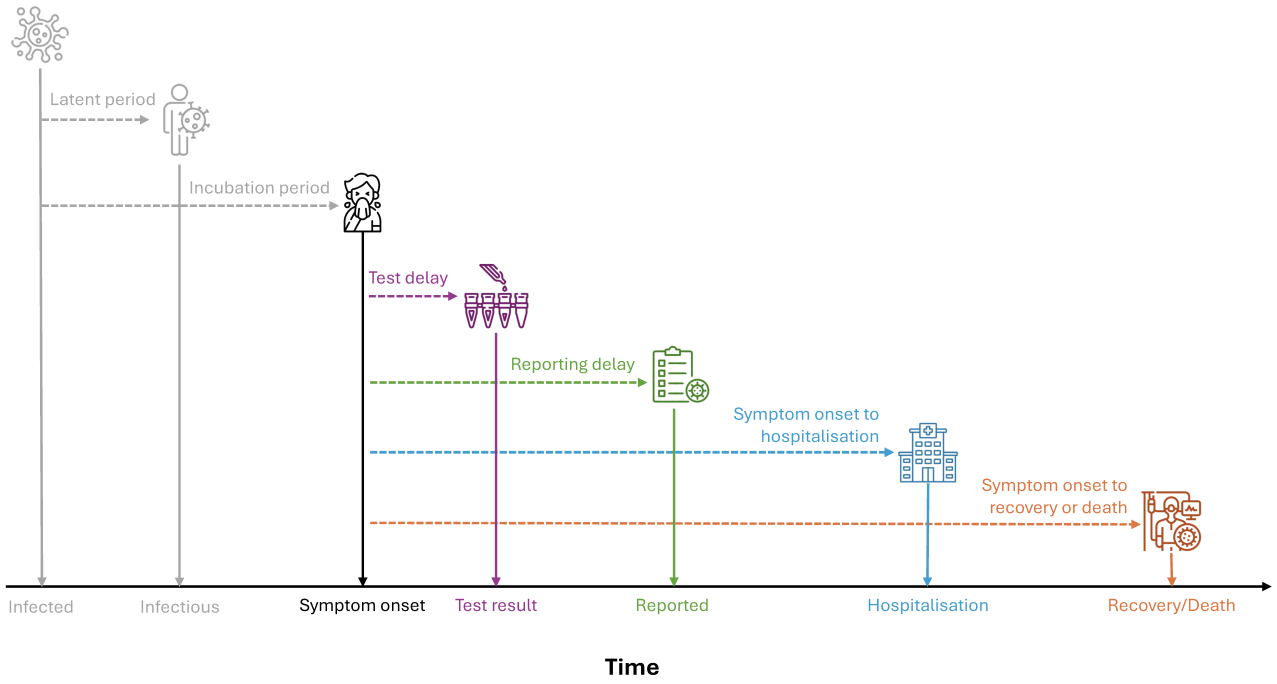


Figure 7.2: **Schematic showing an example of the infection process for an individual and the corresponding epidemiological delays.** Typically, the date of infection is unobserved (grey). In this example, the latent period (time from infection to the onset of infectiousness) is shorter than the incubation period (time from infection to symptom onset), leading to a period of asymptomatic transmission. The date of symptom onset is easier to observe (black) and can be used to back-calculate incidence by date of infection using estimates of the incubation period. Here, the individual is tested once they become symptomatic (purple). Before cases are ultimately reported (green), delays are introduced by the testing process used to confirm cases and other administrative delays. If cases by date of symptom onset or date of report are unavailable or unreliable, then incidence by date of hospitalisation (blue) or incidence by date of death (orange) may be used instead. Incidence by date of symptom onset can be inferred using the delay from symptom onset to hospitalisation or death, respectively. Notably, data from later stages of this infection timeline represent a progressively smaller proportion of the total cases. For example, not all cases are tested or reported, and only a subset of cases result in hospitalisation or death. Therefore, assumptions about the case ascertainment, case hospitalisation and case fatality ratios may also be required.

Under-reporting can occur for several reasons, including behavioural factors (e.g., differences in healthcare-seeking behaviour due to disparities in access to medical services, perceived severity of symptoms, or fear of stigma), misclassification, and misdiagnosis. It can also result from inadequate surveillance systems, insufficient testing practices, or asymptomatic transmission. This can be worse

in the early stages of outbreaks when rapid diagnostic tests may not be readily available, resources are limited (e.g., shortages of tests and staff) or when awareness of the disease is low. For example, during the 2022 mpox outbreak in Europe, many clinicians were unfamiliar with the clinical presentation of mpox given that the disease was historically endemic to West Africa, leading to missed cases [321].

If an individual is captured by the surveillance system, under-reporting may still occur due to differences in the definitions used to define 'confirmed', 'probable' and 'suspected' cases (Section 1.1.2). These definitions balance sensitivity (correctly identifying true cases) and specificity (correctly excluding non-cases) and can vary depending on the country or region, and over time if the case definition is refined during an outbreak [322]. A more sensitive case definition in one location or time compared to another can lead to higher reported case numbers, even if the actual burden of the disease is similar. Whereas, an overly specific case definition could lead to under-ascertainment of true case incidence. Additionally, some public health bodies may only report confirmed cases in incidence counts, while others include probable or suspected cases.

This issue also applies to the incidence of deaths. During the COVID-19 pandemic, some countries did not report an official definition of death due to COVID-19, and those that did had definitions that varied considerably [323]. Stringent criteria for attributing deaths to COVID-19 led to an underestimation of the true mortality burden. For example, in Russia, a death was only attributed to COVID-19 if SARS-CoV-2 infection was confirmed on autopsy, which was not performed for all deaths [324]. If there is no consensus on the definitions of cases or deaths across locations or time periods, then direct comparisons between them will not be valid. This issue can only be addressed by the standardisation of case definitions and criteria used for testing eligibility, but previous pandemics have shown that this is unlikely to occur in practice. During the COVID-19 pandemic, test capacity varied substantially depending on the country and availability of resources [325], and case definitions were continuously updated in line with further knowledge of SARS-CoV-2 etiology and clinical presentations [326–329]. Therefore, the most feasible solution is to strongly advocate for each country and region to explicitly document the criteria they use to identify and define cases, and how this may vary over time, in order to fully contextualise comparisons. It is important to note that if the level of under-reporting is relatively constant within locations then comparisons of R_t between locations remains feasible. For instance, even if the true incidence of cases differs from what is observed, if the proportion of cases reported remains the same, the trend in the "true" value of R_t should be consistent with the trend in R_t estimates derived from the observed data.

Under-reporting can also follow weekly patterns due to administrative noise, such as 'weekend-effects' (see Chapter 2). As shown in Chapter 5 and Chapter 6, reporting patterns by weekday can vary substantially depending on the epidemic context and should be accounted for when choosing calibration windows for models estimating R_t and forecasting future incidence. In instances where case data reporting is particularly sporadic, such as the early stages of outbreaks, alternative approaches

have included the incorporation of a day of the week effect to account for weekly noise in the data [168], the addition of a time-varying reporting rate [169], and the use of death data, which is typically reported more reliably [117, 158]. The findings in Chapter 5 support the use of death data, as UK COVID-19 deaths were far less influenced by administrative noise in reporting than cases (Figure D.7). However, it is important to be aware of the "severity pyramid" of incidence data [330]. For instance, surveillance is more likely to detect severe cases of a disease given that they present to hospital or die, but this represents only a subset of cases. If a vaccine is widely distributed that dramatically reduces the incidence of hospitalisations or deaths, or the characteristics of a pathogen changes over time, e.g., a new variant emerges that causes less severe disease, then hospitalisations and deaths would be less informative about the magnitude of case incidence over time. These factors must be taken into consideration when determining which incidence data stream to prioritise at different stages of an outbreak.

Even when cases or deaths are ultimately reported, incidence data is often affected by reporting delays, resulting in right censoring. For instance, when using incidence by date of symptom onset, the most recent cases will be missing from the data as they have not yet been reported, and these delays between symptom onset and eventual reporting can be substantial. Using Ebola Virus Disease (EVD) as an example, these delays typically ranged from four days to two weeks, which would lead to corresponding delays in detecting changes in transmissibility and potential underestimation of future cases if the incidence data is used to generate forecasts.

If incidence by date of hospitalisation or death are used, additional epidemiological delays must be accounted for. In Chapter 3, the delay from symptom onset of EVD to admission to care was found to be between 3 and 6 days for 94% (n = 44) of published estimates. Additionally, a random-effects meta-analysis (accounting for both within-study and between study variance) estimated that the delay from symptom onset to death was 9.3 days (95% CI 8.1-10.1 days). These longer delays can make it more difficult to link trends in R_t with contextual changes, such as the implementation of interventions.

Fortunately, Chapter 2 showed that R packages such as EpiNow2, Epidemia and bayEStim addressed these issues by allowing users to add delay distributions [167–169]. For example, EpiNow2 allows users to estimate and account for right truncation by estimating the reporting delay using multiple snapshots of past incidence data [168, 331]. Additionally, incidence by date of infection can be back-calculated from incidence by date of symptom onset using estimates of the incubation period. Similarly, incidence by date of symptom onset can be inferred from incidence by date of hospitalisation or death using estimates of the delay from symptom onset to hospitalisation or death, respectively [168]. However, these delays must be reliably parameterised, otherwise estimates will be biased (see Section 7.2).

While Chapter 2 demonstrates that many developments have been made to improve the applicability of tools to account for issues in disease incidence data, two critical gaps were identified, including

the lack of methods to handle different temporal and spatial scales of incidence data.

The inability to work directly with temporally aggregated (e.g., weekly) incidence data with many existing tools was a key limitation for pathogens that are typically reported weekly, such as influenza and Zika, as well as most notifiable diseases in the UK and the US [271–274]. Moreover, as discussed in Chapter 5 and Chapter 6, many regions and countries stopped reporting daily incidence data for COVID-19 over time [281, 282, 294–297]. Given that incidence data at a finer temporal resolution is often unavailable for many pathogens, a method was needed to maximise the utility of available incidence data without introducing bias.

The EM algorithm in Chapter 5 addressed this issue by allowing daily incidence data to be reconstructed from any temporal aggregation of incidence. I have shown in Chapter 5 and Chapter 6 that reconstructed daily incidence data is able to perform just as well, if not better, than reported daily incidence when generating rapid and accurate estimates of R_t and short- to medium-term forecasts using branching process models. In incidence datasets with a large degree of weekly administrative noise, the findings suggest that reconstructing daily incidence from weekly counts may be a preferable approach, thereby removing noise from the data.

This method has been integrated into the EpiEstim R package, enhancing its ability to be rapidly and seamlessly applied to a broader range of pathogen incidence datasets to estimate R_t. To extend the utility of the method beyond EpiEstim, the reconstructed daily incidence data can be easily extracted and used as an input for other outbreak tools (such as the branching process forecasting model in Chapter 6), depending on user requirements. Overall, this development enhances the versatility of EpiEstim and potentially other outbreak analytics tools, improving their capacity to support real-time outbreak response.

An issue that remains unresolved is the need for tools that account for spatial factors in transmissibility estimates. While many developments have been made to account for spatial factors on an adhoc basis [136, 140, 152, 155, 157, 158, 163, 165, 332], these efforts have not resulted in the development of reusable, open-source software. Given the importance of spatial factors in transmission, addressing this gap should be a priority for future research. Transmissibility in one location can affect neighbouring locations due to population movement. Explicitly modelling these interactions when estimating transmissibility could allow for a more comprehensive assessment of the impact of interventions. For example, developing tools to evaluate how implementing or easing control measures (e.g., social distancing) in one area might influence transmission in neighbouring areas would provide a valuable tool for outbreak response.

7.2 Facilitating rapid and accurate parameterisation is essential for outbreak preparedness

Outbreak models are only as reliable as the parameter values they depend on. In the early stages of outbreak response, it is necessary to quickly determine appropriate model parameter values to ensure timely characterisation of pathogen transmissibility, and to support the effective implementation of interventions, thereby minimising the impact of the disease as much as possible.

If evidence generation to support decision-making is delayed, this can have a drastic impact on the effectiveness of interventions. For example, Knock et al. performed a retrospective modelling study for COVID-19 in England, suggesting that if lockdown was implemented just one week earlier, mortality could have been roughly halved in the first wave (up to July 2020) [333]. Similar findings were reported in the US, suggesting that if non-pharmaceutical interventions had been implemented one to two weeks earlier, a substantial number of cases and deaths could have been averted [334].

Moreover, to promote confidence and trust in the decision-making process, it is important to demonstrate that model assumptions are well supported by evidence from previous outbreaks. Knowledge of the key epidemiological and transmission characteristics for known pathogens requires the consolidation of information from historical outbreaks, while the parameterisation of models for newly emerging pathogens typically relies on information from related pathogens. As mentioned in Section 1.2, initial models for COVID-19 had been parameterised using information for pathogens from the same family of viruses, including Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS).

Given the impact delays in producing actionable insights can have during the early stages of outbreak response [333, 334], it is crucial to pre-empt the need for scientific evidence by synthesising the available information in advance to facilitate timely model parameterisation. There is a need for an easily accessible, centralised resource that consolidates key transmission and epidemiological characteristics for known pathogens.

To address this issue, the Pathogen Epidemiology Review Group (PERG) at Imperial College London, in collaboration with the WHO and the London School of Hygiene and Tropical Medicine, has undertaken an initiative to systematically review all WHO priority pathogens [190]. This includes COVID-19, Marburg virus disease, Crimean-Congo haemorrhagic fever, Lassa fever, EVD, MERS, SARS, Nipah and henipaviral diseases, Rift Valley fever and Zika [193]. Alongside the systematic reviews, PERG sought to create a standardised extraction database and the epireview R package, designed to reproduce analyses for all reviewed pathogens [189].

To act as a case study and template for this resource, I led a systematic review of EVD transmission and epidemiological characteristics, in addition to all known transmission models that have been used to model EVD outbreaks up to 7th July 2023 (Chapter 3). The extracted information included parameter

values that may be needed to parameterise a simple transmission model, such as the basic and effective reproduction numbers, the serial interval, generation time, other epidemiological delays, seroprevalence, attack rates, the CFR, the Infection Fatality Ratio (IFR), mutation rates and risk factors for infection or death. In total, 1280 parameters and 295 models were extracted from 522 papers.

This work was critical in initiating the core functionalities of the epireview R package, which is intended to be a living resource, continually updated as new data become available [189]. Furthermore, the standardised database created as part of the EVD work has been used as a template for other diseases, and is continually refined to accommodate the specific requirements of each pathogen under investigation. At the time of writing, systematic reviews have been completed for Marburg virus disease, Lassa fever and SARS [335–337]. Reviews for Zika and Nipah are currently ongoing.

For EVD, the extracted parameter estimates varied depending on contextual factors (e.g., the country and outbreak), as well as the specific species of EVD, and were summarised accordingly. Random-effects meta-analyses, which account for heterogeneity in estimates within and across studies, should provide the optimal estimates for use in EVD models. However, despite the vast amount of data extracted, there was a lot of heterogeneity in how parameters were reported, making it difficult to directly compare and summarise estimates. Furthermore, meta-analyses could only be performed for a limited number of parameters due to insufficient data available in the required format (see Appendix B.3.6). Based on these findings, it would be useful for public health agencies to recommend best-practice guidelines for reporting epidemiological parameters in the literature to ensure that we can make the most of published data in future analyses.

This study also highlighted the lack of information published for EVD species other than Zaire, such as Bundibugyo, Sudan and Taï Forest. The vast majority of parameter estimates were also made in the context of the West African EVD outbreak. This demonstrates that published information is not always going to be comprehensive, but compiling all available evidence in one place is still a critical resource for rapidly parameterising models during future outbreaks.

In addition to acting as a resource for parameterising models, compiling historical outbreak information into open-source databases is also valuable for determining "How does this outbreak compare to previous outbreaks?". As discussed in Section 1.2, these analyses are particularly valuable when determining whether outbreaks which are larger than normal are caused by differences in pathogen or transmission characteristics, or simply due to the inadequacy of control measures.

For EVD, the 2014-16 outbreak in West Africa far exceeded the scale of all previous outbreaks combined [65]. Transmissibility estimates and the clinical timeline of infection were similar to previous outbreaks, suggesting that the scale was driven by insufficient control measures and specific characteristics of the affected populations, such as the high level of cross-border mobility in Guinea, Liberia and Sierra Leone [65]. Therefore, these analyses are critical for understanding the factors driving current outbreaks and guiding effective responses.

7.3 Maximising the utility of models in policy-relevant decision making

The outputs of epidemiological analyses are used to inform policy decisions for governments and public health agencies globally. For example, considering the outputs of the Imperial College COVID-19 response team alone (including reports, journal articles, pre-prints and news articles), 43% (n = 217) of these outputs were cited in one or more policy documents from 41 countries or regions in 26 different languages [338]. Of these 217 outputs, there was a mean of 8 policy document citations per output, demonstrating the impact of outbreak analytics in evidence-based decision making.

Given that estimates of R_t are an important component of the evidence used to inform policy, it was surprising to find that very few R packages had been created to directly answer important practical or logistical questions during outbreaks (Chapter 2). One such question that arises when an outbreak is coming to an end is whether to declare that an outbreak is over and the associated risk of epidemic resurgence. Historically this has relied on a predefined time period to pass (e.g., the maximum duration of two incubation periods) without an observed case [339, 340]. The reliability of these arbitrary guidelines in end-of-outbreak declarations has been questioned since they do not account for potential periods of remission (which can result in recrudescence of symptoms for diseases like EVD) and under-reporting [341]. In order to provide policy-makers with a more quantitative threshold of confidence that an outbreak is over, EpiFilter estimates the probability that a disease has been eliminated in real-time [124, 162, 171]. Another question is the impact of non-pharmaceutical interventions on transmission, such as social distancing and mobility restrictions, which has been addressed by Epidemia using a logistic regression framework [169]. Many key questions remain, therefore, developing open-source software tools to directly answer policy-relevant questions was identified as a promising avenue for future research.

As the COVID-19 pandemic evolved so did the virus itself, and one of the most common challenges for policymakers was quantifying how much worse a new variant was to those currently circulating, and determining whether interventions needed to be intensified in response. In late 2020, the WHO began to characterise SARS-CoV-2 variants as either Variants Of Interest (VOIs) or Variants Of Concern (VOCs) [342]. VOIs are variants with genetic changes that make it more transmissible, severe, or less easily detected or treated [343]. VOCs meet the same definition, but must also meet at least one additional criterion, including: causing more severe disease, having a greater impact on healthcare systems, or evidence of reduced vaccine efficacy against it [343].

Some of the most notable variants of SARS-CoV-2 include Alpha, Beta, Gamma, Delta and Omicron [91, 92]. Each variant and its associated sub-variants posed unique challenges, for example, Alpha, Beta and Gamma each showed evidence of greater transmissibility than the wild-type virus or other variants circulating at the time and were more likely to cause severe disease [344–347]. Delta was more transmissible than previous variants and appeared to increase the risk of hospitalisation in the

unvaccinated [348–350] and Omicron spread even more easily than Delta, showing increased ability to evade immunity, but is thought to have caused less severe disease [305, 351]. Given that specific interventions may be more suited to certain variants, characterising variants as quickly as possible was a key priority for policymakers.

In Chapter 4, EpiEstim was extended to estimate the transmission advantage of new pathogen variants in real-time, using only the reported incidence data for each variant alongside the estimated serial interval (which can be the same or differ by variant). Previous approaches to estimate the transmission advantage of variants, such as those using phylodynamic models, relied on whole-genome sequencing data, which introduced delays into the analytical pipeline. However, when specific biomarkers (e.g., the S-gene for SARS-CoV-2) are sufficient to distinguish between variants, EpiEstim can rely on routinely collected incidence data, eliminating the need to wait for whole-genome sequencing.

Retrospective analysis of COVID-19 data demonstrated that the method produces comparable estimates to those generated by more complex tools. Additionally, EpiEstim allows users to analyse incidence data for multiple pathogen variants across multiple locations simultaneously, taking less than a few minutes to run. This rapid assessment of differences in transmission potential is crucial in the early stages of variant emergence to directly inform policy decisions. It is also simpler, more accessible and better suited to low-resource settings compared to phylodynamic models.

Following this development, EpiEstim demonstrated its ability to provide actionable insights to policymakers by being used to estimate the transmission advantage of the Delta variant in the UK government response to COVID-19 [241].

While this extension addresses an important policy question, there is still an urgent need to address other recurring policy priorities. For instance, estimating the risk or extent of resurgence when non-pharmaceutical interventions are relaxed, understanding the impact of easing restrictions in one area on neighbouring locations (as mentioned in Section 7.1), or extending the framework for logistical planning in hospitals. Given the impact of inadequate resource planning on patient outcomes [352], this should be prioritised for future research. For example, historical trends in hospital admission data, combined with estimates of length of stay and admission forecasts, could be used to determine whether bed capacity is likely to be exceeded. To ensure that outbreak analytics tools align with evolving policy needs across different phases of an outbreak, it is necessary to consult with policymakers so that their insights can guide the development of methods and tools.

7.4 Increased functionality should not be at the expense of usability

During outbreak response, methods and software tools must balance functionality (i.e., features and capabilities) with usability to ensure that they are accessible enough to be widely and properly applied by a range of users in real-time. If tools are too complex, poorly documented, or slow to execute, this

will hinder their reusability and uptake by the research community.

Open-source code is a critical aspect of open science, which is a movement to ensure research integrity through accessibility, reusability and transparency in research [353]. In Chapter 3, it was disappointing to discover that, out of the 295 EVD models extracted, only 13% (n = 37) referenced open-source code. Upon verification, code was accessible for just 8% (n = 23) of the models. Of these, only 14 were accompanied by instructions or a README, highlighting that very few were reusable. The absence of open-source, well-documented code hinders collaboration and increases inefficiencies in outbreak response. For instance, lack of knowledge sharing can lead to duplicated efforts, wasting valuable time in the initial response to outbreaks. In the first three months of the COVID-19 pandemic, only 15% of COVID-19-related preprints on bioRxiv and medRxiv provided raw data or code to reproduce the analyses [354].

One of the best ways to facilitate efficiency and reusability in analyses is by packaging code into well-documented, open-source packages for widely used software. For example, open-source R packages, distributed in public repositories such as The Comprehensive R Archive Network (CRAN) or GitHub [84–86]. An evaluation of the R packages identified in Chapter 2 revealed that there are issues with ease of package installation, speed and the availability of adequate documentation or tutorials (Table A.2).

These principles of usability were kept in mind during the development of EpiEstim extensions. The EM algorithm used in Chapter 5 and Chapter 6 was intentionally designed to prioritise speed by avoiding computationally intensive stochastic reconstructions of incidence data. Instead, the method provides a single deterministic reconstruction that is both fast and robust, meeting the demands of real-time application. The multi-variant extension is computationally efficient given that it leverages Gibbs sampling, which enables a partially analytical solution to the estimation process (Chapter 4). Therefore, it takes only a few minutes to run on a standard laptop with average specifications.

As part of both projects, vignettes were created to provide tutorials and describe how to use the tools alone and also in a full analytical pipeline [355, 356]. In light of the findings from Chapter 2, the main EpiEstim vignette was also updated to include more context, examples and FAQs [357].

Of course, it is equally as important that outbreak analytics tools are scientifically and statistically robust to reliably support evidence-based decision making. The adequate documentation and transparency of tools ensures that the end user can understand the methods implemented and its limitations, and verify that they are methodologically sound and appropriate for their specific dataset.

While there are some excellent open-access resources that showcase various analysis methods for epidemiologists [358], there remains a need for a centralised resource to help identify the most appropriate tools for different settings and contexts. To assist end users in selecting alternative software tools that may be better suited to their specific data or analytical needs for estimating R_t, Table 2.2 was made available online, and is intended to be a dynamic resource [359]. Expanding on such resources to include example workflows showing the integration of appropriate R packages within outbreak analytics

pipelines in different contexts would further enhance accessibility.

7.5 Further Limitations & Future Research

This thesis has primarily centred around transmissibility estimation using branching process models, which represents just one aspect of the vast field of outbreak analytics. There is also a focus on packaged code in the R programming language. While R is one of the most commonly used programming languages for outbreak epidemiology, outbreak tools have been developed in other languages, such as Python [360, 361]. Many of the challenges identified, such as issues with incidence data and parameterisation, also impact forecasting, scenario modelling, analysis using more complex models (e.g., compartmental and individual based models) and tools implemented in other programming languages. In the future, a broader focus would ensure that a wider range of tools can meet the demands of future outbreaks.

Moreover, the original scoping review (Chapter 2) was conducted at the end of 2020 and the landscape of tools evolves quickly. As discussed, it is important to continue to evaluate the common issues encountered by users of new and current outbreak analytics tools and address them. Future research should build on the findings highlighted in Chapter 2, including the need to account for spatial factors in transmissibility (Section 7.1) and to develop tools that directly answer more policy-relevant questions (Section 7.3). To achieve this effectively, collaboration among interdisciplinary teams of researchers, software developers, and, crucially, end users and policy/decision makers is needed to ensure that tools are both scientifically robust and practically applicable.

This thesis mainly focuses on diseases that are transmitted through direct contact or respiratory transmission pathways, including COVID-19, EVD and influenza. It would be interesting to explore tools used to respond to diseases with different transmission pathways. For example, the ongoing global mpox outbreak is most commonly linked to sexual encounters [362]. Additionally, climate change is anticipated to expand the range of vector-borne diseases, like malaria and dengue [363]. The presence of a vector introduces more factors to consider, including environmental variables like climate, which can influence vector behaviour, survival and ability to transmit pathogens. Moreover, interventions for vector-borne diseases target both humans and the vector itself. In response to malaria, insecticide-treated bed nets are deployed to reduce mosquito populations, while mass drug administration programmes target the parasite in human populations. Therefore, responses to these diseases may require integrating data from a wider variety of sources, including environmental and entomological surveillance, which would pose their own challenges.

Future research should also consider the barriers to the adoption of tools in low-resource settings, where there may be less technical infrastructure or computational capacity. Building capacity and fostering autonomy within these settings is crucial, as local teams are best placed to understand the

unique social, cultural and logistical contexts of an outbreak. In light of recent substantial cuts to foreign aid, and the instability of global health funding more generally, strengthening local autonomy and resilience should be considered a high priority [364]. Building on successful platforms such as the Global Health Training Centre [365] and the Epidemiologist R Handbook [358], which support equitable access to global health training, can help strengthen the foundation for more locally driven outbreak analytics. Leveraging these established platforms and their communities can help foster sustained engagement and collaborations which could enable us to gain insights into the strengths and limitations of current analytical tools. Adapting outbreak tools to meet the specific needs of these settings, along with the exchange of expertise, data and technical support between low and high resource settings, is necessary for creating sustainable and effective outbreak response networks globally.

7.6 Conclusion

This thesis has advanced outbreak preparedness by improving analytical tools, methods and resources for real-time outbreak response. Through the development of new methodologies, such as the extension of EpiEstim to estimate the transmission advantage of emerging variants, the EM algorithm to reconstruct daily incidence data, and the open-source database capable of rapidly parameterising models for EVD, this thesis addresses critical challenges encountered in the field of outbreak analytics. By combining statistical robustness, functionality and usability, these tools can be effectively applied across diverse outbreak contexts.

Greater emphasis should be placed on the wider benefit of the tools we create. Hypothetically, one could have a perfect model or tool, but if it is not adequately documented, its utility is limited to the research group or the individual who developed it. Strengthening the capacity for outbreak response, particularly in low-resource settings, should not only focus on building robust surveillance and healthcare systems, but also knowledge and resource sharing for outbreak analytics. Continuous innovation and adaptation is essential to address the evolving challenges posed by diverse pathogens and outbreak scenarios, ensuring preparedness for future pandemics.

Bibliography

[1] WHO. *WHO EMRO | Infectious Diseases | Health Topics*. World Health Organization - Regional Office for the Eastern Mediterranean. URL: http://www.emro.who.int/health-topics/infectious-diseases/index.html (visited on 11/13/2024).

[2] WHO. *Global Health Estimates: Leading Causes of Death*. URL: https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates/ghe-leading-causes-of-death (visited on 11/13/2024).

[3] WHO. *WHO Fact Sheet: The Top 10 Causes of Death*. URL: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death (visited on 11/13/2024).

[4] Jeremy Hawker et al. *Communicable Disease Control and Health Protection Handbook*. wiley, Jan. 1, 2018. ISBN: 978-1-119-32779-0. DOI: 10.1002/9781119328070. URL: http://www.scopus.com/inward/record.url?scp=85173852737&partnerID=8YFLogxK (visited on 11/10/2024).

[5] WHO. *Ebola Outbreak 2014-2016 - West Africa*. URL: https://www.who.int/emergencies/situations/ebola-outbreak-2014-2016-West-Africa (visited on 11/19/2024).

[6] CDC. *Malaria | CDC Yellow Book 2024*. URL: https://wwwnc.cdc.gov/travel/yellowbook/2024/infections-diseases/malaria (visited on 11/13/2024).

[7] *Tuberculosis (TB) Fact Sheet*. URL: https://www.who.int/news-room/fact-sheets/detail/tuberculosis (visited on 11/13/2024).

[8] *Influenza (Seasonal)*. URL: https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal) (visited on 11/14/2024).

[9] Rachel E. Baker et al. "Infectious Disease in an Era of Global Change". In: *Nature Reviews Microbiology* 20.4 (Apr. 2022), pp. 193–205. ISSN: 1740-1534. DOI: 10.1038/s41579-021-00639-z. URL: https://www.nature.com/articles/s41579-021-00639-z (visited on 11/05/2024).

[10] David N. Fisman. "Seasonality of Infectious Diseases". In: *Annual Review of Public Health* 28 (2007), pp. 127–143. ISSN: 0163-7525. DOI: 10.1146/annurev.publhealth.28.021406.144128. pmid: 17222079.

[11] Walter R. Dowdle. *The Principles of Disease Elimination and Eradication.* URL: https://www.cdc.gov/mmwr/preview/mmwrhtml/su48a7.htm (visited on 11/13/2024).

[12] Anna A. Minta. "Progress Toward Measles Elimination — Worldwide, 2000–2022". In: *MMWR. Morbidity and Mortality Weekly Report* 72 (2023). ISSN: 0149-21951545-861X. DOI: 10.15585/mmwr.mm7246a3. URL: https://www.cdc.gov/mmwr/volumes/72/wr/mm7246a3.htm (visited on 11/14/2024).

[13] CDC. *Global Measles Outbreaks.* Global Measles Vaccination. Nov. 8, 2024. URL: https://www.cdc.gov/global-measles-vaccination/data-research/global-measles-outbreaks/index.html (visited on 11/13/2024).

[14] Petra Klepac et al. "Six Challenges in the Eradication of Infectious Diseases". In: *Epidemics* 10 (Dec. 12, 2014), p. 97. DOI: 10.1016/j.epidem.2014.12.001. pmid: 25843393. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC7612385/ (visited on 11/14/2024).

[15] *Ebola Virus Disease.* URL: https://www.who.int/health-topics/ebola (visited on 11/12/2024).

[16] WHO. *Smallpox.* URL: https://www.who.int/health-topics/smallpox (visited on 11/14/2024).

[17] WHO. *History of Smallpox Vaccination.* URL: https://www.who.int/news-room/spotlight/history-of-vaccination/history-of-smallpox-vaccination (visited on 11/14/2024).

[18] *Outbreak Toolkit.* URL: https://www.who.int/emergencies/outbreak-toolkit (visited on 11/05/2024).

[19] *Toolkit for Investigation and Response to Food and Waterborne Disease Outbreaks with an EU Dimension.* URL: https://www.ecdc.europa.eu/en/publications-data/toolkit-investigation-and-response-food-and-waterborne-disease-outbreaks-eu (visited on 11/05/2024).

[20] *Communicable Disease Outbreak Management: Operational Guidance.* GOV.UK. URL: https://www.gov.uk/government/publications/communicable-disease-outbreak-management-operational-guidance/communicable-disease-outbreak-management-operational-guidance (visited on 11/05/2024).

[21] *Epidemic Curves – Outbreak Toolkit.* URL: https://outbreaktools.ca/background/epidemic-curves/ (visited on 11/15/2024).

[22] Bruce Alberts et al. "Introduction to Pathogens". In: *Molecular Biology of the Cell. 4th Edition.* Garland Science, 2002. URL: https://www.ncbi.nlm.nih.gov/books/NBK26917/ (visited on 11/05/2024).

[23] CDC. *About Prion Diseases*. Prion Diseases. Apr. 22, 2024. URL: https://www.cdc.gov/prions/about/index.html (visited on 11/05/2024).

[24] Leon Gordis. *Epidemiology E-Book*. Elsevier Health Sciences, Nov. 14, 2013. 418 pp. ISBN: 978-1-4557-4251-6. Google Books: 7YX6AQAAQBAJ.

[25] *Coronavirus Disease (COVID-19): How Is It Transmitted?* URL: https://www.who.int/news-room/questions-and-answers/item/coronavirus-disease-covid-19-how-is-it-transmitted (visited on 11/05/2024).

[26] *WHO Fact Sheet: Ebola Virus Disease*. URL: https://www.who.int/news-room/fact-sheets/detail/ebola-virus-disease (visited on 11/11/2024).

[27] WHO. *Zika Virus*. URL: https://www.who.int/news-room/fact-sheets/detail/zika-virus (visited on 11/20/2024).

[28] *Describing Epidemiologic Data | Epidemic Intelligence Service | CDC*. Sept. 24, 2019. URL: https://www.cdc.gov/eis/field-epi-manual/chapters/Describing-Epi-Data.html (visited on 11/05/2024).

[29] *Case Definitions – Outbreak Toolkit*. URL: https://outbreaktools.ca/background/case-definitions/ (visited on 11/12/2024).

[30] *WHO COVID-19 Case Definition*. URL: https://www.who.int/publications/i/item/WHO-2019-nCoV-Surveillance_Case_Definition-2022.1 (visited on 11/12/2024).

[31] Michael B. Gregg. *Field Epidemiology*. Oxford University Press, 2008. 593 pp. ISBN: 978-0-19-531380-2. Google Books: 67y0PBZ9LIMC.

[32] *Conducting a Field Investigation | Epidemic Intelligence Service | CDC*. Aug. 17, 2023. URL: https://www.cdc.gov/eis/field-epi-manual/chapters/Field-Investigation.html (visited on 11/12/2024).

[33] World Health Organization Regional Office for the Eastern Mediterranean. *Epidemiological Data Analysis for the Early Warning Alert and Response Network [EWARN] in Humanitarian Emergencies : A Quick Reference Handbook*. World Health Organization. Regional Office for the Eastern Mediterranean, 2019. ISBN: 978-92-9022-252-1. URL: https://iris.who.int/handle/10665/325830 (visited on 11/12/2024).

[34] *Trend Analysis Guidance for Surveillance Data*. URL: https://www.ecdc.europa.eu/en/publications-data/trend-analysis-guidance-surveillance-data (visited on 11/12/2024).

[35] Jonathan A. Polonsky et al. "Outbreak Analytics: A Developing Data Science for Informing the Response to Emerging Pathogens". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 374.1776 (May 20, 2019), p. 20180276. DOI: 10.1098/rstb.2018.0276. URL: https://royalsocietypublishing.org/doi/10.1098/rstb.2018.0276 (visited on 11/10/2024).

[36] *Outbreak Surveillance and Response in Humanitarian Emergencies: WHO Guidelines for EWARN Implementation*. URL: https://www.who.int/publications/i/item/outbreak-surveillance-and-response-in-humanitarian-emergencies-who-guidelines-for-ewarn-implementation (visited on 11/12/2024).

[37] *COVID-19 Cases | WHO COVID-19 Dashboard*. datadot. URL: https://data.who.int/dashboards/covid19/cases (visited on 11/12/2024).

[38] *How ECDC Collects and Processes COVID-19 Data*. URL: https://www.ecdc.europa.eu/en/covid-19/data-collection (visited on 11/12/2024).

[39] Nina B. Masters et al. "Fine-Scale Spatial Clustering of Measles Nonvaccination That Increases Outbreak Potential Is Obscured by Aggregated Reporting Data". In: *Proceedings of the National Academy of Sciences* 117.45 (Nov. 10, 2020), pp. 28506–28514. DOI: 10.1073/pnas.2011529117. URL: https://www.pnas.org/doi/full/10.1073/pnas.2011529117 (visited on 11/12/2024).

[40] Tania M. Alarcon Falconi et al. "Effects of Data Aggregation on Time Series Analysis of Seasonal Infections". In: *International Journal of Environmental Research and Public Health* 17.16 (Aug. 13, 2020), p. 5887. DOI: 10.3390/ijerph17165887. pmid: 32823719. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC7460497/ (visited on 11/12/2024).

[41] Cheryl L. Gibbons et al. "Measuring Underreporting and Under-Ascertainment in Infectious Disease Datasets: A Comparison of Methods". In: *BMC Public Health* 14.1 (Feb. 11, 2014), p. 147. ISSN: 1471-2458. DOI: 10.1186/1471-2458-14-147. URL: https://doi.org/10.1186/1471-2458-14-147 (visited on 11/05/2024).

[42] Nancy H. L. Leung. "Transmissibility and Transmission of Respiratory Viruses". In: *Nature Reviews Microbiology* 19.8 (Aug. 2021), pp. 528–545. ISSN: 1740-1534. DOI: 10.1038/s41579-021-00535-6. URL: https://www.nature.com/articles/s41579-021-00535-6 (visited on 11/09/2024).

[43] Kenrad E. Nelson and Carolyn Masters Williams. *Infectious Disease Epidemiology: Theory and Practice*. Jones & Bartlett Publishers, 2014. 984 pp. ISBN: 978-0-7637-9533-7. Google Books: SnYGIlHpix0C.

[44] *Marburg Virus Disease*. URL: https://www.who.int/news-room/fact-sheets/detail/marburg-virus-disease (visited on 11/12/2024).

[45] Viviana Simon, David D. Ho, and Quarraisha Abdool Karim. "HIV/AIDS Epidemiology, Pathogenesis, Prevention, and Treatment". In: *Lancet* 368.9534 (Aug. 5, 2006), p. 489. DOI: 10.1016/S0140-6736(06)69157-5. pmid: 16890836. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC2913538/ (visited on 11/05/2024).

[46] Ingrid Chen et al. ""Asymptomatic" Malaria: A Chronic and Debilitating Infection That Should Be Treated". In: *PLOS Medicine* 13.1 (Jan. 19, 2016), e1001942. ISSN: 1549-1676. DOI: 10.1371/journal.pmed.1001942. URL: https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001942 (visited on 11/05/2024).

[47] Quirine A. Ten Bosch et al. "Contributions from the Silent Majority Dominate Dengue Virus Transmission". In: *PLoS pathogens* 14.5 (May 2018), e1006965. ISSN: 1553-7374. DOI: 10.1371/journal.ppat.1006965. pmid: 29723307.

[48] Jon C Emery et al. "The Contribution of Asymptomatic SARS-CoV-2 Infections to Transmission on the Diamond Princess Cruise Ship". In: *eLife* 9 (Aug. 24, 2020). Ed. by Marc Lipsitch, Eduardo Franco, and Marc Lipsitch, e58699. ISSN: 2050-084X. DOI: 10.7554/eLife.58699. URL: https://doi.org/10.7554/eLife.58699 (visited on 11/05/2024).

[49] US Department of Health and Human Services. *What Is the Difference between Isolation and Quarantine? | HHS.Gov*. URL: https://www.hhs.gov/answers/public-health-and-safety/what-is-the-difference-between-isolation-and-quarantine/index.html (visited on 11/19/2024).

[50] T. Déirdre Hollingsworth, Roy M. Anderson, and Christophe Fraser. "HIV-1 Transmission, by Stage of Infection". In: *The Journal of Infectious Diseases* 198.5 (Sept. 1, 2008), pp. 687–693. ISSN: 0022-1899. DOI: 10.1086/590501. URL: https://doi.org/10.1086/590501 (visited on 11/09/2024).

[51] WHO. *Tuberculosis*. URL: https://www.who.int/news-room/questions-and-answers/item/tuberculosis (visited on 11/09/2024).

[52] Paul E. M. Fine. "The Interval between Successive Cases of an Infectious Disease". In: *American Journal of Epidemiology* 158.11 (Dec. 1, 2003), pp. 1039–1047. ISSN: 0002-9262. DOI: 10.1093/aje/kwg251. URL: https://doi.org/10.1093/aje/kwg251 (visited on 11/09/2024).

[53] Fabrice Carrat et al. "Time Lines of Infection and Disease in Human Influenza: A Review of Volunteer Challenge Studies". In: *American Journal of Epidemiology* 167.7 (Apr. 1, 2008), pp. 775–785. ISSN: 0002-9262. DOI: 10.1093/aje/kwm375. URL: https://doi.org/10.1093/aje/kwm375 (visited on 11/05/2024).

[54] Dongxuan Chen et al. "Inferring Time-Varying Generation Time, Serial Interval, and Incubation Period Distributions for COVID-19". In: *Nature Communications* 13.1 (Dec. 13, 2022), p. 7727. ISSN: 2041-1723. DOI: 10.1038/s41467-022-35496-8. URL: https://www.nature.com/articles/s41467-022-35496-8 (visited on 11/05/2024).

[55] Gianpaolo Scalia Tomba et al. "Some Model Based Considerations on Observing Generation Times for Communicable Diseases". In: *Mathematical Biosciences* 223.1 (Jan. 1, 2010), pp. 24–31. ISSN: 0025-5564. DOI: 10.1016/j.mbs.2009.10.004. URL: https://www.sciencedirect.com/science/article/pii/S0025556409001679 (visited on 11/08/2024).

[56] Åke Svensson. "A Note on Generation Times in Epidemic Models". In: *Mathematical Biosciences* 208.1 (July 1, 2007), pp. 300–311. ISSN: 0025-5564. DOI: 10.1016/j.mbs.2006.10.010. URL: https://www.sciencedirect.com/science/article/pii/S0025556406002094 (visited on 11/13/2024).

[57] J. Wallinga and M. Lipsitch. "How Generation Intervals Shape the Relationship between Growth Rates and Reproductive Numbers". In: *Proceedings of the Royal Society B: Biological Sciences* 274.1609 (Nov. 28, 2006), p. 599. DOI: 10.1098/rspb.2006.3754. pmid: 17476782. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC1766383/ (visited on 11/09/2024).

[58] Junling Ma. "Estimating Epidemic Exponential Growth Rate and Basic Reproduction Number". In: *Infectious Disease Modelling* 5 (Jan. 1, 2020), pp. 129–141. ISSN: 2468-0427. DOI: 10.1016/j.idm.2019.12.009. URL: https://www.sciencedirect.com/science/article/pii/S2468042719300491 (visited on 11/14/2024).

[59] William Ogilvy Kermack, A. G. McKendrick, and Gilbert Thomas Walker. "A Contribution to the Mathematical Theory of Epidemics". In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 115.772 (Jan. 1997), pp. 700–721. DOI: 10.1098/rspa.1927.0118. URL: https://royalsocietypublishing.org/doi/10.1098/rspa.1927.0118 (visited on 11/09/2024).

[60] Anne Cori et al. "A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics". In: *American Journal of Epidemiology* 178.9 (Nov. 1, 2013), pp. 1505–1512. ISSN: 0002-9262. DOI: 10.1093/aje/kwt133. URL: https://academic.oup.com/aje/article/178/9/1505/89262 (visited on 11/20/2020).

[61] Jacco Wallinga and Peter Teunis. "Different Epidemic Curves for Severe Acute Respiratory Syndrome Reveal Similar Impacts of Control Measures". In: *American Journal of Epidemiology* 160.6 (Sept. 15, 2004), p. 509. DOI: 10.1093/aje/kwh255. pmid: 15353409. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC7110200/ (visited on 11/09/2024).

[62] Anne Cori et al. “A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics”. In: *American Journal of Epidemiology* 178.9 (Nov. 1, 2013), pp. 1505–1512. ISSN: 0002-9262. DOI: 10.1093/aje/kwt133. URL: https://academic.oup.com/aje/article/178/9/1505/89262 (visited on 11/20/2020).

[63] Christophe Fraser. “Estimating Individual and Household Reproduction Numbers in an Emerging Epidemic”. In: *PLOS ONE* 2.8 (Aug. 22, 2007), e758. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0000758. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0000758 (visited on 11/20/2020).

[64] Anne Cori and Adam Kucharski. “Inference of Epidemic Dynamics in the COVID-19 Era and Beyond”. In: *Epidemics* 48 (Sept. 1, 2024), p. 100784. ISSN: 1755-4365. DOI: 10.1016/j.epidem.2024.100784. URL: https://www.sciencedirect.com/science/article/pii/S1755436524000458 (visited on 11/20/2024).

[65] WHO Ebola Response Team. “Ebola Virus Disease in West Africa — The First 9 Months of the Epidemic and Forward Projections”. In: *New England Journal of Medicine* 371.16 (Oct. 16, 2014), pp. 1481–1495. ISSN: 0028-4793. DOI: 10.1056/NEJMoa1411100. URL: https://www.nejm.org/doi/full/10.1056/NEJMoa1411100 (visited on 09/30/2024).

[66] WHO Ebola Response Team. “West African Ebola Epidemic after One Year — Slowing but Not Yet under Control”. In: *New England Journal of Medicine* 372.6 (Feb. 5, 2015), pp. 584–587. ISSN: 0028-4793. DOI: 10.1056/NEJMc1414992. URL: https://www.nejm.org/doi/full/10.1056/NEJMc1414992 (visited on 09/30/2024).

[67] The Ebola Outbreak Epidemiology Team et al. “Outbreak of Ebola Virus Disease in the Democratic Republic of the Congo, April–May, 2018: An Epidemiological Study”. In: *The Lancet* 392.10143 (July 21, 2018), pp. 213–221. ISSN: 0140-6736, 1474-547X. DOI: 10.1016/S0140-6736(18)31387-4. pmid: 30047375. URL: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(18)31387-4/abstract (visited on 06/11/2021).

[68] Flavio Finger et al. “Real-Time Analysis of the Diphtheria Outbreak in Forcibly Displaced Myanmar Nationals in Bangladesh”. In: *BMC Medicine* 17.1 (Mar. 12, 2019), p. 58. ISSN: 1741-7015. DOI: 10.1186/s12916-019-1288-7. URL: https://doi.org/10.1186/s12916-019-1288-7 (visited on 06/11/2021).

[69] Nicholas G. Reich et al. “Collaborative Hubs: Making the Most of Predictive Epidemic Modeling”. In: *American Journal of Public Health* 112.6 (June 2022), pp. 839–842. ISSN: 1541-0048. DOI: 10.2105/AJPH.2022.306831. pmid: 35420897.

[70] Harrison Manley et al. “Combining Models to Generate Consensus Medium-Term Projections of Hospital Admissions, Occupancy and Deaths Relating to COVID-19 in England”. In: *Royal*

Society Open Science 11.5 (May 22, 2024), p. 231832. DOI: 10.1098/rsos.231832. URL: https://royalsocietypublishing.org/doi/10.1098/rsos.231832 (visited on 11/15/2024).

[71] Estee Y. Cramer et al. "The United States COVID-19 Forecast Hub Dataset". In: *Scientific Data* 9.1 (Aug. 1, 2022), p. 462. ISSN: 2052-4463. DOI: 10.1038/s41597-022-01517-w. URL: https://www.nature.com/articles/s41597-022-01517-w (visited on 11/15/2024).

[72] RespiCast - ECDC Forecasting Hub. *RespiCast | ECDC Forecasting Hub*. URL: https://respicast.ecdc.europa.eu/ (visited on 11/15/2024).

[73] Evan L. Ray and Nicholas G. Reich. "Prediction of Infectious Disease Epidemics via Weighted Density Ensembles". In: *PLOS Computational Biology* 14.2 (Feb. 20, 2018), e1005910. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1005910. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005910 (visited on 11/18/2024).

[74] Pierre Nouvellet et al. "A Simple Approach to Measure Transmissibility and Forecast Incidence". In: *Epidemics* 22 (Mar. 2018), pp. 29–35. ISSN: 1878-0067. DOI: 10.1016/j.epidem.2017.02.012. pmid: 28351674.

[75] Katelyn M. Gostic et al. "Practical Considerations for Measuring the Effective Reproductive Number, Rt". In: *medRxiv* (Aug. 28, 2020), p. 2020.06.18.20134858. DOI: 10.1101/2020.06.18.20134858. pmid: 32607522. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC7325187/ (visited on 11/18/2024).

[76] Roy M. Anderson and Robert M. May. *Infectious Diseases of Humans: Dynamics and Control*. Oxford, New York: Oxford University Press, Aug. 27, 1992. 766 pp. ISBN: 978-0-19-854040-3.

[77] Herbert W. Hethcote. "The Mathematics of Infectious Diseases". In: *SIAM Review* 42.4 (Jan. 1, 2000), pp. 599–653. ISSN: 0036-1445. DOI: 10.1137/S0036144500371907. URL: https://epubs.siam.org/doi/abs/10.1137/s0036144500371907 (visited on 06/04/2021).

[78] CDC. *Behind the Model: How Scenario Modeling Can Inform Public Health Decision-Making*. CFA: Behind the Model. Aug. 29, 2024. URL: https://www.cdc.gov/cfa-behind-the-model/php/data-research/scenario-modeling/index.html (visited on 11/18/2024).

[79] UKHSA. *UKHSA Advisory Board: Preparedness for Infectious Disease Threats*. GOV.UK. URL: https://www.gov.uk/government/publications/ukhsa-board-meeting-papers-january-2023/ukhsa-advisory-board-preparedness-for-infectious-disease-threats (visited on 11/15/2024).

[80] WHO. *Preparedness and Resilience for Emerging Threats (PRET)*. URL: https://www.who.int/initiatives/preparedness-and-resilience-for-emerging-threats (visited on 11/15/2024).

[81] WHO. *Preparing for Pandemics*. URL: https://www.who.int/westernpacific/activities/preparing-for-pandemics (visited on 11/15/2024).

[82] *Report of the Review Committee on the Functioning of the International Health Regulations (2005) during the COVID-19 Response*. URL: https://www.who.int/publications/m/item/a74-9-who-s-work-in-health-emergencies (visited on 11/05/2024).

[83] Nita Madhav et al. "Pandemics: Risks, Impacts, and Mitigation". In: *Disease Control Priorities: Improving Health and Reducing Poverty*. Ed. by Dean T. Jamison et al. 3rd. Washington (DC): The International Bank for Reconstruction and Development / The World Bank, 2017. ISBN: 978-1-4648-0527-1 978-1-4648-0528-8. pmid: 30212163. URL: http://www.ncbi.nlm.nih.gov/books/NBK525302/ (visited on 11/05/2024).

[84] *R: The R Project for Statistical Computing*. URL: https://www.r-project.org/ (visited on 11/10/2024).

[85] *The Comprehensive R Archive Network*. URL: https://cran.r-project.org/ (visited on 11/10/2024).

[86] *Build Software Better, Together*. GitHub. URL: https://github.com (visited on 11/10/2024).

[87] *The Epidemiologist R Handbook*. URL: https://www.epirhandbook.com/en/ (visited on 11/10/2024).

[88] Zeinab Abdelrahman, Mengyuan Li, and Xiaosheng Wang. "Comparative Review of SARS-CoV-2, SARS-CoV, MERS-CoV, and Influenza A Respiratory Viruses". In: *Frontiers in Immunology* 11 (Sept. 11, 2020). ISSN: 1664-3224. DOI: 10.3389/fimmu.2020.552909. URL: https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2020.552909/full (visited on 12/04/2024).

[89] Natsuko Imai et al. *Report 1 - Estimating the Potential Total Number of Novel Coronavirus (2019-nCoV) Cases in Wuhan City, China*. Imperial College London. URL: https://www.imperial.ac.uk/medicine/departments/school-public-health/infectious-disease-epidemiology/mrc-global-infectious-disease-analysis/disease-areas/covid-19/report-1-case-estimates-of-covid-19/ (visited on 12/04/2024).

[90] William T. Harvey et al. "SARS-CoV-2 Variants, Spike Mutations and Immune Escape". In: *Nature Reviews Microbiology* 19.7 (July 2021), pp. 409–424. ISSN: 1740-1534. DOI: 10.1038/s41579-021-00573-0. URL: https://www.nature.com/articles/s41579-021-00573-0 (visited on 11/21/2024).

[91] WHO. *COVID-19 Variants | WHO COVID-19 Dashboard*. datadot. URL: https://data.who.int/dashboards/covid19/variants (visited on 11/21/2024).

[92] Mickensone Andre et al. “From Alpha to Omicron: How Different Variants of Concern of the SARS-Coronavirus-2 Impacted the World”. In: *Biology* 12.9 (9 Sept. 2023), p. 1267. ISSN: 2079-7737. DOI: 10.3390/biology12091267. URL: https://www.mdpi.com/2079-7737/12/9/1267 (visited on 11/21/2024).

[93] Erik Volz et al. “Assessing Transmissibility of SARS-CoV-2 Lineage B. 1.1. 7 in England”. In: *Nature* 593.7858 (2021), pp. 266–269. DOI: 10.1038/s41586-021-03470-x.

[94] Nicholas G. Davies et al. “Estimated Transmissibility and Impact of SARS-CoV-2 Lineage B.1.1.7 in England”. In: *Science* 372.6538 (Apr. 2021). DOI: 10.1126/science.abg3055.

[95] Kathy Leung et al. “Early Transmissibility Assessment of the N501Y Mutant Strains of SARS-CoV-2 in the United Kingdom, October to November 2020”. In: *Eurosurveillance* 26.1 (Jan. 2021). DOI: 10.2807/1560-7917.ES.2020.26.1.2002106.

[96] Wan Yang and Jeffrey Shaman. “Development of a Model-Inference System for Estimating Epidemiological Characteristics of SARS-CoV-2 Variants of Concern”. In: *Nature Communications* 12.1 (Dec. 2021), p. 5573. DOI: 10.1038/s41467-021-25913-9.

[97] Shi Zhao et al. “The Co-Circulating Transmission Dynamics of SARS-CoV-2 Alpha and Eta Variants in Nigeria: A Retrospective Modeling Study of COVID-19”. In: *Journal of Global Health* 11 (Dec. 2021), p. 05028. DOI: 10.7189/jogh.11.05028.

[98] Shi Zhao et al. “Quantifying the transmission advantage associated with N501Y substitution of SARS-CoV-2 in the UK: An early data-driven analysis”. In: *Journal of Travel Medicine* 28.2 (Jan. 2021). tex.pmcid: PMC7928809, taab011. DOI: 10.1093/jtm/taab011.

[99] Meera Chand et al. *Investigation of Novel SARS-COV-2 Variant: Variant of Concern 202012/01*. 2021. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/959438/Technical_Briefing_VOC_SH_NJL2_SH2.pdf (visited on 09/01/2021).

[100] Mark S Graham et al. “Changes in Symptomatology, Reinfection, and Transmissibility Associated with the SARS-CoV-2 Variant B.1.1.7: An Ecological Study”. In: *The Lancet Public Health* 6.5 (May 2021), e335–e345. DOI: 10.1016/S2468-2667(21)00055-4.

[101] Chayada Piantham et al. *Estimating the Elevated Transmissibility of the B.1.1.7 Strain over Previously Circulating Strains in England Using GISAID Sequence Frequencies*. medRxiv. Mar. 2021. DOI: 10.1101/2021.03.17.21253775.

[102] Katelyn M. Gostic et al. “Practical considerations for measuring the effective reproductive number, Rt”. In: *PLOS Computational Biology* 16.12 (Dec. 2020), e1008409. DOI: 10.1371/journal.pcbi.1008409.

[103] *Prioritizing Diseases for Research and Development in Emergency Contexts*. URL: https://www.who.int/activities/prioritizing-diseases-for-research-and-development-in-emergency-contexts (visited on 11/08/2024).

[104] Anne Cori et al. *EpiEstim: Estimate Time Varying Reproduction Numbers from Epidemic Curves*. Version 2.2-3. Jan. 7, 2021. URL: https://CRAN.R-project.org/package=EpiEstim (visited on 06/09/2022).

[105] Rebecca K. Nash, Pierre Nouvellet, and Anne Cori. "Real-Time Estimation of the Epidemic Reproduction Number: Scoping Review of the Applications and Challenges". In: *PLOS Digital Health* 1.6 (June 27, 2022), e0000052. ISSN: 2767-3170. DOI: 10.1371/journal.pdig.0000052. URL: https://journals.plos.org/digitalhealth/article?id=10.1371/journal.pdig.0000052 (visited on 06/29/2022).

[106] Christophe Fraser et al. "Pandemic Potential of a Strain of Influenza A (H1N1): Early Findings". In: *Science (New York, N.Y.)* 324.5934 (June 19, 2009), pp. 1557–1561. ISSN: 0036-8075. DOI: 10.1126/science.1176062. pmid: 19433588. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3735127/ (visited on 06/04/2021).

[107] Simon Cauchemez et al. "Middle East Respiratory Syndrome Coronavirus: Quantification of the Extent of the Epidemic, Surveillance Biases, and Transmissibility". In: *The Lancet. Infectious Diseases* 14.1 (Jan. 2014), pp. 50–56. ISSN: 1474-4457. DOI: 10.1016/S1473-3099(13)70304-9. pmid: 24239323.

[108] WHO Ebola Response Team. "Ebola Virus Disease in West Africa — The First 9 Months of the Epidemic and Forward Projections". In: *New England Journal of Medicine* 371.16 (Oct. 16, 2014), pp. 1481–1495. ISSN: 0028-4793. DOI: 10.1056/NEJMoa1411100. pmid: 25244186. URL: https://doi.org/10.1056/NEJMoa1411100 (visited on 06/04/2021).

[109] WHO Ebola Response Team. "West African Ebola Epidemic after One Year — Slowing but Not Yet under Control". In: *The New England journal of medicine* 372.6 (Feb. 5, 2015), pp. 584–587. ISSN: 0028-4793. DOI: 10.1056/NEJMc1414992. pmid: 25539446. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4368109/ (visited on 06/11/2021).

[110] Ahmadou Barry et al. "Outbreak of Ebola Virus Disease in the Democratic Republic of the Congo, April–May, 2018: An Epidemiological Study". In: *The Lancet* 392.10143 (2018), pp. 213–221. ISSN: 0140-6736.

[111] Tom Britton and Gianpaolo Scalia Tomba. "Estimation in Emerging Epidemics: Biases and Remedies". In: *Journal of The Royal Society Interface* 16.150 (Jan. 31, 2019), p. 20180670. DOI: 10.1098/rsif.2018.0670. URL: https://royalsocietypublishing.org/doi/full/10.1098/rsif.2018.0670 (visited on 05/16/2022).

[112] Katelyn M. Gostic et al. "Practical Considerations for Measuring the Effective Reproductive Number, Rt". In: *PLOS Computational Biology* 16.12 (Dec. 10, 2020), e1008409. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1008409. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1008409 (visited on 06/04/2021).

[113] Thibaut Jombart et al. *Projections: Project Future Case Incidence*. Version 0.5.4. Apr. 22, 2021. URL: https://CRAN.R-project.org/package=projections (visited on 06/09/2022).

[114] R. N. Thompson et al. "Improved Inference of Time-Varying Reproduction Numbers during Infectious Disease Outbreaks". In: *Epidemics* 29 (Dec. 1, 2019), p. 100356. ISSN: 1755-4365. DOI: 10.1016/j.epidem.2019.100356. URL: https://www.sciencedirect.com/science/article/pii/S1755436519300350 (visited on 06/04/2021).

[115] MRC Centre for Global Infectious Disease Analysis [@MRC_Outbreak]. *Have You Used EpiEstim? What Do You like? What Would You like to Change? We Are Collecting Feedback & Love Your Input to Improve Package Development @dr_anne_cori @Rebecca_K_Nash @ImperialSPH @imperialcollege Let Us Know What You Think! https://t.co/mr3Esz4A31*. Twitter. Jan. 11, 2021. URL: https://twitter.com/MRC_Outbreak/status/1348586478394339328 (visited on 09/04/2024).

[116] Bimandra A. Djaafara et al. "Quantifying the Dynamics of COVID-19 Burden and Impact of Interventions in Java, Indonesia". In: *medRxiv* (Feb. 15, 2021), p. 2020.10.02.20198663. ISSN: 2019-8663. DOI: 10.1101/2020.10.02.20198663. URL: https://www.medrxiv.org/content/10.1101/2020.10.02.20198663v2 (visited on 06/04/2021).

[117] Bimandra A Djaafara et al. "A Quantitative Framework for Defining the End of an Infectious Disease Outbreak: Application to Ebola Virus Disease". In: *American Journal of Epidemiology* 190.4 (Apr. 6, 2021), pp. 642–651. ISSN: 0002-9262. DOI: 10.1093/aje/kwaa212. URL: https://doi.org/10.1093/aje/kwaa212 (visited on 06/04/2021).

[118] Rachel Esra et al. "Evaluating the Impact of Non-Pharmaceutical Interventions for SARS-CoV-2 on a Global Scale". In: *medRxiv* (Aug. 5, 2020), p. 2020.07.30.20164939. ISSN: 2016-4939. DOI: 10.1101/2020.07.30.20164939. URL: https://www.medrxiv.org/content/10.1101/2020.07.30.20164939v1 (visited on 06/03/2021).

[119] Leandro Pereira Garcia et al. "The Potential Spread of Covid-19 and Government Decision-Making: A Retrospective Analysis in Florianópolis, Brazil". In: *Revista Brasileira de Epidemiologia* 23 (Sept. 30, 2020). ISSN: 1415-790X, 1415-790X, 1980-5497. DOI: 10.1590/1980-549720200091. URL: https://www.scielo.br/j/rbepid/a/WJLGyZfwFkfPGVLMVW5y8ch/?lang=en (visited on 06/04/2021).

[120] Jana S. Huisman et al. "Estimation and Worldwide Monitoring of the Effective Reproductive Number of SARS-CoV-2". In: *medRxiv* (Nov. 30, 2020), p. 2020.11.26.20239368. DOI: 10.1101/2020.11.26.20239368. URL: https://www.medrxiv.org/content/10.1101/2020.11.26.20239368v1 (visited on 06/04/2021).

[121] A. Khosravi et al. "The Basic Reproduction Number and Prediction of the Epidemic Size of the Novel Coronavirus (COVID-19) in Shahroud, Iran". In: *Epidemiology & Infection* 148 (2020). ISSN: 0950-2688, 1469-4409. DOI: 10.1017/S0950268820001247. URL: https://www.cambridge.org/core/journals/epidemiology-and-infection/article/basic-reproduction-number-and-prediction-of-the-epidemic-size-of-the-novel-coronavirus-covid19-in-shahroud-iran/B9ED8A0969C0F36DCC72E4ABB45CCC19 (visited on 06/04/2021).

[122] Chenyu Li, Paola Romagnani, and Hans-Joachim Anders. "Novel Criteria for When and How to Exit a COVID-19 Pandemic Lockdown". In: *Frontiers in Big Data* 3 (2020). ISSN: 2624-909X. DOI: 10.3389/fdata.2020.00026. URL: https://www.frontiersin.org/articles/10.3389/fdata.2020.00026/full (visited on 06/04/2021).

[123] Yige Li, Eduardo A. Undurraga, and José R. Zubizarreta. "Effectiveness of Localized Lockdowns in the COVID-19 Pandemic". In: *medRxiv* (Mar. 30, 2021), p. 2020.08.25.20182071. DOI: 10.1101/2020.08.25.20182071. URL: https://www.medrxiv.org/content/10.1101/2020.08.25.20182071v3 (visited on 06/04/2021).

[124] Kris V. Parag. "Improved Estimation of Time-Varying Reproduction Numbers at Low Case Incidence and between Epidemic Waves". In: *medRxiv* (May 19, 2021), p. 2020.09.14.20194589. DOI: 10.1101/2020.09.14.20194589. URL: https://www.medrxiv.org/content/10.1101/2020.09.14.20194589v3 (visited on 06/04/2021).

[125] David J Price et al. *Estimating the Case Detection Rate and Temporal Variation in Transmission of COVID-19 in Australia: Technical Report 14th April 2020*. Doherty Institute, Apr. 14, 2020, p. 11. URL: https://www.doherty.edu.au/uploads/content_doc/Estimating_changes_in_the_transmission_of_COVID-19_April14-public-release.pdf.

[126] Joaquín Salas. "Improving the Estimation of the COVID-19 Effective Reproduction Number Using Nowcasting". Jan. 25, 2021. arXiv: 2007.09800 [physics, q-bio]. URL: http://arxiv.org/abs/2007.09800 (visited on 06/04/2021).

[127] Katharine Sherratt et al. "Exploring Surveillance Data Biases When Estimating the Reproduction Number: With Insights into Subpopulation Transmission of Covid-19 in England". In: *medRxiv* (Mar. 18, 2021), p. 2020.10.18.20214585. DOI: 10.1101/2020.10.18.20214585. URL: https://www.medrxiv.org/content/10.1101/2020.10.18.20214585v2 (visited on 06/04/2021).

[128] Hugues Turbé et al. “Adaptive Time-Dependent Priors and Bayesian Inference to Evaluate SARS-CoV-2 Public Health Measures Validated on 31 Countries”. In: *Frontiers in Public Health* 8 (2021). ISSN: 2296-2565. DOI: 10.3389/fpubh.2020.583401. URL: https://www.frontiersin.org/articles/10.3389/fpubh.2020.583401/full (visited on 06/04/2021).

[129] Shuo Wang et al. “A Bayesian Updating Scheme for Pandemics: Estimating the Infection Dynamics of COVID-19”. In: *IEEE Computational Intelligence Magazine* 15.4 (Nov. 2020), pp. 23–33. ISSN: 1556-6048. DOI: 10.1109/MCI.2020.3019874.

[130] Bryan Wilder, Michael J. Mina, and Milind Tambe. “Tracking Disease Outbreaks from Sparse Data with Bayesian Inference”. Sept. 12, 2020. arXiv: 2009.05863 [cs, stat]. URL: http://arxiv.org/abs/2009.05863 (visited on 06/04/2021).

[131] Sam Abbott et al. “Estimating the Time-Varying Reproduction Number of SARS-CoV-2 Using National and Subnational Case Counts”. In: *Wellcome Open Research* 5 (Dec. 8, 2020), p. 112. ISSN: 2398-502X. DOI: 10.12688/wellcomeopenres.16006.2. URL: https://wellcomeopenresearch.org/articles/5-112/v2 (visited on 06/02/2021).

[132] Patrice Abry et al. “Spatial and Temporal Regularization to Estimate COVID-19 Reproduction Number R(t): Promoting Piecewise Smoothness via Convex Optimization”. In: *PLOS ONE* 15.8 (Aug. 20, 2020), e0237901. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0237901. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0237901 (visited on 06/04/2021).

[133] Adil Al Wahaibi et al. “COVID-19 Epidemic Monitoring after Non-Pharmaceutical Interventions: The Use of Time-Varying Reproduction Number in a Country with a Large Migrant Population”. In: *International Journal of Infectious Diseases* 99 (Oct. 2020), pp. 466–472. ISSN: 1201-9712. DOI: 10.1016/j.ijid.2020.08.039. pmid: 32829052. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7439014/ (visited on 06/04/2021).

[134] Luis Alvarez et al. “Computing the Daily Reproduction Number of COVID-19 by Inverting the Renewal Equation”. In: *medRxiv* (Apr. 21, 2021), p. 2020.08.01.20165142. DOI: 10.1101/2020.08.01.20165142. URL: https://www.medrxiv.org/content/10.1101/2020.08.01.20165142v5 (visited on 06/04/2021).

[135] Luis Alvarez, Miguel Colom, and Jean-Michel Morel. “Removing Weekly Administrative Noise in the Daily Count of COVID-19 New Cases. Application to the Computation of Rt”. In: *medRxiv* (Nov. 18, 2020), p. 2020.11.16.20232405. DOI: 10.1101/2020.11.16.20232405. URL: https://www.medrxiv.org/content/10.1101/2020.11.16.20232405v1 (visited on 06/04/2021).

[136] Jantien A. Backer and Jacco Wallinga. “Spatiotemporal Analysis of the 2014 Ebola Epidemic in West Africa”. In: *PLOS Computational Biology* 12.12 (Dec. 8, 2016), e1005210. ISSN: 1553-7358.

DOI: 10.1371/journal.pcbi.1005210. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005210 (visited on 06/04/2021).

[137] Samir Bhatt et al. "Semi-Mechanistic Bayesian Modeling of COVID-19 with Renewal Processes". Dec. 29, 2020. arXiv: 2012.00394 [stat]. URL: http://arxiv.org/abs/2012.00394 (visited on 06/04/2021).

[138] Hervé Bourhy et al. "Revealing the Micro-scale Signature of Endemic Zoonotic Disease Transmission in an African Urban Setting". In: *PLOS Pathogens* 12.4 (Apr. 8, 2016), e1005525. ISSN: 1553-7374. DOI: 10.1371/journal.ppat.1005525. URL: https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1005525 (visited on 06/04/2021).

[139] Marcos A. Capistrán, Antonio Capella, and J. Andrés Christen. "Filtering and Improved Uncertainty Quantification in the Dynamic Estimation of Effective Reproduction Numbers". Dec. 3, 2020. arXiv: 2012.02168 [stat]. URL: http://arxiv.org/abs/2012.02168 (visited on 06/04/2021).

[140] James John Chirombo. "Modelling Spatial Processes of Infectious Diseases". Faculty of Health and Medicine, Lancaster University, Sept. 2018. 39-57. URL: https://core.ac.uk/download/pdf/211246237.pdf#page=55.

[141] Benjamin J. Cowling et al. "Impact Assessment of Non-Pharmaceutical Interventions against Coronavirus Disease 2019 and Influenza in Hong Kong: An Observational Study". In: *The Lancet Public Health* 5.5 (May 1, 2020), e279–e288. ISSN: 2468-2667. DOI: 10.1016/S2468-2667(20)30090-6. pmid: 32311320. URL: https://www.thelancet.com/journals/lanpub/article/PIIS2468-2667(20)30090-6/abstract (visited on 06/04/2021).

[142] Andrea De Simone and Marco Piangerelli. "A Bayesian Approach for Monitoring Epidemics in Presence of Undetected Cases". In: *Chaos, Solitons & Fractals* 140 (Nov. 1, 2020), p. 110167. ISSN: 0960-0779. DOI: 10.1016/j.chaos.2020.110167. URL: https://www.sciencedirect.com/science/article/pii/S0960077920305634 (visited on 06/04/2021).

[143] Claire Donnat and Susan Holmes. "Modeling the Heterogeneity in COVID-19's Reproductive Number and Its Impact on Predictive Scenarios". Apr. 14, 2021. arXiv: 2004.05272 [q-bio, stat]. URL: http://arxiv.org/abs/2004.05272 (visited on 06/04/2021).

[144] Charles M. Evans, Joseph R. Egan, and Ian Hall. "Pneumonic Plague in Johannesburg, South Africa, 1904 - Volume 24, Number 1—January 2018 - Emerging Infectious Diseases Journal - CDC". In: *Emerging Infectious Diseases* 24.1 (Jan. 2018). DOI: 10.3201/eid2401.161817. URL: https://wwwnc.cdc.gov/eid/article/24/1/16-1817_article (visited on 06/04/2021).

[145] Neil M. Ferguson et al. "Countering the Zika Epidemic in Latin America". In: *Science* 353.6297 (July 22, 2016), pp. 353–354. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aag0219. pmid: 27417493. URL: https://science.sciencemag.org/content/353/6297/353 (visited on 06/03/2021).

[146] Seth Flaxman et al. "Estimating the Effects of Non-Pharmaceutical Interventions on COVID-19 in Europe". In: *Nature* 584.7820 (7820 Aug. 2020), pp. 257–261. ISSN: 1476-4687. DOI: 10.1038/s41586-020-2405-7. URL: https://www.nature.com/articles/s41586-020-2405-7 (visited on 06/04/2021).

[147] Nick Golding et al. *Estimating Temporal Variation in Transmission of COVID-19 and Adherence to Social Distancing Measures in Australia: Technical Report 5 May 2020*. Technical Report. Doherty Institute, May 5, 2020, p. 19. URL: https://www.doherty.edu.au/uploads/content_doc/Estimating_changes_in_the_transmission_of_COVID-19_May5-public-release1.pdf.

[148] Karthick Kanagarathinam, Ebrahem A. Algehyne, and Kavaskar Sekar. "Analysis of 'earlyR' Epidemic Model and Time Series Model for Prediction of COVID-19 Registered Cases". In: *Materials Today. Proceedings* (Oct. 14, 2020). ISSN: 2214-7853. DOI: 10.1016/j.matpr.2020.10.086. pmid: 33078097. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7556808/ (visited on 06/04/2021).

[149] Karthick Kanagarathinam and Kavaskar Sekar. "Estimation of the Reproduction Number and Early Prediction of the COVID-19 Outbreak in India Using a Statistical Computing Approach". In: *Epidemiology and Health* 42 (May 9, 2020). ISSN: 2092-7193. DOI: 10.4178/epih.e2020028. pmid: 32512670. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7644936/ (visited on 06/04/2021).

[150] Jesse Knight and Sharmistha Mishra. "Estimating Effective Reproduction Number Using Generation Time versus Serial Interval, with Application to Covid-19 in the Greater Toronto Area, Canada". In: *Infectious Disease Modelling* 5 (Jan. 1, 2020), pp. 889–896. ISSN: 2468-0427. DOI: 10.1016/j.idm.2020.10.009. URL: https://www.sciencedirect.com/science/article/pii/S2468042720300634 (visited on 06/04/2021).

[151] You Li et al. "The Temporal Association of Introducing and Lifting Non-Pharmaceutical Interventions with the Time-Varying Reproduction Number (R) of SARS-CoV-2: A Modelling Study across 131 Countries". In: *The Lancet Infectious Diseases* 21.2 (Feb. 1, 2021), pp. 193–202. ISSN: 1473-3099, 1474-4457. DOI: 10.1016/S1473-3099(20)30785-4. URL: https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(20)30785-4/abstract (visited on 06/04/2021).

[152] Jue Tao Lim et al. "Revealing Regional Disparities in the Transmission Potential of SARS-CoV-2 from Interventions in Southeast Asia". In: *Proceedings of the Royal Society B: Biological Sciences* 287.1933 (Aug. 26, 2020), p. 20201173. DOI: 10.1098/rspb.2020.1173. URL: https://royalsocietypublishing.org/doi/10.1098/rspb.2020.1173 (visited on 06/04/2021).

[153] Yang Liu et al. "The Impact of Non-Pharmaceutical Interventions on SARS-CoV-2 Transmission across 130 Countries and Territories". In: *BMC Medicine* 19.1 (Feb. 5, 2021), p. 40. ISSN: 1741-7015. DOI: 10.1186/s12916-020-01872-8. URL: https://doi.org/10.1186/s12916-020-01872-8 (visited on 06/04/2021).

[154] Theodore Lytras et al. "An Improved Method to Estimate the Effective Reproduction Number of the COVID-19 Pandemic: Lessons from Its Application in Greece". In: *medRxiv* (Sept. 22, 2020), p. 2020.09.19.20198028. DOI: 10.1101/2020.09.19.20198028. URL: https://www.medrxiv.org/content/10.1101/2020.09.19.20198028v1 (visited on 06/04/2021).

[155] Miguel A. Martinez-Beneito, Jorge Mateu, and Paloma Botella-Rocamora. "Spatio-Temporal Small Area Surveillance of the Covid-19 Pandemics". Nov. 8, 2020. arXiv: 2011.03938 [stat]. URL: http://arxiv.org/abs/2011.03938 (visited on 06/04/2021).

[156] Swapnil Mishra et al. "On the Derivation of the Renewal Equation from an Age-Dependent Branching Process: An Epidemic Modelling Perspective". June 29, 2020. arXiv: 2006.16487 [q-bio, stat]. URL: http://arxiv.org/abs/2006.16487 (visited on 06/04/2021).

[157] Ta-Chou Ng and Tzai-Hung Wen. "Spatially Adjusted Time-varying Reproductive Numbers: Understanding the Geographical Expansion of Urban Dengue Outbreaks". In: *Scientific Reports* 9.1 (1 Dec. 16, 2019), p. 19172. ISSN: 2045-2322. DOI: 10.1038/s41598-019-55574-0. URL: https://www.nature.com/articles/s41598-019-55574-0 (visited on 06/04/2021).

[158] Pierre Nouvellet et al. "Reduction in Mobility and COVID-19 Transmission". In: *Nature Communications* 12.1 (1 Feb. 17, 2021), p. 1090. ISSN: 2041-1723. DOI: 10.1038/s41467-021-21358-2. URL: https://www.nature.com/articles/s41467-021-21358-2 (visited on 06/04/2021).

[159] Kris V Parag and Christl A Donnelly. "Adaptive Estimation for Epidemic Renewal and Phylogenetic Skyline Models". In: *Systematic Biology* 69.6 (Nov. 1, 2020), pp. 1163–1179. ISSN: 1063-5157. DOI: 10.1093/sysbio/syaa035. URL: https://doi.org/10.1093/sysbio/syaa035 (visited on 06/04/2021).

[160] Kris V. Parag, Benjamin J. Cowling, and Christl A. Donnelly. "Deciphering Early-Warning Signals of the Elimination and Resurgence Potential of SARS-CoV-2 from Limited Data at Multiple Scales". In: *medRxiv* (Jan. 5, 2021), p. 2020.11.23.20236968. DOI: 10.1101/2020.11.23.20236968. URL: https://www.medrxiv.org/content/10.1101/2020.11.23.20236968v2 (visited on 06/04/2021).

[161] Kris V. Parag and Christl A. Donnelly. "Using Information Theory to Optimise Epidemic Models for Real-Time Prediction and Estimation". In: *PLOS Computational Biology* 16.7 (July 1, 2020), e1007990. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1007990. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1007990 (visited on 06/04/2021).

[162] Kris V. Parag et al. "An Exact Method for Quantifying the Reliability of End-of-Epidemic Declarations in Real Time". In: *PLOS Computational Biology* 16.11 (Nov. 30, 2020), e1008478. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1008478. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1008478 (visited on 06/04/2021).

[163] Chawarat Rotejanaprasert et al. "Preliminary Estimation of Temporal and Spatiotemporal Dynamic Measures of COVID-19 Transmission in Thailand". In: *PLOS ONE* 15.9 (Sept. 24, 2020), e0239645. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0239645. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0239645 (visited on 06/04/2021).

[164] Peng Wu et al. "Suppressing COVID-19 Transmission in Hong Kong: An Observational Study of the First Four Months". In: *Research Square [Pre-print]* (June 9, 2020). DOI: 10.21203/rs.3.rs-34047/v1. URL: https://www.researchsquare.com/article/rs-34047/v1 (visited on 06/04/2021).

[165] Takenori Yamauchi et al. "Estimation of the Effective Reproduction Number of Influenza Based on Weekly Reports in Miyazaki Prefecture". In: *Scientific Reports* 9.1 (1 Feb. 22, 2019), p. 2539. ISSN: 2045-2322. DOI: 10.1038/s41598-019-39057-w. URL: https://www.nature.com/articles/s41598-019-39057-w (visited on 06/04/2021).

[166] Shi Zhao et al. "The Long-Term Changing Dynamics of Dengue Infectivity in Guangdong, China, from 2008–2018: A Modelling Analysis". In: *Transactions of The Royal Society of Tropical Medicine and Hygiene* 114.1 (Jan. 6, 2020), pp. 62–71. ISSN: 0035-9203. DOI: 10.1093/trstmh/trz084. URL: https://doi.org/10.1093/trstmh/trz084 (visited on 06/04/2021).

[167] Theodore Lytras. *Estimate Epidemic Effective Reproduction Number in a Bayesian Framework [R Package bayEStim Version 0.0.1]*. URL: https://github.com/thlytras/bayEStim.

[168] Sam Abbott et al. *EpiNow2: Estimate Real-Time Case Counts and Time-Varying Epidemiological Parameters*. Version 1.3.2. Dec. 14, 2020. URL: https://CRAN.R-project.org/package=EpiNow2 (visited on 11/18/2021).

[169] *Epidemia [R Package Version 1.0.0]*. Imperial College London, May 1, 2021. URL: https://github.com/ImperialCollegeLondon/epidemia (visited on 06/07/2021).

[170] UK Government. *Coronavirus in the UK - About the Data*. URL: https://coronavirus.data.gov.uk/details/about-data (visited on 09/22/2021).

[171] Kris Parag. *Kpzoo/EpiFilter*. Jan. 11, 2021. URL: https://github.com/kpzoo/EpiFilter (visited on 06/07/2021).

[172] Kris Parag. *Kpzoo/APEestim*. Dec. 4, 2020. URL: https://github.com/kpzoo/APEestim (visited on 06/07/2021).

[173] Thibaut Jombart et al. *earlyR: Estimation of Transmissibility in the Early Stages of a Disease Outbreak*. Version 0.0.5. Oct. 27, 2020. URL: https://CRAN.R-project.org/package=earlyR (visited on 11/02/2021).

[174] Finlay Campbell et al. *Epicontacts: Handling, Visualisation and Analysis of Epidemiological Contacts*. Version 1.1.2. Oct. 21, 2021. URL: https://CRAN.R-project.org/package=epicontacts (visited on 11/18/2021).

[175] UK Government. *The R Value and Growth Rate*. GOV.UK. URL: https://www.gov.uk/guidance/the-r-value-and-growth-rate (visited on 02/21/2022).

[176] Neil M. Ferguson et al. "Strategies for Containing an Emerging Influenza Pandemic in Southeast Asia". In: *Nature* 437.7056 (7056 Sept. 2005), pp. 209–214. ISSN: 1476-4687. DOI: 10.1038/nature04017. URL: https://www.nature.com/articles/nature04017 (visited on 02/15/2022).

[177] Babak Pourbohloul et al. "Modeling Control Strategies of Respiratory Pathogens". In: *Emerging Infectious Diseases* 11.8 (Aug. 2005), pp. 1249–1256. ISSN: 1080-6040. DOI: 10.3201/eid1108.040449. pmid: 16102315. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3320482/ (visited on 02/15/2022).

[178] Diana Buitrago-Garcia et al. "Occurrence and Transmission Potential of Asymptomatic and Presymptomatic SARS-CoV-2 Infections: A Living Systematic Review and Meta-Analysis". In: *PLoS medicine* 17.9 (Sept. 2020), e1003346. ISSN: 1549-1676. DOI: 10.1371/journal.pmed.1003346. pmid: 32960881.

[179] Oyungerel Byambasuren et al. "Estimating the Extent of Asymptomatic COVID-19 and Its Potential for Community Transmission: Systematic Review and Meta-Analysis". In: *Official Journal of the Association of Medical Microbiology and Infectious Disease Canada* 5.4 (Dec. 1, 2020), pp. 223–234. ISSN: 2371-0888. DOI: 10.3138/jammi-2020-0030. URL: https://jammi.utpjournals.press/doi/full/10.3138/jammi-2020-0030 (visited on 06/07/2021).

[180] *Coronavirus (COVID-19) Testing - Statistics and Research*. Our World in Data. URL: https://ourworldindata.org/coronavirus-testing (visited on 06/07/2021).

[181] *WHO | FluNet*. WHO. URL: http://www.who.int/influenza/gisrs_laboratory/flunet/en/ (visited on 06/07/2021).

[182] CSSEGISandData. *COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University*. July 5, 2022. URL: https://github.com/CSSEGISandData/COVID-19 (visited on 07/05/2022).

[183] *EpiEstim/Vignettes/full_EpiEstim_vignette.Rmd at Master · Mrc-Ide/EpiEstim*. GitHub. URL: https://github.com/mrc-ide/EpiEstim/blob/master/vignettes/full_EpiEstim_vignette.Rmd (visited on 09/04/2024).

[184] Kory D. Johnson et al. "Disease Momentum: Estimating the Reproduction Number in the Presence of Superspreading". In: *Infectious Disease Modelling* 6 (Jan. 1, 2021), pp. 706–728. ISSN: 2468-0427. DOI: 10.1016/j.idm.2021.03.006. URL: https://www.sciencedirect.com/science/article/pii/S2468042721000270 (visited on 03/17/2022).

[185] Andrea Brizzi, Megan O'Driscoll, and Ilaria Dorigatti. "Refining Reproduction Number Estimates to Account for Unobserved Generations of Infection in Emerging Epidemics". In: *Clinical Infectious Diseases* 75.1 (July 2022), e114–e121. ISSN: 1058-4838. DOI: 10.1093/cid/ciac138. (Visited on 03/25/2025).

[186] William Green, Neil Ferguson, and Anne Cori. "Inferring the Reproduction Number Using the Renewal Equation in Heterogeneous Epidemics". In: *Journal of the Royal Society, Interface* 19.188 (Mar. 2022), p. 20210429. ISSN: 1742-5662. DOI: 10.1098/rsif.2021.0429. pmid: 35350879.

[187] John Griffin et al. "Rapid Review of Available Evidence on the Serial Interval and Generation Time of COVID-19". In: *BMJ Open* 10.11 (Nov. 1, 2020), e040263. ISSN: 2044-6055, 2044-6055. DOI: 10.1136/bmjopen-2020-040263. pmid: 33234640. URL: https://bmjopen.bmj.com/content/10/11/e040263 (visited on 03/17/2022).

[188] Rebecca K. Nash et al. "Ebola Virus Disease Mathematical Models and Epidemiological Parameters: A Systematic Review". In: *The Lancet Infectious Diseases* 0.0 (Aug. 2024). ISSN: 1473-3099, 1474-4457. DOI: 10.1016/S1473-3099(24)00374-8. (Visited on 09/30/2024).

[189] *Epireview: Tools to Update and Summarise the Latest Pathogen Data from the Pathogen Epidemiology Review Group (PERG)*. (Visited on 09/30/2024).

[190] *Pathogen Epidemiology Review Group*. https://www.imperial.ac.uk/medicine/departments/school-public-health/infectious-disease-epidemiology/mrc-global-infectious-disease-analysis/related-initiatives/perg/. (Visited on 12/04/2024).

[191] Spinello Antinori et al. "Update on Mpox: A Brief Narrative Review". In: *Le Infezioni in Medicina* 31.3 (2023), p. 269. DOI: 10.53854/liim-3103-1. (Visited on 09/30/2024).

