Measuring the quality and safety of hospital care using specialty-specific indicators based on routinely collected administrative data: a feasibility study

William Lanyon Palmer MA MSc

Department of Primary Care and Public Health

School of Public Health

Faculty of Medicine

Imperial College London

A thesis submitted for the degree of Doctor of Philosophy, March 2014

Declaration of originality

I declare that this thesis is my own work and, where content is not my own, this is appropriately referenced.

William Palmer

Copyright declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work

William Palmer

Acknowledgements

I would like to thank the input of a range of experts who kindly inputted into this research, including Alex Bottle, Carmen Tsang, Charlie Davie, Sue Eve-Jones, and Susan Bewley. While I have been able to attribute the contributions of these experts in the relevant chapters, it is not possible – due to word count alone – to fully recognise the contributions of my two supervisors. Charles Vincent provided sage advice throughout, while Paul Aylin intrinsically knew when to balance pushing me on delivery goals and when I instead needed to talk about nu-folk and cycling apparel. To Karen Taylor: I both blame and thank you for suggesting a PhD (predominantly the latter). Whilst a privilege to do, six years of evening and weekend studying is inescapably draining and I'm indebted to my parents, friends and family for supporting me. Alas, too many of you deserve named recognition to list here.

The PhD was funded, in-part, by the National Audit Office (NAO), who were also my employers during the conduct of the PhD. The NAO played no part in the design, conduct, interpretation and presentation of the results. I am grateful for the support of the Brocher Foundation (www.brocher.ch) during my research fellowship in February and March 2013.

I used Hospital Episode Statistics (HES) dataset, which the Dr Foster Unit at Imperial has permission to hold under Section 251 (formerly Section 60) granted by the National Information Governance Board for Health and Social Care (NIGB, formerly the Patient Information Advisory Group). The unit have approval for using these data for research from the South East Research Ethics Committee.

Abstract

Using administrative data to measure the quality and safety of hospital care offers many opportunities. However, progress has been limited to few countries and predominantly to a small subset of broad measures, such as Hospital Standardised Mortality Rates. In this thesis, I investigate the potential advantages and feasibility – in terms of validity and applicability – of specialty-specific indicators.

In the first part of my PhD work, I examine the case for specialty-specific indicators. I also present potential applications which overcome some of the existing shortcomings of previous uses of indicators based on administrative data.

In the next stage of the project I focus on assessing feasibility by focusing on two specialties – stroke and obstetric care – conducting systematic reviews and consulting with experts to develop two indicator sets. As part of this, I identified the shortcomings in current use of indicators in these specialties. To investigate the limitations of these indicators, I applied the indicator definitions to English hospital administrative data (Hospital Episode Statistics, HES) and evaluated whether they can be used to discriminate between hospitals based on their performance and, importantly, to understand the effect of differences in coding practice.

The final aspect of the research was to investigate alternative applications for the indicators which can overcome some of the shortcomings highlighted in both the prior analyses and existing literature. In doing so, I raise serious, robust shortcomings on the quality and safety of weekend care.

Novel aspects of this research

New indicator sets:	First collation of obstetric and stroke indicators based on administrative data
New application:	First application of some of these indicators to English data
	Applying indicator bundles to improve validity
New methods:	Present innovative methods to investigate bias from coding practice
New insights:	Reveal previously unknown issues on variation of care in England

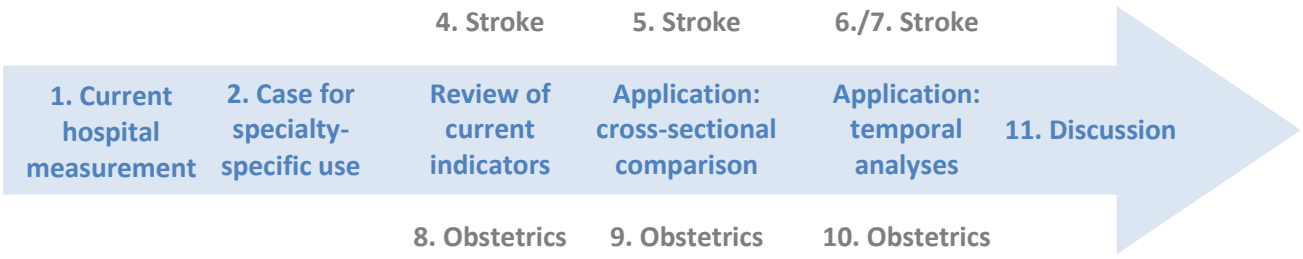


Table of contents

Declaration of originality 2
Copyright declaration 2
Acknowledgements 2
Abstract 3
Table of contents 4
List of figures 11
List of tables 13
List of abbreviations 16

Chapter 1. Quality, safety and their measurement 20
Overview 21
1.1. Introduction to quality and safety in healthcare 22
1.1.1. Improving healthcare 22
1.1.2. Patient safety 22
1.1.3. Quality in healthcare 23
1.1.4. Continuum of quality and safety 23
1.2. Causes, counts and costs 24
1.2.1. Conceptual models 24
1.2.2. Prevalence and preventability 24
1.2.3. Cost of adverse events 25
1.3. Interest of quality and patient safety 27
1.3.1. Policy focus 28
1.3.2. Hospital focus 28
1.3.3. Patient and media focus 29
1.4. Need for indicators 30
1.4.1. Indicators 30
1.4.2. Demand for indicators 32
1.4.3. Strengths and weaknesses of methods for identifying and analysing adverse events 34
1.4.4. No consensus on a preferred technique 36
1.5. Implications for scope 36

Chapter 2. Specialty-specific indicators 37
Overview 38
2.1. Use of administrative data 39
2.1.1. Introduction to administrative data 39
2.1.2. Use of data 40
2.1.3. The development of indicators 41
2.2. Advantages of using administrative data 42
2.2.1. Longitudinal 43
2.2.2. Ease and cost-effectiveness of use 43
2.2.3. Timely 44
2.2.4. Coverage 44
2.2.5. Standardised definitions and rules 45
2.2.6. Basis for Pay for Performance schemes 45
2.2.7. Independent reporting 46
2.3. Limitation of PSIs 46
2.3.1. Attribution of sequence and causation 46
2.3.2. Judgements of responsibility and preventability 47
2.3.3. Problems in coding details 47
2.3.4. Case-mix adjustment 47

2.3.5. Accuracy of coding 48
2.3.6. Some have not been applied to ICD-10 48
2.3.7. Limited evidence on effectiveness 49
2.4. Review of validity of indicators 49
2.4.1. Defining validity of indicators 49
2.4.2. Previous literature on validity 50
2.5. Potential applications of indicators 51
2.5.1. Types of reporting 51
2.5.2. Applying sets of indicators 52
2.5.3. Primarily organisational-wide indicators to date 53
2.6. Advantages of reporting specialty-specific indicators 54
2.6.1. Introduction to concept of specialty 54
2.6.2. Introduction to specialty-specific indicators 55
2.6.3. Difference between specialties 55
2.6.4. Nature of specialties 57
2.6.5. Disadvantages of specialty-specific 58
2.7. Research objectives 59
2.7.1. Hypotheses 59
2.7.2. Thesis outline 59

Chapter 3. Project design and methods 60
Overview 61
3.1. Introduction to study design 62
3.1.1. Introduction to feasibility 62
3.1.2. Addressing limitation in indicator use 62
3.2. Analytical framework 63
3.2.1. Introduction 63
3.2.2. Previous indicator validation 65
3.2.3. The analytical framework 66
3.3. Specialty selection 68
3.3.1. Selection framework 68
3.3.2. Selected specialties 69
3.4. Developing an indicator set 70
3.4.1. Framework for developing indicators 70
3.4.2. Literature review type 72
3.4.3. Previous systematic reviews 73
3.4.4. Literature search resources 74
3.5. Selecting indicators 75
3.5.1. Considerations on indicator selection 75
3.5.2. Defining indicators 77
3.6. Applying indicators 77
3.6.1. Descriptive statistics 78
3.6.2. Regression models 79
3.6.3. Generalised linear models 79
3.6.4. Unadjusted indicator rates 80
3.6.5. Risk adjustment 80
3.6.6. Goodness of fit 80
3.6.7. Standardisation of indicator results 81
3.6.8. Funnel plots 81
3.7. Setting 82
3.7.1. Structure of English NHS 82
3.7.2. Quality and safety in England 82

3.7.3. Focus on acute care 83
3.8. Data 84
3.8.1. Introduction to Hospital Episode Statistics 84
3.8.2. Ethical considerations 85
3.8.3. Software 85

Chapter 4. Stroke literature review 87
Overview 88
4.1. Background on stroke care 89
4.1.1. About stroke 89
4.1.2. Importance 89
4.2. Introduction 91
4.2.1. Stroke as a specialty 91
4.2.2. Stroke measurement 91
4.3. Objectives 92
4.4. Methods 92
4.4.1. Search resources 92
4.4.2. Validation of search results 92
4.4.3. Definitions 92
4.4.4. Search strategy 93
4.5. Review process 94
4.5.1. First stage review criteria 94
4.5.2. Second stage review criteria 95
4.5.3. Data extraction and review 96
4.6. Results 97
4.6.1. Identification of studies 97
4.6.2. Details from studies 100
4.6.3. Study designs 100
4.6.4. Details from datasets 102
4.6.5. Risk models 104
4.6.7. Details from indicators 107
4.7. Discussion 109
4.7.1. Principle findings 109
4.7.2. Source of data 109
4.7.3. Stroke case ascertainment strategies 109
4.7.4. Indicator validity 110
4.7.5. Strengths and limitations 111
4.7.6. Conclusion 112

Chapter 5. Stroke variation and validation 115
Overview 116
5.1. Background 117
5.1.1. Stroke care in England 117
5.1.2. Revisiting limitations in existing literature 118
5.2. Objectives 118
5.3. Methods 118
5.3.1. Selecting indicators 118
5.3.2. Developing indicator definitions 119
5.4. Application and validation of indicators 120
5.4.1. Assumptions 120
5.4.2. Risk-adjustment 122
5.4.3. Statistical testing and validation 122
5.4.4. Sensitivity analysis for coding practice 122

5.4.5. Alignment with analytical framework 124
5.4.6. Analytical packages 124
5.5. Results 125
5.5.1. Importance 125
5.5.2. Precision 125
5.5.3. Construct validity 127
5.5.4. Minimum bias: coding practice 128
5.5.5. Risk model 129
5.6. Discussion 129
5.6.1. Summary of findings 129
5.6.2. Amenable: scope for improvement 130
5.6.3. Precision 130
5.6.4. Minimum bias 131
5.6.5. Errors in underlying data 132
5.6.6. Face and content validity 132
5.6.7. Construct validity 133
5.6.8. Strengths of this study to other stroke measurement efforts 134
5.6.9. Criterion validity 134
5.6.10. Conclusion 134

Chapter 6. Stroke temporal trends (longitudinal) 136
Overview 137
6.1. Background 138
6.1.1. National trends 138
6.1.2. Regional analysis 138
6.2. Objectives 138
6.3. Methods 139
6.3.1. Indicators 139
6.3.2. Analysis 139
6.4. Results 140
6.4.1. National 140
6.4.2. Regional trends 143
6.5. Discussion 145
6.5.1. Longitudinal 145
6.5.1. Regional analysis 146
6.5.2. Limitations 146
6.5.3. Conclusion 147

Chapter 7. Stroke temporal trends (day of admission) 148
Overview 149
7.1. Background 150
7.1.1. Measuring the ‘weekend effect’ 150
7.2. Objectives 150
7.3. Methods 150
7.3.1. Indicators 150
7.3.2. Weekend analysis 151
7.4. Results 152
7.5. Discussion 159
7.5.1. Conclusion 162

Chapter 8. Obstetrics literature review 165
Overview 166
8.1. Background 167

8.1.1. Background on obstetrics 167
8.1.2. Importance 167
8.1.3. Obstetrics measurement 168
8.2. Objectives 169
8.3. Methods 169
8.3.1. Search resources and validation 169
8.3.2. Definitions 169
8.3.3. Search strategy 171
8.3.4. Search terms 171
8.3.5. Review process 173
8.4. Results 175
8.4.1. Identification of studies that applied administrative data 175
8.4.2. Details from studies applying administrative data 177
8.4.3. Details from datasets 179
8.4.4. Risk models 181
8.4.5. Details from indicators 182
8.5. Discussion 184
8.5.1. Principle findings 184
8.5.2. Indicator validity 185
8.5.3. Strengths and limitations 187
8.5.4. Conclusions 189

Chapter 9. Obstetrics variation and validation 191
Overview 192
9.1. Background 193
9.1.1. Obstetrics care in England 193
9.1.2. Revisiting existing literature 193
9.2. Objectives 196
9.3. Methods 196
9.3.1. Data 196
9.3.2. Selection of indicators 197
9.3.3. Developing indicator definitions 198
9.4. Application and validation of indicators 200
9.4.1. Assumptions 200
9.4.2. Risk adjustment 202
9.4.3. Validation 202
9.4.4. Amenable: associations between organisational factors and performance 203
9.4.5. Alignment with analytical framework 204
9.4.6. Analytical packages 205
9.5. Results 205
9.5.1. Importance 205
9.5.2. Amenable 207
9.5.3. Precision 208
9.5.4. Construct validity 212
9.5.5. Minimum bias: coding practice 213
9.6. Discussion 216
9.6.1. Summary of findings 216
9.6.2. Amenable: scope for improvement 216
9.6.3. Precision 216
9.6.4. Minimum bias 217
9.6.5. Errors in underlying data 218
9.6.6. Face and content validity 219

9.6.7. Criterion validity 219
9.6.8. Construct validity 220
9.6.9. Unit of analysis 221
9.6.10. Strengths of this study to other measurement efforts 221
9.6.11. Conclusion 222

Chapter 10. Obstetrics temporal trends 223
Overview 224
10.1. Background 225
10.1.1. Existing obstetric studies on administrative data 225
10.1.2. Other literature 225
10.2. Objectives 226
10.3. Methods 228
10.3.1. General methodological considerations 228
10.3.2. Additional 'day of the week' methodological considerations 229
10.4. Results 231
10.4.1. Longitudinal analysis 231
10.4.2. Association with day of the week 234
10.5. Discussion 240
10.5.1. Summary of findings on seasonality 240
10.5.2. Day of week 241
10.5.3. Conclusion 243

Chapter 11. Discussion 245
11.1. Summary 246
11.1.1. Overview 246
11.1.2. Summary of findings, by hypothesis 246
11.2. Methodological issues 249
11.2.2. Limitations to development of indicator sets 249
11.2.3. Indicator validation 250
11.2.4. Limitations to findings on shortcomings in care 251
11.3. Future research 251
11.3.2. Developing indicator sets for other specialties 251
11.3.3. Further research within these specialties 254
11.3.4. Developing reporting standards 255
11.3.5. Study designs 256
11.3.5.1. Introduction 256
11.3.5.2. Epidemiological studies 257
11.3.5.3. Quasi-experimental 258
11.4. Data implications 258
11.4.1. Background 258
11.4.2. Recommended interventions 259
11.5. Clinical implications 260
11.5.1. Clinical guidelines and quality improvement 260
11.5.2. Advantages of reporting to a 'specialty team' 261
11.5.2.2. Learning and culture 261
11.5.2.3. Develop new and existing indicators 262
11.6. Policy implications 262
11.6.1. Variations in performance 262
11.6.1.2. Use of benchmarking 262
11.6.2. Addressing issue on association between complications and day of the week 264
11.6.3. Application of indicators by policy makers 264
11.6.3.3. Types of application 265

11.6.3.4. Accountability 267
11.6.3.5. Reimbursement 267
11.6.3.6. Make judgements about priorities 268
11.6.3.7. Case findings 268
11.7. Conclusion 269

References 270

Appendices 296
Appendix A. Outputs from the project 296
Appendix B. Details of previous literature reviews of quality and safety indicators 297
Appendix C: Additional websites searched 301
Appendix D: Stroke indicators 304
Appendix E: Obstetrics indicators 306

List of figures

Figure 1. Trend in publications based on administrative datasets 40
Figure 2. Trends in publication of quality indicators and use of administrative data 41
Figure 3. Dimensions of validity used by AHRQ 50
Figure 4. Suggested indicator applications 52
Figure 5. Indicative recommended stroke care pathway 90
Figure 6. Search terms 94
Figure 7. Inclusion and exclusion of articles in the literature review 98
Figure 8. Number of publications compared to latest year of data 100
Figure 9. Positioning of the stroke indicators across the care pathway 119
Figure 10: Details of stroke indicators used in analysis 120
Figure 11: Assumptions for identifying stroke using Hospital Episode Statistics 121
Figure 12: Funnel plots of hospital-level performance across six indicators spanning the hospital stroke care pathway for 2009-10 126
Figure 13. Performance by month[1] 142
Figure 14. Example of hospital trends 147
Figure 15. Number of stroke admissions, by day of presentation 153
Figure 16. Performance by day of admission 157
Figure 17. Hospital-level standardised mortality ratio, for deaths within 7 days of admission on a weekend in 2009/10 163
Figure 18. Overview of the obstetric care pathway 167
Figure 19. Search strategy 173
Figure 20. Inclusion and exclusion of articles in the literature review 176
Figure 21. Characteristics of literature 177
Figure 22. Trend in annual publication numbers 178
Figure 23. Studies comparing hospital performance using obstetric QIs 195
Figure 24. Obstetric indicators across the care pathway 198
Figure 25. Key diagnosis fields and codes for identifying delivery episode 200
Figure 26. Coding rule examples 201
Figure 27: Funnel plots of trust-level performance across the obstetric indicators, 2011-12 209
Figure 28: Funnel plots of site-level performance across the obstetric indicators, 2011-12 210
Figure 29: Trends in obstetric indicators, by month 232

Figure 30. Proportion of births and maternities and breakdown in number of elective caesarean sections, by day of week 234
Figure 31: Association between performance and day of admission/birth 237
Figure 32: Potential uses of patient safety indicators 265

List of tables

Table 1: National estimates on the prevalence of patient safety events 25
Table 2: Examples of different indicators 31
Table 3: The strengths and weaknesses of methods for measuring quality and safety 35
Table 4. Examples of key quality and patient safety indicator sets 42
Table 5. Methods table 59
Table 6. Recommended study design to mitigate limitations 63
Table 7. Domains of validity 65
Table 8. Analytical framework 67
Table 9. Differences in selected specialties 69
Table 10. Development of indicator sets 71
Table 11. Key distinctions between narrative and systematic reviews 73
Table 12. Selected details from sample of previous review of indicators 74
Table 13. Use of multiple measures and mix of process and outcome measures to address analytical framework 77
Table 14. Statistical tests to address the analytical framework 78
Table 15. Simple statistical tests 78
Table 16. Information extracted 96
Table 17. Characteristics of literature 99
Table 18. Variations in use of ICD-10 diagnosis codes for identifying stroke 103
Table 19. Factors included in risk models 104
Table 20. Summary of key factors used in studies 106
Table 21. Categories of measures 108
Table 22. Summary of implications for study design 114
Table 23. Analytical framework 124
Table 24: Number of trusts identified as having statistically significant above or below average performance 127
Table 25: Coefficient of correlation between pairs of indicators 128
Table 26: Model fit statistics 129
Table 27: Annual trend in performance 141
Table 28. Summary of differences in performance 143
Table 29. Crude rates before and after intervention 144

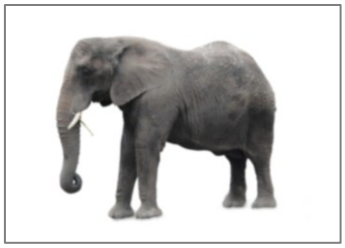
Table 30. Time-Series regression analysis of stroke indicators in London relative to the rest of England 145
Table 31. Characteristics of the study population 153
Table 32. Association between weekday/weekend admission and indicators of quality and safety of care 155
Table 33. The variation in 'weekend effect' across different age groups, for 7-day in-hospital mortality 158
Table 34. The specialty of the first responsible consultant after onset of stroke 159
Table 35. Search terms 171
Table 36. Information extracted 174
Table 37. Example papers 180
Table 38. Key obstetrics indicators 182
Table 39. Risk adjustment in the SimPatIE obstetrics measures 186
Table 40. Selected benchmarks for performance 187
Table 41. Summary of implications of study design 190
Table 42. Details of obstetric indicators used in analysis 198
Table 43. Analytical framework 205
Table 44. Characteristics of mothers in the study population 206
Table 45: Association between birth-to-midwife ratio and performance measures 208
Table 46: Number of trusts identified as having statistically significant above or below average performance 211
Table 47: Trust-level Coefficient of correlation between pairs of indicators 212
Table 48: Site-level coefficient of correlation between pairs of indicators 213
Table 49: Correlation between adjusted and unadjusted rates, and model fit statistics 214
Table 50: Effect of including coding practice in case-mix model 215
Table 51. Comparison of numbers of maternities and births from the HES extract and ONS births register 219
Table 52. Studies using obstetric indicators to investigate temporal issues 227
Table 53. Obstetric indicators 228
Table 54. Recommended levels of consultant presence 230
Table 55. Number of births and maternities and complication rates 231
Table 56. Longitudinal trend and seasonal effect[1] 233
Table 57. Characteristics of the study population 234

Table 58. Association between weekday/weekend admission (or birth) and indicators of quality and safety of care.....236
Table 59 Perinatal rates by delivery rates for common weekend delivery methods.....238
Table 60. Association between business and complication rates, for weekdays April 2010 – March 2012 239
Table 61. Association between compliance with consultant staffing levels and indicators of quality and safety of care.....240
Table 62. Systematic review findings in relation to analytical framework.....253
Table 63. Summary of strengths and weaknesses of the hospital and patient-level approaches 257
Table 64. Key strengths and weakness of common applications of administrative data.....266

List of abbreviations

Agency for Healthcare Research and Quality	**AHRQ**
Agency for Healthcare Research and Quality Patient Safety Indicators	**AHRQ PSIs**
Australian Council for Safety and Quality in Health Care	**ACSQ**
Caesarean Section	**CS**
Canadian Institute for Health Information	**CIHI**
Care Quality Commission	**CQC**
Centre for Patient Safety and Service Quality	**CPSSQ**
Cochrane Library: Database of Systematic Reviews	**CDSR**
Complication Screening Program	**CSP**
Danish National Indicator Project	**NIP**
Database of Abstracts of Reviews of Effectiveness	**DARE**
Deep vein thrombosis	**DVT**
Department of Health (England)	**DH**
Do Not Resuscitate	**DNR**
Excerpta Medica Database	**EMBASE**
Gross Domestic Product	**GDP**
Health Care Financing Administration	**HCFA**
Health Maintenance Organisation	**HMO**
Health Technology Assessment	**HTA**
Healthcare Cost and Utilization Project	**HCUP**
Healthcare Cost and Utilization Project National Inpatient Sample	**HCUP NIS**
Healthcare Cost and Utilization Project Quality Indicators	**HCUP QIs**
Healthcare Cost and Utilization Project Refined Quality Indicators	**HCUP II QIs**
Healthcare Cost and Utilization Project State Inpatient Databases	**HCUP SIDs**
Hospital Episode Statistics	**HES**
Hospital Standardised Mortality Rates	**HSMR**
Institute for Healthcare Improvement	**IHI**
Institute of Medicine	**IOM**
Intensive Care Unit	**ICU**
International Statistical Classification of Diseases and Related Health Problems	**ICD**
International Classification of Diseases, Ninth Revision, Clinical Modification	**ICD-9-CM**
International Classification of Diseases, Tenth Revision	**ICD-10**
International Quality Indicator Project	**IQIP Project**
Joint Commission on Accreditation of Healthcare Organizations	**JCAHO**
Length of stay	**LOS**
Medical Literature Analysis and Retrieval System Online	**MEDLINE**
Medicare Provider Analysis and Review file	**MEDPAR**
National Health Service	**NHS**
National Institute for Health Research	**NIHR**
National Institute for Health and Clinical Excellence	**NICE**
Danish National Indicators Project	**NIP**
National Patient Safety Agency	**NPSA**
National Reporting and Learning System	**NRLS**
Organization for Economic Co-operation and Development	**OECD**
OECD Health Care Quality Indicator Project	**OECD HCQI**
Patient safety indicator(s)	**PSI(s)**
Postpartum haemorrhage	**PPH**

Present on admission	**POA**
Quality Adjusted Life Years	**QALYs**
Quality Indicators	**QIs**
Safety Improvement for Patients In Europe Project	**SimPatIE Project**
Transient Ischaemic Attack	**TIA**
United Kingdom	**UK**
United States America	**USA**
Urinary Tract Infection	**UTI**
Vaginal Birth	**VB**
Vaginal birth with instrument	**VbwI**
Vaginal birth without instrument	**VbwoI**
Veterans Association	**VA**
World Health Organization	**WHO**



Jumbo (1861 – 1885) was an African elephant of enormous proportions. The giant mammal was famed in Paris and London zoos. Despite 100,000 school children writing to Queen Victoria not to sell him, Jumbo was bought by the owner of "The Greatest Show on Earth". The picture has been included here to introduce an idiom, *the elephant in the room,* and a fable, *the blind men and the elephant*, from which you can draw analogies to measuring healthcare quality and safety. However, this thesis perhaps offers a more promising story than that of Jumbo. Unfortunately, the 4 metre tall Jumbo died – reportedly saving a young circus elephant – when crushed by a locomotive. Further, the stuffed Jumbo was destroyed in a fire in 1975.

Part One:

Introduction and concepts

It was six men of Indostan

To learning much inclined,

Who went to see the Elephant

(Though all of them were blind),

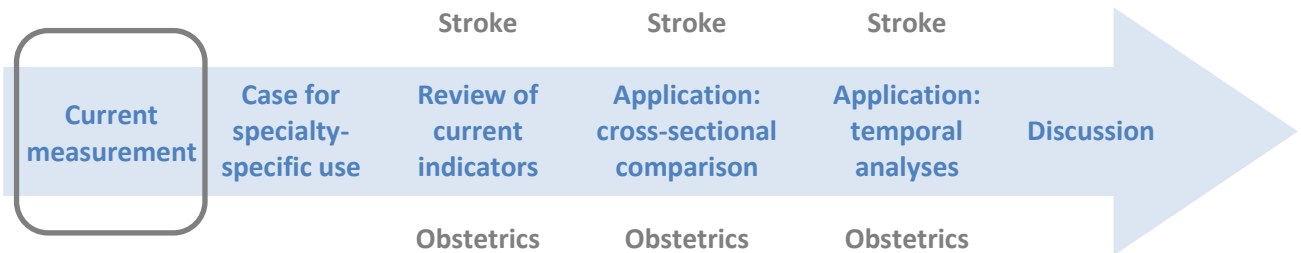
That each by observation

Might satisfy his mind.

(The Blind Men and the Elephant, John Godfrey Saxe 1816-87)

Chapter 1.

Quality, safety and their measurement



Overview

"Declare the past, diagnose the present, foretell the future; practice these acts.

As to diseases, make a habit of two things-to help, or at least to do no harm"

(Epidemics, Book 1, Section XI, Hippocrates, 460-370BC)

Context

This chapter sets out the quality and safety landscape outlining the impact of, and interest in, shortcomings in healthcare. In particular, I highlight the importance of measuring the quality and safety of hospital care. Thereafter, I present the previous techniques used to meet this measurement demand, and provide a short critique of these. The chapter closes by setting out the case for further development of measures of quality and safety.

Methods

A narrative review of the literature

Findings

The quality and safety of healthcare remains a global concern. The literature reveals: significant shortcomings in quality and safety; high impact of these adverse events; potential preventability; and widespread interest. There is scope to develop further healthcare indicators and apply them in a broader array of countries than currently adopts them.

Related papers

Tsang C, Palmer W, Bottle A, Majeed A, Aylin P. A review of patient safety measures based on routinely collected hospital data. *Am J Med Qual* 2011;27(2):154-69.

1.1. Introduction to quality and safety in healthcare

1.1.1. Improving healthcare

There has been considerable progress in health care across the world. From 2000 to 2009, total expenditure on health has increased by a seventh (from 8.2% to 9.4% of total global GDP) and now represents a more significant part of governments accounts, with general government expenditure on health increasing from 13.3% to 14.3% of total government expenditure.[1] People's health is also improving. The World Health Organisation (WHO) estimated that the reduction in the rate of child mortality between 1978 and 2006 is equivalent to over 18 000 children's lives being saved every day.[2] Global life expectancy has also increased by four years, to 68, between 1990 and 2009.[1]

1.1.2. Patient safety

A seminal report by the Institute of Medicine (IoM), published in 1999, outlined the patient safety issue.[3] Despite these improvements, there is a longstanding acknowledgment of the existence of persistent risks to patients' safety. 'The longstanding cornerstone of medicine "do no harm" exists because of the fragility of life and health during medical care encounters'(p1).[4] While health services are treating more patients, due to the ever increasing size and complexity of healthcare, upholding this maxim is proving more and more difficult.[5] Such harm manifests as patient safety events, defined by one organisation as "any unintended or unexpected incident which could have or did lead to harm for one or more patients receiving… care".[6] Patient safety events come under many labels, including: iatrogenic harm, medical errors, nosocomial infections, substandard care, adverse events, critical incidents, preventable or avoidable complications, and sentinel events.

The number of labels reflects the various types of adverse events – such as diagnostic errors, equipment failure, infections, blood transfusion-related injuries, and misinterpretation of other medical orders [7] – and belies the common belief that medical errors usually involve incorrect drug use or wrong-site surgery. Whilst the context within which healthcare services are operating is making the risks more pronounced, the problem of patient safety events is not new, with studies on such incidents being produced throughout the 1950s and an emerging body of evidence by the 1990s.[8] However, despite these early reports, in terms of research, patient safety is still a new and fast growing area.[9]

1.1.3. Quality in healthcare

As a result of developments of the health services and increased scrutiny, expectations have changed. Healthcare is expected to be not just safe but of high quality which can be defined as "the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge."[10] This can include new technologies and procedures. The proliferation of health research and resultant guidelines to instruct the provision of care has helped to codify differences in quality, in terms of processes as well as outcomes. Again 'quality' is a broad term and can include not only evaluation of clinical contact but of patient satisfaction and access.

1.1.4. Continuum of quality and safety

1.1.4.1. Difficulty in disaggregating quality and safety

The changes in perceptions over what is expected from healthcare services redefine the line between quality and patient safety. To demonstrate this problem, consider the thought experiment on misdiagnosis:

Patient A has a rare condition that has ambiguous symptoms. Patient B has a typical case of a common disease that has high diagnosis rates within the population. Both conditions are easily treatable but, on presentation to the health service, both are misdiagnosed and the prescribed treatment makes no effect on Patient A's state but makes Patient B less well.

Given Patient B might have reasonable expectations of being correctly diagnosed and the failure to do so caused harm, most definitions would categorise this as a patient safety event. Whereas, the lower expectations of correct diagnosis and lack of harm would, under most definitions, classify Patient A as quality, but not patient safety, incident. However, the question of what likelihood of correct diagnosis and level of harm (or failure to improve) does the line between quality and safety fall remains subjective. Indeed, this lack of distinction has been highlighted in academic literature.[11,12]

1.1.4.2. Defining quality and safety in this project

Despite difficulties in disaggregating between these two aspects of healthcare delivery, some parts of the remainder of this chapter (particular on estimating prevalence and costs) will focus predominantly on commonly evoked patient safety events and measurement since this is sufficient to describe the importance of measurement, and benefits from being easier to conceptualise given the more dichotomous nature of patient safety events. I did not

attempt to create a potentially burdensome and redundant framework for codifying events along the continuum since both quality and safety have intrinsic interest to all stakeholders, as detailed in section 1.3, p27. For brevity, I use the term 'adverse event' to cover both quality and patient safety incidents.

1.2. Causes, counts and costs

1.2.1. Conceptual models

Many frameworks have been used to describe the causes of patient safety events or shortcomings in quality, although most have focused on the former, such as: Donabedian's interconnected outcome, process and structure;[13] the widely-used[14] categorisation proposed by James Reason with errors as slips, lapse or mistakes;[15] or the long list of causes presented by the World Alliance for Patient Safety including poor training, misdiagnosis and latent failures.[16] Given the different dimensions that can compound to result in an adverse event, any categorisation will struggle to be both mutually exclusive and comprehensively exhaustive. There is, therefore, no agreement on the best scheme for classifying adverse events[14] and many existing categorisations have omissions. For example, Reason's popular scheme fails to capture some commonly cited adverse events, such as equipment failures.

1.2.2. Prevalence and preventability

1.2.2.1. National estimates

The Harvard study in 1984 was the first major investigation into the extent of the patient safety issue and estimated that adverse events, defined in this instance as "injuries caused by medical management, and of the subgroup of such injuries that resulted from negligent or substandard care" (p370), occurred in 3.7% of admissions.[17] A similar, subsequent study undertaken in Australia, using a broader definition of adverse events, estimated an incidence rate of 16.6% of admissions.[18] Although these two studies represent the most exhaustive evaluation of the number of patient safety events in their respective countries, they are still thought to underestimate the actual prevalence.[19] A number of further studies have estimated the prevalence of such events, as outlined in Table 1 below.

Table 1: National estimates on the prevalence of patient safety events

Country	Study	Number of hospital admissions †	Number of adverse events †	Adverse event rate (% patients) †	Adverse event rate (% admissions)
Australia	Quality in Australian Health Care Study, QAHCS (1992)	14 179	2 353 (1 499) ‡	-	16.6 (10.6) ‡
France	(Pilot study)	-	-	-	14.5
New Zealand^	(1998)	6 579	849	12.9	11.2
United Kingdom	(1999-2000)	1 014	119	11.7	10.8
Denmark	(1998)	1 097	176	9.0	9.0
Canada	(2001)	3 720	279	7.5	7.5
United States ~	California Insurance Feasibility Study	-	-	-	4.7
United States	Harvard Medical Practice Study (1984)	30 195	1 133	3.8	3.7
United States	Utah-Colorado Study, UTCOS (1992)	14 565	475 (787) ‡	3.2 (5.4) ‡	2.9

Notes: **^** Included all acute care; **†** Additional information from WHO (2004)[8]; ‡ UTCOS and QAHCS were revised using the same methodology. These comparable results are given in brackets; **~** Study measured 'potentially compensable events'.

Source: Adapted from Vincent (2006), p42.[20]

1.2.2.2. Amenable to healthcare

Further analysis has suggested that most patient safety events identified in the studies of national prevalence were preventable, although "some adverse events are not preventable and they reflect the risk associated with treatment, such as a life-threatening allergic reaction to a drug when the patient had no known allergies to it." [21] In the Harvard case, later analysis estimated that 70% of adverse events were potentially preventable [22] and in the Australian study the equivalent figure was 51%.[18] In a preliminary study in 2001 using a similar method in England, of the adverse events identified (which occurred in 10.8% of admissions) almost half (48%) were judged preventable.[23]

1.2.3. Cost of adverse events

1.2.3.1. Increased use of hospital resources

Patients who are harmed as a result of an adverse event can, and often do, have longer admissions and require additional procedures. For example, in the Harvard and Utah studies,

adverse drug events were associated with an increase of 2.2 and 1.9 days, respectively, in comparison to data from matched controls.[19] When considering only preventable events, the figure for the Harvard study was 4.6 days. Analysis focusing on paediatric care showed that some patient safety events were associated with considerable increases in average length of stay, with an increase of 18 days for decubitus ulcers, 26 days for postoperative sepsis and 30 days for infection as a result of medical care.[24] In addition to the acute care costs, there are also potentially other costs incurred as a result of adverse events, such as in primary care and attendances at accident and emergency when the harm, as can be the case for hospital acquired infection, presents itself after discharge.