[192] Olalekan J. Okesanya et al. “Rise of Marburg Virus in Africa: A Call for Global Preparedness”. In: *Annals of Medicine and Surgery* 85.10 (Sept. 2023), pp. 5285–5290. ISSN: 2049-0801. DOI: 10.1097/MS9.0000000000001257. (Visited on 09/30/2024).

[193] WHO. *Prioritizing Diseases for Research and Development in Emergency Contexts*. https://www.who.int/activities/prioritizing-diseases-for-research-and-development-in-emergency-contexts. (Visited on 09/30/2024).

[194] Heinz Feldmann, Armand Sprecher, and Thomas W. Geisbert. “Ebola”. In: *The New England Journal of Medicine* 382.19 (May 2020), pp. 1832–1842. ISSN: 1533-4406. DOI: 10.1056/NEJMra1901594.

[195] UK Health Security Agency. *Ebola: Overview, History, Origins and Transmission*. https://www.gov.uk/government/publications/ebola-origins-reservoirs-transmission-and-guidelines/ebola-overview-history-origins-and-transmission. (Visited on 09/30/2024).

[196] CDC. *Outbreak History*. https://www.cdc.gov/ebola/outbreaks/index.html. Aug. 2024. (Visited on 09/30/2024).

[197] WHO. *Ebola Outbreak 2014-2016 - West Africa*. https://www.who.int/emergencies/situations/ebola-outbreak-2014-2016-West-Africa. (Visited on 09/30/2024).

[198] Tracey Goldstein et al. “The Discovery of Bombali Virus Adds Further Support for Bats as Hosts of Ebolaviruses”. In: *Nature Microbiology* 3.10 (Oct. 2018), pp. 1084–1089. ISSN: 2058-5276. DOI: 10.1038/s41564-018-0227-2. (Visited on 09/30/2024).

[199] WHO. *Therapeutics for Ebola Virus Disease*. https://www.who.int/publications/i/item/9789240055742. (Visited on 09/30/2024).

[200] Ana Maria Henao-Restrepo et al. “Efficacy and Effectiveness of an rVSV-vectored Vaccine in Preventing Ebola Virus Disease: Final Results from the Guinea Ring Vaccination, Open-Label, Cluster-Randomised Trial (Ebola Ça Suffit!)” In: *The Lancet* 389.10068 (Feb. 2017), pp. 505–518. ISSN: 0140-6736, 1474-547X. DOI: 10.1016/S0140-6736(16)32621-6. (Visited on 09/30/2024).

[201] WHO. *Ebola Virus Disease: Vaccines*. https://www.who.int/news-room/questions-and-answers/item/ebola-vaccines. (Visited on 03/01/2024).

[202] Tsung-Pei Tsou. “Sudan Virus Disease - A Quick Review”. In: *Journal of the Formosan Medical Association = Taiwan Yi Zhi* 123.1 (Jan. 2024), pp. 16–22. ISSN: 0929-6646. DOI: 10.1016/j.jfma.2023.06.001.

[203] Anton Camacho et al. “Temporal Changes in Ebola Transmission in Sierra Leone and Implications for Control Requirements: A Real-time Modelling Study”. In: *PLOS Currents Outbreaks* (Feb. 2015). ISSN: 2157-3999. DOI: 10.1371/currents.outbreaks.406ae55e83ec0b5193e30856b9235ed2. (Visited on 09/30/2024).

[204] Ebola Outbreak Epidemiology Team. “Outbreak of Ebola Virus Disease in the Democratic Republic of the Congo, April–May, 2018: An Epidemiological Study”. In: *The Lancet* 392.10143 (July 2018), pp. 213–221. ISSN: 0140-6736, 1474-547X. DOI: 10.1016/S0140-6736(18)31387-4. (Visited on 09/30/2024).

[205] J. Daniel Kelly et al. “Real-Time Predictions of the 2018–2019 Ebola Virus Disease Outbreak in the Democratic Republic of the Congo Using Hawkes Point Process Models”. In: *Epidemics* 28 (Sept. 2019), p. 100354. ISSN: 1755-4365. DOI: 10.1016/j.epidem.2019.100354. (Visited on 09/30/2024).

[206] Julii Brainard et al. “Risk Factors for Transmission of Ebola or Marburg Virus Disease: A Systematic Review and Meta-Analysis”. In: *International Journal of Epidemiology* 45.1 (Feb. 2016), pp. 102–116. ISSN: 0300-5771. DOI: 10.1093/ije/dyv307. (Visited on 09/30/2024).

[207] Natalie E. Dean et al. “Transmissibility and Pathogenicity of Ebola Virus: A Systematic Review and Meta-analysis of Household Secondary Attack Rate and Asymptomatic Infection”. In: *Clinical Infectious Diseases* 62.10 (May 2016), pp. 1277–1286. ISSN: 1058-4838. DOI: 10.1093/cid/ciw114. (Visited on 09/30/2024).

[208] Maria D. Van Kerkhove et al. “A Review of Epidemiological Parameters from Ebola Outbreaks to Inform Early Public Health Decision-Making”. In: *Scientific Data* 2.1 (May 2015), p. 150019. ISSN: 2052-4463. DOI: 10.1038/sdata.2015.19. (Visited on 09/30/2024).

[209] Gustavo E. Velásquez et al. “Time From Infection to Disease and Infectiousness for Ebola Virus Disease, a Systematic Review”. In: *Clinical Infectious Diseases* 61.7 (Oct. 2015), pp. 1135–1140. ISSN: 1058-4838. DOI: 10.1093/cid/civ531. (Visited on 09/30/2024).

[210] Z. S. Y. Wong et al. “A Systematic Review of Early Modelling Studies of Ebola Virus Disease in West Africa”. In: *Epidemiology & Infection* 145.6 (Apr. 2017), pp. 1069–1094. ISSN: 0950-2688, 1469-4409. DOI: 10.1017/S0950268817000164. (Visited on 09/30/2024).

[211] Saranya A Selvaraj et al. “Infection Rates and Risk Factors for Infection Among Health Workers During Ebola and Marburg Virus Outbreaks: A Systematic Review”. In: *The Journal of Infectious Diseases* 218.suppl_5 (Nov. 2018), S679–S689. ISSN: 0022-1899. DOI: 10.1093/infdis/jiy435. (Visited on 09/30/2024).

[212] Hilary Bower and Judith R. Glynn. "A Systematic Review and Meta-Analysis of Seroprevalence Surveys of Ebolavirus Infection". In: *Scientific Data* 4.1 (Jan. 2017), p. 160133. ISSN: 2052-4463. DOI: 10.1038/sdata.2016.133. (Visited on 09/30/2024).

[213] Suliman Jamiel M. Abdalla, Faraimunashe Chirove, and Keshlan S. Govinder. "A Systematic Review of Mathematical Models of the Ebola Virus Disease". In: *International Journal of Modelling and Simulation* 42.5 (Sept. 2022), pp. 814–830. ISSN: 0228-6203. DOI: 10.1080/02286203.2021.1983745. (Visited on 09/30/2024).

[214] A. M. Rojek et al. "A Systematic Review and Meta-Analysis of Patient Data from the West Africa (2013–16) Ebola Virus Disease Epidemic". In: *Clinical Microbiology and Infection* 25.11 (Nov. 2019), pp. 1307–1314. ISSN: 1198-743X. DOI: 10.1016/j.cmi.2019.06.032. (Visited on 09/30/2024).

[215] J. Kawuki, T. H. Musa, and X. Yu. "Impact of Recurrent Outbreaks of Ebola Virus Disease in Africa: A Meta-Analysis of Case Fatality Rates". In: *Public Health* 195 (June 2021), pp. 89–97. ISSN: 1476-5616. DOI: 10.1016/j.puhe.2021.03.027.

[216] Drifa Belhadi et al. "The Number of Cases, Mortality and Treatments of Viral Hemorrhagic Fevers: A Systematic Review". In: *PLOS Neglected Tropical Diseases* 16.10 (Oct. 2022), e0010889. ISSN: 1935-2735. DOI: 10.1371/journal.pntd.0010889. (Visited on 09/30/2024).

[217] Luke Nyakarahuka et al. "How Severe and Prevalent Are Ebola and Marburg Viruses? A Systematic Review and Meta-Analysis of the Case Fatality Rates and Seroprevalence". In: *BMC Infectious Diseases* 16.1 (Nov. 2016), p. 708. ISSN: 1471-2334. DOI: 10.1186/s12879-016-2045-6. (Visited on 09/30/2024).

[218] Richard FitzJohn, Rob Ashton, and Alex Hill. *Mrc-Ide/Orderly2*. MRC Centre for Global Infectious Disease Analysis.

[219] Sara Balduzzi, Gerta Rücker, and Guido Schwarzer. "How to Perform a Meta-Analysis with R: A Practical Tutorial". In: *Evidence-Based Mental Health* 22.4 (Nov. 2019), pp. 153–160. ISSN: 1468-960X. DOI: 10.1136/ebmental-2019-300117.

[220] Guido Schwarzer, James R. Carpenter, and Gerta Rücker. *Meta-Analysis with R*. Use R! Cham: Springer International Publishing, 2015. ISBN: 978-3-319-21415-3 978-3-319-21416-0. DOI: 10.1007/978-3-319-21416-0. (Visited on 06/02/2025).

[221] Mathias Harrer et al. *Doing Meta-Analysis with R: A Hands-on Guide*. https://bookdown.org/MathiasHarrer/Doing_Meta_Analysis_in_R/. (Visited on 05/29/2024).

[222] Julian P. T. Higgins and Simon G. Thompson. "Quantifying Heterogeneity in a Meta-Analysis". In: *Statistics in Medicine* 21.11 (June 2002), pp. 1539–1558. ISSN: 0277-6715. DOI: 10.1002/sim.1186.

[223] Oyewale Tomori et al. "Serologic Survey among Hospital and Health Center Workers during the Ebola Hemorrhagic Fever Outbreak in Kikwit, Democratic Republic of the Congo, 1995". In: *The Journal of Infectious Diseases* 179.Supplement_1 (Feb. 1999), S98–S101. ISSN: 0022-1899. DOI: 10.1086/514307. (Visited on 09/30/2024).

[224] Placide Mbala-Kingebeni et al. "2018 Ebola Virus Disease Outbreak in Équateur Province, Democratic Republic of the Congo: A Retrospective Genomic Characterisation". In: *The Lancet Infectious Diseases* 19.6 (June 2019), pp. 641–647. ISSN: 1473-3099, 1474-4457. DOI: 10.1016/S1473-3099(19)30124-0. (Visited on 09/30/2024).

[225] Y. H. Li and S. P. Chen. "Evolutionary History of Ebola Virus". In: *Epidemiology & Infection* 142.6 (June 2014), pp. 1138–1145. ISSN: 0950-2688, 1469-4409. DOI: 10.1017/S0950268813002215. (Visited on 09/30/2024).

[226] Jade Vincent Membrebe et al. "Bayesian Inference of Evolutionary Histories under Time-Dependent Substitution Rates". In: *Molecular Biology and Evolution* 36.8 (Aug. 2019), pp. 1793–1803. ISSN: 0737-4038. DOI: 10.1093/molbev/msz094. (Visited on 09/30/2024).

[227] Christine E. Hulseberg et al. "Molecular Analysis of the 2012 Bundibugyo Virus Disease Outbreak". In: *Cell Reports Medicine* 2.8 (Aug. 2021). ISSN: 2666-3791. DOI: 10.1016/j.xcrm.2021.100351. (Visited on 09/30/2024).

[228] Antoine Nkuba-Ndaye et al. "Added Value of an Anti-Ebola Serology for the Management of Clinically Suspected Ebola Virus Disease Patients Discharged as Negative in an Epidemic Context". In: *The Journal of Infectious Diseases* 226.2 (July 2022), pp. 352–356. ISSN: 0022-1899. DOI: 10.1093/infdis/jiac057. (Visited on 09/30/2024).

[229] Kristina M. Busico et al. "Prevalence of IgG Antibodies to Ebola Virus in Individuals during an Ebola Outbreak, Democratic Republic of the Congo, 1995". In: *The Journal of Infectious Diseases* 179.Supplement_1 (Feb. 1999), S102–S107. ISSN: 0022-1899. DOI: 10.1086/514309. (Visited on 09/30/2024).

[230] A. Camacho et al. "Potential for Large Outbreaks of Ebola Virus Disease". In: *Epidemics* 9 (Dec. 2014), pp. 70–78. ISSN: 1755-4365. DOI: 10.1016/j.epidem.2014.09.003. (Visited on 09/30/2024).

[231] Alexis Robert et al. "Control of Ebola Virus Disease Outbreaks: Comparison of Health Care Worker-Targeted and Community Vaccination Strategies". In: *Epidemics* 27 (June 2019), pp. 106–114. ISSN: 1878-0067. DOI: 10.1016/j.epidem.2019.03.001.

[232] Emma Saulnier, Olivier Gascuel, and Samuel Alizon. "Inferring Epidemiological Parameters from Phylogenies Using Regression-ABC: A Comparative Study". In: *PLOS Computational Biology* 13.3 (Mar. 2017), e1005416. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1005416. (Visited on 09/30/2024).

[233] C. L. Althaus et al. "Ebola Virus Disease Outbreak in Nigeria: Transmission Dynamics and Rapid Control". In: *Epidemics* 11 (June 2015), pp. 80–84. ISSN: 1878-0067. DOI: 10.1016/j.epidem.2015.03.001.

[234] Paul Shears and Carrie Garavan. "The 2018/19 Ebola Epidemic the Democratic Republic of the Congo (DRC): Epidemiology, Outbreak Control, and Conflict". In: *Infection Prevention in Practice* 2.1 (Mar. 2020), p. 100038. ISSN: 2590-0889. DOI: 10.1016/j.infpip.2020.100038.

[235] Tini Garske et al. "Heterogeneities in the Case Fatality Ratio in the West African Ebola Outbreak 2013–2016". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 372.1721 (Apr. 2017), p. 20160308. DOI: 10.1098/rstb.2016.0308. (Visited on 09/30/2024).

[236] Patrick Doohan et al. *Lassa Fever Outbreaks, Mathematical Models, and Disease Parameters: A Systematic Review and Meta-Analysis*. Mar. 2024. DOI: 10.1101/2024.03.23.24304596. (Visited on 09/30/2024).

[237] J. O. Lloyd-Smith et al. "Superspreading and the Effect of Individual Variation on Disease Emergence". In: *Nature* 438.7066 (Nov. 2005), pp. 355–359. ISSN: 1476-4687. DOI: 10.1038/nature04153. (Visited on 09/30/2024).

[238] M. D. Hall, M. E. J. Woolhouse, and A. Rambaut. "Using Genomics Data to Reconstruct Transmission Trees during Disease Outbreaks". In: *Revue Scientifique Et Technique (International Office of Epizootics)* 35.1 (Apr. 2016), pp. 287–296. ISSN: 0253-1933. DOI: 10.20506/rst.35.1.2433.

[239] Placide Mbala-Kingebeni et al. "Medical Countermeasures during the 2018 Ebola Virus Disease Outbreak in the North Kivu and Ituri Provinces of the Democratic Republic of the Congo: A Rapid Genomic Assessment". In: *The Lancet. Infectious Diseases* 19.6 (June 2019), pp. 648–657. ISSN: 1474-4457. DOI: 10.1016/S1473-3099(19)30118-5.

[240] Sangeeta Bhatia et al. "Extending EpiEstim to Estimate the Transmission Advantage of Pathogen Variants in Real-Time: SARS-CoV-2 as a Case-Study". In: *Epidemics* (June 21, 2023), p. 100692. ISSN: 1755-4365. DOI: 10.1016/j.epidem.2023.100692. URL: https://www.sciencedirect.com/science/article/pii/S1755436523000282 (visited on 06/26/2023).

[241] *Imperial College London: Delta (B.1.617.2) Transmission in England – Risk Factors and Transmission Advantage, 1 June 2021*. https://www.gov.uk/government/publications/imperial-

college-london-delta-b16172-transmission-in-england-risk-factors-and-transmission-advantage-1-june-2021. (Visited on 01/20/2025).

[242] *Prime minister's statement on coronavirus (COVID-19): 19 december 2020*. Dec. 2020. URL: https://www.gov.uk/government/speeches/prime-ministers-statement-on-coronavirus-covid-19-19-december-2020 (visited on 11/08/2021).

[243] *PM statement at coronavirus press conference: 14 June 2021*. June 2021. URL: https://www.gov.uk/government/speeches/pm-statement-at-coronavirus-press-conference-14-june-2021 (visited on 10/25/2021).

[244] *Tracking SARS-CoV-2 variants*. 2021. URL: https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/ (visited on 10/25/2021).

[245] Raphael Sonabend et al. "Non-pharmaceutical interventions, vaccination, and the SARS-CoV-2 delta variant in England: a mathematical modelling study". In: *The Lancet* (Oct. 2021). DOI: 10.1016/S0140-6736(21)02276-5.

[246] Andrew Rambaut et al. *Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations - SARS-CoV-2 coronavirus / nCoV-2019 Genomic Epidemiology*. Dec. 2020. URL: https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563 (visited on 07/12/2022).

[247] *SARS-CoV-2 variants of concern and variants under investigation in England: Technical Briefing 15*. 2021. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/993879/Variants_of_Concern_VOC_Technical_Briefing_15.pdf (visited on 09/01/2021).

[248] Neil M Ferguson. *B.1.617.2 transmission in England: Risk factors and transmission advantage*. Pages: 14. June 2021. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/993159/S1270_IMPERIAL_B.1.617.2.pdf (visited on 07/11/2022).

[249] Samuel Alizon et al. "Rapid spread of the SARS-CoV-2 Delta variant in some French regions, June 2021". In: *Eurosurveillance* 26.28 (July 2021). DOI: 10.2807/1560-7917.ES.2021.26.28.2100573.

[250] Finlay Campbell et al. "Increased transmissibility and global spread of SARS-CoV-2 variants of concern as at June 2021". In: *Eurosurveillance* 26.24 (June 2021). DOI: 10.2807/1560-7917.ES.2021.26.24.2100509.

[251] Matt J Keeling. *Estimating the transmission advantage for B.1.617.2*. 2021. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/993156/S1269_WARWICKTransmission_Advantage.pdf (visited on 09/15/2021).

[252] Wan Yang et al. "Epidemiological characteristics of the B.1.526 SARS-CoV-2 variant". In: *Science Advances* 8.4 (Jan. 2022). tex.pmcid: PMC8797779, eabm0300. DOI: 10.1126/sciadv.abm0300.

[253] Renato Mendes Coutinho et al. "Model-based estimation of transmissibility and reinfection of SARS-CoV-2 P.1 variant". In: *Communications Medicine* 1.1 (Dec. 2021), p. 48. DOI: 10.1038/s43856-021-00048-6.

[254] Nuno R. Faria et al. "Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil". In: *Science* 372.6544 (May 2021), pp. 815–821. DOI: 10.1126/science.abh2644.

[255] Christiaan H. van Dorp et al. "Estimating the strength of selection for new SARS-CoV-2 variants". In: *Nature Communications* 12.1 (Dec. 2021), p. 7239. DOI: 10.1038/s41467-021-27369-3.

[256] Christiaan H. van Dorp et al. "Global estimates of the fitness advantage of SARS-CoV-2 variant Omicron". In: *Virus Evolution* 8.2 (Oct. 2022), veac089. DOI: 10.1093/ve/veac089.

[257] Marlin D. Figgins and Trevor Bedford. *SARS-CoV-2 variant dynamics across US states show consistent differences in effective reproduction numbers*. medRxiv. Dec. 2021. DOI: 10.1101/2021.12.09.21267544.

[258] Anne Cori. *EpiEstim*. Dec. 2021. URL: https://github.com/mrc-ide/EpiEstim (visited on 11/16/2021).

[259] Andrew Gelman and Donald B. Rubin. "Inference from Iterative Simulation Using Multiple Sequences". In: *Statistical Science* 7.4 (Nov. 1992). DOI: 10.1214/ss/1177011136.

[260] *COVID-19 Point épidémiologique hebdomadaire du 28 janvier 2021*. URL: https://www.santepubliquefrance.fr/content/download/315275/2903017 (visited on 10/25/2021).

[261] *COVID-19 Point épidémiologique hebdomadaire du 4 mars 2021*. URL: https://www.santepubliquefrance.fr/content/download/324805/2944195 (visited on 10/25/2021).

[262] *Tableau synthétique des résultats par vague d'enquêtes*. URL: https://www.santepubliquefrance.fr/content/download/368440/3132368 (visited on 10/25/2021).

[263] *EpiEstim package - rdocumentation*. 2021. URL: https://www.rdocumentation.org/packages/EpiEstim/versions/2.2-4 (visited on 09/19/2021).

[264] Moritz U G Kraemer et al. "Spatiotemporal invasion dynamics of SARS-CoV-2 lineage B.1.1.7 emergence". In: *Science* 373.6557 (2021), pp. 889–895. DOI: 10.1126/science.abj0113.

[265] Robert Hinch et al. “Estimating SARS-CoV-2 variant fitness and the impact of interventions in England using statistical and geo-spatial agent-based models”. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 380.2233 (Aug. 2022), p. 20210304. DOI: 10.1098/rsta.2021.0304. URL: https://royalsocietypublishing.org/doi/10.1098/rsta.2021.0304 (visited on 04/05/2023).

[266] François Blanquart et al. “Selection for infectivity profiles in slow and fast epidemics, and the rise of SARS-CoV-2 variants”. In: *eLife* 11 (May 2022). Ed. by Ben S Cooper and Miles P Davenport, e75791. ISSN: 2050-084X. DOI: 10.7554/eLife.75791. URL: https://doi.org/10.7554/eLife.75791 (visited on 04/05/2023).

[267] Kimihito Ito et al. “Estimating relative generation times and reproduction numbers of Omicron BA.1 and BA.2 with respect to Delta variant in Denmark”. en. In: *Mathematical Biosciences and Engineering* 19.9 (2022). tex.copyright: 2022 The Author(s), pp. 9005–9017. ISSN: 1551-0018. DOI: 10.3934/mbe.2022418. URL: http://www.aimspress.com/rticle/doi/10.3934/mbe.2022418 (visited on 04/05/2023).

[268] Jantien A. Backer et al. “Shorter serial intervals in SARS-CoV-2 cases with Omicron BA.1 variant compared with Delta variant, the Netherlands, 13 to 26 December 2021”. en. In: *Eurosurveillance* 27.6 (Feb. 2022), p. 2200042. ISSN: 1560-7917. DOI: 10.2807/1560-7917.ES.2022.27.6.2200042. URL: https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2022.27.6.2200042 (visited on 04/05/2023).

[269] Edouard Mathieu et al. “A global database of COVID-19 vaccinations”. In: *Nature Human Behaviour* 5.7 (July 2021), pp. 947–953. DOI: 10.1038/s41562-021-01122-8.

[270] Rebecca K. Nash et al. “Estimating the epidemic reproduction number from temporally aggregated incidence data: A statistical modelling approach and software tool”. en. In: *PLOS Computational Biology* 19.8 (Aug. 2023). Publisher: Public Library of Science, e1011439. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1011439. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1011439 (visited on 09/13/2024).

[271] *National flu and COVID-19 surveillance reports: 2021 to 2022 season*. en. URL: https://www.gov.uk/government/statistics/national-flu-and-covid-19-surveillance-reports-2021-to-2022-season (visited on 06/27/2022).

[272] Oscar Pacheco et al. “Zika Virus Disease in Colombia — Preliminary Report”. In: *New England Journal of Medicine* 383.6 (Aug. 2020). Publisher: Massachusetts Medical Society _eprint: https://doi.org/10.1056/NEJMoa1604037, e44. ISSN: 0028-4793. DOI: 10.1056/NEJMoa1604037. URL: https://doi.org/10.1056/NEJMoa1604037 (visited on 06/27/2022).

[273] *Notifiable diseases: weekly reports for 2022*. en. URL: https://www.gov.uk/government/publications/notifiable-diseases-weekly-reports-for-2022 (visited on 06/27/2022).

[274] *Notifiable Infectious Disease Tables | CDC*. en-us. Sept. 2021. URL: https://www.cdc.gov/nndss/data-statistics/infectious-tables/index.html (visited on 07/02/2022).

[275] Rachel E. Baker et al. "Infectious disease in an era of global change". en. In: *Nature Reviews Microbiology* 20.4 (Apr. 2022). Number: 4 Publisher: Nature Publishing Group, pp. 193–205. ISSN: 1740-1534. DOI: 10.1038/s41579-021-00639-z. URL: https://www.nature.com/articles/s41579-021-00639-z (visited on 07/26/2022).

[276] Swapnil Mishra et al. *A COVID-19 Model for Local Authorities of the United Kingdom*. en. Pages: 2020.11.24.20236661. Nov. 2020. DOI: 10.1101/2020.11.24.20236661. URL: https://www.medrxiv.org/content/10.1101/2020.11.24.20236661v1 (visited on 07/01/2022).

[277] *Role of Data Aggregation in Biosurveillance Detection Strategies with Applications from ESSENCE*. URL: https://www.cdc.gov/mmwr/preview/mmwrhtml/su5301a16.htm (visited on 07/02/2022).

[278] Kelly Charniga et al. "Spatial and temporal invasion dynamics of the 2014–2017 Zika and chikungunya epidemics in Colombia". en. In: *PLOS Computational Biology* 17.7 (July 2021). Publisher: Public Library of Science, e1009174. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1009174. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1009174 (visited on 06/27/2022).

[279] Benjamin J. Cowling et al. "Estimation of the serial interval of influenza". In: *Epidemiology (Cambridge, Mass.)* 20.3 (May 2009), pp. 344–347. ISSN: 1044-3983. DOI: 10.1097/EDE.0b013e31819d1092. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3057478/ (visited on 06/24/2022).

[280] Laura Forsberg White et al. "Estimation of the reproductive number and the serial interval in early phase of the 2009 influenza A/H1N1 pandemic in the USA". en. In: *Influenza and Other Respiratory Viruses* 3.6 (2009). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1750-2659.2009.00106.x, pp. 267–276. ISSN: 1750-2659. DOI: 10.1111/j.1750-2659.2009.00106.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1750-2659.2009.00106.x (visited on 06/24/2022).

[281] Qifang Bi et al. "Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study". English. In: *The Lancet Infectious Diseases* 20.8 (Aug. 2020). Publisher: Elsevier, pp. 911–919. ISSN: 1473-3099, 1474-4457. DOI: 10.1016/S1473-3099(20)30287-5. URL: https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(20)30287-5/fulltext (visited on 06/24/2022).

[282] Balram Rai, Anandi Shukla, and Laxmi Kant Dwivedi. "Estimates of serial interval for COVID-19: a systematic review and meta-analysis". In: *Clinical Epidemiology and Global Health* 9 (Jan. 2021), pp. 157–161. DOI: 10.1016/j.cegh.2020.08.007.

[283] J Wallinga and M Lipsitch. "How generation intervals shape the relationship between growth rates and reproductive numbers". In: *Proceedings of the Royal Society B: Biological Sciences* 274.1609 (Feb. 2007), pp. 599–604. ISSN: 0962-8452. DOI: 10.1098/rspb.2006.3754. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1766383/ (visited on 06/29/2022).

[284] Pete Riley, Angelia A. Cost, and Steven Riley. "Intra-Weekly Variations of Influenza-Like Illness in Military Populations". In: *Military Medicine* 181.4 (Apr. 2016), pp. 364–368. ISSN: 0026-4075. DOI: 10.7205/MILMED-D-15-00226. URL: https://doi.org/10.7205/MILMED-D-15-00226 (visited on 03/02/2022).

[285] *Cases in the UK | Coronavirus in the UK*. en. URL: https://coronavirus.data.gov.uk/details/cases (visited on 01/09/2022).

[286] H. Nishiura and G. Chowell. "Early transmission dynamics of Ebola virus disease (EVD), West Africa, March to August 2014". en. In: *Eurosurveillance* 19.36 (Sept. 2014). Publisher: European Centre for Disease Prevention and Control, p. 20894. ISSN: 1560-7917. DOI: 10.2807/1560-7917.ES2014.19.36.20894. URL: https://www.eurosurveillance.org/content/10.2807/1560-7917.ES2014.19.36.20894 (visited on 07/05/2022).

[287] Sheikh Taslim Ali et al. "Serial Interval of SARS-CoV-2 Was Shortened over Time by Non-pharmaceutical Interventions". In: *Science* 369.6507 (Aug. 28, 2020), pp. 1106–1109. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.abc9004. pmid: 32694200. URL: https://science.sciencemag.org/content/369/6507/1106 (visited on 06/02/2021).

[288] Elisabeth K. Brockhaus et al. *Why are different estimates of the effective reproductive number so different? A case study on COVID-19 in Germany*. en. Pages: 2023.04.27.23289109. Apr. 2023. DOI: 10.1101/2023.04.27.23289109. URL: https://www.medrxiv.org/content/10.1101/2023.04.27.23289109v1 (visited on 05/09/2023).

[289] Oswaldo Gressani et al. "EpiLPS: A fast and flexible Bayesian tool for estimation of the time-varying reproduction number". en. In: *PLOS Computational Biology* 18.10 (Oct. 2022). Publisher: Public Library of Science, e1010618. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1010618. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1010618 (visited on 05/09/2023).

[290] *mrc-ide/EpiEstim: A tool to estimate time varying instantaneous reproduction number during epidemics*. URL: https://github.com/mrc-ide/EpiEstim (visited on 08/09/2022).

[291] *Short-Term Forecasts of COVID-19 Deaths in Multiple Countries*. URL: https://mrc-ide.github.io/covid19-short-term-forecasts/ (visited on 12/20/2024).

[292] Nalini Chintalapudi et al. "COVID-19 Outbreak Reproduction Number Estimations and Forecasting in Marche, Italy". In: *International Journal of Infectious Diseases* 96 (July 1, 2020), pp. 327–333. ISSN: 1201-9712. DOI: 10.1016/j.ijid.2020.05.029. URL: https://www.sciencedirect.com/science/article/pii/S1201971220303362 (visited on 12/20/2024).

[293] Andrea L. Bertozzi et al. "The Challenges of Modeling and Forecasting the Spread of COVID-19". In: *Proceedings of the National Academy of Sciences* 117.29 (July 21, 2020), pp. 16732–16738. DOI: 10.1073/pnas.2006520117. URL: https://www-pnas-org.iclibezp1.cc.ic.ac.uk/doi/full/10.1073/pnas.2006520117 (visited on 12/20/2024).

[294] *COVID-19 Cases | WHO COVID-19 Dashboard*. datadot. URL: https://data.who.int/dashboards/covid19/cases (visited on 01/05/2025).

[295] Adeel Hassan. "Some U.S. States Are Reducing Daily Reporting of Coronavirus Data, Raising Fears of Blind Spots." In: *The New York Times. U.S.* (Mar. 19, 2022). ISSN: 0362-4331. URL: https://www.nytimes.com/2022/03/19/us/covid-reporting-states.html (visited on 01/05/2025).

[296] *State Data Reporting Reduction Continues Amidst COVID-19 Resurgence*. Johns Hopkins Coronavirus Resource Center. URL: https://coronavirus.jhu.edu/pandemic-data-initiative/news/state-data-reporting-reduction-continues-amidst-covid-19-resurgence (visited on 01/05/2025).

[297] *Download Historical Data (to 14 December 2020) on the Daily Number of New Reported COVID-19 Cases and Deaths Worldwide*. URL: https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide (visited on 01/05/2025).

[298] Michelle L. Holshue et al. "First Case of 2019 Novel Coronavirus in the United States". In: *The New England Journal of Medicine* 382.10 (Mar. 5, 2020), pp. 929–936. ISSN: 0028-4793. DOI: 10.1056/NEJMoa2001191. pmid: 32004427. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7092802/ (visited on 12/19/2024).

[299] WHO. *Coronavirus Disease (COVID-19) Pandemic*. URL: https://www.who.int/europe/emergencies/situations/covid-19 (visited on 12/19/2024).

[300] CDC. *CDC Museum COVID-19 Timeline*. Centers for Disease Control and Prevention. Mar. 15, 2023. URL: https://www.cdc.gov/museum/timeline/covid19.html (visited on 12/21/2024).

[301] *State Reporting Frequencies*. Johns Hopkins Coronavirus Resource Center. URL: https://coronavirus.jhu.edu/data/state-reporting-frequencies (visited on 12/20/2024).

[302] Thomas Hale et al. *Variation in US States' Responses to COVID-19*. BSG-WP-2020/034. Blavatnik School of Government & Oxford University, May 7, 2021. URL: http://www.bsg.ox.ac.uk/research/publications/variation-us-states-responses-covid-19 (visited on 12/20/2024).

[303] Shruti H. Mehta et al. "Holiday Gatherings, Mobility and SARS-CoV-2 Transmission: Results from 10 US States Following Thanksgiving". In: *Scientific Reports* 11.1 (Aug. 30, 2021), p. 17328. ISSN: 2045-2322. DOI: 10.1038/s41598-021-96779-6. URL: https://www.nature.com/articles/s41598-021-96779-6 (visited on 12/19/2024).

[304] Carlos del Rio, Preeti N. Malani, and Saad B. Omer. "Confronting the Delta Variant of SARS-CoV-2, Summer 2021". In: *JAMA* 326.11 (Sept. 21, 2021), pp. 1001–1002. ISSN: 0098-7484. DOI: 10.1001/jama.2021.14811. URL: https://doi.org/10.1001/jama.2021.14811 (visited on 01/05/2025).

[305] Pryanka Relan et al. "Severity and Outcomes of Omicron Variant of SARS-CoV-2 Compared to Delta Variant and Severity of Omicron Sublineages: A Systematic Review and Metanalysis". In: *BMJ Global Health* 8.7 (July 2023). ISSN: 2059-7908. DOI: 10.1136/bmjgh-2023-012328. (Visited on 01/15/2025).

[306] *National COVID-19 Preparedness Plan*. The White House. Mar. 2022. URL: https://www.whitehouse.gov/covidplan/ (visited on 01/05/2025).

[307] Ensheng Dong, Hongru Du, and Lauren Gardner. "An Interactive Web-Based Dashboard to Track COVID-19 in Real Time". In: *The Lancet Infectious Diseases* 20.5 (May 1, 2020), pp. 533–534. ISSN: 1473-3099, 1474-4457. DOI: 10.1016/S1473-3099(20)30120-1. pmid: 32087114. URL: https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(20)30120-1/fulltext (visited on 12/19/2024).

[308] *COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University*. Dec. 18, 2024. URL: https://github.com/CSSEGISandData/COVID-19 (visited on 12/19/2024).

[309] *U.S. Census Bureau QuickFacts: United States*. URL: https://www.census.gov/quickfacts/fact/table/ (visited on 01/21/2025).

[310] Zachary J. Madewell et al. "Rapid Review and Meta-Analysis of Serial Intervals for SARS-CoV-2 Delta and Omicron Variants". In: *BMC infectious diseases* 23.1 (June 26, 2023), p. 429. ISSN: 1471-2334. DOI: 10.1186/s12879-023-08407-5. pmid: 37365505.

[311] Yuke Wang and Peter Teunis. "Strongly Heterogeneous Transmission of COVID-19 in Mainland China: Local and Regional Variation". In: *Frontiers in Medicine* 7 (June 19, 2020). ISSN: 2296-

858X. DOI: 10.3389/fmed.2020.00329. URL: https://www.frontiersin.org/journals/medicine/articles/10.3389/fmed.2020.00329/full (visited on 12/23/2024).

[312] Sangeeta Bhatia et al. "Retrospective Evaluation of Real-Time Estimates of Global COVID-19 Transmission Trends and Mortality Forecasts". In: *PloS One* 18.10 (2023), e0286199. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0286199. pmid: 37851661.

[313] *Mrc-Ide/Jointlyr*. MRC Centre for Global Infectious Disease Analysis, Aug. 3, 2022. URL: https://github.com/mrc-ide/jointlyr (visited on 12/19/2024).

[314] Nikos I. Bosse et al. "Scoring Epidemiological Forecasts on Transformed Scales". In: *PLOS Computational Biology* 19.8 (Aug. 29, 2023), e1011393. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1011393. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1011393 (visited on 12/20/2024).

[315] Nikos I. Bosse et al. "Evaluating Forecasts with Scoringutils in R". In: *arXiv* (2022). DOI: 10.48550/ARXIV.2205.07090. URL: https://arxiv.org/abs/2205.07090.

[316] Alexander Jordan, Fabian Krüger, and Sebastian Lerch. "Evaluating Probabilistic Forecasts with scoringRules". In: *Journal of Statistical Software* 90.12 (2019), pp. 1–37. DOI: 10.18637/jss.v090.i12.

[317] Hannah Hoffman. *Analysis & Updates | How Day-of-Week Effects Impact COVID-19 Data*. The COVID Tracking Project. URL: https://covidtracking.com/analysis-updates/how-day-of-week-effects-impact-covid-19-data (visited on 01/03/2025).

[318] Eric C. Schneider. "Failing the Test — The Tragic Data Gap Undermining the U.S. Pandemic Response". In: *New England Journal of Medicine* 383.4 (July 23, 2020), pp. 299–302. ISSN: 0028-4793. DOI: 10.1056/NEJMp2014836. URL: https://www.nejm.org/doi/full/10.1056/NEJMp2014836 (visited on 01/03/2025).

[319] Adam Spannaus et al. "Inferring the Spread of COVID-19: The Role of Time-Varying Reporting Rate in Epidemiological Modelling". In: *Scientific Reports* 12.1 (June 24, 2022), p. 10761. ISSN: 2045-2322. DOI: 10.1038/s41598-022-14979-0. URL: https://www.nature.com/articles/s41598-022-14979-0 (visited on 01/05/2025).

[320] Amanda Jean Meadows et al. "Historical Trends Demonstrate a Pattern of Increasingly Frequent and Severe Spillover Events of High-Consequence Zoonotic Viruses". In: *BMJ Global Health* 8.11 (Nov. 2, 2023). ISSN: 2059-7908. DOI: 10.1136/bmjgh-2023-012026. pmid: 10.1136/bmjgh-2023-012026. URL: https://gh.bmj.com/content/8/11/e012026 (visited on 01/09/2025).

[321] ECDC. *Monkeypox Multi-Country Outbreak - First Update*. https://www.ecdc.europa.eu/en/publications-data/monkeypox-multi-country-outbreak-first-update. July 2022. (Visited on 01/13/2025).

[322] Tim K Tsang et al. “Effect of Changing Case Definitions for COVID-19 on the Epidemic Curve and Transmission Parameters in Mainland China: A Modelling Study”. In: *The Lancet Public Health* 5.5 (May 2020), e289–e296. ISSN: 2468-2667. DOI: 10.1016/S2468-2667(20)30089-X. (Visited on 01/13/2025).

[323] André Peralta-Santos. *Assessment of COVID-19 Surveillance Case Definitions and Data Reporting in the European Union*. Tech. rep. European Parliament.

[324] Owen Dyer. “Covid-19: Russia Admits to Understating Deaths by More than Two Thirds”. In: *BMJ* 371 (Dec. 2020), p. m4975. ISSN: 1756-1833. DOI: 10.1136/bmj.m4975. (Visited on 01/13/2025).

[325] Edouard Mathieu et al. “Coronavirus (COVID-19) Testing”. In: *Our World in Data* (May 2020). (Visited on 01/14/2025).

[326] World Health Organization. “Global Surveillance for human infection with novel coronavirus (2019-nCoV): interim guidance, 31 January 2020”. In: WHO/2019-nCoV/SurveillanceGuidance/2020.3 (2020). (Visited on 01/14/2025).

[327] World Health Organization. “Global Surveillance for COVID-19 Caused by Human Infection with COVID-19 Virus: Interim Guidance, 20 March 2020”. In: WHO/2019-nCoV/SurveillanceGuidance/2020.6 (2020). (Visited on 01/14/2025).

[328] World Health Organization. “Public Health Surveillance for COVID-19: Interim Guidance, 7 August 2020”. In: WHO/2019-nCoV/SurveillanceGuidance/2020.7 (2020). (Visited on 01/14/2025).

[329] Amitabh B. Suthar et al. “Coronavirus Disease Case Definitions, Diagnostic Testing Criteria, and Surveillance in 25 Countries with Highest Reported Case Counts”. In: *Emerging Infectious Diseases* 28.1 (Jan. 2022), pp. 148–156. ISSN: 1080-6040. DOI: 10.3201/eid2801.211082. (Visited on 01/14/2025).

[330] Robert Verity et al. “Estimates of the Severity of Coronavirus Disease 2019: A Model-Based Analysis”. In: *The Lancet Infectious Diseases* 20.6 (June 2020), pp. 669–677. ISSN: 1473-3099, 1474-4457. DOI: 10.1016/S1473-3099(20)30243-7. (Visited on 01/20/2025).

[331] *Workflow for Rt Estimation and Forecasting*. https://epiforecasts.io/EpiNow2/articles/estimate_infections_ (Visited on 01/14/2025).

[332] Zhenwei Zhou et al. “Estimation of Heterogeneous Instantaneous Reproduction Numbers with Application to Characterize SARS-CoV-2 Transmission in Massachusetts Counties”. In: *PLOS Computational Biology* 18.9 (Sept. 2022), e1010434. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1010434. (Visited on 01/20/2025).

[333] Edward S. Knock et al. “Key Epidemiological Drivers and Impact of Interventions in the 2020 SARS-CoV-2 Epidemic in England”. In: *Science Translational Medicine* 13.602 (July 2021), eabg4262. DOI: 10.1126/scitranslmed.abg4262. (Visited on 01/15/2025).

[334] Sen Pei, Sasikiran Kandula, and Jeffrey Shaman. “Differential Effects of Intervention Timing on COVID-19 Spread in the United States”. In: *Science Advances* 6.49 (Dec. 2020), eabd6370. DOI: 10.1126/sciadv.abd6370. (Visited on 01/15/2025).

[335] Gina Cuomo-Dannenburg et al. “Marburg Virus Disease Outbreaks, Mathematical Models, and Disease Parameters: A Systematic Review”. In: *The Lancet Infectious Diseases* 0.0 (Nov. 2023). ISSN: 1473-3099, 1474-4457. DOI: 10.1016/S1473-3099(23)00515-7. (Visited on 02/27/2024).

[336] Patrick Doohan et al. “Lassa Fever Outbreaks, Mathematical Models, and Disease Parameters: A Systematic Review and Meta-Analysis”. In: *The Lancet Global Health* 12.12 (Dec. 2024), e1962–e1972. ISSN: 2214-109X. DOI: 10.1016/S2214-109X(24)00379-6. (Visited on 12/06/2024).

[337] Christian Morgenstern et al. *Severe Acute Respiratory Syndrome (SARS) Mathematical Models and Disease Parameters: A Systematic Review and Meta-Analysis*. Aug. 2024. DOI: 10.1101/2024.08.13.24311934. (Visited on 12/06/2024).

[338] Sabine L. van Elsland et al. “Policy Impact of the Imperial College COVID-19 Response Team: Global Perspective and United Kingdom Case Study”. In: *Health Research Policy and Systems* 22.1 (Nov. 2024), p. 153. ISSN: 1478-4505. DOI: 10.1186/s12961-024-01236-1. (Visited on 01/15/2025).

[339] *WHO Recommended Criteria for Declaring the End of the Ebola Virus Disease Outbreak*. https://www.who.int/publications/m/item/who-recommended-criteria-for-declaring-the-end-of-the-ebola-virus-disease-outbreak. (Visited on 01/16/2025).

[340] WHO. *Lassa Fever – Nigeria*. https://www.who.int/emergencies/disease-outbreak-news/item/27-may-2016-lassa-fever-nigeria-en. (Visited on 01/16/2025).

[341] Bimandra A Djaafara et al. “A Quantitative Framework for Defining the End of an Infectious Disease Outbreak: Application to Ebola Virus Disease”. In: *American Journal of Epidemiology* 190.4 (Apr. 2021), pp. 642–651. ISSN: 0002-9262. DOI: 10.1093/aje/kwaa212. (Visited on 06/14/2025).

[342] *Tracking SARS-CoV-2 Variants*. https://www.who.int/activities/tracking-SARS-CoV-2-variants. (Visited on 01/15/2025).

[343] *Coronavirus Disease (COVID-19): Variants of SARS-COV-2*. https://www.who.int/news-room/questions-and-answers/item/coronavirus-disease-(covid-19)-variants-of-sars-cov-2. (Visited on 01/15/2025).

[344] Nicholas G. Davies et al. “Estimated Transmissibility and Impact of SARS-CoV-2 Lineage B.1.1.7 in England”. In: *Science* 372.6538 (Apr. 2021), eabg3055. DOI: 10.1126/science.abg3055. (Visited on 01/15/2025).

[345] *Estimates of Severity and Transmissibility of Novel SARS-CoV-2 Variant 501Y.V2 in South Africa*. https://cmmid.github.io/topics/covid19/sa-novel-variant.html. Jan. 2021. (Visited on 01/15/2025).

[346] Finlay Campbell et al. “Increased Transmissibility and Global Spread of SARS-CoV-2 Variants of Concern as at June 2021”. In: *Eurosurveillance* 26.24 (June 2021), p. 2100509. ISSN: 1025-496X. DOI: 10.2807/1560-7917.ES.2021.26.24.2100509. (Visited on 01/15/2025).

[347] Qianhang Xia et al. “Case Fatality Rates of COVID-19 during Epidemic Periods of Variants of Concern: A Meta-Analysis by Continents”. In: *International Journal of Infectious Diseases* 141 (Apr. 2024). ISSN: 1201-9712. DOI: 10.1016/j.ijid.2024.01.017. (Visited on 01/15/2025).

[348] Aziz Sheikh et al. “SARS-CoV-2 Delta VOC in Scotland: Demographics, Risk of Hospital Admission, and Vaccine Effectiveness”. In: *The Lancet* 397.10293 (June 2021), pp. 2461–2462. ISSN: 0140-6736, 1474-547X. DOI: 10.1016/S0140-6736(21)01358-1. (Visited on 01/15/2025).

[349] David N. Fisman and Ashleigh R. Tuite. “Evaluation of the Relative Virulence of Novel SARS-CoV-2 Variants: A Retrospective Cohort Study in Ontario, Canada”. In: *CMAJ* 193.42 (Oct. 2021), E1619–E1625. ISSN: 0820-3946, 1488-2329. DOI: 10.1503/cmaj.211248. (Visited on 01/15/2025).

[350] *Hospital Admission and Emergency Care Attendance Risk for SARS-CoV-2 Delta (B.1.617.2) Compared with Alpha (B.1.1.7) Variants of Concern: A Cohort Study - The Lancet Infectious Diseases*. https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(21)00475-8/fulltext. (Visited on 01/15/2025).

[351] Srijan Chatterjee et al. “A Detailed Overview of SARS-CoV-2 Omicron: Its Sub-Variants, Mutations and Pathophysiology, Clinical Characteristics, Immunological Landscape, Immune Escape, and Therapies”. In: *Viruses* 15.1 (Jan. 2023), p. 167. ISSN: 1999-4915. DOI: 10.3390/v15010167. (Visited on 01/15/2025).

[352] Carl O. Eriksson et al. “The Association Between Hospital Capacity Strain and Inpatient Outcomes in Highly Developed Countries: A Systematic Review”. In: *Journal of General Internal Medicine* 32.6 (June 2017), pp. 686–696. ISSN: 1525-1497. DOI: 10.1007/s11606-016-3936-3. (Visited on 07/02/2025).

[353] UKRI. *Ensuring Open Research*. https://www.ukri.org/what-we-do/supporting-healthy-research-and-innovation-culture/open-research/. (Visited on 01/16/2025).

[354] Josip Strcic et al. "Open Data and Data Sharing in Articles about COVID-19 Published in Preprint Servers medRxiv and bioRxiv". In: *Scientometrics* 127.5 (May 2022), pp. 2791–2802. ISSN: 1588-2861. DOI: 10.1007/s11192-022-04346-1. (Visited on 01/16/2025).

[355] *EpiEstim for Aggregated Incidence Data*. https://mrc-ide.github.io/EpiEstim/articles/EpiEstim_aggregated (Visited on 01/16/2025).

[356] *MV-EpiEstim*. https://mrc-ide.github.io/EpiEstim/articles/MV_EpiEstim_vignette.html. (Visited on 01/16/2025).

[357] *EpiEstim Vignette*. https://mrc-ide.github.io/EpiEstim/articles/full_EpiEstim_vignette.html. (Visited on 01/16/2025).

[358] *The Epidemiologist R Handbook*. https://www.epirhandbook.com/en/. (Visited on 01/16/2025).

[359] *Alternative Software for Estimating the Reproduction Number*. https://mrc-ide.github.io/EpiEstim/articles/alternative_software.html. (Visited on 01/16/2025).

[360] *Outbreak-Info/Python-Outbreak-Info: Python Package to Access the Genomics and Epidemiology Data and Research Library Metadata Compiled and Standardized on Outbreak.Info*. https://github.com/outbreak-info/python-outbreak-info. (Visited on 01/21/2025).

[361] *Emory-Melody/EpiLearn*. Emory Melody Lab. Jan. 2025. (Visited on 01/21/2025).

[362] WHO. *2022-24 Mpox (Monkeypox) Outbreak: Global Trends*. https://worldhealthorg.shinyapps.io/mpx_global/#53_Case_profile_(overall). (Visited on 01/23/2025).

[363] William M. de Souza and Scott C. Weaver. "Effects of Climate Change and Human Activities on Vector-Borne Diseases". In: *Nature Reviews Microbiology* 22.8 (Aug. 2024), pp. 476–491. ISSN: 1740-1534. DOI: 10.1038/s41579-024-01026-0. (Visited on 01/23/2025).

[364] The White House. *Reevaluating And Realigning United States Foreign Aid*. https://www.whitehouse.gov/presidential-actions/2025/01/reevaluating-and-realigning-united-states-foreign-aid/. Jan. 2025. (Visited on 06/12/2025).

[365] *Global Health Training Centre*. https://globalhealthtrainingcentre.tghn.org/. (Visited on 06/12/2025).

[366] S Abbott. *Temporal Variation in Transmission during the COVID-19 Outbreak in Italy*. CMMID Repository. Mar. 17, 2020. URL: https://cmmid.github.io/topics/covid19/italy-time-varying-transmission.html (visited on 06/02/2021).

[367] Manuel Adrian Acuña-Zegarra, Mario Santana-Cibrian, and Jorge X. Velasco-Hernandez. "Modeling Behavioral Change and COVID-19 Containment in Mexico: A Trade-off between Lockdown and Compliance". In: *Mathematical Biosciences* 325 (July 1, 2020), p. 108370. ISSN: 0025-5564. DOI: 10.1016/j.mbs.2020.108370. URL: https://www.sciencedirect.com/science/article/pii/S0025556420300596 (visited on 06/02/2021).

[368] Carly Adams et al. "Quantifying the Roles of Vomiting, Diarrhea, and Residents vs. Staff in Norovirus Transmission in U.S. Nursing Home Outbreaks". In: *PLOS Computational Biology* 16.3 (Mar. 25, 2020), e1007271. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1007271. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1007271 (visited on 06/02/2021).

[369] OE Adeyeri et al. "Modelling COVID-19 Cases in Nigeria: Forecasts, Uncertainties, Projections and the Link with Weather". In: *Pre-print* (July 14, 2020). DOI: 10.21203/rs.3.rs-41193/v1. URL: https://www.researchsquare.com (visited on 06/02/2021).

[370] Sergios Agapiou et al. "Modeling of Covid-19 Pandemic in Cyprus". Oct. 5, 2020. arXiv: 2010.01927 [stat]. URL: http://arxiv.org/abs/2010.01927 (visited on 06/02/2021).

[371] Mohammad Aghaali et al. "Estimation of the Serial Interval and Basic Reproduction Number of COVID-19 in Qom, Iran, and Three Other Countries: A Data-driven Analysis in the Early Phase of the Outbreak". In: *Transboundary and Emerging Diseases* (June 16, 2020). ISSN: 1865-1674. DOI: 10.1111/tbed.13656. pmid: 32473049. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7300937/ (visited on 06/02/2021).

[372] Tashrik Ahmed. "Improving Disease Surveillance Programs in Liberia". Thesis. Johns Hopkins University, July 15, 2019. URL: https://dspace-prod.mse.jhu.edu/handle/1774.2/62045 (visited on 06/02/2021).

[373] G. S. Amrutha, Abhibhav Sharma, and Anudeepti Sharma. "A Comparative Study to Find a Suitable Model for an Improved Real-Time Monitoring of The Interventions to Contain COVID-19 Outbreak in The High Incidence States of India". In: *medRxiv* (Sept. 15, 2020), p. 2020.09.14.20190447. DOI: 10.1101/2020.09.14.20190447. URL: https://www.medrxiv.org/content/10.1101/2020.09.14.20190447v1 (visited on 06/02/2021).

[374] Alexander Arnon, John Ricco, and Kent Smetters. "Epidemiological and Economic Effects of Lockdown". In: *Brookings Papers on Economic Activity* (2020), pp. 61–99. ISSN: 0007-2303. JSTOR: 27059297. URL: https://www.jstor.org/stable/27059297 (visited on 10/03/2024).

[375] Francisco Arroyo Marioli et al. "Tracking R of COVID-19: A New Real-Time Estimation Using the Kalman Filter". In: *Available at SSRN 3581633* (2020).

[376] Luis Alfredo Bautista Baibás et al. "COVID-19 Effective Reproductive Ratio Determination: An Application, and Analysis of Issues and Influential Factors". In: *medRxiv* (Aug. 12, 2020), p. 2020.07.15.20154039. DOI: 10.1101/2020.07.15.20154039. URL: https://www.medrxiv.org/content/10.1101/2020.07.15.20154039v2 (visited on 06/02/2021).

[377] DJ Darwin R. Bandoy and Bart C. Weimer. "Pandemic Dynamics of COVID-19 Using Epidemic Stage, Instantaneous Reproductive Number and Pathogen Genome Identity (GENI) Score: Modeling Molecular Epidemiology". In: *medRxiv* (Mar. 20, 2020), p. 2020.03.17.20037481. ISSN: 2003-7481. DOI: 10.1101/2020.03.17.20037481. URL: https://www.medrxiv.org/content/10.1101/2020.03.17.20037481v1 (visited on 06/02/2021).

[378] Changjun Bao et al. "COVID-19 Outbreak Following a Single Patient Exposure at an Entertainment Site: An Epidemiological Study". In: *Transboundary and Emerging Diseases* (2020). ISSN: 1865-1674.

[379] Deepankar Basu et al. "A Comprehensive Public Health Evaluation of Lockdown as a Non-Pharmaceutical Intervention on COVID-19 Spread in India: National Trends Masking State Level Variations". In: *medRxiv* (2020).

[380] Christelle Baunez et al. "The Acceleration Index as a Test-Controlled Reproduction Number: Application to COVID-19 in France". In: *medRxiv* (Dec. 2, 2020), p. 2020.12.01.20241570. DOI: 10.1101/2020.12.01.20241570. URL: https://www.medrxiv.org/content/10.1101/2020.12.01.20241570v2 (visited on 06/02/2021).

[381] Amin Bemanian et al. "Investigating the Trajectory of the COVID-19 Outbreak in Milwaukee County and Projected Effects of Relaxed Distancing". In: *WMJ: official publication of the State Medical Society of Wisconsin* 119.2 (2020), pp. 84–90. ISSN: 2379-3961.

[382] Ritwik Bhaduri, Subhrajyoty Roy, and Sankar K Pal. "Rough-Fuzzy CPD: A Gradual Change Point Detection Algorithm". In: *arXiv preprint arXiv:2010.06370* (2020). arXiv: 2010.06370.

[383] Muhammed Umar Bibi, Saad Ahmed Aliyu, and Idris M Jega. "Analysing Secondary Infections of Coronavirus Disease 2019 across the Geopolitical Zones of Nigeria Using Estimated Time Dependent Reproduction Number". In: *Sriwijaya Journal of Environment* 5.2 (2020), pp. 103–110. ISSN: 2527-3809.

[384] Philip Bittihn et al. "Local Measures Enable COVID-19 Containment with Fewer Restrictions Due to Cooperative Effects". In: *EClinicalMedicine* 32 (Feb. 1, 2021). ISSN: 2589-5370. DOI: 10.1016/j.eclinm.2020.100718. URL: https://www.thelancet.com/journals/eclinm/article/PIIS2589-5370(20)30462-4/abstract (visited on 06/02/2021).

[385] AI Blokh et al. "Epidemic potential of COVID-19 in Omsk region and assessment of the anti-epidemic measures. [Russian]". In: *Fundamental and Clinical Medicine* 5.3 (3 Sept. 29, 2020), pp. 8–17. ISSN: 2542-0941. DOI: 10.23946/2500-0764-2020-5-3-8-17. URL: https://fcm.kemsmu.ru/jour/article/view/293 (visited on 06/02/2021).

[386] Yacong Bo et al. "Effectiveness of Non-Pharmaceutical Interventions on COVID-19 Transmission in 190 Countries from 23 January to 13 April 2020". In: *International Journal of Infectious Diseases* 102 (Jan. 2021), pp. 247–253. ISSN: 1201-9712. DOI: 10.1016/j.ijid.2020.10.066. pmid: 33129965. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7598763/ (visited on 06/02/2021).

[387] Katarína Bod'ová et al. "Time-Adjusted Analysis Shows Weak Associations Between BCG Vaccination Policy and COVID-19 Disease Progression". In: *medRxiv* (May 6, 2020), p. 2020.05.01.20087809. DOI: 10.1101/2020.05.01.20087809. URL: https://www.medrxiv.org/content/10.1101/2020.05.01.20087809v1 (visited on 06/02/2021).

[388] Benjamin Born, Alexander M. Dietrich, and Gernot J. Müller. "The Lockdown Effect: A Counterfactual for Sweden". In: *PLOS ONE* 16.4 (Apr. 8, 2021), e0249732. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0249732. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0249732 (visited on 06/02/2021).

[389] Patrick Bryant and Arne Elofsson. "The Effect of Opening up the US on COVID-19 Spread". In: *medRxiv* (July 4, 2020), p. 2020.07.03.20145649. DOI: 10.1101/2020.07.03.20145649. URL: https://www.medrxiv.org/content/10.1101/2020.07.03.20145649v1 (visited on 06/02/2021).

[390] Yaset Caicedo-Ochoa et al. "Effective Reproductive Number Estimation for Initial Stage of COVID-19 Pandemic in Latin American Countries". In: *International Journal of Infectious Diseases* 95 (June 1, 2020), pp. 316–318. ISSN: 1201-9712. DOI: 10.1016/j.ijid.2020.04.069. pmid: 32360941. URL: https://www.ijidonline.com/article/S1201-9712(20)30285-X/abstract (visited on 06/02/2021).

[391] Mauricio Canals et al. "Epidemic Trends, Public Health Response and Health System Capacity: The Chilean Experience in Four Months of the COVID-19 Pandemic". In: *Revista Panamericana de Salud Pública* 44 (Aug. 17, 2020). ISSN: 1020-4989. DOI: 10.26633/RPSP.2020.99. pmid: 32821259. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7429930/ (visited on 06/02/2021).

[392] M Canals. *Informe Covid 19. Chile al 03/05/2020 - Escuela de Salud Pública - Universidad de Chile*. URL: http://www.saludpublica.uchile.cl/noticias/163053/informe-covid-19-chile-al-03052020 (visited on 06/02/2021).

[393] M Canals. *Report Covid 19. Chile to 04/25/2020 - School of Public Health - University of Chile.* URL: http://www.saludpublica.uchile.cl/noticias/162949/informe-covid-19-chile-al-25042020 (visited on 06/02/2021).

[394] Andrea E. Cassidy-Bushrow et al. "Social Distancing during the COVID-19 Pandemic: Quantifying the Practice in Michigan – a "Hotspot State" Early in the Pandemic – Using a Volunteer-Based Online Survey". In: *BMC Public Health* 21.1 (Jan. 29, 2021), p. 245. ISSN: 1471-2458. DOI: 10.1186/s12889-021-10287-w. URL: https://doi.org/10.1186/s12889-021-10287-w (visited on 06/02/2021).

[395] Robert Challen et al. "Meta-Analysis of the SARS-CoV-2 Serial Interval and the Impact of Parameter Uncertainty on the COVID-19 Reproduction Number". In: *medRxiv* (Nov. 20, 2020), p. 2020.11.17.20231548. DOI: 10.1101/2020.11.17.20231548. URL: https://www.medrxiv.org/content/10.1101/2020.11.17.20231548v2 (visited on 06/02/2021).

[396] Robert Challen et al. "Estimates of Regional Infectivity of COVID-19 in the United Kingdom Following Imposition of Social Distancing Measures". In: *medRxiv* (Nov. 13, 2020), p. 2020.04.13.20062760. DOI: 10.1101/2020.04.13.20062760. URL: https://www.medrxiv.org/content/10.1101/2020.04.13.20062760v3 (visited on 06/02/2021).

[397] S. D. Chamberlain et al. "Real-Time Detection of COVID-19 Epicenters within the United States Using a Network of Smart Thermometers". In: *medRxiv* (Apr. 10, 2020), p. 2020.04.06.20039909. DOI: 10.1101/2020.04.06.20039909. URL: https://www.medrxiv.org/content/10.1101/2020.04.06.20039909v1 (visited on 06/02/2021).

[398] Yung-Wai Desmond Chan et al. "Transmission Dynamics, Serial Interval and Epidemiology of COVID-19 Diseases in Hong Kong under Different Control Measures". In: *Wellcome Open Research* 5 (Nov. 9, 2020), p. 91. ISSN: 2398-502X. DOI: 10.12688/wellcomeopenres.15896.2. URL: https://wellcomeopenresearch.org/articles/5-91/v2 (visited on 06/02/2021).

[399] Luis Fernando Chaves et al. "COVID-19 Basic Reproduction Number and Assessment of Initial Suppression Policies in Costa Rica". In: *Mathematical Modelling of Natural Phenomena* 15 (2020), p. 32. ISSN: 0973-5348, 1760-6101. DOI: 10.1051/mmnp/2020019. URL: https://www.mmnp-journal.org/articles/mmnp/abs/2020/01/mmnp200065/mmnp200065.html (visited on 06/02/2021).

[400] Tiange Chen et al. "Quantitative Effects of Entry Restrictions and Travel Quarantine on the Next Wave of COVID-19: Case Studies of China and Singapore". In: *SSRN Electronic Journal* (2020). ISSN: 1556-5068. DOI: 10.2139/ssrn.3622386. URL: https://www.ssrn.com/abstract=3622386 (visited on 06/02/2021).

[401] Qing Cheng et al. “Heterogeneity and Effectiveness Analysis of COVID-19 Prevention and Control in Major Cities in China through Time-Varying Reproduction Number Estimation”. In: *Scientific Reports* 10.1 (1 Dec. 15, 2020), p. 21953. ISSN: 2045-2322. DOI: 10.1038/s41598-020-79063-x. URL: https://www.nature.com/articles/s41598-020-79063-x (visited on 06/02/2021).

[402] Nalini Chintalapudi et al. “COVID-19 Outbreak Reproduction Number Estimations and Forecasting in Marche, Italy”. In: *International Journal of Infectious Diseases* 96 (July 1, 2020), pp. 327–333. ISSN: 1201-9712. DOI: 10.1016/j.ijid.2020.05.029. pmid: 32437930. URL: https://www.ijidonline.com/article/S1201-9712(20)30336-2/abstract (visited on 06/02/2021).

[403] Ka Chun Chong et al. “Monitoring Disease Transmissibility of 2019 Novel Coronavirus Disease in Zhejiang, China”. In: *International Journal of Infectious Diseases* 96 (July 2020), pp. 128–130. ISSN: 1201-9712. DOI: 10.1016/j.ijid.2020.04.036. pmid: 32417744. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7227576/ (visited on 06/02/2021).

[404] Ka Chun Chong et al. “Transmissibility of Coronavirus Disease 2019 in Chinese Cities with Different Dynamics of Imported Cases”. In: *PeerJ* 8 (Nov. 6, 2020). ISSN: 2167-8359. DOI: 10.7717/peerj.10350. pmid: 33194456. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7651459/ (visited on 06/02/2021).

[405] Mario Santana Cibrian, Jorge Velasco-Hernandez, and Ruth Corona-Moreno. *Reporte al 7 de Septiembre de 2020 de La Estimación Del Semáforo Epidemiológico de Querétaro*. Instituto de Matemáticas Unidad Juriquilla UNAM (IM), Sept. 14, 2020.

[406] Ruth Corona-Moreno et al. *Nowcasting Para El Número Reproductivo Instantáneo Rt Para La República Mexicana, La Ciudad de México, Sus Alcaldías, Algunas Ciudades y Estados de México Usando Datos Hasta El 08 de Agosto*. Nodo Multidisciplinario de Matematicas Aplicadas: Instituto de Matematicas UNAM Juriquilla, Sept. 9, 2020. URL: https://www.researchgate.net/publication/344178938_Nowcasting_R_t_CDMX_15Agosto.

[407] Ruth Corona-Moreno, Nancy Gonzalez, and Jorge Velasco-Hernandez. *Tendencias de Rts de La Epidemia de COVID-19 En México: Reporte 20 de Septiembre 2020*. Nodo Multidisciplinario de Matematicas Aplicadas: Instituto de Matematicas UNAM Juriquilla, Sept. 23, 2020. URL: https://www.researchgate.net/profile/Jorge-Velasco-Hernandez/publication/344347054_Rts_CDMX_Edos_Ciudades_06Septiembre_Reported_21Septiembre/links/5f6ab61192851c14bc8e2eb9/Rts-CDMX-Edos-Ciudades-06Septiembre-Reported-21Septiembre.pdf.

[408] Ruth Corona-Moreno, Nancy Gonzalez, and Jorge Velasco-Hernandez. *Tendencias de Rts de La Epidemia de COVID-19 En México: Reporte 05 de Octubre 2020*. Nodo Multidisciplinario de Matematicas Aplicadas: Instituto de Matematicas UNAM Juriquilla, Oct. 7, 2020. URL: https://www.researchgate.net/publication/344526176_Rts_CDMX_Edos_Ciudades_19Septiembre_Reported_05Octubre.

[409] Ruth Corona-Moreno, Nancy Gonzalez, and Jorge Velasco-Hernandez. *Tendencias de La Epidemia de COVID-19 En México: Reporte 9 de Noviembre 2020*. Nodo Multidisciplinario de Matematicas Aplicadas: Instituto de Matematicas UNAM Juriquilla, Nov. 13, 2020. URL: https://www.researchgate.net/publication/345815689_Rts_CDMX_Edos_Ciudades_24Octubre_Reported_09Noviembre.

[410] Ruth Corona-Moreno, Nancy Gonzalez, and Jorge Velasco-Hernandez. *Tendencias de La Epidemia de COVID-19 En México: Reporte 18 de Octubre 2020*. Nodo Multidisciplinario de Matematicas Aplicadas: Instituto de Matematicas UNAM Juriquilla, Oct. 21, 2020. URL: https://www.researchgate.net/profile/Jorge-Velasco-Hernandez/publication/344814471_Rts_CDMX_Edos_Ciudades_04Octubre_Reportado_19Octubre/links/5f919c0292851c14bcde7bb2/Rts-CDMX-Edos-Ciudades-04Octubre-Reportado-19Octubre.pdf.

[411] Ruth Corona-Moreno et al. *Tendencias de Rts de La Epidemia de COVID-19 En México: Reporte 31 de Agosto 2020*. Nodo Multidisciplinario de Matematicas Aplicadas: Instituto de Matematicas UNAM Juriquilla, Sept. 14, 2020. URL: https://www.researchgate.net/publication/344233508_Rts_CDMX_Edos_Ciudades_29Agosto_Reported_13Septiembre.

[412] Ruth Corona-Moreno, Mario Santana Cibrian, and Jorge Velasco-Hernandez. *Reporte al 19 de Septiembre de 2020 de La Estimación Del Semáforo Epidemiológico de Querétaro*. Nodo Multidisciplinario de Matematicas Aplicadas: Instituto de Matematicas UNAM Juriquilla, Sept. 19, 2020. URL: https://www.researchgate.net/publication/344347160_Reporte_al_19_de_septiembre_de_2020_de_la_estimacion_del_semaforo_epidemiologico_de_Queretaro.

[413] Ruth Corona-Moreno, Mario Santana Cibrian, and Jorge Velasco-Hernandez. *Reporte al 5 de Octubre de 2020 de La Estimación Del Semáforo Epidemiológico de Querétaro*. Nodo Multidisciplinario de Matematicas Aplicadas: Instituto de Matematicas UNAM Juriquilla, Oct. 5, 2020. URL: https://www.researchgate.net/publication/344526419_Reporte_al_5_de_octubre_de_2020_de_la_estimacion_del_semaforo_epidemiologico_de_Queretaro.

[414] Ruth Corona-Moreno, Mario Santana Cibrian, and Jorge Velasco-Hernandez. *Estimación Del Semáforo Epidemiológico de Querétaro: Reporte al 19 de Octubre de 2020*. Nodo Multidisciplinario de Matematicas Aplicadas: Instituto de Matematicas UNAM Juriquilla, Oct. 19, 2020.

[415] Cristóbal Cuadrado et al. “Impact of Small-Area Lockdowns for the Control of the COVID-19 Pandemic”. In: *medRxiv* (May 9, 2020), p. 2020.05.05.20092106. DOI: 10.1101/2020.05.05.20092106. URL: https://www.medrxiv.org/content/10.1101/2020.05.05.20092106v1 (visited on 06/02/2021).

[416] Tingting Cui et al. “Chinese Residents’ Perceptions of COVID-19 During the Pandemic: Online Cross-sectional Survey Study”. In: *Journal of Medical Internet Research* 22.11 (Nov. 25, 2020). ISSN: 1439-4456. DOI: 10.2196/21672. pmid: 33152684. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7690970/ (visited on 06/02/2021).

[417] Dimitrije Čvokić. “Анализа Ситуације и Предвиђања у Вези Са Епидемијом САРС-КоВ-2 Вируса у Републици Српској На Дан 27. Мај 2020. [Serbian]”. In: (May 21, 2020). DOI: 10.31219/osf.io/4yrtu. URL: https://osf.io/4yrtu/ (visited on 06/03/2021).

[418] Samy Dana et al. *Brazilian Modeling of COVID-19(BRAM-COD): A Bayesian Monte Carlo Approach for COVID-19 Spread in a Limited Data Set Context*. May 17, 2020. DOI: 10.1101/2020.04.29.20081174. URL: https://www.medrxiv.org/content/10.1101/2020.04.29.20081174v3 (visited on 10/03/2024). Pre-published.

[419] Samath Dharmaratne et al. “Estimation of the Basic Reproduction Number (R0) for the Novel Coronavirus Disease in Sri Lanka”. In: *Virology Journal* 17.1 (Oct. 7, 2020), p. 144. ISSN: 1743-422X. DOI: 10.1186/s12985-020-01411-0. URL: https://doi.org/10.1186/s12985-020-01411-0 (visited on 06/03/2021).

[420] Amy Dighe et al. “Response to COVID-19 in South Korea and Implications for Lifting Stringent Interventions”. In: *BMC Medicine* 18.1 (Oct. 9, 2020), p. 321. ISSN: 1741-7015. DOI: 10.1186/s12916-020-01791-8. URL: https://doi.org/10.1186/s12916-020-01791-8 (visited on 06/03/2021).

[421] José Alexandre Felizola Diniz-Filho et al. “The Effective Reproductive Number (Rt) of COVID-19 and Its Relationship with Social Distancing”. In: *medRxiv* (Dec. 29, 2020), p. 2020.07.28.20163493. DOI: 10.1101/2020.07.28.20163493. URL: https://www.medrxiv.org/content/10.1101/2020.07.28.20163493v2 (visited on 06/03/2021).

[422] Hasan Durmuş, Mehmet Enes Gökler, and Selma Metintaş. “The Effectiveness of Community-based Social Distancing for Mitigating the Spread of the COVID-19 Pandemic in Turkey”. In: *Journal of Preventive Medicine and Public Health* 53.6 (Nov. 2020), pp. 397–404. ISSN: 1975-8375. DOI: 10.3961/jpmph.20.381. pmid: 33296579. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7733747/ (visited on 06/03/2021).

[423] Francisco H. C. Felix and Juvenia Fontenele. “Instantaneous R Calculation for COVID-19 Epidemic in Brazil”. In: *medRxiv* (Apr. 29, 2020), p. 2020.04.23.20077172. DOI: 10.1101/2020.04.

23.20077172. URL: https://www.medrxiv.org/content/10.1101/2020.04.23.20077172v1 (visited on 06/03/2021).

[424] Denise Stéphanie de Almeida Ferreira et al. "Transitivity of COVID-19, Based on Parameters R_0 and R_((t)) in Pombos/PE Municipality, Brazil". In: *Research, Society and Development* 9.10 (10 Sept. 20, 2020), e1139108441–e1139108441. ISSN: 2525-3409. DOI: 10.33448/rsd-v9i10.8441. URL: https://rsdjournal.org/index.php/rsd/article/view/8441 (visited on 06/04/2021).

[425] Força-Tarefa de Modelagem da COVID-19. *Análise Do Efeito Das Medidas de Contenção à Propagação Da COVID-19 Em Belo Horizonte (23/03 a 29/03)*. Federal University of Minas Gerais, Apr. 2, 2020. URL: https://ufmg.br/storage/2/5/a/7/25a7163c7fb5575ab6d81b5a05bfd844_15863100172762_298779967.pdf (visited on 01/11/2021).

[426] Oliver Forsyth. "Uncertainty and Lockdown in COVID-19: An Incomplete Information SIR Model". In: *Covid Economics* 52 (Oct. 15, 2020), pp. 1–38. URL: https://cepr.org/file/9808.

[427] Danilo Franco et al. "Early Transmission Dynamics, Spread, and Genomic Characterization of SARS-CoV-2 in Panama". In: *Emerging Infectious Diseases* 27.2 (Feb. 2021), pp. 612–615. ISSN: 1080-6040. DOI: 10.3201/eid2702.203767. pmid: 33496228. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7853578/ (visited on 06/04/2021).

[428] Oscar Frausto-Martínez et al. "COVID-19, Storms, and Floods: Impacts of Tropical Storm Cristobal in the Western Sector of the Yucatan Peninsula, Mexico". In: *Sustainability* 12.23 (23 Jan. 2020), p. 9925. DOI: 10.3390/su12239925. URL: https://www.mdpi.com/2071-1050/12/23/9925 (visited on 06/04/2021).

[429] Jucarlos Rufino de Freitas et al. "Estimativa do número efetivo de reprodução de SARS-CoV-2 em Vitória de Santo Antão/PE, Brasil". In: *Research, Society and Development* 9.9 (9 Sept. 7, 2020), e794997922–e794997922. ISSN: 2525-3409. DOI: 10.33448/rsd-v9i9.7922. URL: https://rsdjournal.org/index.php/rsd/article/view/7922 (visited on 06/04/2021).

[430] Isaac Chun-Hai Fung et al. "SARS-CoV-2 Transmission in Alberta, British Columbia, and Ontario, Canada, January 1-July 6, 2020". In: *medRxiv* (Aug. 26, 2020), p. 2020.07.18.20156992. DOI: 10.1101/2020.07.18.20156992. URL: https://www.medrxiv.org/content/10.1101/2020.07.18.20156992v2 (visited on 06/04/2021).

[431] Enrique Galbán-García and Pedro Más-Bermejo. "COVID-19 in Cuba: Assessing the National Response". In: *MEDICC Review* 22.4 (Oct. 10, 2020), pp. 29–34. URL: https://mediccreview.org/covid-19-in-cuba-assessing-the-national-response/ (visited on 06/04/2021).

[432] Song Gao et al. "Association of Mobile Phone Location Data Indications of Travel and Stay-at-Home Mandates With COVID-19 Infection Rates in the US". In: *JAMA Network Open* 3.9 (Sept. 8, 2020), e2020485. ISSN: 2574-3805. DOI: 10.1001/jamanetworkopen.2020.20485. URL: https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2770249 (visited on 06/04/2021).

[433] Aurnab Ghose et al. "Community Prevalence of Antibodies to SARS-CoV-2 and Correlates of Protective Immunity in an Indian Metropolitan City". In: *medRxiv* (Nov. 30, 2020), p. 2020.11.17.20228155. ISSN: 2022-8155. DOI: 10.1101/2020.11.17.20228155. URL: https://www.medrxiv.org/content/10.1101/2020.11.17.20228155v2 (visited on 06/04/2021).

[434] Governo do Estado do Rio Grande do Norte Secretaria de Estado da Saúde Pública. *Recomendações do comitê de especialistas da SESAP-RN para o enfrentamento da pandemia pela COVID-19 [Recommendation No. 023/2021]*. URL: https://portalcovid19.saude.rn.gov.br/wp-content/uploads/2020/05/recomendacoes-comite-23.pdf (visited on 06/04/2021).

[435] Nicolò Gozzi et al. "Estimating the Effect of Social Inequalities in the Mitigation of COVID-19 across Communities in Santiago de Chile". In: *medRxiv* (Oct. 13, 2020), p. 2020.10.08.20204750. DOI: 10.1101/2020.10.08.20204750. URL: https://www.medrxiv.org/content/10.1101/2020.10.08.20204750v1 (visited on 06/04/2021).

[436] Patrick Grice, Stephen Grice, and Richard S Laugesen. *Calculating the Effective Reproduction Number for COVID-19 Using a New Process for Various Countries*. CloseAssociate and University of Illinois at Urbana-Champaign, May 26, 2020, p. 19. URL: https://e2a6b380-4bd5-43a8-b33f-68e981870108.usrfiles.com/ugd/e2a6b3_093ee4ee50fc442b92c29614902487c1.pdf.

[437] Alex Guevara. "Mathetical Modeling COVID 19 Loja Province". In: *Pre-print* (Apr. 22, 2020). URL: https://www.researchgate.net/publication/340846424_Mathetical_modeling_COVID_19_Loja_province (visited on 01/30/2021).

[438] Mohak Gupta et al. "Transmission Dynamics of the COVID-19 Epidemic in India and Modeling Optimal Lockdown Exit Strategies". In: *International Journal of Infectious Diseases* 103 (Feb. 1, 2021), pp. 579–589. ISSN: 1201-9712. DOI: 10.1016/j.ijid.2020.11.206. URL: https://www.sciencedirect.com/science/article/pii/S1201971220325224 (visited on 06/04/2021).

[439] Giorgio Guzzetta et al. "Impact of a Nationwide Lockdown on SARS-CoV-2 Transmissibility, Italy - Volume 27, Number 1—January 2021 - Emerging Infectious Diseases Journal - CDC". In: *Emerging Infectious Diseases* 27.1 (), pp. 267–270. DOI: 10.3201/eid2701.202114. URL: https://wwwnc.cdc.gov/eid/article/27/1/20-2114_article (visited on 06/04/2021).

[440] W. S. Hart, P. K. Maini, and R. N. Thompson. "High Infectiousness Immediately before COVID-19 Symptom Onset Highlights the Importance of Contact Tracing". In: *medRxiv* (Nov. 23, 2020),

p. 2020.11.20.20235754. DOI: 10.1101/2020.11.20.20235754. URL: https://www.medrxiv.org/content/10.1101/2020.11.20.20235754v1 (visited on 06/04/2021).

[441] Agus Hasan et al. "A New Estimation Method for COVID-19 Time-Varying Reproduction Number Using Active Cases". In: *medRxiv* (Jan. 30, 2021), p. 2020.06.28.20142158. DOI: 10.1101/2020.06.28.20142158. URL: https://www.medrxiv.org/content/10.1101/2020.06.28.20142158v2 (visited on 06/04/2021).

[442] Nils Haug et al. "Ranking the Effectiveness of Worldwide COVID-19 Government Interventions". In: *Nature Human Behaviour* 4.12 (12 Dec. 2020), pp. 1303–1312. ISSN: 2397-3374. DOI: 10.1038/s41562-020-01009-0. URL: https://www.nature.com/articles/s41562-020-01009-0 (visited on 06/04/2021).

[443] Thomas Hotz et al. "Monitoring the Spread of COVID-19 by Estimating Reproduction Numbers over Time". Apr. 18, 2020. arXiv: 2004.08557 [q-bio, stat]. URL: http://arxiv.org/abs/2004.08557 (visited on 06/04/2021).

[444] Fu-Chang Hu. "The Estimated Time-Varying Reproduction Numbers during the Ongoing Pandemic of the Coronavirus Disease 2019 (COVID-19) in 12 Selected Countries Outside China". In: *medRxiv* (May 14, 2020), p. 2020.05.10.20097154. DOI: 10.1101/2020.05.10.20097154. URL: https://www.medrxiv.org/content/10.1101/2020.05.10.20097154v1 (visited on 06/04/2021).

[445] Mirsada Hukić and Mirza Ponjavić. *COVID-19 Pandemic in Bosnia and Herzegovina: March – June 2020*. Academy of Sciences and Arts of Bosnia and Herzegovina, Aug. 2020. ISBN: 978-9926-410-58-2. URL: https://publications.anubih.ba/handle/123456789/710 (visited on 06/04/2021).

[446] Mirsada Hukic et al. "SARS-CoV-2 Virus Outbreak and the Emergency Public Health Measures in Bosnia and Herzegovina: January – July 2020". In: *Bosnian Journal of Basic Medical Sciences* 21.1 (Feb. 2021), pp. 111–116. ISSN: 1512-8601. DOI: 10.17305/bjbms.2020.5081. pmid: 33091331. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7861623/ (visited on 06/04/2021).

[447] Yoshio Kajitani and Michinori Hatayama. "Explaining the Effective Reproduction Number of COVID-19 through Mobility and Enterprise Statistics: Evidence from the First Wave in Japan". In: *medRxiv* (Dec. 4, 2020), p. 2020.10.08.20209643. DOI: 10.1101/2020.10.08.20209643. URL: https://www.medrxiv.org/content/10.1101/2020.10.08.20209643v2 (visited on 06/04/2021).

[448] Michelle Kendall et al. "Epidemiological Changes on the Isle of Wight after the Launch of the NHS Test and Trace Programme: A Preliminary Analysis". In: *The Lancet Digital Health*

2.12 (Dec. 1, 2020), e658–e666. ISSN: 2589-7500. DOI: 10.1016/S2589-7500(20)30241-7. pmid: 33078140. URL: https://www.thelancet.com/journals/landig/article/PIIS2589-7500(20)30241-7/abstract (visited on 06/04/2021).

[449] Wee Chian Koh, Lin Naing, and Justin Wong. "Estimating the Impact of Physical Distancing Measures in Containing COVID-19: An Empirical Analysis". In: *International Journal of Infectious Diseases* 100 (Nov. 1, 2020), pp. 42–49. ISSN: 1201-9712. DOI: 10.1016/j.ijid.2020.08.026. pmid: 32800859. URL: https://www.ijidonline.com/article/S1201-9712(20)30655-X/abstract (visited on 06/04/2021).

[450] Kaushalendra Kumar, Wahengbam Bigyananda Meitei, and Abhishek Singh. *Projecting the Future Trajectory of COVID-19 Infections in India Using the Susceptible-Infected-Recovered (SIR) Model*. An Analytical Paper for Policymakers 7. International Institute for Population Sciences [Ministry of Health & Family Welfare, Government of India], May 10, 2020. URL: https://www.iipsindia.ac.in/sites/default/files/iips_covid19_pfti.pdf.

[451] Christopher K C Lai et al. "Epidemiological Characteristics of the First 100 Cases of Coronavirus Disease 2019 (COVID-19) in Hong Kong Special Administrative Region, China, a City with a Stringent Containment Policy". In: *International Journal of Epidemiology* 49.4 (Aug. 1, 2020), pp. 1096–1105. ISSN: 0300-5771. DOI: 10.1093/ije/dyaa106. URL: https://doi.org/10.1093/ije/dyaa106 (visited on 06/04/2021).

[452] Ramanan Laxminarayan et al. "Epidemiology and Transmission Dynamics of COVID-19 in Two Indian States". In: *Science* 370.6517 (Nov. 6, 2020), pp. 691–697. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.abd7672. pmid: 33154136. URL: https://science.sciencemag.org/content/370/6517/691 (visited on 06/04/2021).

[453] Whanhee Lee et al. "COVID-19 in South Korea: Epidemiological and Spatiotemporal Patterns of the Spread and the Role of Aggressive Diagnostic Tests in the Early Phase". In: *International Journal of Epidemiology* 49.4 (Aug. 1, 2020), pp. 1106–1116. ISSN: 0300-5771. DOI: 10.1093/ije/dyaa119. URL: https://doi.org/10.1093/ije/dyaa119 (visited on 06/04/2021).

[454] Kathy Leung et al. "First-Wave COVID-19 Transmissibility and Severity in China Outside Hubei after Control Measures, and Second-Wave Scenario Planning: A Modelling Impact Assessment". In: *The Lancet* 395.10233 (Apr. 25, 2020), pp. 1382–1393. ISSN: 0140-6736, 1474-547X. DOI: 10.1016/S0140-6736(20)30746-7. pmid: 32277878. URL: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)30746-7/abstract (visited on 06/04/2021).

[455] Kathy Leung et al. "No Detectable Surge in SARS-CoV-2 Transmission Attributable to the April 7, 2020 Wisconsin Election". In: *American Journal of Public Health* 110.8 (Aug. 2020), pp. 1169–1170. ISSN: 0090-0036. DOI: 10.2105/AJPH.2020.305770. pmid: 32552029. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7349432/ (visited on 06/04/2021).

[456] Kathy Leung, Joseph T. Wu, and Gabriel M. Leung. “Real-Time Tracking and Prediction of COVID-19 Infection Using Digital Proxies of Population Mobility and Mixing”. In: *Nature Communications* 12.1 (1 Mar. 8, 2021), p. 1501. ISSN: 2041-1723. DOI: 10.1038/s41467-021-21776-2. URL: https://www.nature.com/articles/s41467-021-21776-2 (visited on 06/04/2021).

[457] Marc Lipsitch, Keya Joshi, and Sarah E. Cobey. “Comment on Pan A, Liu L, Wang C, et al., "Association of Public Health Interventions With the Epidemiology of the COVID-19 Outbreak in Wuhan, China," JAMA, Published Online April 10, 2020, Doi:10.1001/Jama.2020.6130.” In: (Apr. 21, 2020). URL: https://dash.harvard.edu/handle/1/42660128 (visited on 06/04/2021).

[458] Quan-Hui Liu et al. “The COVID-19 Outbreak in Sichuan, China: Epidemiology and Impact of Interventions”. In: *PLOS Computational Biology* 16.12 (Dec. 28, 2020), e1008467. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1008467. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1008467 (visited on 06/04/2021).

[459] Tao Liu et al. “Time-Varying Transmission Dynamics of Novel Coronavirus Pneumonia in China”. In: *bioRxiv* (Feb. 13, 2020), p. 2020.01.25.919787. DOI: 10.1101/2020.01.25.919787. URL: https://www.biorxiv.org/content/10.1101/2020.01.25.919787v2 (visited on 06/04/2021).

[460] Wei Liu et al. “Geographic Information System Methods Tell Spatiotemporal Transmission and Its Drivers of COVID-19 in Wuhan, China at Street-Level”. In: *Pre-print* (Nov. 11, 2020). DOI: 10.21203/rs.3.rs-104986/v1. URL: https://www.researchsquare.com/article/rs-104986/v1 (visited on 06/04/2021).

[461] Xiaoshuang Liu et al. “Differential Impact of Non-Pharmaceutical Public Health Interventions on COVID-19 Epidemics in the United States”. In: *Pre-print* (Dec. 28, 2020). DOI: 10.21203/rs.3.rs-60056/v3. URL: https://www.researchsquare.com/article/rs-60056/v3 (visited on 06/04/2021).

[462] Zhu Liu, Hao Yin, and Daniel Kammen. “The Interaction between Population Age Structure and Policy Interventions on the Spread of COVID-19”. In: *Pre-print* (Aug. 24, 2020). DOI: 10.21203/rs.3.rs-57858/v1. URL: https://www.researchsquare.com/article/rs-57858/v1 (visited on 06/04/2021).

[463] Fred G. Manrique-Abril, Cristian Téllez-Piñerez, and Mario Pacheco-López. “Estimation of Time-Varying Reproduction Numbers of COVID-19 in American Countries with Regards to Non-Pharmacological Interventions”. In: *F1000Research* 9 (July 31, 2020), p. 868. ISSN: 2046-1402. DOI: 10.12688/f1000research.25309.1. URL: https://f1000research.com/articles/9-868/v1 (visited on 06/04/2021).

[464] Mohamed Masoud, Gihan Gewaifel, and Nahla Gamaleldin. "Transmissibility and Mortality Trends of COVID-19 Epidemic in Egypt". In: *Alexandria Journal of Medicine* 56.1 (Jan. 1, 2020), pp. 189–195. ISSN: 2090-5068. DOI: 10.1080/20905068.2020.1845442. URL: https://doi.org/10.1080/20905068.2020.1845442 (visited on 06/04/2021).

[465] José Menéndez. "A Poor-Man's Approach to the Effective Reproduction Number: The COVID-19 Case". In: *medRxiv* (Apr. 27, 2020), p. 2020.04.22.20076430. DOI: 10.1101/2020.04.22.20076430. URL: https://www.medrxiv.org/content/10.1101/2020.04.22.20076430v1 (visited on 06/04/2021).

[466] Andrew C. Miller et al. "Statistical Deconvolution for Inference of Infection Time Series". In: *medRxiv* (Oct. 20, 2020), p. 2020.10.16.20212753. DOI: 10.1101/2020.10.16.20212753. URL: https://www.medrxiv.org/content/10.1101/2020.10.16.20212753v1 (visited on 06/04/2021).

[467] Octavio Miramontes. *Entendamos el COVID-19 en México*. Universidad Nacional Autónoma de México, Sept. 4, 2020, p. 37. URL: http://scifunam.fisica.unam.mx/mir/corona19/covid19.pdf.

[468] Giovenale Moirano et al. "Approaches to Daily Monitoring of the SARS-CoV-2 Outbreak in Northern Italy". In: *Frontiers in Public Health* 8 (2020). ISSN: 2296-2565. DOI: 10.3389/fpubh.2020.00222. URL: https://www.frontiersin.org/articles/10.3389/fpubh.2020.00222/full (visited on 06/04/2021).

[469] Giovenale Moirano, Maurizio Schmid, and Francesco Barone-Adesi. "Short-Term Effects of Mitigation Measures for the Containment of the COVID-19 Outbreak: An Experience From Northern Italy". In: *Disaster Medicine and Public Health Preparedness* (), pp. 1–2. ISSN: 1935-7893. DOI: 10.1017/dmp.2020.119. pmid: 32327001. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7211796/ (visited on 06/04/2021).

[470] Seong-Geun Moon et al. "Time-Variant Reproductive Number of COVID-19 in Seoul, Korea". In: *Epidemiology and Health* 42 (June 28, 2020). ISSN: 2092-7193. DOI: 10.4178/epih.e2020047. pmid: 32660220. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7644928/ (visited on 06/04/2021).

[471] Xavier Mora. "El nombre de reproducció de la COVID-19 i el model SIR. L'efecte dels retards de comptabilització [Catalan]". In: *Materials matemàtics* 2020.2 (May 25, 2020), p. 19. ISSN: 1887-1097. URL: https://mat.uab.cat/web/matmat/wp-content/uploads/sites/23/2020/06/v2020n02.pdf.

[472] Pilar Moreno et al. "An Effective COVID-19 Response in South America: The Uruguayan Conundrum". In: *medRxiv* (July 27, 2020), p. 2020.07.24.20161802. DOI: 10.1101/2020.07.24.

20161802. URL: https://www.medrxiv.org/content/10.1101/2020.07.24.20161802v1 (visited on 06/04/2021).

[473] Christopher P. Morley et al. "Social Distancing Metrics and Estimates of SARS-CoV-2 Transmission Rates: Associations Between Mobile Telephone Data Tracking and R". In: *Journal of Public Health Management and Practice* 26.6 (November/December 2020), pp. 606–612. ISSN: 1078-4659. DOI: 10.1097/PHH.0000000000001240. URL: https://journals.lww.com/jphmp/Fulltext/2020/11000/Social_Distancing_Metrics_and_Estimates_of.17.aspx (visited on 06/04/2021).

[474] Kamalich Muniz-Rodriguez et al. "Time-Varying Reproduction Numbers of COVID-19 in Georgia, USA, March 2-June 14, 2020". In: *medRxiv* (July 28, 2020), p. 2020.05.19.20107219. DOI: 10.1101/2020.05.19.20107219. URL: https://www.medrxiv.org/content/10.1101/2020.05.19.20107219v3 (visited on 06/04/2021).

[475] Salihu S. Musa et al. "Estimation of Exponential Growth Rate and Basic Reproduction Number of the Coronavirus Disease 2019 (COVID-19) in Africa". In: *Infectious Diseases of Poverty* 9.1 (July 16, 2020), p. 96. ISSN: 2049-9957. DOI: 10.1186/s40249-020-00718-y. URL: https://doi.org/10.1186/s40249-020-00718-y (visited on 06/04/2021).

[476] F. Najafi et al. "Serial Interval and Time-Varying Reproduction Number Estimation for COVID-19 in Western Iran". In: *New Microbes and New Infections* 36 (June 14, 2020). ISSN: 2052-2975. DOI: 10.1016/j.nmni.2020.100715. pmid: 32566233. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7293842/ (visited on 06/04/2021).

[477] Huy G. Nguyen and Tuan V. Nguyen. "An Epidemiologic Profile of COVID-19 Patients in Vietnam". In: *medRxiv* (Apr. 15, 2020), p. 2020.04.10.20061226. ISSN: 2006-1226. DOI: 10.1101/2020.04.10.20061226. URL: https://www.medrxiv.org/content/10.1101/2020.04.10.20061226v1 (visited on 06/04/2021).

[478] Ontario Agency for Health Protection and Promotion (Public Health Ontario). *Enhanced Epidemiological Summary: COVID-19 Regional Incidence and Time to Case Notification in Ontario*. Ontario Agency for Health Protection and Promotion: Public Health Ontario, p. 181. URL: https://www.publichealthontario.ca/-/media/documents/ncov/epi/covid-19-regional-epi-summary-report.pdf?la=en.

[479] Mutiu Abimbola Oyinlola, Tolulope Osayomi, and Oluwatosin Adeniyi. "Empirical Modelling of Confirmed COVID-19 Cases in Nigeria: Forecasts and Implications". In: *Social Sciences & Humanities Open [Pre-print]* (May 12, 2020), p. 16. DOI: 10.2139/ssrn.3596095. URL: https://papers.ssrn.com/abstract=3596095 (visited on 06/04/2021).

[480] Gustavo Nicolás Páez et al. *Alternativas En La Estimación Del Número Básico de Reproducción de Una Enfermedad*. Sociedad Colombiana de Matemáticas, p. 23. URL: http://scm.org.co/wp-content/uploads/2020/09/Paez-et-al.-Alternativas-en-estimaci%C3%B3n-del-n%C3%BAmero-de-reproducci%C3%B3n-.pdf (visited on 01/30/2021).

[481] An Pan et al. "Association of Public Health Interventions With the Epidemiology of the COVID-19 Outbreak in Wuhan, China". In: *JAMA* 323.19 (May 19, 2020), p. 1915. ISSN: 0098-7484. DOI: 10.1001/jama.2020.6130. URL: https://jamanetwork.com/journals/jama/fullarticle/2764658 (visited on 06/04/2021).

[482] Sang Woo Park et al. "Potential Role of Social Distancing in Mitigating Spread of Coronavirus Disease, South Korea". In: *Emerging Infectious Diseases* 26.11 (Nov. 2020), pp. 2697–2700. ISSN: 1080-6040. DOI: 10.3201/eid2611.201099. pmid: 32795385. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7588540/ (visited on 06/04/2021).

[483] Hector Pasten and Jorge Castillo Sepúlveda. *El pedal acelerador de la epidemia*. Facultad de Matemáticas: Pontificia Universidad Católica de Chile, p. 4. URL: http://www.mat.uc.cl/~hector.pasten/preprints/Re.pdf (visited on 01/26/2021).

[484] Angel Paternina-Caicedo et al. "Social Interventions Can Lower COVID-19 Deaths in Middle-Income Countries". In: *medRxiv* (Apr. 23, 2020), p. 2020.04.16.20063727. DOI: 10.1101/2020.04.16.20063727. URL: https://www.medrxiv.org/content/10.1101/2020.04.16.20063727v2 (visited on 06/04/2021).

[485] Zhihang Peng et al. "Linking Key Intervention Timings to Rapid Declining Effective Reproduction Number to Quantify Lessons against COVID-19". In: *Frontiers of Medicine* (June 4, 2020), pp. 1–7. ISSN: 2095-0217. DOI: 10.1007/s11684-020-0788-3. pmid: 32495288. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7269685/ (visited on 06/04/2021).

[486] Christian S. Perone. "Analysis of the SARS-CoV-2 Outbreak in Rio Grande Do Sul / Brazil". July 26, 2020. arXiv: 2007.10486 [physics, q-bio]. URL: http://arxiv.org/abs/2007.10486 (visited on 06/04/2021).

[487] Virginia E. Pitzer et al. "The Impact of Changes in Diagnostic Testing Practices on Estimates of COVID-19 Transmission in the United States". In: *medRxiv* (Apr. 24, 2020). DOI: 10.1101/2020.04.20.20073338. pmid: 32511624. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7276051/ (visited on 06/04/2021).

[488] Jean-Philippe Platteau and Vincenzo Verardi. "How to Exit Covid-19 Lockdowns: Culture Matters". In: *Covid Economics* 23 (May 28, 2020), pp. 1–57. URL: https://cepr.org/file/9098.

[489] Amod K. Pokhrel, Yadav P. Joshi, and Sopnil Bhattarai. "Epidemiological Trend of COVID-19 in Nepal and the Importance of Social Distancing to Contain the Virus". In: *Applied Science and Technology Annals* 1.1 (1 June 30, 2020), pp. 15–25. ISSN: 2717-5014. DOI: 10.3126/asta.v1i1.30268. URL: https://www.nepjol.info/index.php/asta/article/view/30268 (visited on 06/04/2021).

[490] Dagoberto Pulido et al. "Geospatial Spread of the COVID-19 Pandemic in Mexico". June 13, 2020. arXiv: 2006.07784 [physics, q-bio]. URL: http://arxiv.org/abs/2006.07784 (visited on 06/04/2021).

[491] Soumik Purkayastha, Maxwell Salvatore, and Bhramar Mukherjee. "Are Women Leaders Significantly Better at Controlling the Contagion during the COVID-19 Pandemic?" In: *Journal of health and social sciences* 5.2 (June 2020), pp. 231–240. ISSN: 2499-2240. pmid: 32875269. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7457824/ (visited on 06/04/2021).

[492] Jinjun Ran et al. "The Changing Patterns of COVID-19 Transmissibility during the Social Unrest in the United States: A Nationwide Ecological Study with a before-and-after Comparison". In: *One Health* 12 (June 1, 2021), p. 100201. ISSN: 2352-7714. DOI: 10.1016/j.onehlt.2020.100201. URL: https://www.sciencedirect.com/science/article/pii/S2352771420303025 (visited on 06/04/2021).

[493] Gergely Röst et al. "Early Phase of the COVID-19 Outbreak in Hungary and Post-Lockdown Scenarios". In: *Viruses* 12.7 (7 July 2020), p. 708. DOI: 10.3390/v12070708. URL: https://www.mdpi.com/1999-4915/12/7/708 (visited on 06/04/2021).

[494] David Rubin et al. "Association of Social Distancing, Population Density, and Temperature With the Instantaneous Reproduction Number of SARS-CoV-2 in Counties Across the United States". In: *JAMA Network Open* 3.7 (July 23, 2020), e2016099. ISSN: 2574-3805. DOI: 10.1001/jamanetworkopen.2020.16099. URL: https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2768570 (visited on 06/04/2021).

[495] Martin Rypdal et al. "Modelling Suggests Limited Change in the Reproduction Number from Reopening Norwegian Kindergartens and Schools during the COVID-19 Pandemic". In: *PLOS ONE* 16.2 (Feb. 25, 2021), e0238268. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0238268. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0238268 (visited on 06/04/2021).

[496] Sukhyun Ryu et al. "Effect of Nonpharmaceutical Interventions on Transmission of Severe Acute Respiratory Syndrome Coronavirus 2, South Korea, 2020 - Volume 26, Number 10—October 2020 - Emerging Infectious Diseases Journal - CDC". In: *Emerging Infectious Diseases* 26.10 (Oct. 2020). DOI: 10.3201/eid2610.201886. URL: https://wwwnc.cdc.gov/eid/article/26/10/20-1886_article (visited on 06/04/2021).

[497] Sukhyun Ryu et al. "Epidemiology and Control of Two Epidemic Waves of SARS-CoV-2 in South Korea". In: *SSRN Electronic Journal [Pre-print]* (Oct. 15, 2020). DOI: 10.2139/ssrn.3684500. URL: https://papers.ssrn.com/abstract=3684500 (visited on 06/04/2021).

[498] Lizet Sanchez et al. "How Mathematical Approaches Could Help Decision-Making to Epidemic Control? The Successful Experience against COVID-19 in Cuba". In: *SSRN Electronic Journal [Pre-print]* (June 17, 2020). DOI: 10.2139/ssrn.3629271. URL: https://papers.ssrn.com/abstract=3629271 (visited on 06/04/2021).

[499] Héctor Eduardo Sánchez Vargas et al. "Modelación Físico-Matemática Para La Toma de Decisiones Frente a La COVID-19 En Cuba". In: *Retos de la Dirección* 14.2 (Dec. 2020), pp. 54–85. ISSN: 2306-9155. URL: http://scielo.sld.cu/scielo.php?script=sci_abstract&pid=S2306-91552020000200054&lng=es&nrm=iso&tlng=es (visited on 06/04/2021).

[500] Luis Santamaría and Joaquín Hortal. "COVID-19 Effective Reproduction Number Dropped during Spain's Nationwide Dropdown, Then Spiked at Lower-Incidence Regions". In: *Science of The Total Environment* 751 (Jan. 10, 2021), p. 142257. ISSN: 0048-9697. DOI: 10.1016/j.scitotenv.2020.142257. URL: https://www.sciencedirect.com/science/article/pii/S0048969720357867 (visited on 06/04/2021).

[501] Mario Santana-Cibrian et al. "Lifting Mobility Restrictions and the Effect of Superspreading Events on the Short-Term Dynamics of COVID-19". In: *Mathematical Biosciences and Engineering* 17.mbe-17-05-330 (5 2020), pp. 6240–6258. ISSN: 1551-0018. DOI: 10.3934/mbe.2020330. URL: http://www.aimspress.com/rticle/doi/10.3934/mbe.2020330 (visited on 06/04/2021).

[502] Palash Sashittal et al. "Characterization of SARS-CoV-2 Viral Diversity within and across Hosts". In: *bioRxiv* (May 13, 2020), p. 2020.05.07.083410. DOI: 10.1101/2020.05.07.083410. URL: https://www.biorxiv.org/content/10.1101/2020.05.07.083410v1 (visited on 06/04/2021).

[503] Jeremie Scire et al. "Reproductive Number of the COVID-19 Epidemic in Switzerland with a Focus on the Cantons of Basel-Stadt and Basel-Landschaft". In: *Swiss Medical Weekly* 150.1920 (May 4, 2020). DOI: 10.4414/smw.2020.20271. URL: https://smw.ch/article/doi/smw.2020.20271 (visited on 06/04/2021).

[504] Kerem Şenel et al. "Instantaneous R for COVID-19 in Turkey: Estimation by Bayesian Statistical Inference [Türkiye'de COVID-19 Için Anlık R Hesaplaması: Bayesyen Istatistiksel Çıkarım Ile Tahmin]". In: *Turkiye Klinikleri Journal of Medical Sciences* 40.2 (2 June 18, 2020), pp. 127–131. ISSN: 1300-0292 (Print) 2146-9040 (Online). URL: http://dx.doi.org/10.5336/medsci.2020-76462 (visited on 06/04/2021).

[505] Mounir Ould Setti and Ari Voutilainen. "Social Distancing with Movement Restrictions and the Effective Replication Number of COVID-19: Multi-Country Analysis Based on Phone Mobility Data". In: *medRxiv* (Oct. 12, 2020), p. 2020.10.08.20209064. DOI: 10.1101/2020.10.08.20209064. URL: https://www.medrxiv.org/content/10.1101/2020.10.08.20209064v1 (visited on 06/04/2021).

[506] Daniel Severo et al. "Nota Técnica Dos Modelos Implementados Pelo Coletivo Covid19br Para Projeções de Cenários Futuros Da Pandemia COVID-19 No Brasil". Apr. 25, 2020. arXiv: 2004.13488 [physics, q-bio]. URL: http://arxiv.org/abs/2004.13488 (visited on 06/04/2021).

[507] Eunha Shim, Amna Tariq, and Gerardo Chowell. "Spatial Variability in Reproduction Number and Doubling Time across Two Waves of the COVID-19 Pandemic in South Korea, February to July, 2020". In: *International Journal of Infectious Diseases* 102 (Jan. 1, 2021), pp. 1–9. ISSN: 1201-9712. DOI: 10.1016/j.ijid.2020.10.007. URL: https://www.sciencedirect.com/science/article/pii/S1201971220322104 (visited on 06/04/2021).

[508] Vitor Silva and Nicole Paul. "Potential Impact of Earthquakes during the 2020 COVID-19 Pandemic". In: *Earthquake Spectra* 37.1 (Feb. 1, 2021), pp. 73–94. ISSN: 8755-2930. DOI: 10.1177/8755293020950328. URL: https://doi.org/10.1177/8755293020950328 (visited on 06/04/2021).

[509] Balbir B. Singh et al. "Public Health Interventions Slowed but Did Not Halt the Spread of COVID-19 in India". In: *Transboundary and Emerging Diseases* n/a.n/a (). ISSN: 1865-1682. DOI: 10.1111/tbed.13868. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/tbed.13868 (visited on 06/04/2021).

[510] Gurpreet Singh et al. "Time-Dependent Dynamic Transmission Potential and Instantaneous Reproduction Number of COVID-19 Pandemic in India". In: *medRxiv* (July 16, 2020), p. 2020.07.15.20154971. DOI: 10.1101/2020.07.15.20154971. URL: https://www.medrxiv.org/content/10.1101/2020.07.15.20154971v1 (visited on 06/04/2021).

[511] Hyunjin Son et al. "Epidemiological Characteristics of and Containment Measures for COVID-19 in Busan, Korea". In: *Epidemiology and Health* 42 (June 1, 2020). ISSN: 2092-7193. DOI: 10.4178/epih.e2020035. URL: http://www.e-epih.org/journal/view.php?number=1102 (visited on 06/04/2021).

[512] Jean-Paul R. Soucy et al. "Estimating Effects of Physical Distancing on the COVID-19 Pandemic Using an Urban Mobility Index". In: *medRxiv* (May 24, 2020), p. 2020.04.05.20054288. DOI: 10.1101/2020.04.05.20054288. URL: https://www.medrxiv.org/content/10.1101/2020.04.05.20054288v3 (visited on 06/04/2021).

[513] Wilson Cabral de Sousa, Demerval Aparecido Gonçalves, and Dafne Brito Cruz. "COVID-19: Local/Regional Inequalities and Impacts over Critical Healthcare Infrastructure in Brazil". In: *Ambiente & Sociedade* 23 (July 3, 2020). ISSN: 1414-753X, 1414-753X, 1809-4422. DOI: 10.1590/1809-4422asoc20200114vu2020L3ID. URL: https://www.scielo.br/j/asoc/a/jcQ8s5jTXVKNLFVYP8nQ5yt/?lang=en (visited on 06/03/2021).

[514] Jessica E. Stockdale et al. "Long Time Frames to Detect the Impact of Changing COVID-19 Control Measures". In: *medRxiv* (June 16, 2020), p. 2020.06.14.20131177. DOI: 10.1101/2020.06.14.20131177. URL: https://www.medrxiv.org/content/10.1101/2020.06.14.20131177v1 (visited on 06/04/2021).

[515] Khouloud Talmoudi et al. "Estimating Transmission Dynamics and Serial Interval of the First Wave of COVID-19 Infections under Different Control Measures: A Statistical Analysis in Tunisia from February 29 to May 5, 2020". In: *BMC Infectious Diseases* 20.1 (Dec. 2, 2020), p. 914. ISSN: 1471-2334. DOI: 10.1186/s12879-020-05577-4. URL: https://doi.org/10.1186/s12879-020-05577-4 (visited on 06/04/2021).

[516] Tiago Tamagusko and Adelino Ferreira. "Data-Driven Approach to Understand the Mobility Patterns of the Portuguese Population during the COVID-19 Pandemic". In: *Sustainability* 12.22 (22 Jan. 2020), p. 9775. DOI: 10.3390/su12229775. URL: https://www.mdpi.com/2071-1050/12/22/9775 (visited on 06/04/2021).

[517] Biao Tang et al. "Lessons Drawn from China and South Korea for Managing COVID-19 Epidemic: Insights from a Comparative Modeling Study". In: *medRxiv* (Mar. 13, 2020), p. 2020.03.09.20033464. DOI: 10.1101/2020.03.09.20033464. URL: https://www.medrxiv.org/content/10.1101/2020.03.09.20033464v1 (visited on 06/04/2021).

[518] Ali Issakou Malam Tchole et al. "Epidemic and Control of COVID-19 in Niger: Quantitative Analyses in a Least Developed Country". In: *Journal of Global Health* 10.2 (Dec. 2020). ISSN: 2047-2978. DOI: 10.7189/jogh.10.020513. pmid: 33312506. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7719275/ (visited on 06/04/2021).

[519] Cristian Tebé et al. "COVID19-world: A Shiny Application to Perform Comprehensive Country-Specific Data Visualization for SARS-CoV-2 Epidemic". In: *BMC Medical Research Methodology* 20.1 (Sept. 21, 2020), p. 235. ISSN: 1471-2288. DOI: 10.1186/s12874-020-01121-9. URL: https://doi.org/10.1186/s12874-020-01121-9 (visited on 06/04/2021).

[520] Pham Quang Thai et al. "The First 100 Days of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Control in Vietnam". In: *Clinical Infectious Diseases* 72.9 (May 1, 2021), e334–e342. ISSN: 1058-4838. DOI: 10.1093/cid/ciaa1130. URL: https://doi.org/10.1093/cid/ciaa1130 (visited on 06/04/2021).

[521] Teresa To et al. “Correlation of Ambient Temperature and COVID-19 Incidence in Canada”. In: *The Science of the Total Environment* 750 (Jan. 1, 2021), p. 141484. ISSN: 0048-9697. DOI: 10.1016/j.scitotenv.2020.141484. pmid: 32829260. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7402211/ (visited on 06/04/2021).

[522] Alemtsehai A. Turasie. “Temporal Dynamics in COVID-19 Transmission: Case of Some African Countries”. In: *Advances in Infectious Diseases* 10.3 (3 Apr. 29, 2020), pp. 110–122. DOI: 10.4236/aid.2020.103011. URL: http://www.scirp.org/Journal/Paperabs.aspx?paperid=101989 (visited on 06/04/2021).

[523] Bryan Valcarcel et al. “The Effect of Early-Stage Public Health Policies in the Transmission of COVID-19 for South American Countries”. In: *Revista Panamericana de Salud Pública* 44 (Nov. 20, 2020). ISSN: 1020-4989. DOI: 10.26633/RPSP.2020.148. pmid: 33245294. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7679045/ (visited on 06/04/2021).

[524] Fabian Valka and Carla Schuler. “Estimation and Interactive Visualization of the Time-Varying Reproduction Number Rt and the Time-Delay from Infection to Estimation”. In: *medRxiv* (Sept. 22, 2020), p. 2020.09.19.20197970. DOI: 10.1101/2020.09.19.20197970. URL: https://www.medrxiv.org/content/10.1101/2020.09.19.20197970v1 (visited on 06/04/2021).

[525] Joan Valls et al. “COVID19-Tracker: una aplicación Shiny para analizar datos de la epidemia de SARS-CoV-2 en España”. In: *Gaceta Sanitaria* 35.1 (Jan. 1, 2021), pp. 99–101. ISSN: 0213-9111. DOI: 10.1016/j.gaceta.2020.04.002. URL: https://www.sciencedirect.com/science/article/pii/S0213911120300856 (visited on 06/04/2021).

[526] Timothy G. Vaughan et al. “Estimates of Outbreak-Specific SARS-CoV-2 Epidemiological Parameters from Genomic Data”. In: *medRxiv* (Sept. 14, 2020), p. 2020.09.12.20193284. DOI: 10.1101/2020.09.12.20193284. URL: https://www.medrxiv.org/content/10.1101/2020.09.12.20193284v1 (visited on 06/04/2021).

[527] Padmanaban Venkatesan. “National and State Wise Estimate of Time Varying Reproduction Number for COVID-19 in India during the Nationwide Lockdown”. In: *medRxiv* (May 6, 2020), p. 2020.05.01.20087197. DOI: 10.1101/2020.05.01.20087197. URL: https://www.medrxiv.org/content/10.1101/2020.05.01.20087197v1 (visited on 06/04/2021).

[528] Daniel Villela. “How Limitations in Data of Health Surveillance Impact Decision Making in the COVID-19 Epidemic”. In: *SciELO Pre-prints* (Oct. 9, 2020). DOI: 10.1590/SciELOPreprints.1313. URL: https://preprints.scielo.org/index.php/scielo/preprint/view/1313/version/1407 (visited on 06/04/2021).

[529] Kai Wang et al. “Real-Time Estimation of the Reproduction Number of the Novel Coronavirus Disease (COVID-19) in China in 2020 Based on Incidence Data”. In: *Annals of Translational*

Medicine 8.11 (June 2020). ISSN: 2305-5839. DOI: 10.21037/atm-20-1944. pmid: 32617309. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7327374/ (visited on 06/04/2021).

[530] Ke-wei Wang et al. "Transmission Potential and Forecasting of the Number of Coronavirus Disease 2019 Cases in Hubei Province, China". In: *Research Square [Pre-print]* (Nov. 23, 2020). DOI: 10.21203/rs.3.rs-36755/v2. URL: https://www.researchsquare.com/article/rs-36755/v2 (visited on 06/04/2021).

[531] Qingan Wang et al. "Could the Ambient Higher Temperature Decrease the Transmissibility of COVID-19 in China?" In: *Environmental Research* 193 (Feb. 2021), p. 110576. ISSN: 0013-9351. DOI: 10.1016/j.envres.2020.110576. pmid: 33279494. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7713592/ (visited on 06/04/2021).

[532] Justin Wong et al. "Epidemiological Investigation of the First 135 COVID-19 Cases in Brunei: Implications for Surveillance, Control, and Travel Restrictions". In: *The American Journal of Tropical Medicine and Hygiene* 103.4 (Oct. 2020), pp. 1608–1613. ISSN: 0002-9637. DOI: 10.4269/ajtmh.20-0771. pmid: 32815514. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7543844/ (visited on 06/04/2021).

[533] Martin C. S. Wong et al. "Stringent Containment Measures without Complete City Lockdown to Achieve Low Incidence and Mortality across Two Waves of COVID-19 in Hong Kong". In: *BMJ Global Health* 5.10 (Oct. 1, 2020), e003573. ISSN: 2059-7908. DOI: 10.1136/bmjgh-2020-003573. pmid: 33028700. URL: https://gh.bmj.com/content/5/10/e003573 (visited on 06/04/2021).

[534] Lee Worden et al. "Estimation of Effects of Contact Tracing and Mask Adoption on COVID-19 Transmission in San Francisco: A Modeling Study". In: *medRxiv* (June 11, 2020). DOI: 10.1101/2020.06.09.20125831. pmid: 32577672. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7302226/ (visited on 06/04/2021).

[535] Fan Xia et al. "Differences in How Interventions Coupled with Effective Reproduction Numbers Account for Marked Variations in COVID-19 Epidemic Outcomes". In: *Mathematical Biosciences and Engineering* 17.5 (2020), pp. 5085–5098. ISSN: 1551-0018. DOI: 10.3934/mbe.2020274. URL: http://www.aimspress.com/rticle/doi/10.3934/mbe.2020274 (visited on 06/04/2021).

[536] Jianpeng Xiao et al. "The Time-Varying Transmission Dynamics of COVID-19 and Synchronous Public Health Interventions in China". In: *International Journal of Infectious Diseases* 103 (Feb. 2021), pp. 617–623. ISSN: 1201-9712. DOI: 10.1016/j.ijid.2020.11.005. pmid: 33181330. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7836966/ (visited on 06/04/2021).

[537] Jie Xu et al. "Associations of Stay-at-Home Order and Face-Masking Recommendation with Trends in Daily New Cases and Deaths of Laboratory-Confirmed COVID-19 in the United States". In: *Exploratory Research and Hypothesis in Medicine* (July 8, 2020), pp. 1–10. ISSN:

2472-0712. DOI: 10.14218/ERHM.2020.00045. pmid: 32734759. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7361445/ (visited on 06/04/2021).

[538] Fook Fah Yap and Minglee Yong. "Implementation of An Online COVID-19 Epidemic Calculator for Tracking the Spread of the Coronavirus in Singapore and Other Countries". In: *medRxiv* (June 4, 2020), p. 2020.06.02.20120188. DOI: 10.1101/2020.06.02.20120188. URL: https://www.medrxiv.org/content/10.1101/2020.06.02.20120188v1 (visited on 06/04/2021).

[539] Xinhua Yu. "Impact of Mitigating Interventions and Temperature on the Instantaneous Reproduction Number in the COVID-19 Pandemic among 30 US Metropolitan Areas". In: *One Health* 10 (Dec. 1, 2020), p. 100160. ISSN: 2352-7714. DOI: 10.1016/j.onehlt.2020.100160. URL: https://www.sciencedirect.com/science/article/pii/S2352771420302615 (visited on 06/04/2021).

[540] Pei Yuan et al. "Efficacy of "Stay-at-Home" Policy and Transmission of COVID-19 in Toronto, Canada: A Mathematical Modeling Study". In: *medRxiv* (Oct. 21, 2020), p. 2020.10.19.20181057. DOI: 10.1101/2020.10.19.20181057. URL: https://www.medrxiv.org/content/10.1101/2020.10.19.20181057v1 (visited on 06/04/2021).

[541] Juanjuan Zhang et al. "Evolving Epidemiology and Transmission Dynamics of Coronavirus Disease 2019 Outside Hubei Province, China: A Descriptive and Modelling Study". In: *The Lancet Infectious Diseases* 20.7 (July 1, 2020), pp. 793–802. ISSN: 1473-3099, 1474-4457. DOI: 10.1016/S1473-3099(20)30230-9. pmid: 32247326. URL: https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(20)30230-9/abstract (visited on 06/04/2021).

[542] Yue Zhang et al. "The Real Time Effective Reproductive Number for COVID-19 in the United States". In: *medRxiv* (May 13, 2020), p. 2020.05.08.20095703. DOI: 10.1101/2020.05.08.20095703. URL: https://www.medrxiv.org/content/10.1101/2020.05.08.20095703v1 (visited on 06/04/2021).

[543] Hongwei Zhao et al. "COVID-19: Short Term Prediction Model Using Daily Incidence Data". In: *medRxiv* (Nov. 24, 2020), p. 2020.11.23.20237024. DOI: 10.1101/2020.11.23.20237024. URL: https://www.medrxiv.org/content/10.1101/2020.11.23.20237024v1 (visited on 06/04/2021).

[544] Shi Zhao and Xue Liang. "A Re-Analysis to Identify the Structural Breaks in COVID-19 Transmissibility during the Early Phase of the Outbreak in South Korea". In: *International Journal of Infectious Diseases* 100 (Nov. 1, 2020), pp. 10–11. ISSN: 1201-9712. DOI: 10.1016/j.ijid.2020.08.061. pmid: 32861827. URL: https://www.ijidonline.com/article/S1201-9712(20)30690-1/abstract (visited on 06/04/2021).

[545] Yu Zhao et al. "Analysis of the Transmissibility Change of 2019-Novel Coronavirus Pneumonia and Its Potential Factors in China from 2019 to 2020". In: *BioMed Research International* 2020 (May 19, 2020), e3842470. ISSN: 2314-6133. DOI: 10.1155/2020/3842470. URL: https://www.hindawi.com/journals/bmri/2020/3842470/ (visited on 06/04/2021).

[546] Lingling Zheng et al. "China's Effective Control and Other Countries' Uncharted Challenge against COVID-19: An Epidemiological and Modelling Study". In: *medRxiv* (May 2, 2020), p. 2020.04.28.20083899. DOI: 10.1101/2020.04.28.20083899. URL: https://www.medrxiv.org/content/10.1101/2020.04.28.20083899v1 (visited on 06/04/2021).

[547] Liting Zhu et al. "Meteorological Impact on the COVID-19 Pandemic: A Study across Eight Severely Affected Regions in South America". In: *The Science of the Total Environment* 744 (Nov. 20, 2020), p. 140881. ISSN: 0048-9697. DOI: 10.1016/j.scitotenv.2020.140881. pmid: 32674022. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7352107/ (visited on 06/04/2021).

[548] Zian Zhuang et al. "Preliminary Estimates of the Reproduction Number of the Coronavirus Disease (COVID-19) Outbreak in Republic of Korea and Italy by 5 March 2020". In: *International Journal of Infectious Diseases* 95 (June 1, 2020), pp. 308–310. ISSN: 1201-9712. DOI: 10.1016/j.ijid.2020.04.044. URL: https://www.sciencedirect.com/science/article/pii/S1201971220302599 (visited on 06/04/2021).

[549] Benjamin D. Dalziel et al. "Unreported Cases in the 2014-2016 Ebola Epidemic: Spatiotemporal Variation, and Implications for Estimating Transmission". In: *PLOS Neglected Tropical Diseases* 12.1 (Jan. 22, 2018), e0006161. ISSN: 1935-2735. DOI: 10.1371/journal.pntd.0006161. URL: https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0006161 (visited on 06/03/2021).

[550] International Ebola Response Team et al. "Exposure Patterns Driving Ebola Transmission in West Africa: A Retrospective Observational Study". In: *PLoS Medicine* 13.11 (Nov. 15, 2016). ISSN: 1549-1277. DOI: 10.1371/journal.pmed.1002170. pmid: 27846234. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5112802/ (visited on 06/04/2021).

[551] Thomas D Kirsch et al. "Impact of Interventions and the Incidence of Ebola Virus Disease in Liberia—Implications for Future Epidemics". In: *Health Policy and Planning* 32.2 (Mar. 2017), pp. 205–214. ISSN: 0268-1080. DOI: 10.1093/heapol/czw113. pmid: 28207062. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6279138/ (visited on 06/04/2021).

[552] Moritz U. G. Kraemer et al. "Dynamics of Conflict during the Ebola Outbreak in the Democratic Republic of the Congo 2018–2019". In: *BMC Medicine* 18.1 (Apr. 27, 2020), p. 113. ISSN: 1741-7015. DOI: 10.1186/s12916-020-01574-1. URL: https://doi.org/10.1186/s12916-020-01574-1 (visited on 06/04/2021).

[553] Seikh Taslim Ali, A. S. Kadi, and Neil M. Ferguson. "Transmission Dynamics of the 2009 Influenza A (H1N1) Pandemic in India: The Impact of Holiday-Related School Closure". In: *Epidemics* 5.4 (Dec. 2013), pp. 157–163. ISSN: 1755-4365. DOI: 10.1016/j.epidem.2013.08.001. pmid: 24267871. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4088951/ (visited on 06/02/2021).

[554] Sheikh Taslim Ali et al. "Ambient Ozone and Influenza Transmissibility in Hong Kong". In: *European Respiratory Journal* 51.5 (May 1, 2018). ISSN: 0903-1936, 1399-3003. DOI: 10.1183/13993003.00369-2018. pmid: 29563172. URL: https://erj.ersjournals.com/content/51/5/1800369 (visited on 06/02/2021).

[555] Jacob Carey. "Characterizing Risk of Transmission of Respiratory Viruses among School Children Using Proxies of Infectious Contact and Genetic Data". Thesis. Johns Hopkins University, Apr. 25, 2016. URL: https://dspace-prod.mse.jhu.edu/handle/1774.2/39486 (visited on 06/02/2021).

[556] B. D. Dalziel et al. "Long-Range Local Influenza Forecasts via Distributed Syndromic Monitoring: Preliminary Results". In: *medRxiv* (June 9, 2020), p. 2020.06.07.20078956. DOI: 10.1101/2020.06.07.20078956. URL: https://www.medrxiv.org/content/10.1101/2020.06.07.20078956v1 (visited on 06/03/2021).

[557] Anne Ewing et al. "Contact, Travel, and Transmission: The Impact of Winter Holidays on Influenza Dynamics in the United States". In: *The Journal of Infectious Diseases* 215.5 (Mar. 1, 2017), pp. 732–739. ISSN: 0022-1899. DOI: 10.1093/infdis/jiw642. pmid: 28031259. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5853779/ (visited on 06/03/2021).

[558] Jonggul Lee and Eunok Jung. "A Spatial–Temporal Transmission Model and Early Intervention Policies of 2009 A/H1N1 Influenza in South Korea". In: *Journal of Theoretical Biology* 380 (Sept. 7, 2015), pp. 60–73. ISSN: 0022-5193. DOI: 10.1016/j.jtbi.2015.05.008. URL: https://www.sciencedirect.com/science/article/pii/S0022519315002362 (visited on 06/04/2021).

[559] Anupong Sirirungreung et al. "The 2009 Pandemic Influenza A Virus in an Outbreak during 2014 in Samut Prakan Province, Thailand | OSIR Journal". In: *Outbreak, Surveillance and Investigation Reports* 8.3 (Sept. 2015), pp. 1–7. URL: http://www.osirjournal.net/index.php/osir/article/view/21 (visited on 06/04/2021).

[560] Paul A. Gastañaduy et al. "Assessment of the Status of Measles Elimination in the United States, 2001–2014". In: *American Journal of Epidemiology* 185.7 (Apr. 1, 2017), pp. 562–569. ISSN: 0002-9262. DOI: 10.1093/aje/kww168. pmid: 28338902. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7110139/ (visited on 06/04/2021).

[561] Paul A Gastañaduy et al. "Impact of Public Health Responses During a Measles Outbreak in an Amish Community in Ohio: Modeling the Dynamics of Transmission". In: *American Journal of Epidemiology* 187.9 (Sept. 2018), pp. 2002–2010. ISSN: 0002-9262. DOI: 10.1093/aje/kwy082. pmid: 29635277. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6118071/ (visited on 06/04/2021).

[562] Paul A. Gastañaduy et al. "Factors Associated With Measles Transmission in the United States During the Postelimination Era". In: *JAMA Pediatrics* 174.1 (Jan. 1, 2020), p. 56. ISSN: 2168-6203. DOI: 10.1001/jamapediatrics.2019.4357. URL: https://jamanetwork.com/journals/jamapediatrics/fullarticle/2755836 (visited on 06/04/2021).

[563] Shi Zhao et al. "Modelling the Measles Outbreak at Hong Kong International Airport in 2019: A Data-Driven Analysis on the Effects of Timely Reporting and Public Awareness". In: *Infection and Drug Resistance* 13 (June 17, 2020), pp. 1851–1861. DOI: 10.2147/IDR.S258035. URL: https://www.dovepress.com/modelling-the-measles-outbreak-at-hong-kong-international-airport-in-2-peer-reviewed-fulltext-article-IDR (visited on 06/04/2021).

[564] Finlay Campbell et al. "When Are Pathogen Genome Sequences Informative of Transmission Events?" In: *PLOS Pathogens* 14.2 (Feb. 8, 2018), e1006885. ISSN: 1553-7374. DOI: 10.1371/journal.ppat.1006885. URL: https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1006885 (visited on 06/02/2021).

[565] Megan O'Driscoll et al. "A Comparative Analysis of Statistical Methods to Estimate the Reproduction Number in Emerging Epidemics, With Implications for the Current Coronavirus Disease 2019 (COVID-19) Pandemic". In: *Clinical Infectious Diseases* (ciaa1599 Oct. 20, 2020). ISSN: 1058-4838. DOI: 10.1093/cid/ciaa1599. URL: https://doi.org/10.1093/cid/ciaa1599 (visited on 06/04/2021).

[566] Lin Zhao et al. "Outbreak of COVID-19 and SARS in Mainland China: A Comparative Study Based on National Surveillance Data". In: *BMJ Open* 10.10 (Oct. 1, 2020), e043411. ISSN: 2044-6055, 2044-6055. DOI: 10.1136/bmjopen-2020-043411. pmid: 33060093. URL: https://bmjopen.bmj.com/content/10/10/e043411 (visited on 06/04/2021).

[567] Qifang Bi et al. "The Epidemiology of Cholera in Zanzibar: Implications for the Zanzibar Comprehensive Cholera Elimination Plan". In: *The Journal of infectious diseases* 218 (suppl_3 2018), S173–S180. ISSN: 0022-1899.

[568] Anton Camacho et al. "Cholera Epidemic in Yemen, 2016–18: An Analysis of Surveillance Data". In: *The Lancet Global Health* 6.6 (June 1, 2018), e680–e690. ISSN: 2214-109X. DOI: 10.1016/S2214-109X(18)30230-4. pmid: 29731398. URL: https://www.thelancet.com/journals/langlo/article/PIIS2214-109X(18)30230-4/abstract (visited on 06/02/2021).

[569] Shi Zhao et al. "Associations between Public Awareness, Local Precipitation, and Cholera in Yemen in 2017". In: *The American Journal of Tropical Medicine and Hygiene* 101.3 (Sept. 2019), pp. 521–524. ISSN: 0002-9637. DOI: 10.4269/ajtmh.18-1016. pmid: 31333154. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6726958/ (visited on 06/04/2021).

[570] Mark F. Olson et al. "High Rate of Non-Human Feeding by Aedes Aegypti Reduces Zika Virus Transmission in South Texas". In: *Viruses* 12.4 (Apr. 17, 2020). ISSN: 1999-4915. DOI: 10.3390/v12040453. pmid: 32316394. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7232486/ (visited on 06/04/2021).

[571] T. Alex Perkins et al. "Heterogeneous Local Dynamics Revealed by Classification Analysis of Spatially Disaggregated Time Series Data". In: *Epidemics* 29 (Dec. 1, 2019), p. 100357. ISSN: 1755-4365. DOI: 10.1016/j.epidem.2019.100357. URL: https://www.sciencedirect.com/science/article/pii/S1755436518301622 (visited on 06/04/2021).

[572] Finlay Campbell et al. "Bayesian Inference of Transmission Chains Using Timing of Symptoms, Pathogen Genomes and Contact Data". In: *PLOS Computational Biology* 15.3 (Mar. 29, 2019), e1006930. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1006930. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1006930 (visited on 06/04/2021).

[573] Thibaut Jombart et al. "Bayesian Reconstruction of Disease Outbreaks by Combining Epidemiologic and Genomic Data". In: *PLOS Computational Biology* 10.1 (Jan. 23, 2014), e1003457. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1003457. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003457 (visited on 06/04/2021).

[574] Sheikh Taslim Ali et al. "Meteorological Drivers of Respiratory Syncytial Virus Infections in Singapore". In: *Scientific Reports* 10.1 (1 Nov. 24, 2020), p. 20469. ISSN: 2045-2322. DOI: 10.1038/s41598-020-76888-4. URL: https://www.nature.com/articles/s41598-020-76888-4 (visited on 06/02/2021).

[575] Hirono Otomaru et al. "Transmission of Respiratory Syncytial Virus Among Children Under 5 Years in Households of Rural Communities, the Philippines". In: *Open Forum Infectious Diseases* 6.3 (Mar. 11, 2019). ISSN: 2328-8957. DOI: 10.1093/ofid/ofz045. pmid: 30882012. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6411217/ (visited on 06/04/2021).

[576] Sun Hee Park et al. "Epidemiologic Parameters of the Middle East Respiratory Syndrome Outbreak in Korea, 2015". In: *Infection & Chemotherapy* 48.2 (June 2016), pp. 108–117. ISSN: 2093-2340. DOI: 10.3947/ic.2016.48.2.108. pmid: 27433381. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4945720/ (visited on 06/04/2021).

[577] Xu-Sheng Zhang et al. “Estimating and Modelling the Transmissibility of Middle East Respiratory Syndrome CoronaVirus during the 2015 Outbreak in the Republic of Korea”. In: *Influenza and Other Respiratory Viruses* 11.5 (Sept. 2017), pp. 434–444. ISSN: 1750-2640. DOI: 10.1111/irv.12467. pmid: 28703921. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5598245/ (visited on 06/04/2021).

[578] Ellsworth M Campbell et al. “Detailed Transmission Network Analysis of a Large Opiate-Driven Outbreak of HIV Infection in the United States”. In: *The Journal of Infectious Diseases* 216.9 (Nov. 1, 2017), pp. 1053–1062. ISSN: 0022-1899. DOI: 10.1093/infdis/jix307. pmid: 29029156. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5853229/ (visited on 06/02/2021).

[579] Michael Carter-Gates et al. “Implications of Increasing Atlantic Influence for Arctic Microbial Community Structure”. In: *Scientific Reports* 10.1 (1 Nov. 6, 2020), p. 19262. ISSN: 2045-2322. DOI: 10.1038/s41598-020-76293-x. URL: https://www.nature.com/articles/s41598-020-76293-x (visited on 06/02/2021).

[580] Luis F. Chaves et al. “Malaria Elimination in Costa Rica: Changes in Treatment and Mass Drug Administration”. In: *Microorganisms* 8.7 (June 30, 2020). ISSN: 2076-2607. DOI: 10.3390/microorganisms8070984. pmid: 32630155. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7409053/ (visited on 06/02/2021).

[581] Chenxi Dai et al. “Epidemics and Underlying Factors of Multiple-Peak Pattern on Hand, Foot and Mouth Disease inWenzhou, China”. In: *Mathematical Biosciences and Engineering* 16.4 (2019), pp. 2168–2188. ISSN: 1551-0018. DOI: 10.3934/mbe.2019106. URL: http://www.aimspress.com/rticle/doi/10.3934/mbe.2019106 (visited on 06/03/2021).

[582] AshleyG. Jordan et al. “Mycoplasma Bovis Outbreak in New Zealand Cattle: An Assessment of Transmission Trends Using Surveillance Data”. In: *Transboundary and Emerging Diseases* 68.6 (2021), pp. 3381–3395. ISSN: 1865-1682. DOI: 10.1111/tbed.13941. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/tbed.13941 (visited on 10/03/2024).

[583] Valeria P. Martínez et al. ““Super-Spreaders” and Person-to-Person Transmission of Andes Virus in Argentina”. In: *New England Journal of Medicine* (Dec. 2, 2020). DOI: 10.1056/NEJMoa2009040. URL: https://www.nejm.org/doi/10.1056/NEJMoa2009040 (visited on 06/04/2021).

[584] Margarita Pons-Salort et al. “The Seasonality of Nonpolio Enteroviruses in the United States: Patterns and Drivers”. In: *Proceedings of the National Academy of Sciences* 115.12 (Mar. 20, 2018), pp. 3078–3083. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1721159115. pmid: 29507246. URL: https://www.pnas.org/content/115/12/3078 (visited on 06/04/2021).

[585] Christian Julián Villabona-Arenas et al. "Epidemiological Dynamics of an Urban Dengue 4 Outbreak in São Paulo, Brazil". In: *PeerJ* 4 (Apr. 5, 2016), e1892. ISSN: 2167-8359. DOI: 10.7717/peerj.1892. URL: https://peerj.com/articles/1892 (visited on 06/04/2021).

[586] Shi Zhao et al. "Phase-Shifting of the Transmissibility of Macrolide-Sensitive and Resistant Mycoplasma Pneumoniae Epidemics in Hong Kong, from 2015 to 2018". In: *International Journal of Infectious Diseases* 81 (Apr. 1, 2019), pp. 251–253. ISSN: 1201-9712. DOI: 10.1016/j.ijid.2019.02.030. URL: https://www.sciencedirect.com/science/article/pii/S1201971219300943 (visited on 06/04/2021).

[587] Shi Zhao et al. "Modelling the Effective Reproduction Number of Vector-Borne Diseases: The Yellow Fever Outbreak in Luanda, Angola 2015–2016 as an Example". In: *PeerJ* 8 (Feb. 27, 2020). ISSN: 2167-8359. DOI: 10.7717/peerj.8601. pmid: 32149023. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7049463/ (visited on 06/04/2021).

[588] *Tools to Make Developing R Packages Easier*. URL: https://devtools.r-lib.org/ (visited on 02/24/2022).

[589] *R Package Installation from Remote Repositories, Including GitHub*. URL: https://remotes.r-lib.org/ (visited on 02/24/2022).

[590] *JAGS - Just Another Gibbs Sampler*. URL: https://mcmc-jags.sourceforge.io/ (visited on 03/17/2022).

[591] *JAGS: Just Another Gibbs Sampler*. SourceForge. URL: https://sourceforge.net/projects/mcmc-jags/ (visited on 03/17/2022).

[592] *Stan*. stan-dev.github.io. URL: //mc-stan.org/ (visited on 03/17/2022).

[593] *RStan*. stan-dev.github.io. URL: //mc-stan.org/users/interfaces/rstan (visited on 03/17/2022).

[594] *RStan Getting Started · Stan-Dev/Rstan Wiki*. GitHub. URL: https://github.com/stan-dev/rstan (visited on 03/17/2022).

[595] *Flu Pandemic*. URL: https://imperialcollegelondon.github.io/epidemia/articles/flu.html (visited on 03/17/2022).

[596] UK Health Security Agency. *Ebola: Overview, History, Origins and Transmission*. 2023.

[597] Maria D Van Kerkhove et al. "A Review of Epidemiological Parameters from Ebola Outbreaks to Inform Early Public Health Decision-Making". In: *Scientific data* 2.1 (2015), pp. 1–10.

[598] A.M. Rojek et al. "A Systematic Review and Meta-Analysis of Patient Data from the West Africa (2013–16) Ebola Virus Disease Epidemic". In: *Clinical Microbiology and Infection* 25.11 (2019), pp. 1307–1314. ISSN: 1198-743X. DOI: 10.1016/j.cmi.2019.06.032.

[599] J. Kawuki, T.H. Musa, and X. Yu. "Impact of Recurrent Outbreaks of Ebola Virus Disease in Africa: A Meta-Analysis of Case Fatality Rates". In: *Public Health* 195 (2021), pp. 89–97. ISSN: 0033-3506. DOI: 10.1016/j.puhe.2021.03.027.

[600] Luke Nyakarahuka et al. "How Severe and Prevalent Are Ebola and Marburg Viruses? A Systematic Review and Meta-Analysis of the Case Fatality Rates and Seroprevalence". In: *BMC infectious diseases* 16 (2016), pp. 1–14.

[601] Drifa Belhadi et al. "The Number of Cases, Mortality and Treatments of Viral Hemorrhagic Fevers: A Systematic Review". In: *PLOS Neglected Tropical Diseases* 16.10 (Oct. 2022), pp. 1–25. DOI: 10.1371/journal.pntd.0010889.

[602] Julii Brainard et al. "Risk Factors for Transmission of Ebola or Marburg Virus Disease: A Systematic Review and Meta-Analysis". In: *International Journal of Epidemiology* 45.1 (Nov. 2015), pp. 102–116. ISSN: 0300-5771. DOI: 10.1093/ije/dyv307. eprint: https://academic.oup.com/ije/article-pdf/45/1/102/24170592/dyv307.pdf.

[603] Saranya A Selvaraj et al. "Infection Rates and Risk Factors for Infection among Health Workers during Ebola and Marburg Virus Outbreaks: A Systematic Review". In: *The Journal of Infectious Diseases* 218.suppl_5 (Sept. 2018), S679–S689. ISSN: 0022-1899. DOI: 10.1093/infdis/jiy435. eprint: https://academic.oup.com/jid/article-pdf/218/suppl{\textbackslash}_5/S679/26689050/jiy435.pdf.

[604] Julian Higgins and James Thomas. *Cochrane Handbook for Systematic Reviews of Interventions*. https://training.cochrane.org/handbook/current. (Visited on 05/02/2024).

[605] Siyu Cai, Jie Zhou, and Jianxin Pan. "Estimating the Sample Mean and Standard Deviation from Order Statistics and Sample Size in Meta-Analysis". In: *Statistical Methods in Medical Research* 30.12 (Dec. 2021), pp. 2701–2719. ISSN: 1477-0334. DOI: 10.1177/09622802211047348.

[606] *Données de laboratoires pour le dépistage : Indicateurs sur les variants (SI-DEP) - data.gouv.fr*. URL: https://www.data.gouv.fr/fr/datasets/donnees-de-laboratoires-pour-le-depistage-indicateurs-sur-les-variants/ (visited on 04/05/2023).

[607] Sebastian Funk et al. "Assessing the performance of real-time epidemic forecasts: A case study of Ebola in the Western Area region of Sierra Leone, 2014-15". In: *PLoS Computational Biology* 15.2 (2019). Publisher: Public Library of Science San Francisco, CA USA, e1006785.

[608] Nikos I Bosse et al. "Evaluating forecasts with scoringutils in R". In: *arXiv preprint arXiv:2205.07090* (2022).

[609] *Projections package - : Project future case incidence*. 2021. URL: https://cran.r-project.org/web/packages/projections/index.html (visited on 04/05/2023).

[610] R Core Team. *R: A language and environment for statistical computing.* Vienna, Austria, 2013. URL: `http://www.R-project.org/`.

Appendix A

Supplementary Information: Real-time estimation of the epidemic reproduction number: scoping review of the applications and challenges

The contents of this appendix has been adapted from supplementary information previously published in PLOS Digital Health [105].

A.1 Overview

In this supplementary information, we provide more detailed results from our scoping review, including additional exploration to determine the usability of each identified R package and tool, more in depth analyses of our questionnaire data, and the questionnaire itself.

A.2 Scoping review: additional results

Our scoping review identified 281 papers that used an unmodified (n=242) or modified (n=54) version of EpiEstim. There were a variety of diseases under investigation, but the most common were COVID-19, followed by Ebola Virus Disease (EVD) and Influenza (Figure A.1).

Between 2013 and December 2020, 242 papers were identified that used an unmodified version of EpiEstim or the underlying method (Table A.1). The majority (82%) of these papers were investigating COVID-19 (n=198).

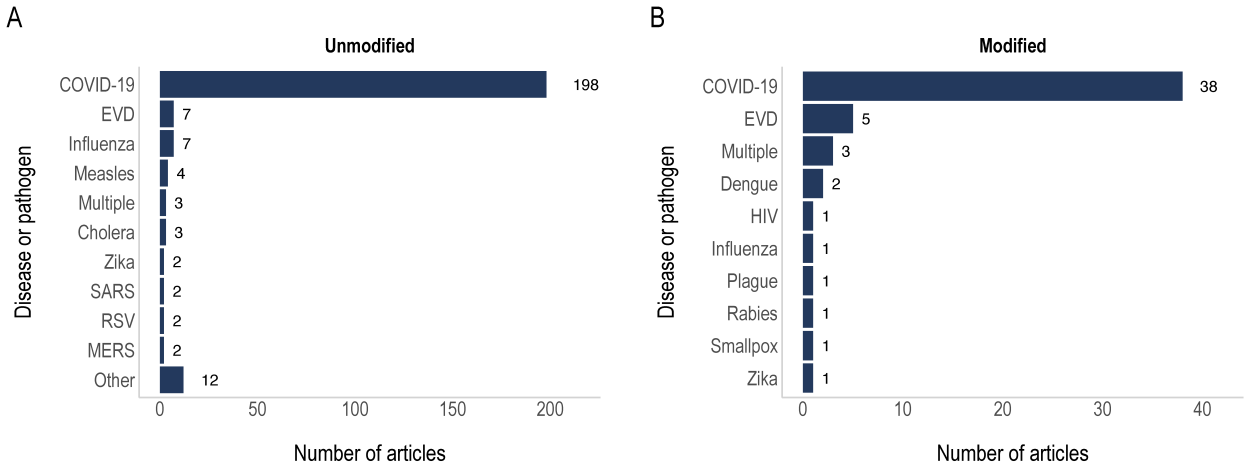


Figure A.1: **The disease or pathogen under investigation in A) the papers that used an unmodified version of the EpiEstim package or the method and B) the papers that used a modified version of the approach or package.** The category "multiple" refers to papers where more than one disease or pathogen were investigated. Note the diseases are different in both panels.

A.3 Exploration of R package and tool usability

We performed some additional exploration to determine the usability of each R package and tool (Table A.2 & Table 2.2 in the main text). This was assessed by the ease of installation, the availability of adequate documentation or tutorials, and their speed. Table A.2 provides the full breakdown for how each score was defined.

A.3.1 Ease of installation

Ease of installation was determined by the availability of installation instructions, the options for how the R package could be installed, the R package dependencies, and whether other system dependencies were required for either running or installing the packages. As EpiFilter is not an R package, some of the criteria in this section are not applicable. To use EpiFilter, the repository needs to be cloned from GitHub [171].

Installation instructions

Installation instructions may be available through an adequate README file or website, e.g. a "quick start" or introductory page. Instructions were found for all R packages except for APEestim and bayEStim.

Options for installation

Installation is easiest if the R package is available on the The Comprehensive R Archive Network (CRAN), by simply using the "install.packages()" function. earlyR, epicontacts and EpiNow2 are available on CRAN. All remaining R packages can be installed using the devtools R package, which

Table A.1: **Summary table of the papers identified that used an unmodified version of EpiEstim.**

Disease or disease agent investigated	**n (total = 242)**	**References**
COVID-19	198	[116, 118–130, 287, 366–548]
Ebola Virus Disease (EVD)	7	[108, 117, 372, 549–552]
Influenza	7	[553–559]
Measles	4	[560–563]
Multiple diseases	3	[564–566]
Cholera	3	[567–569]
Zika	2	[570, 571]
Severe Acute Respiratory Syndrome (SARS)	2	[572, 573]
Respiratory Syncytial Virus (RSV)	2	[574, 575]
Middle East Respiratory Syndrome (MERS)	2	[576, 577]
Other	12	[112, 368, 578–587]

enable packages to be installed and built directly from source (e.g. GitHub) [588, 589]. If the package has compiled code, as is the case for Epidemia, then devtools requires that the user has a working development environment (e.g. Rtools for Windows and Xcode for Mac).

R package dependencies

R package dependencies have the potential to complicate installation, particularly if there are many dependencies or if package versions are specified. For example, certain packages and/or package versions may not be compatible with the version of R that the user has installed. The chance of issues arising increases with the greater number of package dependencies, but we chose 10 dependencies as an arbitrary cut off for the usability score.

Additional system dependencies required for running or installing the packages

Three of the packages: bayEStim, Epidemia, and EpiNow2, require additional system (or software) dependencies, in addition to R packages. bayEStim depends on the R package rjags to run, which itself requires users to download and install the programme Just Another Gibbs Sampler (JAGS) [590, 591]. As bayEStim has no documentation, this process is not clear, especially for less experienced users. Epidemia and EpiNow2 use the Stan probabilistic programming language and depend on the R package rstan, which provides an R interface that allows Stan models to be fit [592, 593]. To install rstan, users are required to install a recent C++ compiler and configure R to use it. Both Epidemia and EpiNow2 provide installation instructions including links to the rstan wiki [594].

A.3.2 Documentation and tutorials

We have determined whether each R package or tool has function documentation, a vignette or worked examples. At a minimum, function documentation including details on function arguments and outputs should be made available. APEstim and bayEStim only provide function documentation, but no examples, which would make both difficult to use. All the remaining packages and tool have a vignette (or an equivalent, such as a webpage with tutorials) with one or more worked examples.

A.3.3 Speed

Computational speed has been assessed by each author of this study with different computer specifications (see Table A.2). The main function in each package/tool was taken from provided examples (if available) and wrapped in the system.time() R function to measure the execution time. The main function estimated the reproduction number for each R package/tool except for epicontacts, which estimates the serial interval. We chose the following classifications: <10 seconds = very good, 10 seconds - 5 minutes = good, >5 minutes = poor. The classification allocated to each package was based on the agreement of at least 2 out of the 3 computers. We note that such direct comparisons of the runtimes of the different models may not be fair, as the examples provided by each package which we have used to assess speed vary in terms of the dataset used, model complexity, and dimensionality of the reproduction number to estimate. Nevertheless, we assume that examples will always be relatively simple and therefore their computational speed may be a good overall indicator of speed of reproduction number estimation in general using a given package.

In our experience, of those that estimated the reproduction number, the fastest packages were APEestim, earlyR and EpiFilter, all taking <10 seconds. Epidemia took from <1 minute to nearly 4 minutes, depending on whether the model was basic (uses renewal equation to propagate infections) or extended (adds variance to this process). Interestingly, the extended model took less time to run, which as discussed on the package website, may be due to the posterior distribution being easier to sample from [595]. Meanwhile, estimating the reproduction number using the epinow() function in EpiNow2 took longer, from 3 up to 13 minutes. This took much longer when estimating the reproduction number for each geographical region in turn using regional_epinow(), ranging from 8 to 25 minutes. Both Epidemia and EpiNow2 give an indication of progress whilst estimation is ongoing. This is particularly important for EpiNow2 given the lengthy execution time, but the progress bar only refreshes after long intervals, so the user is not updated regularly.

Table A.2: **Usability of each R package or tool* identified in the scoping review.** This table shows a full breakdown of how the classifications (very good = ✓✓, good = ✓, poor = X) were determined for the "additional exploration" section of Table 2.2 within the main text. For the 'ease of installation' and 'documentation and tutorials' sections, each criterion was allocated a score, shown in squared brackets, and the overall classification was determined by the sum of the scores. For the 'speed' section, each author used the system.time() function in R to determine the run time of the main function of the package available in the provided examples.** The classification (<10s = ✓✓, >10s - 5min = ✓, >5min = X) was decided based on the time category agreed on by at least 2 out of the 3 computers.

Theme	Usability criterion	**APEestim** *(v 0.0.1)*	**bayEStim** *(v 0.0.1)*	**earlyR** *(v 0.0.5)*	**epicontacts** *(v 1.1.2)*	**Epidemia** *(v 1.0.0)*	**EpiFilter*** *(*Tool)*	**EpiNow2** *(v 1.3.2)*
Ease of installation	Installation instructions available (e.g., detailed README file or webpage) **[1]**	0	0	1	1	1	0	1
	Installation via CRAN **[1]**	0	0	1	1	0	NA	1
	Installation via devtools **[1]** *(Note: devtools may require the installation of Rtools (Windows) or Xcode (Mac))*	1	1	1	1	1	NA	1
	Number of package dependencies **[if n<10 score = 1]**	1 (n=0)	1 (n=3)	1 (n=5)	1 (n=7)	0 (n=18)	1 (n=0)	0 (n=23)
	Separate prerequisite installations or configurations required outside of R **[if no prerequisites, score = 1]**	1	0 (JAGS 4.x.y (rjags))	1	1	0 (Configure C++ toolchain (rstan))	NA (Download github repository)	0 (Configure C++ toolchain (rstan))
	Overall ease of installation Scoring: [4-5] = ✓✓ [2-3] = ✓ [0-2] = X	✓	X	✓✓	✓✓	X	X+	✓
Documentation and tutorials	Function documentation (description of inputs and outputs) **[1]**	1	1	1	1	1	1	1
	Vignette available (either included within the package or online) **[1]**	0	0	1	1	1	1	1
	Worked example(s) available **[1]**	0	0	1	1	1	1	1
	Overall documentation and tutorials Scoring: [3] = ✓✓ [2] = ✓ [0-1] = X	X	X	✓✓	✓✓	✓✓	✓✓	✓✓
Speed	Speed of estimation (estimation of the reproduction number for all except for epicontacts, which estimates the serial interval) Based on provided examples** and the system.time() function on three computers (C1-3) with the following specifications: **C1:** MacOS (2 GHz Quad-Core Intel Core i5) 16GB RAM **C2:** MacOS (M1) 32GB RAM **C3:** Windows (i7) 16GB RAM	apeEstim() (Taken from EpiFilter vignette) *Flu example* C1: 0.227s C2: 0.118s C3: 0.19s *SARS example* C1: 0.206s C2: 0.106s C3: 0.22s	NA (No examples provided)	get_R() C1: 0.022s C2: 3s C3: 0.02s	get_pairwise() C1: 0s C2: 0s C3: 0s	epim() Basic model (fm1) C1: 2m 29s C2: 51s C3: 2m 17s Extended model (fm2) C1: 43s C2: 17s C3: 52s	epiFilter() *Flu example* C1: 0.053s C2: 0.022s C3: 0.24s *SARS example* C1: 0.018s C2: 0.007s C3: 0.22s	epinow() C1: 7m 50s C2: 3m 22s C3: 13m 23s regional_epinow() C1: 16m 36s C2: 7m 47s C3: 25m 02s
	Overall speed Scoring based on what the majority of C1-3 found: <10s = ✓✓ >10s – 5min = ✓ >5min = X	✓✓	NA	✓✓	✓✓	✓++	✓✓	X++

+ Given that EpiFilter is not an R package, some of the criteria were not applicable. We have scored it with an X as it requires cloning a github repository.

++ Function includes a progress bar or some other indication of progress

** Source of examples: APEestim and EpiFilter: https://github.com/kpzoo/EpiFilter/blob/master/R%20files/vignette.R; earlyR: https://cran.r-project.org/web/packages/earlyR/vignettes/earlyR.html; epicontacts: https://cran.r-project.org/web/packages/epicontacts/vignettes/epicontacts.html; Epidemia: https://imperialcollegelondon.github.io/epidemia/articles/flu.html; EpiNow2: https://github.com/epiforecasts/EpiNow2

A.4 Questionnaire: additional results

The full questionnaire is shown in the next section. In total, there were 17 responses to the questionnaire from respondents in 11 different countries, including the US, Canada, France, Indonesia, Austria, Bermuda, Germany, India, Peru, Uruguay, and the UK (Figure A.2). In an attempt to reach a wide variety of people, the questionnaire was distributed to a contact list of known EpiEstim users, in addition to being shared via the Imperial Medical Research Council (MRC) Centre for Global Infectious Disease Analysis Twitter account (with over 141,000 followers). Nonetheless, the number of respondents was small, and it is possible that the sample may be more representative of epidemiologists, or those with a high level of training in the field, as opposed to other researchers who may experience greater barriers in terms of usability. Despite these limitations, the questionnaire helped us to limit the risk of publication bias and was useful to reinforce findings from our literature review whilst ensuring that we didn't overlook any issues that would be less apparent from the literature search alone.

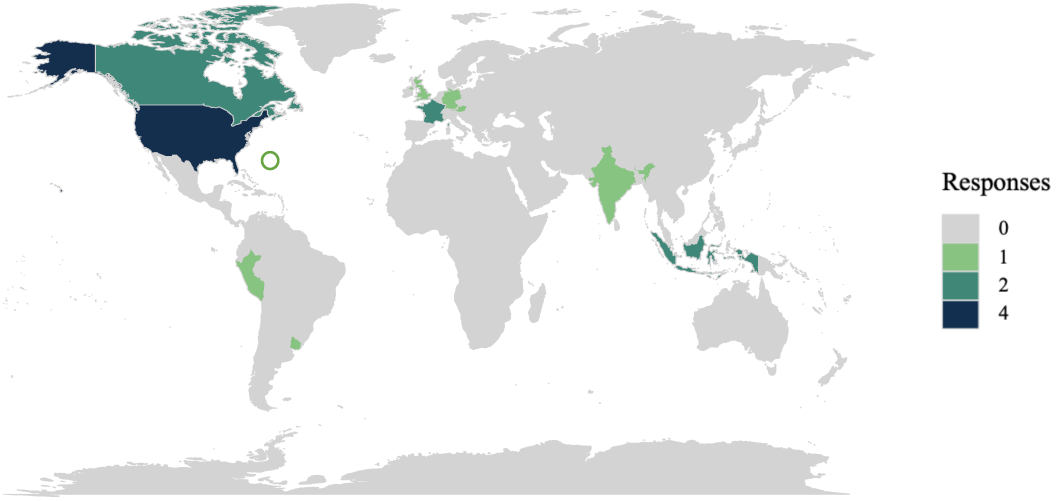


Figure A.2: **Map showing the country in which each questionnaire respondent is based.** The majority of responses were from the US (n=4), followed by Canada (n=2), France (n=2) and Indonesia (n=2). There was one response from each of Austria, Bermuda (circled), Germany, India, Peru, Uruguay, and the UK.

Most respondents were academic researchers, public health consultants or health professionals (Figure A.3). The vast majority (94.1%) used EpiEstim for the purpose of real-time estimation of the reproduction number. Over half (58.8%) had used EpiEstim to retrospectively estimate the reproduction number, whilst the purpose of the analysis was less frequently for real-time incidence forecasting (18%) (Figure A.3). One respondent used EpiEstim for model calibration.

All questionnaire respondents had performed analysis using EpiEstim to investigate COVID-19 and some also used it for Ebola Virus Disease, Norovirus and Influenza (Figure A.4). In terms of input datasets, all respondents had access to daily case incidence data, 24% had daily death data, and 12% had weekly case incidence data (Figure A.4). Only one respondent reported having data that distinguished between local and imported cases and none reported having data containing infector-

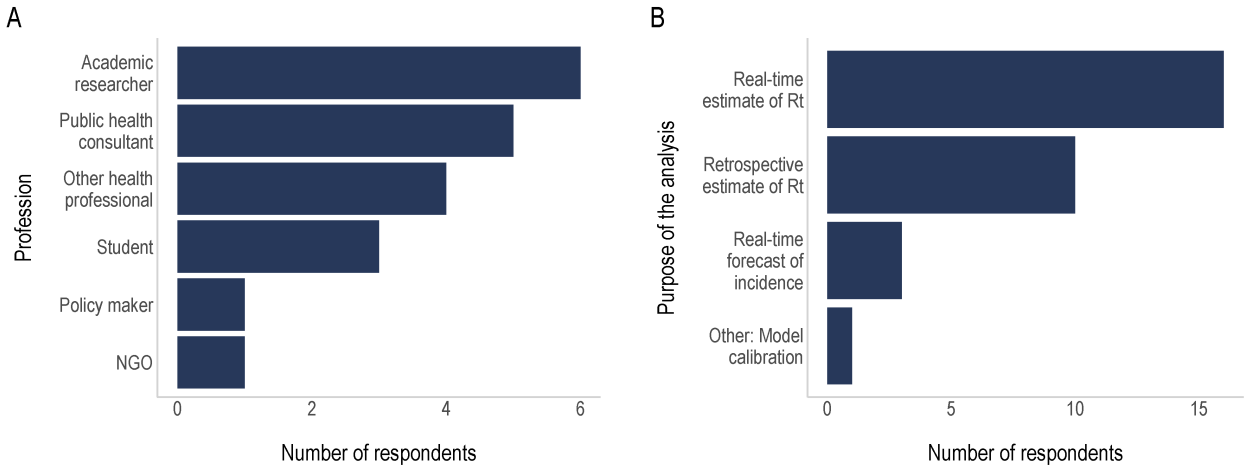


Figure A.3: **The profession of each questionnaire respondent (A) and the purpose of their analysis (B).** Respondents could select more than one answer for both questions.

infected pairs.

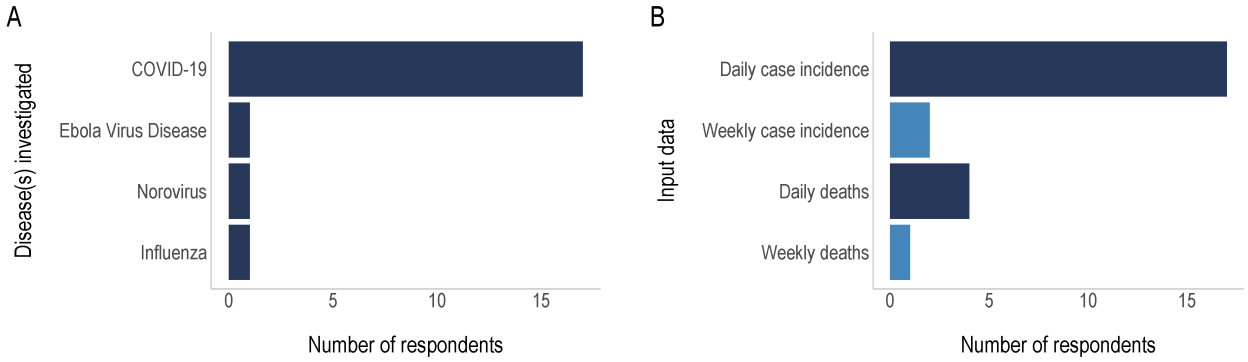


Figure A.4: **Disease(s) investigated by each respondent (A) and the categories of input data (B).** Respondents could select more than one answer for both questions.

The analyses predominantly informed academic research (53%), intervention planning (35%), governmental situation reports (29%) and logistical planning (29%) (Figure A.5).

On a scale from 1 to 5 (1 representing "badly" and 5 "very well") respondents were asked how well EpiEstim met their needs. All respondents selected between 3 and 5, with 94% selecting either 4 or 5 (Figure A.5). Despite this, twelve of the respondents (71%) reported having a technical or methodological issue when using EpiEstim.

Almost half of respondents (47%) said they thought none of EpiEstim's features could be improved, whereas 35% thought that usability could be improved (Figure A.6). One respondent (6%) thought speed could be improved and 24% selected "other" and mentioned either issues understanding how to use the package properly (n=2), wanting extra features (n=1) or issues with compatibility (n=1).

Although when asked directly, only 35% of respondents said usability was an issue (Figure A.6), users' responses were classified as such (Table A.3) if they reported an issue that suggested a lack of



Figure A.5: **Broad reason for the use of EpiEstim.** Respondents could select more than one answer for this question.

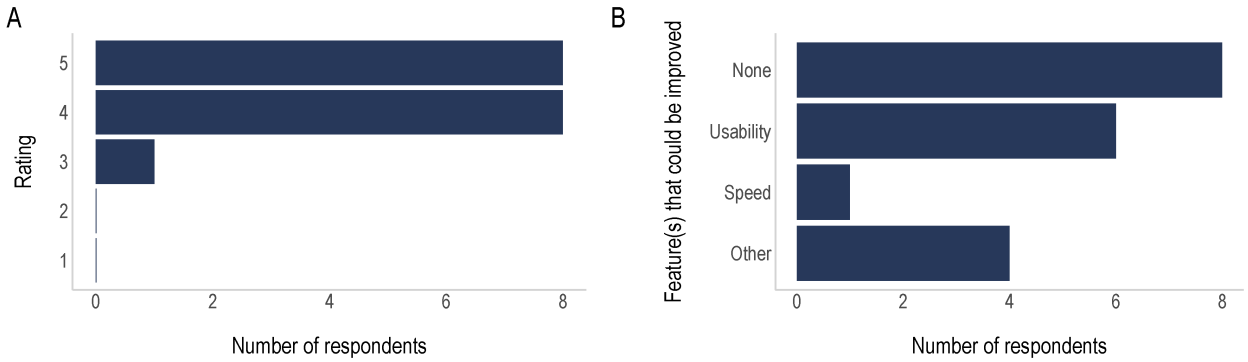


Figure A.6: **Questionnaire responses to A) how well the package met the needs of each respondent on a scale from 1 to 5 (1: "badly", 5: "very well"), and B) "Which features do you think could be improved?".**

understanding of the package. For example, one respondent reported a technical issue as being unable to estimate Rt for day 1 of the outbreak, but EpiEstim can only start estimating Rt on at least day 2 because it requires data on past incidence [183].

The questionnaire highlighted similar challenges to the scoping review, including issues with incidence data, geographical factors and suggestions to extend the capability of the package to address practical and logistical issues (Table A.3). Respondents wanted to account for reporting delays and time-varying reporting rates, to use weekly as well as daily incidence, different formats of data, and more intuitive plotting options for the results. Regarding geographical issues, they suggested enabling easier reproduction number estimates for different regions simultaneously. For more practical purposes, they proposed expanding the methodology to allow for the projection of hospital bed occupancy, thereby assisting logistical planning in hospitals. As discussed in the main text, the questionnaire also revealed an additional three themes: usability, speed, and compatibility.

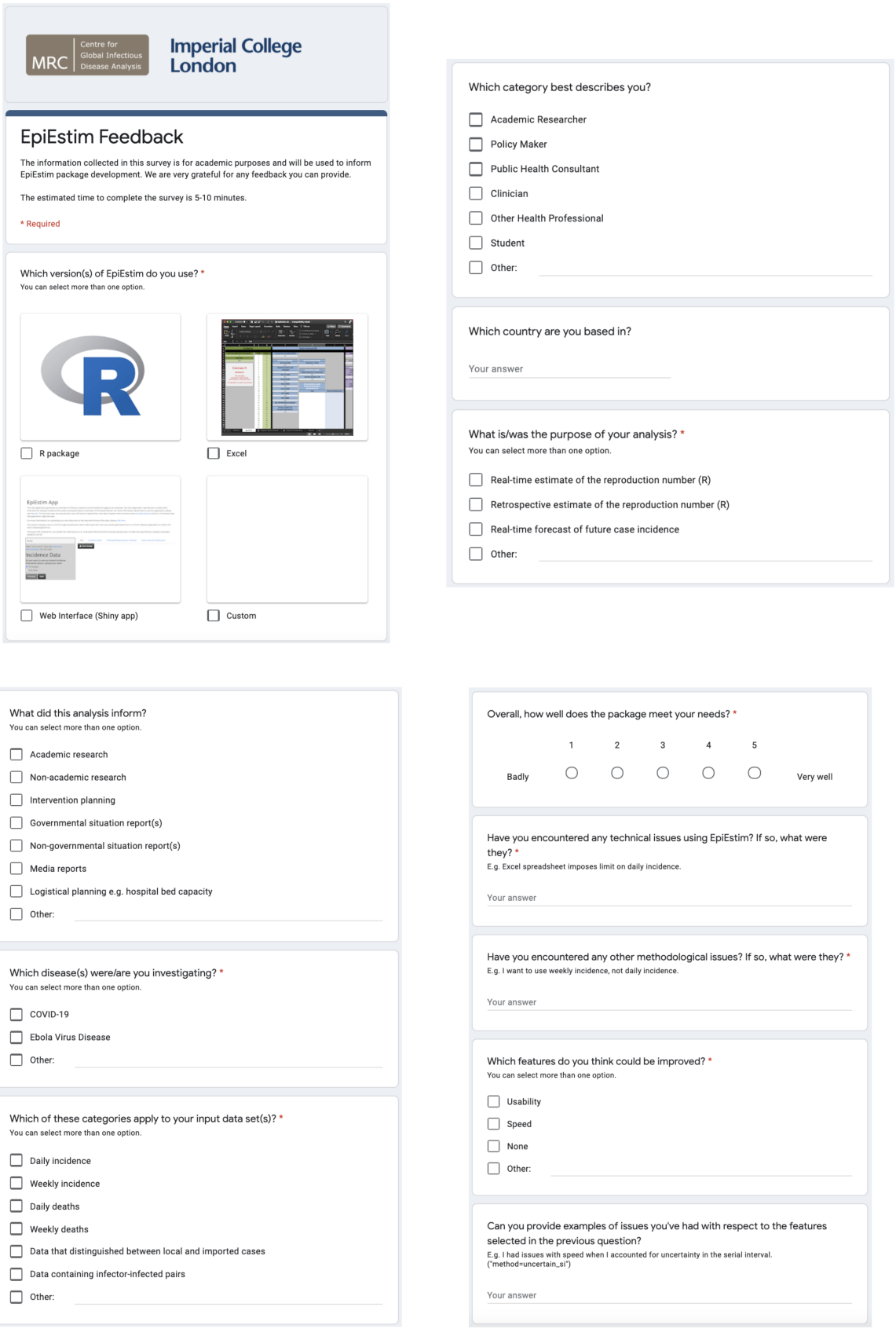
It is important to note that, as shown in Figure A.7, for four of the survey questions that required

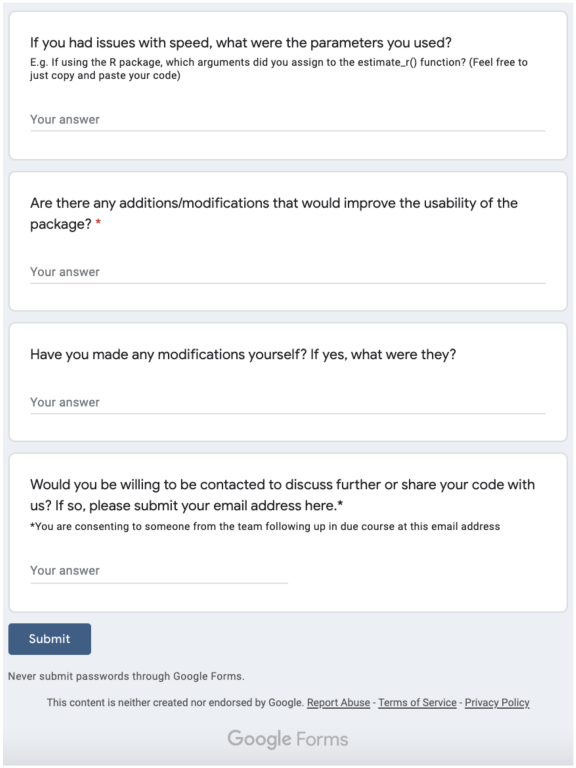
typed answers as opposed to multiple choice, we used example responses to ensure that questions were clear for respondents. These examples were based on issues previously reported in correspondence with users of EpiEstim. It is possible that these examples may have biased the results by making respondents more likely to state these particular issues compared to others.

Table A.3: **Summary of the issues and suggestions reported in questionnaire feedback categorised by broad theme.** n is the number of respondents and % is the percentage of the 17 respondents who reported or made a suggestion regarding the issue.

Theme	**Issue**	**Suggestion**	n	%
Incidence data	• Unable to account for reporting delays or reporting rates • Cannot use weekly aggregated incidence data		2	12
Geographical factors		• Allow users to model different regions at the same time	1	6
Practical or logistical applications		• Extend capability to aid logistical planning in hospitals	1	6
Usability	• Issues understanding how to use the package properly	• Add a pre-processing function to help novice users when setting up parameters • More resources with examples of EpiEstim implementation • Provide more intuitive plotting options	9	53
Speed	• Slow when accounting for uncertainty in the serial interval distribution		3	18
Compatibility	• Unable to use certain software or alternative programming languages	• Examples of how EpiEstim can be used in a workflow with other R packages	3	18

Figure A.7: **Questionnaire**




If you had issues with speed, what were the parameters you used?
E.g. If using the R package, which arguments did you assign to the estimate_r() function? (Feel free to just copy and paste your code)
Your answer
Are there any additions/modifications that would improve the usability of the package? *
Your answer
Have you made any modifications yourself? If yes, what were they?
Your answer
Would you be willing to be contacted to discuss further or share your code with us? If so, please submit your email address here.*
*You are consenting to someone from the team following up in due course at this email address
Your answer
Submit
Never submit passwords through Google Forms.
This content is neither created nor endorsed by Google. Report Abuse - Terms of Service - Privacy Policy
Google Forms

Appendix B

Supplementary Information: Ebola Virus Disease mathematical models and epidemiological parameters: a systematic review and meta-analysis

The contents of this appendix has been adapted from supplementary information previously published in The Lancet Infectious Diseases [188].

B.1 Overview

In this document, we present additional details and results from our systematic review of mathematical models and epidemiological parameters of Ebola Virus Disease (EVD). We first list the 38 known outbreaks of EVD in Appendix B.2. We then present details of the search terms, inclusion and exclusion criteria, data extraction process, and quality assessment strategy (Appendix B.3). Appendix B.4.1 gives an overview of the epidemiological parameters extracted. Subsequent sections (Appendices B.4.2 to B.4.6) detail results on specific parameters, including summaries across all studies. Transmission models identified in our review are described in Appendix B.4.7.

B.2 Ebola outbreaks

Table B.1: **Outbreaks of Ebola Virus Disease up to January 2023.** This table has been adapted from [596] * = laboratory-acquired infection, ** = imported infection, ± = now South Sudan, ⋄ = cases reported in Guinea, Liberia, Sierra Leone, Mali, Nigeria, Senegal, Italy, Spain, UK & USA.