1.2.3.2. Increased mortality

As well as harm to patients, adverse events can result in death. Analysing the effects on mortality over a range of patient safety events, Miller and Zhan (2004) found an increased rate of in-hospital deaths.[24] The varying effects identified by the study ranged from those for the patient safety event of birth trauma, with an increased rate of mortality with odds ratio 1.3 compared with infants who did not have trauma at birth, to iatrogenic pneumothorax (odds ratio of 7.5) and postoperative respiratory failure (odds ratio of 76.6). The authors of this analysis, as with that presented in the previous paragraph, are not able to control for potential bias, such as in the scenario where an unmeasured case-mix factor is associated with an increased likelihood both of an adverse event being identified (due to more obvious poorer outcome in those with complex comorbidities) and death. In another study investigating the proportion of active-care patient deaths that were preventable, almost a quarter (23%) were rated by a clinician panel as at least possibly preventable by optimal care, with 6% deemed to be probably or definitely preventable.[25]

1.2.3.3. Financial costs

Adverse events are also an important issue due to their financial cost. An event may introduce additional costs through, for example: the hospital being reimbursed for additional bed days and procedures; legal fees; a burden on the workforce to support the additional care required; emotional cost to any practitioner involved; and any physical or mental harm on the patient. To date, few studies have looked at the societal and personal costs with focus remaining on hospital charges. Drawing again on the findings of the Harvard and Utah studies, adverse drug events have been associated with an increased average cost per admission of $2 262 and $2 595 ($4 685 when considering only preventable adverse drug events).[19] These figures are dwarfed by a US study on paediatric care which found,

across a range of patient safety events, excess charges per discharge ranged from $30 000 to $140 000 depending on the type of patient safety events.[26]

1.2.3.4. National estimates of costs

Further to estimating the effects on patient safety events at patient and hospital levels, efforts have been made to extrapolate the burden on a national scale. Many of these estimates are based on the studies, outlined in Table 1 (p25), that evaluated the prevalence of adverse events. An IoM report estimated that between 44 000 and 98 000 deaths per year were due to medical errors in US hospitals,[27] whereas a news report stated that medical mistakes killed 180 000 people a year.[28] However, McDonald and colleagues (2000) argued that these estimates are likely to be overestimates; in particular, since there is a lack of consideration of the expected risk of in-hospital death in the absence of a medical error.[29] Further questions are cast over the validity of these estimates by the fact that the extrapolation are made by people not involved in the original reports and so the original notions of preventability are missed.[25] Specifically, there are problems associated with the extrapolation of results from small studies to a national level and comparisons between countries are affected by the variations in definitions used in the original studies. A more sophisticated approach, which is relevant to the discussion at the end of the thesis (paragraph 11.6.3.6, p268), might be to assess quality-adjusted life years (QALYs) lost.

As well as the estimates based on the effects on mortality rates, a number of efforts have been made to estimate the cost of patient safety events on a national level. In the US, solely within paediatric care, excess charges were estimated to be in the region of $1 billion.[30] One report suggests that the national annual cost of patient safety events is approximately $37.6 billion. Of this about half of the costs ($17 billion) are associated with preventable errors, and half of this expenditure (on preventable medical errors) is accounted for through direct health care costs.[27] In Australia, a study based on 2003-04 data estimated that the burden was in the region of Aus$2 billion (equivalent to around US $1.5 billion).[18,31]

1.3. Interest of quality and patient safety

There has been an increased focus on quality and safety from all stakeholders. This increased interest is a results of a multitude of factors, including: changing risks;[32] dramatic reforms in healthcare systems and organisational structures;[33] the introduction of legislation governing quality and safety [34] and policies such as linking adverse events to healthcare

providers' remuneration;[33] and increased awareness - both in terms of certain types of events such as hospital acquired infections[35] and due to high-profile adverse events.[36]

1.3.1. Policy focus

1.3.1.1. Government priority

Perhaps as a result of these startling figures on the national effects of poor quality and patient safety events, many governments have begun to increase the priority they place on reducing such events.[37] Some governments have demonstrated their intensions by introducing incentives to reduce adverse events. One indication of this change is the number of new agencies and legislation designed to improve patient safety, such as the National Patient Safety Agency (NPSA) which was established in England in 2001 (although the function transferred to the newly formed independent national body, the NHS Commissioning Board – now NHS England – in the reforms in June 2012).[34]

The social insurance program administered by the government in the USA, have introduced an initiative designed to improve patient safety whereby hospitals are not fully reimbursed if the patient is affected by certain patient safety events.[38] In England, in 2010 the government announced a policy to levy financial penalties on hospitals for care that results in an emergency readmission. As the subject of these policies, this highlights the importance of adverse events to hospitals. This rise in national action may also have been driven by the increased public and media attention, which is discussed below. This is coupled with the increased focus on the health services in general due to, for example, the problems associated with aging populations, rising costs, lack of accountability and inequalities. [32]

1.3.1.2. Intra-governmental priority

Another possible driver for the raised priority is interest taken by international organisations. For example, in 2004 the WHO, following pressure that the Organisation and its member states paid closer attention to patient safety, set up the World Alliance for Patient Safety.

1.3.2. Hospital focus

1.3.2.1. Importance to hospitals

The additional length of stay and procedures outlined above have an operational cost to hospitals due to the burden on their resources, both in terms of capacity and workforce. For example, a 2007 study conducted in Japanese hospitals estimated there were 6,240 person-hours dedicated to patient safety practices per 100-beds each year, including about 1,141

person-hours for infection control alone.[39] The study included incident reporting, external audit, management of medication and maintenance of medical equipment. It did, however, not include the staff hours associated with the additional lengths of stay and procedures. Moreover, the increasingly common policy of patient choice, whereby the person is able to select who provides their healthcare, means any hospital with a reported poor record on adverse events is likely to miss out on possible treatment opportunities and the funding attached to them.

1.3.2.2. Importance of acute care

Acute care is an important part of any healthcare service. Around a third of personal health care expenditure in the USA is committed towards hospital care, with this proportion having increased recently.[40] In England, in 2012-13, half of local health spend was on "general and acute", "accident and emergency" and "maternity" (£47 billion out of £91 billion), which is more than double that spent on primary care (£21 billion on GP Services, community prescribing costs, dental services).[41] Moreover, the estimates given earlier on the prevalence of patient safety events are predominantly based on the hospital setting and, as such, it's a known area of importance.

In particular, the acuity means the likelihood of a perceptible change following an adverse event is important as it aids outcome measurement whilst also reiterating the high impact (and therefore the need to address it).

1.3.3. Patient and media focus

1.3.3.1. High profile cases

A number of high-profile patient safety events have raised the public's awareness of medical errors. In England, the death of a child at Bristol Royal Infirmary following surgery - which was undertaken against the advice of anaesthetists, some surgeons and the Department of Health (DH) - was widely reported in the media and led to an external inquiry. A subsequent investigation into the cases of 53 children treated at the hospital, led to three doctors being found guilty of serious professional misconduct. Other countries have had similar cases capture the media and public's attention. For example, the case of Willie King in Florida, who had the 'wrong leg' removed, and that of Libby Zion, who died following the reaction of a drug administered by an overworked, unsupervised junior doctor, highlighted patient safety issues in the US.[36]

1.3.3.2. Emerging areas of focus

The interest of patients and the media has been further intensified by the emergence of healthcare associated infections as an issue in health service delivery. Studies into the prevalence of infections, combined with high-profile outbreaks, have focused this interest. Both Methicillin-resistant Staphylococcus aureus (MRSA) and Clostridium difficile (C-diff) have emerged as serious issues in healthcare provision[42] and are increasingly seen as amenable to better healthcare.[35]

1.3.3.3. Increased awareness

Further to the media coverage of high-profile incidents, the prevalence means that many people have personal experience of adverse events. A survey conducted in the US found that around 42% of people reported a medical error having affected themselves, a friend or relative.[43] As a result, there is heightened concern about such errors, with respondents to the survey reporting the health care system as 'moderately safety'. Another survey, conducted by the American Society of Health-System Pharmacists, found that over half of Americans are 'very concerned' about either being given the wrong medicine (61%) or complications from a medical procedure (56%).[44]

1.4. Need for indicators

Given the scale of the issues as outlined above, there is a clear need for measurement efforts to help understand prevalence, causes and preventability. In particular the demand comes from a range of sources, as outlined below.

1.4.1. Indicators

1.4.1.1. History of measurement

While the focus on measurement has intensified, it is not new; such studies can be dated to the first anaesthetic death in 1848,[14,45,46] or perhaps even earlier with Florence Nightingale reporting in the mid-1800s that patients in London had a greater risk of dying if hospitalised than if cared for at home.[47] The increased focus on quality and safety was not instantaneously matched by additional measurement efforts. As expressed by Mainz (2003), there is a need for measurement efforts to catch up with the interest in quality and safety:

> "Much of the interest in quality of care has developed in response to recent dramatic transformations of health care systems, accompanied by new organisational structures and reimbursement strategies that may affect the quality of care. However, only of late has systematic evidence about quality of care begun to be collected in most health care systems." (p523)[33]

1.4.1.2. Definition of indicator

Indicators of the quality and safety of healthcare have been defined in several ways.[33] For the purpose of this project, I use the definition based on that proffered by Joint Commission on Accreditation of Healthcare Organizations (JCAHO),[48] as set out below. The requirements of a good indicator are discussed later (paragraph3.2.3, p66).

'Indicator' refers to a quantified measure that can be used to monitor and evaluate the quality or safety of important governance, management, clinical, and support functions that affect patient outcomes.

1.4.1.3. Types of indicators

These indicators can take many different forms in terms of the format of their numerical output and the subject they evaluate. They can measure sentinel events or rate-based incidents and can be related to structure, process or outcome of healthcare (Table 2). I have also added 'proxy outcome' which is often ignored in the literature as a possible indicator. Each of the types of indicators has advantages and disadvantages. For example, while outcome measures can evaluate actual health improvements they are susceptible to influence from factors other than the standard of healthcare, such as environmental aspects.

Table 2: Examples of different indicators

Type of indicator	Detail	*Example*
Sentinel events	Count of never events	*Number of patient who die during surgery*
Rate-based indicator	Proportion of admissions with preferred treatment or outcome	*The proportion of patients in hospital during a study period which acquire a healthcare associated infection*
Structural	Pertaining to organisational resources	*Evidence that hospital uses specialist wards*
Process	Access to specific treatments	*Proportion of trauma patients receiving urgent scans*
Outcome	Relating to patient's health following intervention	*Mortality or emergency readmission rates*
Proxy outcome	Treatment suggesting a certain outcome	*Patients receiving blood transfusions following treatment (suggesting post-operative haemorrhage)*
Intermediate outcome	Biological indicator associated with higher risk of poor outcome	*Proportion of patients with high blood pressure*

Source: adapted from Table 2 and 3 in Mainz 2003 [33]

1.4.1.1. Indicator bundles

While inherently interesting, outcome measures have been criticised for their reliance on appropriate risk-adjustment models to account for case-mix, with suggestions that – for the purpose of monitoring hospital performance – non-mortality outcomes or process measures should be developed.[49] As far back at 1992, there was a recognition that most efforts had concentrated primarily on mortality rates.[50]

Of course these types of measures can be, and have been, used in conjunction with one another with, for instance, one US hospital using a: "framework to monitoring patient safety, combining a valid rate-based measures to evaluate outcomes and processes of case, and non-rate-based measures to evaluate structure and context of care." (p659)[51] Indeed, in a review of methods for measuring errors and adverse event endorsed a comprehensive monitoring system using a combination of methods.[52] The advantages of applying a range of measures is discussed later (paragraph 2.5.2, p52).

1.4.2. Demand for indicators

1.4.2.1. Under-reporting

McIntyre and Popper noted in 1983 that "errors need to be recorded and analysed if we are to discover why they occurred and how they could be prevented" (p1919) [535353524949]; however, emerging evidence suggests that there is considerable underreporting from the current systems. For instance, a study found that in England, only 7% of patient safety incidents identified using a case note review were included in the routine reporting system, with an even lower positive predictive value (5%) for incidents resulting in harm.[54] Whilst not at such a high level, a review of underreporting by Douglas and Pronovost presented findings from studies from across the USA and UK that showed a majority of errors go unreported, with evidence that more serious errors are often not reported. The authors conclude that this underreporting has serious consequences: plaguing the knowledge of the issues; skewing interest toward those events that are reported; and leading to an inability to generalise findings.[55]

1.4.2.2. Increasing demand for measures

As a result of this increased focus, all these stakeholders – from the public to governments and international organisations – have demanded the tools to help analyse the problem and to facilitate improvements. There has been a simultaneous increase in importance to healthcare providers, commissioners, regulators, and patients of assessing the quality of

care.[33,56] The increased focus on quality improvements – related to, for instance, the aging populations, higher prevalence of chronic diseases, the introduction of new medical technologies, and increased public expectation[57,58] – has been coupled with an increased demand for measures of quality and safety.

1.4.2.3. Measurement efforts remain short

Yet the research to date has failed to satisfy the need; "the literature [on measuring adverse events] consists mostly of descriptive studies. Articles on many of the topics of interest are rare, and the studies are of variable quality" (p 10).[59] Whilst studies have given an indication of the scale of the problem, without better measurement some aspects will inevitably be ignored and go unaddressed with quality and patient safety failings remaining the elephant in the room.

1.4.2.4. Government and intra-governmental demand for measures

Such measures are also needed as the foundation of some governmental policies. The requirement is particularly conspicuous in the healthcare systems which have implemented a patient choice agenda, such as the NHS in England. Further to fitting in with current strategies, Governments also require this information to set future priorities:

> "Greater priority should be given to clinical areas where there is evidence that the quality of care is either variable or substandard, so that areas with a substantial potential for quality improvement are chosen." (pi6)[60]

And the move towards measuring quality of care has not only been a function of Governments' requirements, rather the lead has been taken by international organisations, such as the WHO and Organisation for Economic Cooperation and Development (OECD).[32]

1.4.2.5. Increase in guidelines

Many indicators – especially process measures – are based on clinical guidelines. Where there is sufficient clinical evidence, these guidelines are based on a review of these existing findings or elsewhere the standards are based on a formal process to gain appropriate clinical consensus.[33]

There has been a proliferation in clinical guidelines. For instance, in England, the National Institute for Health and Clinical Excellence (NICE) which was only formed in 1999 had, by 2013, 165 clinical guidelines (coupled with guidance for 381 interventional procedures, 275 technology appraisals and 42 public health areas).[61]

1.4.2.6. Demand for new approaches

As well as new measures (paragraph 1.4.2.2, p32) a more general demand has been articulated for new approaches to addressing adverse events. The implication from a number of authors is that measures have to be further developed: "Efforts to improve patient safety have fallen short of IoM's ambitious goal" (p850)[62] and "the overall approach to patient safety…and definitional issues… remains debated"(p1868).[63] Indeed the message that the current approach is inadequate has also been echoed in the UK.[64] In particular, there has been a demand for more systematic and on-going methods:

> "Adverse events are systematically and frequently reported in European countries and analyses are made in order to learn from incidents and improve safety. This is resource consuming work, which creates a major need to measure dimensions of safety on an ongoing systematic basis, implement learning organisations, demonstrate ongoing safety improvement, determine when lapses in patient safety occur, and document positive effect of the efforts made." (p4) [65]

1.4.3. Strengths and weaknesses of methods for identifying and analysing adverse events

A number of different techniques have been used for measuring adverse events, including: review of medical records; interviews with healthcare providers; direct observation; active clinical surveillance; morbidity and mortality conferences; incident reporting system; external audit and confidential enquiry; studies on claims and complaints; autopsy reports; IT and electronic medical records; and administrative data analysis.

The advantages and disadvantages of the different methods are summarised in Table 3. In 2003, two reviews were published on the relatively advantages and disadvantages of methods for assessing patient safety. [52,59] With other, more recent, comparisons of strengths and weaknesses having been published,[66,67] I do not restate these in depth here.

Table 3: The strengths and weaknesses of methods for measuring quality and safety

Category	Method	Strengths	Weaknesses
Ad hoc studies on epidemiology designs and systematic data collections	Review of medical records (case note review)	Uses readily available data Commonly used High reliability of extracted data	Judgments about adverse events not reliable Medical records are incomplete Hindsight bias Format not readily usable for research
	Interviews with healthcare providers	Potentially provides additional data on root causes	Hindsight bias* Reporting bias*
	Direct observation	Potentially accurate and precise Provides data otherwise unavailable Detects more active errors than other methods	Time consuming and expensive Difficult to train reliable observers Potential concerns about confidentiality Possible to be overwhelmed with information
	Active clinical surveillance	Potentially accurate and precise for adverse events	Time consuming and expensive
Methods based on reporting	Morbidity and mortality conferences	Can suggest contributory factors Familiar to healthcare providers	Hindsight bias Reporting bias Focused on diagnostic errors Infrequently used
	Incident reporting system	Provide multiple perspectives over time Can be a part of routine operations	Reporting bias Hindsight bias Data often unavailable for research
	External audit and confidential enquiry	Provides mechanism for identifying root causes Integrates multiple data sources	Validity of the data untested Cost for using data uncertain
	Studies on claims and complaints	Provides multiple perspectives (patients, providers, lawyers)	Hindsight bias Reporting bias Non-standard source of data
Analysis of routinely collected data	Autopsy reports	Provides data otherwise unavailable (e.g. on missed-diagnoses) High reliability of data	Biased data as not systematically undertaken
	IT and electronic medical records	Inexpensive after initial investment Monitors in real time Integrates multiple data sources	Susceptible to programming and/or data entry errors Expensive to implement
	Administrative data analysis	Use readily available data Commonly used	Judgements about adverse events not reliable Medical records are incomplete Hindsight bias

Note: * assessed by author

Source: Thomas et al (2003),[52] Vincent (2006),[20] Michel (2003),[59] Zhan et al (2003).[68]

1.4.4. No consensus on a preferred technique

There is no consensus over which is the best method for measuring adverse events and neither is this surprising, since the utility value of each method depends on the purpose, whether that be identifying rates of incidents, ascertaining causes or acting as a warning system.[20]

Despite the range of methods available for identifying and analysing adverse events, quality and safety shortcomings remain a significant issue and there is little consensus over how best to address it. A study commissioned by the WHO, completed in 2003, identified 262 articles relating to methods for identifying and analysing adverse events.[59] The report evaluated the effectiveness of each of the methods against criteria defined by the WHO [69] which highlighted a lack of a preferred method. This conclusion was echoed in a further study published that year.[52] That is not to say that the WHO think that having no preferred method means the pursuit of measuring adverse events is futile; conversely it has called for improved efforts, so "making them visible"(p4).[70]

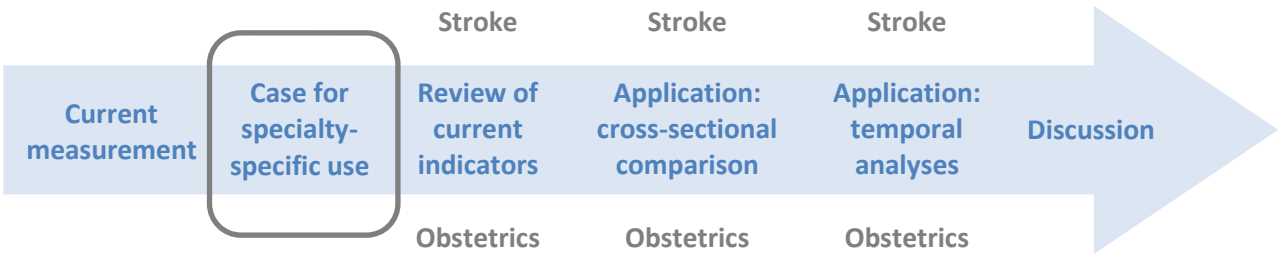
1.5. Implications for scope

This chapter has set out the case for addressing the shortcomings in quality and safety and provided the rationale for investigating new approaches, especially those with potential to be active, on-going surveillance tools. The chapter also explained that the study incorporates measures of what would be typically classified as quality measures alongside patient safety indicators due to this difficulty to disaggregate between the two types and since they both have intrinsic interest. The estimates on prevalence point to a lack of consensus on the scale. This suggests value in both further national efforts and exploratory work on their preventability, scale and causes.

The next chapter sets out the case for using indicators based on administrative data which were scored promisingly in the above-mentioned WHO-commissioned review of methods[59] to meet this demand for an on-going measurement tool. In particular, the chapter focuses on the potential benefits of applying such indicators to specific specialties.

Chapter 2.

Specialty-specific indicators



Overview

"You can't manage what you can't measure"

quoted on p1, Weingart (2003)[71]

Context

The previous chapter presented the argument for measuring quality and safety in healthcare and the importance within that of acute care. In this chapter the existing literature is reviewed to identify potential tools for addressing this unmet need.

Methods:

A narrative literature review

Findings

There is a strong rationale for using administrative data, including that it has the potential to meet the demand for active surveillance tools for quality and safety. Furthermore, specialty-level development and application of quality and safety indicators – rather than the organisation-wide measures which dominate the current debate – can overcome some of the known limitations of such indicators.

What this chapter adds

Introduces the concept of specialty-specific indicators

First explicit rationale for using specialty-specific indicators based on administrative data

Related papers

Tsang C, Palmer W, Bottle A, Majeed A, Aylin P. A review of patient safety measures based on routinely collected hospital data. Am J Med Qual 2011;27(2):154-69.

2.1. Use of administrative data

2.1.1. Introduction to administrative data

2.1.1.1. Definition of administrative data

Administrative data potentially encompass a large array of collection methods and output and the term remains ill-defined. These differences in administrative hospital datasets include variation in the fields and specific codes for diagnoses, procedures, and timings.

'Administrative data' are, for the purpose of this project, defined using the following inclusion criteria:

- routinely collected data covering more than one hospital;
- used for administrative purposes (e.g. for reimbursement); and
- can be replicated using the internationally-recognised ICD9 or ICD10 diagnoses coding framework

2.1.1.2. Common datasets

The most widely used administrative data in research include those of the Agency for Healthcare Research and Quality's (AHRQ) Healthcare Costs and Utilization Project (HCUP), in particular:

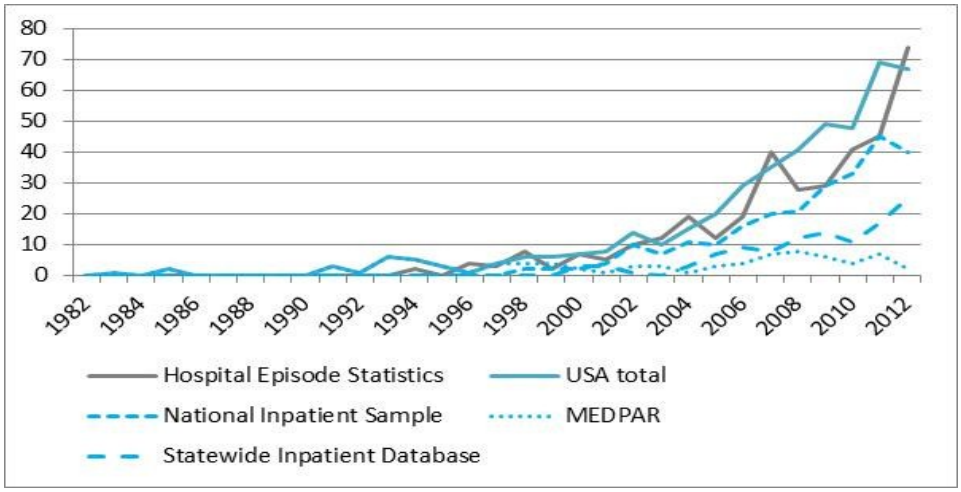
- the State Inpatient Databases (HCUP SID). Currently, 44 States participate in the HCUP SID, encompassing 95% of all USA's community hospital discharges, with data available from 1990;
- the Nationwide Inpatient Sample (HCUP NIS). The largest all-payer inpatient care database in the USA, containing approximately 8 million hospital stays from around 1,000 hospitals, equating to a 20% stratified sample of USA's community hospitals, with data available from 1988.[72]

Within the recent history of research on quality and safety, there has been a rise in the use of administrative data. To demonstrate the use of some key datasets, I conducted a general search of the literature.[i] The results give an indication of the rise in use of such datasets and, while the

[i] I performed a search of title, abstract and key words fields within Medline (Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) 1946 to Present), limited to the names (and common abbreviations or alternate spellings) of key datasets in the USA and, for England, Hospital Episode Statistics.

limitations of the search strategy render the results imprecise, the conclusion that there has been an increase in the publications based on administrative data is unquestionable (Figure 1).

Figure 1. Trend in publications based on administrative datasets



Notes: USA total is a sum of counts for the three non-HES datasets

2.1.2. Use of data

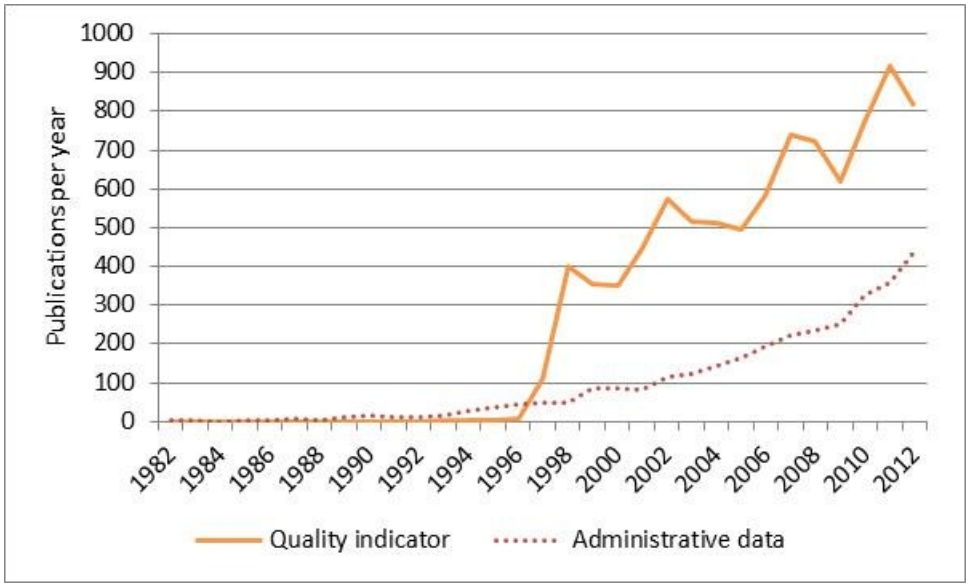
2.1.2.1. Background in non-healthcare settings

The use of administrative data to assess quality has been pioneered by the aviation, petrochemical and nuclear energy industries with their so-called High Reliability Organisations (HROs).[73] Whilst the use of administrative data to measure healthcare outcomes is not new; the heightened attention being paid to quality and safety in healthcare is increasing the demand for data.[27] Such data were used first, in the 1970s, to show variations in care, and its use developed to outcomes measurements in the 1980s before moving on to quality and patient safety since the 1990s.[68]

2.1.2.2. Growth in indicators

Coinciding with a growth in quality indicators, there has been a rise in the use of administrative data. To exemplify this I searched Medline (Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) 1946 to Present) to give an indication of growth in the use of quality indicators and administrative data. I limited the search to the (MESH) indexing term “quality indicator, healthcare” and search term “administrative data” on the title, abstract and key word fields.

Figure 2. Trends in publication of quality indicators and use of administrative data



Administrative data are collected in many countries and there is scope to expand the use of these data to flag issues in the quality and safety of hospital care, both in terms of wider and better coverage of the indicators applied and a broader range of health services subjected to evaluation.

2.1.3. The development of indicators

2.1.3.1. Historical overview of indicator development

Since the development of the original quality indicators in the early 1990s, many medical providers, healthcare organisations and researchers have contributed to their advancement and understanding of their efficacy. A paper in 2001[74] suggested that the number of quality indicators had risen to over 200, and by 2009 one initiative, the International Quality Improvement Programme,[75] had identified nearly 700. However, these numbers should be interpreted with caution as a small variation on an existing algorithm for a measure could be classified as a new indicator thus inflating overall counts.

2.1.3.2. Key indicator sets

A key set of indicators was proposed by the Complications Screening Program (CSP) for screening for variations in quality in 1992.[50] Since then, the Agency for Healthcare Research and Quality (AHRQ) has been at the forefront of indicator development. Examples of key indicator sets are given in Table 4.

Table 4. Examples of key quality and patient safety indicator sets

	Abbreviation	Reference	Example of indicators/groups	Key associated studies
Complication Screening Program	CSP indicators	Iezzoni et al (1992)[50]	27 indicators, e.g.: Septicemia, wound infection, cellulitis or decubitus ulcer.	Iezzoni et al (1997)[76] Lawthers et al (2000)[77]
HCUP Quality Indicators (1994)	HCUP QIs	Ball et al (1998) [78]	33 indicators, e.g.: Hysterectomy, low birth weight, mortality after knee replacement	Johantgen et al (1998)[79]
Refined HCUP Quality Indicators	HCUP II QIs	Davies et al (2001)[80]	25 provider-level, e.g.: stroke mortality, VBAC rates, CE volume	
HCUP Patient Safety Indicators	HCUP PSIs	Miller et al (2001) [81]	12 groups, e.g.: transfusion reaction, postoperative infection, obstetric misadventure	
AHRQ Patient Safety Indicators	AHRQ PSIs	McDonald et al (2002) [82]	20 hospital-level indicators, e.g.: decubitus ulcer and birth trauma	Romano et al (2003)[83] Zhan et al (2003) [84]
Indicators for Patient Safety, OECD	OECD HCQI	Millar et al (2004) [85]	21 indicators, e.g.: Wrong blood type; ventilator pneumonia; birth trauma	McLoughlin et al (2006)[86]
Safety Improvement for Patients in Europe	SimPatIE	Kristensen et al (2007) [65]	42 indicators with 24 recommended, e.g.: wrong site surgery; complications of anaesthesia; decubitus ulcer	

Notes: VBAC - Vaginal Birth After Caesarean; CE – Carotid Endarterectomy

2.1.3.3. Development of indicators

In the editorial introduction to the first major set of quality indicators – CSP, Iezzoni et al (1992) – the commentator states that they "believe the algorithms will prove difficult to improve upon."(p361)[50] However, there have been a number of improvements in the indicators to address the limitations of the previous iterations. For example, in 2001 AHRQ published a new set of indicators to address some of the shortcomings of some of the original Healthcare Cost and Utilization Project (HCUP) Quality Indicators, such as their lack of: severity or risk adjustment; measures covering non-surgical medical care; and reliability.[87] Despite these developments, even by 2008, this "valuable source of information on patient activity remain[ed] under-utilised."(p2)[88]

2.2. Advantages of using administrative data

In a review of available methods for measuring adverse events, Michel highlights that: (1) rapid assessment methods are needed; (2) the relevance of the method depends on the patient safety

measurement goal; and (3) the relevance of a method depends on the availability of the data.[59] The first and third of these points are covered in the advantages set out below, while the second is integral to the discussion about the purpose of the reporting, raised in Chapter 11. There are many advantages offered by using administrative data to evaluate quality and safety, include those listed below. Advantages are, of course, contextual but those set out here can be viewed in the light of surveillance (a purpose whose rationale is set out above).

2.2.1. Longitudinal

Administrative data are typically continuous.[68,89] This property permits two additional routes of investigation over snapshot cross-sectional studies: tracking the performance of providers over time (multiple cross-sectional studies) or tracking patients over time (cohort studies) by linking of patients' records from different admissions. This longitudinal nature is particularly useful for identifying, by looking at specific readmissions, an adverse event for which the symptoms do not present until after discharge.

While this project focuses on acute care, it is worth noting that this longitudinal property of the data allows the researcher to make inferences about what happens outside of hospitals, which has been an area of increasing academic interest. As Weingart and colleagues [90] pointed out, little is known outside hospitals, but administrative data have the potential to record, through emergency admissions and readmissions, an adverse event which may occur outside the non-acute setting. Already, patient safety projects are looking to extend the focus of their research to beyond purely inpatient admissions.[56]

2.2.2. Ease and cost-effectiveness of use

2.2.2.1. Accessibility of data

Administrative data are generally collected to support monitoring or reimbursement activities [33] and, therefore, need to be readily accessible. This advantage has been pointed out in the literature [71] [68] and, more directly, one paper concluded that such information represented "the most readily available set of data systems for tracking patient safety"(p9).[34]

2.2.2.2. Computer readable

The monitoring and reimbursement purposes of administrative datasets involve analysis of the information they contain. This property of being in computer-readable and analysis-ready format makes the use of the data for studying adverse events easier.[91]

2.2.2.3. Cost-effective

Similar to the advantages of being accessible and computer readable, an additional benefit stems from the administrative data having already been collected and so can usually be accessed at little or no cost.[24,33,68,71,73,92] "Administrative datasets provide a substantially cheaper… way of monitoring rates of adverse events"(p24).[93] International examples have shown that the costs of running nationwide programmes for disseminating information from these indicators are also relatively low. For example, the Danish National Indicator Project (NIP), which monitors clinical indicators (both quality and patient safety indicators) across six diseases, costs approximately US$750 000 nationwide per year. The project team estimate that the marginal cost for adding additional disease areas to the project would be in the region of US$75 000 [56]. In England, the HES database costs an estimated £1 per record to collect compared with around £10 - £60 per record for clinical registers.[89]

2.2.3. Timely

Since the monitoring and reimbursement needs to be done in a timely way, the administrative data are provided regularly and with limited time lag.[89] For example, in the NHS in England, the underlying data that are the basis for the Hospital Episode Statistics (HES) administrative dataset are usually updated every month. As such "administrative datasets provide a… more timely way of monitoring rates of adverse event"(p24).[93]

2.2.4. Coverage

2.2.4.1. Large populations

Another attraction of administrative data are that they covers large populations[71,89,91] with, in many cases, the information covering every inpatient procedure offered in hospitals. Again taking England as an example, HES include not only care given by NHS hospitals but also care provided to NHS patients by the independent sector and care given to private patients in NHS hospitals. In 2007-08, the HES dataset included more than 15 million episodes of inpatient care and over 54 million outpatient attendances.[94]

2.2.4.2. Increasingly large proportions of healthcare

Historically, administrative datasets focused predominantly on subsets of inpatient care. However, in response to the demands for increased data to support initiatives such as pay-for-performance within hospitals, there is increasing coverage of services included in administrative data. For example, in the NHS, national data on outpatient episodes and A&E attendances are also recorded, accounting for approximately 12 million and 40 million respectively each year.

2.2.5. Standardised definitions and rules

Administrative data are based on standardised definitions and rules which have been improved through a number of iterations. The first international classification edition was adopted in 1893. The role of updating the framework was passed to the WHO at its creation in 1948, and regulations adopted in 1967 stipulate that Member States use the updated classifications. The 10th revision of International Classification of Diseases (ICD-10) was introduced in 1990 and is used in over 100 countries.[95] However, there are some notable exceptions, with the USA predominantly using an updated ICD-9 framework. These have been developed over a number of years with many iterations and now offer a relatively comprehensive list of codes and associated guidance.

2.2.6. Basis for Pay for Performance schemes

2.2.6.1. Resulting in improved coding

Administrative data are increasingly used for more rigorous performance monitoring and as the basis for reimbursement schemes. Although this does increase the incentive for gaming, it is also a motivation for increased depth of coding, for example coding more secondary diagnoses. Further, the use of administrative data for reimbursement schemes is often associated with an augmented data audit regime, as was the case in the NHS where the Audit Commission have assumed more work in researching coding accuracy.[ii] The expectation is, therefore, that the accuracy and depth of coding is improving.[34]

A systematic review, published in 2012, on discharge coding accuracy of routinely collected data in Great Britain found that primary diagnosis accuracy has improved from 73.8% to 96.0% (p=0.02) since the introduction of Payment by Results in 2002. The authors concluded that such data are sufficiently robust for use in research and managerial decision making.[96]

2.2.6.2. Some adverse events can be costed

This reimbursement aspect has other advantages. Given that healthcare systems are resource constrained, a full understanding of cost implications is important for policy makers to choose priorities. The use of administrative data to reimburse hospitals means that, when a possible adverse event is identified, the cost associated with that admission can be readily estimated. As a result, the economic impact of adverse events can be approximated,[59] and hospitals and governments can make more informed decisions on resources or service redesign.

[ii] In 2014, legislation to abolish the Audit Commission was passed as law, with a planned closure date of 1 April 2015. However, the organisations involved in setting the Payment by Results framework (Monitor, NHS England and Department of Health) will all have an invested interest in ensuring good quality data.

2.2.7. Independent reporting

The individuals who code the medical records are independent from the team involved in the healthcare procedures. This arrangement has some limitations, such as the lack of knowledge of the care provided increasing the chance of miscoding, which are discussed below. However, whilst this does not mean that the reporting is entirely independent since the clinicians are still involved in recording the procedures on the medical record, indicators based on such data at least offer some level of independence, thus meeting one of the WHO's criterion for good indicators.[97]

2.3. Limitation of PSIs

"The sensual eye is just like the palm of the hand.

The palm has not the means of veering the whole of the beast."

(The Elephant in the dark, Rumi, 1207-1273)

Whilst indicators of quality and safety based on administrative data offer a number of intrinsic advantages as already discussed, some limitations have been raised. These concerns are often linked to the information not being specifically designed to analyse patient safety. The most prominently cited limitations, which constrain the possible applications rather than making administrative data redundant with respect to analysing patient safety, are outlined below.

2.3.1. Attribution of sequence and causation

2.3.1.1. Difficulty to attribute

While administrative data are longitudinal and include dates for procedures, the datasets offer limited information on the timing of events.[81,93,98] For instance, it is difficult to accurately deduce the sequence of events, including when a diagnosis was made,[68] and therefore doubt can remain over whether harm was related to quality of care, an adverse event or a pre-existing morbidity.[34,40] The HES dataset in England, for example, does not explicitly distinguish whether a safety incident (flagged by a diagnosis or procedural code) was due to medical error rather than an existing complication which was present on admission to the hospital.