Year	Country	Location	Species
1976	DRC	Mongala Province	Zaire
1976	South Sudan	Western and Central Equatoria State	Sudan
1976	United Kingdom*	Wiltshire, England	Sudan
1977	DRC	Sud-Ubangi Province	Zaire
1979	South Sudan	Western Equatoria State	Sudan
1994	Gabon	Ogooué-Ivindo Province	Zaire
1994	Côte d'Ivoire	Taï National Park	Tai Forest
1995	DRC	Kwilu Province	Zaire
1996	Russia*	Unknown	Zaire
1996(a)	Gabon	Ogooué-Ivindo Province	Zaire
1996(b)	Gabon	Ogooué-Ivindo Province	Zaire
1996	South Africa**	Johannesburg	Zaire
2000-2001	Uganda	Northern and Western regions	Sudan
2001-2002	Gabon & Republic of the Congo	La Zadié, Ivindo and Mpassa Districts (Gabon) and Cuvette-Ouest Department (Republic of the Congo)	Zaire
2003(a)	Republic of the Congo	Cuvette-Ouest Department	Zaire
2003(b)	Republic of the Congo	Cuvette-Ouest Department	Zaire
2004	Sudan±	Western Equatoria State	Sudan
2004	Russia*	Unknown	Zaire
2005	DRC	Cuvette-Ouest Department	Zaire
2007	DRC	Kasai (formerly Kasaï-Occidental)	Zaire
2007	Uganda	Western Region	Bundibugyo
2008-2009	DRC	Kasai (formerly Kasaï-Occidental)	Zaire
2011	Uganda	Central Region	Sudan

Continued on the next page

Year	Country	Location	Species
2012	Uganda	Western Region	Sudan
2012	DRC	Haut-Uélé Province (formerly Orientale)	Bundibugyo
2012-2013	Uganda	Central Region	Sudan
2013-2016	West-Africa$^{\diamond}$		Zaire
2014(b)	DRC	Équateur Province	Zaire
2017	DRC	Bas-Uélé Province	Zaire
2018	DRC	Équateur Province	Zaire
2018-2020	DRC (& Uganda)	North Kivu, Ituri and South Kivu Provinces	Zaire
2020	DRC	Équateur Province	Zaire
2021(a)	DRC	North Kivu Province	Zaire
2021	Guinea	Nzérékoré Region	Zaire
2021(b)	DRC	North Kivu Province	Zaire
2022(a)	DRC	Équateur Province	Zaire
2022(b)	DRC	North Kivu Province	Zaire
2022-2023	Uganda	Central, Eastern and Western regions	Sudan

B.3 Additional Methods

B.3.1 Study Selection

We searched PubMed and Web of Science initially on 8$^{\text{th}}$ March 2019 and then updated our search on 7$^{\text{th}}$ July 2023 using the following search terms: Ebola AND ((transmissi* OR epidemiolog*) OR (model* NOT imag*) OR (severity OR "case fatality ratio*" OR CFR OR "case fatality rate*" OR "mortality rate*" OR "attack rate*") OR ("infectious period*" OR "serial interval*" OR "incubation period*" OR "generation time*" OR "generation interval*" OR "latent period*" OR latency) OR (heterogeneit* OR superspread* OR "super spread*" OR super-spread* OR overdispersion OR overdispersed OR over-dispersion OR over-dispersed OR "over dispersion" OR "over dispersed") OR (infectivity OR infectiousness OR "growth rate*" OR "reproduction number*" OR "reproductive number*" OR R0 OR "reproduction ratio*" OR "reproductive rate*") OR ("pre-existing immunity" OR serological OR serology OR serosurvey*) OR (evolution* OR mutation* OR substitution*) OR (outbreak* OR cluster* OR epidemic*) OR ("risk factor*")).

We used the online platform Covidence to perform title screening, abstract screening, and full text

review. Our inclusion and exclusion criteria are listed in full in Table B.2. Each paper was screened by two reviewers, with disagreements resolved by consensus. No reason was recorded for exclusion at the title and abstract stage. At the full text screening stage, a reason for exclusion was selected from the exclusion criteria.

B.3.2 Inclusion and exclusion criteria

Following our PROSPERO registration CRD42023393345, our inclusion / exclusion criteria are given in Table B.2. We did not extract outbreak size information since the Ebola outbreaks are well documented (see Table B.1 for a full list [596]). We excluded correspondence, letters to editors with the exception of Nature Letters, which published short articles reporting primary research and conference proceedings, unless they were known to be peer-reviewed.

We excluded reviews at the full-text extraction stage but checked the cited literature to ensure that we had included all relevant papers.

Table B.2: **Inclusion and exclusion criteria**

Inclusion	**Exclusion**
• Measures/estimates of human: Reproduction numbers (R, R_0, R_t, r, R_e), growth rate (r), doubling times, generation time, serial interval, incubation/latent period, Case Fatality Ratio (CFR), attack rate, mutation rate (e.g., from phylogenetic study), overdispersion, risk factors (risk and the measure). • Mathematical or statistical model of transmission. • Measures of seroprevalence and negative seroprevalence in humans. • Nature letters.	• Non-English language publication. • Studies of co-infections. • Animal studies. • Qualitative studies (e.g., KAP studies). • Pathogen not the primary focus of study. • Duplicates. • Does not match any of the inclusion criteria. • In-vitro studies. • Non-peer-reviewed publications, conference proceedings, abstracts, posters, letters to the editor.

B.3.3 Backwards citation screening

We identified 179 reviews during our screening process and chose to check we had included all the references from the 11 papers with systematic or meta-analysis in their title that were on relevant topics. We also included one additional review [597] which we knew was highly cited for modelling parameters.

Table B.3: **Review papers used in backwards citation screening**

Parameter	Reviews
Case fatality ratios	Van Kerkhove, 2015 [597]; Rojek, 2019 [598]; Kawuki, 2021 [599]; Nyakarahuka, 2016 [600]; Belhadi, 2022 [601]
Risk factors	Van Kerkhove, 2015 [597]; Brainard, 2016 [602]; Selveraj, 2018 [603]
Incubation period	Van Kerkhove, 2015 [597]; Velasquez, 2015 [209]
Latent period	Velasquez, 2015 [209]
Serial interval	Van Kerkhove, 2015 [597]
Other delays	Van Kerkhove, 2015 [597]
Reproduction number	Van Kerkhove, 2015 [597]
Secondary attack rates	Dean, 2016 [207]
Seroprevalence	Bower, 2017 [212]; Nyakarahuka, 2016 [600]; Rojek, 2019 [598]
Modelling reviews	Wong, 2017 [210]; Abdalla, 2022 [213]

B.3.4 Data extraction

We extracted the data using a bespoke Microsoft Access (Version 2305) database. The extractors extracted data using three forms that then populated the data tables. The exact questions asked in the form and full list of parameters extracted can be found in the Wiki of out GitHub repository for our package (`https://github.com/mrc-ide/epireview`).

Article

We extracted meta information about the paper including the first author, title, journal, digital object identifier (DOI) and page numbers where available to identify the papers. Each article was scored using seven questions about the quality or risk of bias of the paper (see SM Section B.3.5).

Parameters

Our extraction focused on parameters related to seroprevalence, transmission, natural history, severity and evolution. We extracted the following information for each category:

- Seroprevalence
  - seroprevalence in countries with and without reported outbreaks,
  - risk factors for seropositivity.
- Transmission
  - the basic and effective reproduction numbers,
  - growth rate,
  - doubling time,
  - attack rate,

  - secondary attack rate (defined as the probability that an infection occurs among susceptible people within a specific group for example household),
  - overdispersion (defined as overdispersion in the offspring distribution) and
  - risk factors associated for transmission.

- Natural history
  - time periods for the infection process including the serial interval (defined as time between symptoms onset in the infector to symptoms onset in the infectee), generation time (defined as time between infection in the infector and infectee) and incubation period (defined as the time between infection and symptoms onset in an individual),
  - delays beyond symptom onset and
  - delays beyond admission to care.

- Severity
  - case fatality ratios (CFRs) and
  - risk factors associated with severity.

- Evolution
  - evolutionary rates,
  - mutation rates and
  - substitution rates.

We extracted the parameter value type (e.g., mean, standard deviation, median) and uncertainty measure (capturing the precision of estimates) for each parameter when available. If more than three estimates for a single parameter type were presented in a paper, e.g. from different populations, using different methods, or across different time periods, these estimates were extracted as a range of central estimates. Contextual information about the study location and dates, sample size, basic demographic information, and timing of the survey in relation to reported outbreaks were also extracted.

We recorded the methods used for reproduction number estimation (e.g. renewal equations, empirical methods or next generation matrices). For CFRs, we extracted whether the estimation approach was naive or adjusted and recorded numerators and denominators for the CFR and seroprevalence estimates where available. For genomic data, we noted the gene studied if specified and whether the sequence data were available.

We did not extract odds ratio estimates for risk factors because it would be difficult to compare them due to differences in stratifications or reference groups. Instead, we chose to extract an overview of the risk factors explored across studies. This included the outcome of interest e.g. death, the risk

factor for that outcome e.g. sex, whether the estimates were statistically significant and whether they were from univariate or multivariate analyses.

Model

We extracted details of Ebola transmission models that were either fit to data or theoretical only. This included the type of model, transmission routes, whether the model was deterministic or stochastic, assumptions, interventions, the availability of model code, and in the case of a compartmental model, model subclassification (e.g., SIR, SEIR).

We extracted further information from the 37 models that had publicly available code. This included the purpose of the model, health states considered, if uncertainty was included in the model, if the model was fit and the method, the type of data used, the period of the outbreak modelled, if the model was spatial, if it included spillover, what interventions were considered, the language the model was written in, a link to the code and if there were instructions on how to use it.

B.3.5 Quality Assurance (QA)

Given the diversity of papers included in our review, we found that existing quality assessment criteria did not align with our research. We therefore designed a bespoke questionnaire consisting of seven questions SM Table B.4 to assess the quality of each article. Further information about how these questions were interpreted for each parameter can be found in the Wiki of our GitHub repository for our package (`https://github.com/mrc-ide/epireview`).

Table B.4: **Quality assessment questionnaire**

Theme	Questions
Is the methodological/statistical approach suitable? (how the data are used)	• Clear and reproducible • Robust and appropriate for the aim *(subjective criteria)*
Are the assumptions appropriate? (input parameters/assumptions – what goes into the methodology)	• Clear and reproducible • Justified *(published study or analysis of data, objective criteria)*
Are the data appropriate for the selected methodological approach?	• Clearly described and reproducible • Are issues in the data clearly discussed and acknowledged? • Are issues in the data accounted for in the chosen methodological approach?

We observed considerable variability in the QA scores across the extracted parameters (Figure B.1). Among 11 included articles published before 1991, only one had a QA score surpassing 50%. Among the 43 included articles published between 1991 and 2014, no discernible trend in QA scores was observed. However, a surge in articles related to EVD emerged after 2014, coinciding with the West African Ebola Epidemic. Of the 522 articles included in our review, 468 were published after 2014. 36% of these articles had a QA score below 50%.

Given the volume of and variability in the data, we aimed to focus on estimates from high quality articles. Therefore, in the figures and tables presented in the main text, we only included parameters from articles with a QA score of at least 50% (which encompasses just below three-quarters (73%) of all articles). The threshold value was selected to balance the inclusion of as much data as possible with the robustness of our analysis, aiming to minimise the influence of unreliable estimates on the results.

A sensitivity analysis for the meta-analyses showed that the inclusion of parameters from low QA scoring papers produced similar results. All parameters are included in the tables and figures in the Supplementary Material.

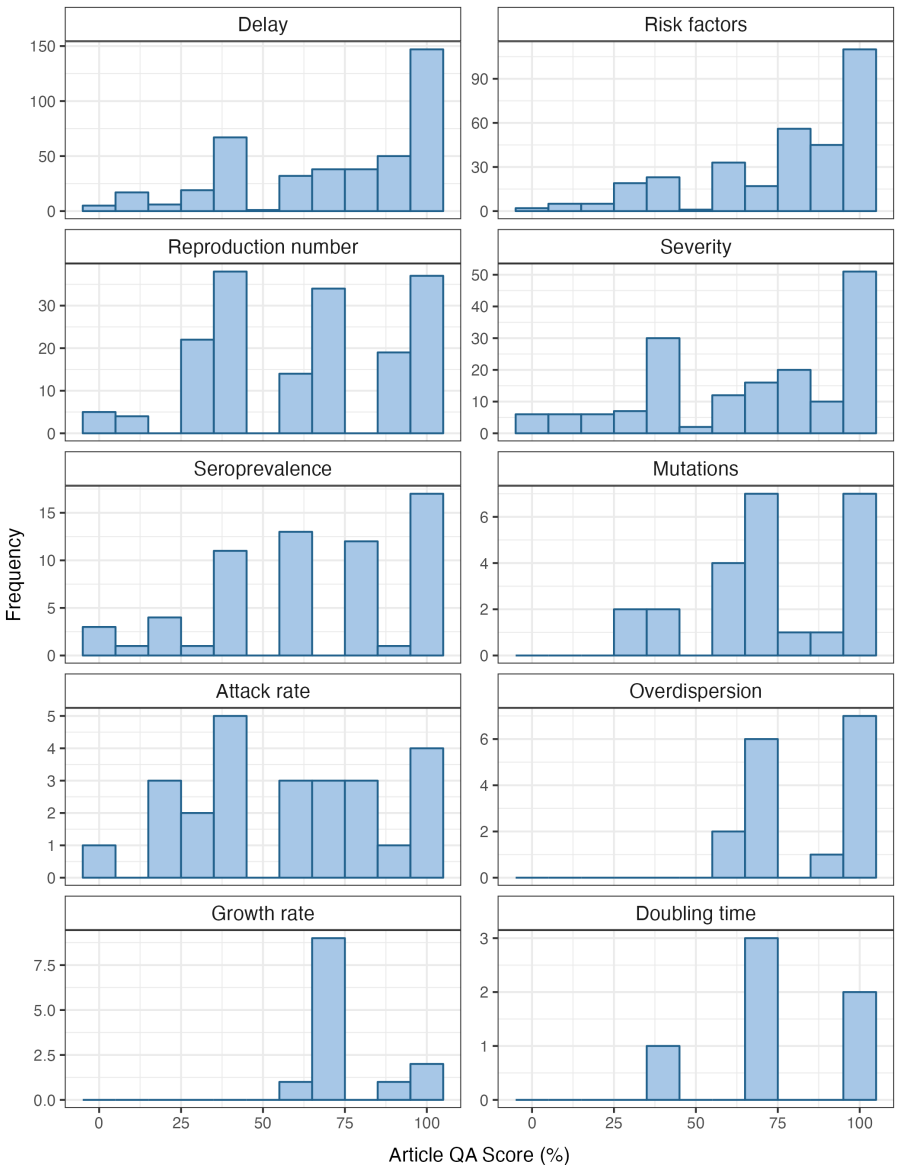


Figure B.1: **QA scores (%) across parameters.**

B.3.6 Data analysis

All analysis was performed in R using orderly2 for workflow management (`https://github.com/mrc-ide/priority-pathogens`) and epireview for review specific functions and data storage.

We did not extract information on EVD outbreaks as they are well documented; instead we assigned an outbreak to each extracted parameter based on the country and time for which the parameter had been estimated.

For meta-analyses, we used the metamean function from the meta R package. [219] Parameters were only included in meta-analyses if studies reported a sample size alongside paired (i) parameter means with a Standard Deviation (SD) of the sample, (ii) means with standard errors, or (iii) medians with a range or interquartile range. For (ii) we converted the standard error to SD using the sample size [604]. For (iii) we used the functionality of metamean to derive the mean and SD using the method outlined by Cai et al. [605], which allows for unknown non-normal distributions. Where data

for multiple EV species were available in this format, we used metamean to perform meta-analyses with species sub-groups with p-values being calculated using a χ^2 test.

For CFR, we chose not to perform a meta-analysis because the study estimates were very context specific. However, we now provide naive syntheses of the heterogeneous estimates, calculated as an unweighted and weighted mean (weighted by the sample size), with recommendation to use the weighted mean.

B.3.7 PRISMA checklist

Table B.5: **PRISMA 2020 Abstracts Checklist**

Section & Topic	Item #	Checklist item	Reported (Yes/No)
Title			
Title	1	Identify the report as a systematic review.	Yes
Background			
Objectives	2	Provide an explicit statement of the main objective(s) or question(s) the review addresses.	Yes
Methods			
Eligibility criteria	3	Specify the inclusion and exclusion criteria for the review.	Yes
Information sources	4	Specify the information sources (e.g. databases, registers) used to identify studies and the date when each was last searched.	Yes
Risk of bias	5	Specify the methods used to assess risk of bias in the included studies.	Yes
Synthesis of results	6	Specify the methods used to present and synthesise results.	Yes
Results			
Included studies	7	Give the total number of included studies and participants and summarise relevant characteristics of studies.	Yes
Synthesis of results	8	Present results for main outcomes, preferably indicating the number of included studies and participants for each. If meta-analysis was done, report the summary estimate and confidence/credible interval. If comparing groups, indicate the direction of the effect (i.e. which group is favoured).	Yes
Discussion			
Limitations of evidence	9	Provide a brief summary of the limitations of the evidence included in the review (e.g. study risk of bias, inconsistency and imprecision).	Yes
Interpretation	10	Provide a general interpretation of the results and important implications.	Yes
Other			
Funding	11	Specify the primary source of funding for the review.	Yes
Registration	12	Provide the register name and registration number.	Yes

Table B.6: **PRISMA 2020 Checklist.** This has been adapted from the published version to align with the location reported within this thesis.

Section & Topic	Item #	Checklist item	Location reported
Title			
Title	1	Identify the report as a systematic review.	Chapter 3
Abstract			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	Table B.5
Introduction			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	Section 3.2
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	Section 3.3
Methods			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	Section 3.3
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	Section 3.3
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	Section 3.3 and Figure 3.1
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	Section 3.3
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	Section 3.3
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	Section 3.3.2 and Appendix B.3.4
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	Section 3.3.2 and Appendix B.3.4

Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	Section 3.3
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	Section 3.3.2
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	Section 3.3.2
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	Section 3.3.2
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	Section 3.3.2
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	Section 3.3.2
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	Section 3.3.2
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	-
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	Section 3.3
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	Section 3.3
Results			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Figure 3.1
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	Available on request
Study characteristics	17	Cite each included study and present its characteristics.	Appendix 2 in [188]
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	Appendix B.3.5
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	Section 3.4
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	Appendix B

	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	Section 3.4
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	Section 3.4
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	Figure B.5
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	Figure B.5
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	Section 3.4
Discussion			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Section 3.5
	23b	Discuss any limitations of the evidence included in the review.	Section 3.5
	23c	Discuss any limitations of the review processes used.	Section 3.5
	23d	Discuss implications of the results for practice, policy, and future research.	Section 3.5
Other Information			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	Section 3.3.1
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	Section 3.3.1
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	n/a
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	[188]
Competing interests	26	Declare any competing interests of review authors.	[188]
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	Section 3.3.3

B.4 Results

B.4.1 Overview of all EVD parameters

Table B.7: **Overview of the total number of parameters extracted within each parameter group** (we do not include parameters in this table where data is available only in the figures).

Parameter Group	Total Parameters
Delay	420
Risk factors	316
Reproduction number	173
Severity	166
Seroprevalence	63
Attack rate	25
Mutations	24
Overdispersion	16
Growth rate	13
Doubling time	6

Table B.8: **Overview of the total number of parameters extracted for each parameter type.** * = "Other delay" encompasses the 32 other delay types that were only extracted from a single study.

Parameter Type	Total Parameters
Delay	
Symptom onset to admission to care	60
Incubation period	52
Symptom onset to death	48
Admission to care to death	33
Other delay*	33
Infectious period	27
Admission to care to death/discharge	22
Serial interval	22
Admission to care to discharge from care	20
Symptom onset to recovery/non-infectiousness	15
Symptom onset to reporting	15
Symptom onset to discharge from care	14
Admission to care to recovery/non-infectiousness	12
Latent period	11
Death to burial	7
Symptom onset to quarantine	5
Symptom onset to seeking care	4
Symptom onset to test	4
Reporting to death	3
Reporting to discharge from care	3
Exposure/infection to admission to care	2
Exposure/infection to death	2
Funeral start to funeral end	2
Symptom onset to antibody detection (IgM/IgG)	2
Symptom onset to other	2
Risk factors	
Risk factors	316
Reproduction number	
Basic reproduction number (R0)	118
Effective reproduction number (Re)	55
Severity	
Case Fatality Ratio (CFR)	166
Seroprevalence	
IgG	36
IFA	13
Unspecified	8
IgM	6
Mutations	
Evolutionary rate	12
Substitution rate	10
Mutation rate	2
Attack rate	
Secondary attack rate	15
Attack rate	10
Overdispersion	
Overdispersion	16
Growth rate	
Growth rate (r)	13
Doubling time	
Doubling time	6

B.4.2 Seroprevalence

Seroprevalence in countries with history of officially reported EVD cases

Table B.9: **Seroprevalence in countries with history of officially reported EVD cases.** * = These are not duplicate entries. One entry corresponds to Zaire antigens, and the other corresponds to Sudan antigens.

Country	Location	Survey date	Central estimate	Central range	Central type	Uncertainty (95% CI)	Population Sample	Sample size	Disaggregated by	Article	QA score (%)
Seroprevalence - IFA											
DRC	Bumba Zone	1 Sep 1977 - 24 Oct 1977					Community based		Age, Other, Sex, Symptoms	International Commission 1978	40.0
DRC	Sud-Ubangi	1981 - 1985	15.00				Contact based	188	Age	Jezek 1999	20.0
DRC	Mabalako, Beni, Bufembo, Katwa	Unspecified	2.30			1.1 - 4	Hospital based	488	Other	Nkuba-Ndaye 2022	80.0
Gabon	Makokou	1994 - Unspecified	0.00				Hospital based			Georges 1999	80.0
Guinea		12 May 2016 - 8 Sep 2017	4.10	3.32 - 20		3.12 - 5.28	Contact based	1,390	Level of exposure, Other, Sex, Symptoms	Diallo 2019	100.0
Sudan	Nzara Maridi	Jun 1976 - Nov 1976		19 - 83			Other		Occupation, Other, Region	WHOInt. Study Team 1978	60.0
Uganda	Karamoja	May 1984	3.00				Community based	132	Region	Rodhain 1989*	20.0
Uganda	Karamoja	May 1984	3.00				Community based	132	Region	Rodhain 1989*	20.0
Seroprevalence - IgG											
DRC	Kikwit	May 1995 - Jul 1995	9.00	4 - 31			Hospital based	429	Occupation	Tomori 1999	60.0
DRC	Kikwit	28 Jul 1995 - 16 Aug 1995	2.20	0 - 11.1			Trade / business based	414	Occupation	Busico 1999	60.0
DRC	Kikwit	18 Aug 1995 - 6 Sep 1995	9.30	0 - 18.1	Other		Community based	161	Age, Occupation, Other, Region, Sex	Busico 1999	60.0
DRC	Watsa region	Aug 2002	18.70		Mean	14.4 - 23.5	Community based	300	Sex	Mulangu 2016	80.0
DRC	Sankuru district	Aug 2007 - Sep 2007	11.00				Community based	3,415		Mulangu 2018	80.0
DRC	Boende, Kabondo-Dianda, Yambuku, North and South Kikwit	Sep 2015 - Aug 2017	8.30		Other		Community based	1,366		Bratcher 2021	100.0
DRC	Boende health zone	Sep 2015 - Nov 2015	28.10				Hospital based	565		Hoff 2019	60.0
DRC	Boende health zone	Sep 2015 - Nov 2015	15.80				Hospital based	565		Hoff 2019	60.0
DRC	Boende	Nov 2015	22.50	0 - 37			Hospital based	582	Age, Level of exposure, Other, Sex	Doshi 2022	80.0
DRC	North Kivu Province, Rubare, Kiwanja Umoja, Biruma, Ntamugenga, Kalengera	May 2017 - Apr 2018	11.03				Household based	272	Region	Goldstein 2020	57.1
DRC	Mbandaka	Jun 2018 - Jul 2018	4.60				Hospital based	539	Other	Shaffer 2022	80.0
DRC	Mbandaka, Equateur Province	12 Nov 2018 - 16 Nov 2018	5.00		Other		Trade / business based	19		Lucas 2020	40.0
Gabon	Lastoursville, Makokou, Mataph-Ngouassa, Moukoro, Doussiéoussou, Doussala	1981 - 1997	1.20	0 - 2.2	Other		Community based	1,147	Region, Sex	Lahm 2007	100.0
Gabon		24 Jan 1996 - 4 Feb 1996	10.20		Mean		Community based	235	Age, Occupation, Sex	Bertherat 1999	40.0
Gabon	Mayibout, Mvadi villages	Feb 1996 - Unspecified		14.9 - 30			Hospital based	205	Region	Georges 1999	80.0
Gabon	Ogooue-Ivindo region	Jul 1997 - Oct 1997	1.00	1 - 1.4	Mean	0.5 - 1.9	Community based	2,533		Hefferman 2005	100.0
Gabon	Rural Gabonese populations	Jun 2005 - Sep 2008	15.30	1.5 - 32.4		14.3 - 16.5	Community based	4,349	Region	Nkoghe 2011	60.0
Gabon		2005 - 2008	15.30				Population based	4,349	Age, Other, Region, Sex	Becquart 2010	100.0
Guinea		Jun 2012 - Oct 2012	0.07				Population based	1,483		Keita 2018	100.0
Guinea		23 Mar 2015 - 11 Jul 2016		59.41 - 99.8			Hospital based		Other, Time	Diallo 2021	100.0
Guinea, Liberia, Sierra Leone	Lassa Diagnostic Laboratory	2007 - 2014	5.20	0 - 10			Other	672	Time	O'Hearn 2016	40.0
Republic of the Congo	Brazzaville, Pointe-Noire, Cuvette-Ouest	Mar 2011 - Jul 2011	2.50	1.6 - 4			Community based	809	Region	Moyen 2015	100.0
Sierra Leone	Bombali District	Feb 2015 - Mar 2015		9.8 - 100			Other		Other, Region	Matopa 2017	60.0
Sierra Leone		Mar 2015 - Dec 2015		21.2 - 27.9			Hospital based	694		Liu 2018	100.0
Sierra Leone		3 Jul 2015 - 10 Sep 2015		0 - 100		0 - 100	Household based	486	Age, Occupation, Other, Sex, Symptoms	Glynn 2017	100.0
Sierra Leone	Sukudu	Oct 2015 - Jan 2016	22.00				Housing estate based	221		Kelly 2018	85.7
Sierra Leone	Kambia District	16 Mar 2016 - 29 Jun 2018	8.40	0 - 27.8	Other	7 - 10	Community based	1,282	Level of exposure, Occupation, Other	Manno 2022	100.0
Sierra Leone	Kono District	Sep 2016 - Jul 2017	9.50				Household based	421		Kelly 2022	40.0
Sierra Leone	Makeni	Mar 2017		40.1 - 97.7	Other		Other	481	Other	Halfmann 2019	40.0
Uganda	Ibanda, Kamwenge, Luwero	Jan 2015 - Feb 2015	2.30				Community based	724	Level of exposure	Nyakarahuka 2020	100.0
Seroprevalence - IgM											
DRC	Kikwit	28 Jul 1995 - 16 Aug 1995	0.00		Other		Trade / business based	414		Busico 1999	60.0
Gabon	Minkebe	27 Dec 1994	27.00				Unspecified	33		Ambiard 1997	20.0
Gabon	Andock, Minkebe, Mékouka, Mvadi, Mayibout villages	Dec 1994 - Unspecified		1.8 - 10			Hospital based		Region, Time	Georges 1999	80.0
Sierra Leone		Mar 2015 - Dec 2015		2.4 - 11			Hospital based	694		Liu 2018	100.0
Seroprevalence - Unspecified											
DRC	Tandala	1978	7.00	1 - 21			Population based	1,096	Age, Region	Heymann 1980	28.6
DRC	Boende health zone	Sep 2015 - Nov 2015	9.50				Hospital based	565		Hoff 2019	60.0
DRC	Boende health zone	Sep 2015 - Nov 2015	2.80				Hospital based	565		Hoff 2019	60.0
Liberia	Bong County	15 Sep 2014 - 4 Jan 2015	42.00				Hospital based	382	Age, Occupation, Other, Region, Sex	Levine 2015	100.0
Sierra Leone	Sukundu	Dec 2014 - Jan 2015	7.50				Contact based	187		Richardson 2016	80.0
South Sudan	Nzara, Yambio	22 Sep 1979 - 9 Oct 1979	18.00	13 - 38			Household based	106	Level of exposure	Baron 1983	14.3

Seroprevalence in countries with no history of officially reported EVD cases

Table B.10: **Seroprevalence in countries with no history of officially reported EVD cases and multi-country analyses**

Country	Location	Survey date	Central estimate	Central range	Central type	Uncertainty (95% CI)	Population Sample	Sample size	Disaggregated by	Article	QA score (%)
Seroprevalence - IFA											
Cameroon	Mbloundou, Lomie, Lolodorf-Bipindi, Pete, Yaounde	1979 - 1980	9.70	3.2 - 23.5	Mean		Community based	1,517	Age, Other, Region	Bourée 1983	40.0
Central African Republic	Nola, Ikaumba, Bozo, Bangassou, Bouar, Obo, Mbre, Birao	Unspecified	21.30				Population based	4,295	Region	Johnson 1993	60.0
Djibouti	Randa, Djibouti City	1987	0.00				Other	160		Salah 1988	0.0
Kenya	Nzoia, Lalsamis, Masinga, Lodwar, Malindi, Kilifi	Unspecified	1.42				Population based	1,899	Region	Johnson 1983	40.0
Multi-country: Africa (n = 6)	Cameroon, Central African Republic, Chad, Republic of the Congo, Equatorial Guinea, Gabon	Jan 1985 - Jun 1987	12.40	1.85 - 32.72	Other		Hospital based	5,070	Region	Gonzalez 1989	42.9
Seroprevalence - IgG											
Central African Republic	Lobaye district, Belemboke, Nola and Bangassou	Dec 1992 - Nov 1995	5.30	1.9 - 16.6			Community based	1,331	Region, Sex, Time	Paul Gonzalez 2000	40.0
Mali	Bamba, Banzana, Soromba	2015		1.5 - 6.1	Other		Other	600	Other	Bane 2021	40.0
Multi-country: Africa (n = 5)	Cameroon, DRC, Republic of the Congo, Ghana, Uganda	1997 - 2012	2.20	0 - 4.4		0.3 - 4.8	Other	2,430	Region	Steffan 2019	100.0
Tanzania		Jun 2018 - Nov 2018	1.80				Other	500	Age, Other, Region, Sex, Symptoms	Rugarabamu 2021	80.0
United Kingdom		16 Dec 2015 - 16 Jun 2016	0.75				Hospital based	268		Houlihan 2017	100.0
Seroprevalence - IgM											
Tanzania		Jun 2018 - Nov 2018	1.60				Other	500	Age, Other, Region, Sex, Symptoms	Rugarabamu 2021	80.0
Tanzania		Jun 2018 - Nov 2018	0.90				Hospital based	308	Age, Region, Sex	Rugarabamu 2022	80.0
Seroprevalence - Unspecified											
Madagascar	Antananarivo, Mandoto, Andasibe, Tsiroanomandidy, Ampijoroa	Unspecified	4.50				Population based	381	Region	Mathiot 1989	0.0
Madagascar	Antananarivo, Mandoto, Andasibe, Tsiroanomandidy, Ampijoroa	Unspecified	0.00				Population based	381	Region	Mathiot 1989	0.0

Risk factors associated with seropositivity

Table B.11: **Risk factors for seropositivity.** Total is the number of parameters extracted for each risk factor for seropositivity.

Risk Factor for Seropositivity	Significant			Not significant			Total
	Adjusted	Not adjusted	Unspecified	Adjusted	Not adjusted	Unspecified	
Age	2	2	4	3	7	2	20
Sex	2	4	2	3	5	3	19
Occupation	1	2	2	1	7	0	13
Contact with animal	3	2	1	3	2	0	11
Close contact	1	2	0	1	4	0	8
Household contact	1	1	0	1	1	1	5
Funeral	1	1	0	1	1	0	4
Non-household contact	0	0	0	0	1	0	1
Other	6	3	4	5	8	1	27

B.4.3 Transmission parameters

Attack rate

Table B.12: **Attack rates**

Country	Survey date	Central estimate (%)	Central range	Uncertainty (95% CI)	Population Sample	Sample size	Population Group	Disaggregated by	Article	QA score (%)
South Sudan, 1976										
Sudan	Jun 1976 - Nov 1976		0.34 - 1.42		Community based	280	Persons under investigation	Region	WHO/Int. Study Team 1978	60.0
DRC, 1995										
DRC	May 1995 - Jul 1995	9	4 - 31		Hospital based	429	Healthcare workers	Occupation	Tomori 1991	100.0
Uganda, 2000-2001										
Uganda	2000		0.01 - 0.238		Community based		Persons under investigation	Region	Okware 2015	20.0
Republic of the Congo, 2005										
Republic of the Congo	18 Apr 2005 - 8 Jul 2005	7.1			Community based		General population	Region	Nkoghe 2011 (b)	42.9
Republic of the Congo	18 Apr 2005 - 8 Jul 2005	8.6			Community based		General population	Region	Nkoghe 2011 (b)	42.9
Republic of the Congo	18 Apr 2005 - 8 Jul 2005	8.7			Community based		General population	Region	Nkoghe 2011 (b)	42.9
Uganda, 2007										
Uganda	29 Nov 2007 - 20 Feb 2008	0.043	0.003 - 0.185		Community based	116	General population	Age, Region, Sex	Wamala 2010	80.0
West Africa 2013-2016										
Guinea	22 Jun 2017 - 9 Jul 2017	7.29		4.38 - 11.28	Household based	247	General population		Timothy 2019	100.0
Liberia	14 May 2014 - 7 Mar 2015	0.25		0.15 - 0.4	Community based	17	General population		Kuehne 2016	28.6
Sierra Leone	May 2014 - Apr 2015	2.7			Household based	1161	General population		Caleo 2018	100.0

Basic reproduction number

Table B.13: **Basic R estimates**

Country	Survey date	Central estimate	Central range	Central type	Uncertainty	Uncertainty type	Method	Disaggregated by	Article	QA score (%)
DRC, 1976										
DRC	Aug 1976 - Nov 1976	4.71	1.34 - 4.71	Median	3.92 - 5.66	95% CI	Compartmental model	Other	Camacho 2014	71.4
DRC, 1995										
DRC	1 Mar 1995 - 16 Jul 1995	1.40		Mean			Compartmental model		Lekone 2006	100.0
DRC	1 Mar 1995 - 12 Jul 1995	1.80		Unspecified			Compartmental model		Zaman 2009	100.0
DRC	1 Mar 1995 - 21 Jul 1995	2.11		Unspecified			Compartmental model		Ndanguza 2013	42.9
DRC	1 Mar 1995 - 21 Jul 1995	2.22		Mean	2.22 - 2.22	95% CrI			Ndanguza 2013	42.9
DRC	1 Mar 1995 - 21 Jul 1995	2.22		Mean	0.13	Standard Deviation	Compartmental model		Ndanguza 2013	42.9
DRC	6 Mar 1995 - 12 Jul 1995	1.83		Median	0.06	Standard Deviation	Compartmental model		Chowell 2004	28.6
DRC	Mar 1995 - Jul 1995		1.5 - 2.08	Mean			Compartmental model	Other	Ward 2023	42.9
DRC	1995	1.93		Median	1.74 - 2.78	IQR	Other		White 2007	85.7
DRC	1995	2.70		Mean	1.9 - 2.8	95% CI	Compartmental model		Legrand 2007	100.0
DRC	1995	2.98		Mean	2.11 - 4.36	95% CrI	Next generation matrix	Occupation	Robert 2019	100.0
Uganda, 2000-2001										
Uganda	20 Aug 2000 - 7 Jan 2001	1.34		Median	0.03	Standard Deviation	Compartmental model		Chowell 2004	28.6
Uganda	2000	2.70		Mean	2.5 - 4.1	95% CI			Legrand 2007	100.0
DRC, 2012										
DRC	2012	0.86		Mean	0.37 - 1.6	95% CrI	Branching process		Choi 2019	42.9
DRC	2012	1.11		Mean	1.06 - 1.17	95% CrI	Other		Choi 2019	42.9
DRC	2012	1.37		Mean	0.85 - 2.02	95% CrI	Branching process		Choi 2019	42.9
West Africa 2013-2016										
Guinea	Dec 2013 - Mar 2016	0.89		Mean			Empirical (contact tracing)	Disease generation	Robert 2019 (b)	100.0
Guinea	1 Jan 2014 - Dec 2014	1.30		Unspecified			Compartmental model		Uekermann 2019	28.6
Guinea	1 Mar 2014 - 18 Mar 2015	1.12		Mean	1.11 - 1.12	95% CI	Compartmental model		Wang 2015	42.9
Guinea	22 Mar 2014 - 29 Aug 2019	1.00		Mean	0.77 - 1.35	Range	Compartmental model		Agusto 2015	57.1
Guinea	22 Mar 2014 - 22 Mar 2015	1.26		Mean	1.22 - 1.29	95% CI	Next generation matrix		Shen 2015	71.4
Guinea	22 Mar 2014 - 29 Aug 2019	1.60		Mean	1.15 - 2.05	Range	Compartmental model		Agusto 2015	57.1
Guinea	22 Mar 2014 - 22 Aug 2014	2.46		Unspecified			Compartmental model		Fisman 2014 (b)	28.6
Guinea	22 Mar 2014 - 25 Jan 2015	4.16		Unspecified			Compartmental model		Li 2015	42.9
Guinea	25 Mar 2014 - 3 May 2015	1.21		Median	1.21 - 1.21	95% CI	Growth rate		Liu 2015	71.4
Guinea	25 Mar 2014 - 25 Oct 2014	1.49		Median	1.48 - 1.5	95% CI	Next generation matrix		Luo 2019	85.7
Guinea	27 May 2014 - 1 Jun 2015		0.85 - 1.22	Unspecified			Compartmental model	Time	Webb 2016	57.1
Guinea	Aug 2014 - May 2015	1.18		Mean	1.17 - 1.19	95% CI	Next generation matrix		Ajelli 2016	71.4
Guinea	Unspecified - 7 Sep 2014	1.24		Unspecified	1.04 - 1.42	95% CI	Other		Taylor 2016	71.4
Guinea, Liberia, Nigeria, Sierra Leone	30 Dec 2013 - 2014		1.2 - 2.02	Mean			Branching process	Region	Aylward 2014	100.0
Guinea, Liberia, Nigeria, Sierra Leone	22 Mar 2014 - 22 Aug 2014	1.78		Unspecified			Compartmental model		Fisman 2014 (b)	28.6
Guinea, Liberia, Sierra Leone	28 Dec 2013 - 3 Oct 2014	1.44	0.75 - 1.92	Other			Compartmental model		Barbarossa 2015	42.9
Guinea, Liberia, Sierra Leone	2013 - 11 Nov 2014	1.21	1.09 - 1.22	Unspecified			Branching process	Region	Evans 2015	42.9
Guinea, Liberia, Sierra Leone	2013 - 2014	1.30	0.36 - 3.37	Mean	0.64	Standard Deviation	Growth rate	Region	Krauer 2016	100.0
Guinea, Liberia, Sierra Leone	1 Jan 2014 - 28 Sep 2014	1.96		Unspecified			Other		Pruyt 2015	71.4
Guinea, Liberia, Sierra Leone	Jan 2014 - Sep 2015		0.98 - 2.39	Unspecified			Next generation matrix	Other, Region	Lin 2020	42.9
Guinea, Liberia, Sierra Leone	22 Mar 2014 - 20 Aug 2014		1.51 - 2.53	Unspecified			Compartmental model	Region	Althaus 2014	42.9
Guinea, Liberia, Sierra Leone	25 Mar 2014 - 3 May 2015	1.35		Median	1.35 - 1.35	95% CI	Growth rate		Liu 2015	71.4

Country	Survey date	Central estimate	Central range	Central type	Uncertainty	Uncertainty type	Method	Disaggregated by	Article	QA score (%)
Guinea, Liberia, Sierra Leone	Mar 2014 - Dec 2014		1.24 - 1.56	Unspecified				Region	Diaz 2018	85.7
Guinea, Liberia, Sierra Leone	6 Jul 2014 - 9 Aug 2014	1.80	1.5 - 2	Unspecified			Compartmental model		Gomes 2014	85.7
Guinea, Liberia, Sierra Leone	2014 - 2015	1.14		Unspecified			Next generation matrix		Long 2018	0.0
Guinea, Liberia, Sierra Leone	2014 - 2015	1.40		Unspecified			Next generation matrix		Denes 2019	57.1
Guinea, Liberia, Sierra Leone	2014		1.4 - 1.9	Unspecified			Compartmental model	Region	Stockdale 2021	42.9
Guinea, Liberia, Sierra Leone	Unspecified - 18 Oct 2014	1.79		Unspecified	1.78 - 1.81	95% CI	Compartmental model		Fisman 2014	28.6
Guinea, Sierra Leone	May 2014 - Jun 2014	5.92		Mean	2.97 - 11.12	HPDI 95%	Genomics		Saulnier 2017	71.4
Liberia	2013 - Dec 2014	2.00		Unspecified			Next generation matrix		Njankou 2018	57.1
Liberia	1 Jan 2014 - Apr 2015	2.50		Unspecified			Compartmental model		Uekermann 2019	28.6
Liberia	1 Mar 2014 - 18 Mar 2015	1.23		Mean	1.23 - 1.23	95% CI	Compartmental model		Wang 2015	42.9
Liberia	17 Mar 2014 - Aug 2014	2.11		Unspecified	1.88 - 2.71	Other	Next generation matrix		Valdez 2015	57.1
Liberia	18 Mar 2014 - 9 Nov 2014	1.40	0.7 - 2	Mean	0.1 - 3.9	95% CI	Compartmental model	Time	Funk 2017	100.0
Liberia	18 Mar 2014 - 9 Nov 2014	4.09		Standard Deviation	1.47 - 8.35	95% CrI	Compartmental model		Funk 2017	100.0
Liberia	23 Mar 2014 - 5 Oct 2014		0.9 - 4.6	Median			Compartmental model	Time	Enanoria 2015	85.7
Liberia	25 Mar 2014 - 3 May 2015	3.02		Median	2.61 - 3.49	95% CI	Growth rate		Liu 2015	71.4
Liberia	27 Mar 2014 - 22 Aug 2014	1.72		Unspecified			Compartmental model		Fisman 2014 (b)	28.6
Liberia	27 Mar 2014 - 25 Oct 2014	1.82		Median	1.77 - 1.83	95% CI	Next generation matrix		Luo 2019	85.7
Liberia	1 May 2014 - 1 Oct 2014	1.76	1.76 - 1.9	Unspecified			Next generation matrix		Khan 2015	28.6
Liberia	2 Jun 2014 - 2015	1.95		Unspecified			Compartmental model		Xie 2019	57.1
Liberia	5 Jun 2014 - 14 Sep 2014		1.71 - 1.94	Unspecified			Compartmental model	Other	Ponce 2019	42.9
Liberia	14 Jun 2014 - 15 Dec 2014	2.49		Unspecified	2.38 - 2.6	95% CI	Next generation matrix		Lewnard 2014	85.7
Liberia	17 Jun 2014 - 23 Sep 2014	1.54		Unspecified			Compartmental model		Webb 2015	71.4
Liberia	17 Jun 2014 - 22 Mar 2015	1.80		Mean	1.77 - 1.83	95% CI	Next generation matrix		Shen 2015	71.4
Liberia	29 Jun 2014 - 7 Oct 2014	2.01		Unspecified			Next generation matrix		Xia 2015	85.7
Liberia	Jul 2014 - Dec 2014	1.70		Median	1.1 - 2.6	95% CI	Empirical (contact tracing)		Lindblade 2015	100.0
Liberia	2014 - 2015	1.67		Median	1.29 - 2.24	95% CrI	Other		Tang 2023	42.9
Liberia	2014	2.22		Unspecified			Compartmental model		Rivers 2014	42.9
Liberia	Unspecified - 1 Oct 2014		1.54 - 1.61	Unspecified			Next generation matrix		Imran 2017	42.9
Liberia	Unspecified - 7 Sep 2014	2.06		Unspecified	1.93 - 2.27	95% CI	Other		Taylor 2016	71.4
Multi-country: Africa, Europe, USA (n = 12)	Jul 2014 - Mar 2015	0.50	0.05 - 2.9	Unspecified	0.2 - 1	90% CI	Branching process	Other	Toth 2015	71.4
Multi-country: Africa, Europe, USA (n = 8)	2014 - 2015	2.10	1.5 - 2.6	Unspecified			Compartmental model	Region	Ivorra 2015	42.9
Nigeria	17 Jul 2014 - 20 Oct 2014	3.20		Mean	0.8 - 16.4	95% CrI	Branching process		Chan 2020	85.7
Nigeria	17 Jul 2014 - 20 Oct 2014	10.00		Mean	3 - 18.5	95% CrI	Branching process		Chan 2020	85.7
Nigeria	20 Jul 2014 - 1 Oct 2014		0.4 - 12	Unspecified			Empirical (contact tracing)	Disease generation	Fasina 2014	71.4
Nigeria	20 Jul 2014 - 20 Oct 2014	9.01		Unspecified	5.22 - 15.55	95% CI	Compartmental model		Althaus 2015	71.4
Nigeria	Jul 2014 - Sep 2014	2.57		Unspecified			Next generation matrix		Do 2016	14.3
Sierra Leone	2013 - Dec 2014	2.50		Unspecified					Njankou 2018	57.1
Sierra Leone	1 Jan 2014 - Feb 2015	1.50		Unspecified			Compartmental model		Uekermann 2019	28.6
Sierra Leone	1 Mar 2014 - 18 Mar 2015	1.18		Mean	1.18 - 1.18	95% CI	Compartmental model		Wang 2015	42.9
Sierra Leone	25 Mar 2014 - 3 May 2015	1.90		Median	1.86 - 1.95	95% CI	Growth rate		Liu 2015	71.4
Sierra Leone	1 May 2014 - 1 Oct 2014	1.49	1.36 - 1.49	Unspecified			Next generation matrix		Khan 2015	28.6
Sierra Leone	12 May 2014 - 13 Nov 2015	1.30		Unspecified	1.26 - 2.53	Range	Other		Chen 2021	71.4
Sierra Leone	19 May 2014 - 11 Jan 2015	1.70		Unspecified			Compartmental model	Region	Li 2017 (b)	57.1

Country	Survey date	Central estimate	Central range	Central type	Uncertainty	Uncertainty type	Method	Disaggregated by	Article	QA score (%)
Sierra Leone	19 May 2014 - 11 Jan 2015	1.90		Unspecified			Other		Li 2020	0.0
Sierra Leone	25 May 2014 - 18 Jun 2014	2.18	1.65 - 2.18	Median	1.24 - 3.55	HPDI 95%	Genomics		Stadler 2014	100.0
Sierra Leone	25 May 2014 - 20 Jun 2014		2.1 - 3.81	Median			Compartmental model		Volz 2014	85.7
Sierra Leone	25 May 2014 - 18 Jan 2015		0.6 - 5.9	Median			Compartmental model	Time	Enanoria 2015	85.7
Sierra Leone	27 May 2014 - 1 Jun 2015		0.72 - 1.77	Unspecified			Compartmental model	Time	Webb 2016	57.1
Sierra Leone	27 May 2014 - 23 Sep 2014	1.26		Unspecified			Compartmental model		Webb 2015	71.4
Sierra Leone	27 May 2014 - 31 Aug 2014	1.29	1.29 - 1.47	Unspecified	1.27 - 1.37	95% CI	Compartmental model		Scarpino 2015	42.9
Sierra Leone	27 May 2014 - 31 Aug 2014	1.40		Unspecified	1.1 - 1.8	HPDI 95%	Genomics		Scarpino 2015	42.9
Sierra Leone	27 May 2014 - 22 Mar 2015	1.61		Mean	1.56 - 1.66	95% CI	Next generation matrix		Shen 2015	71.4
Sierra Leone	27 May 2014 - 25 Oct 2014	1.67		Median	1.59 - 1.76	95% CI	Next generation matrix		Luo 2019	85.7
Sierra Leone	27 May 2014 - 22 Aug 2014	8.33		Unspecified			Compartmental model		Fisman 2014 (b)	28.6
Sierra Leone	May 2014 - Jul 2014		1.21 - 1.6	Median				Other	Duchene 2019	57.1
Sierra Leone	May 2014 - May 2015		1.7 - 4.8	Unspecified			Compartmental model	Region	Getz 2019	71.4
Sierra Leone	Jul 2014 - Aug 2014	1.63		Mean	0 - 4	Range	Empirical (contact tracing)		Ajelli 2015	71.4
Sierra Leone	Jul 2014 - Nov 2014	2.24		Unspecified	1.52 - 4.51	95% CI	Growth rate		Ajelli 2015	71.4
Sierra Leone	Sep 2014 - 2 Feb 2015		0.5 - 8.4	Unspecified			Compartmental model	Region	Kucharski 2015	100.0
Sierra Leone	20 Oct 2014 - 30 Mar 2015	2.00		Unspecified	1.8 - 2.2	95% CI	Other		Lau 2017	85.7
Sierra Leone	20 Oct 2014 - 30 Mar 2015	2.39		Median	2.05 - 2.84	95% CrI	Other		Lau 2017 (b)	100.0
Sierra Leone	28 Nov 2014 - Feb 2015	2.90	2.1 - 4	Mean	1.2 - 6.9	90% CrI	Compartmental model	Time	Funk 2019	100.0
Sierra Leone	1 Dec 2014 - 28 Feb 2015	0.29		Mean	0.11 - 0.55		Empirical (contact tracing)	Level of exposure	Stehling-Ariza 2016	100.0
Sierra Leone	1 Dec 2014 - 28 Feb 2015	0.93		Mean	0.15 - 2.3	95% CI	Empirical (contact tracing)	Level of exposure	Stehling-Ariza 2016	100.0
Sierra Leone	2014 - 2015	1.66		Median	1.31 - 2.15	95% CrI	Other		Tang 2023	42.9
Sierra Leone	2014	1.78		Unspecified			Compartmental model		Rivers 2014	42.9
Sierra Leone	Unspecified - 7 Sep 2014	1.71		Unspecified	1.4 - 1.82	95% CI	Other		Taylor 2016	71.4
Sierra Leone	Unspecified		1.81 - 1.84	Unspecified			Next generation matrix		Imran 2017	42.9
Sierra Leone	Unspecified	2.13		Other			Next generation matrix		Abbate 2016	85.7
Unspecified	Unspecified - 6 Sep 2014		1.51 - 2.26	Unspecified			Unspecified		Hunt 2014	14.3
DRC, 2014										
DRC	26 Jul 2014 - 4 Oct 2014	5.15		Unspecified	3.95 - 6.69	95% CI	Compartmental model		Althaus 2015 (b)	71.4
DRC	2014	0.84		Mean	0.59 - 1.15	95% CrI			Burch 2017	42.9
DRC, 2018-2020										
DRC	3 May 2018 - 12 Sep 2019		1.1 - 1.21	Mean			Compartmental model	Time	Vossler 2022	42.9
DRC	5 Aug 2018 - 2 Feb 2020	1.08		Unspecified			Compartmental model		Potluri 2022	71.4
DRC	5 Aug 2018 - 20 Oct 2019	7.73		Unspecified			Next generation matrix		Mouanguissa 2021	57.1
DRC	Aug 2018 - Sep 2019		0.88 - 1.22	Median	0.76 - 1.31	95% CI	Branching process	Time	Jombart 2020	42.9
DRC	4 Aug 2019 - 10 Jan 2019	1.83		Unspecified			Compartmental model		Hart 2019	71.4
DRC	4 Aug 2019 - 10 Jan 2019	1.90		Unspecified			Compartmental model		Hart 2019	71.4
Uganda, 2022-2023										
Uganda	Aug 2022 - Nov 2022		1.99 - 2.7	Unspecified				Other	Marziano 2023	100.0
Multiple outbreaks										
DRC, Guinea, Liberia, Sierra Leone	1995 - 2016		1.53 - 266.72	Unspecified			Compartmental model	Region, Time	Siewe 2020 (b)	28.6

Table B.14: **Effective R estimates**

Country	Survey date	Central estimate	Central range	Central type	Uncertainty	Uncertainty type	Method	Disaggregated by	Article	QA score (%)
DRC, 1995										
DRC	1 Mar 1995 - 12 Jul 1995	0.73		Unspecified			Compartmental model		Zaman 2009	100.0
West Africa 2013-2016										
Guinea	22 Mar 2014 - 22 Aug 2014	1.04		Mean	0.01	Standard Deviation	Compartmental model		Kiskowski 2014	100.0
Guinea	Mar 2014 - Aug 2014	0.66		Unspecified			Branching process	Other, Region	Kucharski 2016	71.4
Guinea	Mar 2014 - Aug 2014	7.00		Unspecified			Branching process	Other, Region	Kucharski 2016	71.4
Guinea	Apr 2014 - 18 Mar 2015		1.05 - 1.31	Unspecified			Growth rate	Region, Time	Hsieh 2015	28.6
Guinea	Aug 2014 - Feb 2015	2.00		Mean			Empirical (contact tracing)		Valencia 2017	40.0
Guinea	Aug 2014 - Feb 2015	3.00		Mean			Empirical (contact tracing)		Valencia 2017	40.0
Guinea, Liberia, Sierra Leone	30 Dec 2013 - 18 Jul 2015		0 - 3.5	Unspecified			Compartmental model	Other, Region	Santermans 2016	100.0
Guinea, Liberia, Sierra Leone	Dec 2013 - 25 Nov 2014		0.93 - 1	Other			Branching process	Region, Time	Agua-Agum 2015	100.0
Guinea, Liberia, Sierra Leone	2013 - 11 Nov 2014		0.86 - 1.72	Unspecified			Branching process	Region, Time	Evans 2015	42.9
Guinea, Liberia, Sierra Leone	Jan 2014 - Feb 2015	2.30	1 - 7	Mean	1.8 - 2.7	95% CI	Unspecified	Time	Cleaton 2016	42.9
Guinea, Liberia, Sierra Leone	25 Mar 2014 - 22 Mar 2015		0.62 - 1.38	Mean			Growth rate	Region, Time	Wiratsudakul 2016	100.0
Guinea, Liberia, Sierra Leone	1 Jul 2014 - 8 Sep 2014		1.2 - 2.3	Unspecified			Compartmental model	Other, Region	Towers 2014	0.0
Guinea, Liberia, Sierra Leone	30 Dec 2014 - 31 Aug 2014		1.38 - 1.81	Mean				Region	Aylward 2014	100.0
Guinea, Liberia, Sierra Leone	2014 - 2015	1.01		Mean	0.27 - 2.58	Range	Next generation matrix		Denes 2019	57.1
Guinea, Liberia, Sierra Leone	18 Jun 2015 - 4 Aug 2015	2.58	0.64 - 4.74	Unspecified			Empirical (contact tracing)		Tiffany 2017	28.6
Liberia	22 Mar 2014 - 22 Aug 2014	1.17		Mean	0.03	Standard Deviation	Compartmental model		Kiskowski 2014	100.0
Liberia	Apr 2014 - 18 Mar 2015		1.05 - 3.58	Unspecified			Growth rate	Region, Time	Hsieh 2015	28.6
Liberia	27 May 2014 - 21 Dec 2014		0.42 - 2.9	Mean			Compartmental model	Time	Siettos 2015	14.3
Liberia	27 May 2014 - 22 Dec 2014		0.38 - 6.1	Mean			Compartmental model	Other, Time	Russo 2016	42.9
Liberia	7 Jul 2014 - 22 Sep 2014	1.73		Unspecified	1.66 - 1.83	95% CI	Branching process	Symptoms, Other	Yamin 2016	100.0
Liberia	Jul 2014 - Dec 2014	0.10		Median	0.02 - 0.6	95% CI	Empirical (contact tracing)		Lindblade 2015	100.0
Liberia	Dec 2014 - Mar 2015	0.76		Unspecified			Branching process	Other, Region	Kucharski 2016	71.4
Liberia	Dec 2014 - Mar 2015	5.00		Unspecified			Branching process	Other, Region	Kucharski 2016	71.4
Liberia, Sierra Leone	14 Apr 2014 - 11 Oct 2014		1.9 - 2.5	Unspecified			Other	Region	Majumder 2015	28.6
Sierra Leone	22 Mar 2014 - 22 Aug 2014	1.11		Mean	0.02	Standard Deviation	Compartmental model		Kiskowski 2014	100.0
Sierra Leone	27 Apr 2014 - 3 Feb 2016		0.16 - 3.31	Mean			Branching process	Time	Bodine 2018	100.0
Sierra Leone	Apr 2014 - 18 Mar 2015		1.06 - 2.12	Unspecified			Growth rate	Region, Time	Hsieh 2015	28.6
Sierra Leone	22 May 2014 - 3 Dec 2014		1.1 - 2.13	Unspecified			Compartmental model	Time	White 2015	57.1
Sierra Leone	25 May 2014 - 25 Jan 2015		0.44 - 2.22	Mean			Compartmental model	Region, Time	Yang 2015	85.7
Sierra Leone	27 May 2014 - 21 Dec 2014		1.93 - 2.5	Mean			Compartmental model	Time	Siettos 2015	14.3
Sierra Leone	May 2014 - 22 Oct 2014	0.46		Unspecified			Unspecified		Fang 2016	71.4

Country	Survey date	Central estimate	Central range	Central type	Uncertainty	Uncertainty type	Method	Disaggregated by	Article	QA score (%)
Sierra Leone	Jul 2014 - Aug 2014	2.24		Mean			Branching process		Ajelli 2015	71.4
Sierra Leone	11 Oct 2014 - Unspecified	1.30		Unspecified			Empirical (contact tracing)	Other	DeSilva 2019	40.0
Sierra Leone	11 Oct 2014 - Unspecified	4.00		Unspecified			Empirical (contact tracing)	Disease generation	DeSilva 2019	40.0
Sierra Leone	22 Oct 2014 - Sep 2015	0.09		Unspecified			Unspecified		Fang 2016	71.4
Sierra Leone	21 Dec 2014 - 15 Aug 2015		0.68 - 1.98	Mean			Compartmental model	Time	Siettos 2016	42.9
Sierra Leone	2014 - Unspecified	1.26		Unspecified	1.04 - 1.54	HPDI 95%	Genomics		Alizon 2014	0.0
Sierra Leone	18 Jan 2015		0.1 - 0.97	Median			Compartmental model	Region	Camacho 2015	100.0
Sierra Leone	Jul 2015 - Sep 2015	1.20	1.2 - 1.8	Unspecified			Empirical (contact tracing)		Glynn 2018	100.0
Sierra Leone	Oct 2015 - Jan 2016		0.83 - 9	Unspecified			Empirical (contact tracing)	Time	Kelly 2018	85.7
Sierra Leone	Sep 2016 - Jul 2017		0 - 29	Unspecified			Empirical (contact tracing)	Disease generation, Region	Kelly 2022	40.0
DRC, 2014										
DRC	26 Jul 2014 - 7 Oct 2014	0.84		Unspecified	-0.38 - 2.06	95% CI	Empirical (contact tracing)		Maganga 2014	28.6
DRC	26 Jul 2014 - 7 Oct 2014	1.29		Unspecified	-4.71 - 7.29	95% CI	Empirical (contact tracing)		Maganga 2014	28.6
DRC, 2018										
DRC	30 Apr 2018 - 24 May 2018	1.03	0.83 - 1.37	Median	0.83 - 1.37	95% CrI	Branching process		Ahuka-Mundeke 2018	100.0
DRC, 2018-2020										
DRC	30 Apr 2018 - 15 Jan 2019	0.90	0 - 1.8	Median	0.4 - 1.1	95% CI	Branching process		Tariq 2019	100.0
DRC	8 May 2018 - 15 Apr 2019	1.11		Mean			Branching process	Other, Region, Time	Wannier 2019	85.7
DRC	5 Aug 2018 - 27 Jul 2019	2.49		Unspecified			Next generation matrix		Juga 2020	28.6
DRC	2018 - 2020	0.33		Unspecified			Compartmental model		Ouemba Tasse 2022	0.0
DRC	Jan 2019 - Sep 2019		0.4 - 1.1	Median			Next generation matrix	Region, Time	Mizumoto 2019	100.0
DRC	Apr 2019 - Aug 2019	1.06		Mean	1.02 - 1.11	95% CI	Branching process	Other, Region, Time	Kelly 2020	85.7
DRC	12 Jun 2019 - 18 May 2020	1.14		Unspecified			Empirical (contact tracing)		Keita 2023	28.6
DRC	Jul 2019	2.19		Mean			Next generation matrix	Region, Time	Mizumoto 2019	100.0
DRC	29 Sep 2019	1.03		Median	0.59 - 2.12	95% CrI	Next generation matrix	Region	Mizumoto 2019	100.0
DRC	Sep 2019	1.20		Mean			Next generation matrix	Region, Time	Mizumoto 2019	100.0

24% of effective reproduction number estimates in Table B.14 (unfiltered by QA score, n=55) were estimated using branching process models (n=13), 26% compartmental models (n=14), 11% next generation matrices (n=6), 22% contact tracing (n=12), 7% growth rate (n=4), 2% genomics (n=1) and 7% with an unspecified or "other" method (n=4).

Secondary attack rate

Table B.15: **Secondary attack rates.** Note: The three Fang 2016 estimates had no units, so an assumption was made to convert them into percentages.

Country	Survey date	Central estimate (%)	Central range	Uncertainty (95% CI)	Population Sample	Sample size	Population Group	Disaggregated by	Article	QA score (%)
DRC, 1976										
DRC	1976		0.1 - 0.8		Unspecified		Unspecified	Unspecified	Emond 1977	0.0
DRC, 1977										
DRC	1 Sep 1977 - 24 Oct 1977	5.6	3.4 - 27.3		Community based		Other	Disease generation, Level of exposure, Other	International Commission 1978	40.0
South Sudan, 1976										
Sudan	Jun 1976 - Nov 1976	12	0.34 - 14		Household based		Persons under investigation	Disease generation	WHO/Int. Study Team 1978	60.0
Sudan	Jun 1976 - Nov 1976		23 - 81		Household based	17	Persons under investigation	Level of exposure	WHO/Int. Study Team 1978	60.0
DRC 1981-1985*										
DRC	1981 - 1985	15	7 - 19		Contact based	188	General population	Age	Jezek 1999	20.0
DRC, 1995										
DRC	1 Jan 1995 - 3 Jun 1995	16			Household based	200	Persons under investigation		Dowell 1999	85.7
Uganda, 2000-2001										
Uganda	Aug 2000 - Jan 2001	2.5			Contact based	~1000	Persons under investigation	Age, Sex	Okware 2002	40.0
Uganda	21 Oct 2000 - 22 Dec 2000	26			Community based	73	Persons under investigation		Borchert 2011	20.0
West Africa 2013-2016										
Guinea	20 Sep 2014 - 31 Dec 2014		2.1 - 7.9		Community based	152	Persons under investigation	Region	Dixon 2015	80.0
Guinea, Liberia, Sierra Leone	18 Jun 2015 - 4 Aug 2015	65	33 - 89		Contact based	301	Persons under investigation	Region	Tiffany 2017	28.6
Sierra Leone	May 2014 - Sep 2015	5.9	5.6 - 6.2	5 - 7	Household based	1154	Persons under investigation	Other	Fang 2016	71.4
Sierra Leone	May 2014 - 22 Oct 2014	9.3		7.7 - 11	Household based		Persons under investigation		Fang 2016	71.4
Sierra Leone	22 Oct 2014 - Sep 2015	1.7		1 - 3.1	Household based		Persons under investigation		Fang 2016	71.4
Sierra Leone	15 Dec 2014 - 30 Apr 2015	9.9	3.3 - 26.8		Household based	838	Persons under investigation	Age, Level of exposure, Other, Sex, Symptoms, Time	Reichler 2018	80.0
Sierra Leone	Jul 2015 - Sep 2015		18 - 38		Household based	937	Household contacts of survivors	Other	Glynn 2018	100.0

Growth rate

Table B.16: **Growth rates**

Country	Survey date	Central estimate	Unit	Central range	Central type	Uncertainty	Uncertainty type	Disaggregated by	Article	QA score (%)
DRC, 1976										
DRC	1976 - Unspecified	1.43	Per day		Other/Unspecified	0.69 - 2.62	95% CI		Viboud 2016	75.0
West Africa 2013-2016										
Guinea	22 Mar 2014 - 31 Aug 2014	0.01	Per day		Other/Unspecified				Weitz 2015	75.0
Guinea	25 Mar 2014 - 3 May 2015	0.02	Per day		Mean	0	Standard Deviation		Liu 2015	71.4
Guinea, Liberia, Sierra Leone	25 Mar 2014 - 3 May 2015	0.03	Per day		Mean	0	Standard Deviation		Liu 2015	71.4
Liberia	17 Mar 2014 - Aug 2014	0.05	Per day		Other/Unspecified	0	Standard Error		Valdez 2015	57.1
Liberia	25 Mar 2014 - 3 May 2015	0.09	Per day		Mean	0.01	Standard Deviation		Liu 2015	71.4
Liberia	22 Jun 2014 - 31 Aug 2014	0.05	Per day		Other/Unspecified				Weitz 2015	75.0
Sierra Leone	25 Mar 2014 - 3 May 2015	0.05	Per day		Mean	0	Standard Deviation		Liu 2015	71.4
Sierra Leone	28 May 2014 - 31 Aug 2014	0.03	Per day		Other/Unspecified				Weitz 2015	75.0
DRC, 2018-2020										
DRC	6 Sep 2018 - 11 Mar 2019	0.12	Per week		Median	0.11 - 0.14	95% CI		Chowell 2019	85.7
Uganda, 2022-2023										
Uganda	Aug 2022 - Nov 2022	0.08	Per day		Other/Unspecified	0.08 - 0.08	95% CrI		Marziano 2023	100.0
Uganda	Aug 2022 - Nov 2022	0.11	Per day		Other/Unspecified	0.05 - 0.19	95% CrI		Marziano 2023	100.0
Multiple outbreaks										
Uganda, Guinea, Liberia, Sierra Leone	2000 - 2014		Per week	0.08 - 2.38	Other/Unspecified			Region, Time	Viboud 2016	75.0

Doubling time

Table B.17: **Doubling times**

Country	Survey date	Central estimate (Days)	Central range	Uncertainty (95% CI)	Disaggregated by	Article	QA score (%)
West Africa 2013-2016							
Guinea	22 Mar 2014 - 31 Aug 2014	61.0				Weitz 2015	75.0
Guinea, Liberia, Nigeria, Sierra Leone	30 Dec 2014 - 31 Aug 2014		12.84 - 59.75		Region, Time	Aylward 2014	100.0
Guinea, Liberia, Sierra Leone	Dec 2013 - 25 Nov 2014		-3322.71 - 8455.5		Region, Time	Agua-Agum 2015	100.0
Guinea, Sierra Leone	28 Sep 2014 - 11 Nov 2014	22.1		18.9 - 25.59		Tong 2015	42.9
Liberia	22 Jun 2014 - 31 Aug 2014	14.0				Weitz 2015	75.0
Sierra Leone	28 May 2014 - 31 Aug 2014	21.0				Weitz 2015	75.0

Overdispersion

All studies from which overdispersion estimates were extracted assumed a Negative Binomial distribution for the offspring distribution.

Table B.18: **Overdispersion table**

Country	Survey date	Central estimate	Central range	Central type	Uncertainty	Uncertainty type	Method	Disaggregated by	Article	QA score (%)
West Africa 2013-2016										
Guinea	Dec 2013 - Mar 2016	0.310		Mean	0.25 - 0.37	95% CI	Empirical (contact tracing)		Robert 2019 (b)	100.0
Guinea	Mar 2014 - Aug 2014	0.190		Other/Unspecified			Branching process	Other, Region	Kucharski 2016	71.4
Guinea	Mar 2014 - Aug 2014	1.600		Other/Unspecified			Branching process	Other, Region	Kucharski 2016	71.4
Guinea, Liberia, Sierra Leone	Dec 2013 - May 2015	0.030		Other/Unspecified			Branching process		Agua-Agum 2016	100.0
Guinea, Liberia, Sierra Leone	Dec 2013 - May 2015	0.520		Other/Unspecified			Branching process		Agua-Agum 2016	100.0
Liberia	4 Jul 2014 - 2 Sep 2014	2.200		Mean	2 - 2.4	IQR	Branching process		Drake 2015	57.1
Liberia	Dec 2014 - Mar 2015	0.820		Other/Unspecified			Branching process	Other, Region	Kucharski 2016	71.4
Multi-country: Africa, Europe, USA (n = 12)	Jul 2014 - Mar 2015	0.090	0.09 - 0.5	Other/Unspecified	0.03 - 0.2	90% CI	Branching process	Other	Toth 2015	71.4
Sierra Leone	Jul 2014 - Aug 2014	0.450		Other/Unspecified	0.19 - 1.32	95% CI	Empirical (contact tracing)		Ajelli 2015	71.4
Sierra Leone	4 Aug 2014 - 29 Mar 2015	0.065		Mean	0.037 - 0.11	HPDI 95%	Branching process	Time	Zhang 2022	57.1
Sierra Leone	20 Oct 2014 - 30 Mar 2015	0.370		Mean			Other		Lau 2017 (b)	100.0
Sierra Leone	20 Oct 2014 - 30 Mar 2015	0.470		Other/Unspecified			Other		Lau 2017	85.7
DRC, 2018-2020										
DRC	31 Jul 2018 - 26 Apr 2020	0.270		Other/Unspecified	0.2 - 0.33	95% CrI	Empirical (contact tracing)		Polonsky 2021	100.0
Multiple outbreaks										
Multi-country: Africa (n = 6)	2000 - 2015	0.240	0.24 - 2.09	Median	0.06	Standard Error	Empirical (contact tracing)	Other	Taube 2022	71.4

Risk factors associated with transmission

Table B.19: **Risk factors for infection.** Total is the number of parameters extracted for each risk factor for infection.

Risk Factor for Infection	Significant			Not significant			Unspecified significance			Total
	Adjusted	Not adjusted	Unspecified	Adjusted	Not adjusted	Unspecified	Adjusted	Not adjusted	Unspecified	
Close contact	6	4	5	2	1	2	0	0	2	22
Age	5	2	2	0	4	3	1	0	0	17
Sex	2	1	0	5	4	3	1	0	0	16
Funeral	2	3	2	1	2	1	0	0	3	14
Household contact	3	2	0	0	0	1	0	0	1	7
Occupation	2	2	0	1	1	1	0	0	0	7
Non-household contact	1	1	0	0	0	0	0	0	1	3
Hospitalisation	0	2	0	0	0	0	0	0	1	3
Contact with animal	0	0	0	0	1	1	0	0	0	2
Social gathering	0	0	0	0	1	0	0	0	0	1
Other	8	5	5	6	3	6	1	1	2	37

Table B.20: **Risk factors for onward transmission by significance and adjustment.**

Risk Factor for Onward Transmission	Significant	Not significant
	Adjusted	
Age	1	0
First generation of transmission chain	1	0
Funeral	1	0
Hospitalisation	1	0
Sex	1	0
Socioeconomic status	1	0
Survival	1	0
Location (rural/urban)	0	1

B.4.4 Temporal dynamics of disease progression

Time periods for the infection process

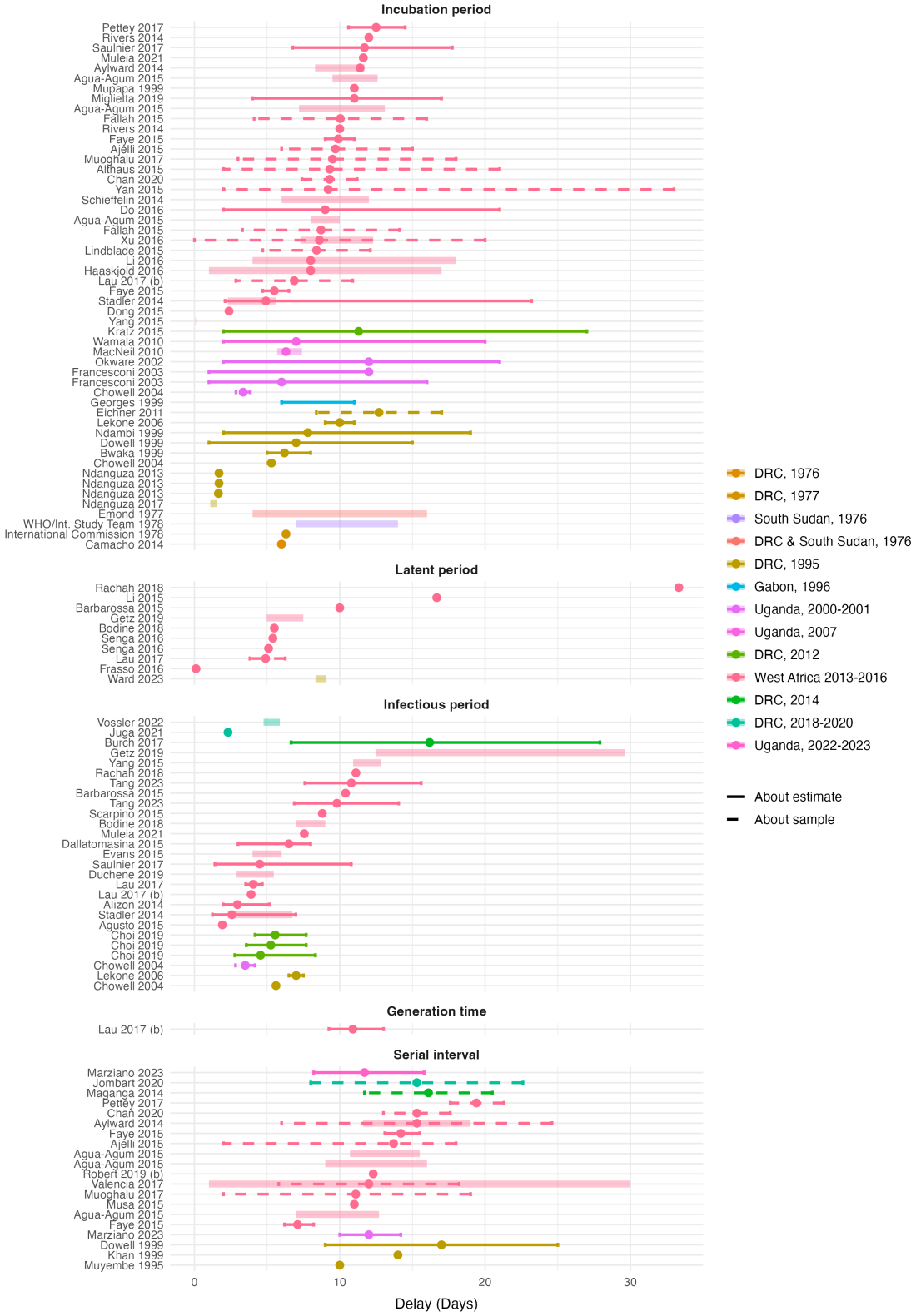


Figure B.2: **Time periods for the infection process**, including the incubation period, latent period, infectious period, generation time and serial interval. This figure excludes one outlier (Martinez et al, 2022). Error bars show uncertainty around the estimate (solid lines) and sample (dashed lines).

Delays beyond symptom onset



Figure B.3: **All delays from symptom onset to various endpoints**, including test, test result, antibody detection, reporting, WHO notification, seeking care, quarantine, admission to care, diagnosis, recovery/non-infectiousness, negative test, undetectable viremia, discharge from care, death, and discharge or death (when no distinction was made between these outcomes). Error bars show uncertainty around the estimate (solid lines) and sample (dashed lines).

Delays beyond admission to care

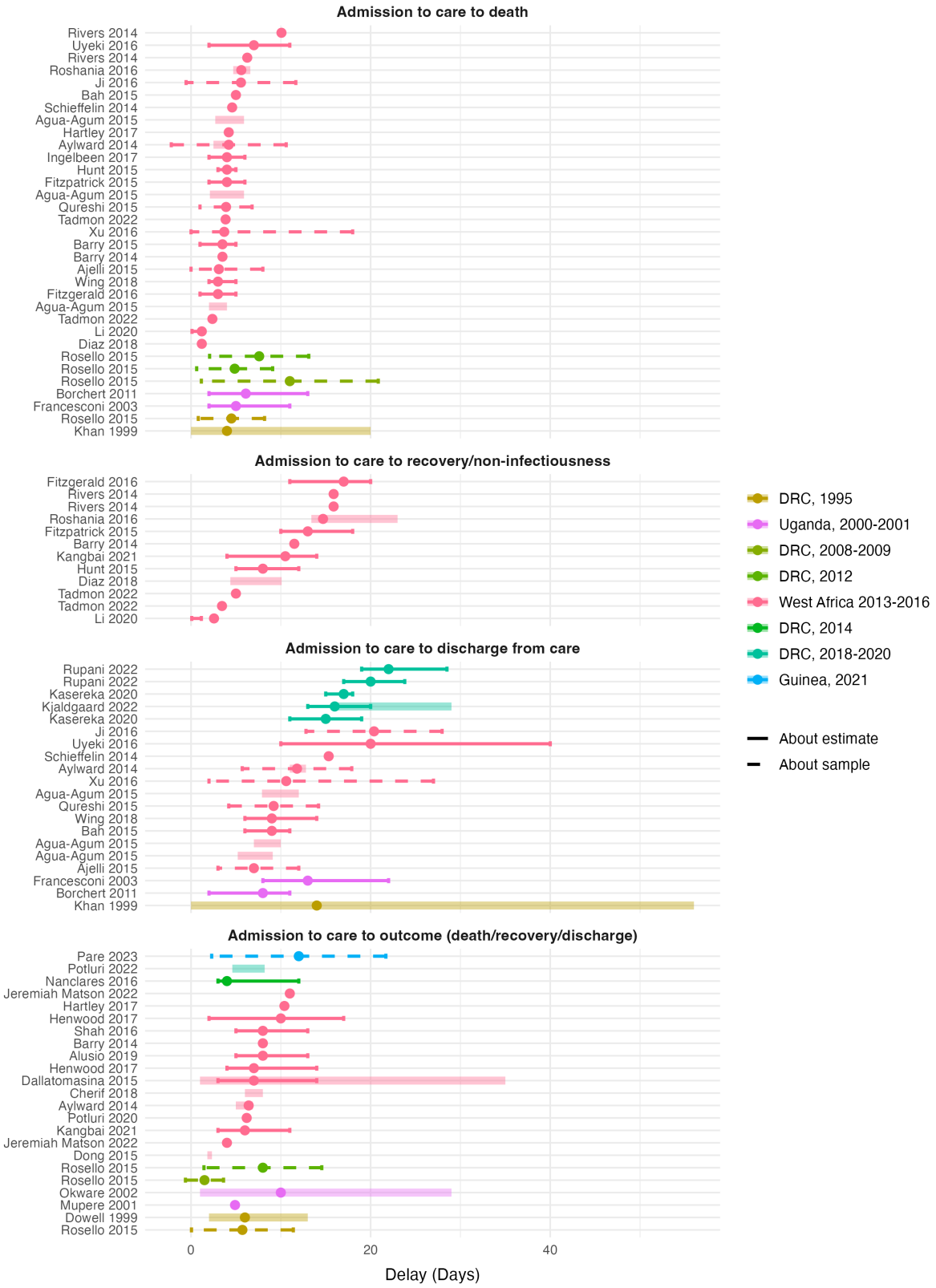


Figure B.4: **Delays beyond admission to care.** Error bars show uncertainty around the estimate (solid lines) and sample (dashed lines)

QA unfiltered meta-analysis by Ebola virus species

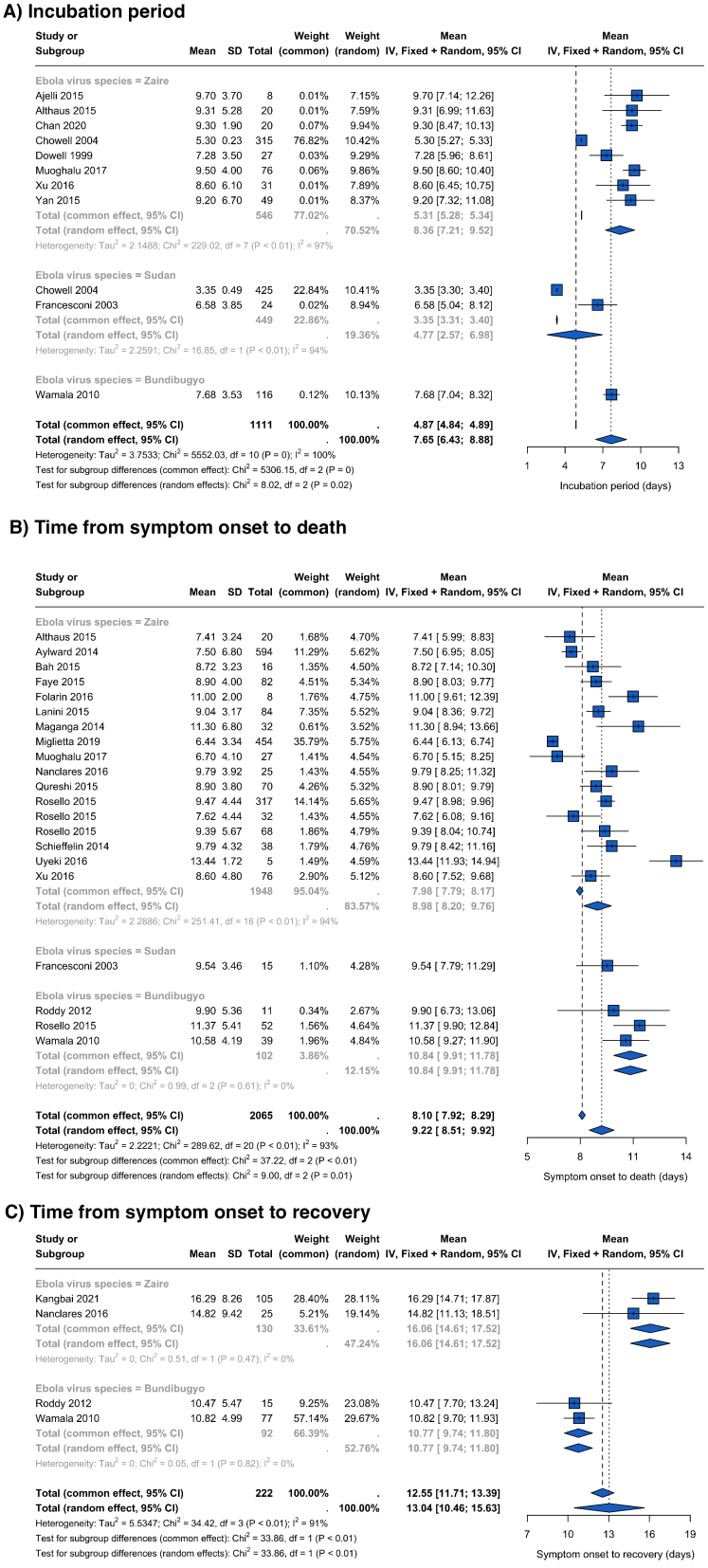


Figure B.5: **Meta-analysis for A) incubation period B) onset-to-death and C) onset-to-recovery regardless of QA score, with a species sub-group analysis**

QA unfiltered meta-analysis

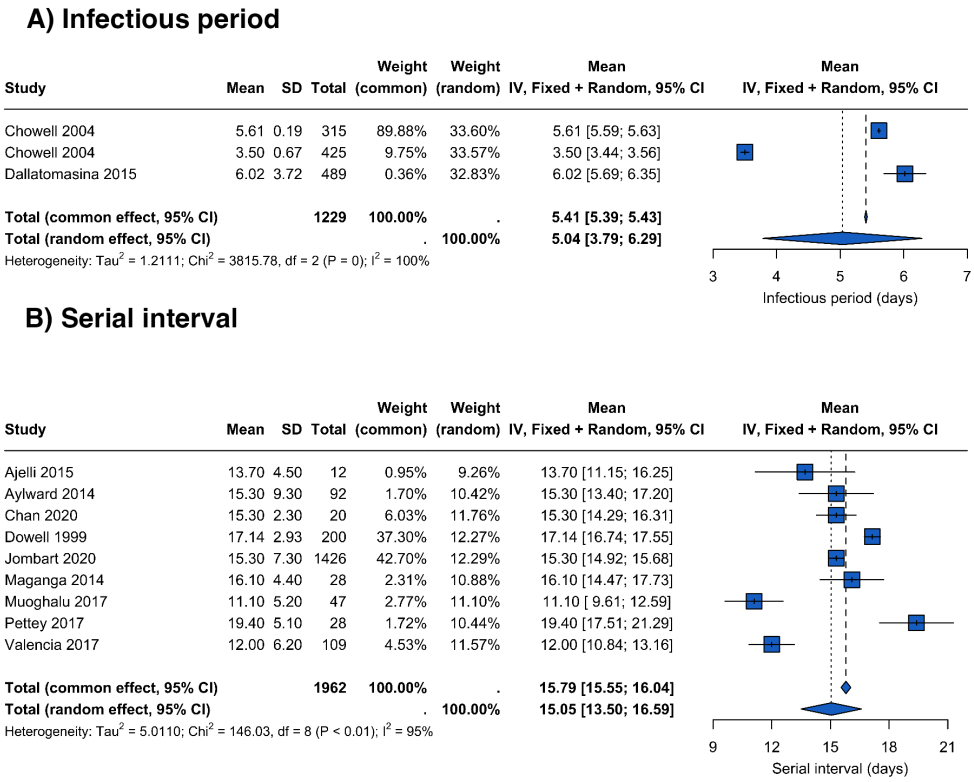


Figure B.6: **Meta-analysis for A) infectious period and B) serial interval regardless of QA score**

B.4.5 Severity

Case Fatality Ratio (CFR)

Table B.21: **Case fatality ratios**

Country	Survey date	Central estimate	Deaths	Cases	Central range	Uncertainty	Uncertainty type	Adjustment	Population Sample	Disaggregated by	Article	QA score (%)
DRC, 1976												
DRC	Aug 1976 - Nov 1976	88				80 - 94	95% CI	Unspecified	Population based		Camacho 2014	71.4
DRC	Aug 1976 - Oct 1976	96	252	262		93 - 98	95% CI	Adjusted	Unspecified	Age, Time	Rosello 2015	100.0
DRC	1 Sep 1976 - 24 Oct 1976	88	280	318				Unspecified	Hospital based		International Commission 1978	40.0
South Sudan, 1976												
Sudan	Jun 1976 - Nov 1976	53	151	284	35 - 67			Naive	Hospital based	Age, Region, Sex, Time	WHO/Int. Study Team 1978	60.0
South Sudan, 1979												
South Sudan	31 Jul 1979 - 6 Oct 1979	65	22	34	0 - 100			Unspecified	Hospital based	Time	Baron 1983	14.3
DRC 1981-1985*												
DRC	1981 - 1985	43	9	21				Naive	Hospital based		Jezek 1999	20.0
Gabon, 1994												
Gabon	Dec 1994 - Feb 1995	59	29	49				Unspecified	Hospital based		Georges 1999	80.0
DRC, 1995												
DRC	1 Jan 1995 - 3 Jun 1995	95	52	55				Unspecified	Household based		Dowell 1999	85.7
DRC	6 Jan 1995 - 16 Jul 1995	81	250	310				Unspecified	Community based		Khan 1999	42.9
DRC	13 Jan 1995 - 24 Aug 1995	77						Unspecified	Population based		Butler 1996	40.0
DRC	Jan 1995 - Jun 1995	78	248	317		73 - 83	95% CI	Adjusted	Unspecified	Age, Time	Rosello 2015	100.0
DRC	1 Mar 1995 - 16 Jul 1995				82 - 83			Unspecified	Population based	Other	Ndanguza 2017	57.1
DRC	9 Apr 1995 - 23 May 1995	74	101	136				Naive	Population based		Muyembe 1995	50.0
DRC	Apr 1995 - Jun 1995	96	14	15				Adjusted	Hospital based		Mupapa 1999	20.0
DRC	Apr 1995 - Unspecified	78	18	23				Unspecified	Hospital based		Ndambi 1999	100.0
DRC	May 1995	81	250	310	69 - 96			Naive	Hospital based	Age, Occupation, Other, Sex, Time	Sadek 1999 (b)	60.0
DRC	1995				54 - 83			Naive	Unspecified	Other	Roels 1999	40.0
Gabon, 1996												
Gabon	Feb 1996 - Unspecified		21	37				Naive	Population based		Amblard 1997	20.0
Gabon	Feb 1996 - Unspecified	68	21	31				Unspecified	Population based		Georges 1999	80.0
Gabon	Jul 1996 - Dec 1996		40	52				Naive	Population based		Amblard 1997	20.0
Gabon	Oct 1996 - Mar 1997	75	45	60				Unspecified	Population based		Georges 1999	80.0
Uganda, 2000-2001												
Uganda	Aug 2000 - Jan 2001	53	224	425				Unspecified			Okware 2002	40.0
Uganda	8 Oct 2000 - 27 Feb 2001	53	224	425	52 - 80			Naive	Population based	Region	Lamunu 2004	40.0
Uganda	21 Oct 2000 - 22 Dec 2000	69	18	26				Naive	Community based	Other	Borchert 2011	20.0
Uganda	Oct 2000 - Feb 2001	40	8	20				Adjusted	Hospital based	Other	Mupere 2001	40.0
Uganda	Oct 2000 - Feb 2001	28	18	64				Adjusted	Hospital based	Other	Mupere 2001	40.0
Uganda	Oct 2000 - Feb 2001	8	7	84				Adjusted	Hospital based	Other	Mupere 2001	40.0
Uganda	2000 - 2001				29 - 77			Unspecified	Unspecified	Age	McElroy 2014	100.0
Uganda	2000				0 - 100			Unspecified			Okware 2015	20.0
Republic of the Congo, 2005												
Republic of the Congo	18 Apr 2005 - 8 Jul 2005	83	10	12				Unspecified	Community based	Region	Nkoghe 2011 (b)	42.9
DRC, 2007												
DRC	Apr 2007 - Oct 2007	74	187			68 - 79	95% CI	Adjusted	Unspecified	Age, Time	Rosello 2015	100.0
Uganda, 2007												
Uganda	Aug 2007 - Dec 2007	32						Unspecified			MacNeil 2011	42.9
Uganda	29 Nov 2007 - 20 Feb 2008	42	11	26				Naive	Hospital based		Roddy 2012	57.1
Uganda	29 Nov 2007 - 20 Feb 2008	25	37	149				Naive	Population based		Roddy 2012	57.1
Uganda	29 Nov 2007 - 20 Feb 2008	34	39	116	29 - 100			Unspecified	Community based	Region, Sex, Other	Wamala 2010	80.0
Uganda	2007 - Unspecified	40	17	43				Adjusted	Community based		MacNeil 2010	85.7
DRC, 2008-2009												
DRC	Nov 2008 - Jan 2009	44	14	32		26 - 62	95% CI	Adjusted	Unspecified	Age, Time	Rosello 2015	100.0

Country	Survey date	Central estimate	Deaths	Cases	Central range	Uncertainty	Uncertainty type	Adjustment	Population Sample	Disaggregated by	Article	QA score (%)
DRC, 2012												
DRC	Jun 2012 - Nov 2012	54	28	52		39 - 68	95% CI	Adjusted	Unspecified	Age, Time	Rosello 2015	100.0
DRC	2 Aug 2012 - 26 Nov 2012	54	28	52	0 - 100			Adjusted	Other	Age, Other, Sex	Kratz 2015	71.4
DRC	Aug 2012 - Oct 2012	74	49	66		62 - 84	95% CI	Adjusted	Unspecified	Age, Time	Rosello 2015	100.0
West Africa 2013-2016												
Guinea	Dec 2013 - 25 Nov 2014	66						Unspecified	Community based		Helleringer 2015	100.0
Guinea	1 Jan 2014 - 29 Mar 2015	46	429	932	40 - 60	43 - 49	95% CI	Adjusted	Hospital based	Region	Rico 2016	80.0
Guinea	1 Jan 2014 - 29 Mar 2015	60	817	1,355	53 - 76			Naive	Community based	Region	Rico 2016	80.0
Guinea	1 Jan 2014 - 29 Mar 2015	40	168	420	23 - 44	35 - 45	95% CI	Adjusted	Hospital based	Region	Rico 2016	80.0
Guinea	1 Jan 2014 - 29 Mar 2015	53	295	553	44 - 60			Naive	Community based	Region	Rico 2016	80.0
Guinea	Jan 2014 - Dec 2015	63	437	695				Adjusted	Hospital based		Cherif 2017	100.0
Guinea	Jan 2014 - Dec 2015	68	1,574	2,313	66 - 81			Adjusted	Hospital based	Age	Cherif 2018	100.0
Guinea	10 Feb 2014 - 11 Aug 2014	54				49 - 63	95% CI	Unspecified	Community based		Faye 2015	85.7
Guinea	1 Mar 2014 - 28 Feb 2015				35 - 49			Unspecified	Hospital based	Time	Faye 2015 (b)	71.4
Guinea	17 Mar 2014 - 29 Mar 2015	60	719	1,205				Naive	Hospital based	Age, Other	Kerber 2016	40.0
Guinea	22 Mar 2014 - 25 Jan 2015	66						Unspecified	Unspecified		Li 2015	42.9
Guinea	25 Mar 2014 - 26 Apr 2014	43	16	37				Adjusted	Hospital based		Bah 2015	28.6
Guinea	25 Mar 2014 - 20 Aug 2014	44	39	90		33 - 54	95% CI	Naive	Hospital based		Barry 2014	14.3
Guinea	Mar 2014 - Unspecified	71						Unspecified	Hospital based		Baize 2014	42.9
Guinea	Mar 2014 - Unspecified	86						Unspecified	Hospital based		Baize 2014	42.9
Guinea	Mar 2014 - Aug 2014	44	39	89				Naive	Hospital based		Barry 2015	57.1
Guinea	Mar 2014 - May 2015	48	70	146	23 - 74			Unspecified	Hospital based	Other, Region, Sex, Symptoms	Sow 2020	80.0
Guinea	20 Sep 2014 - 31 Dec 2014	81	123	152	79 - 84			Naive	Community based	Region	Dixon 2015	80.0
Guinea	29 Nov 2014 - 31 Jan 2015	59						Unspecified	Hospital based		Vernet 2017	71.4
Guinea	1 Dec 2014 - 1 Mar 2015				19 - 81			Unspecified	Population based	Other	Dong 2015	14.3
Guinea	2 Dec 2014 - 23 Feb 2015	61	46	76				Naive	Hospital based		Loubet 2016	0.0
Guinea	22 Jun 2017 - 9 Jul 2017	56	10	18		31 - 78	95% CI	Unspecified	Household based		Timothy 2019	100.0
Guinea, Liberia, Nigeria, Sierra Leone	30 Dec 2013 - 14 Sep 2014	71			32 - 84	69 - 73	95% CI	Adjusted		Age, Occupation, Region, Sex, Symptoms, Time, Other	Aylward 2014	100.0
Guinea, Liberia, Sierra Leone	28 Dec 2013 - 3 Oct 2014	73						Unspecified	Community based		Barbarossa 2015	42.9
Guinea, Liberia, Sierra Leone	28 Dec 2013 - 3 Oct 2014	61						Unspecified	Hospital based		Barbarossa 2015	42.9
Guinea, Liberia, Sierra Leone	30 Dec 2013 - 28 Sep 2015	63			40 - 86	62 - 64	95% CI	Adjusted		Age, Other, Region, Time	Garske 2016	100.0
Guinea, Liberia, Sierra Leone	30 Dec 2013 - 8 Jul 2015				19 - 66			Unspecified	Population based	Region	Santermans 2016	100.0
Guinea, Liberia, Sierra Leone	Dec 2013 - 25 Nov 2014				24 - 58			Naive	Population based	Region, Time	Agua-Agum 2015	100.0
Guinea, Liberia, Sierra Leone	Dec 2013 - 25 Nov 2014				52 - 100			Adjusted	Population based	Age, Occupation, Other, Region, Sex, Time	Agua-Agum 2015	100.0
Guinea, Liberia, Sierra Leone	2013 - 2016	83		33,338		46 - 86	95% CI	Adjusted	Population based	Age, Other, Region, Symptoms	Forna 2019	100.0
Guinea, Liberia, Sierra Leone	2013 - 2016	72		33,338		56 - 80	95% CI	Naive	Population based	Age, Other, Region, Symptoms	Forna 2019	100.0
Guinea, Liberia, Sierra Leone	2013 - 2016	75		18,644		74 - 77	95% CI	Naive	Population based	Age, Other, Region, Symptoms	Forna 2019	100.0
Guinea, Liberia, Sierra Leone	Jan 2014 - Feb 2015	74				68 - 80	95% CI	Unspecified		Age, Level of exposure, Region, Time	Cleaton 2016	42.9
Guinea, Liberia, Sierra Leone	22 Mar 2014 - 20 Aug 2014				48 - 74			Unspecified	Unspecified	Region	Althaus 2014	42.9
Guinea, Liberia, Sierra Leone	24 May 2014 - 30 Sep 2014				39 - 67			Unspecified	Population based	Region	D'Silva 2017	71.4
Liberia	Dec 2013 - 25 Nov 2014	43						Unspecified	Community based		Helleringer 2015	100.0
Liberia	28 Feb 2014 - 1 Dec 2014				42 - 47			Unspecified			Fallah 2015	42.9
Liberia	20 Mar 2014 - 20 Sep 2014	53	330	619				Naive			Ki 2015	0.0
Liberia	20 Mar 2014 - 20 Sep 2014	82						Unspecified			Ki 2015	0.0
Liberia	21 Mar 2014 - 31 Dec 2014	54	229	428		49 - 58	95% CI	Naive		Other	Weppelmann 2016	100.0
Liberia	4 Apr 2014 - 29 Mar 2015	69	2,701	3,897		68 - 71	95% CI	Naive	Population based	Age, Region, Sex, Time	Furuse 2017	100.0
Liberia	14 May 2014 - 7 Mar 2015	59	10	17				Unspecified	Community based		Kuehne 2016	28.6
Liberia	27 May 2014 - 22 Dec 2014				34 - 83			Unspecified	Unspecified	Other, Time	Russo 2016	42.9
Liberia	27 May 2014 - 21 Dec 2014				12 - 70			Adjusted	Population based	Time	Siettos 2015	14.3
Liberia	7 Jul 2014 - 22 Sep 2014	55	76	139	9 - 93	60 - 64	95% CI	Naive		Age, Other, Sex, Symptoms	Yamin 2016	100.0

Country	Survey date	Central estimate	Deaths	Cases	Central range	Uncertainty	Uncertainty type	Adjustment	Population Sample	Disaggregated by	Article	QA score (%)
Liberia	Jul 2014 - Dec 2014	68			50 - 92	60 - 74	95% CI	Unspecified	Hospital based	Age, Other, Time	Lindblade 2015	100.0
Liberia	28 Aug 2014 - 18 Dec 2014				33 - 80			Unspecified	Hospital based	Other	Jeremiah Matson 2022	100.0
Liberia	15 Sep 2014 - 4 Jan 2015	51	82	160				Naive	Hospital based		Levine 2015	100.0
Liberia	2014	50						Unspecified	Unspecified		Rivers 2014	42.9
Liberia, Sierra Leone	15 Sep 2014 - 31 Dec 2015	58	244	424				Unspecified	Hospital based		Alusio 2019	100.0
Liberia, Sierra Leone	15 Sep 2014 - 15 Sep 2015	46	6	13				Adjusted	Hospital based	Other	Henwood 2017	71.4
Liberia, Sierra Leone	15 Sep 2014 - 15 Sep 2015	54	87	162				Unspecified	Hospital based	Other	Henwood 2017	71.4
Liberia, Sierra Leone	15 Sep 2014 - 15 Sep 2015	58	268	465	38 - 93			Adjusted	Hospital based	Age, Region, Sex	Roshania 2016	100.0
Liberia, Sierra Leone	Sep 2014 - Sep 2015	58	244	424				Unspecified	Hospital based	Other	Garbem 2019	50.0
Liberia, Sierra Leone	Sep 2014 - Sep 2015	58						Unspecified	Hospital based	Age, Other, Region, Sex	Peters 2019	85.7
Multi-country: Africa, Europe, USA (n = 10)	Dec 2013 - Mar 2016	40	11,325	28,652				Unspecified		Occupation, Region	Shultz 2016	42.9
Multi-country: Africa, Europe, USA (n = 6)	Jan 2014 - 12 Oct 2014	55	425	236				Unspecified	Unspecified		Hossain 2016	0.0
Multi-country: Europe & USA (n = 9)	Aug 2014 - Dec 2015	18	5	27				Naive	Hospital based		Uyeki 2016	80.0
Nigeria	20 Jul 2014 - 20 Oct 2014	39				14 - 71	95% CrI	Unspecified	Unspecified		Althaus 2015	71.4
Nigeria	20 Jul 2014 - 1 Oct 2014	40	8	20		22 - 61	95% CI	Unspecified	Population based		Fasina 2014	71.4
Nigeria	Jul 2014 - Unspecified	40	7	20	22 - 46			Adjusted	Contact based	Occupation	Musa 2015	60.0
Nigeria	2014				33 - 46			Naive	Population based	Sex	Fawole 2016	40.0
Sierra Leone	Dec 2013 - 25 Nov 2014	31						Unspecified	Community based		Helleringer 2015	100.0
Sierra Leone	1 May 2014 - 31 Jan 2015	69						Unspecified	Hospital based	Occupation	Senga 2016	85.7
Sierra Leone	1 May 2014 - 31 Jan 2015	74						Unspecified	Hospital based	Occupation	Senga 2016	85.7
Sierra Leone	12 May 2014 - 13 Nov 2015				19 - 46			Unspecified		Other, Time	Chen 2021	71.4
Sierra Leone	19 May 2014 - 11 Jan 2015	51				12 - 107	95% CI	Unspecified	Hospital based		Li 2020	0.0
Sierra Leone	23 May 2014 - 31 Jan 2015	74	2,536	3,422		73 - 76	95% CrI	Naive	Population based		Wong 2016	75.0
Sierra Leone	25 May 2014 - 18 Jun 2014	74	64	87	33 - 94			Adjusted	Hospital based	Age, Other, Region, Sex	Schieffelin 2014	100.0
Sierra Leone	27 May 2014 - 21 Dec 2014				24 - 67			Adjusted	Population based	Time	Siettos 2015	14.3
Sierra Leone	28 May 2014 - 31 Dec 2015	46						Unspecified		Other, Time	Potluri 2020	100.0
Sierra Leone	May 2014 - Apr 2015	94	29	31		79 - 99	95% CI	Unspecified	Household based		Caleo 2018	100.0
Sierra Leone	1 Jun 2014 - 7 Nov 2015	37	1,816	4,954	25 - 73			Naive	Population based	Age	Lamunu 2017	40.0
Sierra Leone	23 Jun 2014 - 5 Oct 2014	53	259	488		49 - 58	95% CI	Unspecified	Hospital based	Occupation	Dallatomasina 2015	85.7
Sierra Leone	26 Jun 2014 - 12 Oct 2014	51	270	525	24 - 85			Adjusted	Hospital based	Occupation, Other, Region, Time	Fitzpatrick 2015	80.0
Sierra Leone	26 Jun 2014 - 31 Dec 2014	57	52	91	46 - 82			Unspecified	Hospital based	Age, Other, Region, Sex	Shah 2016	100.0
Sierra Leone	Jun 2014 - Apr 2015				0 - 38			Unspecified	Hospital based	Age, Other, Sex	Kangbai 2019	85.7
Sierra Leone	Jun 2014 - Apr 2015	26	279	1,077	1 - 100			Adjusted	Hospital based	Age, Occupation, Other, Sex, Symptoms	Kangbai 2020	100.0
Sierra Leone	Jun 2014 - Apr 2015	26	248	938	3 - 100			Adjusted	Hospital based	Age, Occupation, Other, Sex, Symptoms	Kangbai 2020 (b)	80.0
Sierra Leone	1 Jul 2014 - 30 Jun 2015	24	93	454	0 - 40			Unspecified	Population based	Region	Miglietta 2019	28.6
Sierra Leone	Jul 2014 - Nov 2014	86	42	49		73 - 94	95% CI	Naive	Population based		Ajelli 2015	71.4
Sierra Leone	14 Aug 2014 - 31 Mar 2015	57	160	282	0 - 100	51 - 63	95% CI	Adjusted	Hospital based	Age, Other, Region, Sex, Symptoms, Time	Fitzgerald 2016	80.0
Sierra Leone	12 Sep 2014 - 23 Feb 2015	44	110	249				Naive	Hospital based	Age, Other, Region, Sex, Symptoms	Theocharopoulos 2017	60.0
Sierra Leone	13 Sep 2014 - 26 Nov 2014	48	99	205	10 - 100			Adjusted	Hospital based	Age, Occupation, Other, Region, Sex	Kangbai 2021	80.0
Sierra Leone	Sep 2014 - Jan 2015	66	142	216				Adjusted	Community based	Other	Crowe 2016	85.7
Sierra Leone	1 Oct 2014 - 21 Mar 2015	49	139	285				Naive	Hospital based		Ji 2016	100.0
Sierra Leone	1 Oct 2014 - 14 Nov 2014	69						Unspecified	Hospital based	Age	Qin 2015	42.9
Sierra Leone	1 Oct 2014 - 9 Nov 2014	60	51	85				Unspecified	Hospital based		Yan 2015	100.0
Sierra Leone	11 Oct 2014 - 9 Dec 2014	51	38	74				Unspecified	Population based		DeSilva 2019	40.0
Sierra Leone	Oct 2014 - Mar 2015	73	98	134				Adjusted	Hospital based	Sex	Li 2016 (b)	100.0
Sierra Leone	Oct 2014 - Apr 2015	66	94	142	55 - 74	58 - 74	95% CI	Unspecified	Population based	Age, Sex	Muoghalu 2017	100.0
Sierra Leone	15 Nov 2014 - 18 Jan 2015	55	76	139	9 - 93			Naive	Hospital based	Age, Other, Sex, Symptoms	Xu 2016	100.0
Sierra Leone	1 Dec 2014 - 28 Feb 2015	64	32	50				Unspecified	Community based	Level of exposure	Stehling-Ariza 2016	100.0
Sierra Leone	1 Dec 2014 - 28 Feb 2015	63	27	43				Unspecified	Community based	Level of exposure	Stehling-Ariza 2016	100.0
Sierra Leone	8 Dec 2014 - 9 Jan 2015	37	55	150	24 - 40			Unspecified	Hospital based	Age, Other, Sex	Hunt 2015	71.4

Country	Survey date	Central estimate	Deaths	Cases	Central range	Uncertainty	Uncertainty type	Adjustment	Population Sample	Disaggregated by	Article	QA score (%)
Sierra Leone	12 Dec 2014 - 14 Mar 2015	46						Unspecified	Hospital based	Age	Rudolf 2017	60.0
Sierra Leone	12 Dec 2014 - 14 Mar 2015	67						Unspecified	Hospital based	Age	Rudolf 2017	60.0
Sierra Leone	13 Dec 2014 - 20 Apr 2015	50	50	101				Naive	Hospital based		Lanini 2015	100.0
Sierra Leone	19 Dec 2014 - 17 Feb 2015	58	18	31	35 - 86			Unspecified	Hospital based	Sex	Haaskjold 2016	85.7
Sierra Leone	21 Dec 2014 - 15 Aug 2015				10 - 39			Unspecified		Time	Siettos 2016	42.9
Sierra Leone	Dec 2014 - Jan 2015	78	28	36		61 - 90	95% CI	Naive	Unspecified		Richardson 2016	80.0
Sierra Leone	2014	75						Unspecified	Unspecified		Rivers 2014	42.9
Sierra Leone	13 Jan 2015 - 5 Apr 2015	57	8	14				Adjusted	Community based		Li 2016	100.0
Sierra Leone	16 Jan 2015 - 10 Mar 2015	47	29	62				Adjusted	Other		Jiang 2017	80.0
Sierra Leone	Jun 2015 - Jan 2016	58	227	395				Naive	Household based		Bower 2016	14.3
DRC, 2014												
DRC	26 Jul 2014 - 7 Oct 2014	74						Unspecified	Community based		Maganga 2014	28.6
DRC	13 Aug 2014 - 8 Sep 2014	62	23	37	62 - 68			Adjusted	Unspecified	Other	Li 2019	71.4
DRC	28 Aug 2014 - 8 Nov 2014	48	12	25				Unspecified	Hospital based		Nanclares 2016	80.0
DRC, 2017												
DRC	27 Mar 2017 - 1 Jul 2017	50	4	8				Naive			Nsio 2019	100.0
DRC, 2018												
DRC	4 Apr 2018 - 27 May 2018	52	25	48				Naive	Community based		Kelly 2019	28.6
DRC	5 Apr 2018 - 30 May 2018	50	25	50	36 - 64	36 - 64	95% CI	Naive	Unspecified		Ahuka-Mundeke 2018	100.0
DRC	5 Apr 2018 - 30 May 2018	56	25	50	39 - 72	39 - 72	95% CI	Adjusted	Unspecified		Ahuka-Mundeke 2018	100.0
DRC, 2018-2020												
DRC	Jul 2018 - May 2020	62	61	579	3 - 94			Unspecified	Hospital based	Age, Occupation	Baller 2022	0.0
DRC	5 Aug 2018 - 2 Feb 2020				50 - 80			Unspecified	Hospital based	Time	Potluri 2022	71.4
DRC	5 Aug 2018 - 2 Feb 2020				50 - 80			Unspecified	Community based	Time	Potluri 2022	71.4
DRC	7 Dec 2018 - 29 Jan 2020	63	157	248				Adjusted	Hospital based		Rupani 2022	80.0
DRC	7 Dec 2018 - 29 Jan 2020	25	34	137				Adjusted	Hospital based		Rupani 2022	80.0
DRC	30 Mar 2019 - 3 Aug 2019	55	117	213				Adjusted	Hospital based		Kasereka 2020	100.0
DRC	30 Mar 2019 - 3 Aug 2019	23	10	44				Adjusted	Hospital based		Kasereka 2020	100.0
DRC	24 Apr 2019 - 14 Oct 2019	45	139	307	41 - 71			Adjusted	Hospital based	Age	Kjaldgaard 2022	100.0
DRC	12 Jun 2019 - 18 May 2020	12	4	32				Unspecified	Other		Keita 2023	28.6
DRC	12 Jun 2019 - 18 May 2020	48	42	87				Unspecified	Community based		Keita 2023	28.6
Uganda	18 Sep 2020 - Dec 2000	71	12	17				Unspecified	Contact based		Francesconi 2003	57.1
Uganda	18 Sep 2020 - Dec 2000	100	10	10				Unspecified	Contact based		Francesconi 2003	57.1
Multiple outbreaks												
DRC	Aug 1976 - Nov 2012	79	526		44 - 96	76 - 82	95% CI	Adjusted	Unspecified	Other	Rosello 2015	100.0
Multi-country: Africa (n = 7)	1976 - 2014	66	1,595	2,411				Unspecified		Region, Time	Shultz 2016	42.9

Risk factors associated with severity

Table B.22: **A) Risk factors associated with death, B) Protective factors associated with recovery.** Total is the number of parameters extracted for each risk/protective factor.

A

Risk Factor for Death	Significant			Not significant			Unspecified significance			Total
	Adjusted	Not adjusted	Unspecified	Adjusted	Not adjusted	Unspecified	Adjusted	Not adjusted	Unspecified	
Age	18	12	11	9	12	4	1	1	0	68
Sex	3	5	2	14	14	7	1	1	0	47
Occupation	1	1	1	6	5	2	0	0	1	17
Hospitalisation	2	2	2	1	0	0	0	0	0	7
Close contact	1	1	0	1	2	0	0	0	0	5
Comorbidity	0	0	0	0	4	1	0	0	0	5
Funeral	1	0	0	0	2	0	0	0	0	3
Household contact	0	0	0	2	1	0	0	0	0	3
Non-household contact	0	0	0	2	0	0	0	0	0	2
Other	26	22	11	21	22	9	2	1	0	114

B

Protective Factor for Recovery	Significant			Not significant			Total
	Adjusted	Not adjusted	Unspecified	Adjusted	Not adjusted	Unspecified	
Age	1	1	0	0	1	1	4
Sex	0	0	0	1	1	1	3
Hospitalisation	0	0	1	0	0	0	1
Other	1	2	0	1	1	0	5

B.4.6 Mutation rates

Table B.23: **Mutation rates**

Country	Survey date	Central estimate	Central range	Central type	Uncertainty	Uncertainty type	Genome site	Population Sample	Sample size	Disaggregated by	Species	Article	QA score (%)
Evolutionary rate (Substitutions/site/year (10^{-4}))													
DRC	May 2007 - Feb 2009		1.5 - 7.5	Other/Unspecified			Whole genome, Glycoprotein, Nucleoprotein	Other			Zaire	Girard 2011	57.1
DRC, Guinea, Liberia, Sierra Leone	2013 - 2016	4.7	1 - 10	Mean	3.4 - 5.7	95% CrI	Whole genome, Partial sequences		85	Other	Zaire	Li 2019	71.4
Guinea, Liberia, Sierra Leone	2013 - 2015	7.3		Other/Unspecified			Glycoprotein		600		Zaire	Dudas 2019	71.4
Guinea, Liberia, Sierra Leone	2013 - 2015	8.2		Other/Unspecified			Whole genome		600		Zaire	Dudas 2019	71.4
Guinea, Mali, Sierra Leone	Dec 2013 - Jan 2015	14.2		Mean	12.2 - 16.2	95% CrI	Whole genome		179		Zaire	Carroll 2015	57.1
Guinea, Sierra Leone	2014 - 2015	12.5		Mean			Partial genome		318		Zaire	Park 2015	100.0
Sierra Leone	May 2014 - Sep 2015	9.6	5 - 9.63	Mean	8.6 - 10.6	95% CrI	Whole genome			Other	Zaire	Whitmer 2018	100.0
Guinea	12 Feb 2021 - 4 Mar 2021	11		Other/Unspecified			Whole genome, Partial genome		12		Zaire	Kalta 2021	71.4
DRC	Aug 1976 - Aug 2018		2.4 - 8.6	Mean			Whole genome	Population based	15	Other, Time	Zaire	Mbala-Kingebeni 2019	100.0
Multi-country: Africa (n = 5)	1976 - 2014	10.75		Median	9.32 - 12.2	HPDI 95%	Glycoprotein		65	Other	Zaire	Azaran 2015	100.0
Multi-country: Africa, Asia, Europe, USA (n = 9)	1976 - 2008	10.93	7.66 - 13.94	Other/Unspecified	0.52 - 24.61	HPDI 95%	Glycoprotein				All species	Li 2014	100.0
Multi-country: Africa, Asia, Europe, USA (n = 9)	1976 - 2008	23.14		Other/Unspecified	16.49 - 31.66	HPDI 95%	Polymerase (L)				All species	Li 2014	100.0
Mutation rate (Substitutions/site/year (10^{-4}))													
Liberia	Sep 2014 - Feb 2015	9.17		Other/Unspecified	0.523	Other	Whole genome	Hospital based	25	Other	Zaire	Kugelman 2015	28.6
Liberia	Sep 2014 - Feb 2015		9.44 - 15.67	Other/Unspecified			Whole genome	Hospital based	25	Other	Zaire	Kugelman 2015	28.6
Substitution rate (Substitutions/site/year (10^{-4}))													
DRC	3 Nov 2007 - 10 Sep 2012	8.4		Mean	1.1 - 21	95% CrI	Whole genome		12		Bundibugyo	Hulsebeng 2021	100.0
Guinea	Feb 2014 - Unspecified	10.7		Mean			Glycoprotein				Zaire	Dudas 2014	60.0
Guinea	Unspecified	2		Mean	1 - 3	HPDI 95%	Partial genome		49		All species	Membrebe 2019	57.1
Guinea, Mali, Sierra Leone	Mar 2014 - Dec 2014		8.7 - 9.1	Mean	6.9 - 11	95% CrI	Whole genome	Hospital based	195		Zaire	Simon-Loriere 2015	71.4
Guinea, Mali, Sierra Leone	2014	13		Mean			Whole genome		106		Zaire	Hoenen 2015	83.3
Guinea, Sierra Leone	28 Sep 2014 - 11 Nov 2014	12.3		Other/Unspecified	10.4 - 14.1	HPDI 95%	Whole genome	Community based	175		Zaire	Tong 2015	42.9
DRC	8 May 2018 - 24 Jul 2018	36		Mean	12.2 - 59.9	HPDI 95%	Whole genome		15		Zaire	Mbala-Kingebeni 2019 (b)	71.4
Multi-country: Africa (n = 5)	15 Aug 1976 - 4 Jun 2018	6.9		Mean	5.5 - 8.3	HPDI 95%	Whole genome		50		Zaire	Mbala-Kingebeni 2019 (b)	71.4
Multi-country: Africa, Asia, Europe, USA (n = 6)	1976 - 1995	0.36		Mean	0.109	Standard Error	Glycoprotein				Unspecified	Suzuki 1997	42.9
DRC	Unspecified	1.4		Mean	0.473 - 2.45	95% CrI	Whole genome				Zaire	Vrancken 2020	85.7

B.4.7 Transmission Models

Table B.24: **Summary of EVD model types.**

Model type	Total
Compartmental	210
Other or combination	49
Branching process	19
Agent/Individual based	17

Table B.25: **Summary of EVD model assumptions.** Total is the number of models with this assumption extracted (assumptions were not extracted for all models).

Assumptions	Total
Compartmental	
Homogeneous mixing	83
Heterogeneity in transmission rates between groups	43
Heterogeneity in transmission rates over time	42
Latent period the same as incubation period	27
Other or combination	
Heterogeneity in transmission rates between groups	15
Heterogeneity in transmission rates over time	11
Homogeneous mixing	9
Latent period the same as incubation period	5
Age dependent susceptibility	1
Branching process	
Heterogeneity in transmission rates over time	4
Heterogeneity in transmission rates between groups	3
Homogeneous mixing	1
Agent/Individual based	
Heterogeneity in transmission rates between groups	9
Heterogeneity in transmission rates over time	4
Latent period the same as incubation period	3
Age dependent susceptibility	2

Table B.26: **Models overview.** * = Zaire, ** = Bundibugyo, *** = Bundibugyo, Sudan, Taï Forest and Zaire

Stochastic or Deterministic	Theoretical model	Interventions	Code available	Assumptions	Article	QA score (%)
Agent/Individual based						
Stochastic	Yes	Behaviour changes, Contact tracing, Hospitals, Safe burials, Treatment Centres		Age dependent susceptibility, Heterogeneity in transmission rates between groups	Ajelli 2018	100.00
Stochastic	Yes	Vaccination		Heterogeneity in transmission rates between groups, Unspecified	Bisanzio 2023	100.00
Stochastic				Heterogeneity in transmission rates between groups and over time	Cope 2014	100.00
Stochastic		Hospitals, Other		Unspecified	Fadlkar 2018	100.00
Stochastic			Yes	Heterogeneity in transmission rates between groups, Latent period the same as incubation period	Kahn 2020	100.00
Stochastic		Other, Safe burials, Treatment Centres		Heterogeneity in transmission rates between groups	Merler 2015	100.00
Stochastic		Vaccination		Heterogeneity in transmission rates between groups	Merler 2016	100.00
Stochastic		Contact tracing, Hospitals, Safe burials		Heterogeneity in transmission rates over time	Rizzo 2016	100.00
Stochastic		Hospitals		Unspecified	Gomes 2014	85.71
Stochastic				Unspecified	Gomez-Barroso 2017	80.00
Stochastic		Contact tracing, Hospitals, Safe burials, Treatment Centres, Vaccination		Age dependent susceptibility, Heterogeneity in transmission rates between groups	Ajelli 2016	71.43
Stochastic	Yes			Heterogeneity in transmission rates between groups	Kustudic 2021	50.00
Stochastic				Unspecified	Fallah 2015	42.86
Stochastic		Other, Safe burials		Heterogeneity in transmission rates over time, Latent period the same as incubation period	Sietios 2016	42.86
Stochastic				Heterogeneity in transmission rates between groups and over time, Latent period the same as incubation period	Uekermann 2019	28.57
Stochastic	Yes	Behaviour changes, Hospitals, Other, Quarantine, Safe burials		Unspecified	Venkatramanan 2018	25.00
				Unspecified	Khaleque 2017	0.00
Branching process						
Stochastic				Unspecified	Djaafara 2021	100.00
Stochastic		Quarantine, Vaccination	Yes	Heterogeneity in transmission rates between groups and over time	Green 2022	100.00
Stochastic				Unspecified	Holbrook 2022	100.00
Stochastic			Yes	Heterogeneity in transmission rates over time	Kelly 2019 (b)	100.00
Stochastic				Unspecified	Nouvellet 2018	100.00
Stochastic		Other, Quarantine	Yes	Heterogeneity in transmission rates between groups	Yamin 2016	100.00
Stochastic				Unspecified	Nishiura 2014	75.00
Stochastic		Quarantine, Safe burials		Unspecified	Fang 2016	71.43
Stochastic		Vaccination		Unspecified	Kucharski 2016	71.43
Stochastic				Unspecified	Toth 2015	71.43
Stochastic		Hospitals, Safe burials		Heterogeneity in transmission rates between groups	Drake 2015	57.14
Stochastic				Heterogeneity in transmission rates over time	Zhang 2022	57.14
Stochastic			Yes	Unspecified	Lee 2022	50.00
Stochastic				Unspecified	Choi 2019**	42.86
Stochastic				Unspecified	Jombart 2020	42.86
Stochastic		Vaccination		Heterogeneity in transmission rates over time	Worden 2019	42.86
Stochastic		Vaccination		Homogeneous mixing	Kelly 2019	28.57
Stochastic				Unspecified	Champredon 2018	25.00
Stochastic	Yes		Yes	Unspecified	Park 2019	25.00
Compartmental						
Deterministic	Yes			Homogeneous mixing, Latent period the same as incubation period, Unspecified	Almocera 2019	100.00
Deterministic	Yes			Heterogeneity in transmission rates between groups	Altaf 2019	100.00
Deterministic	Yes	Other, Vaccination		Homogeneous mixing	Anguelov 2020	100.00
Stochastic				Homogeneous mixing	Asher 2018	100.00
Deterministic	Yes			Homogeneous mixing	Baldassi 2016*	100.00

Stochastic or Deterministic	Theoretical model	Interventions	Code available	Assumptions	Article	QA score (%)
Deterministic	Yes			Homogeneous mixing, Latent period the same as incubation period	Berge 2016	100.00
Deterministic	Yes			Homogeneous mixing, Latent period the same as incubation period	Berge 2018	100.00
Deterministic	Yes	Contact tracing, Quarantine		Homogeneous mixing	Berge 2018 (b)	100.00
Deterministic	Yes			Heterogeneity in transmission rates between groups, Homogeneous mixing	Berge 2018 (c)	100.00
Deterministic	Yes			Heterogeneity in transmission rates between groups, Homogeneous mixing	Berge 2018 (c)	100.00
Deterministic	Yes	Behaviour changes, Other, Vaccination		Homogeneous mixing	Berge 2018 (d)	100.00
Deterministic	Yes	Vaccination		Heterogeneity in transmission rates between groups, Homogeneous mixing	Bhunu 2016	100.00
Deterministic	Yes	Hospitals, Quarantine, Safe burials		Heterogeneity in transmission rates between groups and over time	Blackwood 2016	100.00
Deterministic		Vaccination		Heterogeneity in transmission rates between groups, Latent period the same as incubation period	Bodine 2018	100.00
Stochastic		Hospitals, Safe burials, Treatment Centres		Heterogeneity in transmission rates over time	Camacho 2015	100.00
Deterministic				Homogeneous mixing	Chapwanya 2022	100.00
	Yes	Treatment, Vaccination		Heterogeneity in transmission rates between groups and over time	De la Sen 2017 (b)	100.00
Deterministic	Yes	Safe burials, Treatment, Vaccination		Unspecified	De la Sen 2019	100.00
Stochastic				Homogeneous mixing	Frasso 2016	100.00
Stochastic		Hospitals		Heterogeneity in transmission rates over time	Funk 2017	100.00
Stochastic			Yes	Heterogeneity in transmission rates over time	Funk 2018	100.00
Stochastic			Yes	Heterogeneity in transmission rates over time	Funk 2019	100.00
Stochastic	Yes	Other, Vaccination	Yes	Unspecified	Getz 2015	100.00
Deterministic	Yes	Hospitals, Quarantine, Treatment		Unspecified	Gutfraind 2015	100.00
Stochastic	Yes			Homogeneous mixing	Kamara 2020	100.00
Stochastic		Behaviour changes, Hospitals, Quarantine		Heterogeneity in transmission rates between groups	Ko 2019	100.00
Stochastic			Yes	Unspecified	Kramer 2016	100.00
Stochastic		Treatment Centres		Heterogeneity in transmission rates over time	Kucharski 2015	100.00
Deterministic				Unspecified	Leander 2016	100.00
Deterministic		Vaccination		Heterogeneity in transmission rates between groups and over time, Homogeneous mixing, Latent period the same as incubation period	Lee 2019	100.00
Stochastic		Hospitals, Safe burials		Homogeneous mixing	Legrand 2007	100.00
Stochastic	Yes		Yes	Homogeneous mixing	Lekone 2006	100.00
Deterministic		Other, Quarantine, Treatment, Vaccination		Homogeneous mixing	Mbah 2022	100.00
Deterministic	Yes	Behaviour changes, Safe burials		Unspecified	Mhlanga 2019	100.00
	Yes			Unspecified	Nazir 2020	100.00
Stochastic	Yes			Homogeneous mixing	Nieddu 2017	100.00
Deterministic	Yes	Other		Unspecified	Njankou 2022	100.00
Deterministic	Yes			Heterogeneity in transmission rates between groups	Njankou 2023	100.00
Stochastic		Hospitals, Vaccination		Heterogeneity in transmission rates between groups and over time, Homogeneous mixing	Potluri 2020	100.00
Stochastic				Heterogeneity in transmission rates between groups and over time, Homogeneous mixing, Latent period the same as incubation period	Riad 2019	100.00
Deterministic		Vaccination		Heterogeneity in transmission rates between groups and over time	Robert 2019	100.00
Deterministic			Yes	Heterogeneity in transmission rates over time, Homogeneous mixing	Santermans 2016	100.00
Deterministic	Yes			Heterogeneity in transmission rates between groups	Seidu 2020	100.00
Stochastic				Unspecified	Shakiba 2021	100.00
Deterministic	Yes	Other		Unspecified	Siewe 2020	100.00
Deterministic			Yes	Unspecified	Stadler 2014	100.00
Deterministic	Yes			Latent period the same as incubation period	Verma 2019	100.00
	Yes	Hospitals, Safe burials		Unspecified	Wang 2017	100.00
Deterministic				Unspecified	Wiratsudakul 2016	100.00

Stochastic or Deterministic	Theoretical model	Interventions	Code available	Assumptions	Article	QA score (%)
		Treatment		Unspecified	Zaman 2009	100.00
Deterministic			Yes	Homogeneous mixing	Abbate 2016	85.71
Deterministic	Yes	Behaviour changes	Yes	Homogeneous mixing	Abbate 2016	85.71
Deterministic		Hospitals, Safe burials		Heterogeneity in transmission rates between groups, Homogeneous mixing	Diaz 2018	85.71
Stochastic				Heterogeneity in transmission rates between groups	Enanoria 2015	85.71
Deterministic		Other, Treatment Centres		Homogeneous mixing	Lewnard 2014	85.71
Stochastic				Unspecified	Luo 2019	85.71
Stochastic			Yes	Heterogeneity in transmission rates between groups and over time	Muleia 2021	85.71
Deterministic		Safe burials		Latent period the same as incubation period	Sofonea 2018	85.71
Stochastic		Quarantine, Safe burials, Treatment		Heterogeneity in transmission rates over time	Volz 2014	85.71
Deterministic		Quarantine, Safe burials, Treatment		Heterogeneity in transmission rates over time	Volz 2014	85.71
Deterministic		Quarantine, Safe burials, Treatment		Unspecified	Volz 2014	85.71
		Hospitals, Safe burials		Unspecified	Xia 2015	85.71
Deterministic				Heterogeneity in transmission rates over time, Homogeneous mixing	Yang 2015	85.71
Stochastic				Unspecified	Gomez-Barroso 2017	80.00
Deterministic	Yes	Quarantine		Unspecified	Abbasi 2020	75.00
Deterministic	Yes	Other		Homogeneous mixing	Agusto 2016*	75.00
Deterministic	Yes	Hospitals, Quarantine, Vaccination		Heterogeneity in transmission rates between groups, Homogeneous mixing	Ahmad 2016*	75.00
Deterministic & Stochastic				Unspecified	Atkins 2016	75.00
Deterministic	Yes	Vaccination		Homogeneous mixing	Brettin 2018	75.00
				Unspecified	Burghardt 2016	75.00
		Other		Heterogeneity in transmission rates between groups and over time, Latent period the same as incubation period	Buyuktahtakin 2018	75.00
	Yes	Hospitals		Heterogeneity in transmission rates between groups	Dautel 2021	75.00
Deterministic		Behaviour changes		Heterogeneity in transmission rates over time	Djiomba Njankou 2022	75.00
Deterministic	Yes			Unspecified	Dokuyucu 2020	75.00
Deterministic	Yes			Homogeneous mixing	El Rhoubari 2019	75.00
Deterministic	Yes			Unspecified	Farman 2022 (b)	75.00
Stochastic				Heterogeneity in transmission rates over time	Getz 2018	75.00
Deterministic				Heterogeneity in transmission rates over time	Getz 2018	75.00
Deterministic				Heterogeneity in transmission rates over time	Getz 2018	75.00
Stochastic				Heterogeneity in transmission rates over time	Getz 2018	75.00
Deterministic	Yes	Behaviour changes, Other, Safe burials		Unspecified	Guo 2016	75.00
Deterministic & Stochastic	Yes			Unspecified	Kamara 2021	75.00
Deterministic	Yes	Quarantine		Heterogeneity in transmission rates between groups	Ngwa 2016	75.00
Deterministic	Yes	Other, Quarantine, Treatment, Vaccination		Heterogeneity in transmission rates between groups	Ojoma 2021	75.00
Stochastic	Yes	Contact tracing		Heterogeneity in transmission rates between groups	Shahtori 2018	75.00
Deterministic	Yes			Latent period the same as incubation period	Tahir 2019	75.00
Stochastic		Behaviour changes		Unspecified	Vinson 2016	75.00
		Safe burial, Unspecified		Homogeneous mixing	Weitz 2015	75.00
Deterministic		Contact tracing, Other, Quarantine		Heterogeneity in transmission rates over time, Homogeneous mixing, Latent period the same as incubation period, Unspecified	Althaus 2015	71.43
Deterministic		Contact tracing, Other, Quarantine, Safe burials, Treatment	Yes	Heterogeneity in transmission rates over time, Homogeneous mixing, Latent period the same as incubation period, Unspecified	Althaus 2015 (b)	71.43
Deterministic & Stochastic		Behaviour changes, Hospitals, Safe burials		Heterogeneity in transmission rates between groups and over time, Homogeneous mixing	Camacho 2014	71.43
Deterministic		Other, Quarantine, Treatment Centres	Yes	Heterogeneity in transmission rates between groups	D'Silva 2017*	71.43
Stochastic		Other, Quarantine, Treatment Centres	Yes	Heterogeneity in transmission rates between groups	D'Silva 2017*	71.43

Stochastic or Deterministic	Theoretical model	Interventions	Code available	Assumptions	Article	QA score (%)
Stochastic				Heterogeneity in transmission rates between groups	Fasina 2014*	71.43
Deterministic				Heterogeneity in transmission rates over time	Getz 2019	71.43
Deterministic		Other, Quarantine	Yes	Heterogeneity in transmission rates over time, Homogeneous mixing, Latent period the same as incubation period	Hart 2019	71.43
Stochastic		Vaccination		Unspecified	Potluri 2022	71.43
		Other, Quarantine, Safe burials, Vaccination		Heterogeneity in transmission rates between groups and over time	Shen 2015	71.43
Stochastic		Contact tracing		Unspecified	Webb 2015*	71.43
Deterministic		Behaviour changes, Hospitals, Other, Safe burials		Heterogeneity in transmission rates between groups, Homogeneous mixing	Agusto 2015*	57.14
Deterministic				Unspecified	Li 2017 (b)	57.14
Deterministic & Stochastic				Homogeneous mixing	Mouanguissa 2021	57.14
Stochastic				Homogeneous mixing	Ndanguza 2017	57.14
Deterministic		Treatment Centres		Homogeneous mixing	Njankou 2018	57.14
Stochastic		Hospitals, Behaviour changes, Safe burials		Heterogeneity in transmission rates between groups, Latent period the same as incubation period	Valdez 2015	57.14
		Hospitals		Unspecified	Webb 2016	57.14
Deterministic		Quarantine, Treatment		Homogeneous mixing, Latent period the same as incubation period, Heterogeneity in transmission rates over time	White 2015	57.14
Deterministic		Contact tracing, Hospitals, Other, Vaccination		Heterogeneity in transmission rates over time, Latent period the same as incubation period	Xie 2019	57.14
Deterministic		Other, Safe burials		Unspecified	Agbomola 2022	50.00
Deterministic	Yes	Vaccination		Homogeneous mixing, Latent period the same as incubation period	Ahmad 2020	50.00
Deterministic	Yes	Quarantine		Homogeneous mixing	Ahmad 2021	50.00
Stochastic	Yes	Contact tracing, Other, Quarantine, Safe burials		Unspecified	Bouba 2023	50.00
Stochastic				Unspecified	Brown 2016	50.00
Deterministic	Yes		Yes	Homogeneous mixing	Cheema 2021	50.00
Deterministic		Hospitals, Quarantine		Homogeneous mixing, Latent period the same as incubation period	Chinyoka 2021	50.00
Deterministic	Yes	Treatment, Vaccination		Homogeneous mixing	Chukwu 2020	50.00
Deterministic		Safe burials, Treatment Centres, Vaccination		Heterogeneity in transmission rates between groups and over time	Diakite 2016*	50.00
Deterministic		Quarantine		Heterogeneity in transmission rates between groups, Homogeneous mixing, Latent period the same as incubation period	Diakite 2016	50.00
Deterministic	Yes	Hospitals, Safe burials		Unspecified	Gaffey 2018	50.00
Deterministic		Hospitals, Treatment, Vaccination		Latent period the same as incubation period	Jiang 2017 (b)	50.00
Deterministic		Hospitals, Other, Safe burials		Homogeneous mixing	Juga 2023	50.00
Deterministic	Yes			Unspecified	Lopez 2016	50.00
Deterministic	Yes	Quarantine		Homogeneous mixing	Nisar 2023	50.00
Deterministic	Yes	Vaccination		Homogeneous mixing, Latent period the same as incubation period	Obeng-Kusi 2021	50.00
Deterministic				Unspecified	Ord 2018	50.00
Deterministic	Yes	Behaviour changes, Other, Vaccination		Homogeneous mixing	Rachah 2016	50.00
Deterministic	Yes	Treatment, Vaccination		Homogeneous mixing	Rachah 2016 (b)	50.00
Deterministic	Yes	Vaccination		Homogeneous mixing	Rachah 2017	50.00
	Yes			Unspecified	Rafiq 2020	50.00
Deterministic	Yes	Behaviour changes, Hospitals, Safe burials		Heterogeneity in transmission rates over time, Homogeneous mixing	Roy 2017	50.00
Deterministic	Yes	Behaviour changes, Hospitals, Other, Vaccination		Homogeneous mixing, Latent period the same as incubation period	Seck 2022	50.00
Deterministic				Unspecified	Singh 2023	50.00
Deterministic				Heterogeneity in transmission rates over time	Smirnova 2019	50.00
Deterministic	Yes			Homogeneous mixing	Srivastava 2020	50.00
Deterministic	Yes			Unspecified	Srivastava 2020 (b)	50.00
Deterministic	Yes			Unspecified	Srivastava 2021	50.00
Deterministic	Yes	Vaccination		Homogeneous mixing	Tahir 2018	50.00
Deterministic	Yes			Homogeneous mixing, Latent period the same as incubation period	Tandon 2018	50.00

Stochastic or Deterministic	Theoretical model	Interventions	Code available	Assumptions	Article	QA score (%)
Deterministic	Yes			Unspecified	Telionis 2020	50.00
Deterministic	Yes		Yes	Homogeneous mixing	Tsanou 2017	50.00
Deterministic & Stochastic		Quarantine, Vaccination		Homogeneous mixing	Tulu 2017 (b)	50.00
Deterministic	Yes			Homogeneous mixing, Latent period the same as incubation period	Tulu 2017 (c)	50.00
Stochastic	Yes			Unspecified	Wang 2019	50.00
Stochastic	Yes	Treatment Centres		Heterogeneity in transmission rates between groups	Yin 2021	50.00
Deterministic	Yes	Behaviour changes, Other, Treatment		Heterogeneity in transmission rates between groups	Zakary 2017	50.00
Deterministic		Other		Heterogeneity in transmission rates over time, Homogeneous mixing, Latent period the same as incubation period, Unspecified	Althaus 2014	42.86
Deterministic		Other, Quarantine, Safe burials, Treatment Centres		Heterogeneity in transmission rates over time, Homogeneous mixing, Latent period the same as incubation period	Barbarossa 2015	42.86
Stochastic		Vaccination		Unspecified	Camacho 2015 (b)	42.86
				Unspecified	Evans 2015	42.86
Deterministic		Quarantine, Safe burials		Homogeneous mixing	Imran 2017	42.86
Deterministic		Other	Yes	Homogeneous mixing	Ivorra 2015	42.86
Deterministic		Treatment		Homogeneous mixing	Li 2015	42.86
Deterministic				Unspecified	Ndanguza 2013*	42.86
Stochastic		Hospitals, Safe burials		Heterogeneity in transmission rates between groups, Homogeneous mixing	Ponce 2019	42.86
Deterministic		Vaccination		Homogeneous mixing	Rachah 2018	42.86
Deterministic & Stochastic		Contact tracing, Other, Treatment		Homogeneous mixing, Latent period the same as incubation period	Rivers 2014	42.86
Deterministic			Yes	Heterogeneity in transmission rates over time	Stockdale 2021	42.86
Stochastic			Yes	Heterogeneity in transmission rates over time	Tang 2023	42.86
Stochastic			Yes	Heterogeneity in transmission rates over time	Vossler 2022	42.86
Deterministic				Unspecified	Wang 2015	42.86
Stochastic			Yes	Unspecified	Ward 2023	42.86
Stochastic		Other		Heterogeneity in transmission rates over time	Chowell 2004	28.57
Deterministic		Hospitals		Unspecified	Juga 2020	28.57
Deterministic		Behaviour changes, Hospitals, Other		Homogeneous mixing	Juga 2021	28.57
Deterministic		Quarantine		Heterogeneity in transmission rates between groups	Khan 2015	28.57
Deterministic		Other		Heterogeneity in transmission rates between groups and over time, Homogeneous mixing	Siewe 2020 (b)	28.57
Deterministic		Behaviour changes, Quarantine		Unspecified	Tadmon 2022	28.57
Deterministic				Unspecified	Area 2015*	25.00
Deterministic	Yes			Unspecified	Atangana 2014	25.00
Deterministic	Yes		Yes	Unspecified	Bachinsky 2013	25.00
Deterministic	Yes			Homogeneous mixing	Bhardwaj 2021	25.00
Stochastic				Unspecified	Champredon 2018	25.00
Deterministic	Yes	Contact tracing, Hospitals, Other, Quarantine		Unspecified	Chen 2014***	25.00
	Yes			Unspecified	Dike 2017	25.00
Deterministic	Yes			Homogeneous mixing	Emile Franc Doungmo 2016*	25.00
	Yes			Unspecified	Gong 2019	25.00
Deterministic	Yes	Other		Homogeneous mixing	Grigorieva 2015	25.00
Deterministic	Yes	Other, Quarantine		Homogeneous mixing	Grigorieva 2017	25.00
Deterministic	Yes	Hospitals, Other, Safe burials		Homogeneous mixing	Grigorieva 2018	25.00
Deterministic	Yes			Unspecified	Hasan 2023	25.00
Stochastic				Unspecified	Kaur 2021	25.00
Deterministic		Behaviour changes		Heterogeneity in transmission rates between groups	Levy 2017	25.00

Stochastic or Deterministic	Theoretical model	Interventions	Code available	Assumptions	Article	QA score (%)
Deterministic				Unspecified	Mubayi 2021	25.00
	Yes	Hospitals, Behaviour changes		Heterogeneity in transmission rates between groups	Shah 2019	25.00
Deterministic	Yes	Contact tracing, Hospitals, Quarantine, Safe burials, Vaccination		Unspecified	Tadmon 2022 (b)	25.00
Deterministic & Stochastic		Quarantine, Vaccination		Homogeneous mixing	Tulu 2017	25.00
Deterministic	Yes			Homogeneous mixing	Zafar 2021	25.00
Deterministic		Quarantine, Safe burials		Unspecified	Do 2016	14.29
Deterministic		Other		Heterogeneity in transmission rates between groups	Dong 2015	14.29
	Yes	Safe burials, Vaccination		Heterogeneity in transmission rates over time	Area 2018	0.00
Stochastic	Yes			Unspecified	Atangana 2021	0.00
Deterministic & Stochastic		Quarantine, Vaccination		Unspecified	Din 2022	0.00
	Yes	Quarantine		Unspecified	Farman 2022	0.00
	Yes			Unspecified	Koca 2018	0.00
		Quarantine, Treatment, Vaccination		Unspecified	Li 2015 (b)	0.00
Deterministic				Homogeneous mixing	Li 2020	0.00
	Yes	Quarantine, Vaccination		Unspecified	Liu 2019	0.00
Deterministic		Behaviour changes		Heterogeneity in transmission rates between groups and over time	Long 2018	0.00
Deterministic	Yes	Contact tracing, Quarantine		Homogeneous mixing	Madubueze 2018	0.00
Deterministic	Yes			Unspecified	Momani 2022	0.00
Deterministic	Yes	Other, Treatment, Vaccination		Unspecified	Okyere 2020	0.00
Deterministic		Hospitals, Quarantine, Safe burials		Unspecified	Ouemba Tasse 2022	0.00
Deterministic				Unspecified	Pan 2021	0.00
Deterministic	Yes	Vaccination		Homogeneous mixing	Rachah 2015	0.00
Stochastic	Yes			Heterogeneity in transmission rates between groups	Rashid 2022	0.00
Deterministic	Yes			Unspecified	Raza 2020	0.00
	Yes	Treatment		Unspecified	Shaikh 2023	0.00
Deterministic	Yes			Homogeneous mixing	Singh 2020	0.00
Deterministic				Heterogeneity in transmission rates between groups	Sivaraman 2022	0.00
Deterministic	Yes			Homogeneous mixing	Zhang 2020	0.00
Other or combination						
Deterministic	Yes	Quarantine		Unspecified	Adams 2016	100.00
Stochastic		Vaccination		Heterogeneity in transmission rates between groups	Chowell 2019 (b)	100.00
			Yes	Unspecified	De Maio 2018	100.00
Deterministic				Unspecified	De la Sen 2017	100.00
Deterministic	Yes			Homogeneous mixing	Feng 2016	100.00
Stochastic				Heterogeneity in transmission rates over time	Ganyani 2018	100.00
				Unspecified	Gustafson 2017	100.00
Stochastic	Yes			Unspecified	Jacobsen 2018	100.00
Stochastic			Yes	Heterogeneity in transmission rates between groups	Kiskowski 2014	100.00
Stochastic				Heterogeneity in transmission rates between groups	Kiskowski 2016	100.00
Stochastic			Yes	Heterogeneity in transmission rates between groups and over time	Kraemer 2019	100.00
Stochastic				Heterogeneity in transmission rates between groups and over time, Latent period the same as incubation period	Lau 2017 (b)	100.00
Stochastic		Behaviour changes		Heterogeneity in transmission rates over time, Homogeneous mixing	Lee 2019	100.00
Stochastic	Yes			Unspecified	Pell 2016	100.00
Deterministic			Yes	Heterogeneity in transmission rates between groups	Santermans 2016	100.00
Stochastic				Heterogeneity in transmission rates over time	Shaman 2014	100.00

Stochastic or Deterministic	Theoretical model	Interventions	Code available	Assumptions	Article	QA score (%)
Deterministic				Unspecified	Smirnova 2017	100.00
				Unspecified	Tuitea 2018	100.00
	Yes			Unspecified	Zabinski 2018	100.00
Stochastic		Contact tracing, Quarantine		Heterogeneity in transmission rates over time	Chan 2020	85.71
			Yes	Homogeneous mixing	Chowell 2019	85.71
Stochastic			Yes	Heterogeneity in transmission rates between groups and over time, Latent period the same as incubation period	Lau 2017	85.71
Stochastic				Homogeneous mixing	White 2007*	85.71
	Yes			Unspecified	Mangiarotti 2016*	75.00
Deterministic	Yes			Homogeneous mixing	Upadhyay 2016*	75.00
Deterministic		Other, Vaccination		Unspecified	Chen 2021	71.43
Stochastic		Behaviour changes	Yes	Unspecified	Pruyt 2015	71.43
				Unspecified	Saulnier 2017	71.43
				Unspecified	Taylor 2016	71.43
				Unspecified	Browne 2015	57.14
Stochastic				Heterogeneity in transmission rates between groups and over time	Martinez 2022	57.14
Stochastic				Heterogeneity in transmission rates over time	Park 2022	57.14
Stochastic			Yes	Homogeneous mixing	Vaughan 2019	57.14
Deterministic	Yes			Heterogeneity in transmission rates between groups	Ahmed 2023	50.00
Stochastic				Unspecified	Backer 2016	50.00
Stochastic				Heterogeneity in transmission rates between groups	Chen 2016	50.00
Stochastic	Yes	Vaccination	Yes	Age dependent susceptibility, Heterogeneity in transmission rates between groups	Rachah 2016 (b)	50.00
Deterministic	Yes			Unspecified	Yamazaki 2018	50.00
Stochastic				Unspecified	Burch 2017	42.86
Deterministic				Unspecified	Jombart 2020	42.86
Stochastic				Latent period the same as incubation period	Russo 2016	42.86
Deterministic				Heterogeneity in transmission rates between groups, Homogeneous mixing	Scarpino 2015	42.86
Deterministic		Vaccination		Homogeneous mixing, Latent period the same as incubation period, Unspecified	Fisman 2014	28.57
Deterministic				Homogeneous mixing, Latent period the same as incubation period, Unspecified	Fisman 2014 (b)	28.57
				Unspecified	Hsieh 2015	28.57
Stochastic	Yes	Behaviour changes		Heterogeneity in transmission rates between groups and over time	Halvorsen 2022	25.00
Deterministic	Yes	Contact tracing, Other, Quarantine, Safe burials		Heterogeneity in transmission rates between groups and over time	Ivorra 2020	25.00
Stochastic		Safe burials		Heterogeneity in transmission rates between groups	Siettos 2015*	14.29
Stochastic	Yes			Unspecified	Burkhead 2015	0.00

Table B.27: Summary of EVD models with code.

Article	QA score (%)	Purpose	Model type	Heath states	Uncertainty considered	Fitted?	Fitting Method	Data	Outbreak period	Spatial model	Spillover included	Language	Link	README?
Kahn 2020	100.00	Transmission	Agent/Individual based	Susceptible, exposed, and recovered	No	Yes	MLE	Incidence of cases	All	Yes	No	Python	https://github.com/rek160/Sierra-Leone-Cholera-Ebola	No
Yamin 2016	100.00	Transmission	Branching process	Incubation, early symptomatic, and late symptomatic	Yes	No			Middle	No	No	Mathematica	https://www.acpjournals.org/doi/suppl/10.7326/M14-2255/suppl_file/AIME201501060-00154_supplement6.pdf	No
Park 2019	25.00	Methodological development	Branching process	Not compartmental	No	No			Unspecified	No	No	R	https://zenodo.org/records/2532906	Yes
Green 2022	100.00	Transmission	Branching process	Not compartmental	Yes	Yes	Unclear	Incidence of cases	All	No	No	R	https://github.com/willgreen236/Heterogeneity_transmission	No
Ward 2023	42.86	Methodological development	Compartmental	Susceptible, exposed, infectious, recovered	Yes	Yes	MCMC	Incidence of cases	All	No	No	R	https://github.com/ceward18/BayesSEIR	Yes
Vossler 2022	42.86	Transmission	Compartmental	Susceptible, infectious, recovered	Yes	Yes	MCMC	Incidence of cases	All	No	No	R	https://doi.org/10.5281/zenodo.6104188	Yes
Tang 2023	42.86	Transmission	Compartmental		Yes	Yes	MCMC	Sequencing data	All	No	No	R	https://projecteuclid-org.iclibezp1.cc.ic.ac.uk/journals/supplementalcontent/10.1214/21-AOAS1583/aoas1583suppb.zip	Yes
Stockdale 2021	42.86	Transmission	Compartmental	Susceptible, exposed, infectious, recovered	No	Yes	MLE	Incidence of deaths	All	No	No	R, C++	https://github.com/jessicastockdale/PBLA	Yes
Santermans 2016	100.00	Transmission	Compartmental	Susceptible, exposed, infectious, recovered	Yes	Yes	Other	Incidence of cases AND deaths		Yes	No	R	https://zenodo.org/records/33126	Yes
Muleia 2021	85.71	Characterise spatial spread or spatial heterogeneity	Compartmental	Susceptible, exposed, infectious, recovered	Yes	Yes	MCMC	Incidence of cases	All	No	No	R	https://github.com/RachidMuleia/Sierra_Leone_Data/blob/master/SEIR_epidemic_model_code	Yes but not informative
Kramer 2016	100.00	Characterise spatial spread or spatial heterogeneity	Compartmental	Susceptible (admin unit), infected (admin unit)	Yes	Yes	MLE	Infected status of admin units	Early	Yes	No	R	https://datadryad.org/stash/dataset/doi:10.5061/dryad.k95j3	Yes
Ivorra 2015	42.86	Transmission	Compartmental	infected, Infectious, hospitatlised, dead, buried, recovered	No	Yes	Unclear	Cumulative incidence of cases AND deaths	All	Yes	No	MATLAB	http://www.mat.ucm.es/~ivorra/softbecodis.htm	Yes
Hart 2019	71.43	Forecast	Compartmental	Susceptible, exposed, infectious, recovered	No	Yes	Other	Cumulative incidence of cases	All	No	No	MATLAB	https://github.com/will-s-hart/EbolaVariableSymptoms	No
Funk 2019	100.00	Forecast	Compartmental	Susceptible, exposed, infectious, recovered	Yes	Yes	MCMC	Incidence of cases	All	No	No	R	https://doi.org/10.5281/zenodo.2547701	Yes
Funk 2018	100.00	Forecast	Compartmental	Susceptible, exposed, infectious (unreported and potentially reported), removed (recovered or dead)	Yes	Yes	MCMC	Simulated data	All	No	No	R	http://github.com/sbfnk/ebola_forecasting_challenge	Yes
D'Silva 2017*	71.43	Transmission	Compartmental	Susceptible, exposed, two-stage infection, funeral, recovered, dead	Yes	Yes	Unclear	Incidence of cases AND deaths		Yes	No	MATLAB	https://github.com/epimath/gravity-model-ebola	Yes but not informative
Althaus 2015 (b)	71.43	Transmission	Compartmental	Susceptible, exposed, infectious, recovered, dead	Yes	Yes	MLE	Incidence of cases	All	No	No	R	https://github.com/calthaus/Ebola and supplementary material files	Yes
Abbate 2016	85.71	Transmission	Compartmental	Susceptible, exposed, infectious, recovered, dead	Yes	Yes	MLE	Incidence of cases	Middle	No	No	R	https://doi.org/10.1371/journal.pntd.0004676.s007	Yes
Abbate 2016	85.71	Sexual transmission impact	Compartmental	Susceptible, exposed, infectious, convelescent (sexual transmission from survivors), recovered, dead	Yes	No		Simulated data	Middle	No	No	R	https://doi.org/10.1371/journal.pntd.0004676.s008	Yes
Vaughan 2019	57.14	Methodological development	Other or combination	Susceptible, infectious, recovered	Yes	Yes	MCMC	Sequencing data	All	No	No	R	https://github.com/tgvaughan/ParticleFilterResults	Yes
Santermans 2016	100.00	Transmission	Other or combination	Susceptible, exposed, infectious who receover, recovered, infectious who die, dead infectious, dead not infectious	Yes	Yes	MCMC	Incidence of cases AND deaths	All	Yes	No	R	https://zenodo.org/records/33126	Yes but not informative
Lau 2017	85.71	Methodological development	Other or combination	Susceptible, exposed, infectious, recovered	Yes	Yes	MCMC	Individual times of clinical events (e.g. symptom onset, death)	Middle	Yes	No	R, C++	https://github.com/msylau/A-mechanistic-spatio-temporal-framework-for-modelling-individual-to-individual-transmission-Ebola	Yes but not informative
De Maio 2018	100.00	Transmission	Other or combination	Not compartmental	Yes	Yes	MCMC	Sequencing data	Early	No	No	BEAST2	https://doi.org/10.1371/journal.pcbi.1006117.s002	No

B.5 Pathogen Epidemiology Review Group (PERG) membership

Table B.28: **PERG membership**

First Name	Surname	Affiliation
Aaron	Morris	University of Oxford
Alpha	Forna	University of Georgia
Amy	Dighe	Johns Hopkins
Anna-Maria	Hartner	Imperial College London
Anna	Vicco	Imperial College London
Anne	Cori	Imperial College London
Arran	Hamlet	Imperial College London
Ben	Lambert	University of Oxford
Bethan	Daniels	Imperial College London
Charlie	Whittaker	Imperial College London
Christian	Morgenstern	Imperial College London
Cosmo	Santoni	Imperial College London
Cyril	Geismar	Imperial College London
Dariya	Nikitin	Imperial College London
David	Jorgensen	Imperial College London
Dominic	Dee	Imperial College London
Ed	Knock	Imperial College London
Gina	Cuomo-Dannenburg	Imperial College London
Hayley	Thompson	PATH
Ilaria	Dorigatti	Imperial College London
Isobel	Routledge	UCSF
Janetta	Skarp	Imperial College London
Joseph	Hicks	Imperial College London
Juliette	Unwin	University of Bristol
Kanchan	Parchani	Imperial College London
Keiran	Drake	Imperial College London
Keith	Fraser	Imperial College London
Kelly	Charniga	Imperial College London
Kelly	McCain	Imperial College London
Lily	Geidelberg	Imperial College London

Lorenzo	Cattarino	UKHSA
Mara	Kont	Imperial College London
Marc	Baguelin	Imperial College London
Natsuko	Imai-Eaton	Imperial College London (Wellcome Trust)
Nima	Moghaddas	Imperial College London
Pablo	Guzman	Imperial College London
Patrick	Doohan	Imperial College London
Paula	Christen	Imperial College London
Rebecca	Nash	Imperial College London
Richard	Shepherd	Imperial College London
Rob	Johnson	Imperial College London
Ruth	McCabe	Imperial College London
Sabine	van Elsland	Imperial College London
Sangeeta	Bhatia	UK Health Security Agency
Sequoia	Leuba	Imperial College London
Shazia	Ruybal-Pesantez	Imperial College London
Sreejith	Radhakrishnan	University of Glasgow
Thomas	Rawson	Imperial College London
Tristan	Naidoo	Imperial College London
Zulma	Cucunuba Perez	Pontificia Universidad Javeriana
Jack	Wardle	Imperial College London

Appendix C

Supplementary Information: Extending EpiEstim to estimate the transmission advantage of pathogen variants in real-time: SARS-CoV-2 as a case-study

The contents of this appendix has been adapted from supplementary information previously published in the journal Epidemics [240].

C.1 Overview

In this document, we present details of the SARS-CoV-2 variant-specific incidence data used in the analysis (Appendix C.2) and describe the method used for obtaining estimates of the transmission advantage for SARS-CoV-2, both using a non-parametric approach and using MV-EpiEstim (Appendix C.3). Appendix C.4 shows additional results for the estimation of the transmission advantage of SARS-CoV-2 Variants Of Concern (VOCs), including more detailed results on the estimated transmission advantage of Alpha over the wildtype, but also estimates of the transmission advantages of Beta and Gamma (combined) over the wildtype, and of Delta over Alpha. Appendix C.5 presents an overview of the simulation study used to assess the validity of our method. We describe the methodology used for the simulation and present the range of scenarios we explored, as well as more comprehensive results from our simulation study.

C.2 SARS-CoV-2 variant-specific incidence data

C.2.1 Incidence data from England

We used the daily number of positive tests from England's community SARS-CoV-2 testing system (also called Pillar 2) from 1st September 2020 to 20th June 2021, stratified by NHS region (Figure C.1). The Pillar 2 testing data were shared with Imperial College London by Public Health England. Up to 14th March 2021, we interpreted the number of samples with no S-gene target failure (SGTF) in this data as incidence of the wildtype, and from 14th March 2021 onwards as incidence of the Delta variant. Samples with S-gene failure were considered to be of the Alpha variant throughout. The weekly proportion of reported cases for which the SGTF status was known varied from 57% to 74% over the study period (Figure C.4). We assumed that the daily proportion of the variants (Alpha and wildtype, or Alpha and Delta) was identical in the cases for which SGTF was known (S-gene positive and S-gene negative) and in those for which SGTF was not known. We adjusted the daily incidence of the variants by distributing the cases with unknown S-gene status using this daily proportion.

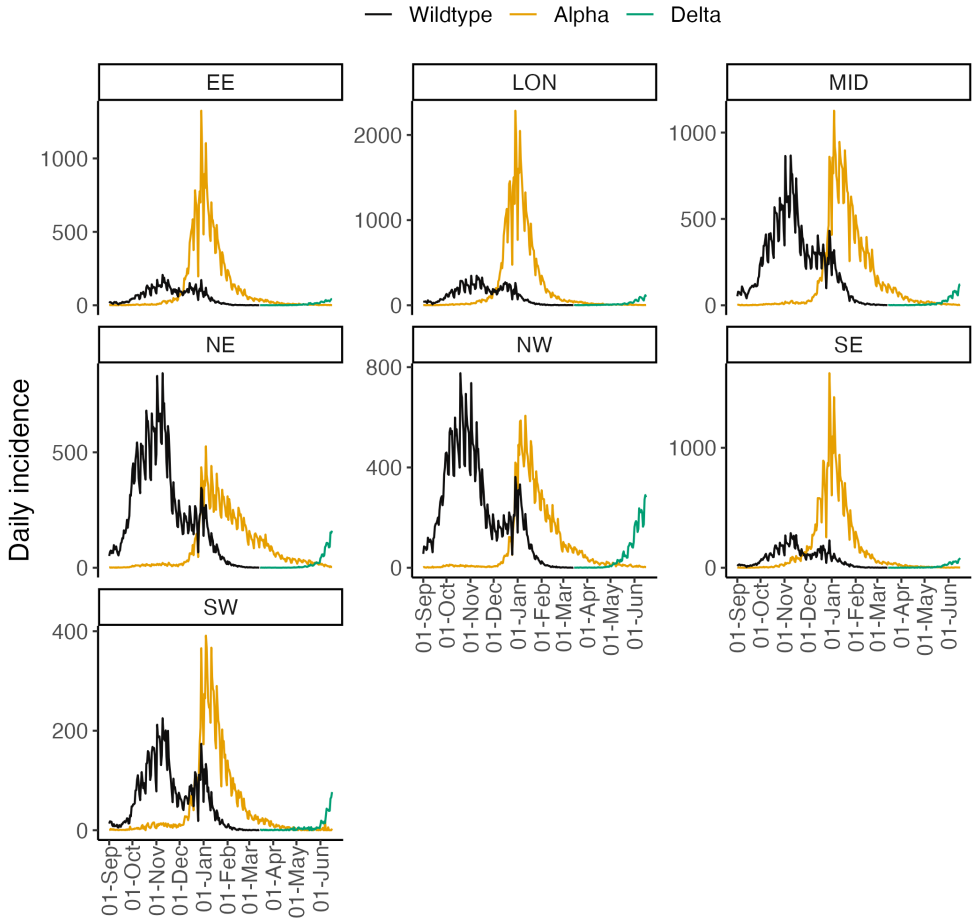


Figure C.1: **Daily incidence of SARS-CoV-2 wildtype (black), Alpha (yellow) and Delta (green) variants in the 7 NHS regions in England.** Note that the y-axis in each panel is different. The NHS England regions are - East of England (EE), London (LON), Midlands (MID), North-East (NE), North-West (NW), South-East (SE), South-West (SW).

C.2.2 Incidence data from France

Santé Publique France reports the age-disaggregated number of PCR tests with Alpha, Beta, and Gamma variants of SARS-CoV-2 at a sub-national level in France [606] with the incidence of Beta and Gamma variants reported as an aggregate. The absence of labelling with a specific VOC was interpreted as an infection with the wildtype. The 18 ADM2 units for which data were reported include metropolitan France and overseas regions. We aggregated the data across all age groups to obtain a daily incidence time series for each variant from 28th February to 30th May 2021 (Figure C.2). The proportion of PCR tests screened for the presence of variants among positive PCRs varied from 66.1% in the week starting 28th February to 26.9% in the week ending 30th May 2021. The variant was know for all cases that were used for estimation.

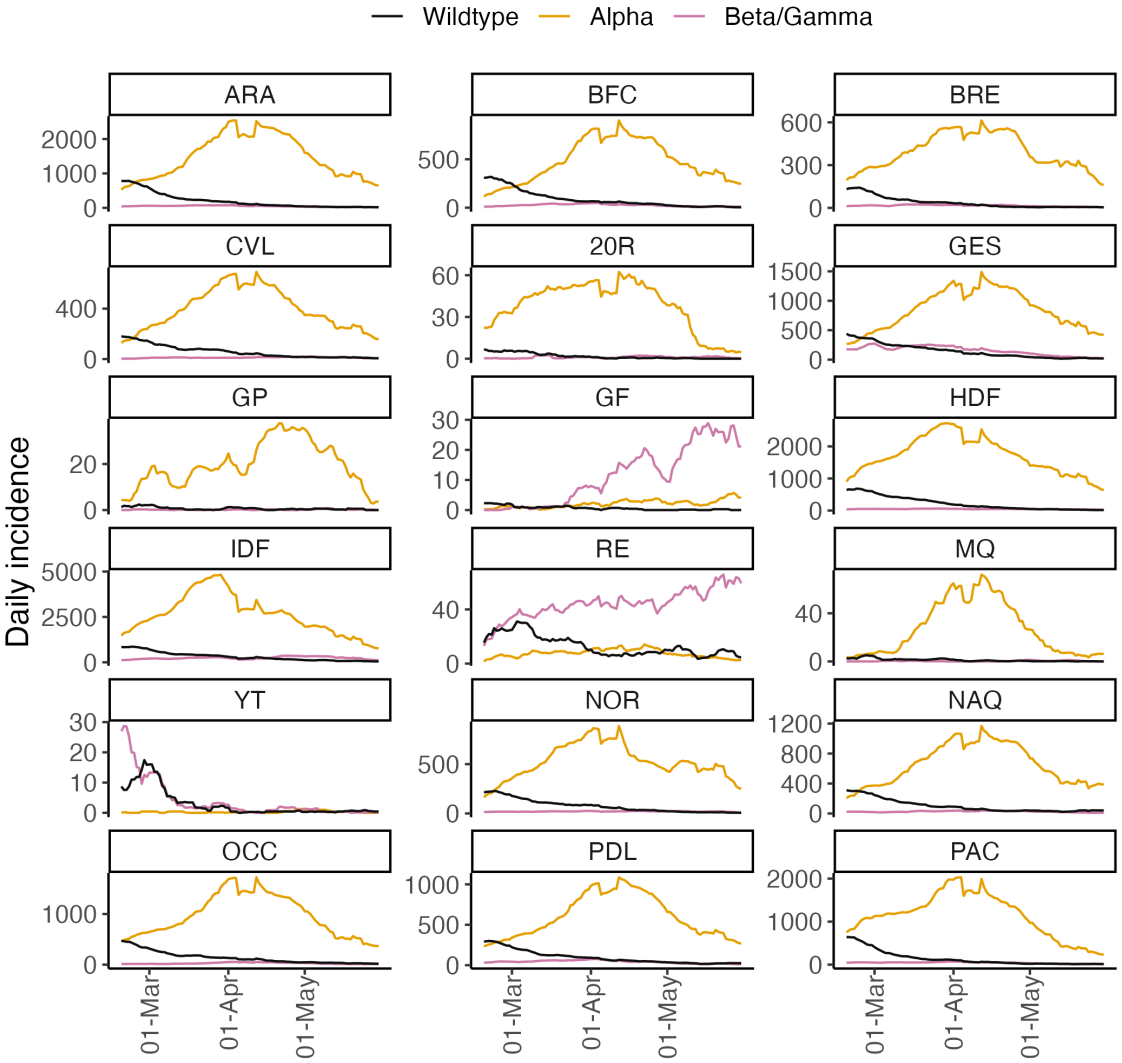


Figure C.2: **Daily incidence of SARS-CoV-2 wildtype (black), Alpha (yellow) and Beta/Gamma (blue) variants for the 18 ADM2 regions in France.** Note that the y-axis in each panel is different. The ADM2 regions are - ARA: Auvergne-Rhône-Alpes, BFC: Bourgogne-Franche-Comté, BRE: Bretagne, CVL: Centre-Val de Loire, 20R: Corse, GES: Grand Est, GP: Guadeloupe, GF: Guyane, HDF: Hauts-de-France, IDF: Île-de-France, RE: La Réunion, MQ: Martinique, YT: Mayotte, NOR: Normandie, NAQ: Nouvelle-Aquitaine, OCC: Occitanie, PDL: Pays de la Loire, and PAC: Provence-Alpes-Côte d'Azur.

C.3 Estimating the effective transmission advantage

C.3.1 Serial interval distribution

Both the non-parametric (see Appendix C.3.2) and the MV-EpiEstim approaches use the discrete distribution of the serial interval (time between symptom onset in a case and their infector) as an input. We assumed a discrete gamma distributed serial interval for SARS-CoV-2 with mean 5.4 days and standard deviation of 1.5 days following [282]. We used the same serial interval distribution across all variants.

C.3.2 Non-parametric approach

To obtain a non-parametric estimate of the effective transmission advantage of a VOC over the reference SARS-CoV-2 variant, we first estimated the daily effective reproduction number independently for each variant (wildtype or VOC) for 18 ADM2 units in France and 7 NHS regions in England using the R package EpiEstim [62]. We used a sliding weekly window, and set the prior R_t to have a mean and a standard deviation of 1.

To exclude region-weeks where the R_t estimates were highly uncertain, we only used estimates from region-weeks where the width of 95% CrI of R_t was less than 0.5. We started estimation of R_t on the week starting on the 11^{th} day. The threshold of 11 days was chosen as it is the 99^{th} percentile of the serial interval distribution. That is, 99% of the cases that were infected by an index cases from day 1 in our analysis are expected to have been observed by day 11. Note that because of these exclusion criteria, some of the non-parametric estimates are missing in the tables shown in Appendix C.4, when no weeks could be included for a particular region or time period.

For each region-week included in the analysis, we drew a sample of 100 values from the posterior distribution of R_t for each variant. Non-parametric estimates of a variant's transmission advantage over a reference variant were obtained by dividing the sampled values from their respective posterior R_t distributions (with random pairing). To account for sub-national variation in R_t profile, estimates at the national level were obtained by pooling the sub-national estimates (thereby giving the same weight to each week of data from any region). To gain insight into the potential temporal heterogeneity of the effective transmission advantage, we divided the incidence time series into four non-overlapping periods of equal duration and estimated the transmission advantage in each period.

C.3.3 Estimation using MV-EpiEstim

We set the priors for both R_t and ϵ to have mean and standard deviation 1. We ran the multi-stage Gibbs sampler for 20,000 iterations. The first 5,000 iterations were discarded as burn-in and thinning was set to keep 1 in 10 iterations, leading to a final posterior sample of size 1,500.

Posterior samples of the transmission advantage were obtained for (i) each region independently

and (ii) nationally but using regional data, by assuming a single underlying transmission advantage and region-specific R_t profiles. Independent estimates of the transmission advantage were obtained for the same non-overlapping time period as for the non-parametric estimates.

To mimic real-time epidemic context and examine how estimates changed as more data became available, we estimated the effective transmission advantage using data available up to successive weeks.

C.4 Estimates of the effective transmission advantages of SARS-CoV-2 variants

C.4.1 Alpha over wildtype

England

Main results for the estimated transmission advantage of Alpha over the wildtype using data from England are shown in the main text Figure 4.1. Unlike in Figure 4.1 (where panel C shows transmission advantages estimated using data from only the week ending on the date specified on the x-axis), Figure C.3C shows additional results where we estimate ϵ using the entire time series from the start of the study period up to the time-point shown on the x-axis (implicitly assuming a transmission advantage that is constant over time) (Figure C.3). This mimics real-time analyses using all the data accumulated up to the present time. Since this approach uses more information, there is less uncertainty in the estimates. However, compared to estimates using only the most recent week of data (4.1C), the estimates shown in Figure C.3C, obtained using progressively more data, smooth out any underlying temporal trend.

To explore the robustness of estimates to under-reporting, we estimated the transmission advantage of Alpha using only 50% of reported cases of Alpha and wildtype. As expected, the estimates with under-reporting are very similar to that obtained using the full data set, albeit with slightly more uncertainty (Figure C.4).