2.3.1.2. Importance of attribution

Reviewing the existing literature reveals conflicting findings on whether this issue on attribution is important. One study investigating the issue of present-on-admission based on 14 selected AHRQ PSIs found that for five indicators the rate would be lower if 'present on admission' was accounted for by the administrative data.[99] A study, in the USA, evaluating the sensitivity and specificity of some complication codes at flagging patient safety incidents found, by retrospectively reviewing the

case notes, that the flagged condition was present on admission in 13% of surgical cases and 58% of medical cases.[100] Another study suggested one-third of discharges were flagged by PSIs because of a diagnosis already present at admission.[101] However, some of the adverse events that are dismissed as present on admission could be the result of a patient being harmed during a previous admission or the failure of the primary or tertiary care sectors, and therefore still of interest. In response to this problem a number of studies have altered the coding structure so that pre-existing conditions are flagged.[102]

2.3.2. Judgements of responsibility and preventability

As well as lacking detailed information on the sequences and causation of an adverse event, neither does administrative data contain any consistent approach for recording whether the harm was preventable or rather a complication of care.[71,93] Although some of the diagnosis codes do explicitly imply preventability and certain adverse events lend themselves to detection through appropriate algorithms, many indicators rely on judgements about whether a sequence of procedures and diagnoses is likely to identify an event which was preventable. This is particularly relevant to indicators of hospital acquired infections, since it may not be clear that the appearance of an infection-related diagnosis code during the admission is the result of a delayed diagnosis of an existing infection or one that has actually been hospital-acquired.

2.3.3. Problems in coding details

The administrative data are based on the coding of case notes using a consistent framework. Further to imperfect information about preventability there are also limitations since the coding frameworks were not developed to specifically analyse quality and patient safety issues and are not comprehensive.[68,81] As a result, the resultant indicators do not represent a comprehensively exhaustive list of measures as analysing administrative data cannot pick up all possible adverse events.[24] As such, the method would be inappropriate for trying to estimate prevalence.[99] It should be noted that ICD-10 has a generally more detailed coding framework than the predecessor ICD-9 – e.g. viral hepatitis expanded from ICD-9 code '070' to ICD-10 codes 'B15'-'B19', and from one to five 3-digit categories[103] so such criticisms raised on the former framework – such as by the Agency for Healthcare Research and Quality [40] – cannot necessarily be generalised to the more comprehensive ICD-10 framework.

2.3.4. Case-mix adjustment

There are also limitations to the extent in which administrative data permit you to adjust for case mix.[81,98] For example, there are a heterogeneity in severity within many of the diagnosis codes,[34] which is a particular issue within medical or psychiatric specialties.[4] Similarly, it is also a problem

where patients often have complex comorbidities.[98,99] In some instances there is no effort made to adjust for case mix, for example the SimPatIE project catalogues 'death in low mortality DRG [disease resource group]' as not requiring any risk-adjustment.[104] Risk adjustment is clearly important and, without it, some of the indicators may be redundant:

> *"the level of the PSIs which monitor harm (e.g. rates of mortality and infections) is usually sensitive to bias caused by the severity of the disease, comorbidities and lifestyle factors. Validity of these indicators is therefore depend on simultaneous collection of patient-related data appropriate for risk adjustment. Empirical evidence suggests that administrative data are not satisfactory in this respect especially when used at clinical (hospital) level"*(p14).[65]

This assessment by Kristensen and colleagues paints a particularly negative picture and is not backed up with any empirical evidence on the effect of such bias.

2.3.5. Accuracy of coding

2.3.5.1. Errors and gaming

Administrative data are collected through the coding of medical records which are the notes made by clinicians on the care given to patients. Therefore the accuracy of the dataset depends on the standard of documentation and accuracy of coding. Also, the primary uses of the administrative data for monitoring and reimbursement might provide a potential incentives for 'gaming'[34] and make it *"unreliable because of coding problems"*(pi9).[105] The issues on coding was raised as far back at 1994 by Romano and Mark[106] and has been reiterated often since.[83,99]

2.3.5.2. Hospital variation

Further to general issues of coding, there is evidence of significant differences in the standard of coding within and across hospitals.[40,81,107] The same study that identified the factors influencing coding standards found that HRG errors – i.e. the degree to which the correct payment was made – ranged across NHS hospital trusts from 0.3 to 52%.[108] Factors that might affect the standard of coding in different hospitals include the depth of coding used (there is considerable variation by hospital in the mean number of codes used per record),[93] level of clinicians involvement, and standards of validation.[108]

2.3.6. Some have not been applied to ICD-10

As the older ICD-9 coding framework is replaced with the revised ICD-10 system, likewise the algorithms used to flag quality and safety events need to be updated. A group of researchers and users of health administrative data from Canada, the United States, Switzerland, Australia, China and

the United Kingdom came together in June 2005, and identified the translation of indicators to ICD-10 as one of the high-priority methodological research areas.[109]

2.3.7. Limited evidence on effectiveness

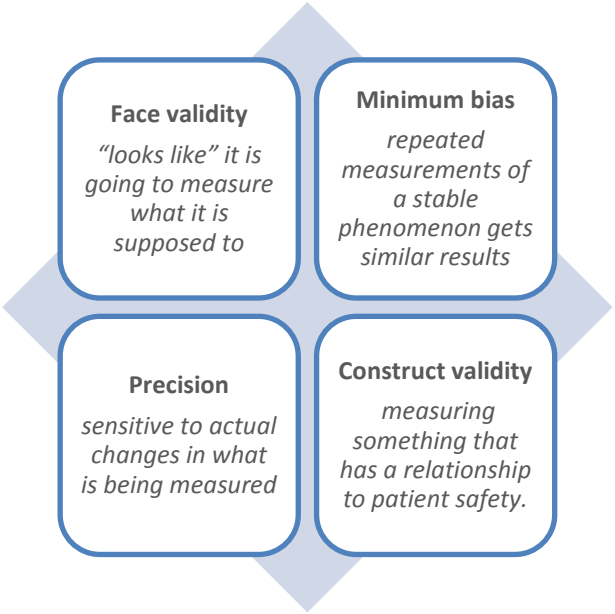
There is limited evidence on whether application of such indicators is effective.[67] This may be a combination of lack of time lag to evaluate this relatively recent development, and difficulties to account for bias created by changing data recording practices over time. However, some of the positive findings to date are drawn out in the discussion (Chapter 11).

2.4. Review of validity of indicators

2.4.1. Defining validity of indicators

Validity is a measure of the effectiveness of indicators. In the project to refine the original AHRQ indicators (HCUP II QI), the following criteria were used to judge the validity of the indicators (Figure 3).[74] These aspects of validity are not comprehensive with, for instance, no assessment of content validity which refers to the extent to which a measure represents all facets of quality or safety being assessed. Moreover, these dimensions of validity do not necessarily carry equal weight, with this being a function of the method being applied. For instance, *"reliability is important when using an indicator to make comparisons among groups or within groups over time"* (p524).[33] A more detailed assessment of validity is given in the following Chapter.

Figure 3. Dimensions of validity used by AHRQ



2.4.2. Previous literature on validity

2.4.2.1. Examples of validation results

Findings from specific studies on validity are not necessarily generalisable due to, for instance: most assessments covering only a subset of the dimensions of validity; differences in data accuracy and completeness between different datasets. Most evaluation have focused on patient safety measures, presumably as the events being investigated are, possibly more significant by nature, often more identifiable than quality differences so making validation easier. A review of literature found that validation of existing indicators had been largely restricted to the USA and the validity of many individual indicators remains unresolved.[88] Of these publications on aspects of validity, the results are mixed in terms of techniques used and the resultant findings, for example:

- by comparing two reporting systems and conducting a case note review, one study found that one of the AHRQ PSIs (PSI-7, selected infections) performed poorly;[110]
- another study compared an indicator based on administrative data to case note review and found that the former identified 23.3 decubitus ulcers per 1000 patients, to a true rate of 6.14, a 74% variance;[111] and
- in the UK, out of 2,150 medical records that were flagged as having an adverse event by a PSIs, 72.5% were found to match the (translated) AHRQ PSI specification.[112]

2.4.2.2. Difficulty in generalising

Different indicators are susceptible to different bias or errors and their validities will vary considerably. The possible exception is face validity, which is less dependent on the actual dataset being employed and, in this case, Rosen and colleagues found that several studies had demonstrated that such indicators have reasonable face validity.[113] Despite this difficulty in generalising findings, some academics inadvisably made sweeping conclusions. A symptom of the issue is the contradictory nature of some conclusions, with some stating that indicators have demonstrated good reliability, based on the consistent results found across studies [113,114] and across years [73,115] whereas, for instance, Naessens and colleagues concluded that "in general, many patient safety indicators do not reliably identify adverse hospital events" (p781).[101]

2.5. Potential applications of indicators

Whilst there are certainly some limitations to the current array of QIs and PSIs, these failings do not rule such indicators either irrelevant or unusable. Rather the floors just limit the extent to which the indicators can be appropriately used. Rivard, Rosen et al noted that "to address the question of how best to use patient safety indicators, one needs to assess both what the indicators can validly be used for as well as what purpose people may want to derive from them"(p1644).[73]

2.5.1. Types of reporting

2.5.1.1. Framework for describing reporting

In general, performance indicators can be used for four different functions: facilitating accountability; monitoring performance; as a learning tool; and forming policy initiatives.[58] In exploring the purpose of indicators in comparing performance of NHS hospital trusts, one paper categorised utilisations of indicators through the type of information they can provide: either 'hard' (formal and quantitative) or 'soft' (informal and subjective) information.[116] This is covered in more detail at the end of the thesis.

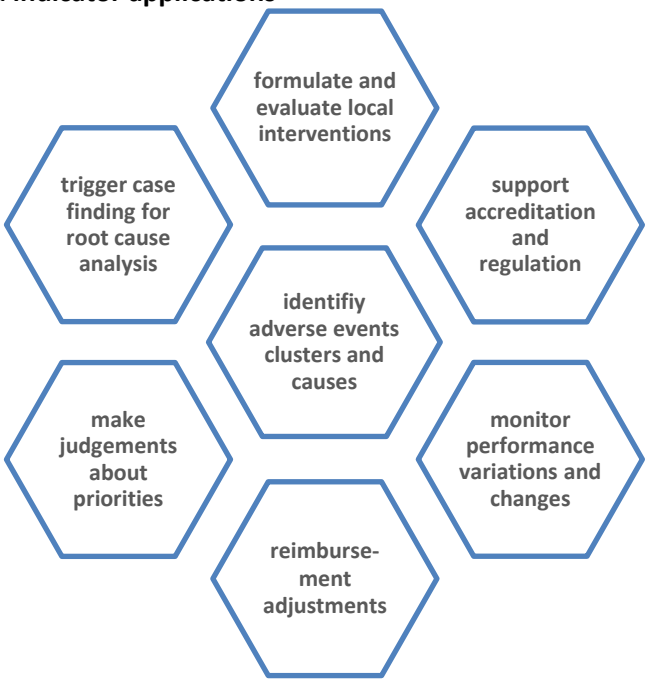
2.5.1.2. Recommended reporting type

The WHO has recommended that reporting systems should be blame-free, concluding that successful systems are: non-punitive; confidential; independent; analysed by experts; credible; timely; systems-orientated; and responsive.[97] This conclusion was echoed, in part, by the team constructing a national set of indicators in Denmark which called for reporting systems to be: confidential, sanction-free, learning oriented.[117]

2.5.1.3. Previously recommended applications

As well as setting conceptual frameworks on the use of such indicators, the literature also contains many examples of specific suggestions for these indicators (Figure 4). For example, one paper suggested that "their simplicity and reliability make them valuable as a higher-level safety performance measure. They offer one means for coordination and integration of patient safety data and activity within and across organizations" (p1633).[73]

Figure 4. Suggested indicator applications



Source: Garcia Armesto (2007),[34] Mainz (2003),[33] Perneger (2006),[118] Rivard (2006),[73] Kristensen (2007),[65] Weingart (2003),[71] Weeks (2008),[119] Bahl (2008),[99] Pulcins (2004),[120] Coffey (2005),[121] Culler (2007),[122] Menachemi (2007),[123] Li (2007).[124]

2.5.2. Applying sets of indicators

The case for bundles of indicators was introduced earlier (paragraph 1.4.1.1, p32) with specific advantages set out below. An explanation of the domains of validity is given in paragraph 3.2.1.2, p64.

2.5.2.1. Improving content validity

To draw on the analogy of the blind men and the elephant, the story – in some of its interpretations – tells of a number of blind men, each grasping at an elephant but all touching different limbs. The result being that the man holding one of the elephant's legs makes an entirely different inference of what the object is to his peer grasping its trunk. Whilst none of the men are able to deduce alone the full nature of the elephant, with enough hands and communication between them a correct identification can be made. This holds true with these indicators which cannot alone give a whole

picture of patient safety, however, a range of indicators used in conjunction with other measurement methods could give a useful depiction.

2.5.2.2. Allowing additional validation techniques

Using a set of indicators also provides the opportunity to undertake additional validation. Specifically, the validation is based on the hypotheses that:

- Two similar measures of the same construct should yield similar results (construct validity)
- Process and outcome measures evaluating the same aspect of care should be correlation (convergent validity);
- An indicator should be associated with "better" measures of performance (criterion validity)

2.5.2.3. The Donabedian model revisited

As explained earlier (paragraph 1.2.1, p24), the commonly cited model by Donabedian regarding patient safety incidents describes the interplay between structure, process and outcomes. By using appropriately focused set of indicators, these various factors can all be investigated, to some extent, to help epidemiological study design.

2.5.3. Primarily organisational-wide indicators to date

2.5.3.1. Legacy of broad indicators

To date the focus of performance monitoring using administrative data has been organisation-wide, cross-specialty indicators, in particular hospital standardised mortality rates (HSMRs). This phenomenon was cited in the rationale for developing the first major set of quality indicators (CSP, Iezzoni et al 1992 [50]) however recent publications have concluded that this limited focus remains the same.[49] A key example of the prominence of these broad measures are the well-publicised HSMRs based on routinely collected hospital episode statistics (HES) for the major acute hospitals in England, which have been published since January 2001 and, indeed, the monitoring of organisational-wide indicators has coincided with some improved results, such as decreases in recorded hospital-standardised mortality rates.[125]

2.5.3.2. Limitations

However, there are substantial limitations to such broad indicators. Given the widely held belief that it is important to report performance indicators to clinicians (as discussed in more detail at the end of this thesis, paragraph 11.5.1, p260) it is surprising that researchers and regulators have persisted in relying predominantly on organisation-wide indicators of performance which have only limited

relevance to an individual clinician. This situation is epitomised by a recent, comprehensive review of PSIs which purposefully excluded studies "primarily on specific disease, diagnoses or treatments"(p14).[88]

2.5.3.3. Application of indicators to specific cohorts

There has been some development of indicators for particular cohorts of patients, such as the long-standing AHRQ Paediatric Quality Indicators which have been available in January, 2006, contains indicators that apply to the special characteristics of the paediatric population. Similarly the first AHRQ HCUP indicator set reported mortality for different groups.[78] Similarly, indicator sets not exclusively based on administrative data – such as the Danish NIP – have commonly included some specialty-specific indicators.[33]

2.6. Advantages of reporting specialty-specific indicators

The following section draws on both the limitations of administrative data (described in this Chapter) and those generic limitations in existing measurement (described in the previous Chapter) to show that specialty-specific indicators offer the potential to overcome these key challenges to some extent. Specialty-specific measurement is topical: as suggested in 2012, "after recent UK policy developments, considerable attention has been focused upon how clinical specialities measure and report on the quality of care delivered to patients".[126]

2.6.1. Introduction to concept of specialty

2.6.1.1. Definition of specialty

'Specialty' refers to a branch of medical science, such as urology or paediatrics. For the purpose of this project, formulating strict criteria of the term specialty is not necessary since the findings on feasibility are generalisable to either broader or narrower groupings. The specialties investigated are defined in more detail in the relevant sections.

In reality, the case put forward and empirical findings on the benefits of specialty indicators hold for those indicators pitched at a lower-then-hospital and higher-than-clinician level.

2.6.1.2. Differentiating categories of specialties

Specialties can be classified according to whether they are:

- surgical *(i.e. an important part of diagnosis of treatment is achieved through major surgical techniques)* or medical *(related to internal medicine; i.e. important part of diagnosis is never achieved through surgery)*;
- the characteristics of patients treated *(e.g. paediatricians treat children)*;
- diagnostic *(e.g. radiology)* or therapeutic *(e.g. geriatrics)*;
- organ-based *(e.g. cardiology)* or technique-based *(e.g. anaesthetics)*.[127]

2.6.2. Introduction to specialty-specific indicators

2.6.2.1. Demand for specialty-specific initiatives

Although few in number, there are specialty-specific systems for addressing patient safety and, where these have been evaluated, some have shown promise. For instance, Wu and colleagues were able to use a specialty-specific system to analyse patient safety issues and, as a result, implement a number of safety initiatives.[128] *"They suggest that, in time, specialty systems such as their own must be integrated with other reporting systems, such as risk management, State or national systems, to minimise duplication of effort and reduce the burden on the staff"*(p69).[20] In fact, the author of the 1999 IoM report that was integral to increasing the priority of patient safety has suggested that specialty-based, focused reporting programs should be developed by other specialties.[129]

2.6.2.2. Previous specialty-specific initiatives

Previous papers on quality and safety indicators have disaggregated the indicators by specialty or diagnosis. In 2006, Mattke and colleagues recommended the use of 12 mental health, 17 cardiac care and 9 diabetes indicators,[130] whilst another paper recommended the use of indicators categorised by diagnosis: stroke, heart failure, schizophrenia, acute gastrointestinal surgery, heart failure, and lung cancer.[56] In the AHRQ HCUP II indicator set, the following areas of healthcare were given specific, separate consideration: cardiovascular, cerebrovascular, other vascular, geriatric, obstetric, paediatric, chronic conditions, diabetes.[80] Similarly, the SimPatIE indicator set included the AHRQ obstetrics indicators and five single indicators for specific diagnoses.[65]

2.6.3. Difference between specialties

The following section highlights some key variations in specialties which suggest that cross-specialty indicators have specific limitations and, as a result, disaggregating performance measurement tools into individual specialties carries advantages.

2.6.3.1. Different cultures

Some specialties have more developed safety cultures. This variation can, in part, be traced historically. The British Ministry of Health set up a committee to examine morbidity and mortality during child birth in the 1930s. Consequently, a local programme was established, involving antenatal clinics, meetings between midwives and family doctors, a new specialist ward and senior hospital clinician post, which reduced the rate of mortality in that area from 9 to 1.7 per 1000 deliveries within 5 years.[20] This area of medicine has continued to pioneer, in relative terms, patient safety.[20] And the importance of this patient safety culture is highlighted elsewhere.[131]

2.6.3.2. Different prevalence

There is no reason to assume that patient safety is an equally important issue, with similar prevalence, across different aspects on healthcare. For example, the effects and causes of adverse events in elective surgical admissions are, by their nature, more easily identified. Evidence suggests that there are significant variations between specialties in the number of adverse events occurring. For example, in the specialty of mental health, analysis of existing PSIs suggests that people with schizophrenia have higher chances of suffering from post-operative respiratory failure, deep vein thrombosis and sepsis. The same study concluded that there is scope for further work to fill in the gaps in knowledge about adverse events within the specialty of mental health.[132]

As a result, indicators are likely to identify more events in some specialties. Analysis using the AHRQ PSI algorithms suggested that 31% of adverse events identified occurred in obstetric hospitalisation although this will be biased to some extent by the selection of indicators.[98]

2.6.3.3. Different impact

Similarly, there is no reason to assume that PSIs are likely to result in equivalent levels of harm between specialties: anaesthetists' mistakes, for example, may be more serious than those of doctors in other specialties;[14] and in critical care, due to the high number of patients with complex and severe illnesses, the effects of any adverse event are more pronounced [133]. As a result of these variations, it would be preferable to prioritise some specialties and disaggregate between those specialties with different levels of impact from patient safety events.

2.6.3.4. Applicability of coding framework

The framework used for coding medical records to create the administrative dataset on which PSIs are based is not comprehensive (as discussed earlier in §2.3.3). As a result, the framework will be naturally better suited to accurately identify adverse events in some specialties. For instance, a study that conducted a case note review to check the validity of results from applying the AHRQ PSIs to

NHS data found that the highest levels of agreement were found with the indicators relating to obstetric care.[112] Consequently, previous work on developing indicators has focused on certain specialties. For example, a project covering hospitals in Queensland, Australia chose to focus on mental health, stroke, emergency departments, COPD and medication.[134]

2.6.3.5. Maturity of existing tools for monitoring

Further to the example of Wu et al, some specialties already have established tools for monitoring adverse events. "These systems are designed to provide information on specific clinical issues, which can be shared within the professional group" (p68).[20] In particular, anaesthesia, neonatal and adult intensive care units have established specific systems.[20,128,129]

2.6.4. Nature of specialties

This section discusses the properties of specialties, in particular as groups of clinicians, that make specialty-specific measures, in theory, appropriate.

2.6.4.1. Group learning

The patient safety literature has underlined the importance of group learning as well as organisational and individual learning.[73] McIntyre and Popper highlighted that "experiences need to be pooled so that doctors may also learn from the errors of others"(p1919).[53] Specialties offer such a group which already have existing channels, such as through morbidity and mortality conferences, to be analysing these initiatives.

2.6.4.2. Specialty cultures

One of the keys to improving patient safety is to ensure that the health workforce develops the will to address the issues.[135] This was shown empirically by Kline and colleagues who found that a more positive culture of patient safety within hospital units was related to lower incident severity.[136] A review of patient safety culture literature performed in 2007 also found that clinicians are a key factor in the patient safety culture.[137] Reporting to clinicians can, itself, be seen as an initiative for improving the patient safety culture. Further to the importance of clinicians in improving the patient safety culture, the same reviewers suggested that groups of people, termed as "sub cultures", such as specialty teams, are also important.

2.6.4.3. Understanding epidemiology and proposing changes in practice

Once a patient safety issue has been identified, a suitable response needs to be developed. Some mistakes in healthcare are as a result of complex processes and structures, such as in medication which involves numerous stages.[138] Given these complexities, specialty teams, and clinicians within them, are the appropriate recipients of the information as they can analyse any issue in the

appropriate context. Certainly, there is recognition that addressing patient safety requires in-depth involvement of clinicians. Leggat and colleagues, for instance, point out the importance of the skill and capacity of healthcare workers in delivering safer patient care.[135] Involving clinicians has been shown to be effective; for example, involving clinical teams in the monitoring of VTE resulted in a number of practical changes to working practices being made.[139]

2.6.4.4. Narrowed focus

There are currently gaps in our knowledge about the epidemiology and effects of adverse events. However, it is clear that to improve patient safety these gaps in our understanding need to be addressed. Mayers-Oakes, in a paper which outlined research in breast cancer, diabetes mellitus, upper gastrointestinal and lower intestinal, concluded that this need could be met by narrowing the focus of indicators.[140] The implication from this conclusion is that concentrating indicators on specialties rather than, say, pan-hospital issues would be beneficial. Certainly, initiatives with less focus have not always been successful. For instance, Weiner and colleagues suggested that hospital-wide initiatives have not always improved patient safety.[141] This narrowed focus can also help improve interpretability and increase validity by selecting either homogenous patients or areas where less case-mix adjustment is needed.[65]

2.6.4.5. Developing new and existing indicators

The increased demand for indicators (1.4.2) that has precipitated from the raised profile of patient safety should be met, at least in part, by developing and validating more indicators.[73,130] Clinicians and specialty teams are well placed to help address this need due to their knowledge and experience. Further to meeting the need for additional information, developing additional indicators would also:

- reduce the focus on particular diagnoses and whether they are coded. This would reduce any ability and incentive to 'game' by not coding accurately or comprehensively; and
- "create the opportunity to test correlations and identify factors from among indicators. This in turn would enable the indicators to depict patient safety on a number of distinct dimensions"(p1647-8).[73]

2.6.5. Disadvantages of specialty-specific

Many indicators will be specific to that specialty and therefore not applicable to other specialties. One implication is therefore that it is difficult to compare results between specialties and create, for example, control groups for evaluations.[142] One solution is to create other counterfactuals by comparing between patient groups, hospitals or time-periods.

By being more focused, the indicators may be less likely to measure aspects of the treatment and outcomes that are not anticipated; conversely, broader measures have a higher propensity to detect unexpected issues.[143] However, increasing the breadth of indicators will usually make controlling for bias more difficult.

2.7. Research objectives

2.7.1. Hypotheses

The use of hospital-level cross-specialty indicators is relatively well established. However, there is scope to expand the use of administrative data, in terms of: using more focused indicators; applying and validating the measures in novel ways; and introducing them to new healthcare services. As such, the project was designed to test the following hypotheses:

1. A specialty-specific approach to developing indicators can result in a more useful and valid bundle of indicators to identify quality and safety issues;
2. Hospital-level comparisons could be improved through better validation techniques (coding practice, indicator comparison and organisational factors); and
3. The temporal nature of administrative data can be used to overcome some of the limitations of such indicators and highlight important quality and safety issues.

2.7.2. Thesis outline

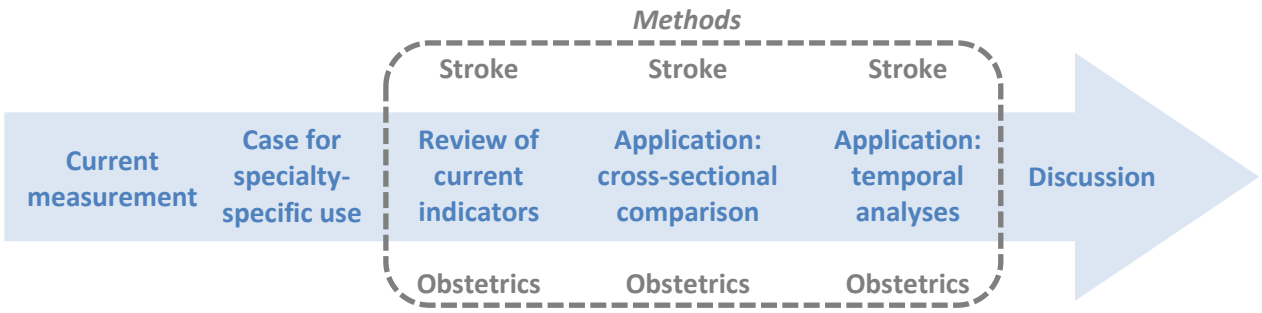
To test these hypotheses, a range of qualitative and quantitative methods were applied (Table 5).

Table 5. Methods table

Hypothesis	Chapter (Key method)
1. A specialty-specific approach to developing indicators can result in a more useful and valid **bundle** of indicators to identify quality and safety issues	**Chapters 4 and 8** (Systematic reviews)
2. Hospital-level comparisons could be improved through more robust **validation** techniques (coding practice, indicator comparison and organisational factors)	**Chapters 5 and 9** (Logistic and multi-level regression modelling)
3. The **temporal** nature of administrative data can be used to overcome some of the limitations of such indicators and highlight important quality and safety issues.	**Chapters 6, 7 and 10** (Logistic and multi-level linear modelling)

Chapter 3.

Project design and methods



Overview

Context

The previous two chapters have set out the case for further work in evaluating the feasibility of using specialty-specific indicators based on administrative data. This chapter presents the analytical approach taken. The design of the individual studies on the two specialties is also driven by the specific literature reviews (Chapters 4 and 7) and the implications on the methods are described in these later specialty-related sections.

Methods

Review of existing analytical approaches to developing and utilising indicators based on administrative data.

Findings

Hospital stroke care and obstetrics are appropriate specialties to use as case studies – due to their importance, evidence of scope to improve, and relative differences – for demonstrating feasibility of applying specialty-specific indicators. A range of methods have been used to identify, develop and present indicators.

What this chapter adds

Set out the analytical opportunity for this new concept of specialty-specific indicators Present a new hybrid approach to evaluating indicators

3.1. Introduction to study design

3.1.1. Introduction to feasibility

Now that the case for specialty-specific indicators based on administrative data has been made, the next step is to ensure that they can be applied in a robust fashion. "It is imperative that clinical indicators are meaningful, scientifically sound, generalisable, and interpretable. To achieve this, clinical indicators must be developed, tested, and implemented with scientific rigour"(pi5).[144] This project investigates the feasibility of using specialty-specific indicators based on administrative data to meet this need.

'Feasibility' is defined, for this project, in broad terms including not only the various dimensions of validity but also the ability to apply these indicators. It is tested here using a reactive approach, namely applying analytical tests to address the gaps in current knowledge (such as assessing the strength of correlations between indicators as a marker of validity) while drawing on previous literature where this is generalisable (such as comparing findings from the indicators to similar research).

3.1.2. Addressing limitation in indicator use

The previous chapter covered the literature on how indicators based on administrative data have been recommended for use, which is covered again in the final chapter. If one accepts the premises that (a) administrative data hold the potential to improve the measurement of quality and safety and (b) the misapplication of administrative data would be a retrograde step, then an implication is that it would be worthwhile to focus the application on mitigating the extant limitations. As such, the studies contained in this monograph are designed to show how to best overcome or evaluate the key limitations from the existing literature, as outlined in Table 6.

Table 6. Recommended study design to mitigate limitations

Para.	Issue	*Detail*	Recommendation on study design
2.3.1	Causation	*Difficulty in attribution of sequence and causation*	**Range of indicator types:** Complement outcome measures with process indicators that inherently are more amenable to hospital performance. **Regression analysis between outcomes, activity and structural factors:** Investigate associations between hospital performance and other hospital factors.
2.3.2	Judging preventability	*Difficulty to assess responsibility and preventability*	**Using comparative analysis:** Evaluate differences between provider performance and longitudinal trends, complemented with statistical analyses, to estimate achievable performance.
2.3.3	Coding detail	*Problems in coding framework not identifying key events*	**Systematic review of existing measures:** Comprehensive evaluation of what indicators exist and which aspects of care can be evaluated.
2.3.4	Risk model	*Need for case-mix adjustment*	**Case-mix adjustment:** Include detail on the risk model **Specialty-level:** Limit to less heterogeneous groups of patients
2.3.5	Coding errors	*Variation in coding practice and accuracy*	**Sensitivity analyses:** Re-run analysis to evaluate potential effect of differences in coding practice and accuracy
2.4.2	USA-bias	*Most of the studies originated from the USA using ICD-9*	**Apply to ICD-10 dataset:** Use a non-USA dataset based on the current ICD-10 coding framework

In summary, the recommendations above are that the individual studies' elements: focus on specialties; apply a bundle of indicators including outcomes and processes; investigate correlations between performance and organisational factors; compare performance between hospitals and over time; outline the case-mix adjustment explicitly; and conduct sensitivity analyses to understand effect of variation in coding practice and accuracy. The specific analytical framework used – which encompasses the addressing of existing shortcomings – is described below.

3.2. Analytical framework

3.2.1. Introduction

3.2.1.1. Issues with validation

It is not possible to have a perfectly validated indicator; even if an indicator's performance was compared to an absolute gold standard then this would be snapshot and not necessarily provide

assurance to all time periods. Some domains of validity "are easier to tap than others; unfortunately, the most challenging domains to tap are often the most useful for health care providers who wish to understand the meaning of the data."(p4)[145] In particular, lack of a gold standard means that 'criterion validity' is not possible. The domains of validity described here are not ordered in a hierarchical fashion; indeed, it is not possible to rank the relative importance of the criteria. Across the domains there are inherent ambiguities and little consensus on their meaning.[142]

Concepts of validity are poorly described in literature on patient outcomes, this is not due to a lack of theoretical discussions but a lack of consistency.[142] I have therefore formulated an analytical framework to evaluate validity and other aspects of feasibility using literature on psychometric testing and patient-based outcome measures and present an interpretation below.

The framework does not represent a comprehensively exhaustive test and, indeed, it would be inefficient to apply every method suggested for taking assurance on validity. In the editorial introduction to the first set of indicators – CSP, Iezzoni et al (1992) – the consensual approach involving a panel of physicians to provide face validity and precision, was questioned as not being replicable. In particular, the physicians did not agree on their interpretation of the presence or not of an adverse event within the medical note abstracts when trying to validate measure results on CSP.[50]

3.2.1.2. Aspects of validation

As discussed at the start of this chapter (paragraph 3.1.1), a key aspect of feasibility is to ensure the indicators are valid (the other aspects of feasibility – and how they are considered – are given later in this chapter). Validity has a number of domains, although these are neither comprehensively exhaustive nor mutually exclusive (Table 7). To organise the plethora of concepts, I have grouped the methods for demonstrating validity (individual 'components') within four global domains (construct, content, face, and criterion).

Table 7. Domains of validity

Validation domain ('global definition')	**Requirement**	**Validation methods** ('component definition')
Construct validity	Two measures of the same construct should yield similar results	**Convergent** validity: compare process and outcome measures **Discriminant** validity: compare to supposedly unrelated measures Comparison to structural factors, e.g. skill mix
Content (logical) validity	Indicator should represent all facets of the subject of evaluation	**Consensual** validity: near consensus of an *expert* panel (for content) or *multi-disciplinary* panel (for face validity). These domains are relatively less setting/purpose-specific, so previous evaluations can often be generalised.
Face validity	Indicator should "look like" it is going to measure what it is supposed to measure	
Criterion (concrete) validity	Indicator should be associated with "better" measures of performance	**Concurrent** validity: comparison to previously validated measure **Predictive** validity: (similar to concurrent) comparison of predicted score and eventual results from validated measure

Source: adapted from concepts described in Romano et al (2007)[145] and Brewer (2000)[146]

3.2.1.3. Study validation

In addition to the aspects of indicator validity, there are domains of study validity. Key domains are: **ecological** validity (that the study design approximates the real-world); **internal** validity (the extent to which a causal conclusion is correct, involving minimising of systematic error); and **external** validity of study, which relates to how generalisable the results are.[146]

3.2.2. Previous indicator validation

Evaluations on how 'good' an indicator performs usually involves the notions of 'validity' and 'reliability'; however, these terms are used in many ways in the literature.[147,148] Some of this variation in the evaluation of indicators might be appropriate since how 'good' an indicator is depends on the context of both the subject and purpose of the measurement. To mitigate this issue of differing contexts, the framework used to evaluate the indicators in this project is based on three previously endorsed frameworks used to evaluate (albeit non-specialty) quality and safety indicators based on administrative data, alongside the domains of validity taken from wider literature:

1. the Danish NIP, where ideal indicators were characterised as being: based on agreed definitions, and described exhaustively and exclusively; highly or optimally specific and sensitive; valid and reliable; discriminatory; relates clearly identifiable events for the user; useful for comparisons; evidence-based.[33]

2. the refinement the HCUP QIs, whereby indicators were required to have: face validity, precision, minimum bias, construct validity, ability to foster quality improvement, should or could be used effectively.[80]

3. OECD HCQI indicator project used: importance (impact on health, policy importance, amenability); scientific soundness (face validity, content validity) and feasibility (data availability, reporting burden).[85]

3.2.3. The analytical framework

3.2.3.1. Methods for evaluation

I synthesised these frameworks into a single hybrid set of criteria that were used to define the project protocol, as listed in the final column (Table 8, with further details on how the specific methods are designed to address this framework given in paragraph 3.6, p77). This is a pragmatic approach built on making efficient use of the data and tools available, since previous efforts – including those of Romano (2007) showed that even a comprehensive assessment can be inconclusive with, for instance, an indicator for obstetric trauma having mixed validity: "Recent evidence on construct validity of this indicator is inconclusive, but the evidence on criterion validity is quite supportive"(p18).[145]

Table 8. Analytical framework

	Domain	Indicator requirement	Study design
Importance	1. Impact	Addresses clear gap, with policy importance	**Systematic reviews** to find potentially important applications. Identifying key variations in indicators performance.
	2. Amenable	Relates to an outcome that can be influenced by healthcare performance	Comparison between key results and **organisational factors**.
Scientific soundness of indicators	3. Face and content validity	Sound clinical or empirical rationale for its use, and represent all facets of the subject of evaluations	**Literature review** to identify indicators with pre-evaluated face/content validity. **Consultation** with experts on new measures (*consensual* validity).
	4. Precision	Relatively large variation among providers that is not due to random variation or patient characteristics. *Includes responsiveness, discriminative (cross-sectional) and evaluative (longitudinal) validity.*	**Hospital results** displayed on **funnel plots** to identify outliers while accounting for random chance.
	5. Construct validity	Supported by evidence of relationship with quality, and related to other similar indicators	Statistical analysis of **correlations** between the set of indicators, including **outcome v. process** indicators (*convergent* validity).
	6. Minimum bias	Unaffected by systematic differences in case-mix	Detailed **risk-adjustment model** applied to indicators and **stratification** of key results.
		Not incentivise gaming of the indicator	**Sensitivity analyses** of influence of coding error and practice.
Applicability	7. Data availability	Applicable to administrative data	**Application** of indicators to show ICD-10 based data is sufficient.
	8. Reporting burden	Minimal reporting burden	**Application** of indicators to show that results can be derived without significant burden.
	9. External and ecological validity	Generalisability of findings	Explicit statement of **assumptions** used.