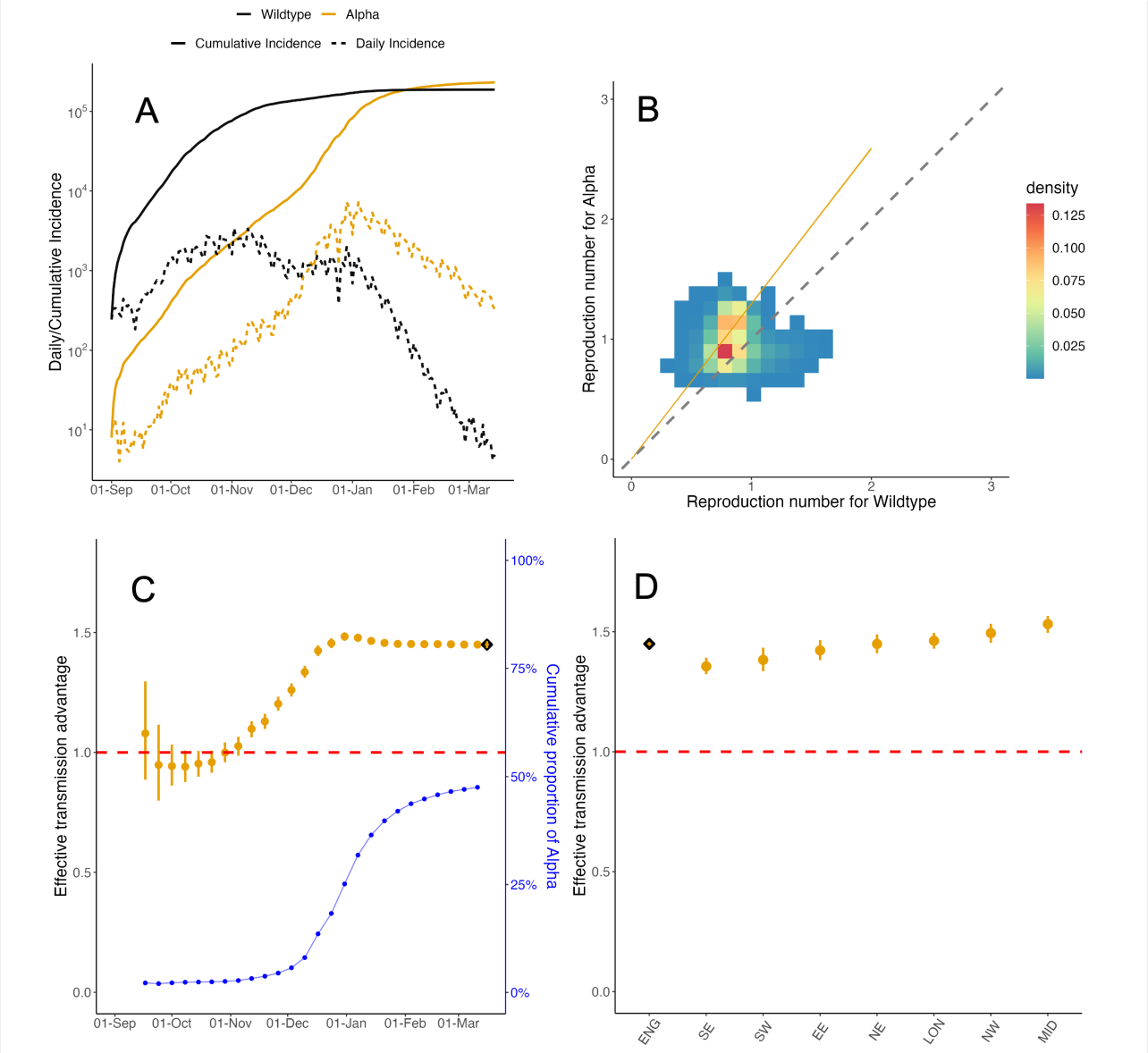


Figure C.3: **Effective transmission advantage of Alpha over wildtype in England. (Note this figure duplicates Figure 4.1 from the main text with the exception of panel C. Further explanation is provided in Appendix C.4.1).** (A) The daily reported incidence of cases of the wildtype (black) and Alpha (yellow) in England from September 2020 to March 2021. (B) The effective reproduction number R_t estimated independently for the wildtype (x-axis) and Alpha (y-axis) on sliding weekly windows. The colour of the cells indicates the density of the draws from the respective posterior distributions of R_t. The dashed diagonal line indicates the $x = y$ threshold. Coloured cells lying above the diagonal line suggest that Alpha is more transmissible. The yellow line denotes the median effective transmission advantage estimated using MV-EpiEstim, assuming no temporal or spatial heterogeneity. 95% CrI were so narrow that they could not be distinguished from the line. (C) Effective transmission advantage estimated using MV-EpiEstim using data from the start of the time series up to the date specified on the x-axis (yellow). The dark blue circles and the horizontal bars denote respectively the mean and 95% binomial confidence interval of the cumulative proportion of incidence of Alpha (right y-axis). Because of the high incidence of both wildtype and Alpha in the study period, the 95% CI are small and hence difficult to distinguish. (D) Effective transmission advantage estimated using MV-EpiEstim for all NHS England regions together (diamond) and separately (solid circles), using data from 1st September 2020 to 14th March 2021. The NHS England regions are - East of England (EE), London (LON), Midlands (MID), North-East (NE), North-West (NW), South-East (SE), South-West (SW). In panels (C) and (D), the solid yellow circles denote the median estimate, the vertical lines indicate the 95% CrI, and the red dashed line denotes the $\epsilon = 1$ threshold.

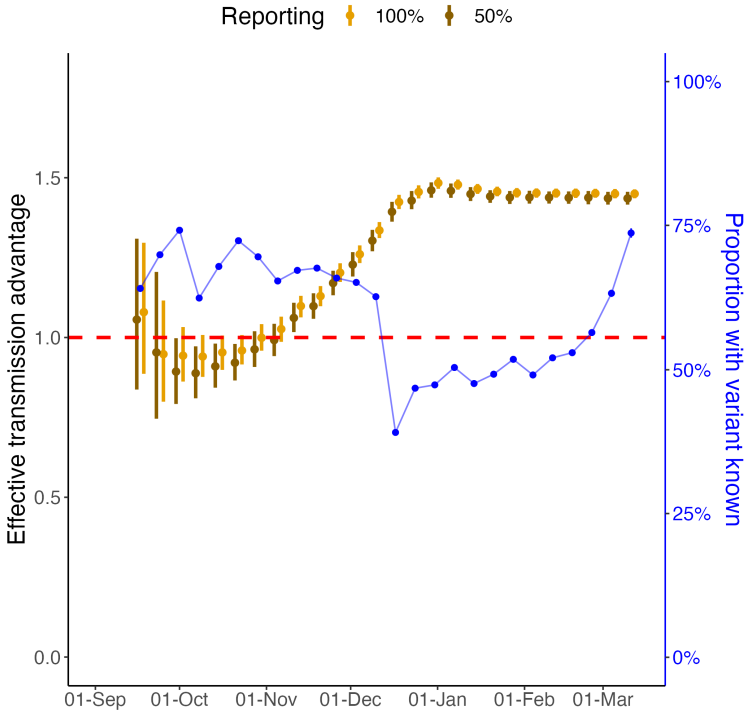


Figure C.4: **Effective transmission advantage of Alpha over wildtype in England from 1st September 2020 to 14th March 2021.** Effective transmission advantage estimated using MV-EpiEstim using all data or 50% of the reported cases in the week ending on the date specified on the x-axis. Yellow circles and vertical lines represent the median and 95% CrI of the estimated transmission advantage using all data. The yellow solid circles and lines are estimates based on use of all reported cases, while the dark orange circles and vertical lines are estimates using only 50% of reported cases of Alpha and wildtype. The red dashed line denotes the $\epsilon = 1$ threshold. The blue circles and the vertical bars denote the mean proportion and the 95% binomial CI of incident cases where the variant was known (right y-axis) in the week of estimation. The estimates and CI were obtained using only cases with known SGTF status. Due to the high incidence of the two variants, the 95% CIs are very narrow and difficult to distinguish.

Table C.1: **Estimates of the effective transmission advantage of SARS-CoV-2 Alpha variant over the wildtype using the non-parametric approach and MV-EpiEstim for 7 NHS regions in England and 4 non-overlapping time periods.** Estimates shown are the posterior median with 95% CrI in parenthesis. Quarters correspond to - Quarter 1: 11$^{\text{th}}$ September - 27$^{\text{th}}$ October 2020; Quarter 2: 27$^{\text{th}}$ October - 12$^{\text{th}}$ December 2020; Quarter 3: 12$^{\text{th}}$ December 2020 - 27$^{\text{th}}$ January 2021; Quarter 4: 27$^{\text{th}}$ January - 14$^{\text{th}}$ March 2021.

Region/Time Period	Non-parametric	MV-EpiEstim
All	1.41 (0.86, 2.01)	1.46 (1.44, 1.47)
East of England	1.38 (1.02, 2.13)	1.43 (1.39, 1.47)
London	1.40 (0.93, 1.82)	1.46 (1.43, 1.50)
Midlands	1.44 (0.92, 2.04)	1.54 (1.50, 1.58)
North East and Yorkshire	1.41 (0.86, 2.05)	1.46 (1.42, 1.50)
North West	1.50 (0.71, 2.08)	1.51 (1.47, 1.55)
South East	1.35 (1.00, 1.87)	1.36 (1.33, 1.39)
South West	1.41 (0.78, 1.82)	1.40 (1.35, 1.45)
Quarter 1	0.89 (0.62, 1.17)	1.03 (0.99, 1.08)
Quarter 2	1.36 (0.82, 1.95)	1.48 (1.45, 1.51)
Quarter 3	1.45 (1.07, 2.02)	1.50 (1.48, 1.52)
Quarter 4	1.39 (0.93, 2.12)	1.28 (1.23, 1.33)

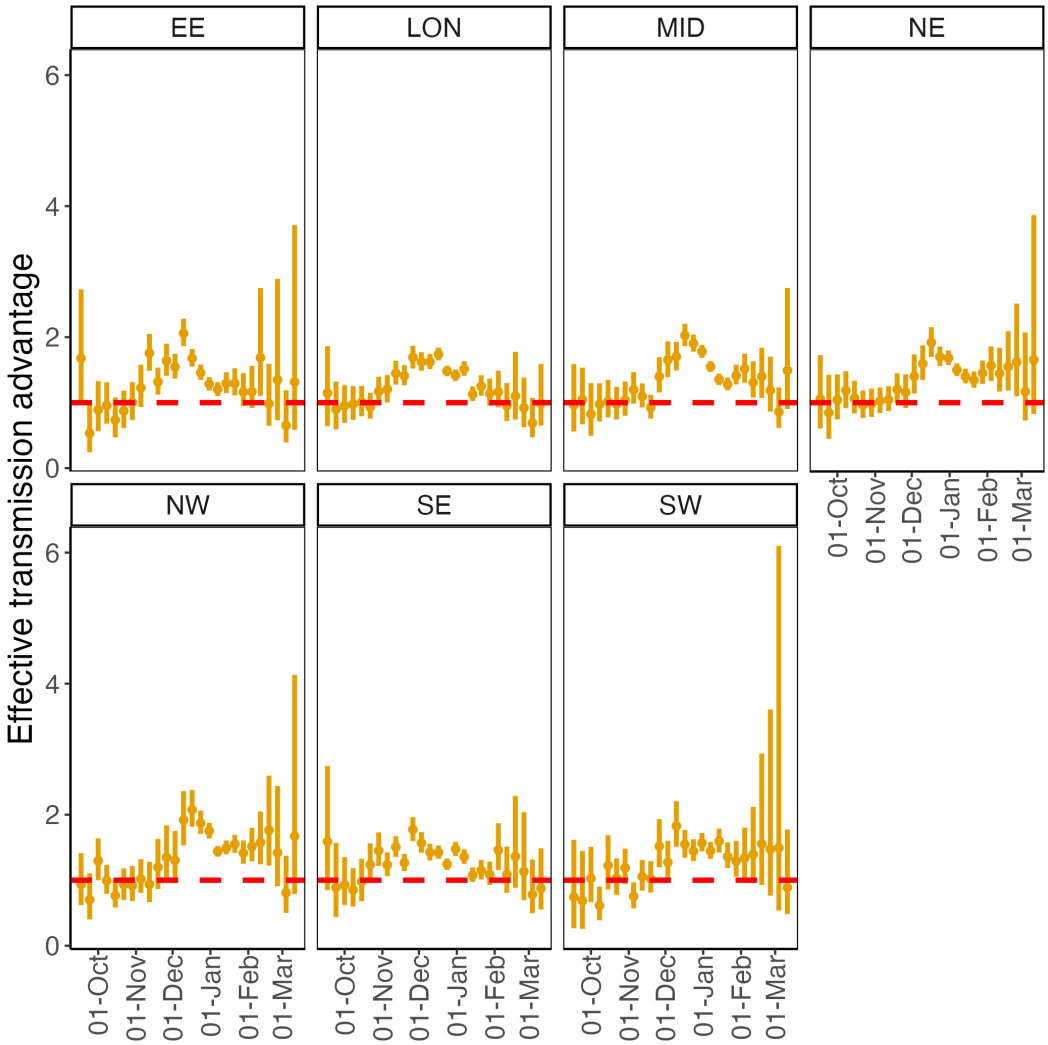


Figure C.5: **Effective transmission advantage of Alpha over wildtype for all NHS England regions from 1st September 2020 to 14th March 2021.** The NHS England regions are - East of England (EE), London (LON), Midlands (MID), North-East (NE), North-West (NW), South-East (SE), and South-West (SW). Each point represents the median estimate from MV-EpiEstim using data in the week ending on the date specified on the x-axis. The vertical yellow lines indicate the 95% CrI, and the red dashed line denotes the $\epsilon = 1$ threshold.

France

Table C.2: **Estimates of the effective transmission advantage of SARS-CoV-2 Alpha variant over the wildtype using the non-parametric approach and MV-EpiEstim for 18 ADM2 regions in France and 4 non-overlapping time periods.** Estimates shown are the posterior median with 95% CrI in parenthesis. Quarters correspond to - Quarter 1: 28th February - 23rd March 2021; Quarter 2: 23rd March - 14th April 2021; Quarter 3: 14th April - 7th May 2021; Quarter 4: 7th May - 20th May 2021.

Region/Time Period	Non-parametric	MV-EpiEstim
All	1.21 (0.75, 1.65)	1.29 (1.29, 1.30)
Auvergne-Rhône-Alpes	1.25 (0.89, 1.67)	1.40 (1.38, 1.43)
Bourgogne-Franche-Comté	1.27 (0.90, 1.78)	1.41 (1.37, 1.46)
Bretagne	1.34 (0.89, 1.78)	1.35 (1.29, 1.40)
Centre-Val de Loire	1.18 (0.79, 1.61)	1.32 (1.27, 1.36)
Corse	-	1.15 (1.00, 1.31)
Grand Est	1.25 (0.70, 1.48)	1.31 (1.28, 1.34)
Guadeloupe	-	1.12 (0.95, 1.35)
Guyane	-	1.33 (1.05, 1.66)
Hauts-de-France	1.19 (0.97, 1.41)	1.28 (1.26, 1.30)
Île-de-France	1.08 (0.87, 1.47)	1.21 (1.20, 1.23)
La Réunion	1.14 (0.67, 2.08)	1.07 (0.98, 1.18)
Martinique	-	1.27 (1.09, 1.49)
Mayotte	-	1.13 (0.73, 1.69)
Normandie	1.21 (0.91, 1.59)	1.31 (1.28, 1.35)
Nouvelle-Aquitaine	1.22 (0.70, 1.62)	1.28 (1.25, 1.31)
Occitanie	1.18 (0.81, 1.50)	1.27 (1.25, 1.30)
Pays de la Loire	1.20 (0.66, 1.55)	1.29 (1.26, 1.32)
Provence-Alpes-Côte d'Azur	1.22 (0.67, 1.58)	1.37 (1.34, 1.40)
Quarter 1	1.45 (1.26, 1.71)	1.42 (1.41, 1.44)
Quarter 2	1.28 (0.99, 1.62)	1.24 (1.22, 1.25)
Quarter 3	1.15 (0.83, 1.51)	1.11 (1.09, 1.13)
Quarter 4	1.00 (0.67, 1.52)	0.97 (0.95, 0.99)

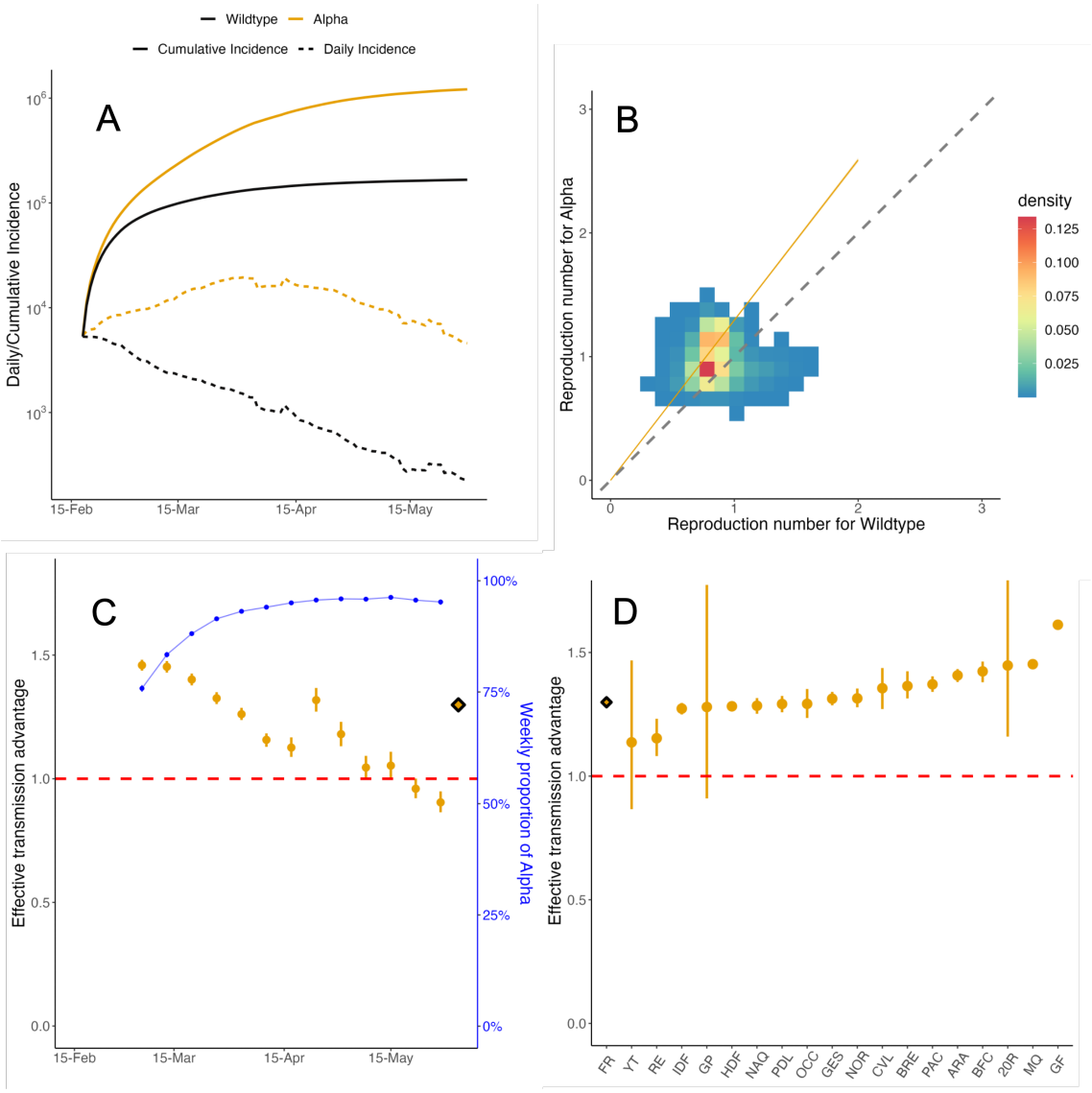


Figure C.6: **Effective transmission advantage of Alpha over wildtype in France** (A) The daily reported incidence of cases of wildtype (black) and Alpha (yellow) in France from 18$^{\text{th}}$ February to 30$^{\text{th}}$ May 2021. (B) The effective reproduction number R_t estimated independently for the wildtype (x-axis) and Alpha (y-axis) on sliding weekly windows. The colour of the cells indicates the density of the draws from the respective posterior distributions of R_t. The dashed diagonal line indicates the $x = y$ threshold. Coloured cells lying above the diagonal line suggest that Alpha is more transmissible. The yellow line denotes the median effective transmission advantage estimated using MV-EpiEstim, assuming no temporal or spatial heterogeneity. 95% CrI were so narrow that they could not be distinguished from the line. (C) Effective transmission advantage estimated from MV-EpiEstim using data in the week ending on the date specified on the x-axis (yellow circles) and the entire time series (diamond). Dark blue circles and the vertical bars denote respectively the mean and 95% binomial confidence interval of the proportion of incidence of Alpha (right y-axis) in the week of estimation. Due to the high incidence of both wildtype and Alpha in the study period, the 95% CI are small and hence difficult to distinguish. (D) Effective transmission advantage estimated using MV-EpiEstim for all ADM2 regions in France together (diamond) and separately (solid circle) using data from 18$^{\text{th}}$ February to 30$^{\text{th}}$ May 2021. ADM2 regions - ARA: Auvergne-Rhône-Alpes, BFC: Bourgogne-Franche-Comté, BRE: Bretagne, CVL: Centre-Val de Loire, 20R: Corse, GES: Grand Est, GP: Guadeloupe, GF: Guyane, HDF: Hauts-de-France, IDF: Île-de-France, RE: La Réunion, MQ: Martinique, YT: Mayotte, NOR: Normandie, NAQ: Nouvelle-Aquitaine, OCC: Occitanie, PDL: Pays de la Loire, and PAC: Provence-Alpes-Côte d'Azur. In panels C-D, solid yellow circles denote the median estimate, vertical lines indicate the 95% CrI, and the red dashed line denotes the $\epsilon = 1$ threshold.

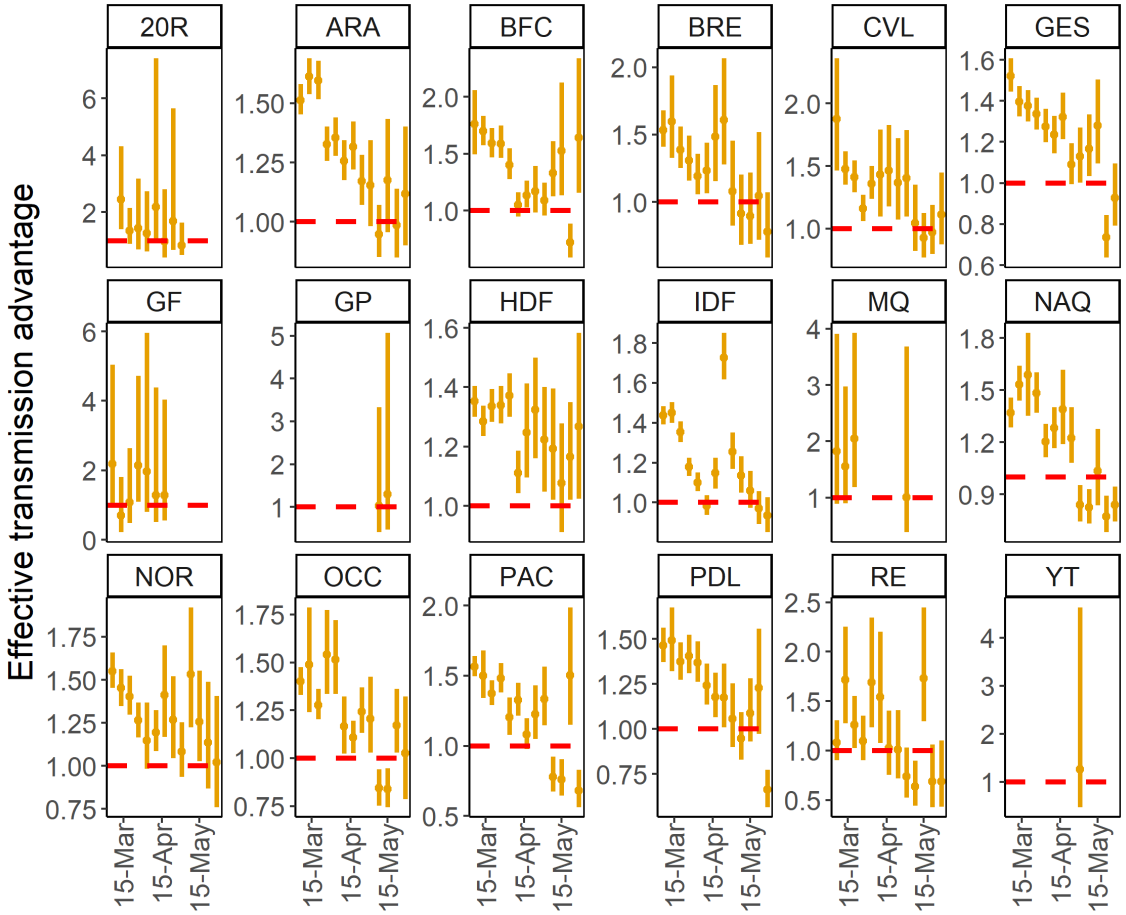


Figure C.7: **Effective transmission advantage of Alpha over wildtype for all ADM2 regions in France from 18th February to 30th May 2021.** The ADM2 regions are - 20R: Corse, ARA: Auvergne-Rhône-Alpes, BFC: Bourgogne-Franche-Comté, BRE: Bretagne, CVL: Centre-Val de Loire, GES: Grand Est, GF: Guyane, GP: Guadeloupe, HDF: Hauts-de-France, IDF: Île-de-France, MQ: Martinique, NAQ: Nouvelle-Aquitaine, NOR: Normandie, OCC: Occitanie, PAC: Provence-Alpes-Côte d'Azur, PDL: Pays de la Loire, RE: La Réunion, and YT: Mayotte. Each point represents the median estimate made using MV-EpiEstim using data in the week ending on the date specified on the x-axis. The vertical yellow lines indicate the 95% CrI, and the red dashed line denotes the ϵ =1 threshold.

C.4.2 Beta and Gamma over wildtype (France)

Table C.3: **Estimates of the combined effective transmission advantage of SARS-CoV-2 Beta and Gamma variants over the wildtype in France using the non-parametric approach and MV-EpiEstim for 18 ADM2 regions in France and 4 non-overlapping time periods.** Estimates shown are the posterior median with 95% CrI in parenthesis. Quarters correspond to - Quarter 1: 28th February - 23rd March 2021; Quarter 2: 23rd March - 14th April 2021; Quarter 3: 14th April - 7th May 2021; Quarter 4: 7th May - 20th May 2021.

Region/Time-period	Non-parametric	MV-EpiEstim
All	1.17 (0.69, 1.74)	1.25 (1.24, 1.27)
Auvergne-Rhône-Alpes	1.15 (0.88, 1.50)	1.29 (1.25, 1.34)
Bourgogne-Franche-Comté	1.22 (0.77, 1.90)	1.32 (1.25, 1.38)
Bretagne	1.26 (0.82, 1.89)	1.28 (1.20, 1.37)
Centre-Val de Loire	1.15 (0.68, 1.73)	1.35 (1.27, 1.44)
Corse	-	1.30 (1.03, 1.62)
Grand Est	1.08 (0.61, 1.31)	1.15 (1.12, 1.17)
Guadeloupe	-	1.07 (0.61, 1.67)
Guyane	-	1.40 (1.14, 1.71)
Hauts-de-France	1.21 (0.96, 1.46)	1.26 (1.22, 1.31)
Île-de-France	1.16 (0.85, 1.72)	1.27 (1.25, 1.29)
La Réunion	1.32 (0.89, 2.10)	1.15 (1.07, 1.22)
Martinique	-	1.30 (0.90, 1.84)
Mayotte	0.95 (0.47, 1.54)	0.83 (0.69, 1.00)
Normandie	1.20 (0.88, 1.70)	1.28 (1.22, 1.35)
Nouvelle-Aquitaine	1.16 (0.53, 1.67)	1.23 (1.17, 1.29)
Occitanie	1.16 (0.71, 1.77)	1.29 (1.24, 1.35)
Pays de la Loire	1.12 (0.64, 1.65)	1.19 (1.14, 1.24)
Provence-Alpes-Côte d'Azur	1.21 (0.65, 1.63)	1.33 (1.28, 1.38)
Quarter 1	1.31 (0.85, 1.78)	1.28 (1.26, 1.30)
Quarter 2	1.17 (0.87, 1.69)	1.15 (1.13, 1.17)
Quarter 3	1.17 (0.81, 1.73)	1.20 (1.17, 1.23)
Quarter 4	1.01 (0.58, 1.75)	0.97 (0.94, 0.99)

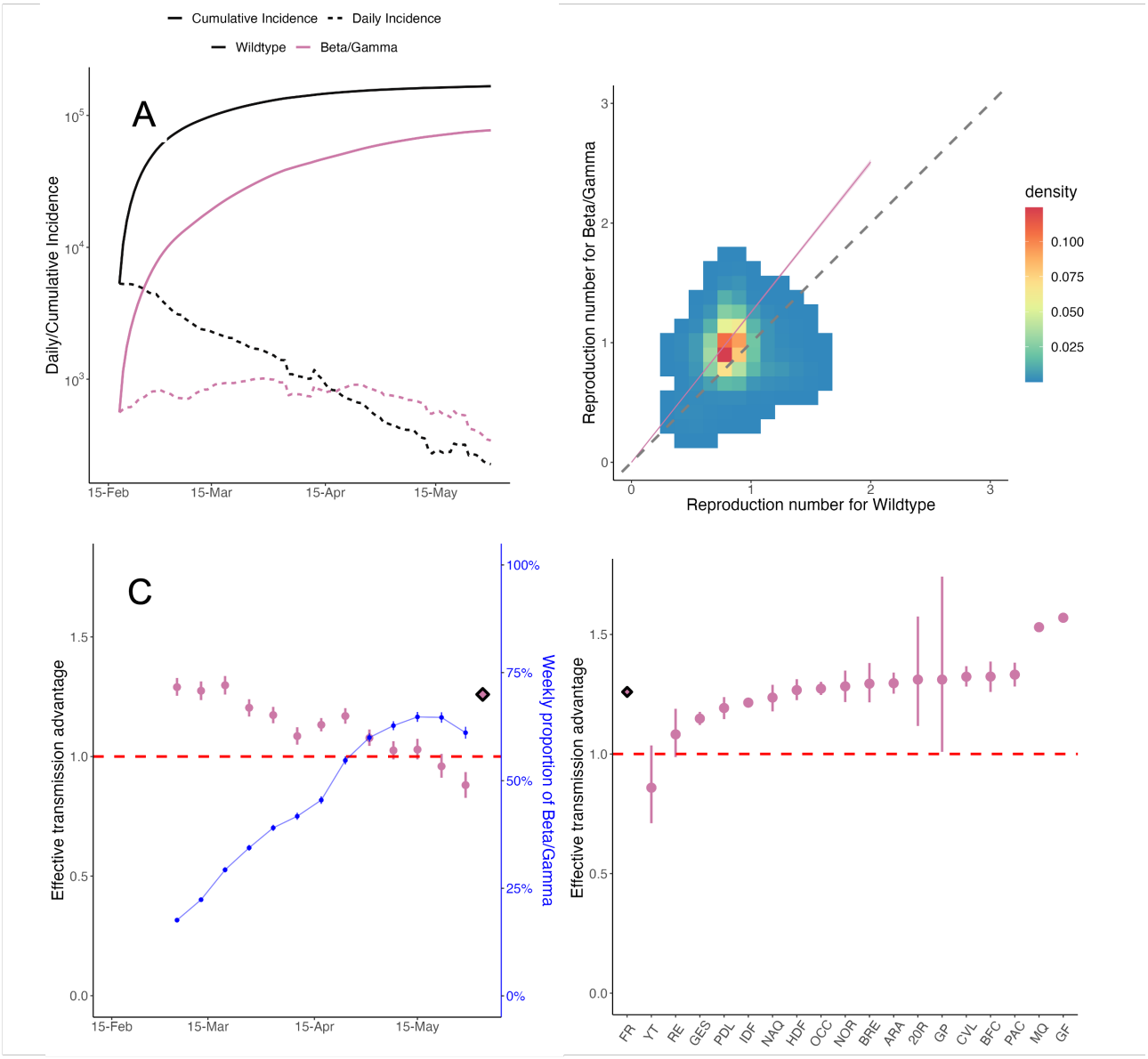


Figure C.8: **Effective transmission advantage of Beta and Gamma (combined) over wildtype in France** (A) The daily reported incidence of cases of the wildtype (black) and Beta/Gamma (blue) in France 18$^{\text{th}}$ February to 30$^{\text{th}}$ May 2021. (B) The effective reproduction number R_t estimated independently for the wildtype (x-axis) and Beta/Gamma (y-axis) on sliding weekly windows. The colour of the cells indicates the density of the draws from the respective posterior distributions of R_t. The dashed diagonal line indicates the $x = y$ threshold. Coloured cells lying above the diagonal line suggest that Beta/Gamma is more transmissible. The pink line denotes the median effective transmission advantage estimated using MV-EpiEstim, assuming no temporal or spatial heterogeneity. 95% CrI were so narrow that they could not be distinguished from the line. (C) Effective transmission advantage estimated from MV-EpiEstim using data in the week ending on the date specified on the x-axis (pink circles) and the entire time series (diamond). The dark blue circles and the vertical bars denote respectively the mean and 95% binomial confidence interval of the proportion of incidence of Beta/Gamma (right y-axis) in the week of estimation. (D) Effective transmission advantage estimated using MV-EpiEstim for all ADM2 regions in France together (diamond) and separately (solid circles) using data from 18$^{\text{th}}$ February to 30$^{\text{th}}$ May 2021. The ADM2 regions are - ARA: Auvergne-Rhône-Alpes, BFC: Bourgogne-Franche-Comté, BRE: Bretagne, CVL: Centre-Val de Loire, 20R: Corse, GES: Grand Est, GP: Guadeloupe, GF: Guyane, HDF: Hauts-de-France, IDF: Île-de-France, RE: La Réunion, MQ: Martinique, YT: Mayotte, NOR: Normandie, NAQ: Nouvelle-Aquitaine, OCC: Occitanie, PDL: Pays de la Loire, and PAC: Provence-Alpes-Côte d'Azur. In panels (C) and (D), the solid pink circles denote the median estimate, the vertical lines indicate the 95% CrI, and the red dashed line denotes the $\epsilon = 1$ threshold.

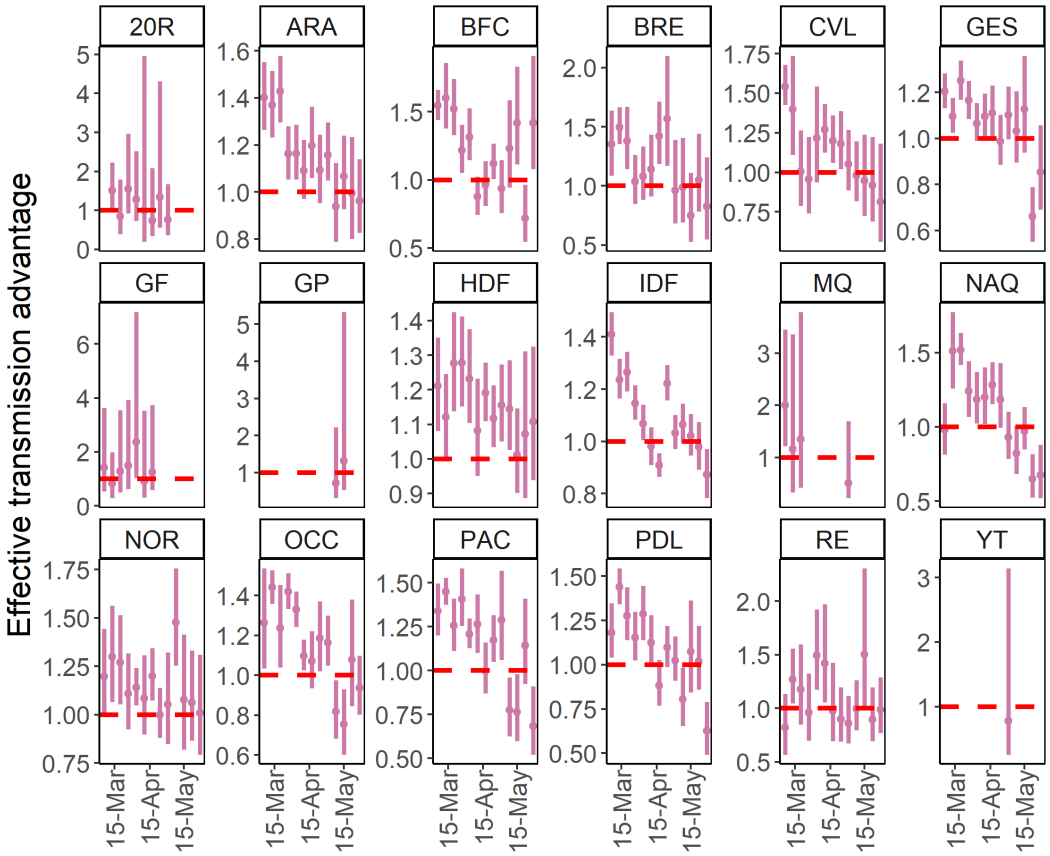


Figure C.9: **Effective transmission advantage of Beta and Gamma (combined) over wildtype for all ADM2 regions in France from 18th February to 30th May 2021.** The ADM2 regions are - 20R: Corse, ARA: Auvergne-Rhône-Alpes, BFC: Bourgogne-Franche-Comté, BRE: Bretagne, CVL: Centre-Val de Loire, GES: Grand Est, GF: Guyane, GP: Guadeloupe, HDF: Hauts-de-France, IDF: Île-de-France, MQ: Martinique, NAQ: Nouvelle-Aquitaine, NOR: Normandie, OCC: Occitanie, PAC: Provence-Alpes-Côte d'Azur, PDL: Pays de la Loire, RE: La Réunion, and YT: Mayotte. Each point represents the median estimate made using MV-EpiEstim using data in the week ending on the date specified on the x-axis. The vertical pink lines indicate the 95% CrI, and the red dashed line denotes the $\epsilon = 1$ threshold.

C.4.3 Delta over Alpha (England)

Table C.4: **Estimates of the effective transmission advantage of SARS-CoV-2 Delta variant over Alpha using the non-parametric approach and MV-EpiEstim for 7 NHS regions in England and 4 non-overlapping time periods.** Estimates shown are the posterior median with 95% CrI in parenthesis. Quarters correspond to - Quarter 1: 25th March - 15rd April 2021; Quarter 2: 15th April - 5th May 2021; Quarter 3: 5th May - 26th May 2021; Quarter 4: 26th May - 16th June 2021.

Region/Time-period	Non-parametric	MV-EpiEstim
All	2.05 (1.09, 6.71)	1.77 (1.69, 1.85)
East of England	-	1.59 (1.37, 1.85)
London	1.23 (0.82, 2.78)	1.51 (1.36, 1.67)
Midlands	1.91 (1.42, 2.77)	1.75 (1.56, 1.95)
North East and Yorkshire	2.55 (1.71, 3.41)	2.12 (1.92, 2.35)
North West	1.84 (1.34, 2.86)	1.87 (1.69, 2.07)
South East	-	1.63 (1.41, 1.86)
South West	7.67 (4.10, 19.21)	1.76 (1.46, 2.13)
Quarter 1	-	1.42 (1.21, 1.66)
Quarter 2	-	1.70 (1.52, 1.90)
Quarter 3	1.75 (0.91, 2.46)	1.71 (1.58, 1.84)
Quarter 4	2.26 (1.42, 7.96)	1.95 (1.81, 2.09)

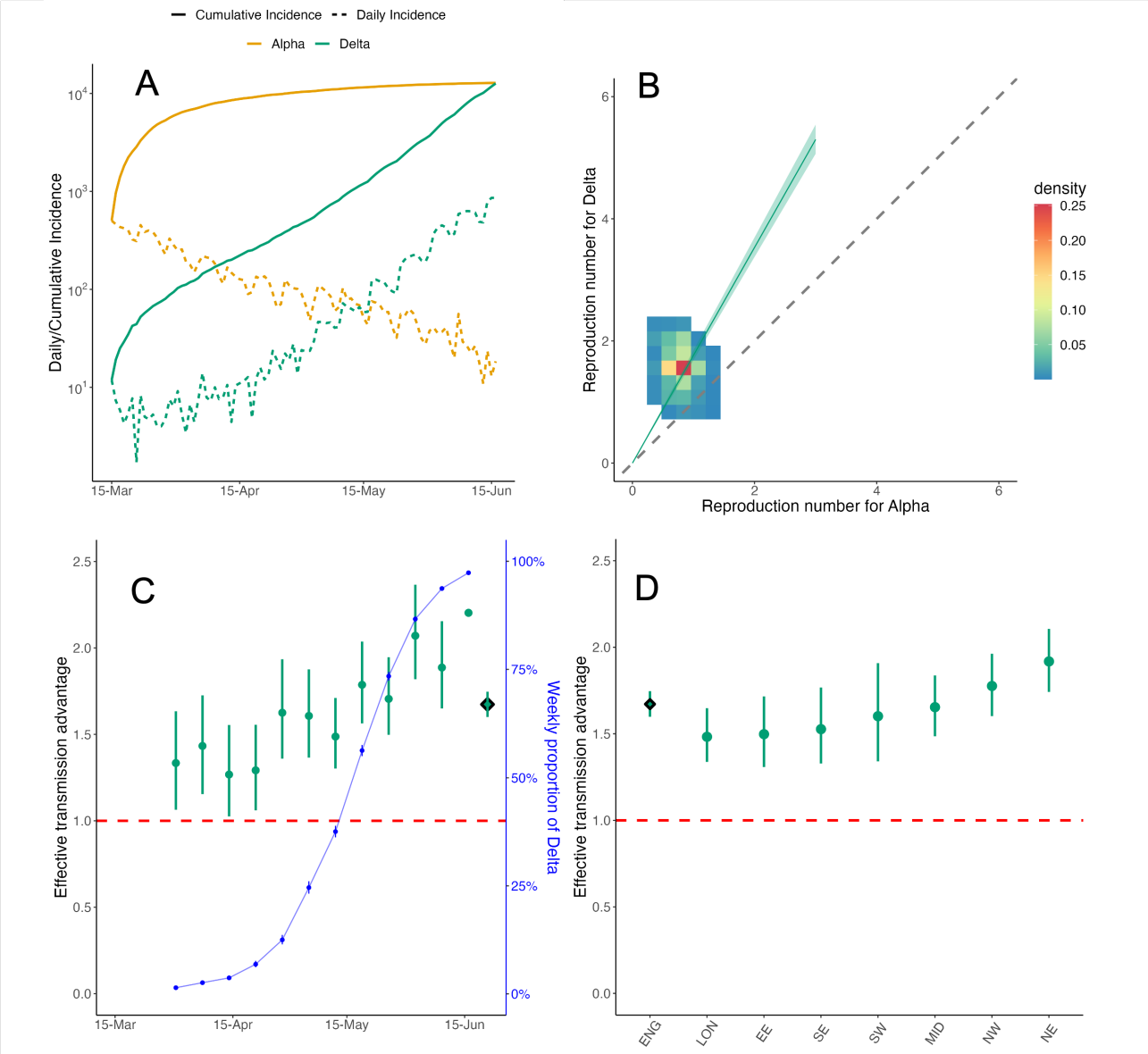


Figure C.10: **Effective transmission advantage of Delta over Alpha in England** (A) The daily reported incidence of cases of Delta (green) and Alpha (yellow) in England from 15$^{\text{th}}$ March to 16$^{\text{th}}$ June 2021. (B) The effective reproduction number R_t estimated independently for Alpha (x-axis) and Delta (y-axis) on weekly sliding windows. The colour of the cells indicates the density of the draws from the respective posterior distributions of R_t. The dashed diagonal line indicates the $x = y$ threshold. Coloured cells lying above the diagonal line suggest that Delta is more transmissible. The green line and the ribbon denote the median and 95% CrI of the effective transmission advantage estimated using MV-EpiEstim, assuming no temporal or spatial heterogeneity. (C) Effective transmission advantage estimated using MV-EpiEstim using data in the week ending on the date specified on the x-axis (green circles) and the entire time series (diamond). The dark blue circles and the vertical bars denote respectively the mean and 95% binomial confidence interval of the proportion of incidence of Delta (right y-axis) in the week of estimation. (D) Effective transmission advantage estimated using MV-EpiEstim for all NHS England regions together (diamond) and separately (solid circles), using data from 15$^{\text{th}}$ March to 16$^{\text{th}}$ June 2021. The NHS England regions are - East of England (EE), London (LON), Midlands (MID), North-East (NE), North-West (NW), South-East (SE), South-West (SW). In panels (C) and (D), the solid green circles denote the median estimate, the vertical lines indicate the 95% CrI, and the red dashed line denotes the $\epsilon = 1$ threshold.

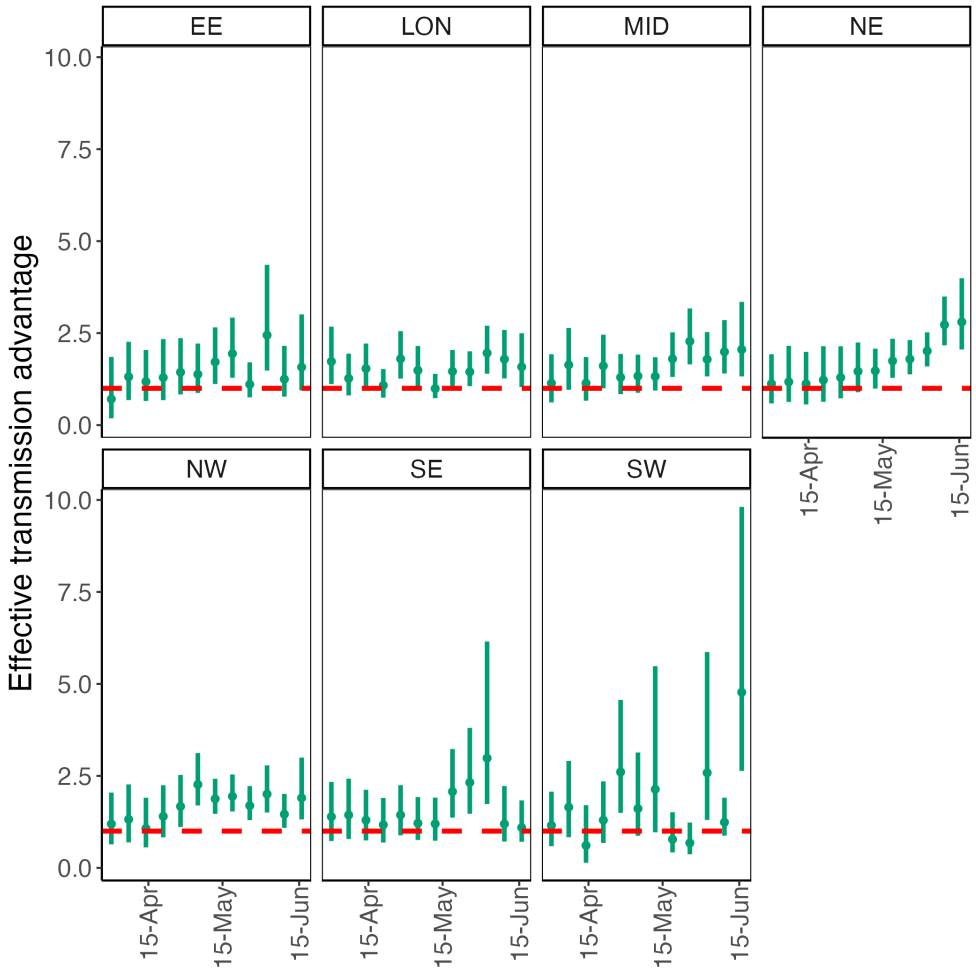


Figure C.11: **Effective transmission advantage of Delta over Alpha for all NHS England regions from 15th March to 16th June 2021.** The NHS England regions are - East of England (EE), London (LON), Midlands (MID), North-East (NE), North-West (NW), South-East (SE), and South-West (SW). Each point represents the median estimate obtained using MV-EpiEstim using data in the week ending on the date specified on the x-axis. The vertical green lines indicate the 95% CrI, and the red dashed line denotes the $\epsilon = 1$ threshold.

C.5 Method performance using simulated data

C.5.1 Simulation approach

We simulated SARS-CoV-2-like incidence data using a branching process where daily incidence is assumed to follow a Poisson distribution:

$$I_t \sim Poisson(R_t \sum_{s=1}^{t-1} I_s \omega_{t-s}), \tag{C.1}$$

where ω_s is the probability mass function of the discrete serial interval.

We assume that the effective reproduction number for the variant is $\epsilon \times R_t$, where R_t is the effective reproduction number for the reference variant and ϵ is the effective transmission advantage of the new variant over the reference. We explored values of $\epsilon > 1$ in all simulation scenarios as $\epsilon < 1$ corresponds to swapping the reference and new variant.

We seeded the epidemic with 20 cases of the reference variant for 10 successive days and 1 case of the new variant on the 10^{th} day. We then simulated forward for an additional 100 days, generating 100 stochastic epidemic trajectories for each simulation scenario and each combination of parameters considered for that scenario (Table C.5). We then estimated the effective transmission advantage using 10, 20, 30, or 50 days of data counted from the 11^{th} day (see Appendix C.3.2).

In each simulation scenario, we assessed the performance of the method using the following metrics:

- Bias, defined as difference between the mean posterior estimate of the effective transmission advantage and its true value. Bias should be as small as possible and will be zero for a perfect model.

- Uncertainty, defined as the posterior Standard Deviation (SD). Models with lower uncertainty are preferable as long as they are unbiased.

- Coverage probability, defined as the proportion of simulations where a given Credible Interval (CrI) of the transmission advantage contained the true value. The 95% coverage probability for a well-calibrated model should be 95%, i.e. the true value will be contained in the 95% CrI in 95% of the simulations (analogous criterion is applicable for 50% coverage probability) [607, 608];

- Classification. We used the posterior distribution of ϵ to classify the new variant as "more transmissible", "less transmissible" than the reference or "unclear" (see methods in main text). To assess the classification performance, we consider the proportion of simulations where the variant is classified correctly (when the true value of ϵ is 1, we consider the correct classification to be 'unclear', see results in Table C.6). A perfect model would always classify the variant correctly. In practice, the threshold posterior quantile used for classifying a variant as more

or less transmissible (see Methods section in the main text) determines the sensitivity of the classification and involves a well-known trade-off with its specificity.

Table C.5: **Parameter values used in the simulations.** For each simulation scenario, we considered all (relevant) combinations of parameter values shown in this table; and for each parameter combination, we simulated 100 data sets. CV: coefficient of variation

Parameter	Values
Reference R_t	1.1, 1.6
Mean of reference serial interval	5.4 days
Standard deviation of reference serial interval	1.5 days
Mean of variant serial interval	Reference serial interval mean × (0.5, 1, 1.5, 2)
CV of variant serial interval	Reference serial interval CV × (0.5, 1, 1.5, 2)
Overdispersion	0.1, 0.5, 1

We first considered a baseline scenario (Appendix C.5.3), where we assumed that the natural history of the reference and the new variants are same. We relaxed this assumption in other scenarios, assuming either that the SI distribution of the variant has a different mean (Appendix C.5.4) or Coefficient of Variation (CV) (Appendix C.5.6), but that the SI distribution of both the new and reference variants are correctly specified in MV-EpiEstim. We then explored a scenario typical of real-time outbreak analysis where the SI distribution (mean or CV) of the variant is different from that of the reference but in the absence of more information, is assumed to be the same as that of the reference (Appendices C.5.5 and C.5.7). We also explored the performance of our method in the presence of superspreading (Appendix C.5.8), extending the simulation to use a negative binomial offspring distribution, i.e.:

$$I_t \sim NegBin(R_t \sum_{s=1}^{t-1} I_s \omega_{t-s}, \kappa \sum_{s=1}^{t-1} I_s \omega_{t-s}), \tag{C.2}$$

where κ is the overdispersion parameter (lower values of κ denoting higher levels of superspreading) and $NegBin(\mu, k)$ denotes a negative binomial distribution with mean μ and variance $\mu + \frac{\mu^2}{k}$. In this formulation, we implicitly assume that the number of secondary infections per index case is distributed as $NegBin(R_t, \kappa)$. The sum of n independant identically distributed $NegBin(\mu, k)$ variables is $NegBin(n\mu, nk)$, yielding equation C.2 when $n = \sum_{s=1}^{t-1} I_s \omega_{t-s}$, $\mu = R_t$ and $k = \kappa$. We use the R package projections [609] to simulate data, which implements superspeading as just described.

Finally, we assessed the sensitivity of MV-EpiEstim to under-reporting (Appendix C.5.9), assuming a constant reporting rate for the reference and the new variant.

The scenarios outlined above used a constant R_t over the period of the simulation. We also explored the effect of time-varying R_t profiles on method performance (Appendix C.5.10). We also considered simulations with two locations with time-varying R_t (Appendix C.5.11), simulating independent epidemics in each location as described above. Finally, we assessed the performance of our

method when estimating the transmission advantage in a scenario where the advantage varied over time (Appendix C.5.12).

C.5.2 Description of figures

The figures that follow in the remainder of Appendix C.5 are composed of four panels. In each figure panel, we present a performance indicator summarised across 100 simulations. In each figure, the top-left panel shows the mean ± SD of the bias in the estimate of the effective transmission advantage. The dashed horizontal line denotes the threshold bias of 0. The top-right panel shows the mean ± SD of the uncertainty in estimates. The bottom-left panel shows the 95% (circles) and 50% (diamonds) coverage probabilities (mean and 95% binomial Confidence Interval (CI)). The dashed horizontal lines denote the threshold values of 0.95 and 0.50. The bottom-right panel shows classification performance (mean and 95% binomial CI). For definition of each performance indicator, see Appendix C.5.1.

C.5.3 Baseline scenario

In this section, we present the results for the baseline scenario. That is, we assume no superspreading, and that both the reference and the new variant have the same natural history. Results are shown for estimates obtained using 10, 20, 30 and 50 days of data in MV-EpiEstim.

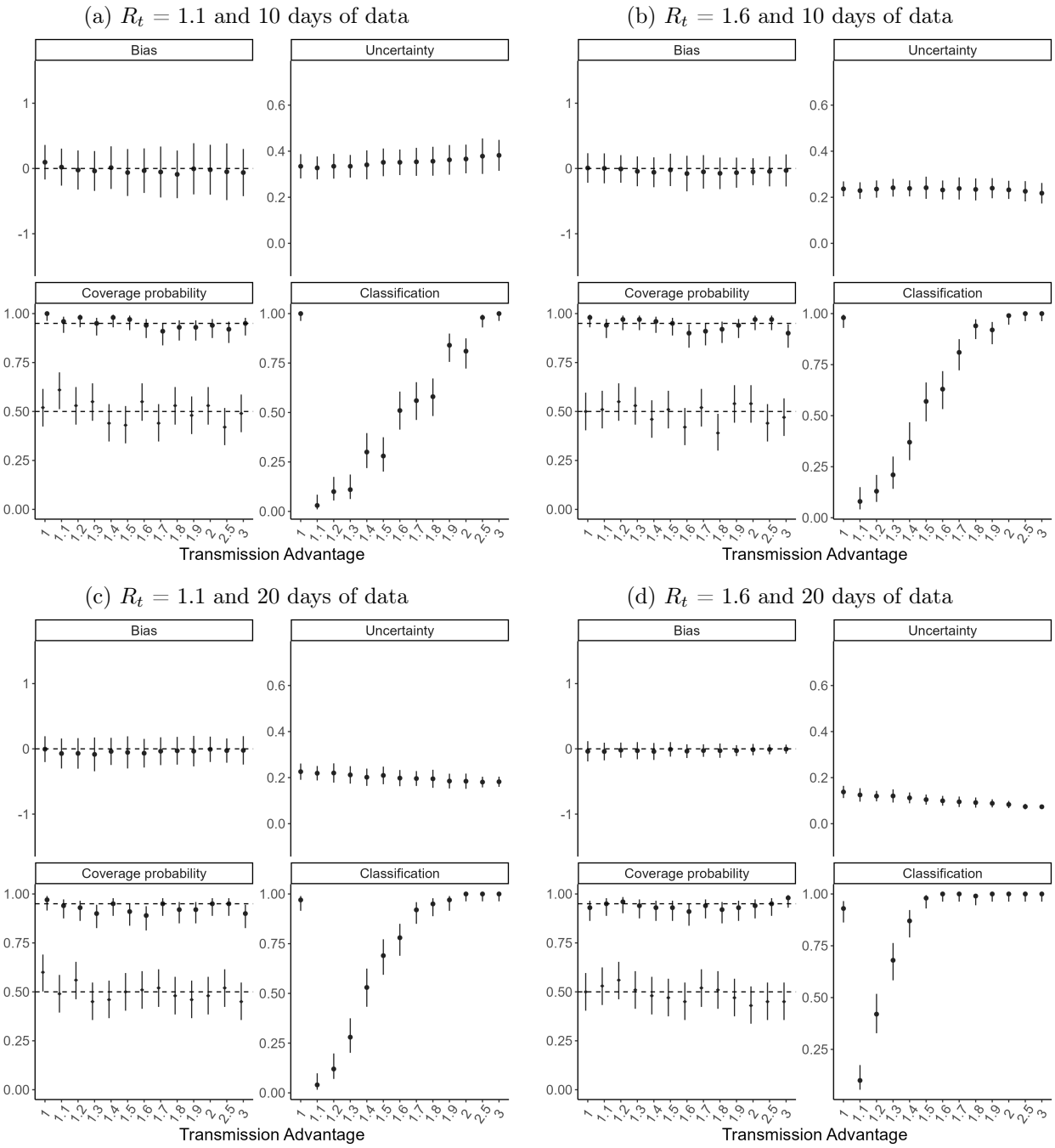


Figure C.12: **Method performance using simulated data assuming the same natural history for the reference and the variant (using 10 or 20 days of incidence data).** In each panel, the dots and vertical bars represent the central estimate and uncertainty respectively. See Appendix C.5.2 for details.

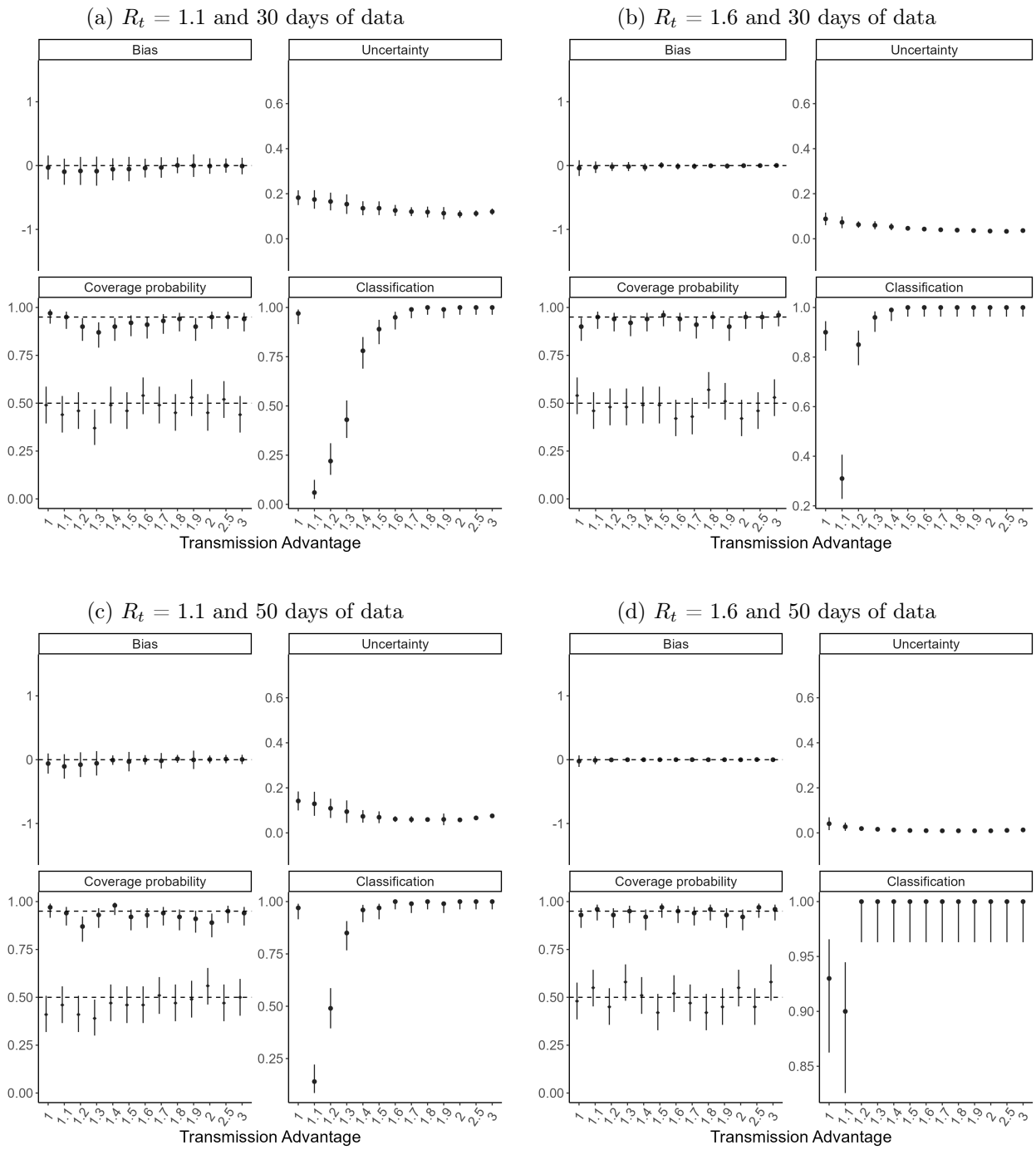


Figure C.13: **Method performance using simulated data assuming the same natural history for the reference and the variant (using 30 or 50 days of incidence data).** In each panel, the dots and vertical bars represent the central estimate and uncertainty respectively. See Appendix C.5.2 for details.

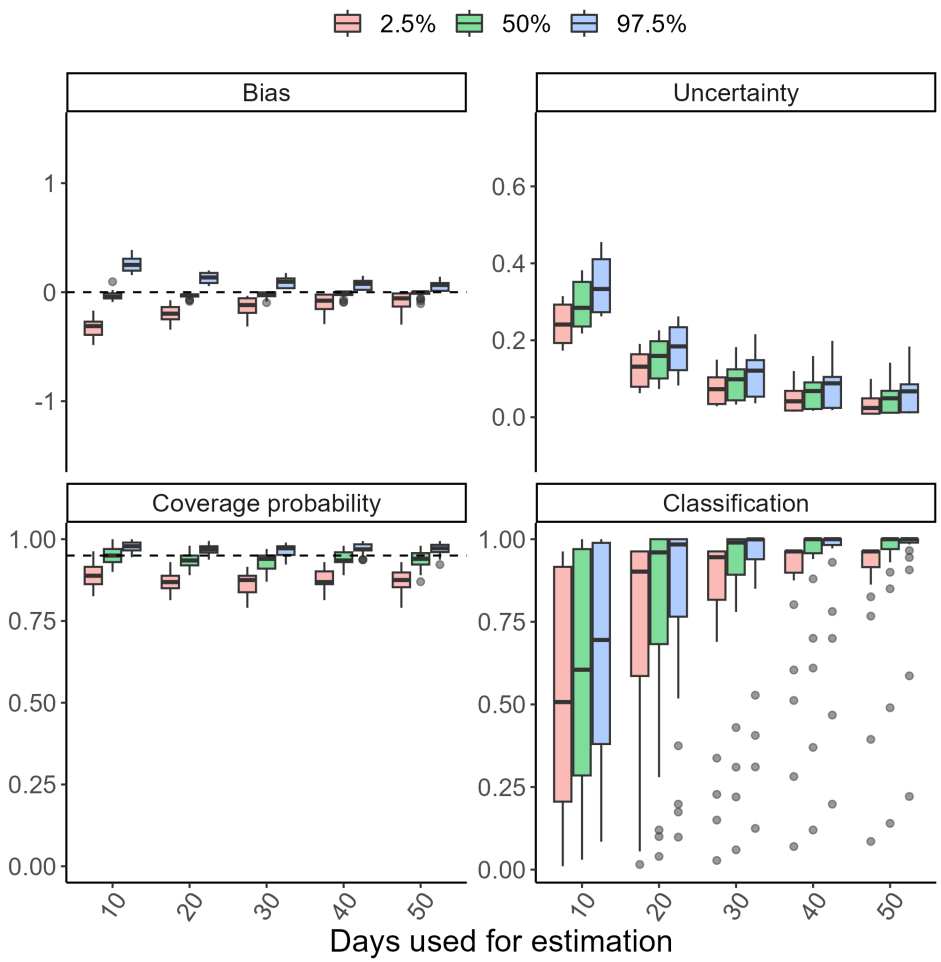


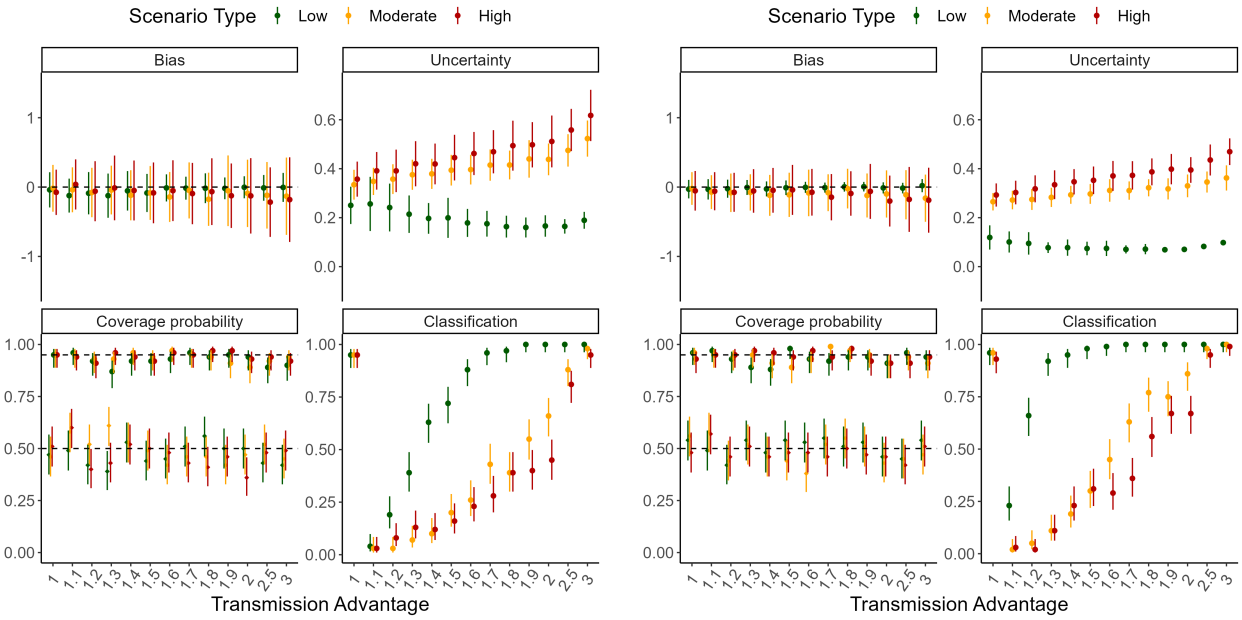
Figure C.14: **Method performance using simulated data assuming the same natural history for the reference and the variant.** In each panel, the boxplots depict the change in the distribution of the 2.5^{th}, 50^{th} and 97.5^{th} quantiles of the corresponding metric as progressively more data are used for estimation. For each metric, each quantile is summarised across all simulations for the values of reference R_t (1.1 and 1.6) and the range of ϵ (1 to 3) in the baseline scenario.

C.5.4 Sensitivity to serial interval mean

In this section, we present results for the scenario where data were simulated assuming different natural history parameters for the reference and the variant. We assumed that the mean serial interval of the variant is $0.5, 1.5$, or 2 times that of the reference. Further, we assumed that the parameters of both the reference and the variant are correctly specified during estimations. Results are shown using $10, 20, 30$, and 50 days of incidence data.

(a) $R_t = 1.1$ and 10 days of data

(b) $R_t = 1.6$ and 10 days of data



(c) $R_t = 1.1$ and 20 days of data

(d) $R_t = 1.6$ and 20 days of data

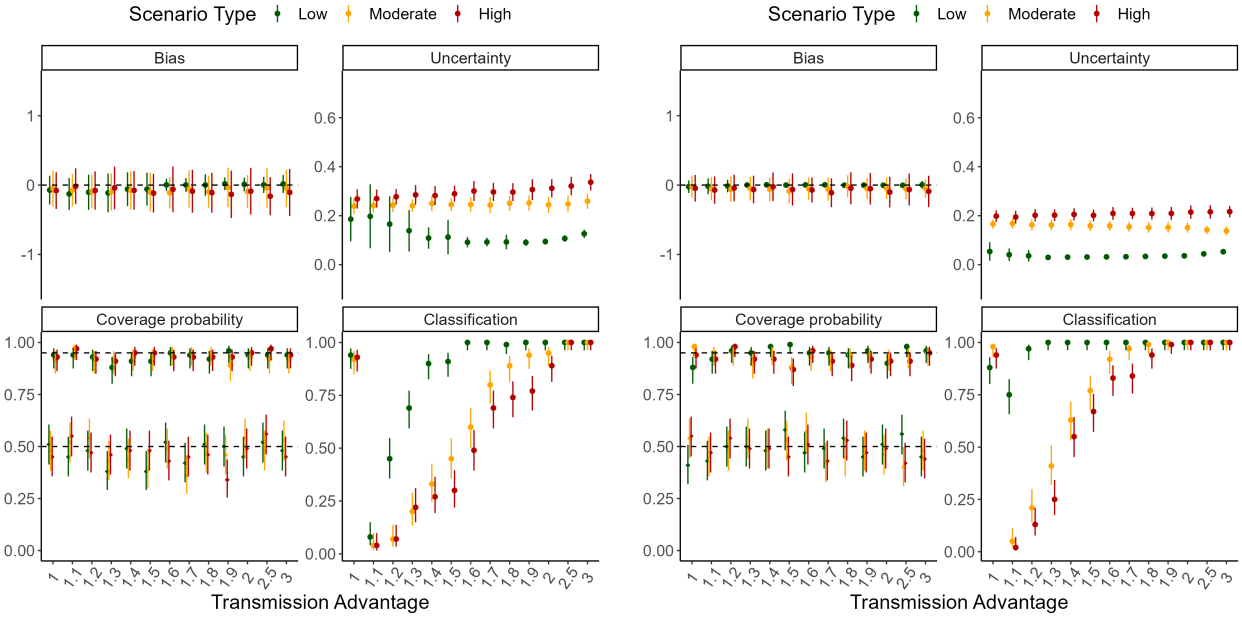


Figure C.15: **Method performance using simulated data assuming different SI mean for the reference and the variant (using 10 or 20 days of incidence data).** The mean serial interval of the variant is assumed to be 0.5 (low), 1.5 (moderate) or 2 (high) times that of the reference. In each panel, the dots and vertical bars represent the central estimate and uncertainty respectively. See Appendix C.5.2 for details.

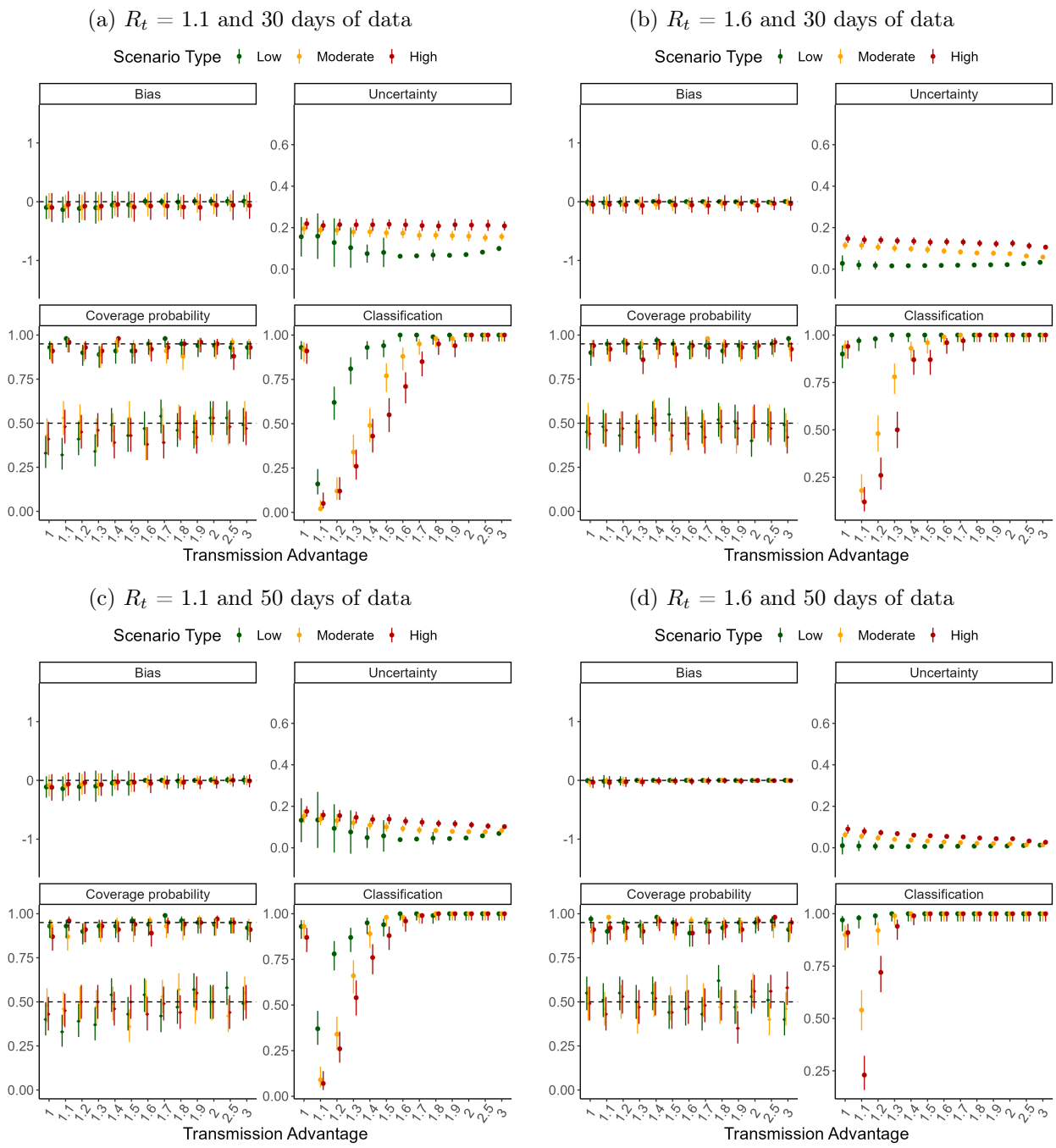


Figure C.16: **Method performance using simulated data assuming different SI mean for the reference and the variant (using 30 or 50 days of incidence data).** The mean serial interval of the variant is assumed to be 0.5 (low), 1.5 (moderate) or 2 (high) times that of the reference. In each panel, the dots and vertical bars represent the central estimate and uncertainty respectively. See Appendix C.5.2 for details.

C.5.5 Misspecification of serial interval mean

In this section, we present results for the scenario where data were simulated assuming different natural history parameters for the reference and the variant. We assumed that the mean serial interval of the variant is $0.5, 1.5$, or 2 times that of the reference. However, we assumed that the parameters of both the reference and the variant are assumed to be the same during estimation. Results are shown using $10, 20, 30$, and 50 days of incidence data.

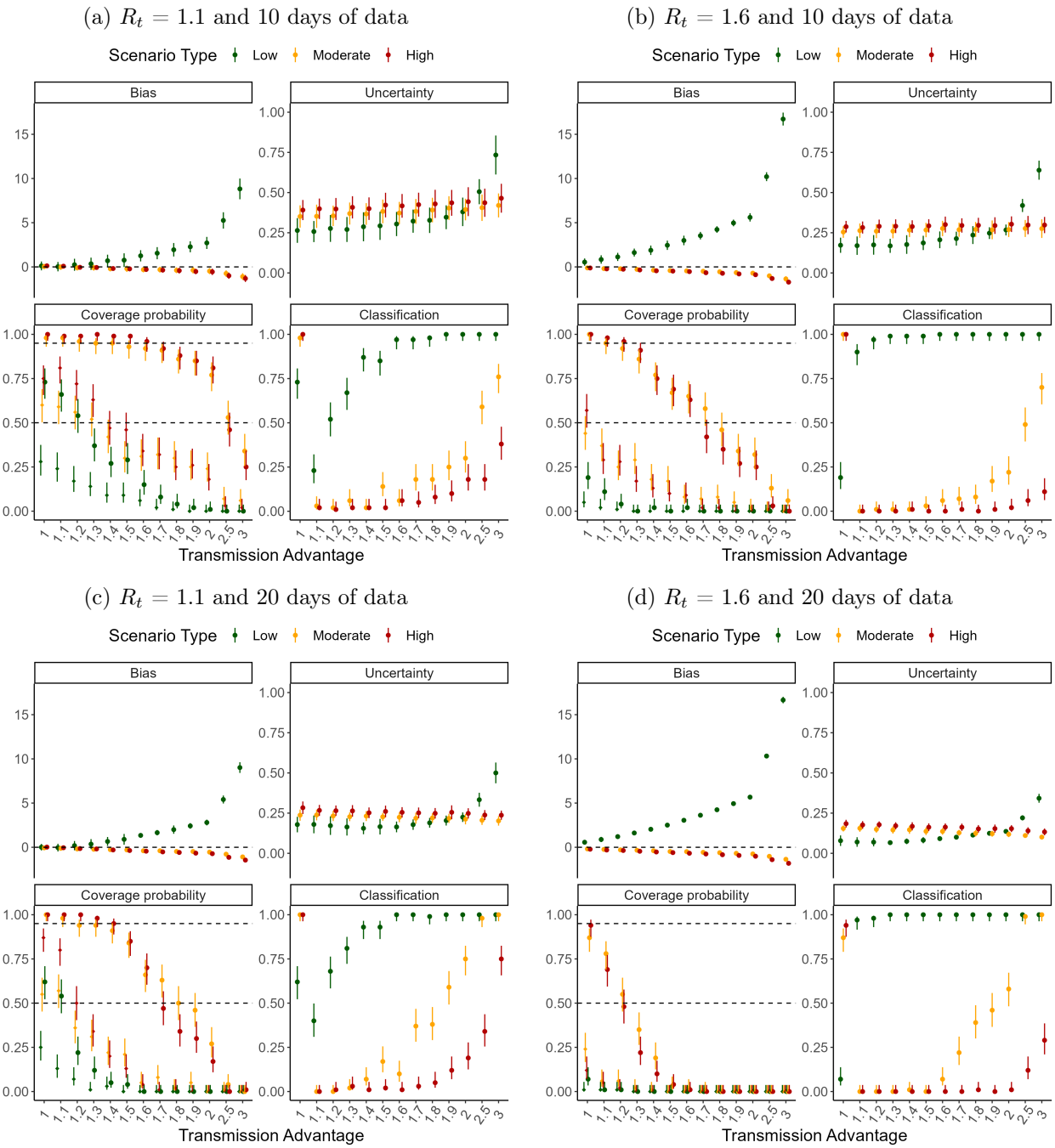
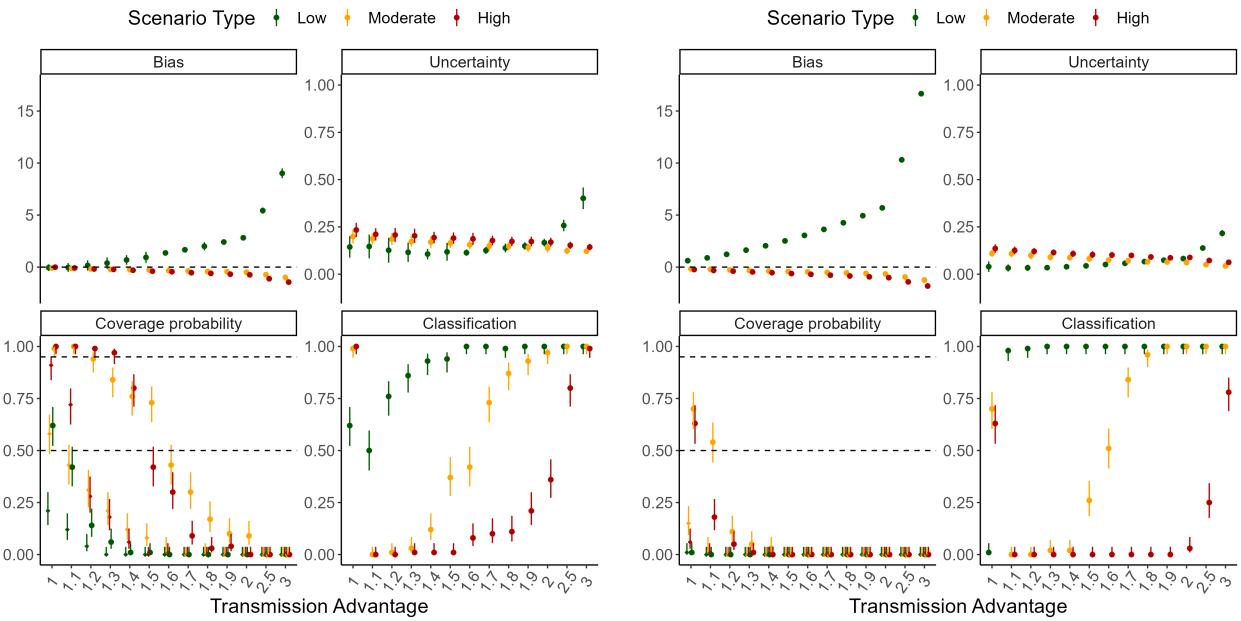


Figure C.17: **Method performance using simulated incidence data when the mean SI of the variant is different and is misspecified during estimation (using 10 or 20 days of incidence data).** The mean SI of the variant is assumed to be 0.5 (low), 1.5 (moderate) or 2 (high) times that of the reference. In each panel, the dots and vertical bars represent the central estimate and uncertainty respectively. See Appendix C.5.2 for details.

(a) $R_t = 1.1$ and 30 days of data

(b) $R_t = 1.6$ and 30 days of data



(c) $R_t = 1.1$ and 50 days of data

(d) $R_t = 1.6$ and 50 days of data

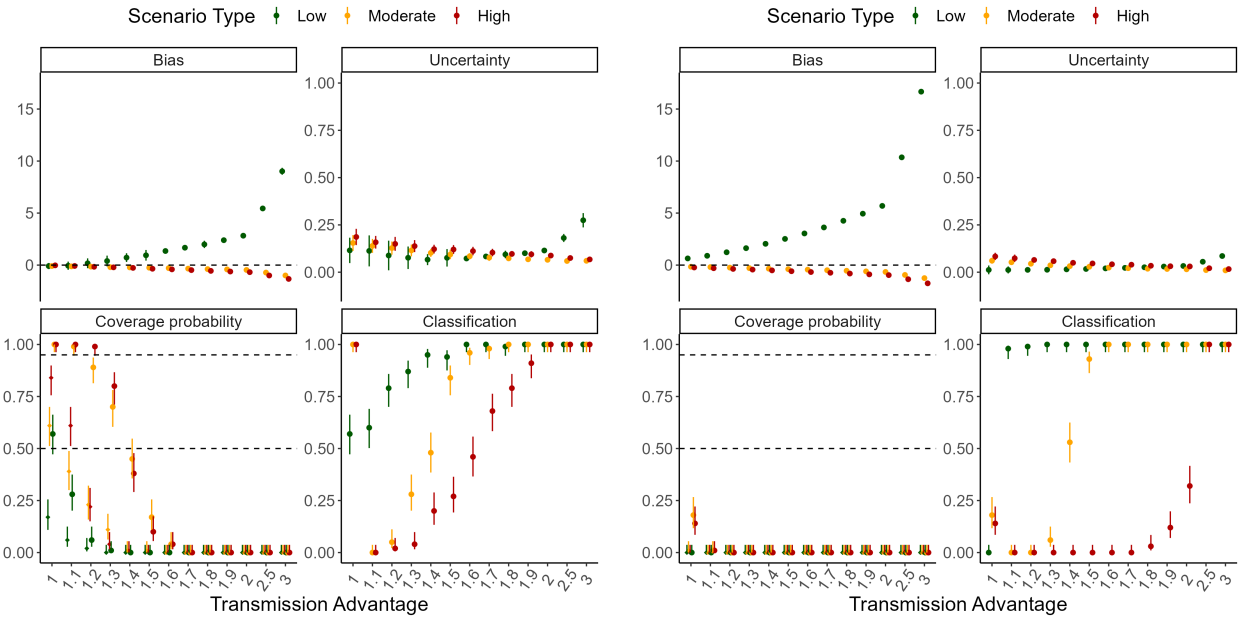


Figure C.18: **Method performance using simulated incidence data when the mean SI of the variant is different and is misspecified during estimation (using 30 or 50 days of incidence data).** The mean SI of the variant is assumed to be 0.5 (low), 1.5 (moderate) or 2 (high) times that of the reference. In each panel, the dots and vertical bars represent the central estimate and uncertainty respectively. See Appendix C.5.2 for details.

C.5.6 Sensitivity to serial interval CV

In this section, we present results for the scenario where data were simulated assuming different natural history parameters for the reference and the variant. We assumed that the CV of the SI distribution of the variant is $0.5, 1.5$, or 2 times that of the reference. Further, we assumed that the parameters of both the reference and the variant are correctly specified during estimation. Results are shown using $10, 20, 30$, and 50 days of data.

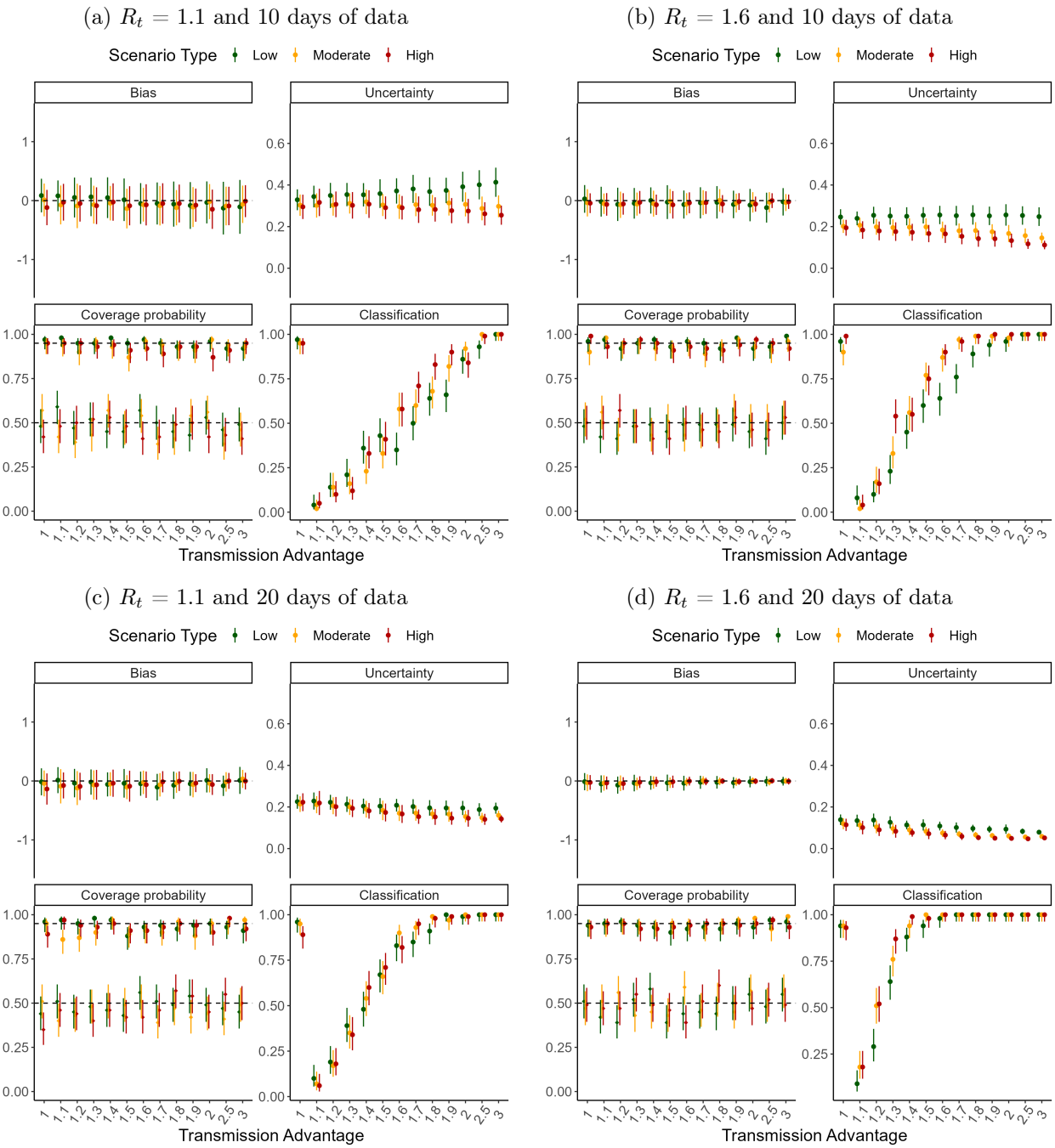


Figure C.19: **Method performance using simulated incidence data when the CV of the SI distribution of the variant is different and is correctly specified during estimation (using 10 or 20 days of incidence data).** The CV of the serial interval distribution of the variant is assumed to be 0.5 (low), 1.5 (moderate) or 2 (high) times that of the reference. In each panel, the dots and vertical bars represent the central estimate and uncertainty respectively. See Appendix C.5.2 for details.