Sources: Adapted concepts presented in Davies et al (2001);[80] Mainz (2003);[33] Millar et al (2004);[85] Romano et al (2007);[145] and Brewer (2000).[146]

Iezzoni and colleagues concluded that indicators "needed to be validated if they were to be used; however, what was not obvious was what constitutes validity in this context"(p366).[50] Some aspects of validity and applicability not covered in the evaluations framework – such as the ability of the indicators to be effective and "work well" in a portfolio of quality and safety metrics – are, instead, accounted for in the overall discussion (Chapter 11) since they relate to the reporting system. In particular, "internal validity" (correct causation) is partially proven through tests on convergent validity which may show, for instance, that providers offering more of a certain treatment or with a

different skill mix have better performance but the causation is also appropriate for assessment by clinicians and, therefore, relates to the reporting system.

3.2.3.2. Limitations to evaluation protocol

The key limitation of this validation process stems from no "gold standard" against which the results can be compared (criterion validity). On the precision element, given the inability to know absolutely whether a quality or safety event has occurred, calculating indicator-level false positive and false negative levels - as suggested in some validation frameworks, such as Mainz 2003[33] and Rosen 2007[149] - is not possible. However, I have mitigated against this by accounting for hospital-level false positive rates using statistical controls and, where available, drawing on previous research, such as the work by Bottle et al (2008) which found, for instance, that from the 2,150 records audited, 72.5% were found to match the PSI specification.[112]

3.3. Specialty selection

3.3.1. Selection framework

Given the aforementioned case to develop and apply specialty-specific indicators, the first stage of the study was to select specialties. Within the NHS, specialties are not mutually exclusive and there is no single agreed list of recognised specialties. However, given the nature of the research presented here, there was no need for precision in the labelling of specialties. By comparing previous lists of specialties (curriculum list, Royal Colleges of Physicians, Surgery, Paediatricians and Child Health, Obstetrics & Gynaecology, Psychiatrists, Anaesthetists, Radiology, Ophthalmology and the Faculty of Occupational Medicine and College of Emergency Medicine), I created a list of 51 recognised specialties within the NHS, excluding general practice and public health which are not hospital-based specialties.

I took a pragmatic approach to selecting which specialties should be evaluated, starting with the domains of 'Impact' and 'Amenable' from the evaluation framework (Table 7, p65). These criteria focusing on importance – or ensuring that there is scope for improvement - have been expressed elsewhere.[144] Given the number of specialties, I took a range of additional considerations in selecting the specialties. To increase the generalisability of the findings on applying such indicators, I purposefully chose specialties with differences, particularly in:

1. culture and current initiatives to improve quality and safety, in particular referring to the NPSA which then had research programmes into four specialties: anaesthesia, neonatal, obstetrics and gynaecology, oncology;[150]

2. the range of treatments and outcomes;
3. extent to which the policy of choice of provider is potentially applicable;
4. the extent to which administrative data has previously been applied, including the availability of suitable indicators was considered, although the research plan did allow for the development of new indicators.

I excluded specialties with only rare events since these are less likely to have a sufficient evidence-base to have robust guidelines and, as such, there would be difficulty in knowing which processes might be improved by monitoring.[144] The decision was made in consultation with a public health doctor.

3.3.2. Selected specialties

Based on the framework, I chose to proceed with stroke and obstetrics specialties. This is not the only pair of specialties that comply with the criteria but uniqueness was not a pre-requisite and proceeding with these two are sufficient to meet the goals of the project with any wider assessment would be outside the scope for a doctoral thesis. The differences - and combined coverage of different facets of healthcare - that led to this decision are summarised in Table 9.

Table 9. Differences in selected specialties

Criteria	Stroke	Obstetrics
Overall burden of disease (incidents/year in England)	110,000 stroke[151]	700,000 births [152]
Type	Medical	Surgical/Medical
Status of specialty	New	Longstanding
Characterisation of measurement	New and predominantly audit based	Longstanding with varied tools
Indicators in CSP (1992)	No	No
Indicators in HCUP QIs	No	Yes
Indicators in AHRQ PSIs	No	Yes
Indicators in Danish NIP[144]	Yes	No
Specific NPSA initiative	No	Yes
Prevalent	Unknown	Yes
Evidence of scope for improvement	Yes	Yes
Specific patients	No	Yes
Patient type	Unwell patients	Predominantly well patients
Choice of provider	Rarely	Potentially

3.4. Developing an indicator set

3.4.1. Framework for developing indicators

Jan Mainz, from the Danish NIP, presented a conference paper in 2001[60] – later published in 2003 as a companion piece to an article defining and characterising the NIP measures [144] – outlining a method for developing evidence-based clinical indicators. Whilst this framework gives only the shortest consideration of the testing of the indicators by application to the data, it provides a useful guide for the initial stage of the project, and the way in which this project complies with the proposed method is summarised in Table 10.

The OECD indicator set which was published at a similar time used a far less comprehensive method for developing an indicator set, involving an expert panel with each member asked to identify 20 measures they felt had the greatest prospects of being selected. Through a series of conference calls and email discussions, the Patient Safety Panel converged on a final list of 21 indicators.[85]

Table 10. Development of indicator sets

Phase	Steps	Sub-steps	Method applied in this study	*Chapters*
Planning	1. Choose the clinical area to evaluate	Establish importance (high volume, cost, variation)	Systematic reviews (secondary purpose)	*4, 8*
		Identify opportunities for clinical intervention	Systematic reviews (secondary purpose)	*4, 8*
	2. Organise the measurement team	Select group participants	Various consultations/ collaborations with clinical specialists, statisticians, and clinical coders	
		Organise and divide tasks		
Development phase	3. Provide an overview of existing evidence and practice	Present documentation and knowledge from the scientific literature for potential indicators	Systematic reviews (primary purpose)	*4, 8*
		Establish consensus about existing knowledge and practice	Systematic reviews (primary purpose)	*4, 8*
	4. Select clinical indicators and standards	Select process indicators	Predetermined criteria for selecting indicators, including that both process and outcome measures are used	*5, 9*
		Select outcome indicators		*5, 9*
		Identify prognostic factors (risk adjustment)	Systematic reviews (primary purpose)	*4, 8*
		Establish consensus and rating procedures	Stroke: online consultation Obstetrics: use of previously endorsed indicators	*5, 9*
	5. Design measure specification	Define indicators and standards	Described in application of indicators for comparing performance. Standards, cohort and risk-model developed through systematic reviews (primary purpose) and testing.	*5, 9*
		Identify target population		*5, 9*
		Determine inclusion and exclusion criteria		*5, 9*
		Devise risk adjustment strategy		*5, 9*
		Identify data sources	Use of English administrative data	*3, 5, 9*
		Describe data collection procedures	Process for collecting English data outlined below	*3*
		Develop a plan	The analytical plan and techniques are described below	*3*
	6. Perform pilot testing		Application of indicators according to analytical framework (Table 8, p67)	*3, 5, 6, 7, 9, 10*

Source: Mainz 2003[144]

3.4.2. Literature review type

3.4.2.1. Meeting requirements of previous indicator development frameworks

The literature review was designed to meet the following development steps set out by Mainz (and as described in Table 10, p71) to: present documentation and knowledge from the scientific literature for potential indicators; establish consensus about existing knowledge and practice; and identify prognostic factors (risk adjustment). To meet these objectives the review would need to be comprehensive and I employed the approach of a systematic review, which are used to collate "all evidence relating to the objectives and to reduce any bias in answering the literature review questions".[153] The literature review was also designed to meet the development steps to establish importance and scope for improvement, although these were not a core objective as they do not require such a robust approach.

3.4.2.2. Gaps in specialty literature

As set out above (Table 8, p67, *Domain 1*) a key component to addressing the feasibility of using specialty-specific indicators is to evaluate the current literature within that specialty and understand what the gaps are. Whilst some of the generic issues with quality and safety indicators are presented in the previous chapter, that discussion also highlighted the variation between specialties which will result in unique advantages and disadvantages for applying administrative data to monitor care.

3.4.2.3. Literature review type

These literature reviews were atypical in that they were not designed to answer a single question – such as the effectiveness of a specific treatment – but rather to understand which, and how, indicators have been applied to date. It was also hoped that the approach would reveal the current shortcomings in the development and utilisation of indicators. Despite the likely heterogeneity of the studies prohibiting formal, quantified meta-analyses, the literature reviews shared the features of a systematic, rather than narrative, review (Table 11). As such, these collations and appraisals of the literature are termed as systematic reviews in this monograph.

Table 11. Key distinctions between narrative and systematic reviews

Core Feature	Narrative Review	Systematic Review	Review protocol
Study question	Often broad in scope.	Often a focused clinical question.	Multiple, specific, non-clinical objectives
Data sources and search strategy	Which databases were searched and search strategy are not typically provided.	Comprehensive search of many databases as well as the so called grey literature. Explicit search strategy provided.	Comprehensive, reproducible search strategy.
Selection of articles for study	Not usually specified, potentially biased.	Criterion-based selection, uniformly applied.	Predetermined criteria applied.
Article review or appraisal	Variable, depending on who is conducting the review.	Rigorous critical appraisal, typically using a data extraction form.	Extraction form used.
Study quality	If assessed, may not use formal quality assessment.	Some assessment of quality is almost always included as part of the data extraction process.	Assess quality of both study (e.g. reporting of assumptions) and indicators (e.g. case-mix model).
Synthesis	Often a qualitative summary.	Quantitative summary (meta-analysis) if the data can be appropriately pooled; qualitative otherwise.	Likely qualitative synthesis due to heterogeneity of study designs and administrative data used.
Inferences	Sometimes evidence-based.	Usually evidence-based.	Inferences all based on extracted information.

Source: Adapted from West et al (2002)[154], Cook et al (1997)[155], Green et al (2008)[156]

3.4.3. Previous systematic reviews

To ensure that the work is not redundant, it is good practice to check whether there are existing or on-going similar reviews.[157] I reviewed existing and on-going reviews by searching the Cochrane Database of Systematic Reviews (CDSR)[c], key websites and making enquiries at key organisations involved in the development, application and validation of indicators. The additional purpose of this process was to both:

- inform the design of the specialty-specific systematic reviews by having a comprehensive understanding of what sources, search terms, inclusion criteria have been used to answer meet similar objectives; and
- use the descriptive results (such as number of indicators collated, dates of studies and countries of origin) from these previous reviews as benchmarks to compare to the results of

[c] Other sites searched included: National Institute for Health and Clinical Excellence (NICE); NIHR Health Technology Assessment (NIHR HTA) programme websites; Campbell Collaboration website; Evidence for Policy and Practice Information (EPPI) Centre website and it's database of systematic and non-systematic reviews of public health interventions (DoPHER).

the specialty-specific systematic reviews and, therefore, allow normative conclusions on the effectiveness of identifying indicators.

Details from key previous literature reviews of quality and safety indicators were collated (Appendix B, p297) with a summary of these previous reviews given in Table 12.

Table 12. Selected details from sample of previous review of indicators

Review	Date of publication	Search terms	Resources	Results
International Quality Improvement Programme (IQIP)	Since 1985	Not stated	Not stated	Nearly 700 indicators covering four care settings
HCUP QIs	1994/1998	Not stated	Not stated	Final list of 33 quality measures
National Indicators of Safety and Quality Project (Australia)	ACSQHC established in 2000	Not stated	Not stated	152 papers on health
AHRQ QIs	2001	MeSH terms "hospital, statistics, and methods" and "quality indicators"	Embase, Medline, hand search, author search	181 articles on potential indicators, with 27 explicitly defining novel indicator
Millar et al (OECD HCQI project)	2004	n/a (reviewed 7 preview indicator sets)	Based on previous indicator searches	21 indicators selected
SimPatIE	2007	Patient safety, indicator, risk, harm and test*, Usage/Use, Apply, Valid	Pubmed and google scholar	42 potential indicators with 24 recommended for all or part of Europe
CPSSQ (Imperial College, London)	2008	Two dimensional search string pertaining to [patient safety indicators] and [adverse events]	Included: Medline, Embase, grey literature, government websites	91 articles on patient safety indicators
Joint Commission's NHQM	2009 (v2.6b) based on work from 1999	Not stated	Not stated	ORYX initiative has over 8,000 measures. CMS chose 21 measures from shortlist of 39

Notes: NHQM: National Hospital Quality Measures; ACQHC: Australian Council for Safety and Quality in Health Care

3.4.4. Literature search resources

3.4.4.1. Electronic databases

Electronic databases were searched to identify both peer-reviewed journal articles and other research such as conference proceedings and unpublished work. There were two main processes regarding the use of electronic databases:

- An extensive systematic search of Institute of Scientific Information (ISI) Web of Science, Medical Literature Analysis and Retrieval System Online (Medline) and Excerpta Medica Database (Embase) databases – the latter two through the OVID interface – from inception to the date of the search.

- Searching databases of previously compiled indicators such as: AHRQ's Clearinghouse (which included measures used by, for instance, the Canadian Institute for Health Information, Australian Council on Healthcare Standards, British Medical Association);[158] OECD's Health Care Quality Indicators (HCQI) project;[130] the EU sponsored SIMPATIE project and EUPHORIC measures;[159] HCUP refined indicators;[80] and Queensland Health: Clinical Practice Improvement Centre measures.[160]

3.4.4.2. Other data sources

To complement the searches of electronic databases, I also used the following methods:

- An online search, including a selection of government and patient safety organisation websites (Appendix C), to identify grey literature, in the form of research and technical papers and reports, conference abstracts, government reports and committee working documents were obtained from this search.

- Searching bibliographies of key articles to identify additional papers which met the eligibility criteria.

The Physiotherapy Evidence Database was not searched as previous research found it did not include new citations relating to these indicators.[161]

3.5. Selecting indicators

3.5.1. Considerations on indicator selection

3.5.1.1. Previous indicator selection

AHRQ had a criteria that "in order to be retained as a potential PSI, at least one of the first three studies needed to demonstrate a positive predictive value of at least 75%, meaning that 3 out of 4 patients identified by the measure did indeed have the complication of interest".[40] However the lack of similar studies in the UK on PSIs let alone other specialty-specific indicators meant that such a strong criterion was not possible. As a looser criterion, AHRQ also stated that indicators should have been used effectively in the past, and/or have high potential for working well with other indicators currently in use. In the updating of the initial indicators, the HCUP II QI project has a criterion that at least 1% of hospitalised patients and/or 20% of providers are captured.[74]

3.5.1.2. Project protocol for indicator selection

The intentional differences in the chosen specialties dictated a different set of criteria for selecting the indicators and any specific criteria are presented in the relevant chapters. However, generic criteria based on the findings from previous literature discussed earlier are that indicators should:

a. constitute a bundle, defined here as at least 6 indicators;

b. include process and outcome indicators;

c. focus on an area with known variation in performance;

d. cover adverse events affecting at least 1% of hospitalised patients and/or 20% of providers.

Alongside the author, an expert in administrative data, a public health doctor and a senior doctor in the relevant specialty fed into the development of the indicator set. I acknowledge that the resulting set of indicators would not necessarily be unique; however, this does not affect the ability to determine the feasibility of applying such indicators. I have demonstrated how these are also aligned with the analytical framework in (Table 13), although the final three domains on applicability ('Data availability', 'Reporting burden', 'External and ecological validity') are not included despite these inherently being partially addressed by the characteristics of indicator bundles offering a wider range of information and yet being based on the same data source.

Table 13. Use of multiple measures and mix of process and outcome measures to address analytical framework

	Domain	a. Multiple measures	b. Process and outcomes	c. Known variation	d. Prevalent
Importance	1. Impact	More comprehensive understanding of impact	Understanding of both impact on hospitals (process) and patients (outcome)	-	Evidence of high impact
	2. Amenable	-	Associations between processes and outcomes suggest amenability	Suggests possibility to improve to peer performance	-
Scientific soundness of indicators	3. Face and content validity	Cover more facets of subject of evaluation	-	-	-
	4. Construct validity	Can test for associations between measures	Can test for associations between process and outcome measures	-	-
	5. Precision	-	-	Ability to identify differences	-
	6. Minimum bias	-	Process measures less likely to be affected by case mix	-	Higher numbers reduce effect of random variation

Notes: Sources and Notes in Table 8, p67.

3.5.2. Defining indicators

An established requirement is that the indicator is based on agreed definitions, and described exhaustively and exclusively.[33] The indicators used in this study are, therefore, set out in detail with key assumptions explained.

3.6. Applying indicators

In this section I describe the statistical methods common to the application of indicators to both specialties. The alignment of these statistical tests with the analytical framework is set out in Table 14, although again the final three domains on applicability ('Data availability', 'Reporting burden', 'External and ecological validity') are not included since these are addressed through the application itself rather than the results.

Table 14. Statistical tests to address the analytical framework

	Domain	Statistical technique	Output
Importance	1. Impact	Descriptive analysis	National rates of adverse events
	2. Amenable	Regression (performance v. structural factors)	Effect size and significance of correlation
Scientific soundness of indicators	3. Face and content validity	n/a	n/a
	4. Precision	Descriptive analysis	Range of scores, by group Mean number of cases, per unit
		Statistical testing	Presentation on funnel plots
	5. Construct validity	Regression	Spearman rank correlation, rho, and significance
	6. Minimum bias	Multi-level regression	Risk-adjusted rates, r^2 (variation explained by model) and c-statistic (discriminatory power at admission level)[162]
		Comparison of risk-adjusted and crude rates	Spearman rank correlation, rho, and significance; % change in outlier performance[163]
		Descriptive analysis	Stratification of key results
		Regression analysis including hospital-level coding variables	Coding- and risk-adjusted rates Spearman rank correlation, rho; and significance % change in outlier performance

Notes: Sources and Notes in Table 8, p67.

3.6.1. Descriptive statistics

For most of the analyses, the first step is to produce crude statistics, including bivariate associations. The methods used are summarised in Table 15.

Table 15. Simple statistical tests

Test	Data type
Pearson chi-squared, X^2	Categorical
Mann Whitney U test	non-normally distributed ordinal
t-tests (ANOVA)	normally distributed continuous
Pearson correlation, r	non-normally distributed continuous
Spearman rank correlation, r	non-normally distributed ordinal

3.6.2. Regression models

Regression analyses were used to calculate odds ratios (ORs) for both unadjusted and case-mix adjusted indicators. The indicators in this project are dichotomous in nature (with each admission either having or not having the complication in question) and, as such modelling is based on the binomial distribution. Each relevant hospital admission is flagged with a "0" or "1", referring to the absence or not of the quality or safety events. The binomial distribution has mean *n.p* and standard deviation *√(n.p.[1-p]),* where *p* is the proportion of cases and *n* the sample size.

The logistic regression can be used to predict the odds at an admission-level (Formula 1) and odds ratio (OR), which is defined as the odds from one group of admissions divided the odds from another group (Formula 2). Odds and ORs have a floor of 0 and no ceiling.

$$Odds_i = \frac{Probility\ (admission\ i\ has\ an\ event)}{Probability\ (admission\ i\ does\ not\ have\ event)}$$ **Formula 1**

$$OR_{jk} = \left(\frac{\sum_{i \in j} Odds_i}{\sum_{i \in k} Odds_i}\right)$$ **Formula 2**

Where:

i = admission within a wider set, say, A; *j,k = groups of admissions within set A;*

Regression analyses are used to predict the presence of an event based on predictor variables, which can be either continuous or categorical data. The expected value of the response variable is fit to these predictor variables, e.g. the probability of death is estimated for different patient age groups. Using natural logarithm transformation (and inverted back using the exponential function), a logistic regression can predict probability of an event as a continuous variable even though the observed events are categorical (binary). The PROC LOGISTIC function is used within the SAS software.

3.6.3. Generalised linear models

When there is evidence that the outcome is affected at different levels – in this case, both patient-level and provider-level, a multi-level model is required. Where this was the case, I fitted generalised linear mixed models using SAS's PROC GLIMMIX function. The clustering of patients may affect results; however, where I report provider-level variation with no hospital level explanatory factor I do not adjust for clustering (in the reported model) since the effects were found to be small.

3.6.4. Unadjusted indicator rates

In the first instance, crude rates (at national and provider levels) were calculated for each indicator. These simple proportion or crude rates (CR) are given by Formula 3 with the 100(1-α)% confidence interval limits given by Formulae 4a and 4b.[164]

$$CR_i = \frac{O_i}{n_i}$$ **Formula 3**

$$CR_{lower} = \frac{O_{lower}}{n} = \frac{O.(1-\frac{1}{9O}-\frac{z}{3\sqrt{O}})^3}{n}$$ **Formula 4a**

$$CR_{upper} = \frac{O_{upper}}{n} = \frac{(O+1).(1-\frac{1}{9(O+1)}+\frac{z}{3\sqrt{(O+1)}})^3}{n}$$ **Formula 4b**

Where:

i = hospital 1,…; n_i = total number of admissions at hospital i

$z = 1 - \alpha/2$ percentile of a standard normal distribution

O_i = observed number of admission at hospital i meeting the numerator criteria (e.g. receiving scan)

3.6.5. Risk adjustment

Previously cited risk factors were collated in the systematic reviews. A priori variables were retained unless they threatened model convergence. This allowed consistency in the case-mix adjustment across the different measures. In other words, all candidate variables were retained, even if not statistically significant, as the goal was confounder control.[162]

3.6.6. Goodness of fit

A common measure of goodness of fit is that described by Hosmer and Lemeshow; however, recent analyses have suggested this test is likely to produce significant results with large data sets and it is unclear how important the imperfect calibration is for a typical hospital in practice.[165] I inspected standard errors for non-convergence, co-linearity, and over-fitting, including only main effects. As reported in the analytical framework above, I reported the c- and r^2-statistics for each model:

- the c-statistic measures how well models discriminate between cases with an adverse event to those without (0.5 indicating no ability to discriminate and 1 for perfect discrimination);
- the r^2-statistic is a measure of how much of the variation is explained by the model (0 indicating none of the variation being explained and 1 indicating all explained).[166]

Model fit was also assessed graphically in plots of the residuals by the predicted values.

3.6.7. Standardisation of indicator results

A standardised ratio is the observed number of events relative to the number of events that would be expected if predictive chance of events was applied to the particular observed population characteristics. A common example is the standardised mortality ratio (SMR). The indirectly standardised ratio (ISR) is given in Formula 5, with confidence intervals given in Formulae 6a and 6b.[167] From these standardised rates, an indirectly adjusted rate (IAR) can be calculated by applying the ISR to the population crude rate (Formula 7)

$$ISR = \frac{O}{E} = \frac{\sum_p O_p}{\sum_p E_p} = \frac{\sum_p O_p}{\sum_p n_p \lambda_p} \qquad \textbf{Formula 5}$$

$$ISR_{lower} = \frac{O_{lower}}{E} \qquad \textbf{Formula 6a}$$

$$ISR_{upper} = \frac{O_{upper}}{E} \qquad \textbf{Formula 6b}$$

$$IAR = ISR.CR_{\text{population}} \qquad \textbf{Formula 7}$$

where:

O_p is the observed number of events in the subject population in group p;

E_p is the expected number of events in the subject population in group p given the standard rates;

n_p is the number of individuals in the subject population in group p;

λ_p is the group-specific rate estimated rate for group p.

O_{upper} and O_{lower} are defined above in Formulae 4a and 4b.

This method assumes both: (1) a homogeneity within each patient group; and (2) that, overall, the expected number of events is that of the observed level. The group-specific rates λ_i can be calculated either directly (subject population rates applied to a standard population) or indirectly (applying standard or calculated rates to the structure of the subject population). This project uses the latter, due to the advantage of being more robust when there small number of events of interest [164] and due to a lack of standard/reference population making the former unfeasible.

3.6.8. Funnel plots

I used funnel plots, with 95% and 99.8% confidence levels.[168] The main advantage of this graphical representation is that it can reflect the fact the confidence intervals for different sizes of providers.

3.7. Setting

England was chosen to evaluate the feasibility of applying specialty-specific indicators, which complied with the inclusion criteria that the setting was outside the USA and based on administrative data which used the ICD-10 diagnosis coding framework. As well introducing the setting, I show below that this choice also complies with the analytical framework criteria of high impact and being amenable.

3.7.1. Structure of English NHS

In England, 60 years after its creation, the National Health Service (NHS) provides 64 million outpatient and A&E attendances and, in general practice, 290 million consultations per year.[5] This care is provided by the 1.35 million staff who work for the range of providers including NHS hospitals (NHS trusts or NHS foundation trusts), GPs, dentists, and private sector and voluntary sector organisations. Following the reforms of 2012, the policy direction is set by the Department of Health (DH), with leadership for the commissioning system assumed through the independent organisation NHS England.[169]

3.7.2. Quality and safety in England

3.7.2.1. Prevalence of quality and safety in England

Recognition of the patient safety issue in England came with the publication of a seminal report in 2000.[64] Five years later, a study by the National Audit Office reported that the cost to the NHS of patient safety events included: £2 billion a year in extra bed days; an additional £1 billion to this expenditure due to hospital acquired infections; and £423 million in settled clinical negligence claims (National Audit Office 2005).

3.7.2.2. Quality and safety regulation regime in England

In England, the Care Quality Commission (CQC) has acted as the independent quality regulator for all health and social care services in England since 2009. Previously, this role was conducted by the Healthcare Commission, the Commission for Social Care Inspection and the Mental Health Act Commission.[170]

A system for reporting patient safety incident reports – the National Reporting and Learning System (NRLS)[d] – was established in 2003, with the intention that the information is used to develop tools and guidance to help local improvements in safety. More than four million incident reports have

[d] The NRLS also cover NHS trusts in Wales

now been submitted by healthcare staff. In 2010, it became mandatory for hospital trust to report serious incidents to the Care Quality Commission.[171]

3.7.2.3. Quality and safety monitoring in England

The monitoring of quality and safety in England is developing; however, few projects have used administrative data. The Government introduced the NHS Outcomes Framework in 2012 (with full role from 2013) to provide an overview of performance and act as a catalyst for driving quality improvement and outcome measurement.[172] The Framework replaced the old regimes such as 'vital signs' and 'national indicators'. The NHS Executive introduced a set of clinical indicators based on administrative data in 1999.[173] Yet a systematic review of published articles on patient safety indicators published in 2008 identified only two studies[23,174] describing adverse event measurement and monitoring in British hospitals.[88]

3.7.3. Focus on acute care

3.7.3.1. Definition of acute setting

'Primary care' services are typically the first contract point within the NHS and provided by, for example, GPs, NHS-walk-in-centres, dentists, and pharmacists. Secondary, or 'acute' care, is provided by hospitals.

3.7.3.2. Framework for identifying the setting

Administrative data systems exist in many healthcare settings, such as family (general) practice and hospitals, and can be used to analyse the quality and safety of the setting that provides the data or other settings. For instance, hospital administrative data can evidently be used to measure hospital performance, but by analysing emergency admission rates, for example, inferences about the quality of community care could be made. The following paragraphs set out why the project focuses on hospital data and hospital performance; although I do not argue that other data resources and settings are less important. To show the value in this scope, I again recall the criteria above on 'importance' regarding the room for improvement and being amenable.

3.7.3.3. Importance for hospitals

As discussed above (paragraph 1.3.2, p28), the importance of quality and safety for hospitals is well established.

3.7.3.4. Amenable to indicator-use

The use of quality and safety indicators based on administrative data is relatively recent, especially outside the USA, and so little work has gone into evaluating whether they are an effective tool to

improve care. [59] However, there have been some promising examples, such as one hospital that detected poor performance on the HCUP indicators and were able to amend their procedures to improve appropriate utilisation.[40] There is also evidence that similar improvement programmes – such as using other computerised clinical information – can be well-received by hospital providers, improve error detection and could reduce the incidence of adverse events in hospitals.[59,175-177]

3.7.3.5. Scope to improve use of indicators in this setting

Importantly, there also remains scope to improve the use of indicators based on administrative data in this setting. A systematic review of patient safety indicators in 2008 identified 165 publications relating to patient safety in hospitals, of which only 51 were original research.[88] Only one cross-sectional study (a survey for determining hospital staff's attitude to safety culture)[178] and on case control study (using administrative data to identify potential quality problems)[90] investigated validation.

3.7.3.6. Findings on quality and safety in English hospitals

The existence of significant variations in quality and safety of care in English hospitals has been well established.[151,152] In particular, annual reports by Dr Foster Intelligence – a joint venture with the Department of Health – have suggested wide-spread variation over a range of conditions and measures.[179]

3.8. Data

3.8.1. Introduction to Hospital Episode Statistics

The Hospital Episode Statistics captured almost 18 million episodes of admitted patient care in 2012-13,[180] with each record covering the continuous period during which the patient is under the care of one consultant (Finished Consultant Episode, FCE). These FCEs are linked together into hospital spells which can be further linked to any spells resulting from the transfer of a patient to another NHS hospital to form 'superspells' (referred to here as admissions). Diagnoses are recorded using the International Statistical Classification of Diseases and Related Health Problems, tenth version (ICD-10) and procedures are coded using the Office of Population Censuses and Survey's Classification of Surgical Operations and Procedures, fourth version (OPCS-4).

This data constitutes Hospital Episode Statistics (HES) which were conceived in 1987, to replace a system based on collecting data from around 10% of patients. The data are cleaned according to published rules.[181]

3.8.2. Ethical considerations

The Dr Foster Unit at Imperial has permission to hold HES data under Section 251 (formerly Section 60) granted by the National Information Governance Board for Health and Social Care (NIGB, formerly the Patient Information Advisory Group). The Unit also have approval for using these data for research from the South East Research Ethics Committee.

3.8.3. Software

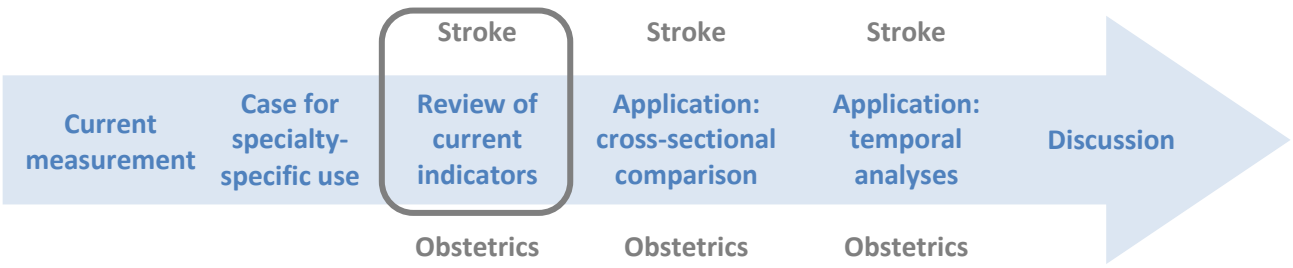
The following software programmes were in the project: Endnote X3, X5 and X6 (Endnote Carlsbad, CA); Microsoft Office Word and Excel 2007 and 2010; and SAS Version 9.2 TS Level 2MO.

Part Two:

Stroke care

Chapter 4.

Stroke literature review



Overview

Context

The previous part of the monograph outlined the rationale and analytical approaches for using routinely-collected administrative data, focused at the specialty-level, to evaluate the quality and safety of hospital care. To understand the feasibility of meaningfully applying the indicators to a particular specialty (acute stroke care), a consolidated understanding of previous measurement efforts is required. This objective was met following the systematic review methodology described in this chapter.

Methods

A systematic review of original studies that applied, evaluated or validated stroke-specific indicators based on administrative data. Indicator and study details were collated.

Findings

The initial search identified 1,999 unique citations and I identified 99 studies, and 20 broad groups of indicators. Most measures had been applied in the USA (n=64) and Canada (n=11) with few studies describing any substantial validation of the indicators, and the application of the indicators – in terms of definitions of stroke and risk-adjustment models – inconsistent. The use of stroke-specific indicators is patchy, in terms of the countries that have applied them and the indicators used. There is potential to make more meaningful inferences about the quality and safety of acute stroke care by applying a more comprehensive set of indicators and being more explicit about the assumptions used.
The literature review followed a similar process to a review of all patient safety measure based on routinely-collected hospital data which identified only a similar magnitude of studies (n = 124) despite its far broader scope.

What this chapter adds

The first comprehensive review of stroke-specific indicators based on administrative data.

The first review of how the stroke indicators have been applied, including strategies for case ascertainment, risk-adjustment and validation.

4.1. Background on stroke care

4.1.1. About stroke

Strokes are caused by an interruption of the blood supply to the brain, either through a clot (*ischaemic*) or bleed (*haemorrhagic*). The subsequent lack of oxygen and nutrients can cause damage to the brain tissue, commonly resulting in sudden weakness or numbness of the face, arm or leg. The outcome from stroke will, along with patient characteristics, depend on the area of the brain affected and the severity of the interruption.[182] An indicative stroke care pathway is included as Figure 5.

4.1.2. Importance

4.1.2.1. Burden of stroke

Stroke is the second leading cause of death worldwide, accounting for an estimated 5.7 million fatalities annually, equivalent to 1 in 10 deaths.[182]

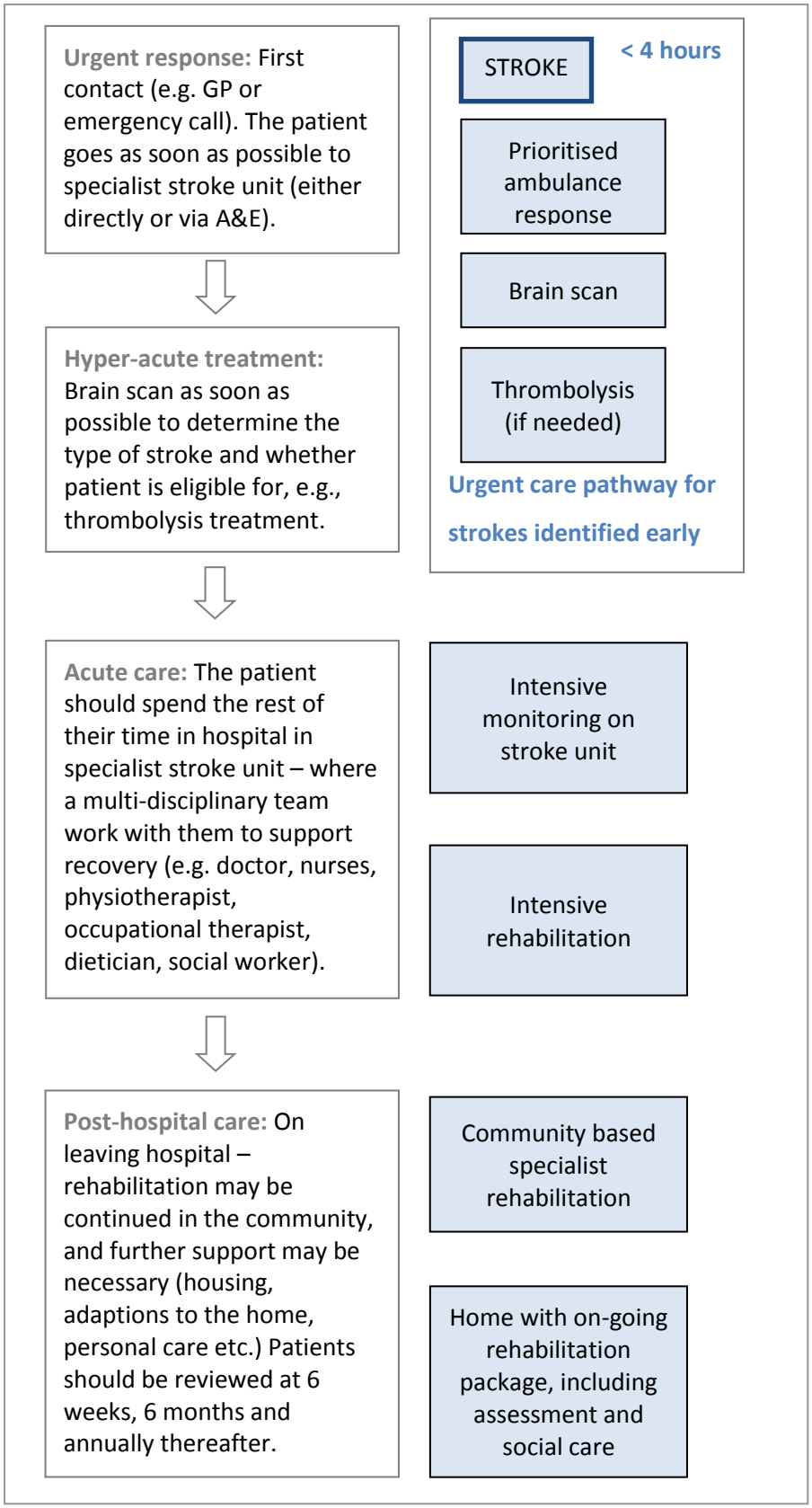
4.1.2.2. Treatment of stroke

Specialist training for stroke care is a recent phenomenon.[183] The development of the area as a specialty has coincided with dramatic changes in the management of stroke care due to burgeoning evidence of effectiveness of certain treatments. For instance, a systematic review in 1997 indicated that organisation of inpatient care in designated wards (stroke units), compared to conventional care, reduced mortality and disability despite no routinely employed medical or surgical interventions at that point.[184,185] However, in the subsequent years such interventions were developed and, for instance, by 2007 the use of thrombolysis "clot busting" treatment for certain stroke types (ischaemic) had been approved in England.[186] This, in turn, changed the importance of stroke patients receiving urgent scans so that their stroke type could be determined within the time window for administering thrombolysis.[151]

4.1.2.3. Amenable to healthcare

Previous research has suggested that stroke outcomes can be improved through the delivery of improved healthcare. Perhaps as a result of the developing nature of stroke care with lag time in some organisations and settings, unacceptable variations in the organisation and delivery of care exists.[187] For instance, a national (English) report also recently highlighted both the enduring variations in hospital care and that some levels of care remain unacceptably low.[151] This phenomenon is not restricted to England.[188]

Figure 5. Indicative recommended stroke care pathway



Source: Adapted from National Audit Office (2010)[151]

4.2. Introduction

4.2.1. Stroke as a specialty

As a specialty, 'Stroke' is a subset of internal medicine dealing with the treatment of strokes. Stroke is a new specialty with attending physicians previously – and often still – being geriatric or neurology specialists.