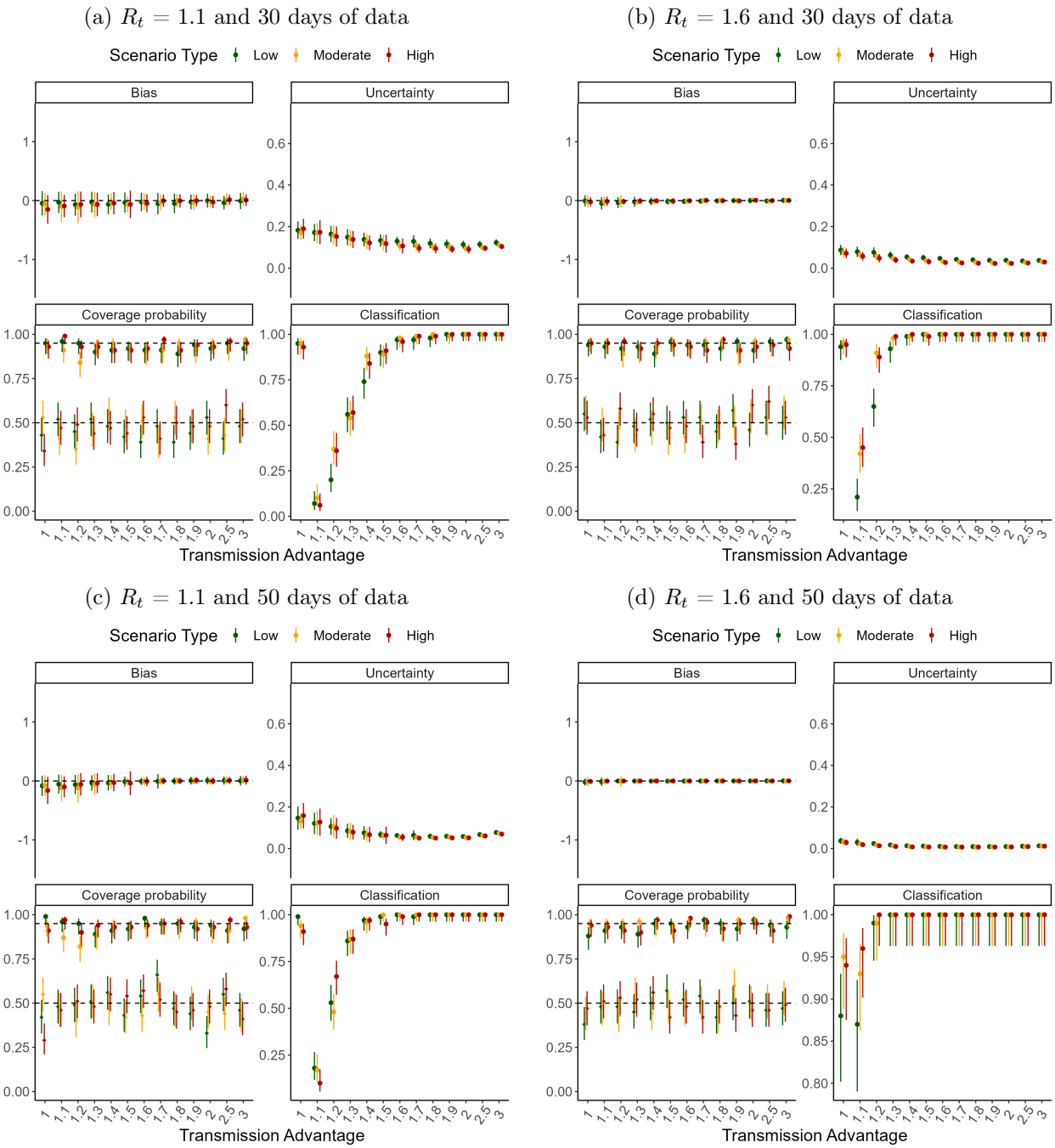


Figure C.20: **Method performance using simulated incidence data when the CV of the SI distribution of the variant is different and is correctly specified during estimation (using 30 or 50 days of incidence data).** The CV of the SI distribution of the variant is assumed to be 0.5 (low), 1.5 (moderate) or 2 (high) times that of the reference. In each panel, the dots and vertical bars represent the central estimate and uncertainty respectively. See Appendix C.5.2 for details.

C.5.7 Misspecification of serial interval CV

In this section, we present results for the scenario where data were simulated assuming different natural history parameters for the reference and the variant. We assumed that the CV of the SI distribution of the variant is $0.5, 1.5$, or 2 times that of the reference. However, we assumed that the parameters of both the reference and the variant are assumed to be the same during estimation. Results are shown using $10, 20, 30$, and 50 days of incidence data.

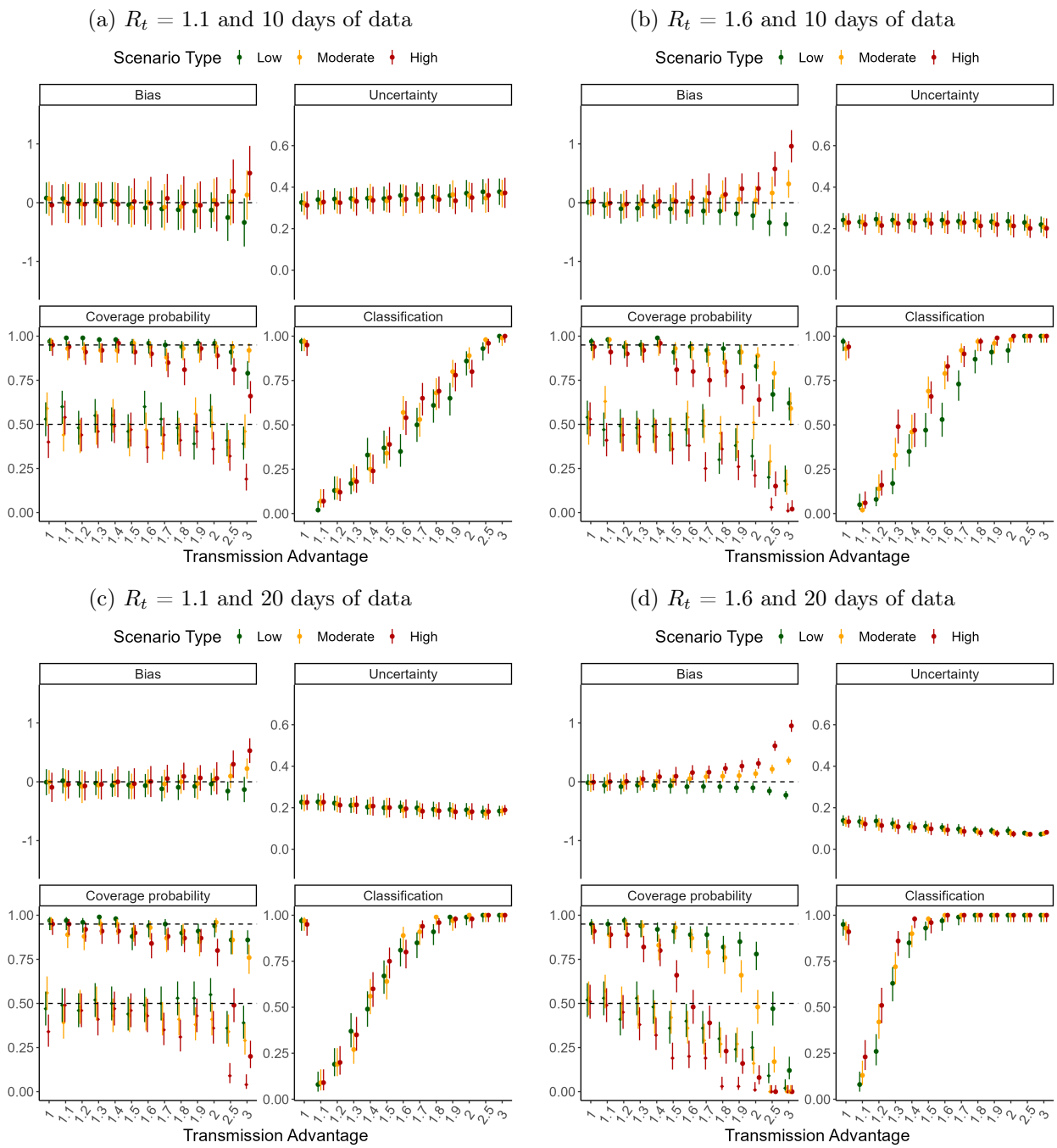
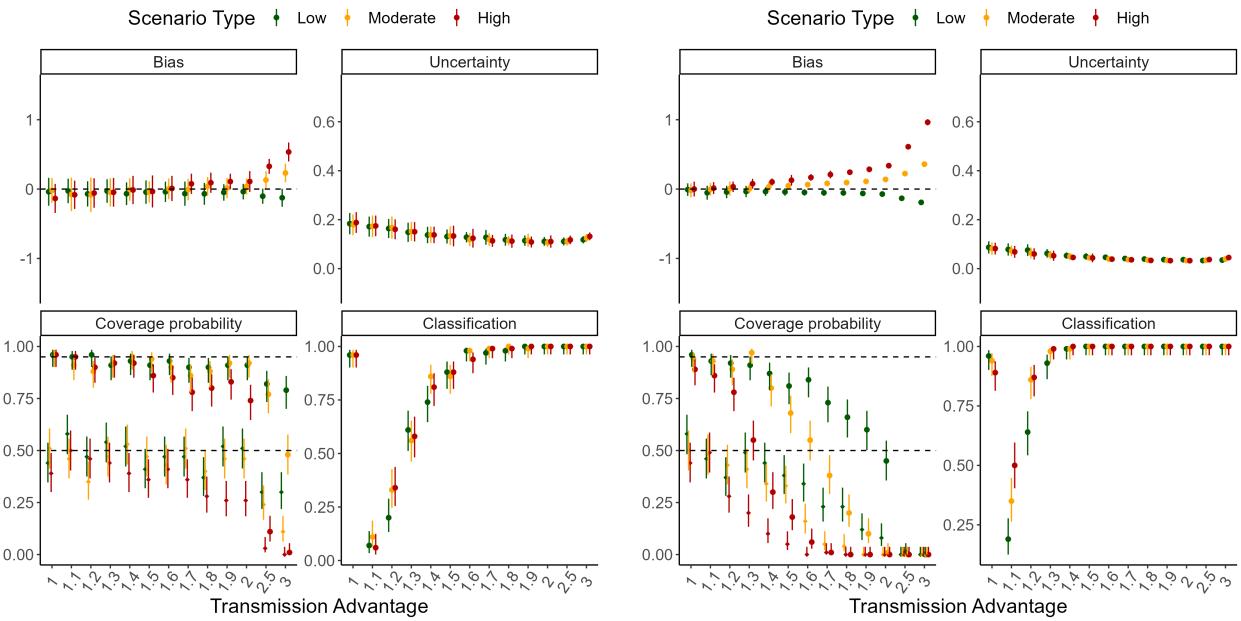


Figure C.21: **Method performance using simulated incidence data when the CV of the SI distribution of the variant is different and is misspecified during estimation (using 10 or 20 days of incidence data).** The CV of the serial interval distribution of the variant is assumed to be 0.5 (low), 1.5 (moderate) or 2 (high) times that of the reference. In each panel, the dots and vertical bars represent the central estimate and uncertainty respectively. See Appendix C.5.2 for details.

(a) $R_t = 1.1$ and 30 days of data

(b) $R_t = 1.6$ and 30 days of data



(c) $R_t = 1.1$ and 50 days of data

(d) $R_t = 1.6$ and 50 days of data

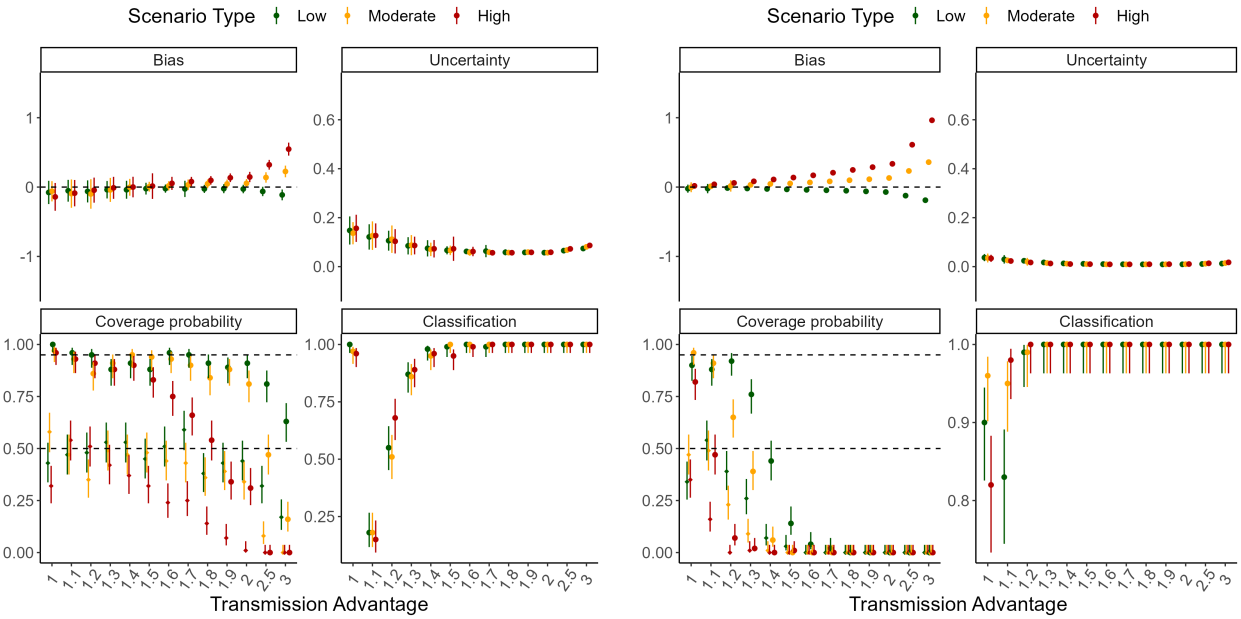


Figure C.22: **Method performance using simulated incidence data when the CV of the SI distribution of the variant is different and is misspecified during estimation (using 30 or 50 days of incidence data).** The CV of the SI distribution of the variant is assumed to be 0.5 (low), 1.5 (moderate) or 2 (high) times that of the reference. In each panel, the dots and vertical bars represent the central estimate and uncertainty respectively. See Appendix C.5.2 for details.

C.5.8 Sensitivity to superspreading

To explore sensitivity of MV-EpiEstim to superspreading (which is not explicitly accounted for in MV-EpiEstim), we used a Negative Binomial distribution as the offspring distribution as in C.2. We simulated data using low (overdispersion parameter $\kappa = 1$), moderate ($\kappa = 0.5$), and high ($\kappa = 0.1$) levels of superspreading. This section presents the performance metrics when 10, 20, 30, and 50 days of incidence data were used for estimation.

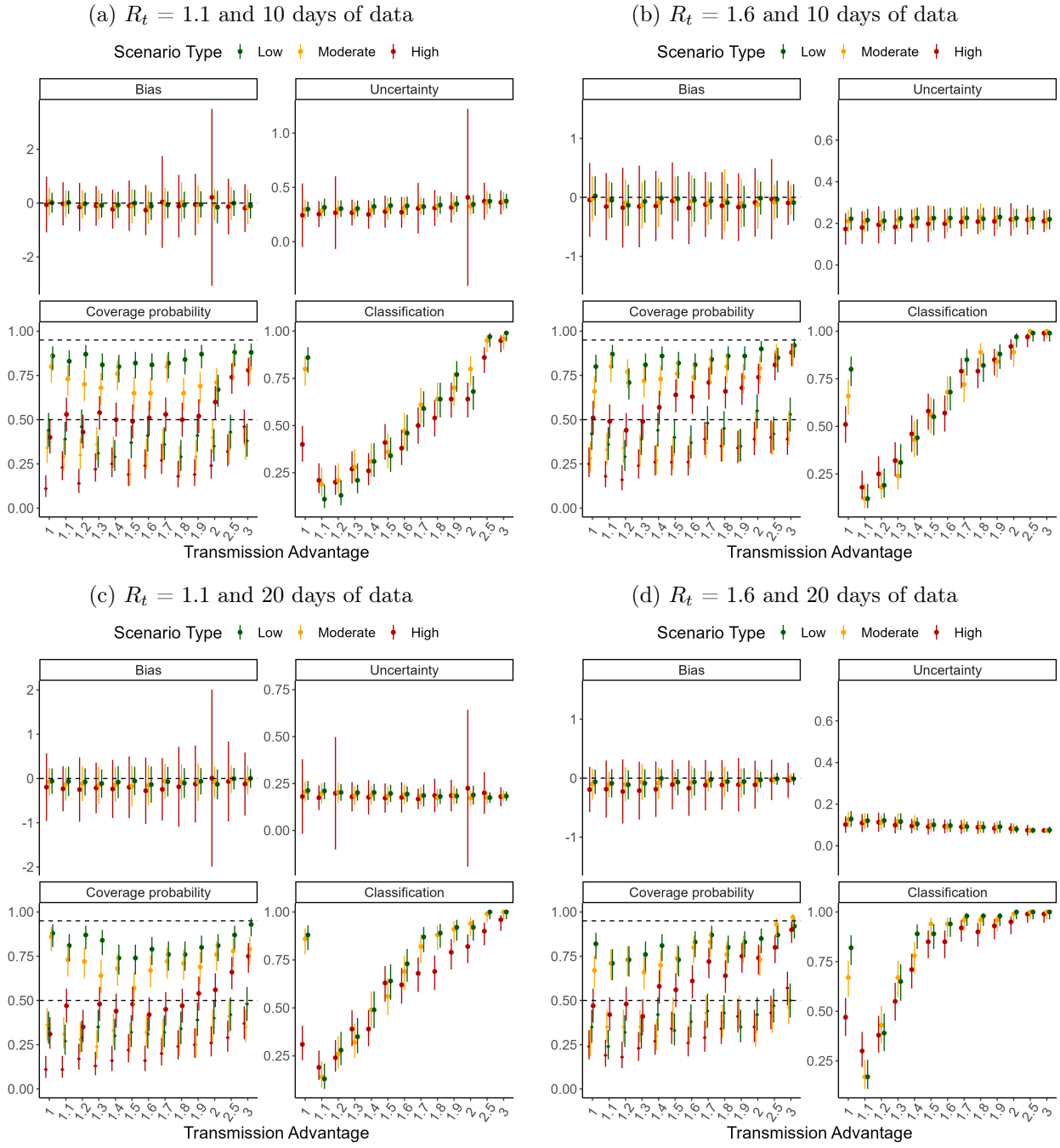
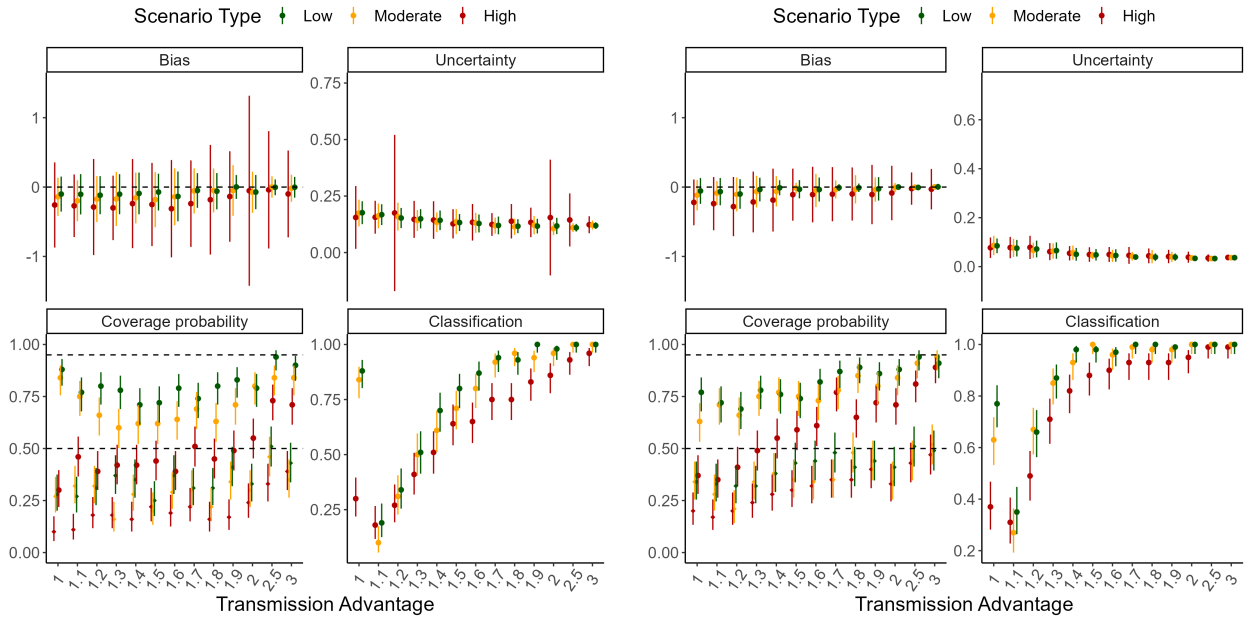


Figure C.23: **Method performance using incidence data simulated with superspreading (using 10 or 20 days of incidence data).** We simulated data with low (overdispersion parameter $\kappa = 1$), moderate ($\kappa = 0.5$) and high ($\kappa = 0.1$) levels of superspreading. In each panel, the dots and vertical bars represent the central estimate and uncertainty respectively. See Appendix C.5.2 for details.

(a) $R_t = 1.1$ and 30 days of data

(b) $R_t = 1.6$ and 30 days of data



(c) $R_t = 1.1$ and 50 days of data

(d) $R_t = 1.6$ and 50 days of data

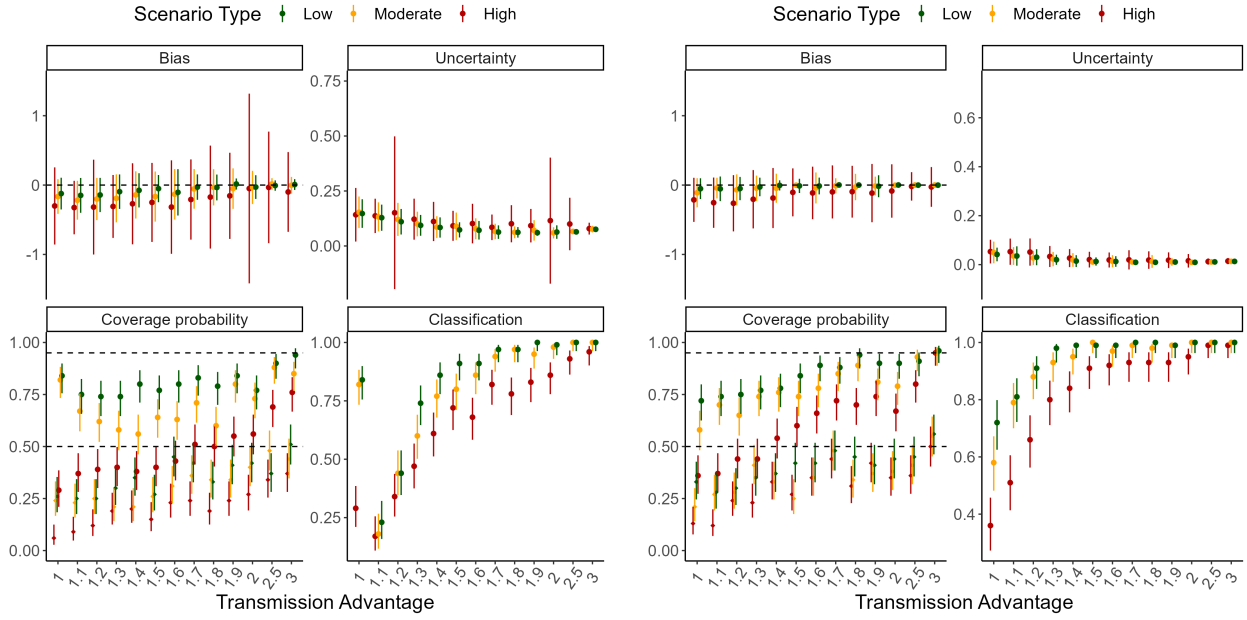


Figure C.24: **Method performance using incidence data simulated with superspreading (using 30 or 50 days of incidence data).** We simulated data with low (overdispersion parameter $\kappa = 1$), moderate ($\kappa = 0.5$) and high ($\kappa = 0.1$) levels of superspreading. In each panel, the dots and vertical bars represent the central estimate and uncertainty respectively. See Appendix C.5.2 for details.

C.5.9 Sensitivity to under-reporting

To explore the sensitivity of our method to under-reporting, we first simulated data as in the baseline scenario (Appendix C.5.3). We then assumed a constant reporting probability for both the reference and the variant and estimated the effective transmission advantage using only the reported cases. We set the reporting probability to 0.2, 0.5, or 0.8. This section presents the performance metrics when 10, 20, 30, and 50 days of incidence data were used for estimation.

(a) $R_t = 1.1$ and 10 days of data

(b) $R_t = 1.6$ and 10 days of data

(c) $R_t = 1.1$ and 20 days of data

(d) $R_t = 1.6$ and 20 days of data

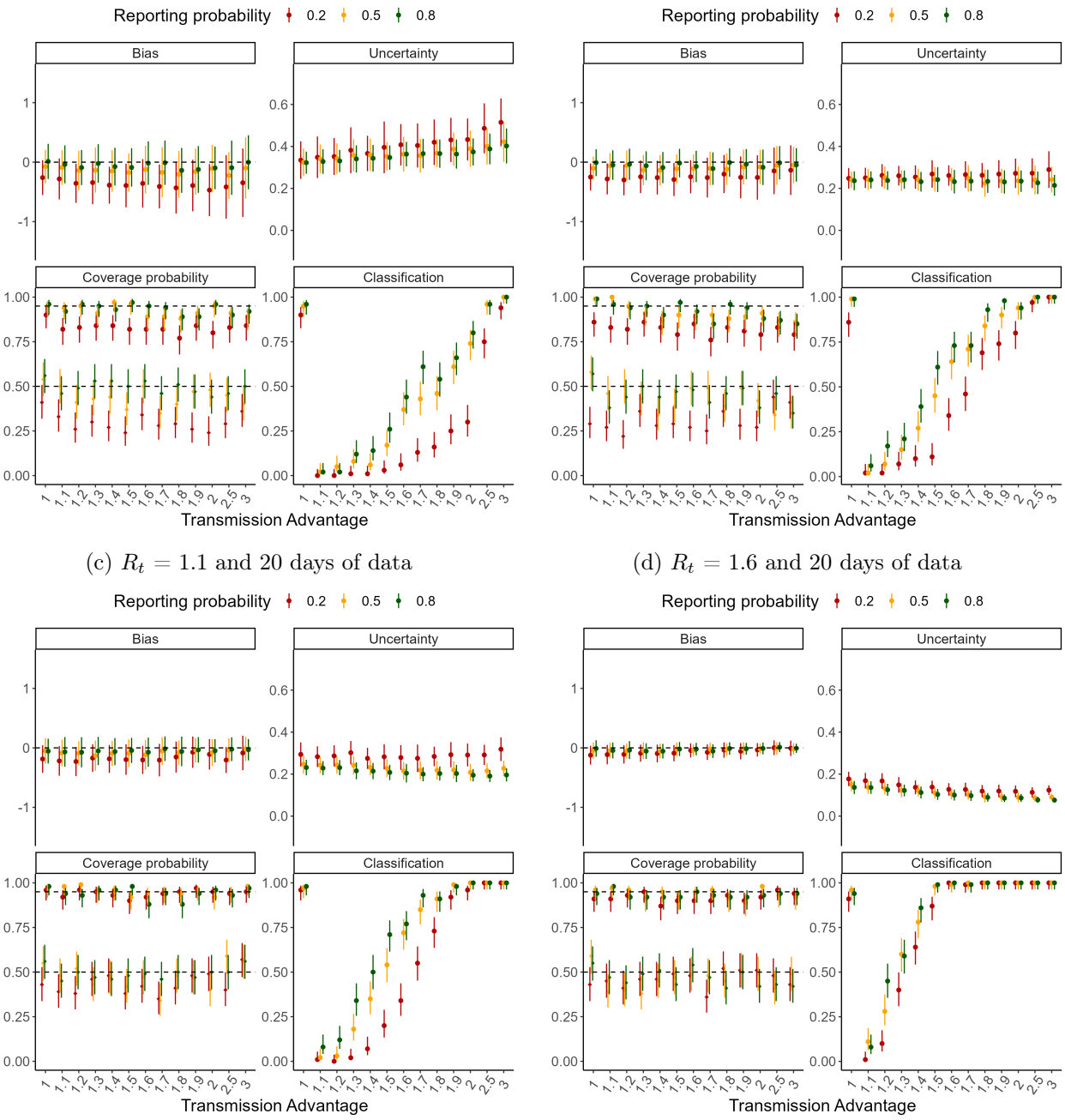
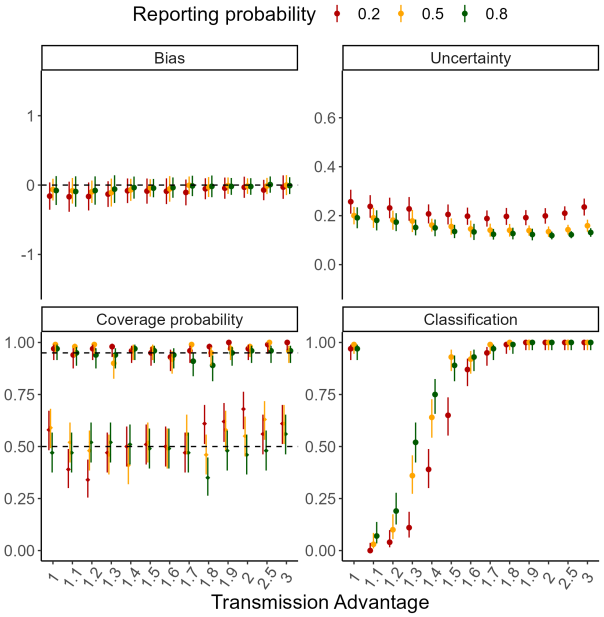
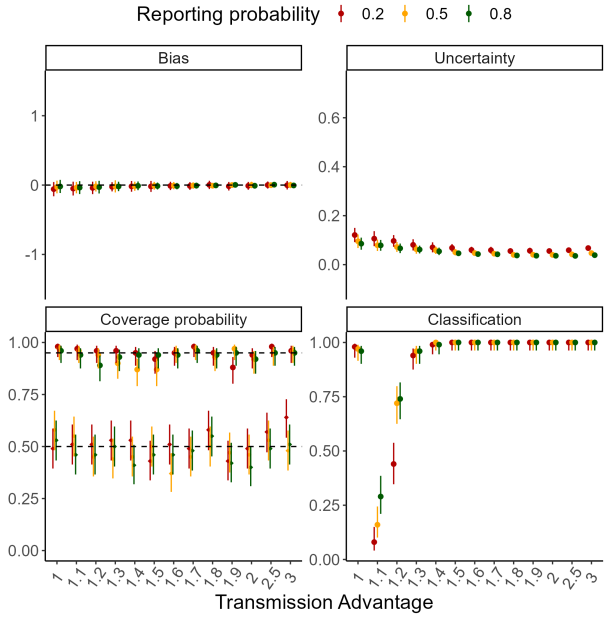


Figure C.25: **Method performance using using simulated incidence data with under-reporting (using 10 or 20 days of incidence data).** We assume that the reporting probability is 0.2, 0.5, or 0.8. In each panel, the dots and vertical bars represent the central estimate and uncertainty respectively. See Appendix C.5.2 for details.

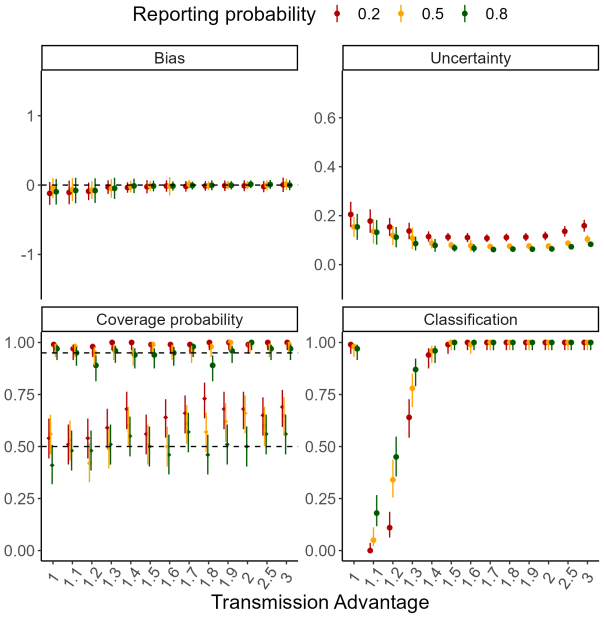
(a) $R_t = 1.1$ and 30 days of data



(b) $R_t = 1.6$ and 30 days of data



(c) $R_t = 1.1$ and 50 days of data



(d) $R_t = 1.6$ and 50 days of data

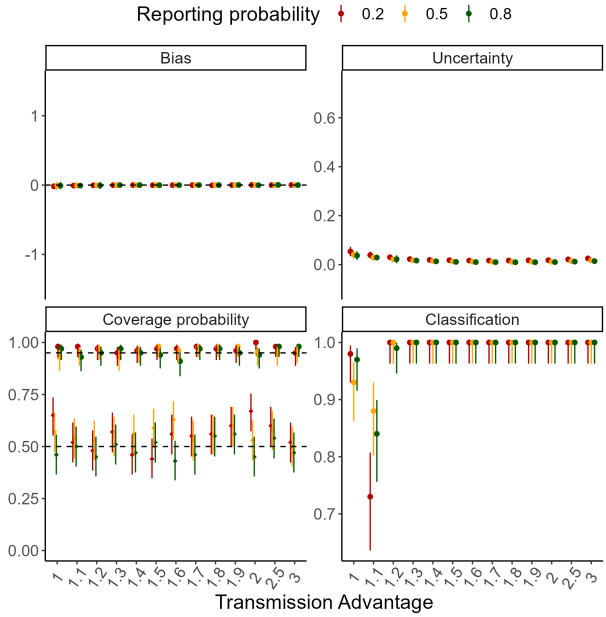


Figure C.26: **Method performance using using simulated incidence data with under-reporting (using 30 or 50 days of incidence data).** We assume that the reporting probability is 0.2, 0.5, or 0.8. In each panel, the dots and vertical bars represent the central estimate and uncertainty respectively. See Appendix C.5.2 for details.

Table C.6: **Classification of the new variant (as unclear, 'less' or 'more' transmissible) when the true transmission advantage is 1.** Note that in this case, 'unclear' is considered as the correct classification (see Appendix C.5.1 for more details). For each scenario, we show the proportion of simulations with each classification in 100 simulations using for 10, 20, 30, 40 or 50 days of data each and across all combinations of relevant for the scenario type (reference R_t 1.1 or 1.6, ϵ varying from 1 to 3, and other relevant parameter). See Table C.5 for a full list of parameters for each scenario.

Scenario	Unclear	Variant less transmissible	Variant more transmissible	Simulations
Baseline	0.96	0.03	0.02	1000
Different SI Mean	0.93	0.05	0.02	3000
Misspecified SI Mean	0.66	0.15	0.20	3000
Different SI CV	0.94	0.04	0.02	3000
Misspecified SI CV	0.94	0.03	0.03	3000
Superspreading	0.64	0.26	0.10	3000
Underreporting	0.96	0.03	0.01	3000

C.5.10 Time-varying R_t

To explore the effect of changing transmission dynamics on the estimates, we simulated data as in the baseline scenario (Appendix C.5.3) but with the reference R_t changing after 30 days from either 1.4 to 1.1, or 1.6 to 1.2. This section presents the performance metrics when 50 days of incidence data were used for estimation (i.e. covering the period both before and after the step-change in R_t).

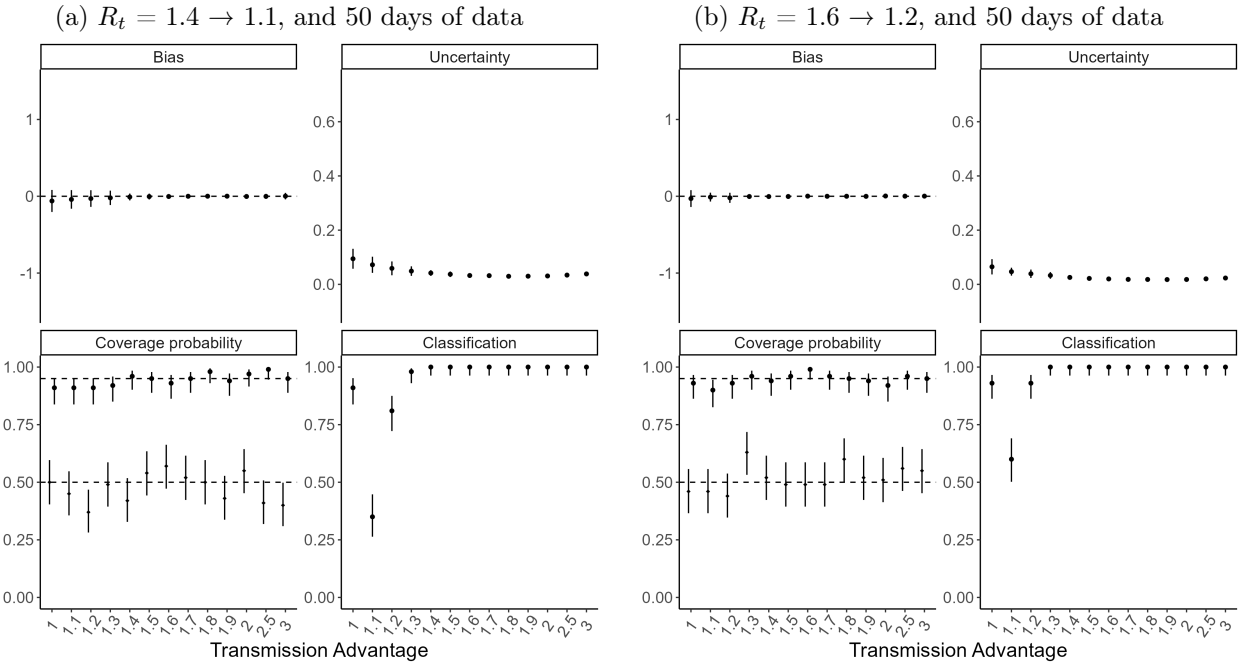


Figure C.27: **Method performance using incidence data simulated with time-varying R_t.** The reference R_t changes after 30 days of the simulation. Here we use 50 days of incidence data for estimation. In each panel, the dots and vertical bars represent the central estimate and uncertainty respectively. See Appendix C.5.2 for details.

C.5.11 Two locations with time-varying R_t

We explored the performance of our method in the presence of changing transmission dynamics when data from two locations are used for estimation, instead of a single location. We simulated independent epidemics in two locations as in the baseline scenario (Appendix C.5.3) but with the R_t profile changing over time. The reference R_t decreased from (i) 1.4 to 1.1, or (ii) 1.6 to 1.2, after 20 days in the first location and after 40 days in the second location. We also explored a further scenario where the decrease in reference R_t is different in the two locations and occurs at different times (from 1.4 to 1.1 after 40 days in the first location, and from 1.6 to 1.2 after 20 days in the second location). Table C.7 presents a summary of the R_t profiles used for simulations. This section presents the performance metrics when 50 days of incidence data were used for estimation (i.e. covering the period both before and after the step-changes in R_t).

Table C.7: **Reference R_t values used to simulate incidence data in scenarios with time-varying R_t profiles.** The method performance results using incidence data generated by the R_t profiles in rows 1, 2 and 3 are shown in Figures C.28a, C.28b and C.29 respectively.

Location 1			Location 2		
Initial R_t	Final R_t	Time of R_t change	Initial R_t	Final R_t	Time of R_t change
1.4	1.1	20 days	1.4	1.1	40 days
1.6	1.2	20 days	1.6	1.2	40 days
1.4	1.1	40 days	1.6	1.2	20 days

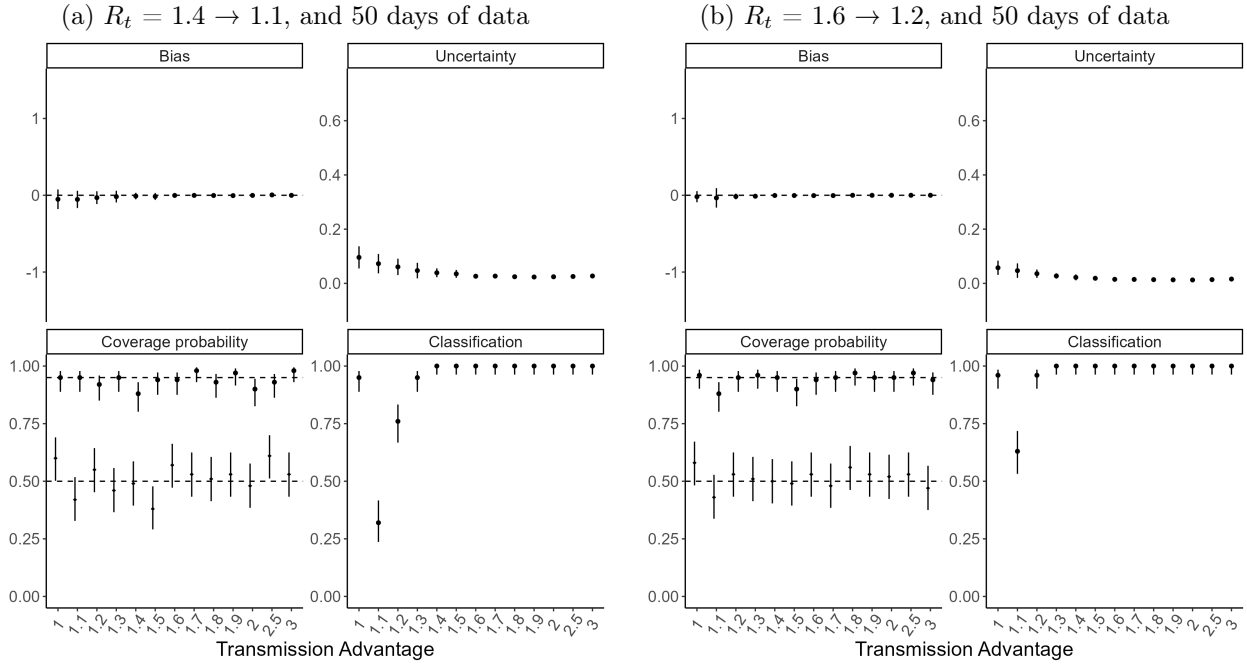


Figure C.28: **Method performance using incidence data simulated with time-varying R_t in two locations. In both locations the reference R_t decreases in the same way, but the change occurs at day 20 of the simulation for the first location and day 40 for the second location.** Here we use 50 days of incidence data for estimation. In each panel, the dots and vertical bars represent the central estimate and uncertainty respectively. See Appendix C.5.2 for details.

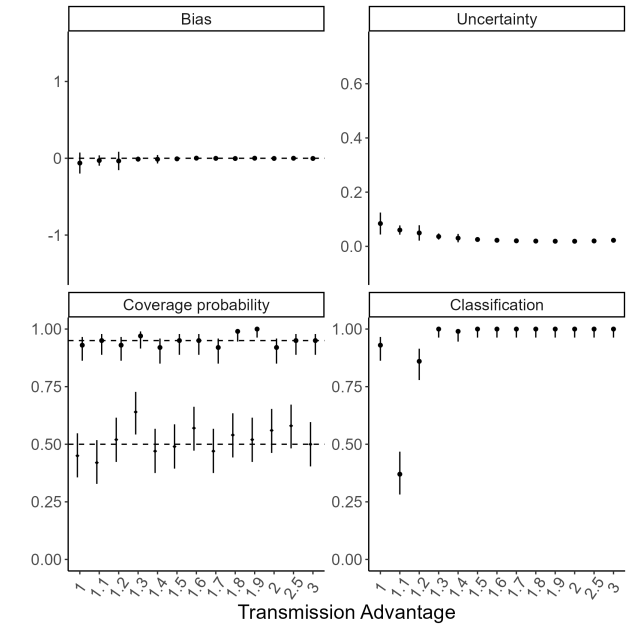


Figure C.29: **Method performance using incidence data simulated with time-varying R_t in two locations.** The reference R_t decreases at day 20 in the first location and day 40 in the second location. Here we use 50 days of incidence data for estimation. In each panel, the dots and vertical bars represent the central estimate and uncertainty respectively. See Appendix C.5.2 for details.

C.5.12 Time-varying transmission advantage

To explore how well the transmission advantage could be estimated in a scenario where ϵ varied over time, we simulated data as in the baseline scenario (Appendix C.5.3) but with a fixed reference R_t of 1.1. Incidence data were simulated with an initial transmission advantage of 1.1 which then increased linearly to 1.5 over a period of 30 days and then remained constant. This section presents the performance metrics over the duration of the simulated incidence. We estimated ϵ every 10 days using the latest 7 days of data, latest 10 days of data, or all available data.

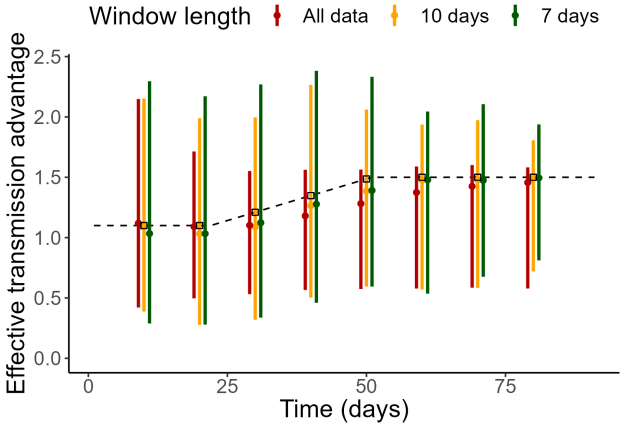


Figure C.30: **Effective transmission advantage estimated using MV-EpiEstim on simulated data in which the value of the transmission advantage changes over time.** The 'true' value of the transmission advantage over time used to simulate the incidence data is shown by the hollow black squares and black dashed line. We show the median (circles) and 95% CrI (lines) of the posterior distribution for ϵ aggregated across all simulations. We estimated ϵ every 10 days into the simulation using the latest 7 days of data (green), latest 10 days of data (yellow), or all the data available up to that point (red). Using the entire time series to estimate ϵ corresponds to ignoring any temporal variability.

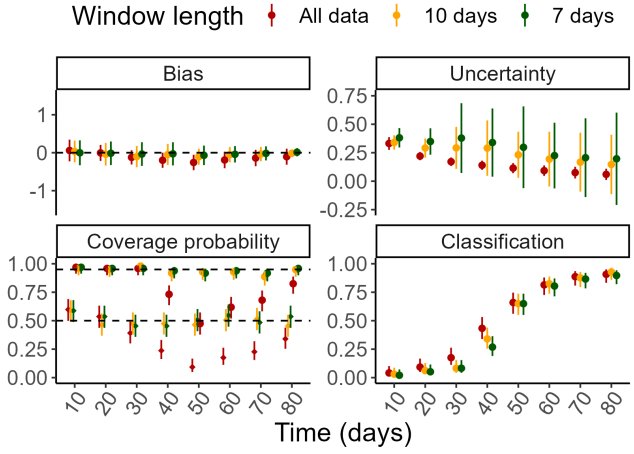


Figure C.31: **Method performance using incidence data simulated with time-varying ϵ in one location.** The reference R_t remains fixed at 1.1 throughout, but the value of ϵ changes over time. It has a value of 1.1 until day 30 of the simulation, at which point it increases linearly to 1.5 over a period of 30 days, then remains constant at 1.5 for the remainder of the simulation. We estimated ϵ every 10 days into the simulation using the latest 7 days of data (green), latest 10 days of data (yellow), or all the data available up to that point (red). Using the entire time series to estimate ϵ corresponds to ignoring any temporal variability. In each panel, the dots and vertical bars represent the central estimate and uncertainty respectively. See Appendix C.5.2 for details.

C.6 Literature review

In total, 66 studies were identified that aimed to compare the transmissibility of SARS-CoV-2 variants. Of those, 53 explicitly provided one or more estimates of the transmission advantage of one variant over the other (see the supplementary database available in [240]). 21 studies provided code, of which only 5 studies provided packaged software, which we list in Table C.8. None of these 5 studies assessed the method performance against simulated data.

Table C.8: **Github repositories for packaged code used in 5 of the studies identified in the literature review.**

Github repository	Number of studies
`https://github.com/mrc-ide/reactidd`	2
`https://github.com/BDI-pathogens/VariantREstimate`	1
`https://mrc-ide.github.io/sircovid/`	1
`https://github.com/haschka/SIER_multivariant_epidemic/`	1

105 transmission advantage estimates were found across the 53 studies. These were divided into six different advantage types, including the effective transmission advantage in: 1) the reproduction number R, 2) the growth rate, 3) the transmission rate, 4) the secondary attack rate, and two further categories which aimed to disentangle the intrinsic transmission advantage in R versus either: 5) immune escape, or 6) the generation time.

The majority of estimates were, as our method, based solely on incidence data (46/105, 44%), which included raw incidence of each variant or total incidence adjusted by the estimated proportion of each variant based on available sequencing data. For those estimates, the underlying models used were generally renewal equation-based models or exponential growth models.

Many estimates were based on using dynamic transmission model-based inference systems (27/105, 26%), such as compartmental models, which typically require additional data (e.g. on hospitalisations, deaths, population size, interventions) and assumptions (e.g. on disease progression and severity, immunity, etc). Phylodynamic models fitted to a combination of incidence and genomic data were used for a couple of estimates (2/105, 1.9%). A number of estimates (17/105, 16.2%) used household surveys to estimate secondary attack rates for each variant.

Overall, there was only 1 study with broad applicability (in that it only uses incidence data) that was also packaged in a ready-to-use tool [265]. We note that this study did explicitly account for overdispersion. Three other packages were identified through the review (Table C.8), but they required a wealth of additional data for fitting and none of those assessed the method performance against simulated data, including a range of transmission advantage scenarios, and a systematic exploration of the impact of mis-specifying the natural history of the new variant, or of the presence of overdispersion or under-reporting on the performance of their method.

A number of other studies explored the trade-off between a change in the generation time and a change in transmission [266, 267], including one with a simulation study assessing the statistical framework performance [266].

See the Supplementary Database published online [240] for all extracted estimates of transmission advantages and hyperlinks to the available code and R packages.

C.7 Code and Data availability

All data and code used in this analysis are available at `https://github.com/mrc-ide/epiestims`. MV-EpiEstim is available in the development version of EpiEstim at `https://github.com/mrc-ide/EpiEstim`.

Appendix D

Supplementary Information: Estimating the epidemic reproduction number from temporally aggregated incidence data

The contents of this appendix has been adapted from supplementary information previously published in PLOS Computational Biology [270].

D.1 Overview

In this supplementary information we present further results from our analyses of influenza-like illness and COVID-19 data. We also present additional analyses from the simulation study, which was used to validate our method using a variety of simulated epidemic scenarios, where R_t either remained constant or varied over time. Here, we explore the application of the method to alternative temporal aggregations of data e.g., 3-day, 10-day, or two-weekly aggregations and discuss the number of iterations generally required to reach convergence when reconstructing daily incidence data using our Expectation Maximisation (EM) algorithm. We also compare our method to an alternative Locally Estimated Scatter plot Smoothing (LOESS) smoothing approach, and present another real-world example using incidence data for Zika, as an example of a disease with non-respiratory transmission.

D.2 Additional results

As described in the main text, reported incidence data for cases of influenza-like illness, COVID-19 cases, and COVID-19 deaths was used to estimate R_t using the original EpiEstim R package [62] as well as our extended method. The reported daily data were artificially aggregated to a weekly timescale to replicate a typical scenario where data are reported on a weekly basis. Both the reported daily data and the reconstructed daily data, which was obtained from the weekly aggregations of incidence using

our EM algorithm, were used to estimate R_t over daily and weekly sliding windows ending on day t.

D.2.1 Influenza

In the influenza case study, both the daily and weekly sliding R_t estimates are smoother when estimated from the reconstructed data as opposed to the reported data (Figure D.1). The prominent weekly oscillations in the daily R_t estimates from the reported data are likely due to the impact of day-to-day variations in reporting, such as weekend effects, which in this dataset may be caused by factors including the military clinic's opening hours or the military personnel's working hours [284]. These intra-weekly fluctuations are lost once the data are aggregated; therefore, the reconstructed incidence and the R_t estimates based on it will be smoother and less affected by this variability (Figure D.1 & Figure D.2A-B).

Weekly sliding R_t estimates account for some of this variation, leading to smoother R_t estimates from both the reported and reconstructed data (Figure D.1B). It is much clearer here that the estimates follow the same general trend, but once again, the R_t estimates are slightly smoother from the reconstructed data.

Bias in the R_t estimates was assessed by computing the absolute difference in the estimates from the reported and reconstructed data (Figure D.2C-D). The greater differences between the R_t estimates correspond to the largest disparities between the reported and reconstructed incidence data (Figure D.2).

D.2.2 COVID-19 cases

In the first case study for COVID-19, the overall trends in R_t over time using the reported and reconstructed daily data were similar for both the daily and weekly sliding R_t estimates (Figure D.3). The daily R_t estimates using the reconstructed data considerably smoothed out the impact of weekend effects, which were prominent in the reported incidence (Figure D.3A, Figure D.4A-B). The weekly sliding R_t estimates are very similar using both datasets, with some minor discrepancies e.g., around early September and mid-December 2020 (Figure D.3B).

When comparing the absolute difference in the R_t estimates made using the reported and reconstructed data, the larger spikes coincide with periods of either low incidence or more prominent weekend effects in the reported data (Figure D.4).

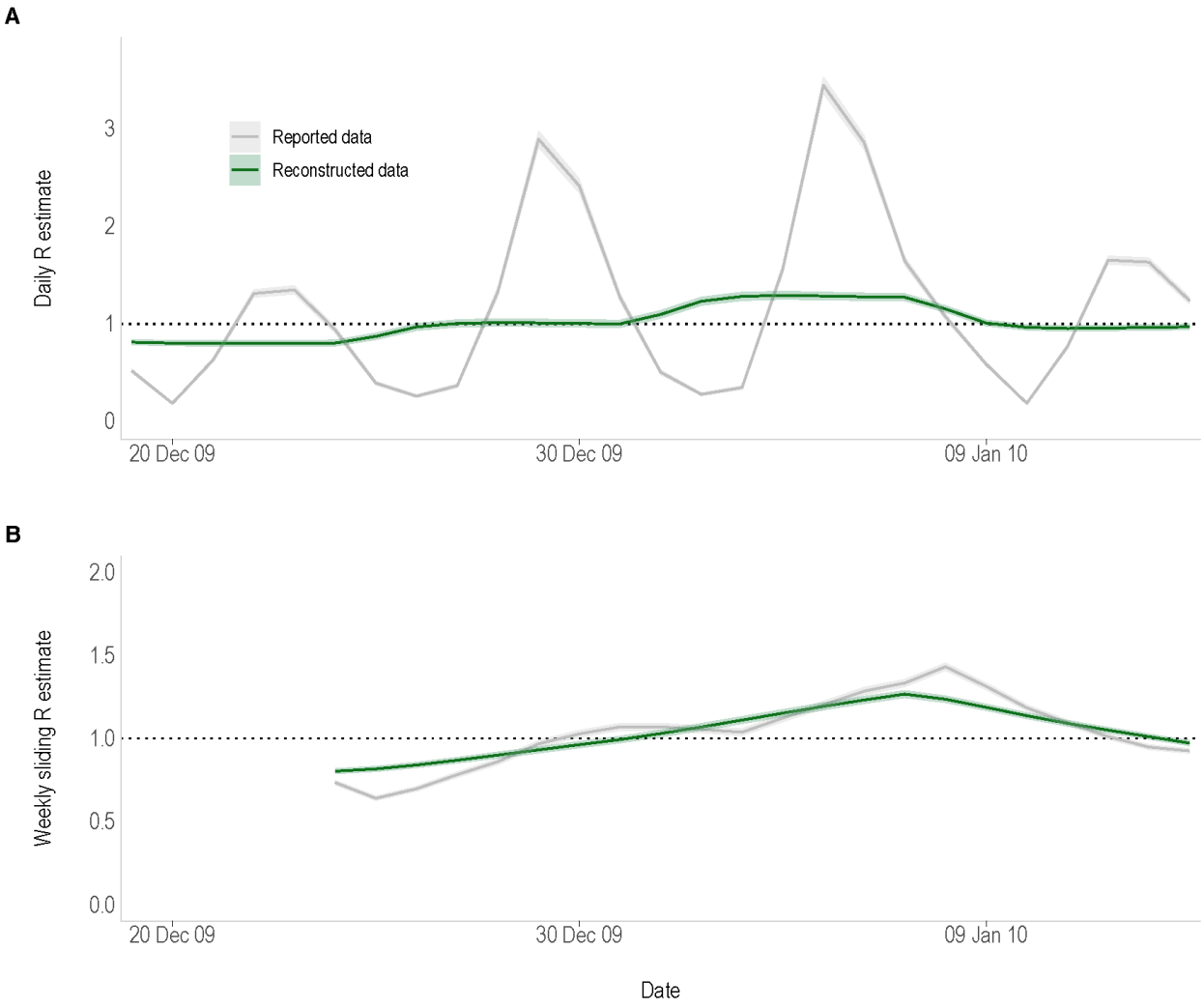


Figure D.1: **The A) daily and B) weekly sliding R_t estimates for influenza based on the reported (grey) and the reconstructed (green) daily data by date of presentation at the military clinic.** R_t estimates start on the first day of the second aggregation window (day 8 – 18th December 2009) and are plotted at the end of the time window. Shading corresponds to the 95% credible interval of the estimates.

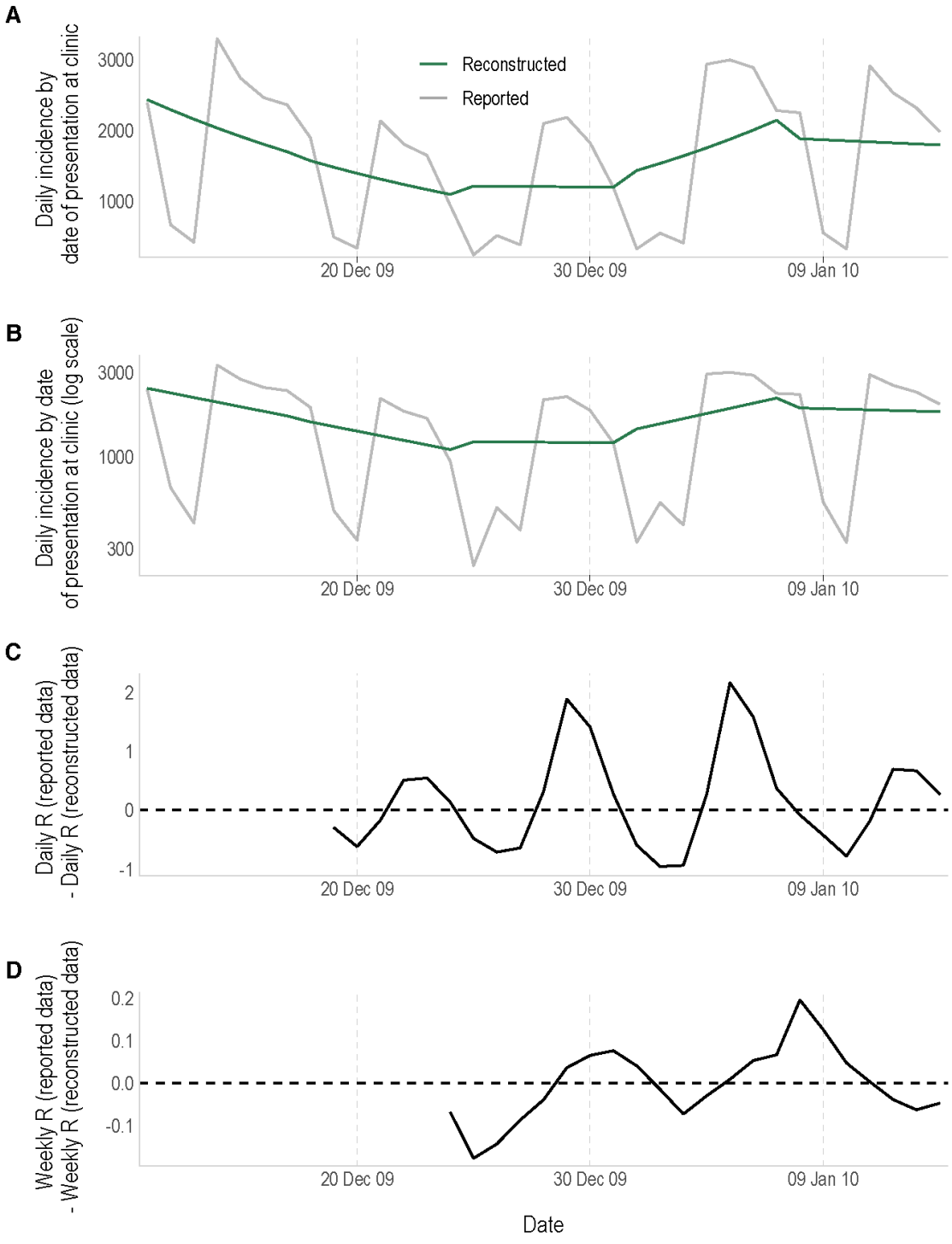


Figure D.2: **A-B) The reported (grey) and reconstructed (green) daily incidence of influenza by the date of presentation at the military clinic, either on a natural scale (A) or log scale (B). C-D) The absolute difference in the C) daily and D) weekly sliding R_t estimates made using the reported daily data and the reconstructed daily data.** Note that the y-axis scale is different in panels C and D.

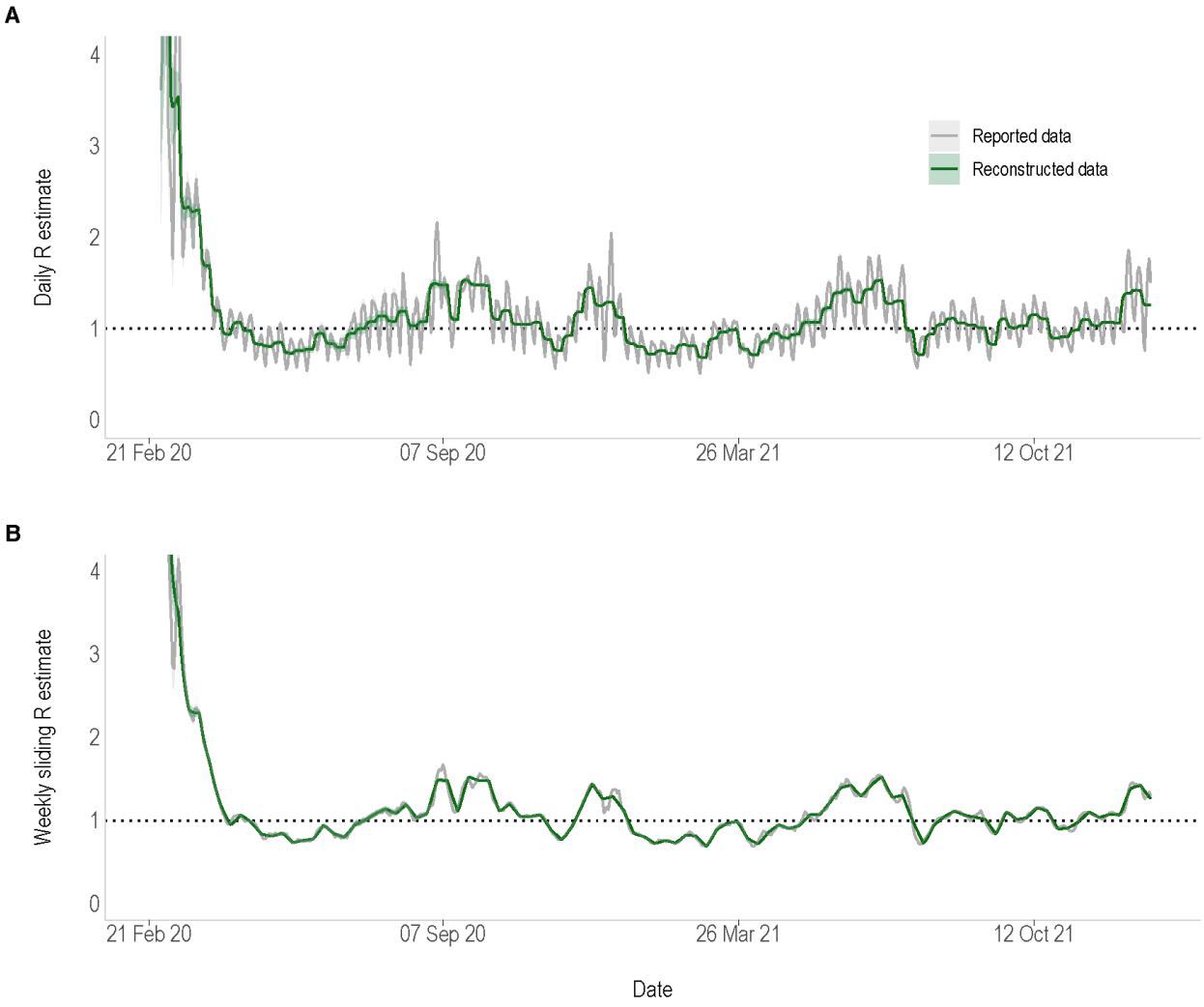


Figure D.3: **The A) daily and B) weekly sliding R_t estimates for COVID-19 cases based on the reported (grey) and the reconstructed (green) daily data by date of specimen.** R_t estimates start on the first day of the second aggregation window (day 8 – 28th February 2020) and are plotted at the end of the time window. Shading corresponds to the 95% credible interval of the estimates. The y-axis has been cropped to a maximum of 4 for clarity.

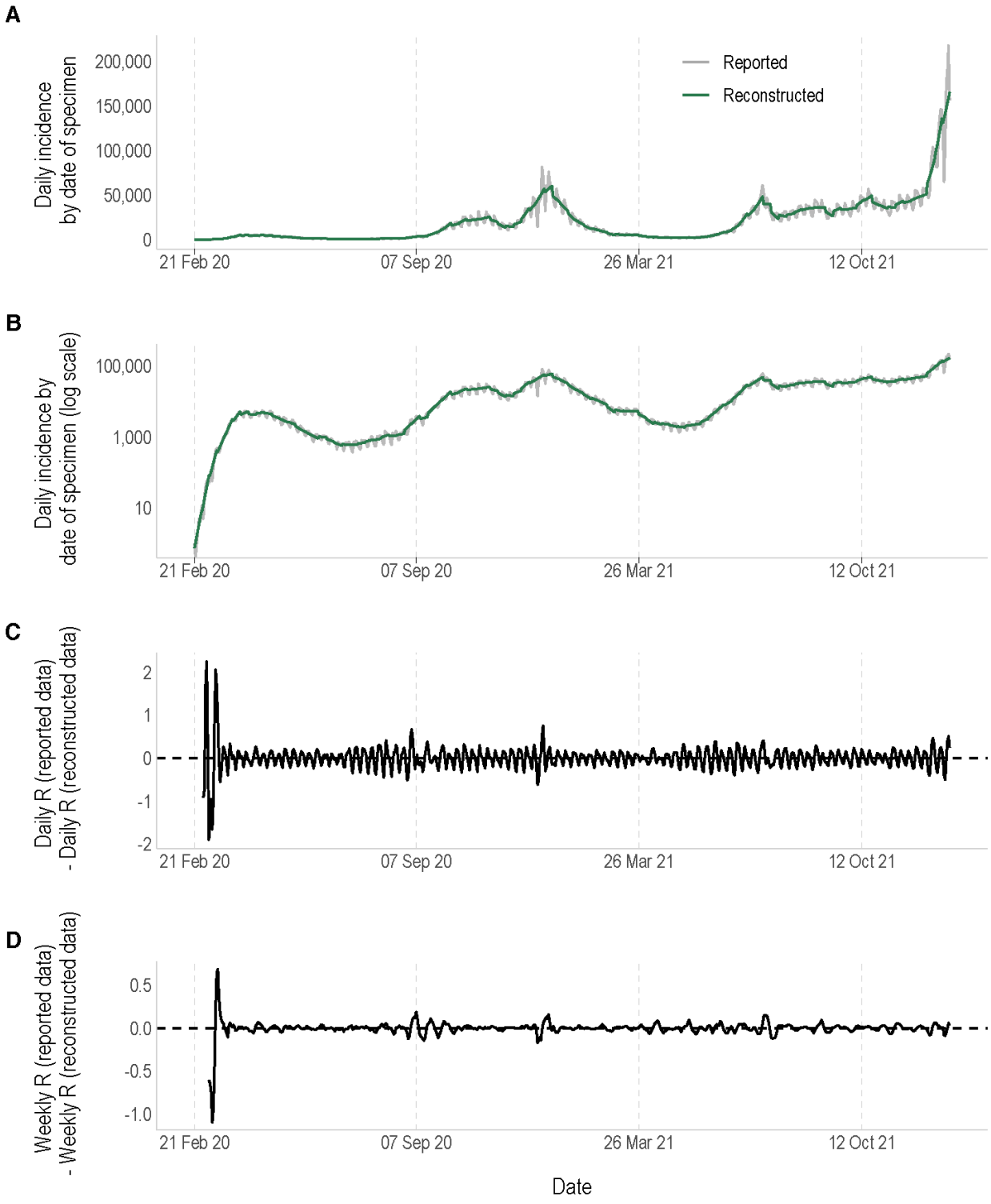


Figure D.4: **A-B) The reported (grey) and reconstructed (green) daily incidence of COVID-19 by the date of specimen, either on a natural scale (A) or log scale (B). C-D) The absolute difference in the C) daily and D) weekly sliding R_t estimates made using the reported data and the reconstructed data.** Note that the y-axis scale is different in panels C and D.

D.2.3 COVID-19 deaths

In the second case study for COVID-19, due to the lower incidence of COVID-19 deaths compared to COVID-19 cases, there is greater uncertainty in the R_t estimates (Figure D.5). There is still some indication of noise in this dataset, but the pattern appears less pronounced and more irregular than the variation caused by weekend effects as seen in the COVID-19 cases data. Unlike the data for flu and COVID-19 cases, further analysis showed that there was no pattern in reporting depending on the day of the week (Figure D.7).

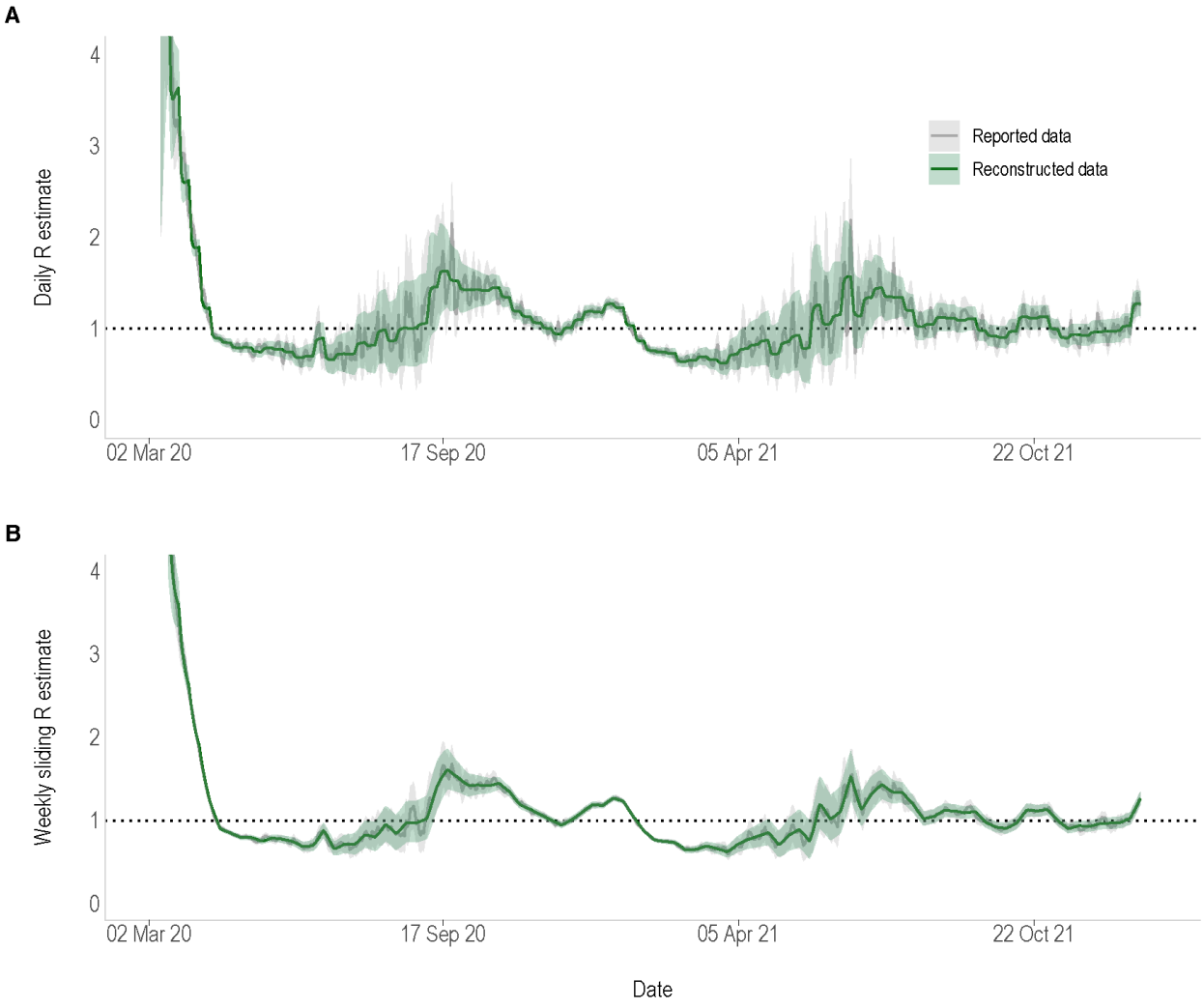


Figure D.5: **The A) daily and B) weekly sliding R_t estimates for COVID-19 deaths based on the reported daily data (grey) and the reconstructed daily data (green) by date of death within 28 days of a positive test.** R_t estimates start on the first day of the second aggregation window (day 8 – 9th March 2020) and are plotted at the end of the time window. Shading corresponds to the 95% credible interval of the estimates. The y-axis has been cropped to a maximum of 4 for clarity.

In contrast with COVID-19 case data, the greatest differences in R_t estimates obtained from death data appear to coincide solely with periods of low incidence (Figure D.6).

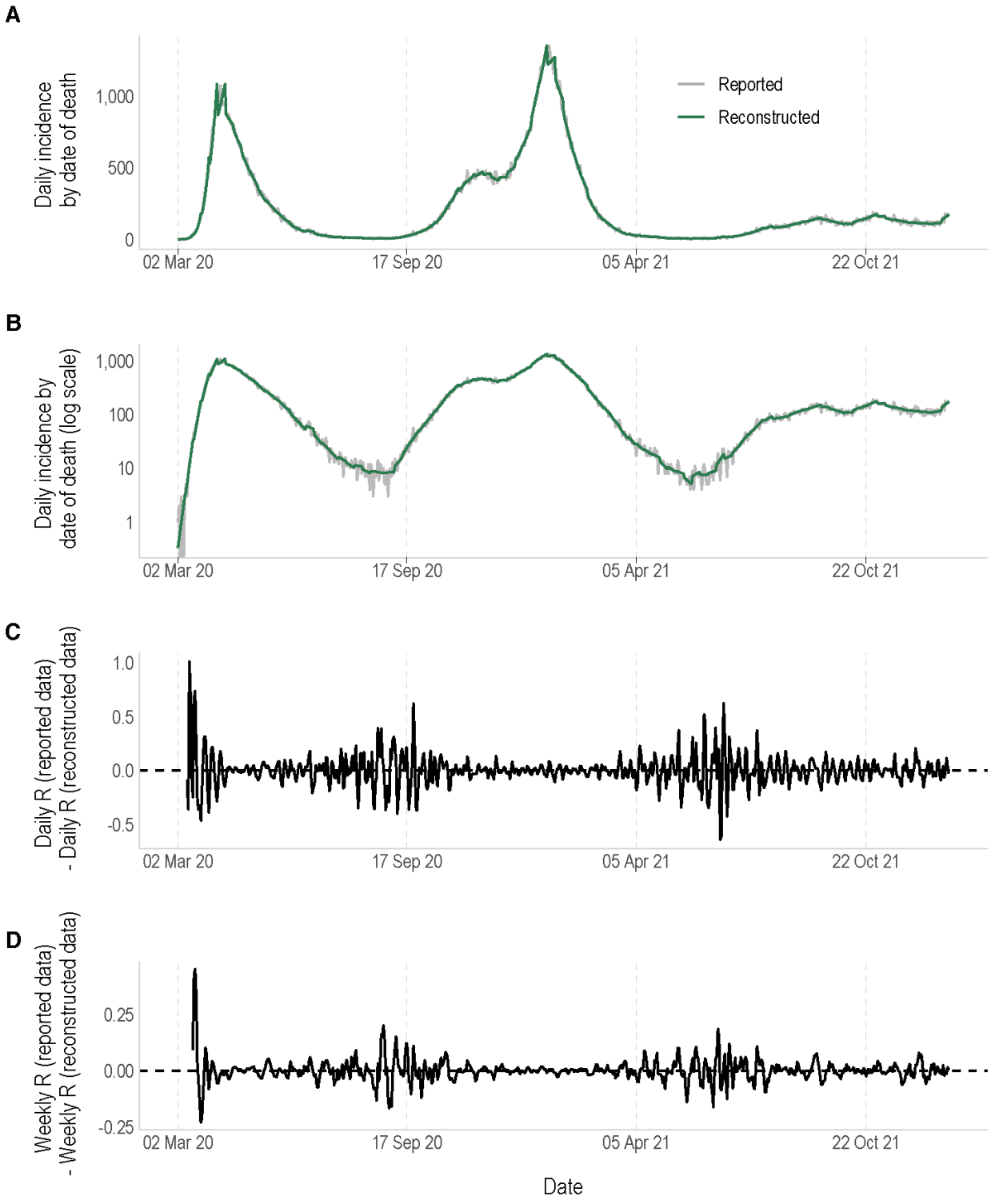


Figure D.6: **A-B) The reported (grey) and reconstructed (green) daily incidence of COVID-19 deaths within 28 days of a positive test, either on a natural scale (A) or log scale (B). C-D) The absolute difference in the C) daily and D) weekly sliding R_t estimates made using the reported data and the reconstructed data.** Note that the y-axis scale is different in panels C and D.

D.2.4 Reported incidence patterns by weekday

There is a clear pattern in the reporting of influenza and COVID-19 cases, with generally higher incidence reported on Mondays, which then declines throughout the week, with the lowest incidence

on the weekends (Figure D.7A-B). Reported incidence appears lower on Fridays for influenza, however this dataset encompasses Christmas Day 2009 and New Year's Day 2010, which both fell on Friday. On the other hand, COVID-19 deaths did not appear to show any pattern in reporting, with no clear deviations from the weekly mean incidence regardless of the weekday (Figure D.7C).

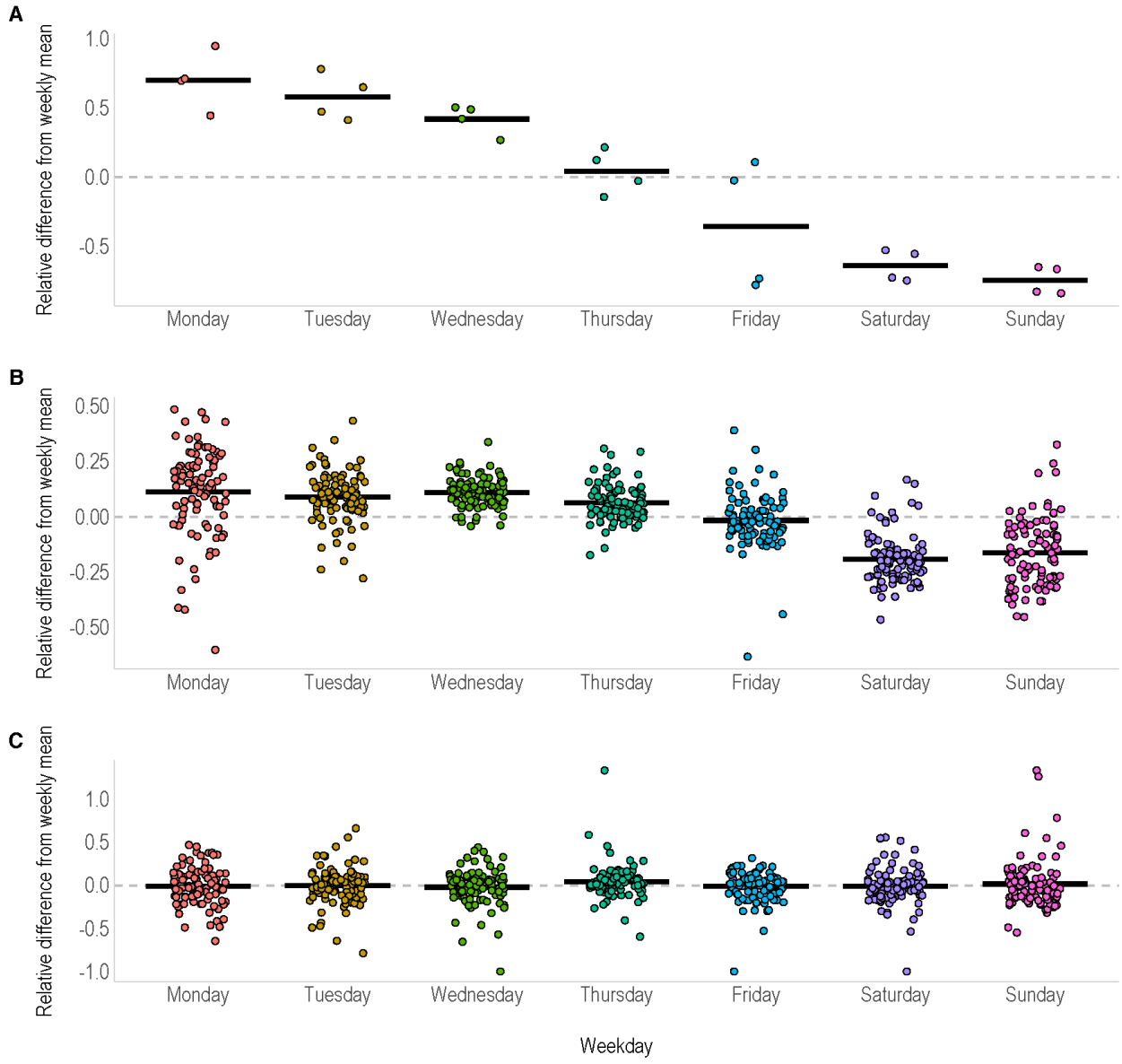


Figure D.7: **The relative difference between the reported incidence for A) influenza, B) COVID-19 cases and C) COVID-19 deaths, on each weekday compared to the mean incidence for that week.** Only data for full calendar weeks are included. Black lines represent the mean relative difference and the dashed grey line indicates zero (corresponding to no difference between the reported incidence on that day compared to the weekly mean).

D.2.5 Classification tables

We directly compared the classification of the epidemic as increasing (95% credible interval of the R_t estimate is above 1), uncertain (95% credible interval encompasses 1), and declining (95% credible interval is below 1), depending on whether the R_t estimates were made using the reported or recon-

structed daily data (Figure S8). The overall agreement in the classification of daily R_t estimates for influenza, COVID-19 cases, and COVID-19 deaths, was 44.4%, 74.4%, and 85.8% respectively. The overall agreement in the classification of weekly sliding R_t estimates was higher for each, with 81.8%, 94.9% and 93.3% agreement respectively. There is a greater correlation between the classification of estimates made using COVID-19 death data, reflected by the dark blue diagonal line across the grid squares. This is likely due to the reported deaths being less affected by weekend effects (see Appendix D.2.4), and therefore the reported and reconstructed data (and the R_t estimates made from them) are more similar (Figure D.8E-F).