4.2.2. Stroke measurement

4.2.2.1. Demand for indicators

There has been widespread demand for better measurement of the burden of stroke and the quality of care administered. The WHO have actively supported countries to understand their stroke mortality rates whilst also highlighting a general lack of quality and safety indicators, particularly in developing countries.[182,189] At a national-level, taking the example of England, the DH have introduced stroke-specific measures of discharge home,[190] survival rates and emergency readmissions to hospitals within 28 days of discharge for stroke.[191] The Department also reiterated that, compared with other countries, there were relatively poor outcomes in England across some measures of stroke and, as such, the area was earmarked as a priority.[192] Academics have also called for indicators to measure stroke care performance.[193,194]

4.2.2.2. Existing measures

To date, the most comprehensive and high-profile study to benchmark stroke care have been based on specific stroke registers and prospective studies. National stroke audits are being used across the world, such as in England, Argentina, and Canada [195] More detail on measurement in the former of these nations (England), is given in the following chapter (paragraph 5.1.1.2, p117). In late 2009, a small indicator set was introduced for hospitals to monitor and review which included in-hospital stroke mortality.[196]

Previous research has suggested that the use of quality indicators in stroke care could improve performance. For instance, improvements have been attributed to the introduction of a registry in the USA with 10 performance measures,[197] another USA-based registry,[198] and the NSSA.[199] Despite the promising outcomes from these performance measurement activities, little use has been made of the vast amounts of detailed, patient-level, administrative data that is routinely collected in many nations to measure the quality and safety of individual specialities such as stroke care. As such, there would be value in documenting what stroke-specific indicators have been used to date and how these have been applied.

4.3. Objectives

To produce a comprehensive list of indicators of the quality and safety of hospital stroke care that can be based on administrative hospital data, and provide a critical assessment of how these have been applied previously. This latter objective is designed to inform the quantitative analyses presented in the following three chapters.

4.4. Methods

4.4.1. Search resources

An outline of the systematic methodology was given in a previous chapter (paragraphs 3.4.2- 3.4.3, p72-). In summary, the approach used in this review is developed from the methodology of Tsang and colleagues on reviewing patient safety measures based on routinely collected hospital data.[67] The method used to search for peer-reviewed journal articles and other research is described earlier (paragraph 3.4.4, p74). An initial search of the databases – designed to identify peer-reviewed journal articles – was performed between May and July 2010 and also later in the review process in May 2011.

4.4.2. Validation of search results

As a further check of completeness, I also reviewed the extracted indicators against measures in an American list of approved stroke and stroke rehab measures, even though they were not designed specifically for administrative data.[200] Where a study was based on a database similar to administrative data – such as a Cleveland-based study [201] which included five measures that could all be applied to administrative data – these measures were cross-referenced against the compiled list of potential indicators to ensure they were already included.

4.4.3. Definitions

4.4.3.1. Definition for administrative data

Recall from paragraph 2.1.1.1 that 'administrative data' is defined, for the purpose of this review and to meet the ambition of the study, using the inclusion criteria: routinely collected data covering more than one hospital; used for administrative purposes (e.g. for reimbursement); and can be replicated using the internationally-recognised ICD-9 or ICD-10 diagnoses coding framework.

As a principle, where a study was based on administrative data linked to another supplementary source, the individual measures described in the paper were reviewed to assess whether they could be replicated with administrative data alone. This ensured that no potentially important indicator

was excluded and that validation studies, where results from administrative data are compared to an augmented or different dataset, were included.

4.4.3.2. Definition of quality and safety

As discussed earlier (paragraph 1.1.4), providing an exact definition of quality and safety is difficult and its application for the purpose of this review on stroke care is possibly best defined by the treatment of contentious indicators:

- ✓ Length of stay is included as, while in some instances this is considered a measure of hospital efficiency rather than treatment effectiveness, some variations – such as a dichotomous indicator of long length of stay – are more likely to be considered a proxy for aspects of quality. One study, which described length of stay as an economic measure, in fact found that it was positively correlated with quality measures.[202]

- × Hospital charges are not considered as a quality indicator since this is a compound measure of patient characteristics, length of stay, and treatments provided – which can be picked up as separate indicators and risk adjustments – and is a more direct measure of hospital efficiency.

- × Clinically associated complications of stroke that are not amenable to healthcare – such as a study on fractures following stroke – are not included.[203]

4.4.4. Search strategy

4.4.4.1. Electronic databases

The following three electronic databases were interrogated: Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) 1946 to Present; Ovid Embase Classic+Embase 1947 to 2011 May 15; and Web of Science (to present).

4.4.4.2. Search terms

The search terms used were informed by previous stroke literature reviews.[204,205] The databases were interrogated using a two-dimensional search string using the 'AND' conjunction to identify papers relating to stroke which mention either a quality indicator or routine data (Figure 6). Medline and Embase were interrogated using both indexing terms (MESH and EMTREE, as shown in the Figure) and also using a two-dimensional search string akin to that used for Web of Science to check if the use of the index terms were systematically omitting results.

Terms for TIAs were also included to accommodate studies which were focused on these 'minor strokes' but may include stroke-specific indicators. As with similar reviews, only English language articles were included, and non-human studies were removed. [67,161]

Figure 6. Search terms

Web of Science Indicator* OR "administrative data" OR "routine data" OR "routinely collected data" OR "hospital episode statistics" OR "hospital episodes statistics" OR "claims data" OR "International Classification of Diseases" OR "International Statistical Classification of Diseases" OR ICD NOT "implantable cardioverter" AND Stroke OR Cerebro*vascular OR CVA OR Apoplexy OR (brain AND vascular) OR (transient isch*mic attack). **Embase** 'health care quality' OR 'clinical indicator' OR 'performance measurement system' OR 'international classification of diseases') AND 'stroke' OR 'transient ischemic attack'. **Medline** 'quality indicators, health care' or 'international classification of disease' AND 'stroke' OR 'ischemic attack, transient'.

4.4.4.1. Duplicates

Duplicate citations were identified using Endnote reference management software by identifying entries with identical authors, date, journal, and title.

4.5. Review process

4.5.1. First stage review criteria

4.5.1.1. Inclusion and criteria for review of indicator use

Recall that literature on quality and safety measures were retained if (1) based on routinely collected hospital administrative data; and (2) applied, evaluated or validated a potential or actual indicator of the quality or safety of hospital stroke care. To meet this latter criterion, an included study had to provide numerical results on performance against the indicator.

Specifically, as a first stage, abstracts were reviewed and candidate papers were excluded if:

- None of the described indicators related to an acute stroke so, for instance, excluding measures relating to surgical procedures for prevention of stroke;

- None of the described indicators explicitly relate to the quality and safety of health care provided, such as measures of cost (and also stroke incidence given this is a wider measure of public health). As a result, readmission for stroke is included[206] but neither stroke incidence nor those complications of stroke not associated with quality (such as Generalised Convulsive Status Epilepticus, GCSE) are included [207]; and
- The study was not based on routinely-collected hospital administrative data (as defined above).

Where there was insufficient detail in the abstracts to ascertain whether the criteria above were met, the study was included in the second stage (full text) review.

4.5.1.1. Inclusion of additional papers describing indicators

Secondly, I separately collated studies which described indicators that could be applied to administrative data but no actual database was used (and so failing the first stage criteria to provide numerical results) to ensure a well-defined comprehensive list of indicators.

4.5.2. Second stage review criteria

For the second stage, the remaining articles had the same exclusion criteria applied to the full texts. This stage provided a form of quality assurance by double-reviewing many of the studies whilst also giving the opportunity to collate a range of additional information from the papers. As well as the details of the publication (such as author, title, year of publication) a standard template was used to extract information relating to the studies (Table 16). Information on validation techniques applied and any key limitations were also recorded where available. The heterogeneity of the studies being considered meant that no meta-analysis would be possible so the results were not systematically recorded.

Table 16. Information extracted

Study section	Data field
Design	Country
	Details of cohort
	Longitudinal
	Key Limitation(s)
Data	Database(s)
	Year(s) of data
	Diagnosis coding framework
Variables	Quality and safety indicator(s)
	Key covariate(s)
Case ascertainment	Diagnosis field(s)
	Diagnosis codes
Statistical methods	Risk adjustment
	Validation technique(s)

4.5.3. Data extraction and review

4.5.3.1. Analytical tools

I catalogued and reviewed the extracted details using Endnote versions X3 and X4 (EndNote Carlsbad, CA) reference management software. A standard form was used to collate the information for the second stage (review of full texts). Similarly, a standard template was used to record the different categories of measures used. Statistical analyses of study details, such as trends in number of publications, were performed using Microsoft Office Excel 2010 (Microsoft Corp., USA).

4.5.3.2. Comparative analyses

In discussing the descriptive statistics on the literature, I use comparisons to previous literature reviews on patient safety indicators, predominantly that of Tsang et al, [66,67] to provide a benchmark to judge whether the specialty-specific nature of the review, and the specialty in question (stroke), have revealed atypical results. A comparison to the obstetrics literature review (Chapter 7) is included in the discussion (Chapter 11).

The information from the studies was compared as a whole and in subsets since certain purposes, for instance, would require specific risk-adjustment models and strategies for identifying stroke episodes. For example, where the purpose of the study is to validate the identification of safety issues against medical notes, it is unlikely that you would want to adjust for case-mix, whereas for comparisons between providers such a risk-adjustment would be important.

4.6. Results

4.6.1. Identification of studies

The initial search identified 1,999 articles of which 206 duplicates were removed. Of the remaining 1,793 articles screened, the full text was retrieved for 169 studies with the most common reason for exclusion being: non-relevant (n=1243); the study describes an indicator not applied to routinely-collected administrative data (n=140); evaluated only the prevalence of stroke (n=78); and investigated only the accuracy of diagnosis coding of stroke (n = 33).

During the second review stage, a further 88 studies were excluded, although 18 were added from the search of grey literature and review of the bibliographies of key articles, resulting in 99 articles (Figure 7). The information extracted from the articles are summarised in Table 17**.**

Figure 7. Inclusion and exclusion of articles in the literature review

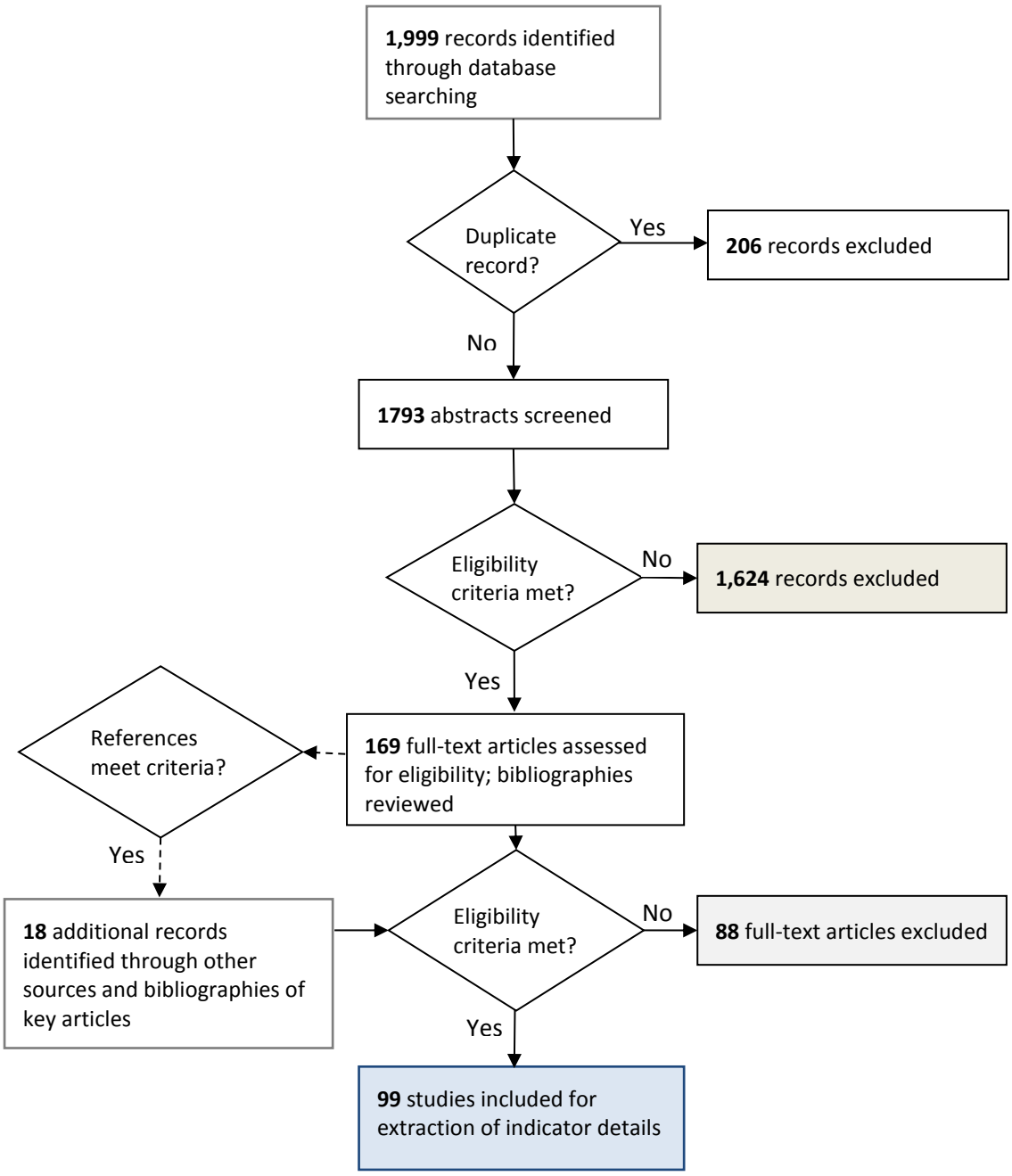


Table 17. Characteristics of literature

	Number of studies	Proportion of all studies	Examples
Data years ^			
Pre-1995	10	10.1%	Ostbye (1997)[208], Shapiro (1994)[209]
1995-1999	24	24.2%	Mosimaneotsile (2000)[210], Polanczyk (2002)[211]
2000-2004	41	41.4%	Kuwubara (2006)[212], Lee (2010)[213]
Post-2004	19	19.2%	Kleindorfer (2009)[214], Lichtman (2011)[215]
Not described	5	5.1%	Tirschwell (1999)[216]
Country of study			
USA	64	64.6%	Mitchell (1996)[217], Borzecki (2010)[218]
Canada	11	11.1%	Field (2004)[219], Johansen (2006)[220]
England	4	4.0%	Laudicella (2008)[221], Lazzarino (2011)[222]
Taiwan	3	3.0%	Lee (2010)[213]
Australia	3	3.0%	Scott (2004)[223]
Multi-national	3	3.0%	Mattke (2006)[130]
Other single-nation studies	11	11.1%	e.g. Netherlands, Slobbe (2008)[224]
Number of primary indicators			
1	62	62.6%	Tonarelli (2010)[225] Tseng (2009)[226]
2 – 3	30	30.3%	Votruba (2006)[227], Williams (2004)[228]
4 – 5	7	7.1%	Smith (2006)[229] Tu (2003)[230]
> 5	0	0.0%	-
Trend analysis			
Longitudinal	30	30.3%	Zhu (2009)[231], Volpp (2006)[232]
Non-longitudinal	69	69.7%	Zhu (2008)[233], Votruba (2006)[227]
Diagnosis coding			
ICD-9	73	73.7%	Adeoye (2011)[234], Andaluz (2008)[235]
ICD-10	11	11.1%	Cho (2009)[236], Gattellari (2009)[237]
Other †	15	15.2%	e.g. both ICD-9 and ICD-10, Westert (2002)[238]
Stroke diagnosis field			
Primary only	57	57.6%	Gregory (2009)[239], Lee (2010)[213]
Primary and some secondary	1	1.0%	Meyer (2009)[240]
Primary and all secondary codes	6	6.1%	Lanska (1994)[241]
DRG-level	3	3.0%	Cleves (1997)[242]
Not described	32	32.3%	Censullo (2008)[243], Crowley (2009)[244]
TOTAL	**99**	**-**	

Notes: ^ Where studies covered multiple years of data, the study is categorised according to the most recent year

† Includes studies where coding framework not described and where both ICD-9 and ICD-10 used.

4.6.2. Details from studies

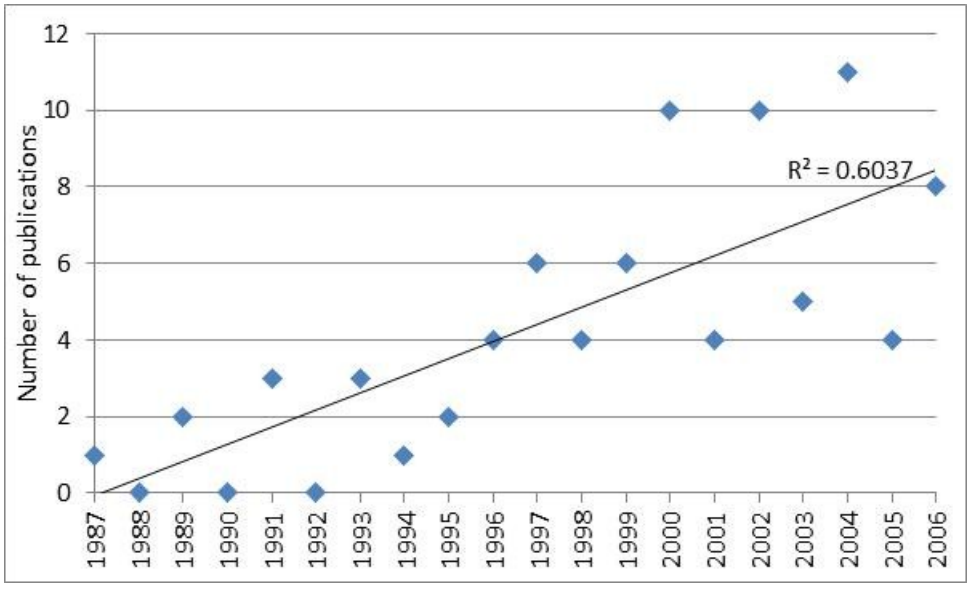
4.6.2.1. Country

Three-in-five published studies were based on data from the USA, with the next most prolific country being Canada (n=11). No other nation had more than 4 studies based on administrative data evaluating this specialty.

4.6.2.2. Year of study

None of the studies were published before 1987; however, this is likely to be an artefact of the commonly used administrative databases not being established before 1988 (paragraph 2.1.1.2, p39). The number of studies has increased over time and around two-thirds (64%) of the studies incorporated data from 2000 or later. Due to time lags in accessing, analysing and publishing studies based on administrative data one might expect a time lag in publishing data and, therefore, the trend was considered before 2007. Using a linear regression of year (of latest data used in each study) against number of studies shows that there has been a substantial and significant increase over time ($\beta = +0.45$, p-value < 0.001) (Figure 8).

Figure 8. Number of publications compared to latest year of data



Note: Year represents the latest date of data used within studies

4.6.3. Study designs

4.6.3.1. Cross-sectional regional analysis

The studies were predominantly cross-sectional in design, with study populations ranging from admissions at a single hospital to multiple national populations and 2 studies presenting the analyses at multiple geographical levels.[214,245] At the most aggregated geographical level, there were 3 cross-

sectional descriptive studies comparing performance across countries.[130,238,246] Other studies used the regional geographical unit either with the purpose of directly comparing performance [206,218,247,248] or investigating particular regional-level covariates such as variations in: competition;[232,249] use of comprehensive regional stroke networks;[250] Health Maintenance Organisation (HMO) penetration;[251] and level of rehabilitation provided by local health authorities within a region.[252]

4.6.3.2. Cross-sectional provider-level analysis

At the provider-level, only two studies were directly focused on highlighting variation between hospitals.[243,245] Rather, these provider-level studies predominantly focused on investigating the effect of particular hospital characteristics. The most prominent hospital characteristics being studied were:

- teaching status, measured either dichotomously [211,217,235,253-258] or using the resident-to-bed ratio as a proxy to differentiate between major teaching, minor teaching, and nonteaching hospitals [257]; and
- size, measured through either number of beds [235,258,259] or throughput [227,247,260-263];
- type, including municipal vs voluntary,[209] HMO vs Fee For Service,[264] Veterans Association (VA) vs non-VA,[218] and non-profit vs for profit vs public ownership;[258] and
- level of stroke specialism, which incorporates studies looking at the association in performance with level of stroke accreditation or existence of a specialist stroke unit. [215,217,231,237,255,265,266]

Further hospital characteristics evaluated were: staffing levels;[236,257] exposure to competition;[249,258] and treatment intensity (cost and treatments).[267]

At a sub-provider level, some studies focused on clinician experience in terms of volume of stroke patients they see [268] and clinician specialty.[217,229,269-271] Other studies used a more investigatory approach, with multiple hypotheses about associations between performance against the indicators and a range of hospital factors,[213] or even a range of patient and hospital factors.[226,261,272-276]

4.6.3.3. Cross-sectional patient characteristic level

Similarly, some investigatory studies looked at how different patient groups performed against the indicators and, for some research, looked across a range of patient factors.[222,277] Other studies investigated associations between performance and specific patient characteristics, such as

ethnicity;[278,279 220,232,280-284] age;[239,280] gender;[280,285] socio-economic status;[239,286] distance to hospital;[227] and insurance status.[287]

Away from patient characteristics, other studies looked at the association between exposure of patients to different treatments,[259 288] with a number of studies looking at the influence of day of admission on outcomes.[253 268 289 244,290]

4.6.3.4. Temporal design

Under a third of studies (30/99) used longitudinal analysis, with a small selection using quasi-experimental design:

- regions, e.g. the effect of reimbursement changes for uncompensated care; [249]
- hospitals, e.g. the association between quality and whether a hospital had been merged/acquired; [291] and
- a single hospital, e.g. how introducing a more specialist service affected performance. [231,266]

4.6.4. Details from datasets

4.6.4.1. Data source

The studies reviewed used a range of different administrative databases, with the most frequent including the:

- AHRQ-sponsored HCUP NIS and HCUP SIDs (see paragraph 2.1.1.2);
- National Patient Care Database's Patient Treatment File, which includes records of all hospital admissions from Veteran Association patients; and
- Medicare Provider Analysis and Review (MEDPAR) file, which includes data from claims for services provided to USA beneficiaries admitted to Medicare facilities, available since 1999.[292]

Some studies used multiple administrative databases – for example, international studies for which there are no cross-national administrative databases[18,33,34] – or, more commonly, administrative hospital data were supplemented with another source, such as mortality records to account for deaths outside the hospital setting. The administrative datasets were found to be heterogeneous with, for instance, many variations in the coding framework for operations/treatments received.

4.6.4.2. Case ascertainment

Almost three-quarters of studies used the ICD-9 framework for coding diagnoses, with only 11 being based on ICD-10. A third of studies did not describe which fields were used to identify strokes, although most that did used only the primary/principle diagnosis (57/66).

Limiting the description to those studies that used the ICD-10 framework, the most commonly used codes were: I61 *(intracerebral haemorrhage)*; I63 *(cerebral infarction)*; I64 *(stroke, no specified as haemorrhage or infarction)* and, to a lesser extent, I60 *(subarachnoid haemorrhage)* and I62 *(other non-traumatic intracranial haemorrhage)*. However, there were numerous variations (Table 18). For instance, one study – described as evaluating all strokes – only used I63.[237] In some instances, the studies set out prognostic differences between the stroke types as a rationale for making exclusions[217,260] while in others the study was explicitly focused on a specific stroke type although no reason for this restriction was given.[206,235]

Table 18. Variations in use of ICD-10 diagnosis codes for identifying stroke

ICD-10 code	Description	Examples of study
Example of additional codes used		
H34.1	Central retinal artery occlusion	Zhu (2009)[213] ; Field (2004)[249]
G45	TIAs and related syndromes	Zhu (2009)[231]
G45, not G45.4	TIA excluding transient global amnesia	Field (2004)[219]
I69	Sequelae of cerebrovascular disease	Katzenellenbogen (2010)[293]
all other I6x codes	*See note**	Reklaitiene (2008) [251]; Slobbe (2008)[224]; Tu (2009)[253]
Examples of codes excluded		
I61	Intracerebral haemorrhage	Saposnik (2007)[202]; Saposnik (2008) [204]
I63.6	Cerebral infarction due to cerebral venous thrombosis, non-pyogenic	Pajunen (2005)[294]
I60	Subarachnoid haemorrhage	Scott (2004)[223]; Mattke (2006)[116]; NHS Information Centre (2011) [180]
I62	Other non-traumatic intracranial haemorrhage	Zhu (2009)[231]; Lazzarino (2011) [228]; Field (2004)[219]
I64	Stroke, no specified as haemorrhage or infarction	Cho and Yun (2009)[236]

Notes: I60 – I64 and I69 described elsewhere, I65-6 (Occlusion and stenosis of precerebral and cerebral arteries, not resulting in cerebral infarction), I67 (Other cerebrovascular diseases), I68 (Cerebrovascular disorders in diseases classified elsewhere).

4.6.4.3. Extract validation

Few studies explicitly validated the accuracy of the underlying data or appropriateness of the assumptions used to extract details of stroke patients. Within the studies reviewed, the only coding items for which the validity was evaluated were the coding of stroke diagnoses [163] and recording of thrombolysis [295,296] including using comparison with non-administrative datasets.[225,234,297] One study excluded hospitals for which the data was deemed of inappropriate accuracy.[263]

4.6.5. Risk models

4.6.5.1. Case mix adjustment

The studies used various adjustments to mitigate the effects of differing case-mix. At its most rudimentary, some studies presented unadjusted rates when making comparisons[219,221,266] and others stratified their findings by age,[130,235] or age and gender.[240,248,294] Some studies used modelling techniques to adjust for these latter two patient demographics.[208,218,220,224,245,298,299] The most common set of variables used in the risk adjustment were age, gender and a measure of comorbidities [251,253] [230,231,233,242,265] however there were many additional variables considered in the studies (Table 19).

Table 19. Factors included in risk models

Factor	Covariate
Patient characteristics	Social deprivation/socio-economic status [253,287,288] Ethnicity [300] [232,283] [284] Resident of country [237] Marital status [163,237,252] Distance to hospital [227] Insurance (payer, including Medicaid status)[229,236,237,239,244,247,262-264,272,281,284,289-291]
Medical history	Place of admission [258] [163,227,288] First vs recurrent stroke [301] Previous admissions: overall [206,220,271] CVD-specific [217] Stroke-specific [263,301] TIA-specific [163,301]
Stroke/treatment details	Stroke type [163,213,239,252,263,279,286,302], including stratified[215,246] *Temporal:* year of treatment [247] [216,264,274,283]; season [290]; and proportion of patients admitted at weekends[271] *Measures of severity:* Use of mechanical ventilation [272,273,281] [163] [213] [229,237,264,273]; Signs of cognitive functions and awareness [237]; gastronomy tube [229,264,272,273,281]; certain surgical operations [213] [264]; hemiplegia or hemiparesis [213] [264,279]; and residual neurologic deficits [213] [264]. Craniotomy [277]; Other conditions (e.g. pneumonia, atrial fibrillation, urinary incontinence) [237]; Severity (late effects of stroke, which includes cognitive deficits. Speech and language deficits an, and paralysis, dysphagia, incontinence, and delirium) [271]; Gastronomy, tracheotomy, interventional intracranial thrombolysis, interventional angioplasty, endarterectomy, operative angioplasty, extracranial-

	intracrancial arterial bypass [212]
Treatment details (e.g. hospital characteristics)	Hospital location [239,260-263,268,277,287,288,290] Hospital size as volume [269,291 211,276,286-288,290] and beds [211,259,271,287] Hospital teaching status [227,260,263,268,275,276] and hospital ownership [211 268] Clustering within hospitals [232,249,257] Physician specialty [261,263,276] To account for the regionalisation on services, studies have also adjusted for whether a patient was transferred between hospitals[217]; regional resources [268]; and competition [268]
Stroke outcome	Discharge destination [275] ICU [277 261,275] Length of stay [272 238,273,277,281] (particularly for time bound indicators, such as mortality within 28 days)

4.6.5.2. Participant exclusions

To account for potential bias from different stroke, patient and hospital characteristics, studies used a range of exclusion criteria to define their participants. The most common criterion was age of patient, although sometimes this was necessitated by the underlying administrative databases being age-specific. Cut-offs for the exclusion of young people included 18,[276] 20,[286,299] 35,[224] and 40.[238] Conversely, other publications excluded older patient , such as those over 105 years old,[286] with one study excluding those over 75 because of the possibility of increasing unreliability of the clinical diagnoses of stroke with advancing age.[303]

Other exclusions related to removing patients with atypical prognosis, such as: lengths of stay over one year;[220] previous admission admitted for CVD in five years before admission;[220,286] patients admitted from long-term care;[257 208] non-residents;[208] transfers;[208,287] admissions not through emergency department;[269,286] patients discharged within 1 day of admission or left the hospital against medical advice.[215,265] Other studies excluded hospitals with low stroke numbers [227,298] or patients for whom diagnosis and treatment codes suggested an alternative primary diagnosis to stroke.[276]

4.6.5.3. Summary of key factors

The following exhibit (Table 20) summarises how different studies have controlled for, or examined, some potential determinants of variation in quality and safety. This incorporates details about the key dependent variables, case-mix adjustment and exclusions from the indicator denominators.

Table 20. Summary of key factors used in studies

Level		Key covariate	Risk-adjusted	Participant exclusions
Regional	Regional networks	Camilo (2005)[250]	e.g. transferred case in Mitchell (1996)[217]	
	Resources	Carinci (2007)[252]	Tung (2009)[268]	
	Competition	e.g. Volpp (2005)[38,39]	Tung (2009)[268]	
Hospital	Location		e.g. Cross (2003)[239,262 260,261,263,268,277,287,288,290]	
	Size (volume)	e.g. Bardach (2002)[35,54-58]	e.g. Gillum (2008)[269,291 211,276,286-288,290]	e.g. Votruba (2006)[227,298]
	Size (beds)	e.g. Andaluz (2008)[44,52,53]	e.g. Polanczyk (2002)[47,53,71,93]	
	Teaching status	e.g. Andaluz (2008)[211,217,235,253-258]	e.g. Saposnik (2007)[227,260,263,268,275,276]	
	Comparison/clustering	e.g. NHS Information Centre (2011)[32,43]	e.g. Volpp (2005)[38,39,51]	
	Ownership		e.g. Polanczyk (2002)[211 268]	
	Specialist centre			
Clinician	Specialty	e.g. Mitchell (1996)[217,229,269-271]	e.g. Saposnik (2008)[56,58,78]	
Stroke	Type	e.g. Borzecki (2010)[209,218,258,264]	e.g. Carinci (2007)[42,58,73,83,90,92,115,117]	
	Year		e.g. Bardach (2002)[247 216,264,274,283]	
	Day of stroke (weekend)	e.g. Bell (2001)[253 268 289 244,290]	Howrey (2011)[271]	
Stroke outcome	Length of stay		e.g. Kind (2007)[272 238,273,277,281]	Johansen (2006)[220]
	Discharge destination		Reed (2001)[275]	
	ICU		e.g. Lee (2007)[277 261,275]	
Patient characteristics	Socio-economic status	e.g. Gregory (2009)[239,286]	e.g. Bell (2001)[45,93,94]	
	Ethnicity	e.g. Lee (1998)[278,279 220,232,280-284]	e.g. Mitchell (2000)[300 232,283 284]	
	Distance to hospital	Votruba (2006)[227]	Votruba (2006)[227]	

Insurance/payer	Shen (2007)[287]	e.g. Bardach (2002)[35,57,58,60,61,66,72,74,86,89,90,95-98]	
Age	Gregory (2009)[239,286]	e.g. Kaprai (2002)[222,286]	e.g. Westert (2002)[224,238,276,286,299,303]

Note: The list gives references as examples and should not be considered as comprehensive

4.6.7. Details from indicators

4.6.7.1. Range of indicators

Across the 99 papers, 166 (not necessarily distinct) indicators were described – an average of 1.7 indicators per paper – with no study having more than 5 indicators. Across the publications there were 20 categories of indicators, which cover: urgent acute care; acute treatment; complications; mortality; discharge; review; post-discharge (Table 21, with a full list given in Appendix D, p304). Through conducting the review, I also identified potential indicators which had yet to be applied to administrative data, such as 'readmission for incontinence',[160] and 'newly institutionalised within 2 weeks of admission'.[304] Mortality, readmission and thrombolysis rates were all often described across the published literature. In line with the inclusion criteria, 'hospital charges' was not regarding as a quality indicator, even though this was common within the included studies [211,227,237,247,273,275,288,290]. Some studies reported compound measures such as: 'complications' which covered either pneumonia, UTI and DVT [243] or pneumonia, PE, DVT [219]; or 'adverse outcome' which combined in-hospital death with discharge destination other than home [284].

4.6.7.2. Indicator validation

Approaches used to take assurance on the validity of the indicator results were both sparse and varied. One of the studies compared performance across different indicators constructed from the same dataset; however, in this instance, few pairs had significant correlation.[223] A more common approach was to compare the results from administrative data to another data set, with the majority looking at mortality rates [240,248,283,285,303] with the other data sets ranging from death registers, to other administrative database [274] and case note reviews at one of the hospitals.[228] One study used a present-on-admission (POA) flag within the administrative data to see the effect on the sensitivity and specificity of the resulting indicator.[298]

The most common form of validation described in the reviewed publications was on the risk-adjustment, with 16 studies using different sets of variables to create the various models for case-mix adjustment within their studies [163,166,209,212,216,227,233,239,242,249,257,272,273,282-284] and another study using competing analytical techniques for modelling risk.[269] Apart from these studies evaluating risk-adjustment, few of the remaining studies explicitly reported the model fit.[226,233,242,244,268,289] One

study investigated the impact of missing data and patient identifiers [130] and two repeating the analysis on a different stroke extract (derived from different sets of stroke codes).[265,290]

Table 21. Categories of measures

Pathway	Outcome measures	Indicator type	Example paper
Urgent acute care	Thrombolysis	Process	Adeoye (2011)[234]
	Urgent brain scan within 24 hours	Process	Lazzarino (2011)[222]
	Emergency hospital transfers	Structure	Kind (2007)[272]
Acute treatment	Depression and other mental health problems following stroke	Outcome	Williams (2004)[228]
	Inpatient rehabilitation services	Process	Kapral (2002)[286]
	Warfarin use	Process	Smith (2005)[264]
	Length of stay	Outcome	Tu (2003)[230]
Complications	Urinary Tract Infections (UTIs)	Outcome	Censullo (2008)[243]
	Pneumonia due to swallowing problems	Outcome	Field (2004)[219]
	Lower respiratory tract infection (LRTI)	Outcome	Laudicella (2008)[221]
	Deep Vein Thrombosis	Outcome	Censullo (2008)[243]
	Pulmonary Embolism	Outcome	Field (2004)[219]
	Clostridium difficile	Outcome	Laudicella (2008)[221]
	Adverse outcome	Compound	Zacharia (2010)[284]
Mortality	Craniotomy: mortality rate	Outcome	Borzecki (2010)[218]
	In-patient death within 30 days of admission	Outcome	Ostbye (1997)[208]
	All-location deaths within 30 days of admission	Outcome	Tirschwell (1999)[216]
Discharge	Return to normal place of residence	Outcome	Votruba (2006)[227]
Post-discharge	Emergency readmission to hospital following treatment from a stroke (all cause)	Outcome	Kamalesh (2007)[305]
	Recurrent stroke within 28 days	Outcome	Tu (2003)[230]

4.7. Discussion

4.7.1. Principle findings

The literature review identified 99 studies that discussed the use of administrative data to measure the quality or safety of hospital stroke care and found 166 indicators, with the most prevalent relating to mortality, readmissions and thrombolysis rates. There has been a general increase in the application of these indicators, with over two-thirds of the studies using data since 2000.