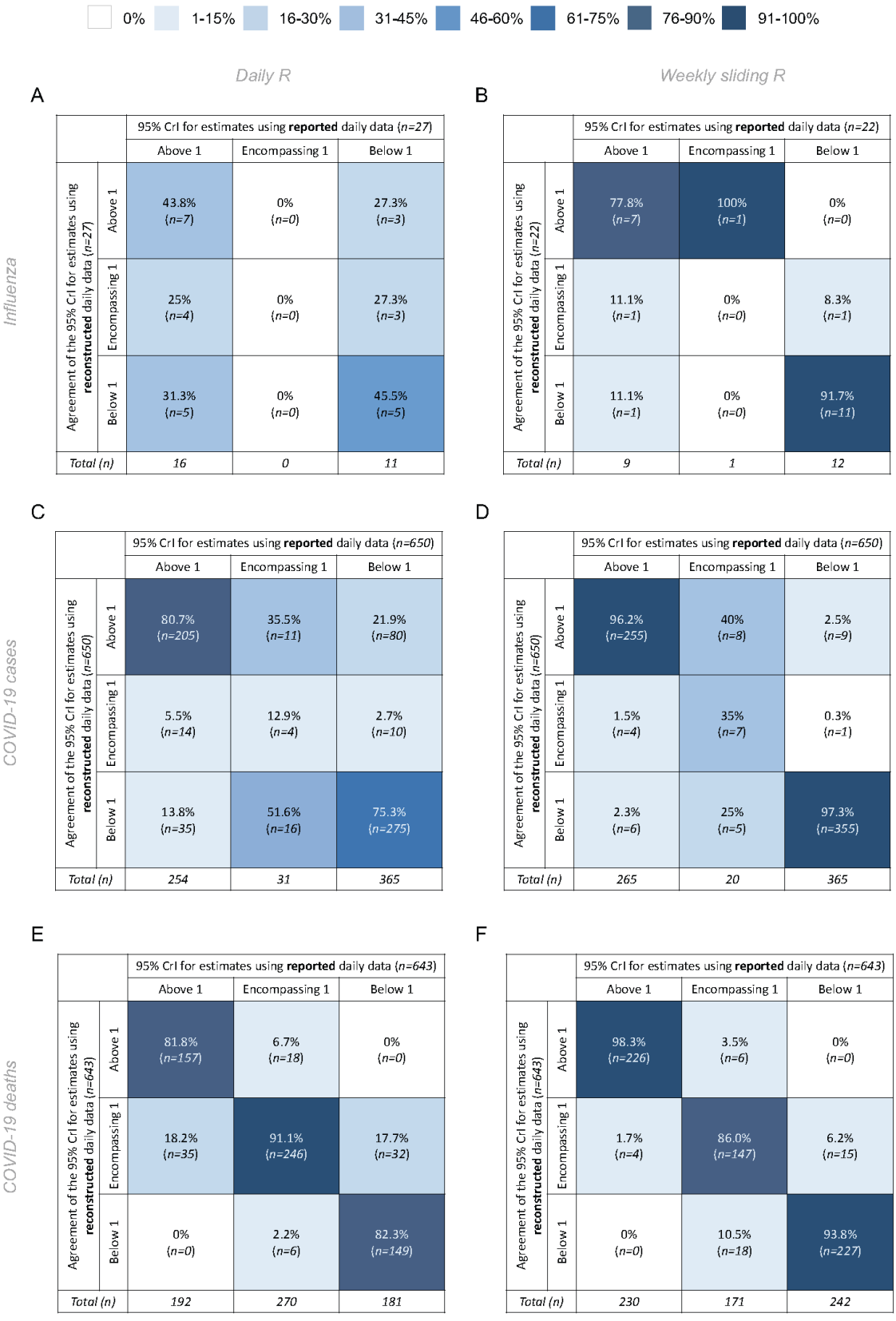


Figure D.8: **Comparison of the percentage agreement in the classification of R_t estimates as above 1 (lower bound of the 95% CrI is above 1), encompassing 1, or below 1 (upper bound of the 95% CrI is below 1), when using reported and reconstructed incidence data. Daily and weekly sliding R_t estimates are compared for influenza cases (A-B), COVID-19 cases (C-D), and COVID-19 deaths (E-F).** Due to low incidence, the first 30 days of data is excluded for both COVID-19 datasets. Darker shades of blue correspond to greater percentage agreement, with a strong correlation represented by a dark diagonal line across the grid squares from top left to bottom right.

D.2.6 Computational time

All real data scenarios took less than 3 seconds to run on MacOS (2 GHz Quad-Core Intel Core i5) 16GB RAM (Table D.1). The influenza cases scenario, with >57,000 cases, took 2 seconds to run. The COVID-19 cases and deaths scenarios, which estimated R_t over 97 and 96 weeks of incidence data, took 3 seconds to run.

Table D.1: **Time taken to estimate R_t in each of the real data scenarios.**

Real data scenario	Time period analysed	Total incidence	Estimation time
Influenza cases	5 weeks	57,351	2s
COVID-19 cases	97 weeks	13,139,522	3s
COVID-19 deaths	96 weeks	149,557	3s

D.3 Simulation study

The simulation study involved assessing the performance of the method in multiple epidemic contexts. These included scenarios where R_t remains constant over time, or where R_t varies over time, with either a sudden stepwise change or a gradual change. For each scenario, 100 epidemic trajectories were stochastically generated using the R package projections [573]. Each epidemic was seeded with 7 days of 10 daily cases and then simulated over 70 days (10 weeks) using a renewal equation based branching process model, where incidence follows a Poisson distribution (equation D.1).

$$I_t \sim Pois\left(R_t \sum_{s=1}^{t-1} I_s g_{t-s}\right) \tag{D.1}$$

The values of R_t estimated from the simulated incidence data were evaluated in terms of their bias, uncertainty and 95% coverage (Table D.2).

Table D.2: **Criteria used to assess the performance of our method in the simulation study.**

Criterion	Definition
Bias	Absolute difference between the mean R_t estimate across the 100 simulations and the true value of R_t that the simulated incidence is based on.
Uncertainty	The mean width (across the 100 simulations) of the 95% credible interval for R_t estimates.
95% coverage	The proportion of R_t estimates across the 100 simulations with the true value of R_t within the 95% credible interval.

D.3.1 Constant R_t

For a scenario where R_t remains constant over time, four values of R_t were considered: 1, 1.25, 1.5 and 1.75. The R_t estimates all recovered the true value that the simulations were based on, with very little difference between the estimates made using the true daily and reconstructed daily data (Figure D.9A-D). There was no evidence of bias in the estimates (Figure D.9E-H), and as expected, the uncertainty declined over time as the number of cases rose (Figure D.9I-L). When R_t was 1, as the seed for the simulated incidence trajectories is relatively low, some of the stochastic trajectories "died out" with case numbers falling to zero, which makes uncertainty appear to increase. The 95% coverage was consistently high (Figure D.9M-P) and on average the true value of R_t was encompassed by the 95% credible interval 96-97% of the time in each scenario. However, the wavy pattern in these plots is likely to correspond to discontinuities in the reconstructed incidence data (Appendix D.3.2, Figure D.10).

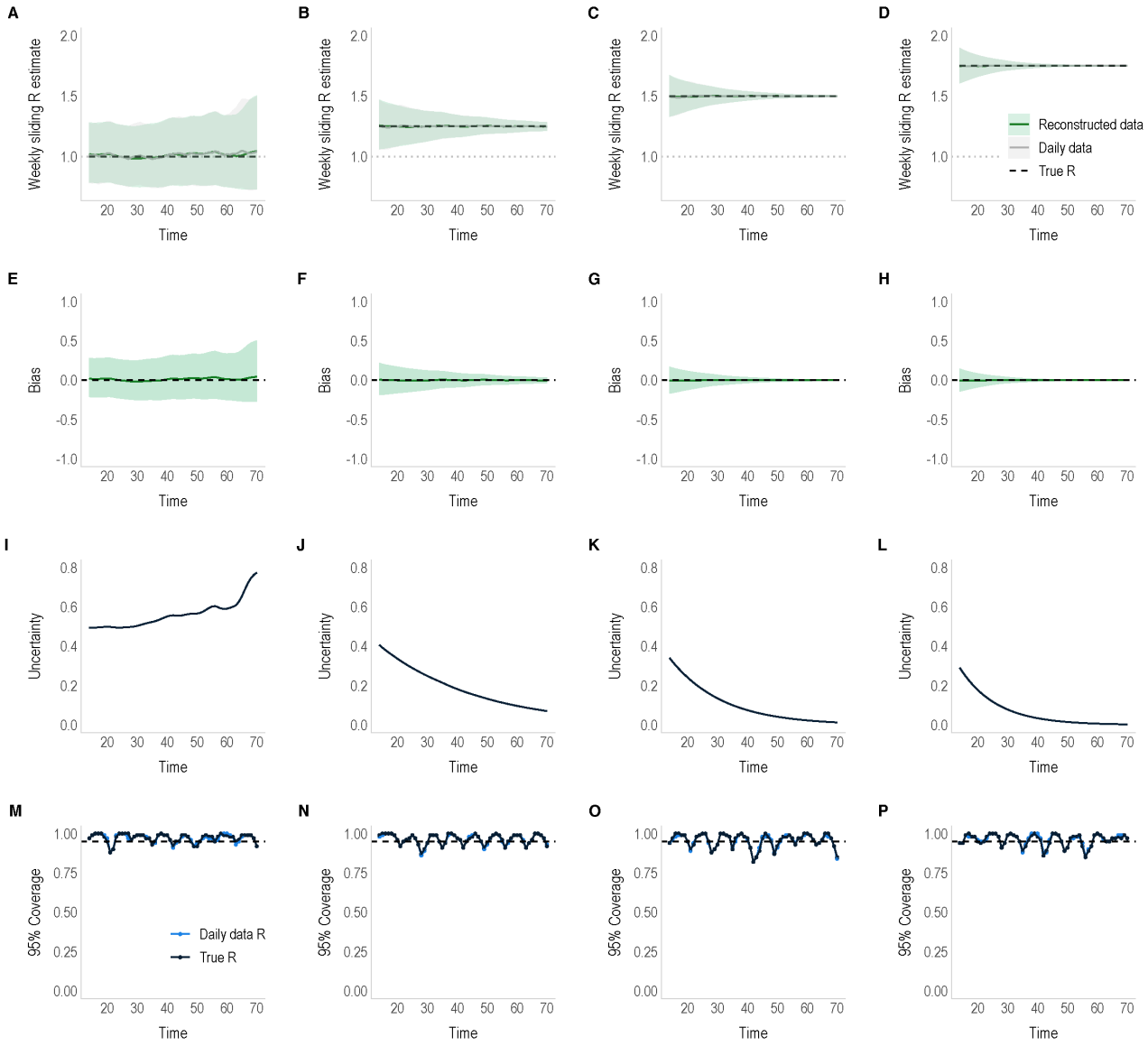


Figure D.9: **Performance of the method when estimating a constant R_t for 100 simulated epidemics.** A-D) The mean weekly sliding R_t estimates using daily data (grey) and reconstructed data (green). The black dashed line represents the true value of R_t that the 100 simulations are based on, which is either A) 1, B) 1.25, C) 1.5, or D) 1.75. The estimates and their 95% credible intervals (shaded areas) are very similar and therefore overlap. E-H) The bias (or absolute difference) in the mean R_t estimated from reconstructed data compared to the true value of R_t. The shaded area corresponds to the 95% quantiles of the bias across the 100 simulations. I-L) The uncertainty (mean width of the 95% credible interval) in the R_t values estimated from reconstructed data. M-P) The proportion of the 100 R_t estimates obtained from reconstructed data where the 95% credible interval encompasses either the true value of R_t (black) or the value of R_t that would have been estimated from daily data (blue). Here, the 95% coverage is similar for both the true value of R_t and the daily data estimate, and therefore they overlap.

D.3.2 Discontinuities in the reconstructed incidence data

As part of the process of reconstructing daily incidence data from aggregated data, the reconstructed incidence is slightly adjusted using a constant (k_w) to ensure that if you were to re-aggregate it, it would still match the original aggregated data used as the input. This can result in discontinuities in the borders between time periods that the data has been aggregated over, which means that the reconstructed incidence is not completely smooth, even when there is no noise in the data (Figure D.10).

When R_t is estimated using sliding windows, these may align perfectly with the time windows that data were aggregated over or they may encompass the border between aggregations, and therefore the discontinuity in the data. As shown in Figure D.9M-P, this can lead to a wavy pattern in the 95% coverage, where the sliding window moves across the incidence data and can be more/less affected by the discontinuities depending on where they fall. This is important to be aware of, but as seen in the real data scenarios (Appendix D.2), reported daily data is often dramatically affected by intra-weekly variations in reporting, such as weekend effects. The reconstructed data, even with these discontinuities, is much smoother than the daily data that is typically reported. Nevertheless, we suggest that the sliding time window used to estimate R_t should be equal to or longer than the length of the aggregation window, to limit the effect of the discontinuities on the estimates.

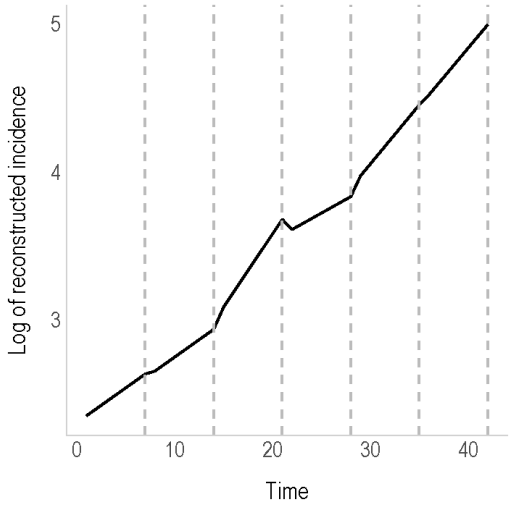


Figure D.10: **An example of reconstructed daily incidence data (on the log scale) with clear discontinuities on the borders between time periods where the data had been aggregated (grey dashed lines).**

D.3.3 Time-varying R_t: sudden change

In the first scenario for time-varying transmissibility, we considered a stepwise decrease or increase in the value of R_t. R_t may suddenly decrease or increase following the rapid implementation or relaxation of stringent control measures, for instance, a strict mass 'lockdown' event.

We considered four decreasing scenarios, where on day 35 of the simulated outbreak, R_t falls from: 1.25 to 0.75, 1.25 to 1, 1.5 to 1.25, or 1.75 to 1.5. The true value of R_t was successfully recovered in all scenarios, except for a week-long delay following the step change (Figure D.11A-D), which corresponds to dips in the bias plots (Figure D.11E-H). This delay is due to the weekly sliding window used for R_t estimation, which would encompass incidence data before and after the step change.

As expected, uncertainty declines as case numbers rise when $R_t > 1$ (Figure D.11I-L). The decrease in R_t from 1.25 to 0.75 replicates a situation where control measures successfully bring R_t below 1,

resulting in falling case numbers and increased uncertainty after day 35 (Figure D.11I). Similarly, when R_t falls from 1.25 to 1, case numbers become stable, leading to a plateau in the uncertainty (Figure D.11J).

Importantly, despite not recovering the true value of R_t in the week following the step change, a high proportion of the estimates recovered the value of R_t that would have been estimated from the reported daily data (Figure D.11M-P). There is a slightly larger drop in the 95% coverage when R_t falls from 1.75 to 1.5 (Figure D.11P), however, this is due to higher case numbers in this scenario resulting in very narrow 95% credible intervals. Overall, the 95% credible interval encompassed the true value of R_t 85-90% of the time, which rose to 95-97% if the week following the step change was excluded.

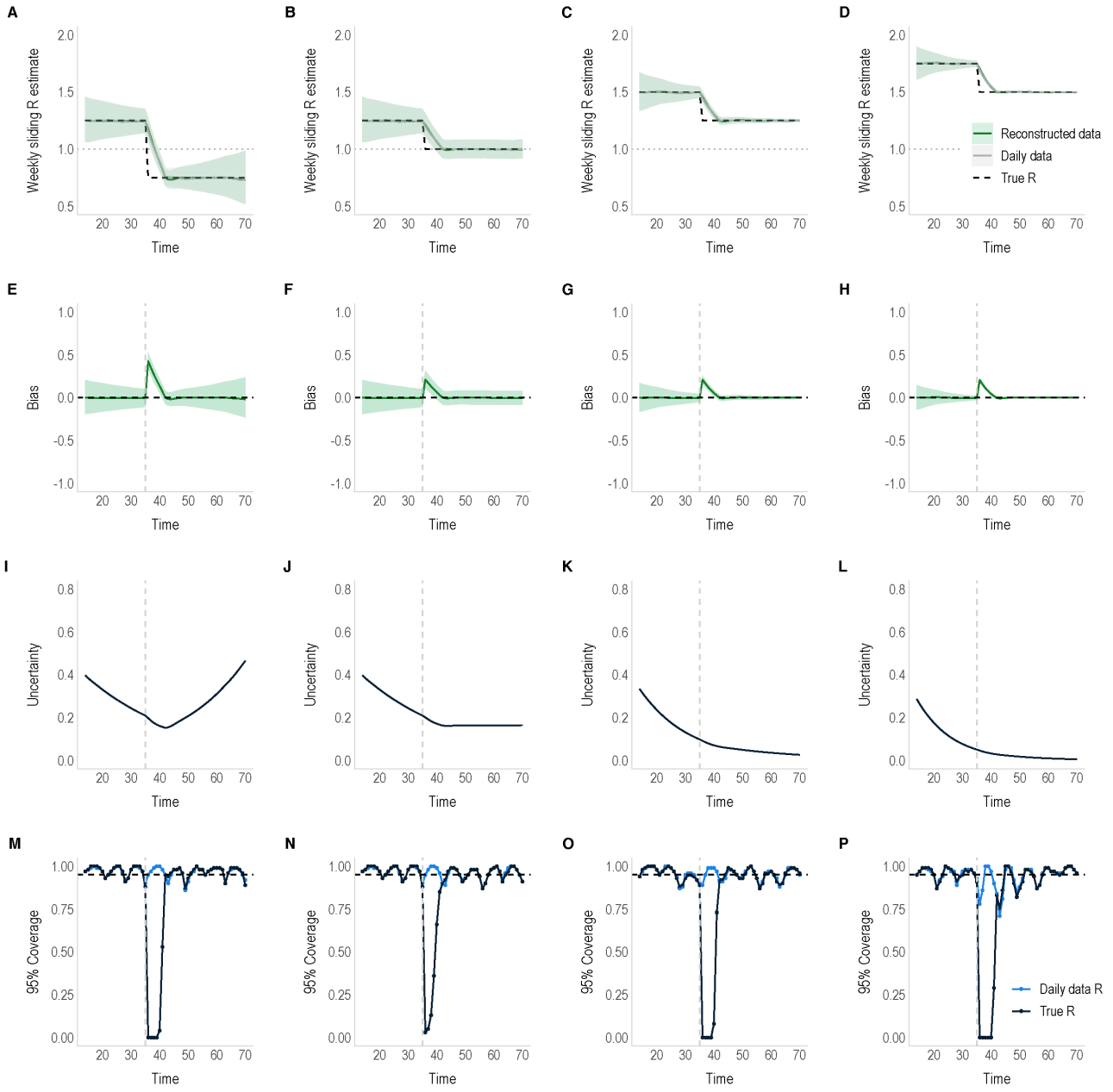


Figure D.11: **Performance of the method when estimating a time varying R_t for 100 simulated epidemics. In this scenario, R_t suddenly decreases on day 35 (grey dashed line) and remains constant before and after the step change.** A-D) The mean weekly sliding R_t estimates using daily data (grey) and reconstructed data (green). The black dashed line represents the true value of R_t that the 100 simulations are based on, which is either a decrease from A) 1.25 to 0.75, B) 1.25 to 1, C) 1.5 to 1.25, or D) 1.75 to 1.5. The grey dotted line represents the threshold of $R_t = 1$. The estimates (plotted at the end of each time window) and their 95% credible intervals (shaded area) are very similar and therefore overlap. E-H) The bias (or absolute difference) in the mean R_t estimated from reconstructed data compared to the true value of R_t. The shaded area corresponds to the 95% quantiles of the bias across the 100 simulations. I-L) The uncertainty (mean width of the 95% credible interval) in the R_t values estimated from reconstructed data. M-P) The proportion of the 100 R_t estimates made using reconstructed data where the 95% credible interval encompasses either the true value of R_t (black) or the value of R_t that would have been estimated from daily data (blue).

A sudden stepwise increase in R_t is perhaps more difficult to explain in a real-world context, but it is also shown here for the sake of completion. Three scenarios for a stepwise increase in R_t were considered: 1 to 1.25, 1.25 to 1.5, and 1.5 to 1.75. Similarly, the true value of R_t was well recovered

in each scenario, but there was a week-long delay following the step change (Figure D.12A-F). The uncertainty was stable when R_t is 1 and falls as case numbers rise (Figure D.12G-I). Finally, despite the drop in 95% coverage for the true value of R_t, the estimates from reconstructed data recovered what would have been estimated from simulated daily data (Figure D.12J-L). Across these scenarios, the 95% credible interval encompassed the true value of R_t 87-96% of the time, rising to 95-97% if the week following the step change is excluded.

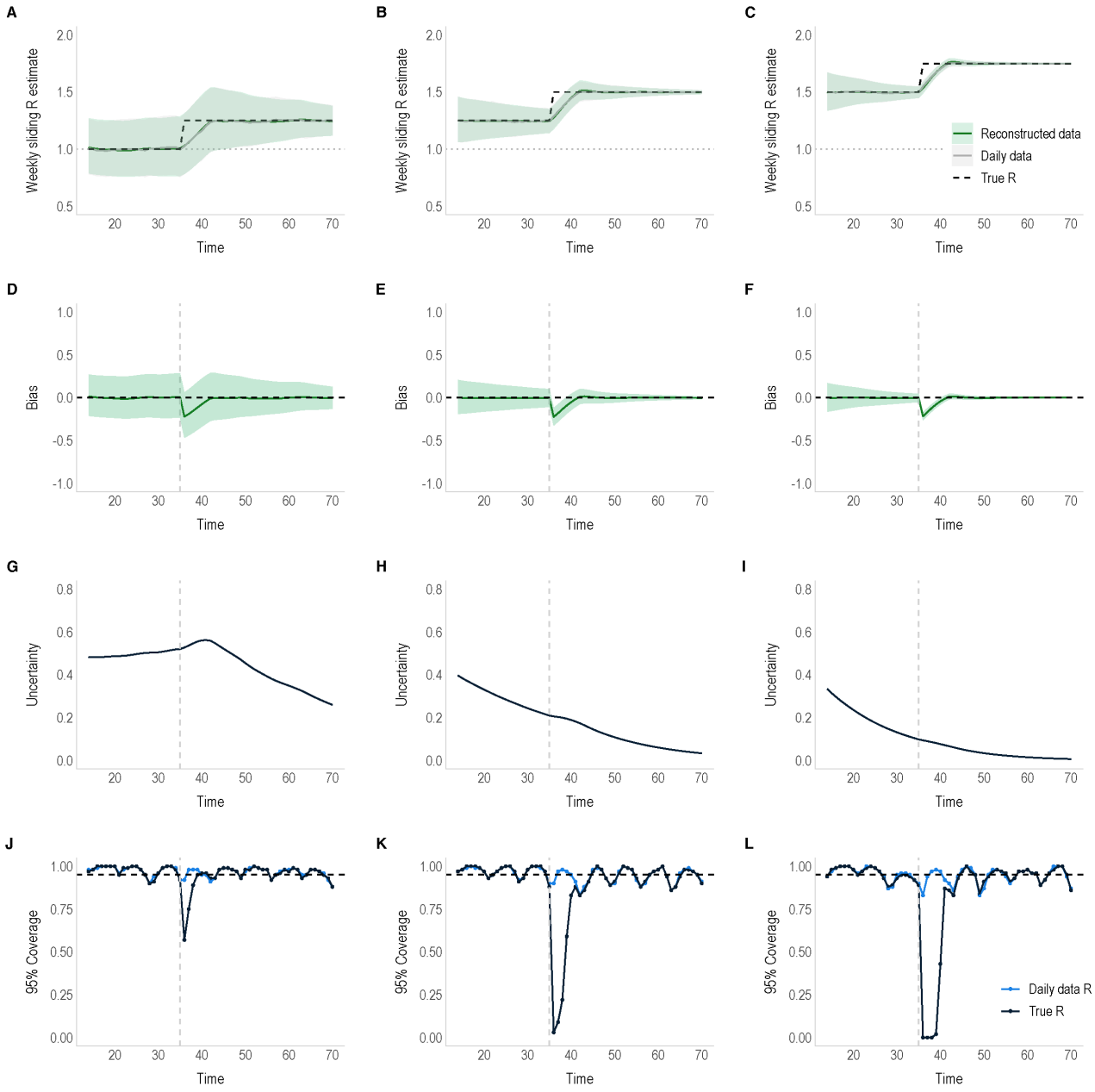


Figure D.12: **Performance of the method when estimating a time varying R_t for 100 simulated epidemics. In this scenario, R_t suddenly increases on day 35 (grey dashed line) and remains constant before and after the step change.** A-C) The mean weekly sliding R_t estimates using simulated daily data (grey) and reconstructed data (green). The black dashed line represents the true value of R_t that the 100 simulations are based on, which is either an increase from A) 1 to 1.25, B) 1.25 to 1.5, or C) 1.5 to 1.75. The grey dotted line represents the threshold of $R_t = 1$. The estimates (plotted at the end of each time window) and their 95% credible intervals (shaded area) are very similar and therefore overlap. D-F) The bias (or absolute difference) in the mean R_t estimated from reconstructed data compared to the true value of R_t. The shaded area corresponds to the 95% quantiles of the bias across the 100 simulations. G-I) The uncertainty (mean width of the 95% credible interval) in the R_t values estimated from reconstructed data. J-L) The proportion of the 100 R_t estimates made using reconstructed data where the 95% credible interval encompasses either the true value of R_t (black) or the value of R_t that would have been estimated from daily data (blue).

D.3.4 Time-varying R_t: gradual change

Changes in transmissibility are often more gradual over time. For example, R_t could slowly decrease as a population becomes more aware of a circulating pathogen and modifies their behaviour accordingly, or

R_t could increase in response to the gradual easing of restrictions, such as social distancing measures, or a gradual decline in compliance to such restrictions. In the following scenarios, we considered a gradual change in R_t that occurred over the course of 30 days, starting on day 20 and ending on day 50. Four scenarios were considered for a gradually decreasing R_t, where R_t falls from: 1.25 to 0.75, 1.25 to 1, 1.5 to 1.25, and 1.75 to 1.5 (Figure D.13). Three scenarios were considered for a gradually increasing R_t, where R_t increases from: 1 to 1.25, 1.25 to 1.5, and 1.5 to 1.75.

As above, during the period of gradual change, the estimates were affected by the lag due to the weekly sliding window for R_t estimation (Figure D.13A-D & Figure D.14A-C). This means that R_t was slightly overestimated during the 30 days that R_t was gradually decreasing and slightly underestimated when R_t was gradually increasing (Figure D.13E-H & Figure D.14D-F). The uncertainty declined as case numbers rose (Figures D.13I-L & D.14G-I), increased when R_t fell below 1 (Figure D.13I) and plateaued when $R_t = 1$ (Figures D.13J & D.14G). The lag in the estimates caused some notable drops in the 95% coverage (Figures D.13M-P & D.14J-L). This is particularly prominent when incidence is high and credible intervals are small, such as in the gradual decrease in R_t from 1.75 to 1.5 scenario, where the true value of R_t was encompassed by the 95% credible interval 61% of the time (Figure D.13P). For all other gradually decreasing scenarios and the gradually increasing scenarios, true R_t was recovered between 80-94% and 77-96% of the time respectively. It is important to note, however, that even when substantial dips in coverage occurred, the bias remained small, meaning there was very little difference between the estimate and the true value of R_t despite not being exactly the same (Figure D.13E-H).

D.3.5 Influence of R_t plotting time relative to the time window used for estimation

By default, the EpiEstim R package estimates R_t over weekly sliding time windows ending at time t, at which point R_t is plotted. An alternative to this, is to plot R_t in the middle of the time window, centred around t. This would mean that R_t estimates cannot be made in real-time, i.e., if a weekly window was used you can only estimate R_t at t + 3.5. However, the advantage is that retrospective R_t estimates would be less influenced by the lag corresponding to the length of the time window.

To demonstrate this, we reconsider the scenarios with a gradual and sudden change in R_t. When gradual changes in R_t are plotted in the middle of the time window (t − 3.5), the true value of R_t is well recovered and the 95% coverage is considerably improved in comparison to estimates plotted at the end of the time window (Figure D.15). Overall, the 95% credible intervals of the gradual R_t estimates plotted at the mid-point of the time window encompassed the true value of R_t 94-97% and 88-96% of the time for the increasing and decreasing scenarios respectively (compared to 77-96% and 61-94% when R_t is plotted at the end of the time window). In the gradual change from 1.5 to 1.75 and 1.75 to 1.5 scenarios, there is still a dip in coverage which occurs when case numbers are so high that the 95% credible intervals become extremely narrow (Figure D.15C & J). Therefore, when the sliding window encompasses incidence data before R_t plateaus, it means R_t is slightly under- or overestimated

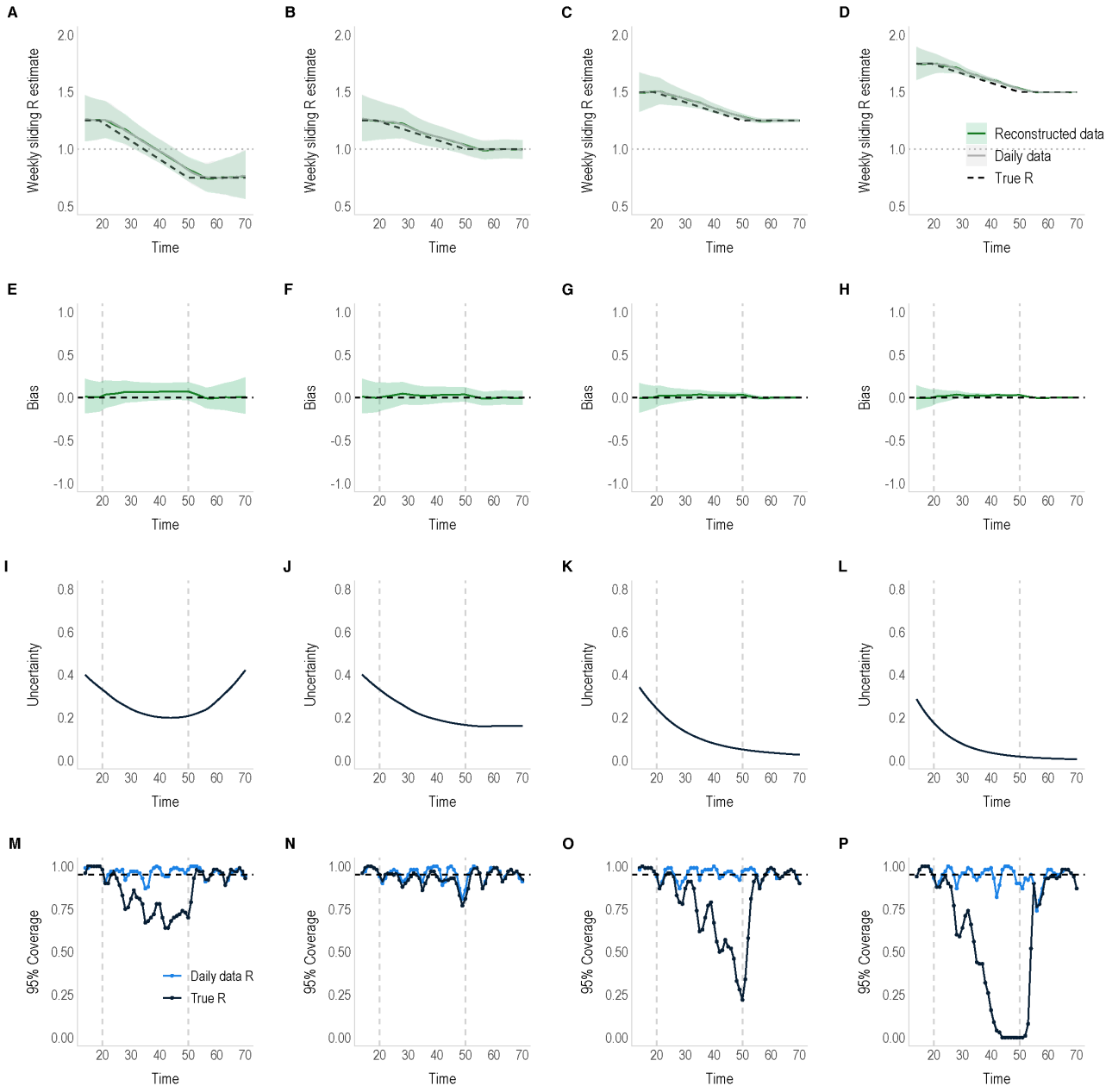


Figure D.13: **Performance of the method when estimating a time varying R_t for 100 simulated epidemics. In this scenario, R_t gradually decreases over the course of 30 days (day 20 to day 50, shown with grey dashed lines), and remains constant before and after the change.** A-D) The mean weekly sliding R_t estimates using daily data (grey) and weekly data (green). The black dashed line represents the true value of R_t that the 100 simulations are based on, which is either a decline from A) 1.25 to 0.75, B) 1.25 to 1, C) 1.5 to 1.25, or D) 1.75 to 1.5. The grey dotted line represents the threshold of $R_t = 1$. The estimates (plotted at the end of each time window) and their 95% credible intervals (shaded area) are very similar and therefore overlap. E-H) The bias (or absolute difference) in the mean R_t estimated from weekly data compared to the true value of R_t. The shaded area corresponds to the 95% quantiles of the bias across the 100 simulations. I-L) The uncertainty (mean width of the 95% credible interval) in the R_t values estimated from weekly data. M-P) The proportion of the 100 R_t estimates made using weekly data where the 95% credible interval encompasses either the true value of R_t (black) or the value of R_t that would have been estimated from daily data (blue).

for increasing and decreasing R_t respectively. However, it is evident that the mean estimates are negligibly different to the true value, despite not being exactly the same.

In the stepwise change in R_t scenarios, there is also some improvement in 95% coverage (Fig-

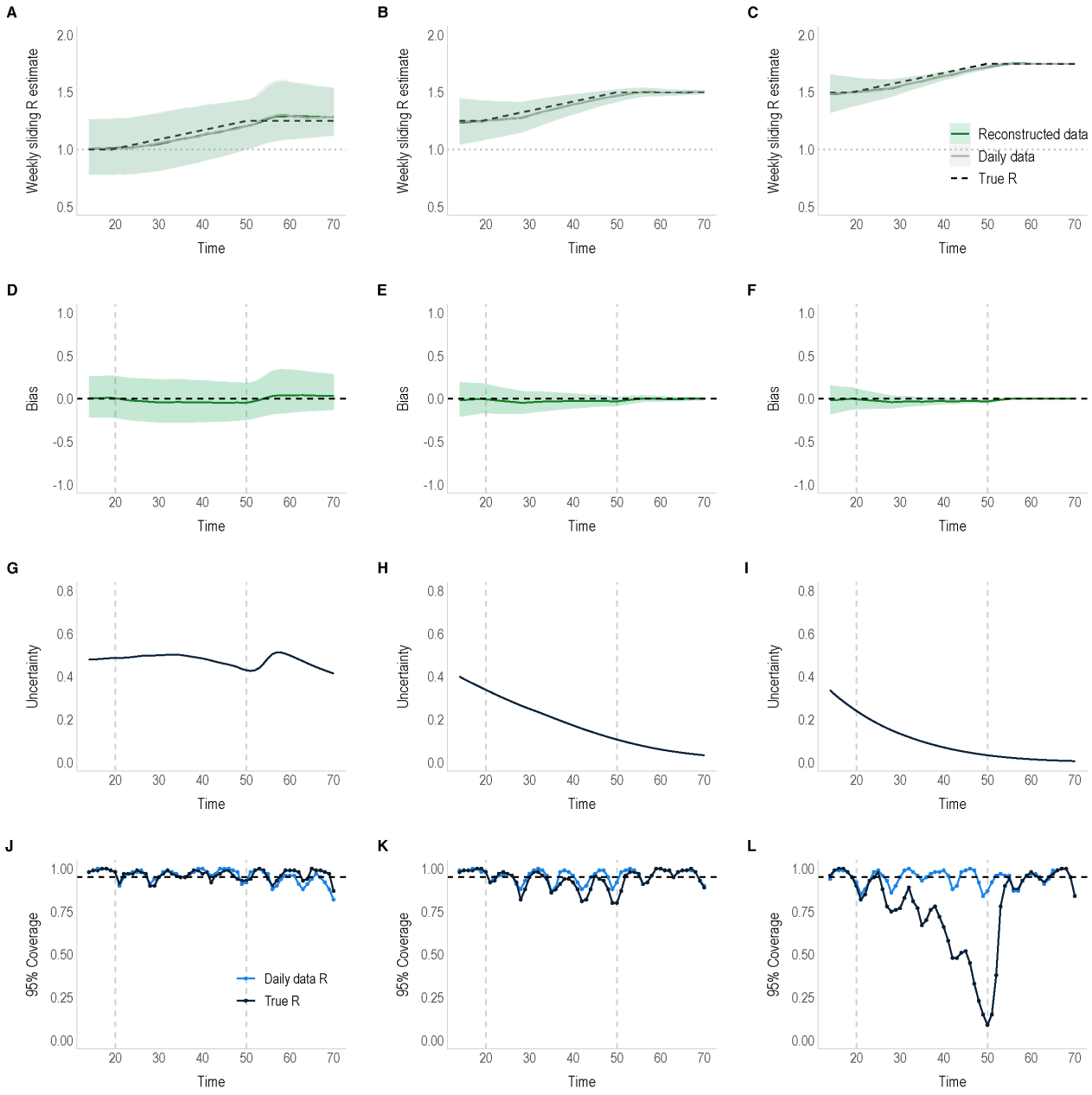


Figure D.14: **Performance of the method when estimating a time varying R_t for 100 simulated epidemics. In this scenario, R_t gradually increases over the course of 30 days (day 20 to day 50, shown with grey dashed lines), and remains constant before and after the change.** A-C) The mean weekly sliding R_t estimates using daily data (grey) and weekly data (green). The black dashed line represents the true value of R_t that the 100 simulations are based on, which is either an increase from A) 1 to 1.25, B) 1.25 to 1.5, or C) 1.5 to 1.75. The grey dotted line represents the threshold of $R_t = 1$. The estimates (plotted in the middle of each time window) and their 95% credible intervals (shaded area) are very similar and therefore overlap. D-F) The bias (or absolute difference) in the mean R_t estimated from weekly data compared to the true value of R_t. The shaded area corresponds to the 95% quantiles of the bias across the 100 simulations. G-I) The uncertainty (mean width of the 95% credible interval) in the R_t values estimated from weekly data. J-L) The proportion of the 100 R_t estimates made using weekly data where the 95% credible interval encompasses either the true value of R_t (black) or the value of R_t that would have been estimated from daily data (blue).

ure D.16). There is still a lag corresponding to the length of the time window, but now the change in R_t is detected earlier and the dip in coverage is centred around the time change on day 35. As above, the dips in coverage are larger when case numbers are very high, and the 95% credible intervals are small. The 95% credible interval of estimates plotted at the mid-point of the time window encompassed

the true value of R_t 88-96% and 85-93% of the time for stepwise increasing and decreasing scenarios respectively, which is a modest improvement compared to 87-96% and 85-90% when R_t is plotted at the end of the time window.

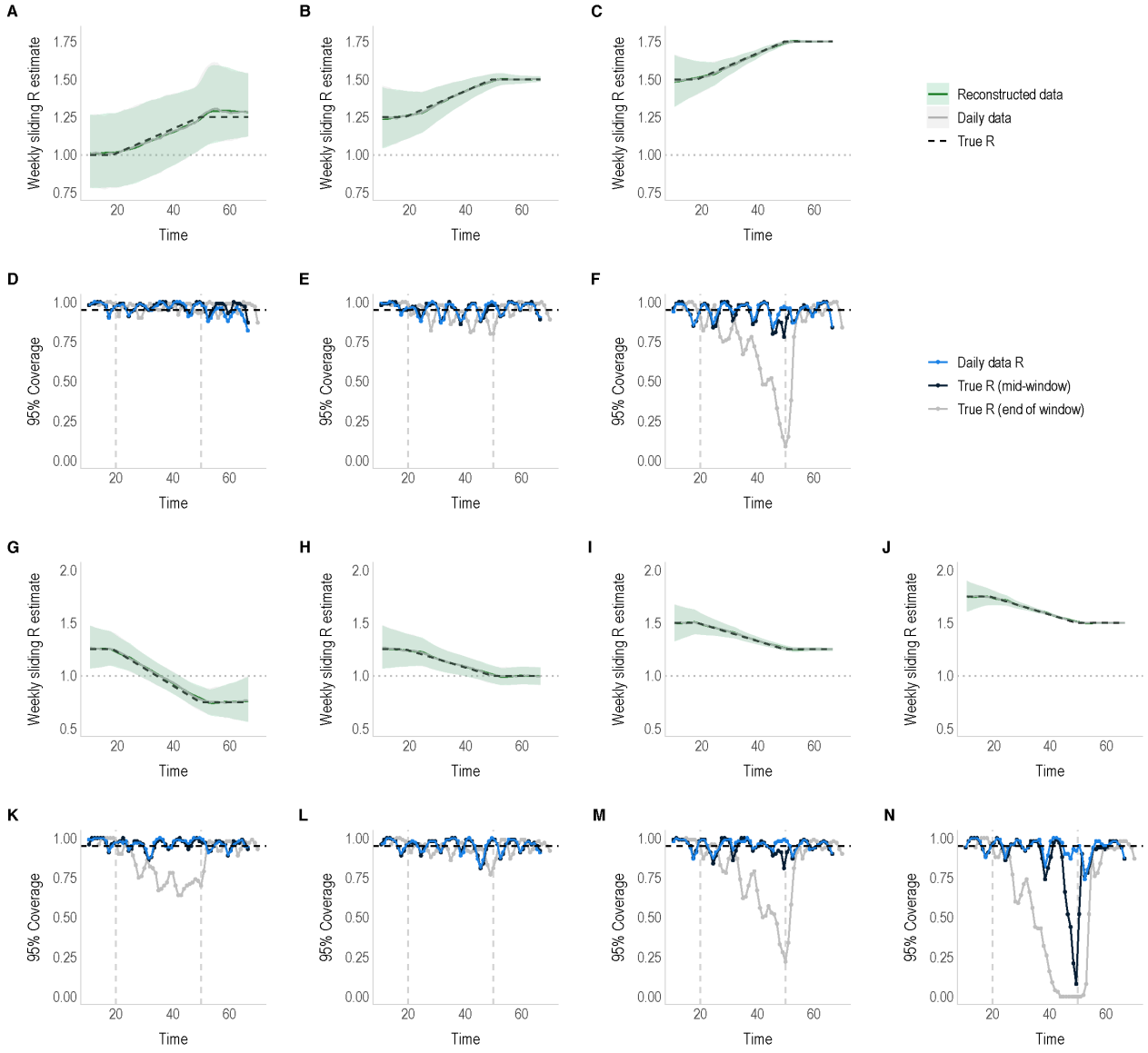


Figure D.15: **Plotting the gradual change in R_t scenarios for 100 simulated epidemics in the middle of the time window. Here, the gradually increasing (A-C) and decreasing (G-J) R_t estimate plots are exactly as in figures S13 and S14, except they are plotted at t - 3.5.** Each 95% coverage plot (D-F & K-N) corresponds to the plot directly above and shows the proportion of the 100 R_t estimates made using weekly data where the 95% credible interval encompasses either the true value of R_t (black) or the value of R_t that would have been estimated from daily data (blue). As a reference, the 95% coverage of the true value of R_t for the R_t estimates plotted at the end of the time window (see figures D.13 & D.14) are shown in grey.

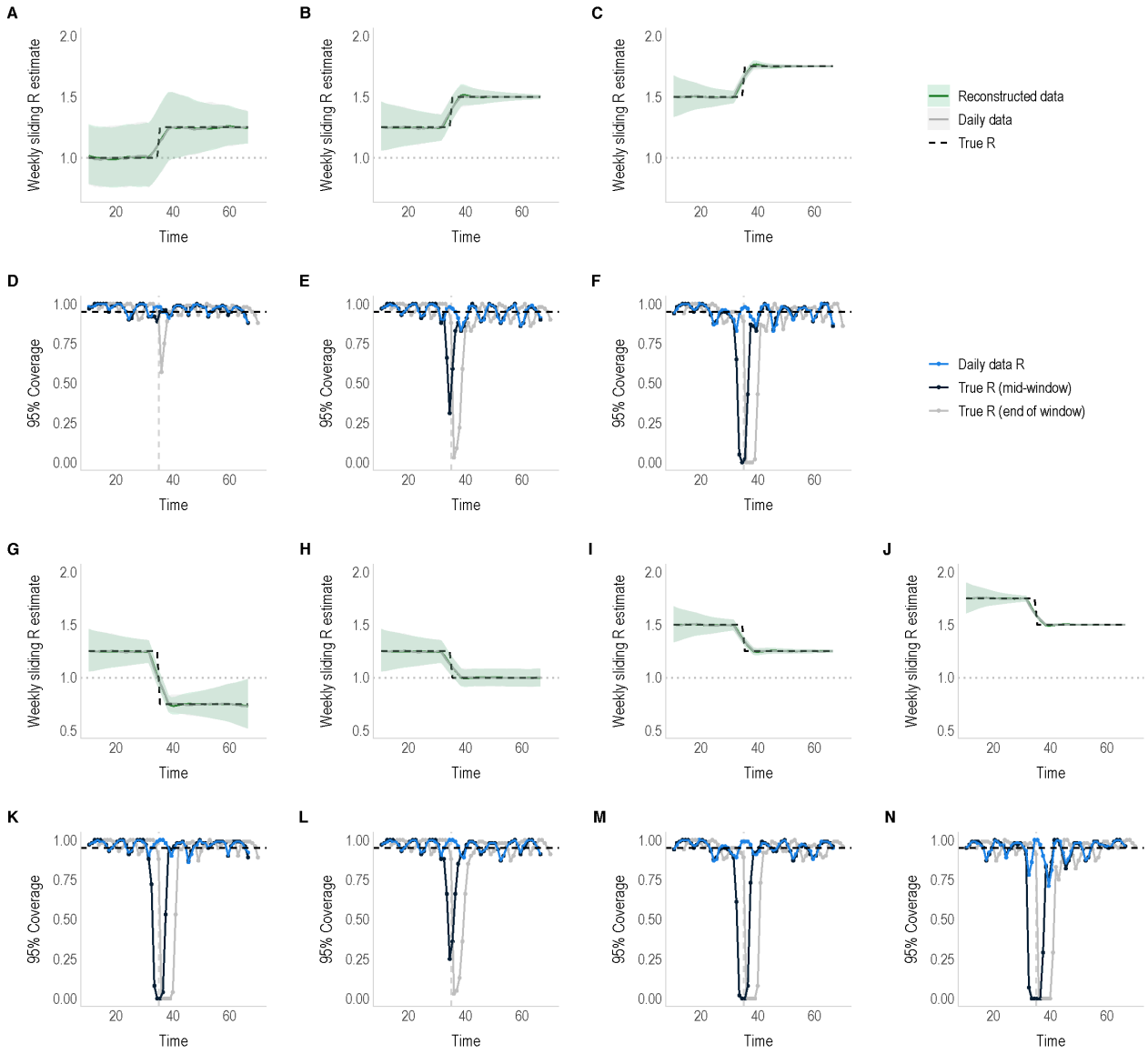


Figure D.16: **Plotting the stepwise change in R_t scenarios for 100 simulated epidemics in the middle of the time window. Here, the stepwise increasing (A-C) and decreasing (G-J) R_t estimate plots are exactly as in figures S11 and S12, except they are plotted at t - 3.5.** Each 95% coverage plot (D-F & K-N) corresponds to the plot directly above and shows the proportion of the 100 R_t estimates made using weekly data where the 95% credible interval encompasses either the true value of R_t (black) or the value of R_t that would have been estimated from daily data (blue). As a reference, the 95% coverage of the true value of R_t for the R_t estimates plotted at the end of the time window (see figures D.11 & D.12) are shown in grey.

D.3.6 Number of iterations

For each scenario, the EM algorithm used to reconstruct the incidence (and in turn estimate R_t) converged over a small number of iterations, with negligible differences beyond 5 iterations (Figure D.17). Given that the computational time is so short (see Appendix D.2.6), the default number of iterations was set to 10 in the R package. However, a convergence check ensures that the final iteration of the reconstructed daily incidence does not differ from the previous iteration beyond a tolerance of 10^{-6}, and the number of iterations can be modified by the user.

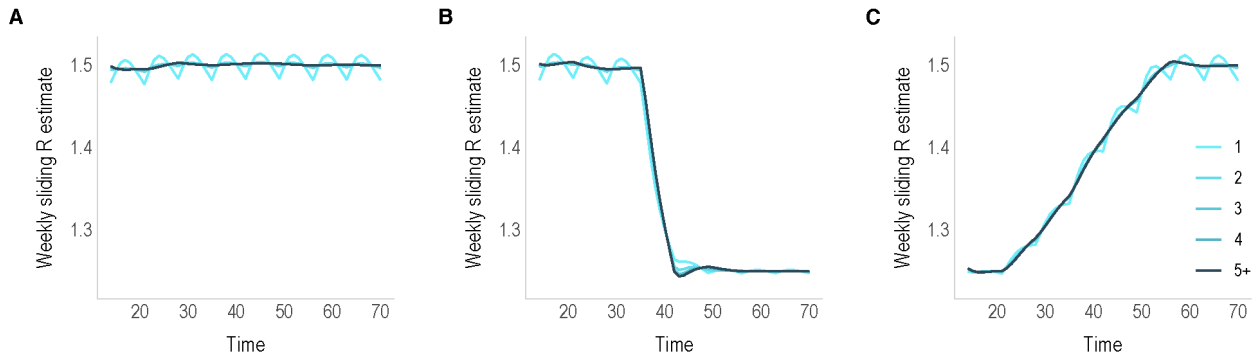


Figure D.17: **The mean weekly sliding R_t estimates generated after each of the 10 iterations from 100 stochastically simulated epidemics** for the following scenarios: A) constant R_t of 1.5, B) stepwise decrease in R_t from 1.5 to 1.25, and C) gradual increase in R_t from 1.25 to 1.5.

D.3.7 Different temporal aggregations

In addition to weekly incidence, the method can be successfully applied to other temporal aggregations of data. Here, we showcase the method applied to data that has been aggregated to 3-day, 10-day and 14-day timescales. First, we consider the ideal scenario where the end of the aggregation window aligns perfectly with the R_t step-change on day 35 (Figure D.18). Then we consider how the accuracy of the estimations would be affected if the step-change fell in the middle of the aggregation window (Figure D.19).

When the aggregation window is aligned with the step-change, there is only the usual lag in detecting temporal changes in R_t, corresponding to the width of the sliding time window used for the estimation (Figure D.18 B, E, H & K). However, when aggregation windows misalign with the step-change, there is a further lag (Figure D.19 H & K). This is because, in the process of daily incidence reconstruction, the growth rate is assumed constant within each aggregation window. When aggregation windows misalign with the step-change, the growth rate is essentially smoothed out when the daily incidence is reconstructed. This is important to be aware of when using larger aggregations of data, as the loss of some temporal resolution in R_t estimates will be unavoidable.

As mentioned in Appendix D.2.2, we recommend that the user ensures that the sliding window used to estimate R_t is equal to or longer than the length of the aggregation window. Although a longer sliding window leads to a further lag in R_t estimates for the stepwise change in R_t scenarios (Panels H & K in Figures D.18 & D.19), the R_t estimates generated are smoother. This is because they are less affected by discontinuities in the reconstructed incidence data, the result of which can be clearly seen in the gradual change in R_t scenarios (Panels I & L in Figures D.18 & D.19), where estimates made using 7-day sliding windows appear wavy.

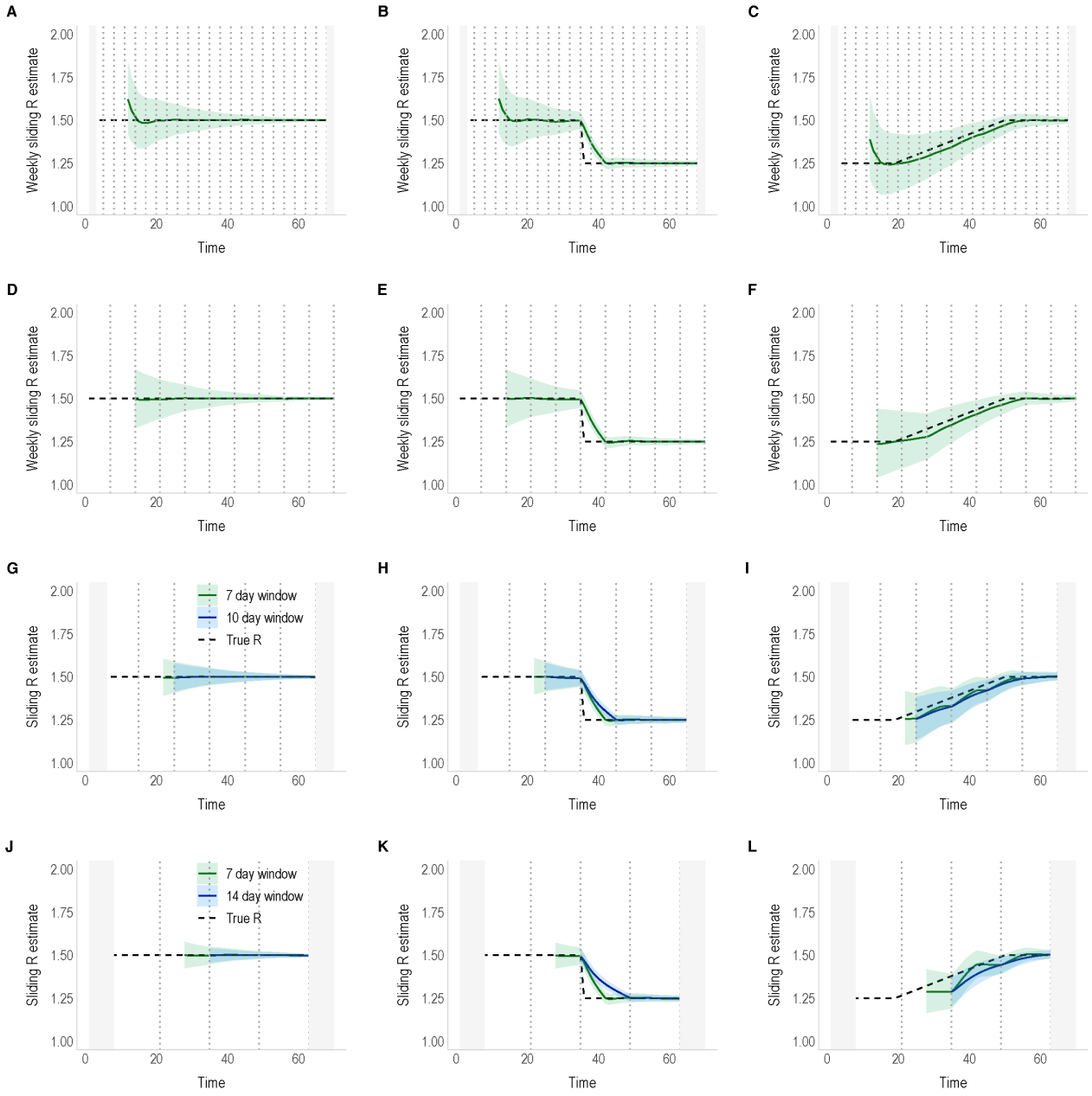


Figure D.18: **Performance of the method using alternative aggregations of incidence data when the end of aggregation windows *align* with the step change on day 35.** We consider the following scenarios: a constant R_t of 1.5 (left), stepwise decrease in R_t from 1.5 to 1.25 (middle), and a gradual increase in R_t from 1.25 to 1.5 (right), with data aggregated over A-C) 3 days, D-F) 7 days, G-I) 10 days, and J-L) 14 days. For the 10-day and 14-day aggregations of data, we compare estimates made using weekly sliding windows (green) with sliding windows of length matching that of the aggregation of data (blue). For all scenarios, R_t estimates are plotted at the end of the sliding time window. Grey dotted lines show the end of each aggregation window and the grey shaded areas are the time periods excluded in order to align the aggregation windows (if necessary).

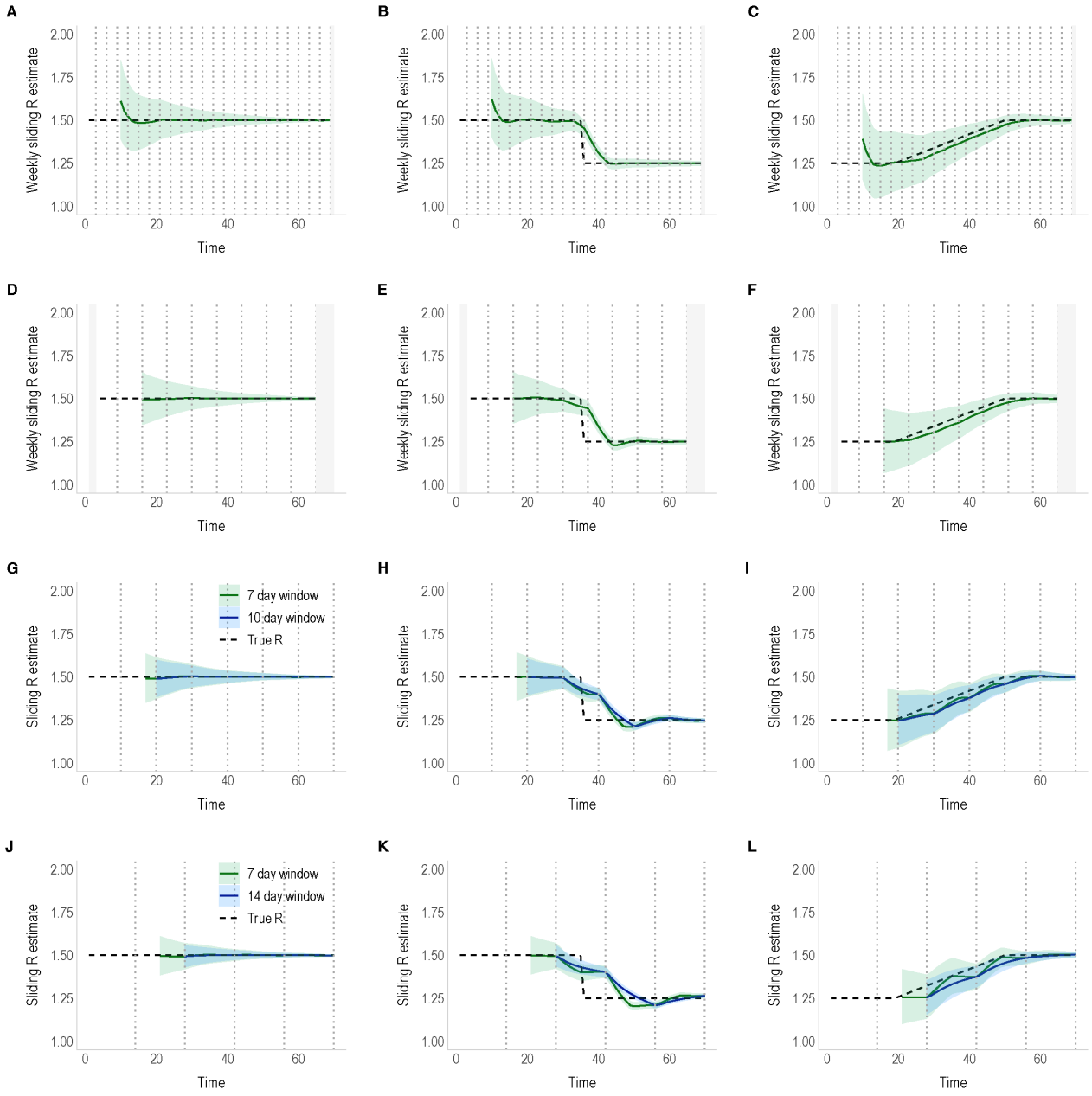


Figure D.19: **Performance of the method using alternative aggregations of incidence data when the end of aggregation windows *do not align* with the step change on day 35.** We consider the following scenarios: a constant R_t of 1.5 (left), stepwise decrease in R_t from 1.5 to 1.25 (middle), and a gradual increase in R_t from 1.25 to 1.5 (right), with data aggregated over A-C) 3-days, D-F) 7-days, G-I) 10-days, and J-L) 14-days. For the 10-day and 14-day aggregations of data, we compare estimates made using weekly sliding windows (green) with sliding windows of length matching that of the aggregation of data (blue). For all scenarios, R_t estimates are plotted at the end of the sliding time window. Grey dotted lines show the end of each aggregation window and the grey shaded areas are the time periods excluded in order to misalign the aggregation windows (if necessary).

D.4 Alternative approach

Here, we compare our EM algorithm method to a LOESS approach, implemented using the 'loess()' function from the stats R package [610].

LOESS fits a smooth curve through datapoints using local regression. It is not designed to disaggregate aggregated data, but rather it interpolates between observations to infer missing data. In addition, it can only interpolate between observations and not beyond them. For use here, we therefore had to make a few assumptions. To address the first issue, we fit LOESS curves to the naïve disaggregation of incidence for each week (e.g. 70 cases in a week / 7 days = 10 cases per day). We then need to decide which day of the week the "observation" is made on (acting as the smoothing points e.g. 10 cases observed on Monday), which influences which days are excluded from the reconstruction because they fall outside the range of the "observations". Finally, we need to decide which degree of the polynomials to be used (varying between 0, 1, and 2).

Using the influenza dataset, we consider two options for smoothing points, either that they fall at the end of the aggregation windows (days 7, 14, 21, 28, 35 - option A), or in the middle of the aggregation windows (days 3, 10, 17, 24, 31 - option B). We then fit the LOESS curves to those "observations" (Figure D.20).

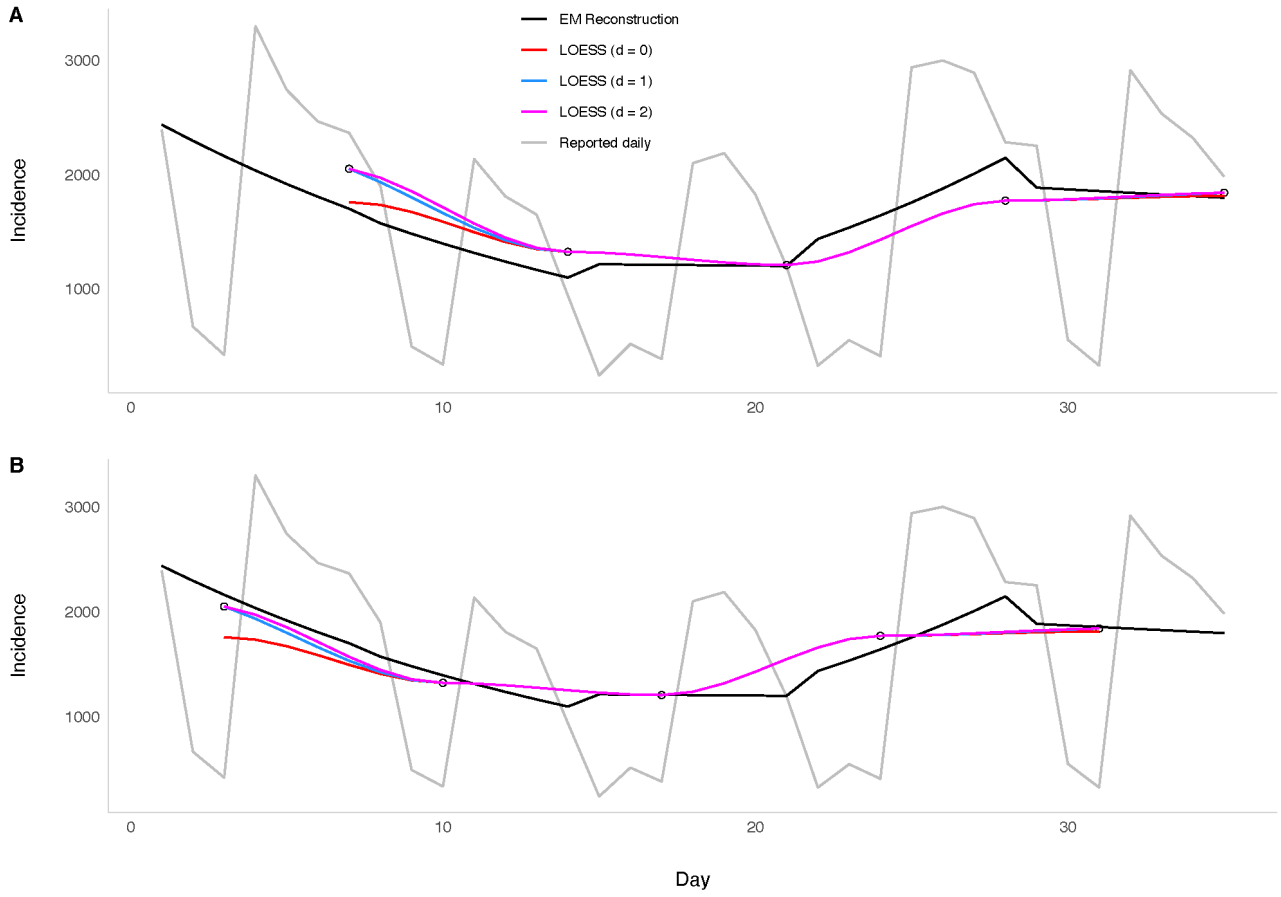


Figure D.20: **Comparison between the reconstructed daily incidence generated by the EM algorithm and LOESS smoothing**, based on smoothing points (black circles) at either the end (A – days 7, 14, 21, 28 and 35) or middle (B - days 3, 10, 17, 24, 31) of aggregation time windows. The "observed" incidence at the smoothing points was computed as the naïve disaggregation of the weekly incidence for that week. The reported daily incidence of influenza (grey), the daily reconstructed incidence generated by the EM algorithm (black), and the daily reconstructed incidence generated by LOESS smoothing with the degree of the polynomial set to either 0 (red), 1 (blue), or 2 (magenta). Note that LOESS smoothing does not allow interpolation beyond the observations.

For option A, data in the first week cannot be reconstructed, whilst for option B, the first two days and final four days of incidence cannot be reconstructed. By using the aggregation mid-points as smoothing points, option B could be a fairer comparison to the incidence reconstructed by the EM algorithm, however, as incidence cannot be reconstructed up to the final day, this approach could not be used in real-time analysis.

As shown, LOESS smoothing generates smooth daily incidence curves without any discontinuities between aggregation windows (see Appendix D.3.2). However, unlike LOESS, our method preserves the aggregated incidence totals supplied by the user. Table D.3 illustrates the differences in the weekly totals when the reconstructed daily incidence generated by either the EM algorithm or by LOESS smoothing is re-aggregated.

Table D.3: **Differences between the re-aggregated reconstructed daily incidence data generated by either the EM algorithm method or LOESS smoothing.** "True weekly" is the weekly aggregations of the reported daily influenza dataset that would be supplied by the user.

Incidence	**Week 1**	**Week 2**	**Week 3**	**Week 4**	**Week 5**
True weekly	14353	9264	8450	12399	12885
EM reconstruction	14353	9264	8450	12399	12885
Option A) Smoothing points at the end of time window					
LOESS (degree = 0)	NA	10570 (+1306)	8798 (+348)	10706 (-1693)	12565 (-320)
LOESS (degree = 1)	NA	11029 (+1765)	8798 (+348)	10706 (-1693)	12637 (-248)
LOESS (degree = 2)	NA	11237 (+1973)	8798 (+348)	10706 (-1693)	12656 (-229)
Option B) Smoothing points at the midpoint of time window					
LOESS (degree = 0)	NA	9228 (-36)	9186 (+736)	12308 (-91)	NA
LOESS (degree = 1)	NA	9247 (-17)	9186 (+736)	12320 (-79)	NA
LOESS (degree = 2)	NA	9275 (+11)	9186 (+736)	12331 (-68)	NA

The differences in the re-aggregated reconstructed daily incidence data are substantial with both options. Reaching up to 21% mismatch for option A and 9% for option B.

We then explored the extent to which these discrepancies in the reconstructed incidence may affect the reproduction number estimates (Figure D.21). R_t estimates from the LOESS reconstructions do not follow the trend in R_t from the daily reported data as closely as R_t estimates from the reconstructed incidence obtained from the EM algorithm.

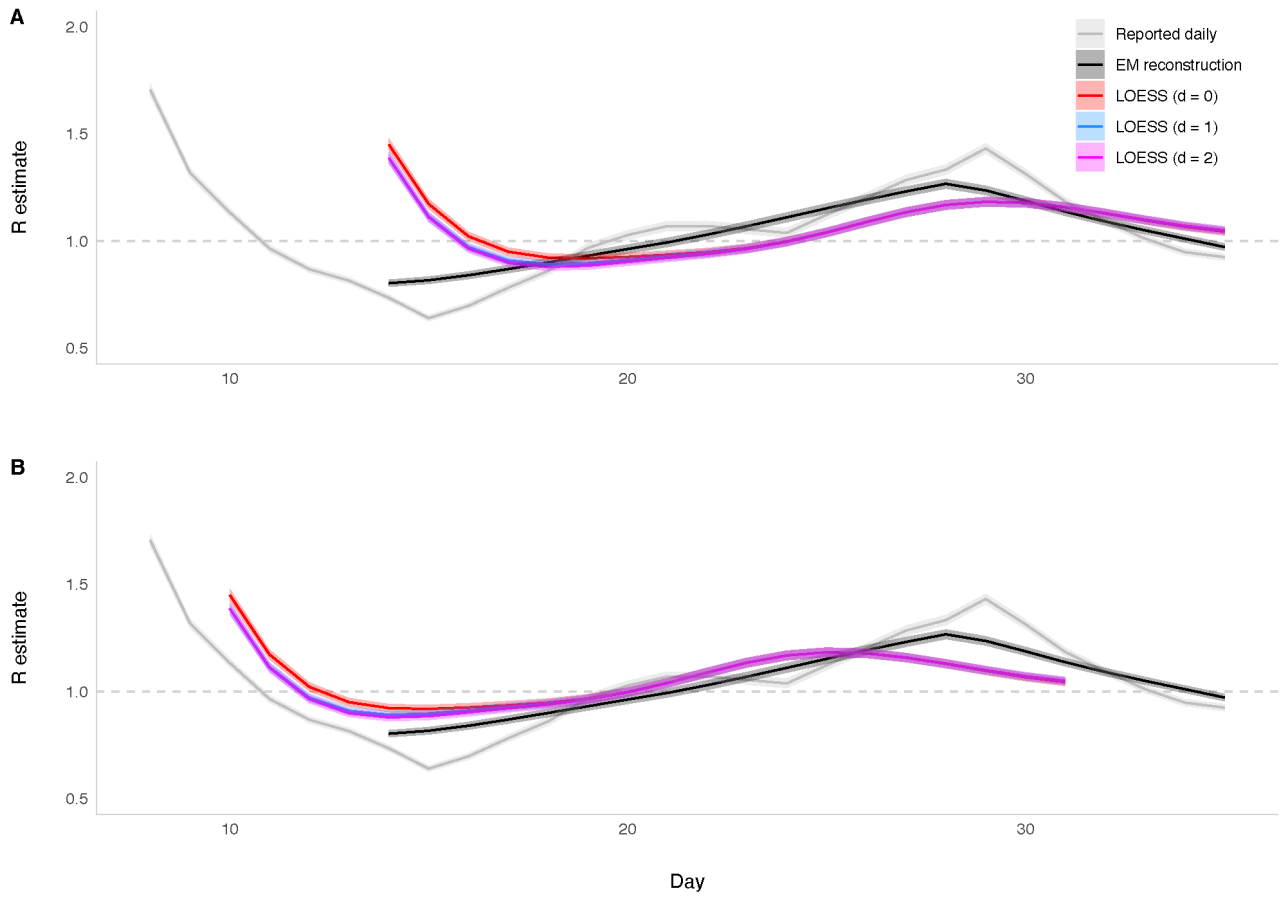


Figure D.21: **Comparison between R_t estimates made using reconstructed daily incidence generated by the EM algorithm (black), or by LOESS smoothing with the degree of the polynomial set to either 0 (red), 1 (blue), or 2 (magenta).** The LOESS approach is based on smoothing points at the end (A) or middle (B) of aggregation time windows.

Bias in the R_t estimates was assessed by computing the absolute difference in estimates from the reported daily data and the reconstructed daily data, generated by the EM algorithm or by LOESS smoothing (Figure D.22). Despite slightly better performance around day 20 with option B (Figure D.22B), we show that estimates using LOESS reconstructions introduce more bias into R_t estimates than those estimated from incidence reconstructed by the EM algorithm. Therefore, we believe it is important to preserve the integrity of the data as much as possible in the reconstruction of incidence.

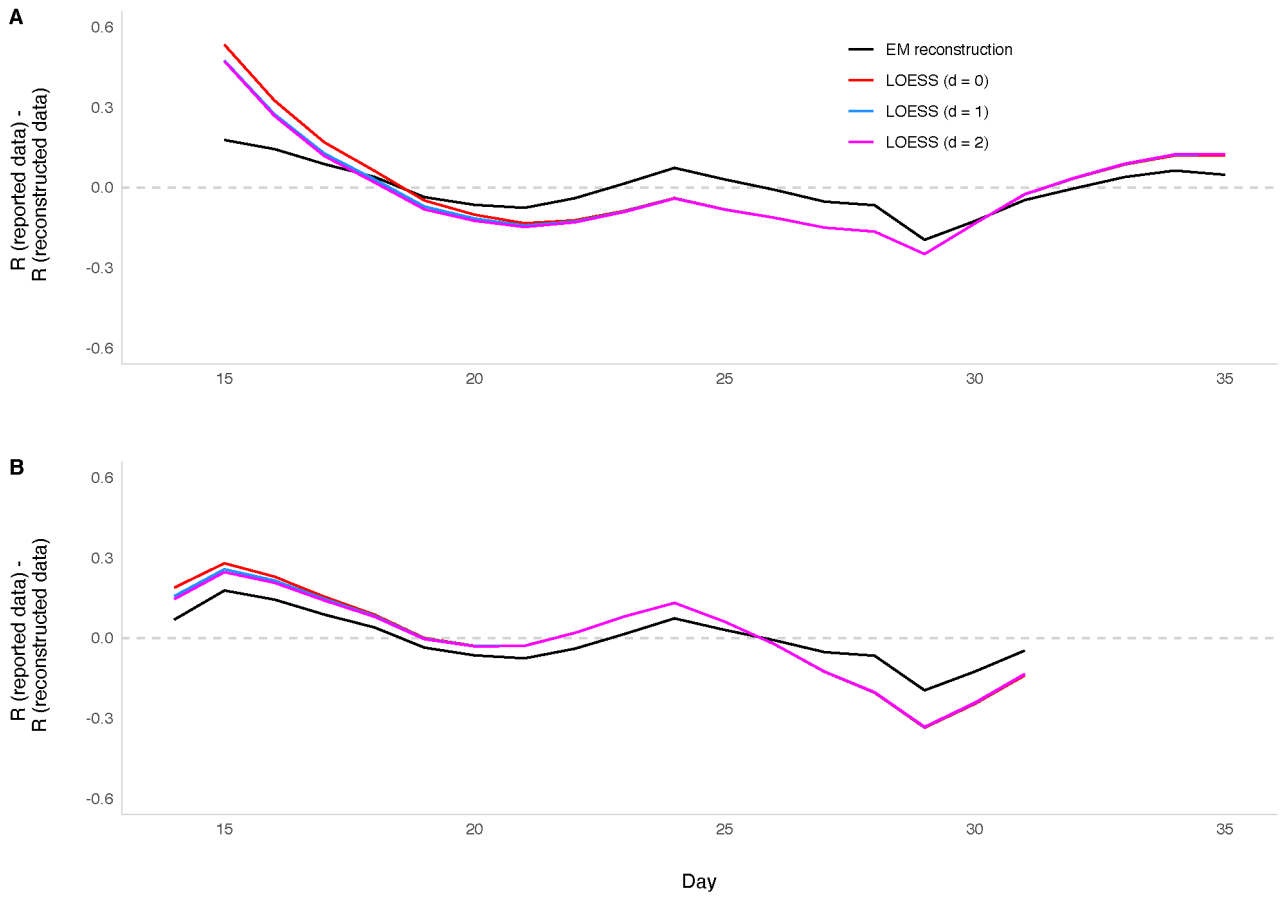


Figure D.22: **The absolute difference between R_t estimates made using the reported daily data and the reconstructed daily incidence, generated by the EM algorithm (black), or by LOESS smoothing with the degree of the polynomial set to either 0 (red), 1 (blue), or 2 (magenta).** The grey dashed line represents an absolute difference of 0. The LOESS approach is based on smoothing points at the end (A) or middle (B) of aggregation time windows.

D.5 Zika virus disease

In order to consider a disease with non-respiratory transmission, we also present an example using real-world data for the vector-borne Zika virus disease. As no daily incidence data for Zika was available to us, weekly incidence of Zika in Rio de Janeiro, Brazil, from 14th November 2015 to 15th April 2016, was obtained from Ferguson et al 2016 [145]. A workaround approach, that is commonly used when applying the original EpiEstim method to temporally aggregated data, involves supplying the SI at the same timescale as the weekly incidence data used. We compared our EM approach to this workaround, assuming a mean SI of 20 days and standard deviation of 7.4 days for the reconstructed incidence generated by the EM approach, and a mean SI of 2.86 weeks and standard deviation 1.06 weeks for the workaround method (Figure D.23) [145].

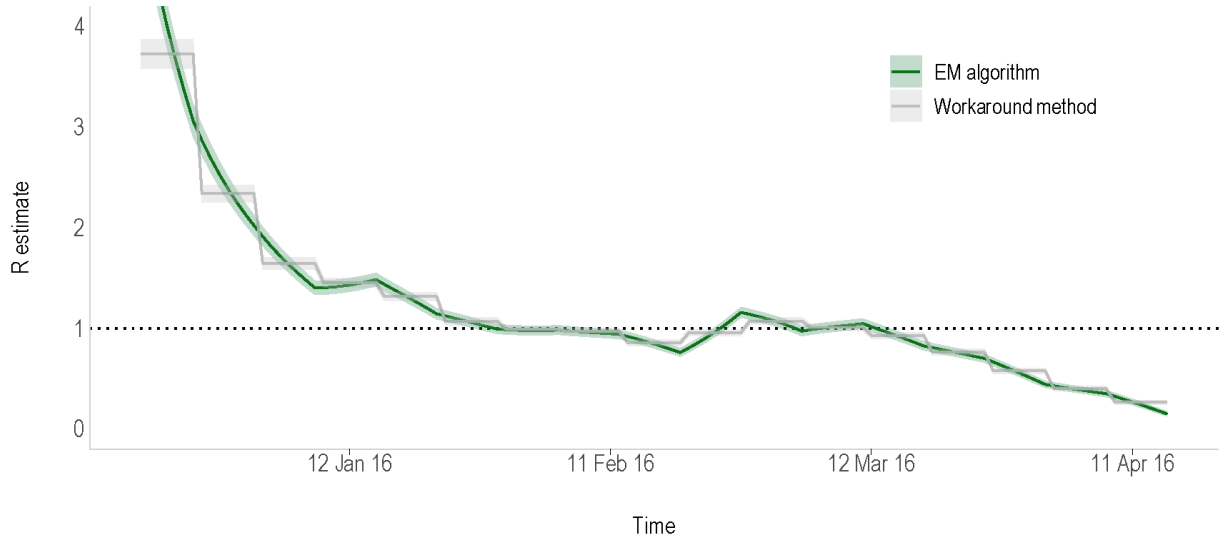


Figure D.23: **Comparison of R_t estimates obtained using a workaround approach (where a weekly SI is supplied to match the incidence data) and the EM approach.** For clarity, we show the R_t estimates between 19th December 2015 and 15th April 2016, when R_t is estimated to have fallen below 4.

We demonstrate that the reconstructed incidence generated by the EM algorithm produces estimates that are consistent with those from the workaround method. However, as R_t can now be estimated from daily reconstructed incidence, the estimates have greater temporal resolution.

It is also important to note that this workaround method cannot be used if the serial interval is shorter than the time windows that data are aggregated over, further highlighting the benefit of our novel EM approach.

D.6 Alternative analysis pipelines

Our EM approach to reconstruct daily incidence data uses EpiEstim within the algorithm, and thus the framework of our method relies on the likelihood used in EpiEstim. In this paper, we focused on producing a consistent statistical framework for R_t estimation, however, it is also possible to use the reconstructed daily incidence generated by our EM algorithm in other analysis pipelines which use alternative methods for R_t estimation.

As an example, we present an analysis comparing R_t estimation using EpiEstim with an alternative R_t estimation method, EpiLPS,[289] using the same set of reconstructed daily incidence data. Using data from our influenza case study, we supply EpiLPS with the reconstructed incidence generated as an output of our EM approach and use the default EpiLPS parameters to then estimate R_t (Figure D.24). EpiLPS estimates are highly influenced by the weekend effects in the reported data (Figure D.24A), and more so than weekly EpiEstim R_t estimates (Figure D.24B). However, as is the case with EpiEstim, the reconstructed incidence smooths out the oscillations in the EpiLPS R_t estimates (Figure D.24A). We note that by default, EpiEstim plots R_t estimates at the end of the time window used, so to be able

to fairly compare the R_t estimates generated by EpiEstim and EpiLPS, we shift the EpiLPS estimates by 3 days to compare the mid-points of estimation windows (Figure D.24B). We demonstrate that the R_t estimates produced by both approaches are very similar, with wider credible intervals using EpiLPS. However, we note that unlike EpiEstim this approach uses data beyond time t for smoothing, and although it has great value in retrospective analysis, this is not a real-time approach.

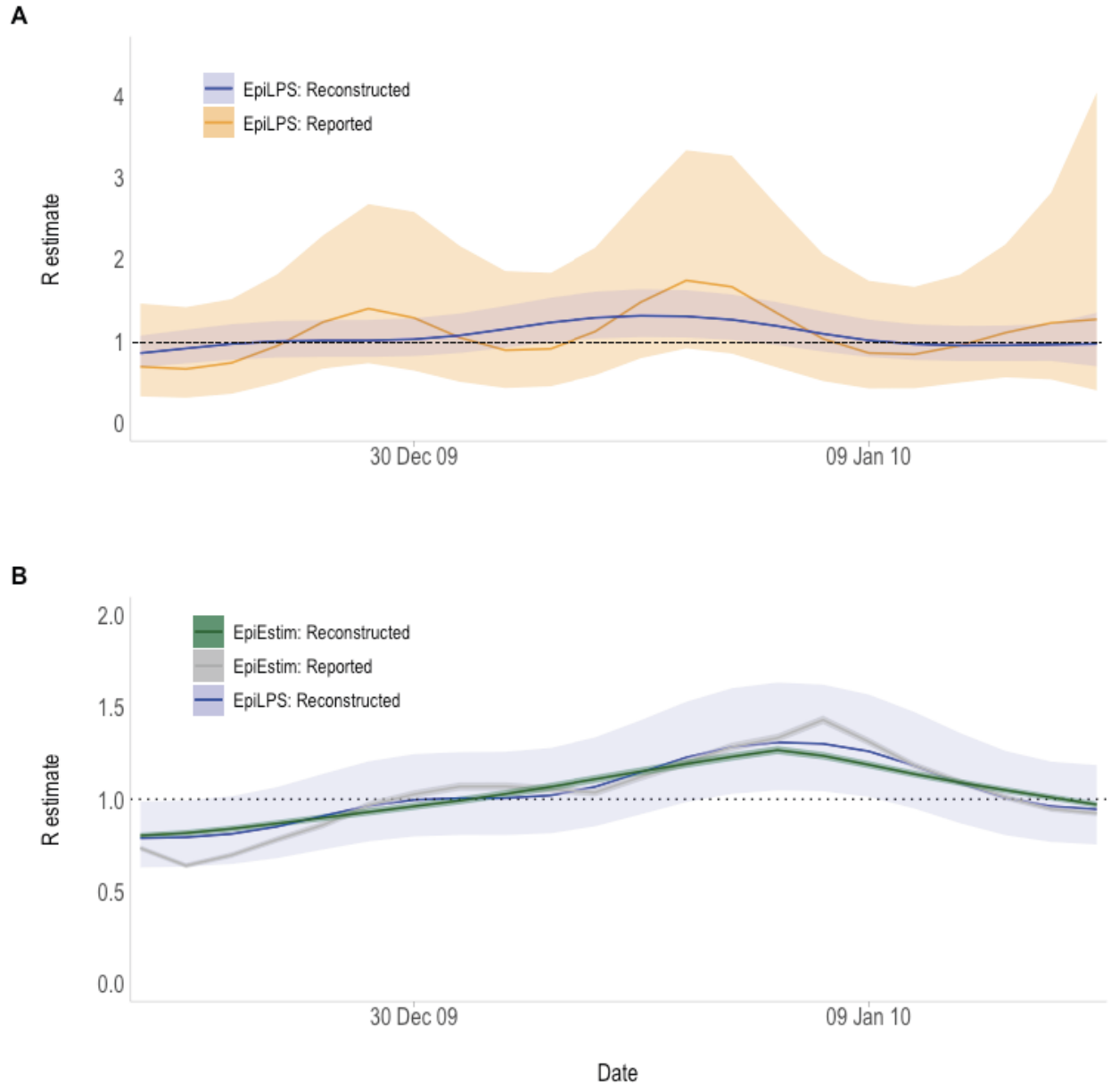


Figure D.24: **Supplying reconstructed daily influenza incidence data to EpiLPS.** A) Comparison between default EpiLPS R_t estimates using reported daily incidence (orange) and daily incidence reconstructed using our EM algorithm (blue). B) Comparison between R_t estimates generated using EpiEstim with daily reported (grey) and daily reconstructed (green) incidence, with EpiLPS R_t estimates using daily incidence reconstructed using our EM algorithm (blue) shifted by 3 days. Lines represent mean R_t estimates, with the coloured shading representing 95% credible intervals.