4.7.2. Source of data

Three-quarters of the studies were from North America (USA or Canada) who use ICD-9 and a combination of ICD-9 and ICD-10 diagnosis codes, respectively. As such, only a small minority of studies, equating to 12%, use ICD-10 diagnosis codes. There is no one-to-one mapping between the two frameworks although many of the codes have equivalents. This, however, highlights that application of the measures and findings cannot be generalised everywhere. The reason for the prominence of research within North America may be an artefact of a greater use of established administrative datasets in these countries and, also, a lack of national audit data in USA (to act as an alternative data source).

4.7.3. Stroke case ascertainment strategies

4.7.3.1. Diagnosis fields

The review identified studies that used all diagnoses fields, and one [240] that used first to fifth diagnosis positions, with these strategies likely to increase the sensitivity but reduce specificity of the algorithm for identifying strokes within the administrative database. However, among those studies that did describe which diagnosis fields were interrogated, the majority used only the primary/principle diagnosis, which is in keeping with previous findings on the use of administrative databases for quality research.[306] Within Australian administrative health data, the primary diagnoses are rarely missing and stroke diagnoses appearing in administrative data sets are accurate 93% of the time.[237] Similar results have been found with ICD-9 databases.[307] Moreover, other studies have added that using the first diagnosis field only increases the probability that the patient had suffered a stroke on the date of admission.[219,308] This is also consistent with coding guidance in the UK which advises that acute stroke should be coded in the primary diagnosis field.[309]

4.7.3.2. Stroke diagnosis codes

From reviewing the literature, it was also evident that there was no consensus over what diagnosis codes were used in the studies on administrative data. As an example, looking just at the research

based on ICD-10 databases, the comprehensiveness of the strategy for identifying stroke admissions range from using just one 3-digit group of diagnosis code (I61) to using ten (I60 – I69). This is significant since certain stroke types have distinct care pathways. For instance, ischaemic strokes can potentially be treated using thrombolytic drug treatment; however, this would result in an adverse outcome if administered in treating a haemorrhagic stroke. Similarly, subarachnoid haemorrhagic strokes will likely receive substantially different urgent care to the other emergency stroke types.

A 2005 study found that stroke coding was equally good for both ICD-9 and ICD-10 and while there were some differences in coding of the different types of stroke, these variations were not statistically significant and the most atypical coding performance was not for actual stroke (rather for TIAs). [310] However, not all strokes will be either accurately diagnosed or, even where they are, coded correctly or specifically enough. As a result, there is variable use of the unspecified or ill-defined stroke codes (e.g. I64 in ICD-10 and 436 *Acute, but ill-defined, cerebrovascular disease* in ICD-9). Therefore, a decision has to be made whether to use more limited codes to identify strokes, e.g. excluding those recorded as subarachnoid haemorrhages, which will increase the specificity of the denominator but might increase bias resulting from variable use of the unspecified stroke codes.

One study specifically looked at the effect of using two different case ascertainment strategies on an outcome indicator (mortality). The risk-adjusted models for cohorts produced from the highly specific and, separate, highly sensitive sampling strategies were both robust and had some impact on respective rankings of the regional networks against the outcome measure, and whereas two outliers were identified in one cohort, none were identified in the other.[163]

4.7.4. Indicator validity

4.7.4.1. Risk-adjustment

The importance of adjusting for case-mix in research on stroke care using administrative data is longstanding.[209] However, continuing the theme of the findings of this review, there was unexplained variations in the approaches to adjusting for stroke type and both patient- and hospital-characteristics. Studies have found that adjusting for socio-economic status is inadequate without more comprehensive measures being included; however, even this level of risk-adjustment is more sophisticated than that applied in many of the studies.[311] Likewise many studies did not adjust for comorbidities despite research showing that it is useful for predicting risk-adjusted in-hospital case-fatality in stroke outcome studies.[233]

4.7.4.2. Coding of covariates, activity and outcomes

Accurate coding – including of patient and hospital characteristics, treatment activity and outcomes – is inherently important if the data are to be used to identify true variations in the quality and safety of care. A number of studies evaluating the accuracy of coding for stroke patients – looking at variables such as coding of comorbidities[310,312,313] and nonmedical and socio-demographic data elements[269,314] – were identified through the initial search of citations and subsequent review of bibliographies and, in general they showed that the coding was of reasonable accuracy. [310] However, the results are neither appropriate for synthesising – since heterogeneity between the studies mean that it is not possible to undertake any meta-analysis of the results – nor generalisable – since there are likely differences between the coding practice in different countries and changes since these non-contemporaneous studies were published.

One commonly utilised indicator that directly relies on coding completeness is thrombolysis rate. A number of studies on this topic that were referenced in the reviewed studies with, in general, high specificity and mixed results on sensitivity, although again these are not necessarily generalisable due to the heterogeneity of databases, differing national incentives to complete them correctly and limited coverage of some of these studies.[225,295,296]

4.7.4.3. Indicator validation

As well as the 'upstream' validation of the case ascertainment strategies and robustness of the risk-adjustment model, the literature review revealed some examples of 'downstream' explicit validation of the indicator results, for instance by comparing the results with indicators derived from other sources. Despite the existence for various frameworks of taking assurance on the validity of such indicators,[130,315] none of the studies described a comprehensive validation strategy, certainly not in comparison to this study's analytical framework (Table 8, p67).

4.7.5. Strengths and limitations

4.7.5.1. Strengths

This study represents the most comprehensive review of measures of the quality and safety of stroke care using administrative data. An obvious strength of this literature review process which focused on a particular specialty is that it seems to be a sensitive strategy for identifying relevant research and, as such, provides more assurance that the resultant citations are comprehensive. For instance, the literature review followed a similar process to a review of all patient safety measure based on routinely-collected hospital data which identified only a similar magnitude of studies (n = 124) despite its far broader scope.[67]

4.7.5.2. Limitations of the review process

The limitations of this review process include the potential bias introduced from the use of a single reviewer and exclusion on non-English language articles. The exclusion of non-English articles was consistent with a previous review of patient safety measures [67] which found that such citations were limited in number. No formal meta-analysis could be undertaken because of the heterogeneity in study methodologies and underlying data frameworks. I did not use a template to assess the quality of the papers since there was a very diverse range of purposes across the studies and, as already discussed, validity needs to be proportionate to the purpose. Given the strong evidence on the existence of publication bias,[316] studies using indicators which did not identify significant results when assessing the quality and safety of stroke care may have been missed by the search strategy in this study.

4.7.5.3. Limitation of review protocol

The literature review again reiterated the differences in extant administrative data and how they are used, such as in data-linkage studies. This makes the formulation and application of strict inclusion/exclusion criteria for such literature review difficult. In keeping with the review criteria for this research as described earlier, a study [317] based on Uniform Data System for Medical Rehabilitation (UDSMR) was excluded, since the indicators in the study used this specific rehabilitation data collection only and were not readily reproducible with other administrative data sources, whereas a study which linked this to the Medicare Provider and Analysis and Review files (MedPAR) administrative database was included, since the measure described could be recreated.[318] Similarly, a study with measures using linked data with variables including physical activity (not commonly included in administrative datasets) [319] was excluded; however, a study which supplemented the hospital administrative data with clinical register data [255] was included since the measures could still be replicated - albeit with some minor amendments - using purely administrative data.

4.7.6. Conclusion

4.7.6.1. Contribution to knowledge of indicators

Looking at a specialty in detail, had advantages over a broader scope; not least, to better identify existing indicators. The study has highlighted a paucity of use of such indicators and a lack of transparency and consensus on key assumptions.

4.7.6.2. Recommended uses

The studies in the review have been used for a range of purposes, from comparing performance across countries to evaluating the impact of the implementation of a specialist stroke unit within a single hospital. Administrative data do not include all the process measures collected in bespoke registers, such as provision of anti-thrombolytics, however they cover the main outcomes – such as including all three key outcomes (length of stay, mortality and discharge destination) from one register.[320] And indeed, in some cases, registers had to be linked to administrative data to get details of areas such as clinician experience.[255] However, continued monitoring of stroke mortality has been described as "essential" [321] [322] and, more generally, surveillance of stroke care is necessary to understand and meet the demand for stroke care [323] with the current lack of measures having been cited as partly attributable to the limitations on the adoption of best practice. [324]

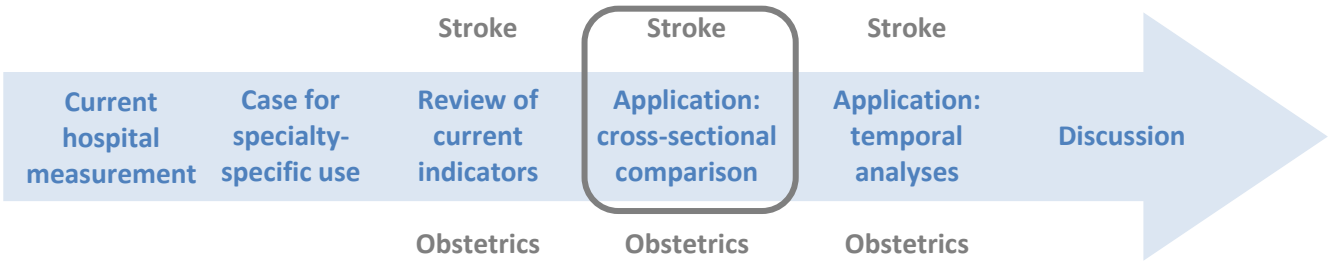
As a result, administrative data has been recommended as a tool for identifying potential quality problem deserving further review.[310,323,325] Moreover, there is some evidence to suggest that such measures can bring meaningful change with, for instance, a decline in stroke mortality rates associated with the introduction of an initiative to measure hospital performance, although this was not statistically significant (p=0.17).[326] While the literature suggests there is a role for stroke-specific indicators based on administrative data, there is scope for the current raft of measures to be improved in their construction and application. Table 22 sets out six key recommendations for future stroke research based on the findings of the review, which are enacted in the following three chapters.

Table 22. Summary of implications for study design

Finding	Implication	Recommendation
Despite the recommendation of using indicators in passive surveillance of stroke services, the majority of research has been restricted to North America, and using ICD-9 databases.	Barriers to generalising findings across borders mean that little is known about the use of such indicators in other countries, especially those using ICD-10.	*Key study designs should be replicated using data from outside USA or Canada to identify any **potential outliers**.*
Two-thirds of studies applied only one indicator.	Overall level of quality of care at a provider-level cannot be deduced from the results of only a small set of indicators.[223]	*To use the depth of information within the administrative data to construct more **comprehensive set of indicators**.*
There is a lack of consensus and transparency about the assumptions used in applying the indicators to administrative data	Barrier to generalising, validating or benchmarking results.	*To explicitly set out the **assumptions** used in studies, such as the choice of diagnosis codes and fields.*
While concerns remain over bias due to differences in coding practice, none of the literature reviewed explicitly applied sensitivity analyses on effect of variations in coding.	The effect of variation in coding practice is unknown.	*To analyse the potential effect of **coding variation**.* *To consider areas known to be potentially important and where comparisons are less affected by coding practice (e.g. **weekend** care)*
Less than a third of studies looked at trends in performance over time.	Potential shortcomings in quality and safety are not being identified.	*Research into the potential opportunities for monitoring quality and safety using **longitudinal** designs.*

Chapter 5.

Stroke variation and validation



Overview

Context

The review of stroke literature described in the previous chapter revealed a paucity of studies evaluating stroke services using administrative data outside Northern America, despite suggestions that this would be a useful surveillance tool. Across the research, there was a lack of transparency of the assumptions used with little use of the potential richness of indicators available or opportunities to validate results. In this chapter I aim to address these limitations by examining the feasibility of evaluating stroke care in England using routinely collected administrative data.

Methods

Six indicators, spanning the hospital care pathway, were applied to English data in 2009/10. Logistic regression analyses were used to adjust for case-mix, and funnel plots to look at potential outliers. I investigated the effect of variations in hospitals' data coding practice.

Findings

The indicators were sufficiently sensitive to identify significant variations in performance as recorded in the data, with 181 occurrences of hospitals performing statistically differently than expected at the 99.8% significance level across the six indicators. Differences in coding practice appeared to only partially explain the variation. The work showed that administrative data can be used to indicate potential concerns around quality and safety, in specific areas, across the whole stroke care pathway. The chapter discusses work – on demographic differences – based on one of the indicators which emphasises the impact of stroke care and suggests that there is scope to improve.

What this chapter adds

Most comprehensive assessment of stroke services using administrative data in any country

The most sophisticated evaluation of validity of any stroke assessment using administrative data

Acknowledgements and related papers

Advice was provided by: Sue Eve-Jones (UK Professional Association of Clinical Coders) on clinical coding; and Charlie Davie provided advice on clinical implications. Antonio Lazzarino undertook the data extraction and analysis for the ethnographic study (reported in the discussion).

Palmer WL, Bottle A, Davie C, Vincent CA, Aylin P. Meeting the ambition of measuring the quality of hospitals' stroke care using routinely collected administrative data: a feasibility study. *International Journal for Quality in Health Care* 25(4):429-36

Lazzarino AI, Palmer WL, Bottle A, Aylin P (2011) Inequalities in Stroke Patients' Management in English Public Hospitals: A Survey on 200,000 Patients. *PLoS One.* 6(3): e17219

5.1. Background

5.1.1. Stroke care in England

5.1.1.1. Priority of stroke care measurement

Stroke costs the National Health Service (NHS) in England £3 billion a year in direct care costs, and is the largest cause of adult disability.[151] In July 2010 the Government set out its policy intentions for the NHS including to measure performance "against results that really matter", exemplifying this with stroke survival rates and emergency readmissions to hospital within 28 days of discharge for stroke.[191] The proposals also reiterated that, compared with other countries, the NHS has achieved relatively poor outcomes in some measures of stroke and, as such, stroke was earmarked as one of the two areas with the most scope for improvement.[192] More recently, the NHS's operating framework for 2011/12 highlighted the scope for improving stroke outcomes, in particular through access to timely scanning and, if clinically indicated, thrombolysis.[327] In recent years, the Department of Health in England have also introduced payments for patients receiving thrombolysis,[328] early scans, and being admitted directly onto stroke units.[329]

A national report also recently highlighted both the enduring variations in hospital care and that some levels of care remain unacceptably low.[151] This phenomenon is not restricted to England,[188] and as a result research has also called for indicators to measure stroke care performance.[193,194]

5.1.1.2. Existing measurement of stroke in England

To date, the most comprehensive and high-profile study to benchmark stroke care in English hospitals has been the National Stroke Sentinel Audit (NSSA). The NSSA, which was first run in 1998, benchmarks hospitals' stroke care over both organisational and clinical domains, based on a survey of hospital stroke teams and medical record review of around 60 consecutive patients. A further bespoke data collection was required after the introduction of mandatory measures regarding stroke care in the NHS operating framework for 2008-09 to 2010-11.[330] Since 2000, the National Centre for Health Outcomes Development (NCHOD) has published a compendium of some 300 indicators, including three indicators based on administrative data specific to hospital stroke care (number of admissions, readmissions and mortality), with a further six measures relating to stroke prevention.[331]

5.1.2. Revisiting limitations in existing literature

The review of literature of stroke measures suggested that, to address the shortcomings of existing research, future efforts should focus on (excluding the recommendations relating to temporal study designs which are covered in the following chapter):

1. applying key study designs for surveillance outside the USA or Canada to identify potential quality issues;
2. using a more comprehensive set of indicators than has been done to date;
3. being explicit about the assumptions used; and
4. understanding the potential bias from variation in coding practice.

5.2. Objectives

In this chapter I investigate the feasibility of using HES data to evaluate the quality and safety of stroke care at a hospital level focusing on the ability of measures to identify important effects in a robust fashion.

5.3. Methods

5.3.1. Selecting indicators

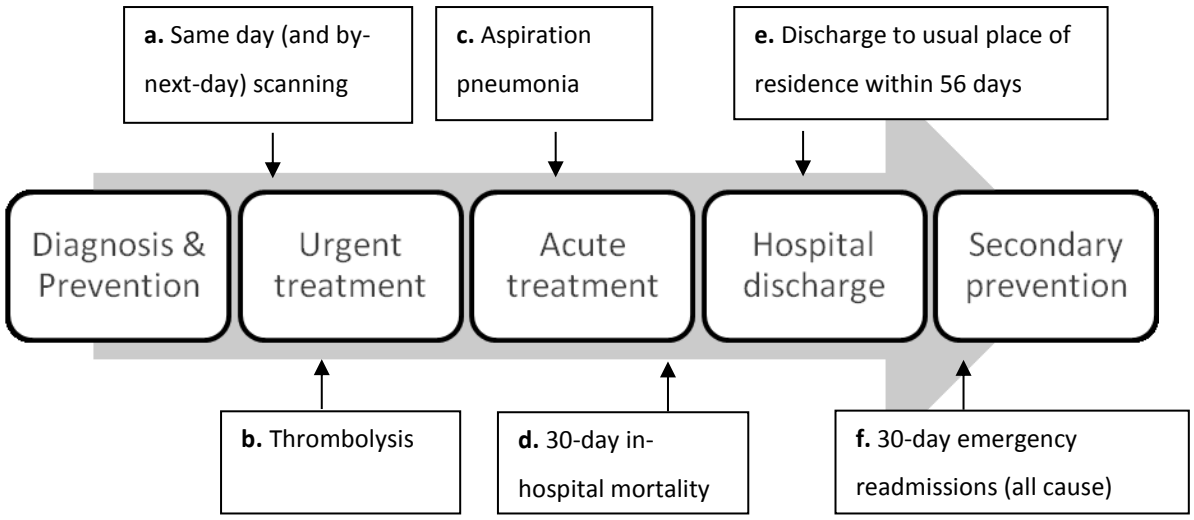
5.3.1.1. Selection criteria revisited

Recall the criteria listed earlier (paragraph 3.5.1.2, p76) for the selection of indicators. For this set of studies on the stroke specialty, the experts involved in getting a consensus on the indicators are listed in paragraph 5.3.2.1, p119.

5.3.1.2. Selection of indicators

The literature review described in the previous chapter identified 20 broad categories of indicators with the potential to be applied to administrative data in England. Following the selection criteria, a subset of six indicators was chosen for application here. The indicators were chosen to cover the hospital stroke medical care pathway and include: process measures (e.g. scanning rates); outcome measures that are proxies for certain aspects of care (e.g. pneumonia rates, with higher rates indicating fewer patients receive swallow assessments); and outcome measures (e.g. emergency readmission rates) (Figure 9).

Figure 9. Positioning of the stroke indicators across the care pathway



5.3.2. Developing indicator definitions

5.3.2.1. Development process

The selected indicators were then refined using a multi-faceted approach.

1. I derived draft definitions using details from algorithms applied in existing literature, identified through the systematic review outlined in the previous chapter.

2. I used an iterative consensual approach to ascertain feedback from a multidisciplinary panel of expert, consisting of a senior stroke physician (who was also involved in the indicator selection), a public health doctor, a clinical coding specialist (who had previously worked on developing indicator definitions) and administrative data experts (from Dr Foster Intelligence and Imperial's Dr Foster Unit).

3. The revised definitions were published online – through the Health Service Journal website – as part of an open consultation on clinical indicators.[332]

4. Where necessary the panel of experts were consulted on addressing comments raised on the definitions from the open consultation.

5.3.2.2. Final definitions

The finalised indicator definitions are included below in Figure 10.

Figure 10: Details of stroke indicators used in analysis

	Indicator	Rationale	**Numerator** (discharges meeting the inclusion/exclusion criteria for the denominator with...)	Exclusions
a	Same day and by-next-day scanning	Brain scanning should be performed immediately (when indicated) or as soon as possible (NICE 2008). Hospitals receive incentive payment for immediate scanning.	OCPS codes for CT or MRI brain scan (U05.1/2 or U21.1/2 with Z01.9)	Patients who die on day of admission (same day scan) or within day (by-next-day scan)
b	Thrombolysis	Measure of provision of thrombolysis. Hospitals receive payment for providing thrombolysis.	OPCS codes for thrombolysis (fibronolytic drugs, X83.3)	Patients outside thrombolysis licence age range of 18 – 80
c	Aspiration pneumonia	Aspiration pneumonia is an indicator of no swallow assessment or related care	ICD-10 codes for aspiration pneumonia: J69.0 (due to food and vomit) and J69.8 (other solids and liquids)	Pneumonia codes recorded in episodes that end before the first episode with a stroke diagnosis
d	30-day in-hospital mortality	Some mortality following stroke is potentially avoidable	Flag for death at discharge and length of stay after stroke <30	None
e	Discharge to usual place of residence within 56 days	Proxy measure for successful outcome of rehabilitation and availability of on-going care	Length of stay after stroke <56 days, and admission source and discharge destination suggesting return to usual place of residence	Admissions which end in death
f	30-day emergency readmissions (all cause)	Some readmissions are potentially avoidable	Emergency admissions within 0-29 days of discharge	Admissions which end in death

5.4. Application and validation of indicators

5.4.1. Assumptions

The next stage was to extract records of stroke admissions from 1 April 2009 to 31 March 2010. The assumptions used in the algorithm for identifying strokes in this study - based on previous studies, consultation with clinical coders and review of coding guidance - are listed in Figure 11.[309,333] The assumptions include: diagnosis codes, fields and type; identification of the date of the stroke; treatment of transfers, including linkage into superspell admissions and assigning a responsibility to the hospital; and treatment of in-hospital strokes.

Figure 11: Assumptions for identifying stroke using Hospital Episode Statistics

Issue	Assumption	Rationale	Alternatives	Implication
Diagnosis codes	ICD-10 codes I60 – I64	Review of literature/ clinical coding advice	ICD-10 codes G08.X, G45 – 46, H34, I65 – 69, Q25.8, Q28, R47 have also been used to identify patients in stroke studies	*Incorrect stroke numbers:* the chosen criteria are quite strict, in comparison with most studies
Diagnosis field	HES primary diagnosis field only	Review of literature/ clinical coding advice and guidance	Some or all of the secondary diagnosis codes	*Underestimate total number of strokes:* those incorrectly coded in the secondary fields are excluded
Defining stroke type	Strokes categorised by 4-digits of first stroke diagnosis code meeting inclusion criteria (above)	Subsequent stroke codes could refer to secondary strokes	[1] Final stroke diagnosis; [2] Supersede coding of I64.X (not specified) if another stroke code is listed; [3] Three-digit codes	*Incorrect classification of stroke*: where stroke is subsequently re-diagnosed, strokes labelled under the initial, provisional diagnosis. This will affect case mix adjustments
Date of stroke	Start date of first spell with stroke diagnosis	Coding guidance	[1] Start date of episode with first stroke diagnosis; [2] Start date of admission	*Incorrect stroke date:* In-hospital strokes not recorded as a new spell may have early stroke date. Strokes not diagnosed in the first spell of admission may have late stroke date
Linking spells	Episodes of care linked to same superspell	Reduce double-counting	Analyse by spell rather than superspell	*Incorrect stroke numbers:* Superspells with >1 spell are not counted more than once
Responsible hospital	First hospital of admission	First hospital are responsible for onward transfers	Hospital providing majority of (e.g. most bed days) or key aspects of care	*Performance assigned to wrong hospital:* if local pathways involve hospital transfers, the performance of the initial hospital will be influenced by the care of subsequent hospital
Type of admission	Emergency and elective	To not exclude in-hospital strokes	Emergency admissions only	*Include non-acute stroke admissions* as some episodes with primary diagnosis of stroke might relate to treatment for a previous stroke

5.4.2. Risk-adjustment

5.4.2.1. Overview of case-mix adjustment

Once an extract had been obtained using the above criteria, the first step was to identify outliers, which involved applying the indicator definitions to obtain denominators and numerators for each hospital. A logistic regression was used to calculate an expected number of numerator events based on the case mix for each hospital to account for: age, sex, socio-economic deprivation quintile, number of previous admissions, co-morbidities (Charlson index), month of discharge, ethnic group, source of admission (including whether admitted as an emergency or elective patient), and stroke type (4-digit ICD-10 diagnosis code). Charlson index was fitted as a continuous variable, with all others fitted as categorical variables. The risk model would not converge for some specifications and, for these, admission source was omitted from the case-mix adjustment. The methods used are described in more detail in paragraphs 3.6.2 to 3.6.6.

5.4.2.2. Reporting of process measures

For most process indicators, risk adjustment plays a smaller role than it does for outcome measures.[144] Process measures (scanning and thrombolysis) are reported here as unadjusted (crude) rates, although adjusted rates were also calculated to check the validity of this approach.

5.4.3. Statistical testing and validation

Crude and standardised rates were plotted using funnel plots with 95% and 99.8% control limits and identified outliers (paragraph 3.6.8, p81). Hospitals' performance across the different indicators was compared to evaluate our hypothesis that certain indicators would be correlated. This involved 21 pairwise comparisons (the scanning measure was considered as two separate measures – same day and by next day scans), with further detail on this method included in the discussion below. Since these were hypothesised as being correlated and not a un-hypothesised statistical trawling exercise, Bonferroni correction or other adjustment to p-values for multiple comparisons was not used; however, I do report which correlations are significant at the 99.8% level.

5.4.4. Sensitivity analysis for coding practice

There is a risk that hospitals' performance for these indicators might be largely affected by variation in the way they code their data rather than due to differences in quality and safety. As such, the final stage of analysis was to investigate, at a hospital level, the consistency of coding practice and evaluate the relationship between any coding bias and hospital performance. The three coding practices investigated, by comparing to performance in a relevant indicator, were:

[1] 'coding depth', with the coding practice in some hospitals increasing the likelihood that they record secondary diagnoses and, therefore, identify complications and comorbidities;

[2] use of the ICD-10 diagnosis code I64 - *Stroke, not specified*, since it may bias the risk adjustment for the outcome measures; and

[3] date of stroke, since some hospitals may be more likely to assign a stroke code to the episode when the stroke was formally diagnosed whilst others would assign a stroke code to the date of admission.

For the latter practice, I calculated the average time-lags, by hospital, between the date used to signify the stroke in this study (start date of the first spell with a stroke diagnosis) and both the start date of the first episode with a stroke diagnosis and the start date for the superspell.

The effect on the sensitivity of the measures when including a hospital-level variable to account for variations in coding practice was investigated by fitting generalised linear models and again plotting on a funnel chart to identify outliers.

5.4.5. Alignment with analytical framework

Table 23. Analytical framework

	Domain	Study design	Method described	Results described
Importanc	1. Impact	**Descriptive, national rates**	-	**5.5.1.1**
	2. Amenable	**Descriptive, range**	-	**5.5.1.2**
		Demographic-level variation		**5.6.2**
Scientific soundness of indicators	3. Face and content validity	**Consultation**	**5.3.2.1**	**5.3.1**
	4. Precision	**Hospital results** displayed on **funnel plots**	**5.4.3**	**5.5.2**
	5. Construct validity	**Correlations** between the set of indicators, including **outcome v. process** indicators	**5.4.3**	**5.5.3**
	6. Minimum bias	**Risk-adjustment model**	**5.4.2**	**5.5.5**
		Sensitivity analyses of influence of coding error and practice.	**5.4.4**	**5.5.4**
Applicability	7. Data availability	**Application** of indicators to show ICD-10 based data is sufficient.	**By implication**	
	8. Reporting burden	**Application** of indicators.	**By implication**	
	9. External and ecological validity	Explicit statement of **assumptions** used.	**5.4.1**	

5.4.6. Analytical packages

All regression analyses were conducted using SAS version 9.2 using either the PROC LOGISTIC or PROC GLIMMIX procedures.

5.5. Results

5.5.1. Importance

5.5.1.1. National-level results

Across 147 acute English NHS hospitals, 91 936 stroke admissions were identified in the period April 2009 to March 2010. Of these, 2 522 (2.7%) died on the same day as admission, 15 846 (17.2%) died within 30 days of admission and 19 721 (21.5%) died before discharge. Of those patients meeting the inclusion criteria (discussed in Figure 11), 69.7% were scanned within one day of admission, 2.6% received thrombolysis, 5.3% had aspiration pneumonia, 72.8% were discharged to their normal place of residence, and 11.0% were readmitted as an emergency within 30 days of discharge.

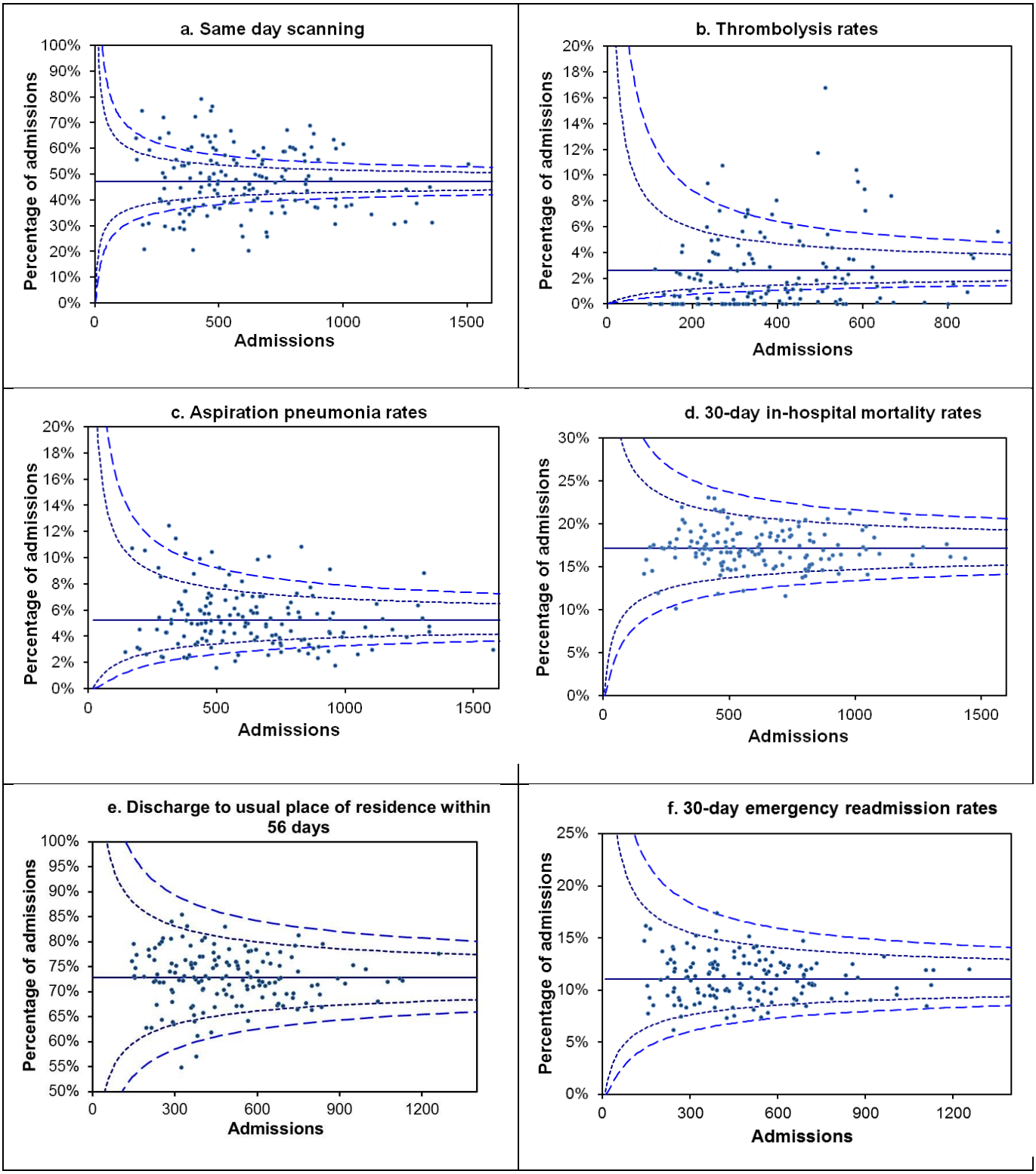
5.5.1.2. Amenable

The wide range in performance suggests scope for improvement. For instance, the rate of scanning by-next-day ranged from 41.5% to 86.9% and for 30-day in-hospital mortality from 10.1% to 23.1%. Similar work I conducted as part of a team using the scanning indicators is reported in the discussion (paragraph 5.6.2).

5.5.2. Precision

Displaying the hospital-level data on funnel plots highlighted the variation in performance (Figure 12). With the exception of the measure for emergency readmissions, all the indicators identified at least one hospital as having performance outside the 99.8% control limits (Table 24). The number of stroke admissions per hospital ranged from 171 to 1532.

Figure 12: Funnel plots of hospital-level performance across six indicators spanning the hospital stroke care pathway for 2009-10



Note: Each dot represents a hospital. The horizontal line refers to national average; short-gauge dotted line refers to p<0.025 significance level; long-gauge dotted line refers to p<0.001 significance level.

Table 24: Number of trusts identified as having statistically significant above or below average performance

Label	Indicator	National rate	Range	Number of hospitals higher than average[2]		Number of hospitals lower than average[2]	
		147 hospitals	[min, max]	**p<0.001**	p<0.025	**p<0.001**	p<0.025
a	Same day scan	47.1%	[20.4% – 79.3%]	**38**	47	**38**	50
a.1	By-next-day scan	69.7%	[41.5% - 86.9%]	**13**	32	**13**	28
a.2	Combined scan[1]	n/a	n/a	**9**	23	**10**	24
b	Thrombolysis	2.6%	[0% - 16.8%]	**13**	24	**63**	75
c	Aspiration pneumonia	5.3%	[1.6% - 12.5%]	**10**	21	**14**	27
d	30-day in-hospital mortality	17.2%	[10.1% - 23.1%]	**0**	8	**3**	10
e	Discharge to usual place of residence within 56 days	72.8%	[54.9% - 85.4%]	**0**	7	**2**	9
f	30-day emergency readmissions (all cause)	11.0%	[6.1% - 17.3%]	**0**	4	**0**	9

Note 1: Combined scan (a.2) was calculated as p<0.001 if both scan indicators had this value, and p<0.025 if both indicators at least p<0.025.

Note 2: Better care is indicated by higher than average performance for measures a, b, and e, and lower than average performance in measures c, d and f.

5.5.3. Construct validity

Across the indicators, there were six pairs of indicators which had a statistically significant correlation at the 95% level, of which two were significant at the 99.8% level (Table 25).

Table 25: Coefficient of correlation between pairs of indicators

	By-next-day scan	Thrombolysis	Aspiration pneumonia	30-day in-hospital mortality	Discharge to usual place of residence within 56 days	30-day emergency readmissions (all cause)
Same day scan	0.570**	0.040*	0.035*	-0.040*	-0.023	0.004
By-next-day scan		0.050	0.014	-0.045*	0.002	-0.021
Thrombolysis			-0.078	-0.048	0.228	-0.084
Aspiration pneumonia				-0.067	-0.720**	0.075
30-day in-hospital mortality					-0.355	-0.042
Discharge to usual place of residence within 56 days						0.001

Note: Correlation significant at the 95% level marked with '*'; those at 99.8% level with '**'. Coefficient is Pearson's r.

5.5.4. Minimum bias: coding practice

5.5.4.1. Coding depth

At a hospital level, the average number of distinct diagnosis codes used per admission ranges from 5.0 to 10.7. There was a statistically significant but weak correlation ($p = 0.002$, $R^2 = 0.067$) between coding depth (measured as average number of unique diagnosis codes in an admission) and performance against the aspiration pneumonia measure. Of the 25 hospitals identified at the 99.8% level in the original regression, 20 (80.0%) were again flagged as outliers at this significance level when coding practice was included in the regression, with no hospitals flagged as having statistically different performance highlighted in the reverse performance category.

5.5.4.2. Case ascertainment: episodes

Secondly, across hospitals the proportion of strokes diagnosed as ICD-10 code I64 – *Stroke Unspecified* varied from 0.2 to 42.6%; however, there is only a weak correlation ($p = 0.125$, $R^2 = 0.016$) between use of this code and performance in the outcome measure of in-hospital mortality within 30 days. There was a statistically significant but weak correlation ($p = 0.04$, $R^2 = 0.028$) between the proportion of patients without a specific stroke diagnosis and the hospital's one-day scanning rates; this association was expected as a scan is required to determine whether a stroke is ischaemic or haemorrhagic.

5.5.4.3. Case ascertainment: date

Nationally, the date-of-stroke calculations (start date of spell with stroke diagnosed, as described in Figure 11 and in the methodology above) were the same for 97.4% (and within one day for 97.9%) of patients, as compared with using the superspell start date. In comparison, when using the date of the first stroke episode as the alternative, the corresponding figures were 88.7% (93.5%). At a hospital level, there was a statistically significant but weak linear correlation ($p = 0.006$, $R^2 = 0.051$) between the proportion of patients identified as having uncertain stroke date and performance in this measure. Of the 76 hospitals identified at the 99.8% level in the original regression, 50 (65.8%) were again flagged as outliers at this significance level when coding practice was included in the regression, with one further hospital (1.3%) flagged as having statistically high rather than low scan rates.

5.5.5. Risk model

The case-mix models were better at predicting 30-day in-hospital mortality ($c = 0.735$, $r^2 = 0.158$) than readmissions ($c = 0.622$, $r^2 = 0.039$).

Table 26: Model fit statistics

		Model fit statistics	
Label	**Indicator**	**c**	**r^2**
a	Same day scan	0.616	0.062
a.1	By-next-day scan	0.645	0.100
b	Thrombolysis	0.771	0.108
c	Aspiration pneumonia	0.661	0.049
d	30-day in-hospital mortality	0.735	0.158
e	Discharge to usual place of residence within 56 days	0.664	0.095
f	30-day emergency readmissions (all cause)	0.622	0.039

Note: r2: the max-rescaled r-square score reported from PROC LOGISTIC

5.6. Discussion

5.6.1. Summary of findings

The results show the potential for using hospital administrative data, meeting the Government's intention to measure stroke care and, moreover, to highlight potentially significant variations in the quality and safety across the care pathway. Six measures of quality and safety, covering the acute stroke care pathway, were applied to English hospital administrative data identifying 91 936 strokes. Five of the six indicators identified hospitals with statistically outlying performance at the 99.8%

level. In particular, the analysis reiterated the high impact of the subject matter, with large numbers of potential adverse events.

5.6.2. Amenable: scope for improvement

For most of the measures there is no clinical consensus or guidelines for what actual levels are acceptable and, in these cases, hospitals should strive to operate at the level of the best performers. Indeed, regulators and guideline-setters could use this analysis to suggest levels of performance that are attainable. The exception to this rule is in access to a scan, for which extant guidelines recommend that all stroke patients should receive brain imaging "within a maximum of 24 hours after onset of symptoms" and, as such, the actual performance of 69.7% of patients receiving a scan within one day of admissions is unacceptable.[334] While some of this shortfall in performance against the measure could be due to incomplete coding of the procedure, audit data in England suggest a similar figure.[335] The level of performance also falls short of the Danish NIP proposed a benchmark that providers should give at least Proportion of patients who have a CT/MR scan on the day of admission.[336] This project also suggested a standard of a maximum of 15% of stroke patients dying within 30 days which, again, was not met in England where 17.2% died within this timeframe.[336]

Work I conducted as part of a team during the development of the indicator set and investigated the inequalities on this scanning indicator further since being a process measure and with only limited and explicit contraindications for the procedure, the expectation is for limited variation in a good performance health service. This analysis, again on English HES data but using a different time period (2008/09) showed that young patients were more likely to be scanned quickly than old patients (Adjusted OR for 10 year increase of age = 0.90; P<0.001; 95% CI = 0.89–0.91); the relations for comorbidity were also similar.[222] These findings again support the case that improvements can be made.

5.6.3. Precision

By using funnel plots and this control limit, I was able to account for random variation. This technique was more appropriate than one stroke study which identified outliers as being those 10% from expected and, therefore, not accounting for the scale of random variation being linked to the number of cases.[243] If the deviation were entirely due to random variation, you would only expect to identify one outlier for every 3 to 4 measures (given there are 147 hospitals in the sample). Further to chance, differences in performance may be due to case mix, how the data were collected, or quality of care.[337] Given the ambition was to compare hospitals on the basis of the last of these factors, the previous two also need to be accounted for.

5.6.4. Minimum bias

5.6.4.1. Adjusting for case-mix

Previous work has highlighted the importance of adjusting for case mix when comparing performance in stroke care.[338] I accounted for case mix using patient-level logistic regression to calculate expected number of events for the outcome measures. However, at a patient level, some significant case-mix factors for stroke, such as severity of stroke and pre-stroke function,[339] are not directly recorded within the data and therefore some of the variation may still be caused by differences in case-mix. If we hypothesis that there are small areas of extreme case-mix but, at a larger geographical level, case-mix is more homogenous coupled with units and trusts predominantly treating patients from their local area then it holds that any bias from this potential 'categorisation error' will be further diminished by the large number of stroke patients admitted at each hospital, ranging from 171 to 1532.

I report some model fit statistics which suggest that the case-mix models were better at predicting mortality ($r^2 = 0.158$) than readmissions ($r^2 = 0.039$) or pneumonia ($r^2 = 0.049$). This information is not for making conclusion on the appropriateness of the case-mix adjustment – since it is not possible to disaggregated the effect of unmeasured differences in case-mix to differences in quality of care provided – but should be considered when interpreting the results.

One specific issue relating to case mix originates from stroke care being increasingly delivered in regional networks, whereby certain hospitals are responsible for the urgent care of patients, whilst the other hospitals may take responsibility for rehabilitation. In these cases, the ambulance service has bypass protocols so that patients requiring urgent treatment are taken directly to the designated hospital. Therefore, the hospital designated to provide urgent care is more likely to receive a higher percentage of patients eligible for thrombolysis and, as such, their thrombolysis rates will be higher. In contrast, hospitals that are not designated to provide urgent care receive fewer eligible patients and, as a result, their performance against the measures of urgent stroke care are likely to be worse.

5.6.4.2. Accounting for variation in coding practice

The variation in performance due to how data were collected is harder to disaggregate; some of the possible variations in coding practice, for example, would mimic recognised alternative hospital pathways. However, I showed that three coding issues that were central to the assumptions for extracting the data explained only a small proportion of the differences in performance against the measures.

5.6.5. Errors in underlying data

There are few recent studies specifically investigating the accuracy of coding stroke care in healthcare systems using the ICD-10 framework; however, one article on coding in England found that coding of stroke diagnoses was excellent and elsewhere administrative data have been recommended for use in tracking progress and identifying problems for further review.[340,325] Previous studies have suggested coding is improving and, specifically in the instance of stroke care, the fact that some of the codes are new - for example the scanning procedure codes were only introduced in 2006 - means that it can be expected that there will initially be under-use (and so under-reporting) against these codes and that recording will improve over time.[341]

The analysis of coding practice highlighted variations in the data recording practice of hospitals, even where guidance exists. For example, the current ICD-10 Clinical Coding Instruction Manual "directs the coder that on emergency admissions for strokes it is of paramount importance that the coder assigns the code for stroke in the primary position". [333] However, variations in the number of stroke codes recorded in secondary diagnosis fields above what would be expected from different prevalence of stroke as a co-morbidity suggest there are, from hospital to hospital, differing proportions of stroke recorded in secondary diagnosis fields.

An English study evaluating the accuracy of the coding of stroke diagnosis found that the cause of errors was predictable, with confusion in coding of different types of strokes. This implies that further guidance and training may be needed to ensure consistency in coding.[340] One recent suggestion to facilitate an improvement is to develop better relationships between coders and clinicians.[306] This paper explicitly outlines the assumptions used for identifying strokes in HES data and, therefore, can form the basis of a debate within the medical and coding practitioner community about how to further develop these indicators and the dataset itself.

5.6.6. Face and content validity

Whilst previous studies have used hospital administrative data to measure the performance of aspects of stroke care, none have brought together the multiple facets of the care pathway and instead mostly focused on one area, such as mortality, and on national trends rather than hospital comparisons. A combination of process and outcome measures was used in this study, so benefiting from the advantages of process measures (which tend to be more sensitive to differences in the quality of care and offer a clear action for improvement) and outcome measures (greater intrinsic interest, high face validity, and can reflect all aspects of care, including those that are otherwise difficult to measure such as technical expertise and operator skill). [337,342] Face validity can be derived from the fact that the indicators appeared in existing literature or clinical guidelines.

As described in the previous chapter, overall level of quality of care at a provider-level cannot be deduced from the results of only a small set of indicators;[223] however, two-thirds of studies within the literature review applied only one indicator. This study included six indicators covering the acute care pathway and, therefore, is likely to have better content validity (extent to which a measure represents all facets of the subject being investigated) over these previous more limited evaluations.

5.6.7. Construct validity

In the absence of a gold standard dataset to compare results, some assurance on validity can be taken by comparison with other indicators of performance. I investigated whether there were associations between hospitals' performance across the different measures. In the existing literature, one study compared performance across different indicators constructed from the same dataset; however, in this instance, few pairs had significant correlation.[223] One would expect some correlation between good performance across all the indicators with, for example, good management of a stroke unit likely to affect all the quality and safety indicators to some extent. I also predicted that there were two specific clinical explanations, with correlations expected if either:

1. two indicators measure different events on a defined clinical pathway; or
2. an outcome indicator measures the results of a process indicators.

Indeed these two categories could explain the significant correlations that I identified between: same day scanning and next day scanning (positive correlation); same day scanning and thrombolysis (positive); scanning rates and mortality (negative); and pneumonia and discharge to usual place of residence within 56 days (negative). However, the causation in this latter correlation, between a proxy outcome measure for swallow assessments and discharge destination, is unclear as developing pneumonia is likely to increase length of stay (and so decrease likelihood of discharge within 56 days) whilst a delayed discharge would inherently increase the chance of developing pneumonia in hospital. Further analysis reveals that at least a third (32.1%) of these pneumonia were diagnosed within 4 days of stroke; however, since the date of diagnosis is not specifically recorded in HES data this figure represents an absolute minimum and, as such, I cannot infer any causal trends between length of stay and pneumonia rates.

The one unexpected, statistically significant result was the positive correlation between same day scanning and aspiration pneumonia rates. I expected that hospitals with high scanning rates would also be likely to have good processes to ensure urgent swallow assessments and, therefore, have lower rates of aspiration pneumonia (so a negative correlation). A plausible explanation for this unexpected result is that a hospital with more comprehensive coding practices is more likely to

record both the scanning procedure code and pneumonia diagnosis code compared to a hospital with less rigorous coding.

5.6.8. Strengths of this study to other stroke measurement efforts

In the NSSA, there is already a comprehensive tool for measuring hospital performance in stroke care. However, the underlying data are self-reported, and the NSSA's two-year reporting cycle coupled with time-lag before publication prohibits its use as a real-time monitoring tool. A recent positive development has been the introduction of the Stroke Improvement National Audit Programme (SINAP) which collects real-time data on stroke patients albeit this only covers the first three days of care and does not include all hospitals. The relative advantages of using administrative data were established earlier (paragraph 2.2, p42), but recall that HES has the advantage of being: longitudinal; timely; covering all hospital admissions; and being relatively cheap, costing £1 per record to collect compared with around £10 - £60 per record for clinical registers.[89] Whilst NCHOD have made use of some of the potential advantages of administrative data, in assessing a limited number of aspects of stroke care, their indicators are only updated annually and at the time of writing, hospital mortality from stroke was two years out of date.

5.6.9. Criterion validity

As well as these considerations on whether the indicators' algorithms are functioning as intended, further assurance needs to be taken about whether the results of the indicators reflect actual performance. The most common process for validating the results is to compare against a gold standard. The most applied and comprehensive method for assessing hospital stroke care in England is the NSSA, which suggests it would be the most appropriate dataset to compare results based on administrative data. However, direct comparison with the published NSSA results is not easily achieved due to differences in definitions. In addition there is more statistical uncertainty associated with the NSSA clinical data, given that it is based on a small sample of around 60 patients per hospital.

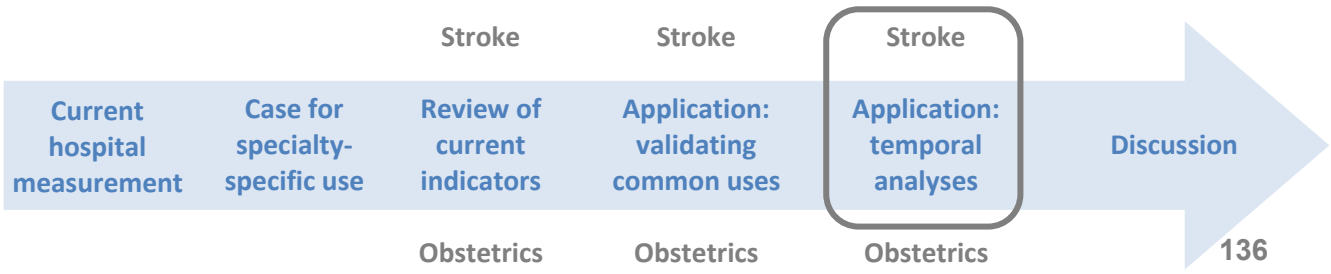
5.6.10. Conclusion

This study shows that hospital administrative data have the potential for benchmarking the standard of hospitals performance over the acute stroke care pathway, in a timely way and without increasing the bureaucratic burden on hospitals. The additional analysis described in the discussion - on how scanning levels were dependent on patient characteristics - also support the argument that things could be improved.

Whilst neither the lack of credible gold standard nor evidence of bias from coding practice render the indicators redundant, it does temper the extent to which they could be used, such as for regulating hospitals. This study shows that HES provides the facility to record some key process and outcome measures across the care pathways in a cheap and timely manner. The chapter gives a comprehensive overview of feasibility and this should determine future application of these indicators. On the more sophisticated end of the scale, these results could be linked to structural measures, such as existence of specialist stroke units or a 24-hours service, to investigate their effectiveness (and also further evaluate the extent to which performance is amenable). With suitable modification and further development, these data could also be used routinely as part of the National Stroke Sentinel Audit and replace some aspects of the self-assessment.

Chapter 6.

Stroke temporal trends (longitudinal)



Overview

Context

The literature review revealed that less than a third of studies on stroke care had used the longitudinal nature of administrative data. This chapter investigates potential longitudinal applications of indicators to address both the gaps in the literature and also to provide evidence on the whether performance is amenable to how services are structured.

Methods

Application of indicators covering the acute stroke care pathway to investigate changes in performance over time, at both a national and regional level. Where appropriate, results were adjusted for case-mix using logistic regression and significant changes in performance were identified using control charts. Specifically, interrupted time series analysis with a concurrent control group study design was used in the regional analysis to assess the impact of the Hyper Acute Stroke Units (HASUs) in London.

Finding

The longitudinal analyses highlighted improvements in performance and a seasonal effect for some measures. Compared with areas outside London, 7 day in-hospital deaths rate reduced significantly following the restructuring of services, as did aspiration pneumonia. However, same day brain scans showed a small but significant reduction following the intervention, as well as a slowing down in the rate of increase.

What this chapter adds

The first discussion of how using temporal analyses can overcome commonly cited concerns on the application of administrative data and can be used to evaluate the impact of healthcare initiatives.

Acknowledgements and related papers

I undertook the national longitudinal analysis. The regional study was led by Roxanna Alexandra with all authors (listed below) involved in the study design, analysis and manuscript revision of the regional analysis.

Alexandrescu R, Lee JT, Bottle A, Palmer W, Aylin P. *Effect of stroke services reorganisation in London: a controlled interrupted time-series analysis* (submitted for publication).

6.1. Background

6.1.1. National trends

The review of stroke literature (Chapter 4) highlighted the unmet potential to use the longitudinal nature of administrative data to highlight potential quality and safety issues within stroke care. Measuring performance over time – rather than a snapshot cross-sectional analysis – might also mitigate against some of the hospital-level variations in unmeasured case-mix and coding practice. This was underlined by the hospital-level cross-sectional analysis (Chapter 5) which was found to be, at least in part, biased by variations in hospital-level coding practice. In England, previous studies have suggested that stroke care has improved but the trend in many aspects of stroke treatment and outcomes remains unclear.[151,343]

6.1.2. Regional analysis

Such trend analysis is particularly useful when evaluating any change in structure or process of a particular provider unit, such as the implementation of regional stroke units, and this is investigated by presenting results of the evaluation of a regional reconfiguration. Prior to 2010, provision of stroke care in London was complex, with only 53% of patients treated on a dedicated stroke ward and care spread across a number of units.[344] To improve the quality of service, eight Hyper Acute Stroke Units (HASUs) were established in London from February 2010. The units, which are dedicated to treating stroke patients, are open 24 hours, seven days a week to offer immediate access to stroke investigations and imaging, including CT brain scan and clot-busting thrombolysis drugs. These units replaced the previous disparate services for acute stroke care.

6.2. Objectives

In this chapter I investigate the potential to use the longitudinal nature of administrative data to identify important shortcomings in quality and safety in a way that overcomes some of the known limitations to such indicators. The regional analysis was conducted as a case example to show that the indicators can be used to assess the impact of interventions, in this case the HASU policy in London, and in doing so show that quality and safety can be influenced by the way services are provided. This part of the study is not intended as a full evaluation of the HASU policy but to demonstrate the potential of the indicator set I developed.

6.3. Methods

6.3.1. Indicators

The indicators used for this part of the study are similar to those used in the previous chapter (paragraph 5.3.1.2, p118) and were identified in the literature review (Chapter 4). The only amendment made to these indicators, for the purpose of this weekend part of the study, was to use 7-day rather than 30-day mortality. The indicators include both process and outcome measures:

a) Receiving a brain scan on the day of admission;

b) Receiving thrombolysis treatment;

c) Diagnosed with aspiration pneumonia (complication) in hospital;

d) 7-day in-hospital mortality;

e) Discharge to usual place of residence within 56 days; and

f) 30-day emergency readmission (all cause).

6.3.2. Analysis

6.3.2.1. Data extract

For the national trends, the details of stroke admissions from 1 April 2007 to 31 December 2009 were extracted from the Hospital Episode Statistics (HES) database. The indicator definitions were applied to the extract to obtain denominators and numerators, categorised for month of admission for stroke. For the regional analysis, data from 1 April 2006 to 31 March 2012 were extracted, to account for the differing demands on this analysis and the availability of additional data at the point that this analysis was conducted.

6.3.2.2. Risk-adjustment

Variables expected to influence the outcome of the association between the quality and safety indicators and day of admission were extracted: age; sex; socio-economic deprivation quintile; number of previous admissions; co-morbidities - Charlson index with weights derived from all England admissions;[345] ethnic group; source of admission; and stroke type. The method for risk-adjustment is covered in paragraph 3.6.2, p79.

6.3.2.3. National data analysis

National performance across the six measures was produced, disaggregated by month and, where appropriate, multiple linear regressions were used to adjust to the effect of the covariates. Results

were also displayed by plotting – using April 2007 as a reference – odds ratios and 95% confidence intervals, by month of admission for stroke.

6.3.2.4. Regional data analysis

For the regional analysis, the quarterly unadjusted rates for the process and outcome indicator before and after the implementation of HASU policy were plotted. Using an interrupted time series (ITS) negative binomial regression model, the impact of HASU policy was assessed. While taking the underlying trend into account, this model estimates both the level and slope changes of the time trend after the policy change. A dummy variable (coded as 0 for the period before and 1 after implementation of the policy) and a continuous variable (set to 0 prior HASU policy, then equal to the length of time since HASU policy was implemented) were included. As we use quarterly data, the beginning of the reorganisation period is modelled as January instead of February (and ending in June).

To remove potential effect of other national wide policies on our outcomes, a comparison group was included in the ITS model using data from rest of England. To do so, a dummy variable representing region (coded as 1 for London and 0 for other areas in England) was included, and the model was fitted with interaction terms between the dummy variable for region and post-intervention step/slope terms.[346,347] As the model estimates the effect of HASU after adjusting for a concurrent comparison group, a negative/positive change in the step (or slope) will indicate a larger decrease/increase in the rates in the intervention group (London) compared with the control group (the rest of England).

6.3.2.5. Software

Statistical analyses were conducted using SAS version 9.2 software's PROC GENMOD and PROC LOGISTIC procedure for regression analysis.

6.4. Results

6.4.1. National

6.4.1.1. Overview of results

The number of stroke admissions rose by 9.8% from 85,287 in 2007/08 to 93,621 in 2009/10 (Table 27). Between these two years, there were reported improvements in scanning (from 29.2% in 2007/08 to 46.5% in 2009/10), thrombolysis (0.3 to 2.6%), 30-day in-hospital mortality (20.1 to 17.1%), and discharge to usual place of residence (68.7 to 72.6%). Conversely, reported performance

got worse for aspiration pneumonia (4.8 to 5.2%) and 30-day all-cause emergency readmissions (9.2 to 11.05).

Table 27: Annual trend in performance

Label	Indicator	2007/08	2008/09	2009/10 ^
		n = 85,287	*n = 89,712*	*n =93,621*
a	Same day scan	29.2	37.4	46.5
a.1	By-next-day scan	50.4	60.0	68.7
b	Thrombolysis	0.3	1.1	2.6
c	Aspiration pneumonia	4.8	5.0	5.2
d	30-day in-hospital mortality	20.1	18.8	17.1
e	Discharge to usual place of residence within 56 days	68.7	70.2	72.6
f	30-day emergency readmissions (all cause)	9.2	10.6	11.0

Note: **^** Proportions for 2009/10 may differ slightly from those reported in the previous chapter, where the results were limited to only those patients admitted to recognised acute trusts.

6.4.1.2. National longitudinal trends

Disaggregating national performance by month suggests there are different types of longitudinal trends (Figure 13):

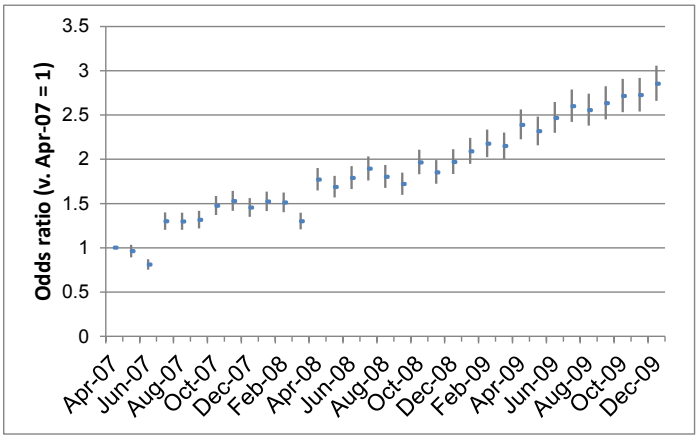
1. **Steady increases**. There has been a steady, statistically significant increase in scanning rates (panel a) over time, which can be closely approximated by a linear regression (R-squared 0.96, $\beta = +0.06$, $p < 0.001$). Similarly, thrombolysis rates have increased gradually over time (panel b).

2. **Overall trend with seasonal effect**. Mortality rates have also improved over time (R-squared 0.28, $\beta = -0.006$, $p=0.002$) but there is also evidence of a substantial seasonal effect (panel d) with rates highest around December. The pattern is similar for the rate of patients being discharged to usual place of residence (panel e) where there has been a general improvement with a seasonal increase around July.

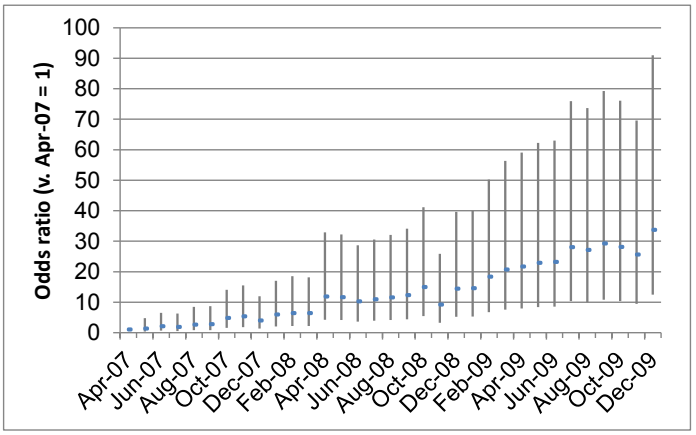
3. **No clear trend**. Figures for pneumonia (panel c) and readmission (panel f) rates showed only minimal seasonal or overall longitudinal trend.

Figure 13. Performance by month[1]

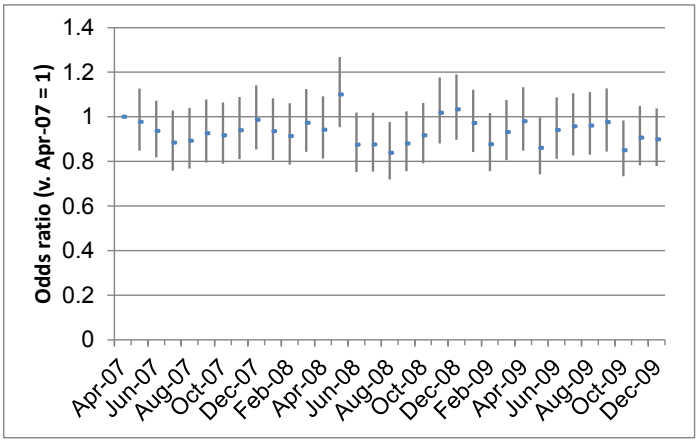
a. Immediate scan (unadjusted)



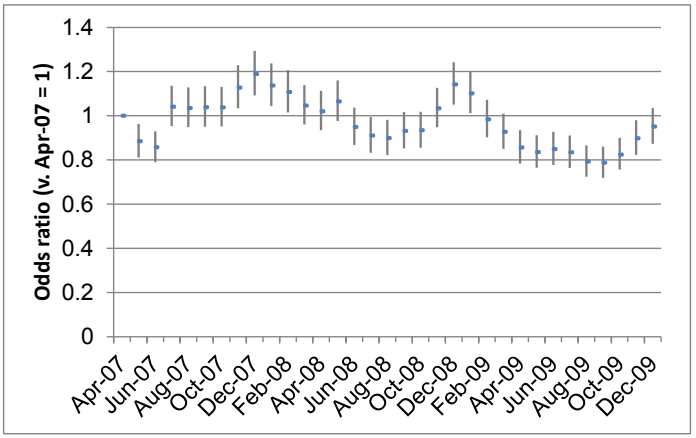
b. Thrombolysis (unadjusted)[2]



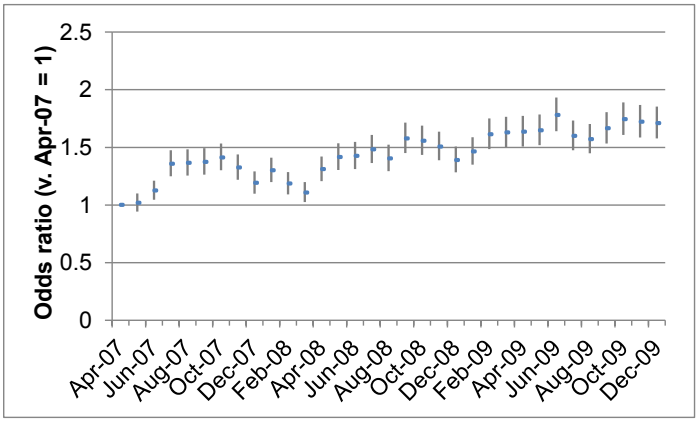
c. Pneumonia (adjusted)



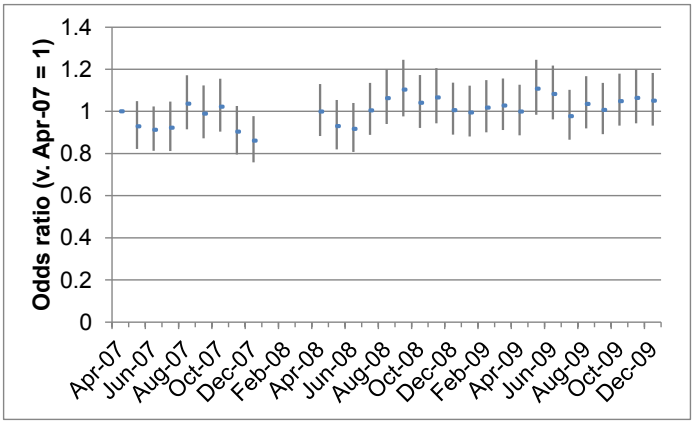
d. Death within 30 days (adjusted)



e. Discharge to usual place of residence (adjusted)



f. Readmission (adjusted)[3]



Note: 1. Data points represent odds ratios, with April 2007 used as a reference (1.00); vertical ranges, 95% confidence intervals.

2. The low number/rate for the reference day on the Thrombolysis chart results in large OR for later months; this is discussed later

3. For the indicator of 30-day emergency readmissions, data for Jan-08 to Mar-08 has been removed due to data issues

6.4.2. Regional trends

The regional analysis included 536,034 stroke admissions in England over the 6-year period to March 2012. Around one-in-nine (11.5%) of these admissions occurred in London. The average annual admission rate in London was 42,375 per year prior to the intervention. This admission rate increased slightly (by 4%) post intervention in London but remained static in the rest of England. The national rates are given in Table 28. Over the entire study period, London performed better than the rest of England on scanning, thrombolysis, and mortality rates, with the inverse true for pneumonia, discharge of usual place of residence, and emergency readmission rates.

Table 28. Summary of differences in performance

Indicator	National	London	Rest of England
Same day scanning	41.6%	49.3%	40.6%
Thrombolysis	2.7%	3.6%	2.5%
Aspiration pneumonia	5.2%	6.1%	5.1%
7-day in-hospital mortality	9.8%	8.0%	10.0%
Discharge to usual place of residence within 56 days	71.9%	71.3%	72.0%
30-day emergency readmissions	10.7%	11.7%	10.5%

Table 29 sets out the crude rates before and after the intervention for both London and the rest of England. In the these regions, there were improvements in scanning rates, thrombolysis, mortality and discharge destination. Conversely, readmission rates increased in both London and the rest of England. While crude rates of aspiration pneumonia decreased in London (6.3% to 5.6%) they increased elsewhere (4.9% to 5.4%).

Table 29. Crude rates before and after intervention

	London, rate (number)		**Rest of England**, rate (number)	
Indicator	Before	After	Before	After
	n = 37 051	n = 19 270	n = 289 491	n = 143 952
Same day scanning	38.3% (14 194)	62.5% (12 047)	30.2% (87 533)	55.3% (79 644)
Thrombolysis	0.7% (265)	4.9% (947)	0.5% (1 541)	3.3% (4 688)
Aspiration pneumonia	6.3% (2 326)	5.6% (1 088)	4.9% (14 087)	5.4% (7 721)
7-day in-hospital mortality	8.7% (3 235)	6.8% (1 311)	10.6% (30 792)	8.8% (12 702)
Discharge to usual place of residence within 56 days	55.9% (20 707)	59.9% (11 540)	53.3% (154 259)	60.2% (86 645)
30-day emergency readmissions	8.9% (3 282)	10.4% (2 006)	7.7% (22 176)	8.9% (12 827)

Notes: 'Before' denotes April 2006 to December and 'After' denotes July 2010 to March 2012

Table 30 shows the results of the interrupted time series analysis for London, relative to the rest of England. There was a significant step change fall in aspiration pneumonia (OR 0.81) and 7 day in-hospital death (0.93). Same day brain scans appeared to decrease relative to England (0.91). There was a decrease in slope relative to the rest of England (0.95), however there was also a reduction in slope for same-day brain scans (0.98) and discharge to usual place of residence within 56 day (0.99).

Table 30. Time-Series regression analysis of stroke indicators in London relative to the rest of England

	Step – change at time of intervention in London vs England excl. London			***Slope*** *– change for post intervention trend in London vs England excl. London*		
Indicator	**OR**	**95% CI**	**p-value**	***OR***	***95% CI***	***p-value***
Same day scanning	0.91	(0.88-0.94)	**< 0.0001**	*0.98*	*(0.97-0.99)*	***< 0.0001***
Thrombolysis	1.11	(0.95-1.29)	0.182	*0.98*	*(0.96-1.00)*	*0.091*
Aspiration pneumonia	0.81	(0.74-0.87)	**< 0.0001**	*0.95*	*(0.93-0.96)*	***< 0.0001***
7-day in-hospital mortality	0.93	(0.86-0.99)	**0.035**	*0.99*	*(0.97-1.00)*	*0.109*
Discharge to usual place of residence within 56 days	0.96	(0.93-0.99)	**0.002**	*0.99*	*(0.98-0.99)*	***0.001***
30-day emergency readmissions	1.02	(0.96-1.09)	0.457	*1.01*	*(0.99-1.02)*	*0.157*

Note: Selected time periods based on pre-intervention trend linearity: Same-day brain scan (July2007-March 2012); Thrombolysis (October 2007-March 2012); Aspiration pneumonia (April 2007-March 2012); 7 day in-hospital death (April 2007-March 2012); Discharge to usual place of residence 56 day (April 2008-March 2012); 30 day emergency readmissions (April 2007-March 2012). Transition period (January to June 2010) removed.

6.5. Discussion

6.5.1. Longitudinal

The longitudinal analysis showed that administrative data could be used to identify potentially important trends in stroke care quality and safety. In particular, the analysis showed significant improvements over four of the six indicators. Interestingly, the analysis also showed a strong seasonal trend in stroke outcomes of death rates and proportion of stroke patients discharged to their usual place of residence. This is an area which, to date, has been poorly researched.

A key strength of this on-going and longitudinal application of administrative data is that it is not be possible with the prominent existing tools for monitoring stroke care in England; the NSSA is

collected over a short period of time and the Stroke Improvement National Audit Programme (SINAP) only covers the urgent part of the care pathway.

6.5.1. Regional analysis

The results of the interrupted time series analysis are consistent with the expected effect if the reorganisation of stroke services had improved the quality of care in London. Importantly 7 day in-hospital deaths showed a significant step reduction following the restructuring of services, as did aspiration pneumonia when compared to trends in the rest of England. Interestingly, same day brain scans showed a small but significant reduction compared with the rest of England following the intervention, as well as a slowing down in the rate of increase. This may be due to a possible ceiling effect, i.e., a limit beyond which any improvement is less likely to continue. In any case, London scan rates are consistently higher than the rest of England.

The analysis supports the previously (paragraph 5.6.2, p130) made argument that performance against the stroke indicator set is amenable. A wider evaluation and discussion about the value of HASUs is beyond the scope of this project and, rather, this analysis was presented to support the evidence against the analytical framework item of 'impact'.

6.5.2. Limitations

6.5.2.1. Changes in coding practice

This application mitigates against the bias from variations in coding practice between hospitals, as even if there is evidence of a change in distribution of patients between hospitals over time then this clustering can be accounted for in such regression analyses. As discussed earlier, the level of thrombolysis coding might be expected to increase due to financial incentives attached with the recording of this treatment and, also, due to the low stating point because thrombolysis has only recently been licensed in England.

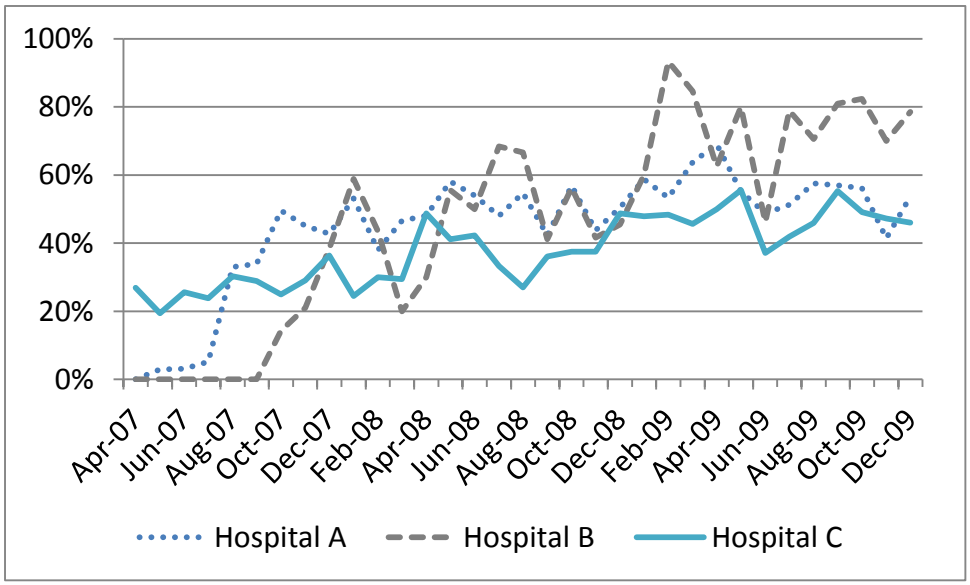
6.5.2.2. Interpreting comparison over time

The analysis also highlighted the potential issues with making comparisons over time. In particular, the choice of using the first month as the reference was unhelpful when interpreting the change in thrombolysis performance since this first month was an extreme performance and likely to be a consequence of a lack of incentive at the time to code thrombolysis correctly and the treatment not being formally approved. Indeed, the odds ratio are substantially different depending on what reference month is chosen; for example, with reference to April 2007 (month 1) the odds ratios for month 2 (May 2007) and month 33 (December 2009) are 2.109 (95% confidence interval 0.680 – 6.539) and 33.670 (12.458 – 91.004), whereas if April 2008 (month 13) is used as reference then the

estimated odds ratios are more precise: 0.114 (0.049 – 0.268) and 2.846 (2.027 – 3.994), respectively.

The issues are not isolated to just thrombolysis nor national-level analysis. Statistical control charts are common techniques for monitoring performance in healthcare.[348] Figure 14 shows performance for three anonymised hospitals (labelled Hospital A, B, C which are based on actual data from the study) against the same day scanning code. Taking the rule that a trend of six of more consecutive increases (or decreases) signifies unexpected variation[349], then Hospital A would be the only hospital identified as a potential 'good performing' outlier although would not necessarily be the obvious interpretation when comparing the trends and overall performance levels. The issue described is particularly problematic when changes in coding practice could mimic actual improvements (i.e. a hospital which is aiming to improve the recording of scans will automatically appear to be performing better).

Figure 14. Example of hospital trends

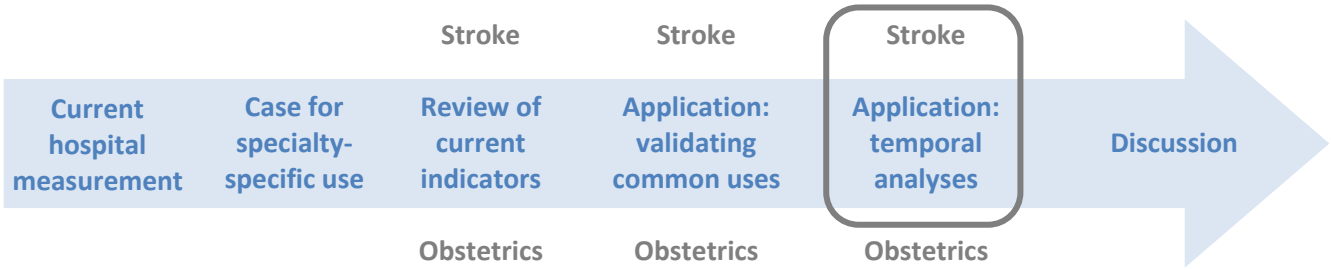


6.5.3. Conclusion

This chapter show that it is possible to address some of the potential bias from case-mix and coding practice using longitudinal analyses; however, given the changes in coding practice over time which coincides with potential changes in service provision (and therefore hard to detect) this application is still susceptible to bias. The regional analysis had the advantage of providing evidence that some of the shortcomings in care are amenable to health care organisation.

Chapter 7.

Stroke temporal trends (day of admission)



Overview

Context

The literature review revealed that no study had comprehensively assessed weekend care. This chapter investigates potential temporal applications of indicators to address both the gaps in the literature and also overcome the shortcomings in validity relating to cross-sectional application of these indicators as identified in the previous chapter.

Methods

Application of indicators covering the acute stroke care pathway to investigate the association between day of admission and adverse events. Where appropriate, results were adjusted for case-mix using logistic regression and significant changes in performance were identified using control charts.

Finding

The day-of-the-week analysis showed performance across five of the six measures was significantly lower at weekends at the 99.8% level. In particular, the rate of 7-day in-hospital mortality for Sunday admissions was 11.0% (adjusted odds ratio 1.26; 95% Confidence Interval 1.16 – 1.37 with Monday as reference) compared to an average of 8.9% for weekday admissions.

What this chapter adds

This chapter provides the most comprehensive evaluation of the “weekend effect” in stroke care.

Acknowledgements and related papers

Advice was provided by: Alex Bottle on statistical methods; and Charlie Davie on clinical implications.

Palmer WL, Bottle A, Davie C, Vincent CA, Aylin P (2012) Dying for the weekend: the association between day of hospital presentation and the quality and safety of stroke care. *Archives of Neurology*. 69(10):1296-1302

7.1. Background

7.1.1. Measuring the 'weekend effect'

Further to using the longitudinal nature of administrative data, an alternative temporal approach which could mitigate some of the known coding issues is to evaluate the association between quality of care and day of admission. Previous studies, across a range of countries, have identified higher mortality in patients admitted on weekends across a range of medical conditions; a phenomenon termed the 'weekend effect'.[253,350-353] This calls into question the idea that quality of care is equal irrespective of when you present at hospital. Similarly, there have also been a limited number of international studies investigating the effect specifically in stroke care that have suggested poorer access to treatments and worse outcomes at weekends, including increased mortality and fewer patients returning to their usual place of residence.[244,354-358] However, other studies have not identified a significant association between day of admission and stroke mortality rates.[253,289,359,360] Therefore, debate remains over the existence and extent of the weekend effect in stroke care. The studies on out-of-hours stroke care are limited in number and most of these fail to stand up to the criticism of primarily focusing on short-term mortality and therefore not capturing wider aspects of the quality and safety of care.[361]

This chapter is unique in providing a comprehensive and current assessment of the degree to which quality and safety of stroke care is affected by whether a patient is admitted at the weekends, highlighting the potential for identifying clinically important issues using this readily available resource.

7.2. Objectives

In this chapter I investigate the potential to use the longitudinal nature of administrative data to identify important shortcomings in quality and safety in a way that overcomes some of the known limitations to such indicators.

7.3. Methods

7.3.1. Indicators

The indicators used for this part of the study are similar to those used in the previous chapter (paragraph 5.3.1.2, p118) and were identified in the literature review (Chapter 4). The only amendment made to these indicators, for the purpose of this weekend part of the study, was to use 7-day rather than 30-day mortality in keeping with the suggestion that it is a more appropriate

timeframe to evaluate the association between day of admission and mortality.[354] The indicators include both process and outcome measures:

a) Receiving a brain scan on the day of admission;

b) Receiving thrombolysis treatment;

c) Diagnosed with aspiration pneumonia (complication) in hospital;

d) 30 (or 7)-day in-hospital mortality;

e) Discharge to usual place of residence within 56 days; and

f) 30-day emergency readmission (all cause).

7.3.2. Weekend analysis

7.3.2.1. Data extract

For the weekend study, stroke admissions from 1 April 2009 to 31 March 2010 only were used. The indicator definitions were applied to the extract to obtain denominators and numerators, categorised for day of admission for stroke. The weekend was defined as the period from midnight on Friday to midnight on Sunday, with all other times defined as weekdays (the time of admission is not captured in HES).

7.3.2.2. Data analysis

Variables expected to influence the outcome of the association between the quality and safety indicators and day of admission were extracted: age; sex; socio-economic deprivation quintile; number of previous admissions; co-morbidities - Charlson index with weights derived from all England admissions;[345] ethnic group; source of admission; and stroke type. Where only one figure is given, process measures were reported as unadjusted rates (paragraph 5.4.2.2, p122). The method for risk-adjustment is covered in paragraph 3.6.2, p79.

Descriptive analysis of the patients, categorised by weekday or weekend admission for stroke were produced. Initially unadjusted (crude) rates for each of the six indicators were calculated for weekday and weekend admissions (these are plotted for the process measures: access to a scan and thrombolysis). Multiple logistic regression analyses were used to account for the covariates and estimate adjusted odds, across each indicator, for weekend compared with weekday admissions. Results were also displayed by plotting – using Monday as a reference – odds ratios and 95% confidence intervals, by day of the week of admission for stroke.

Analyses were carried out using SAS Version 9.2, using the PROC LOGISTIC procedure for regression analyses. I did not adjust for the clustering of patients within hospital as the hospital-level effects were found to be small.

7.3.2.3. Estimating impact

Using regression analysis on just weekday admissions, probabilities of in-hospital death and discharge to usual place of residence for different patient characteristics were derived. By matching these probabilities for each weekday admission, based on the patient's characteristics, indirectly standardised estimates for the outcomes as if those weekend patients had had similar rates as their weekday counterparts were calculated, with these particular results reported in the discussion.

7.3.2.4. Consultant specialty

To investigate one of the possible organisational factors that could cause variations in the performance against the process and outcome measures, an additional indicator – *specialty of responsible consultant* – as a proxy for whether a patient is admitted to a designated stroke unit was calculated.

7.4. Results

7.4.1.1. Descriptive overview

Across English NHS hospitals, I identified 93 621 stroke admissions in the period April 2009 to March 2010. Of these, 8 722 (9.3%) died in-hospital within 7 days of admission and 16 013 (17.1%) died within 30 days. Of those patients meeting the inclusion criteria for each indicator, 46.5% were scanned on the day of admission, 2.6% received thrombolysis, 5.2% had aspiration pneumonia, 72.6% were discharged to their normal place of residence, and 11.0% were readmitted within 30 days of discharge.

Of the stroke admissions, 24.9% (23 297) were admitted at weekends. The number of stroke patients admitted declined throughout the week from 15.6% on Monday to 12.4% on Sunday (Figure 15).

Figure 15. Number of stroke admissions, by day of presentation

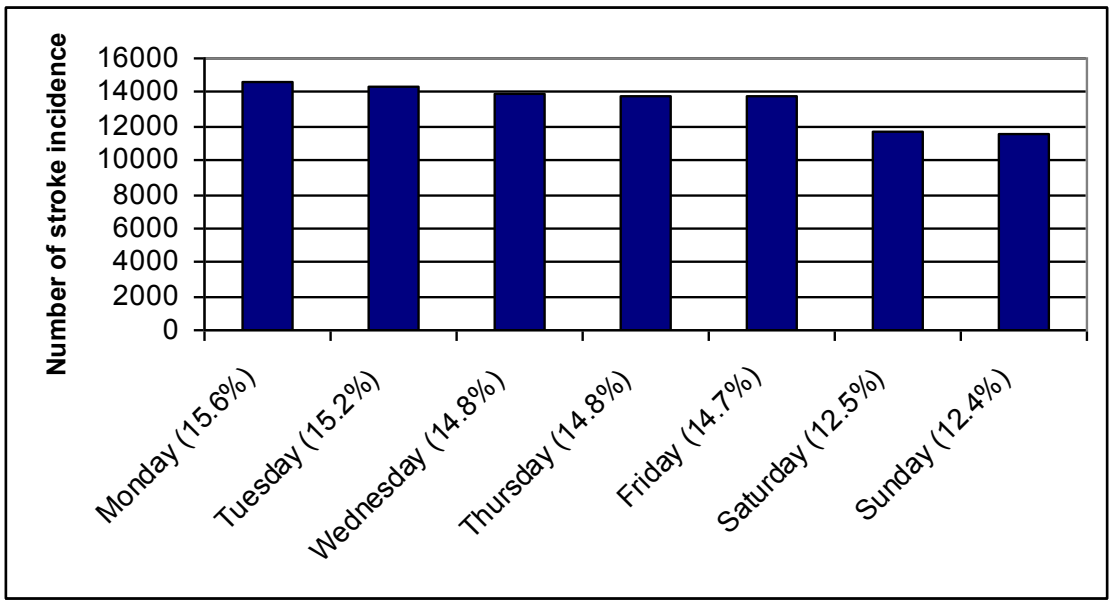


Table 31 gives a description of the characteristics of the patients included in the study.

Table 31. Characteristics of the study population

	Weekday	**Weekend**	
Characteristic	% of admissions (number of cases)		**p-value**
Age of admission in years (s.d)	73.8 (14.8)	74.5 (14.6)	0.048
Sex			< 0.001
Male	*48.5 (34 096)*	*47.2 (10 990)*	
Female	*51.5 (36 228)*	*52.8 (12 307)*	
Index of multiple deprivation quintile			0.983
1 least deprived	*16.8 (11 836)*	*16.9 (3 947)*	
2	*20.6 (14 507)*	*20.7 (4 817)*	
3	*21.4 (15 048)*	*21.3 (4 958)*	
4	*20.9 (14 706)*	*21.1 (4 903)*	
5 most deprived	*19.4 (13 668)*	*19.3 (4 491)*	
6 unclassified	*0.8 (559)*	*0.8 (181)*	
Charlson index of comorbidity			0.100
0 no comorbidity	*45.6 (32 034)*	*45.4 (10 573)*	
*1-4**	*13.8 (9 715)*	*14.0 (3 256)*	

*5-9**	*8.6 (6 050)*	*8.6 (1 991)*	
*10-14**	*18.7 (13 138)*	*18.8 (4 385)*	
*15-24**	*9.4 (6 629)*	*9.3 (2 168)*	
*25-50 highest comorbidity**	*3.9 (2 758)*	*3.9 (924)*	
Number of admissions in previous 12 months			0.157
0	*69.6 (48 960)*	*70.1 (16 327)*	
1	*19.3 (13 586)*	*19.3 (4 484)*	
2	*6.6 (4 650)*	*6.2 (1 444)*	
3	*4.5 (3 128)*	*4.5 (1 042)*	
Source of admission			< 0.001
*Admitted from home**	*89.2 (62 755)*	*89.9 (20 939)*	
*Hospital Transfer**	*9.7 (6 799)*	*9.1 (2 115)*	
*Other or Unknown**	*1.1 (770)*	*1.0 (243)*	
Emergency			0.001
non-emergency admission	*5.6 (3 925)*	*2.77 (645)*	
emergency admission	*94.4 (66 399)*	*97.2 (22 652)*	
Stroke Diagnosis			0.038
I60	*5.1 (3 570)*	*5.0 (1 175)*	
I61	*11.2 (7 864)*	*11.8 (2 738)*	
I62	*6.2 (4 376)*	*5.8 (1 352)*	
I63	*58.6 (41 188)*	*58.5 (13 628)*	
I64	*19.0 (13 326)*	*18.9 (4 404)*	
Total	**75.1 (70324)**	**24.9 (23297)**	

Notes: $\chi 2$ test for association, $p<0.001$

* grouped for purposes of this table

7.4.1.2. Association between admission at weekends and performance

Table 32 shows the results of the association between weekday/weekend admission and performance in the six measures of quality and safety. There were statistically significant associations in five of the indicators (the exception being readmissions), all of which were consistent with a lower standard of care at weekends. The largest effects were seen in the lower rates of same-day scanning (unadjusted odds ratio [OR] 0.83; 95% Confidence Interval 0.81 – 0.86), thrombolysis (unadjusted OR 0.82; 0.73 – 0.92) and higher rates of in-hospital mortality (adjusted OR 1.18; 1.12 – 1.24) for weekend strokes.

Table 32. Association between weekday/weekend admission and indicators of quality and safety of care

	Weekday admission		**Weekend admission**			
Indicator	Unadjusted % (number)	Adjusted rate, %	Unadjusted % (number)	Adjusted rate, %	p-value	OR (95% CI) (weekday as reference)
Same day scanning	47.6 (32521)	47.8	43.1 (9778)	42.6	<0.0001^	0.833 (0.808 – 0.858)^
Thrombolysis	2.69 (1152)	2.70	2.27 (313)	2.23	0.001^	0.819 (0.730 – 0.919)^
Aspiration pneumonia	5.04 (3545)	5.08	5.70 (1329)	5.57	0.003	1.105 (1.035 – 1.180)
7-day in-hospital mortality	8.89 (6250)	8.98	10.61 (2472)	10.30	<0.0001	1.178 (1.120 – 1.240)
Discharge to usual place of residence within 56 days	73.0 (40692)	72.9	71.3 (12746)	71.3	<0.0001	0.918 (0.883 – 0.954)
30-day emergency readmissions within	11.1 (6165)	11.1	11.0 (1971)	11.0	0.801	0.993 (0.940 – 1.049)

Notes:

^ For Unadjusted rates. The p-value and OR for the other measures are for adjusted rates

CI denotes confidence interval.

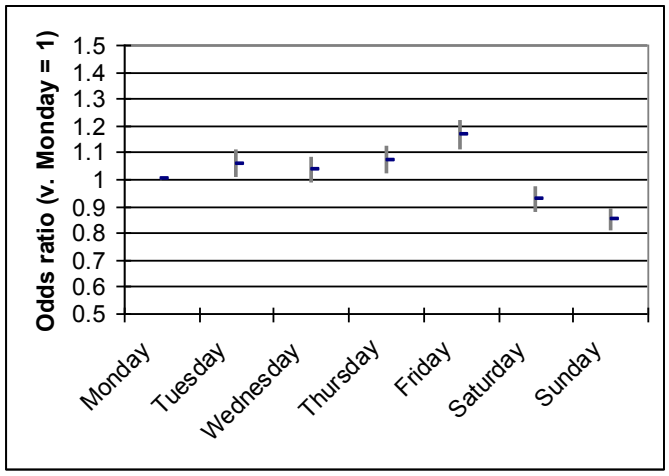
7.4.1.3. Association between day of admission and performance

Figure 16 shows the odds ratios by day of admission, with Monday as a reference, for the six indicators. For the measure of same-day scanning, the highest performing day was Friday with 49.8% of patients accessing a scan (unadjusted OR 1.16; 1.11 – 1.22 with Monday as the reference) whereas Sunday had the lowest rate at 42.0% (unadjusted OR 0.80; 0.88 – 0.97). For thrombolysis, the best performing day was Wednesday when 2.9% of stroke patients received the treatment (unadjusted OR 1.16; 0.98 – 1.37) with Sunday the worst at 2.2% (unadjusted OR 0.82; 0.68 – 0.99), however this latter result is no longer statistically significant when adjusting for case mix (adjusted OR 0.84; 0.70 – 1.02).

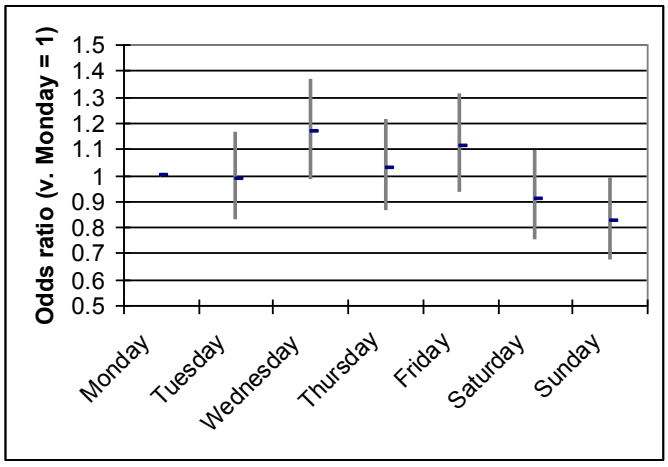
Rates of aspiration pneumonia were higher on Saturdays and Sundays (both 5.7%) than the Monday reference (5.0%); however this was not significant at the 5% level with p-values of 0.12 and 0.05 respectively. The highest rate of 7-day in-hospital mortality was for patients admitted for stroke on Sundays (adjusted OR 1.26; 1.16 – 1.37) at 11.0% compared with an average for weekday admissions of 8.9%. Both Saturday (adjusted OR 0.93; 0.87 – 0.99) and Sunday (adjusted OR 0.93; 0.87 – 0.99) had lower rates of discharge to usual place of residence within 56 days at 71.4% and 71.2% respectively, with an average for weekday admission of 73.0%. The association between day of admission and emergency readmission rates were not significant (p-value 0.60).

Figure 16. Performance by day of admission

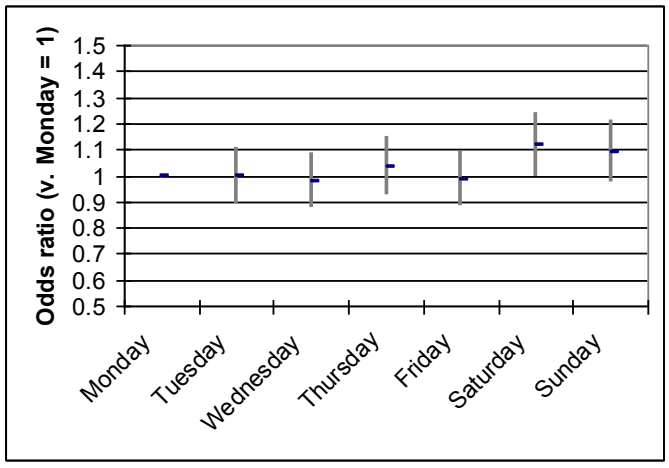
a. Immediate scan (unadjusted)



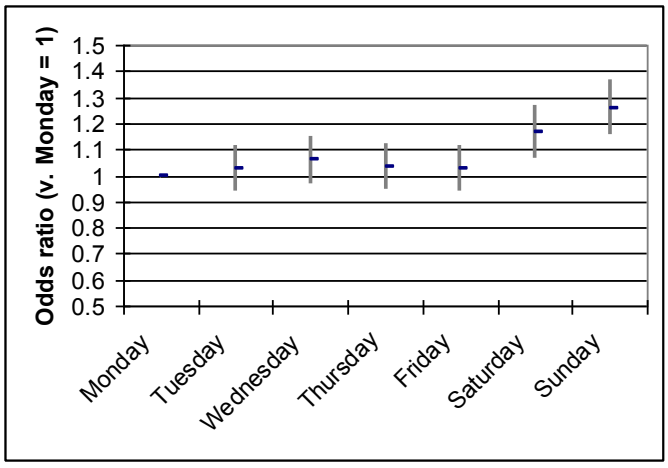
b. Thrombolysis (unadjusted)



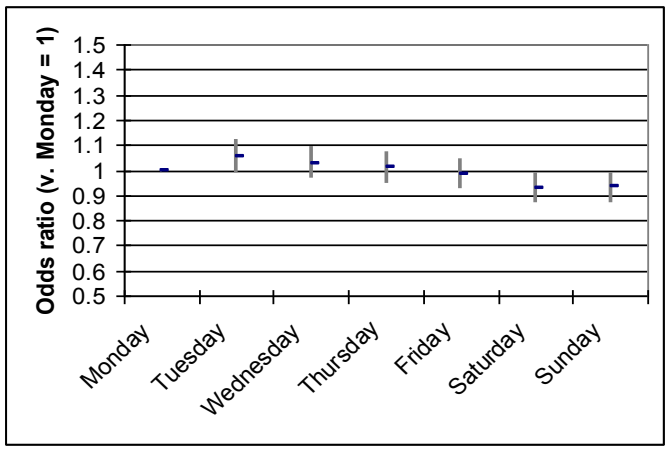
c. Pneumonia (adjusted)



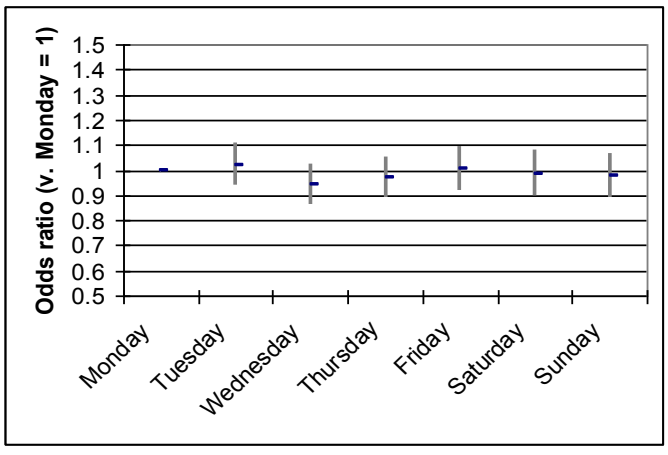
d. Death within 7 days (adjusted)



e. Discharge to usual place (adjusted)



f. Readmission (adjusted)



7.4.1.4. Age stratified results

There was also some evidence that the weekend effect might be more pronounced for certain patient groups, with a more prominent effect on 7-day in-hospital mortality for patients aged 0 to 44 (adjusted OR 1.61; 1.23 – 2.12) than for patients aged 85 and over (1.13; 1.03 – 1.23) (Table 33).

Table 33. The variation in 'weekend effect' across different age groups, for 7-day in-hospital mortality

Age group	Weekday *unadjusted rate, % (admissions)*	Weekend	Odds ratio v. weekday rate, adjusted (95% CI)
0 to 44	5.40 (3 221)	8.82 (986)	1.612 (1.225 – 2.123)
45 to 54	6.17 (4 523)	9.26 (1 414)	1.463 (1.170 – 1.829)
55 to 64	6.47 (8 344)	7.36 (2 527)	1.097 (0.918 – 1.311)
65 to 74	7.25 (14 075)	8.30 (4 565)	1.111 (0.978 – 1.261)
75 to 84	9.25 (22 809)	10.89 (7 660)	1.209 (1.108 – 1.320)
Over 85	12.26 (17 352)	13.91 (6 145)	1.128 (1.033 – 1.232)

Notes: CI denotes confidence interval.

7.4.1.5. Association between consultant specialty and performance

The most common specialty of consultants responsible for the first episode of care in hospital following a patient's stroke was 'general medicine' (48.1%) followed by 'geriatric medicine' (26.3%). **Table 34** sets out the difference in specialty of responsible consultants between weekday and weekend strokes, with the latter slightly more likely to be under the responsibility of a general medic or neurosurgeon, and less likely under geriatric medics. Aggregating the specialties into two groups showed that weekend strokes are less likely to be initially under the responsibility of neuro- or geriatric specialists (33.0% with general medics, A&E medics and other specialties accounting to 67.0%) compared to weekday admissions (33.6%), although this was not statistically significant at the 95% confidence level (p-value 0.06).

Table 34. The specialty of the first responsible consultant after onset of stroke

Specialty	Weekday, % (admissions)	Weekend, % (admissions)	Total, % (admissions)
Neurology	2.89 (2 034)	2.70 (629)	2.84 (2 663)
Neurosurgery	4.31 (3 029)	4.42 (1 029)	4.33 (4 058)
Geriatric medicine	26.42 (18 583)	25.85 (6 022)	26.28 (24 605)
Accident and Emergency	3.06 (2 151)	3.36 (782)	3.13 (2 933)
General medicine	47.99 (33 749)	48.41 (11 277)	48.09 (45 026)
Other	15.33 (10 778)	15.27 (3 558)	15.31 (14 336)

7.5. Discussion

7.5.1.1. Summary of findings

We examined more than 90 000 patients admitted for stroke in England across a one-year period. The study demonstrated that patients admitted for stroke at the weekend have an increased risk of dying in hospital within the first 7 days. Moreover, the study also suggests that weekend care was different in terms of both access to treatments – with lower rates of urgent brain scans and thrombolysis – and wider outcome measures. Even after adjusting for various measures of case mix, weekend stroke admissions had higher rates of in-hospital complications (aspiration pneumonia) and a lower proportion of patients being discharged to their usual place of residence within 56 days.

7.5.1.2. Impact: magnitude of weekend effect

I calculated, as described in the methods section, that there are some 350 potentially avoidable in-hospital deaths within 7 days and an additional 650 people could be discharged to their usual place of residence within 56 days if the performance seen at weekdays was replicated at weekends.

There are also reasons to suggest that the inequality of care is more pronounced than identified in this study. If the effect is indeed caused by a staff deficiency and a lack of resources, then you would expect poorer quality and safety at all out-of-hours periods during the week, including bank holidays and weekday evenings and nights. A Dutch study, for instance, found stroke mortality rates were higher, in comparison with the Monday day shift, during Sunday and Monday evening shifts and during all night shifts.[354] If this is the case, then the out-of-hours periods during weekdays are

masking some of effect. Our analysis was limited to out-of-hours days, since there is not a time stamp within the data.

7.5.1.3. Amenable: association with resource levels

There are several possible explanations for these findings, including: fewer clinical staff working in hospitals on weekends, with those who do work often having less experience and lower familiarity with the patients;[253,244] and less accessibility of resources,[351] such as radiologists to operate the scanners. This study provided some evidence to support the theory that one of the contributing factors to the 'weekend effect' might be a decreased proportion of stroke patients being admitted under the responsibility of stroke specialists at weekends. However, as 'stroke' is not a category of consultant specialty under the coding framework, the study had to define stroke-specialists as geriatricians, neurologists and neurosurgeons, as opposed to general medics, A&E consultants and other specialties. The association between consultant specialty and day of admission also provides some evidence that the 'weekend effect' is amenable to the provision of healthcare. To some extent, this finding is support by previous suggesting that stroke patients cared for by neurologists have better outcomes than those treated by other specialties.[217]

7.5.1.1. Amenable: experience of addressing the weekend effect

More work is required to identify the cause of this 'weekend effect'. However, looking at international experience provides evidence for optimism. Stroke care in Sweden in 1990-95 had similar levels of disparity to the current situation England with adjusted weekend/weekday odds ratios of 7-day in-hospital mortality of 1.14 (95% Confidence Interval 1.09 – 1.20) yet they managed to decrease the variation to 1.07 (1.01 – 1.14).[358] The study also identified increases over this period in the proportion of patients discharged to their usual place of residence within 56 days.[358]

7.5.1.1. Amenable: reorganising stroke services

Furthermore, solutions have been proposed to address the 'weekend effect'. For instance, maintaining a more consistent level of activity – even if this requires hospital staff to be paid at higher levels during weekends – is sometimes economical.[351] Previous studies have also demonstrated that it is possible to provide equal access to thrombolysis at weekends and that the use of comprehensive stroke centres and participation in stroke clinical improvement programmes can ameliorate the problem.[356,359,360] In particular, regional reconfigurations represent a promising strategy for providing consistent levels of access to stroke care.

7.5.1.2. Precision

The indicators identified a wide range of performance, by day of admission, with statistically significant differences in five of the six indicators.

7.5.1.3. Minimum bias

HES data do not include information on the time of admission and therefore I was not able to investigate the wider issue of the quality of out-of-hours care in this study. Other limitations should also be noted. Firstly, the administrative database used gives only limited information on the severity of stroke, which has been suggested to be conceivably higher at weekends.[362,363] Whilst I used a number of variables, such as Charlson index and number of previous admissions, to mitigate for any bias in case mix between admissions on different days of the week, there is no information on, for example, patients' level of consciousness. A further limitation is that the administrative data can contain errors; however, you would not expect this to be different in the coding of weekday or weekend admissions.

I stratified one of the key results – in-hospital mortality – by age group to further investigate potential bias. Chances of death were higher for each of the six age groups for weekend admissions, which again provides some assurance that the results are not an artefact of bias. However, as outlined at the start of the results section, there are differing numbers of stroke patients on each day of the week which could suggest differences in average case-mix; if differences in severity are not recorded in the data then there is a risk of bias.

7.5.1.4. Construct validity

Five of the six indicators identified evidence of the weekend effect, and further analyses revealed lower level of specialist staffing at weekends, which provides evidence of construct validity. However, in lieu of having multiple comparable units – as with the cross-sectional analysis in the previous chapter – I have provided some assurance on validity by comparing consistency with previous literature.

The results are consistent with previous findings on the limited availability of stroke services at weekends in England, and that this is indeed reflected in lower levels of access. An audit of hospital care in 2010 highlighted that nursing practice was presumably leaving many patients in bed over a weekend which may be highly undesirable, including increasing the risk of venous thrombosis (blood clots). Furthermore, the same audit found that 43% of sites that provided thrombolysis services, either on or off site, were not able to provide the service 24/7.[343] Similarly, a national study reported that despite improvements in access to urgent brain scanning, rates at weekends and evenings

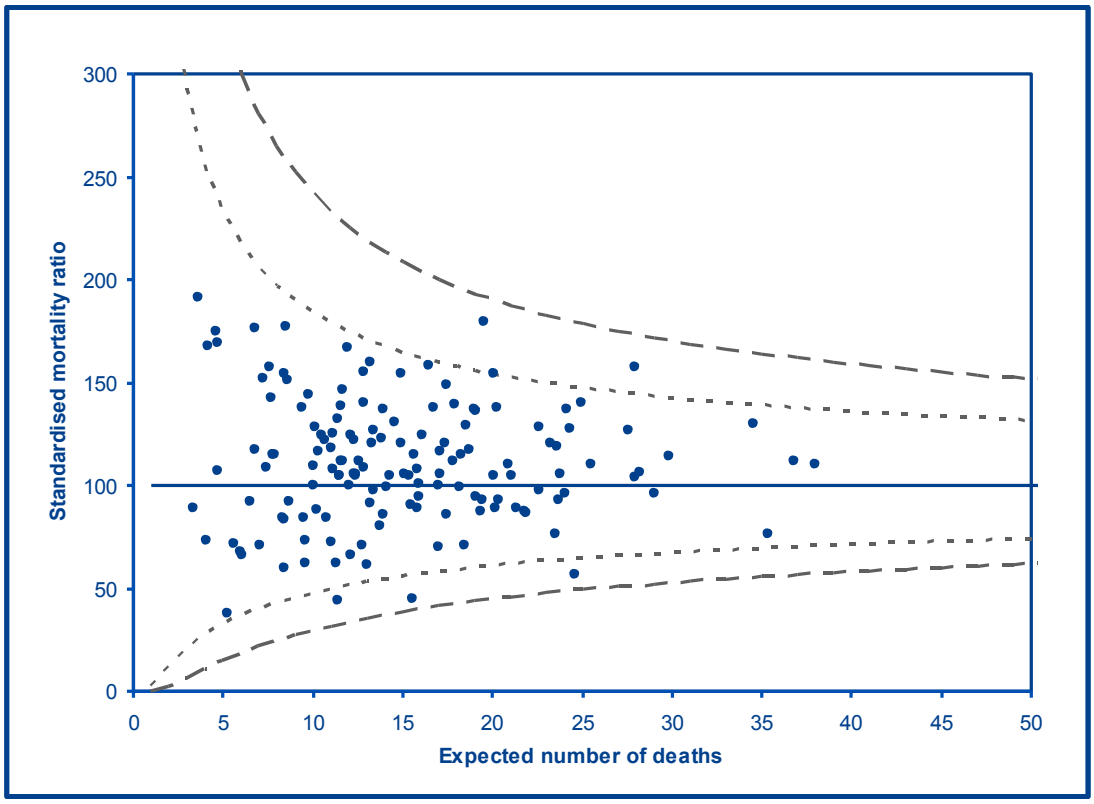
remains significantly more limited.[151] A study by Rudd and colleagues also found that patients were less likely to be admitted directly to a stroke unit (OR 0.77; 95% Confidence Interval 0.69-0.86) at weekends.[364] Further to these findings on urgent stroke treatments, a recent report found that pathways allowing fast track transfer out of hospital covered only 59% of areas at weekends, compared to 75% at other times.[365]

The findings are also consistent with some studies from other countries, including the surprising difference in the number of stroke admissions falling from Monday to Sunday, which in fact tallies with the results from a population based register of stroke incidence in Finland.[366] A Swedish report also highlighted a lower proportion of patients admitted at weekends being discharged to their usual place of residence within 56 days (OR 0.96; 0.93 - 0.99).[358] Similarly the rate of ischaemic stroke patients in a US study receiving screening for dysphagia (which could prevent aspiration pneumonia) falling from 6.1% on weekdays to 5.5% at weekends.[356] The 'weekend effect' for 7-day in-hospital mortality (adjusted OR 1.18; 1.12 – 1.24) is also not dissimilar to studies from other countries, such as Sweden (1.07; 1.01 to 1.14), USA (1.14; 1.05 – 1.25), and Canada (1.14; 1.02 – 1.26).[244,355,358]

7.5.1. Conclusion

Further work is needed to understand what organisational factors might influence the 'weekend effect' and to investigate centres that have reduced the disparities in access and outcome in out-of-hours care. A starting point for this is to allow hospitals to compare the extent of the 'weekend effect' in their organisation to that in their peers, as exemplified in Figure 17.

Figure 17. Hospital-level standardised mortality ratio, for deaths within 7 days of admission on a weekend in 2009/10



Note: Each dot represents a hospital. The horizontal line refers to national average; short-gauge dotted line refers to p<0.025 significance level; long-gauge dotted line refers to p<0.001 significance level.

There is also scope to extend this analysis to other specialties, with similar results having also been found in a limited number of other clinical areas, such as pulmonary embolism, hip fractures and upper gastro-intestinal haemorrhage.[354] A greater understanding of the issue will also require better data, and the inclusion of an out-of-hours admission flag for hospital administrative data should be considered.

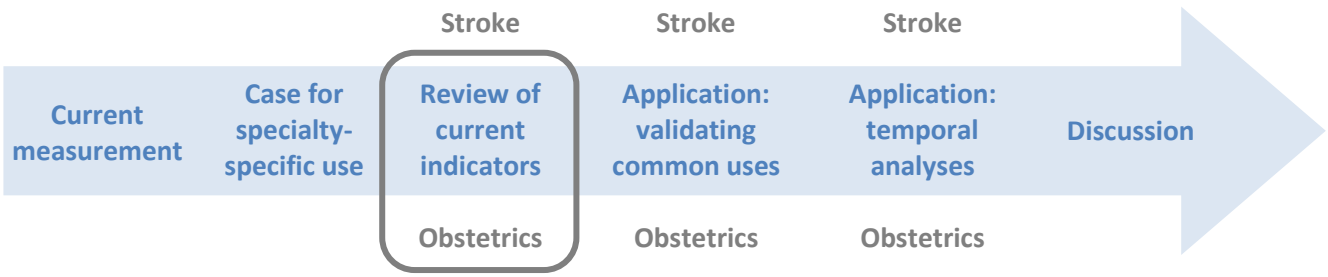
The analysis by day of week addresses most of the coding and risk-adjustment issues raised in the previous chapters and, therefore, permits more robust conclusions in what is, coincidentally, a very high impact area.

Part Three:

Obstetrics care

Chapter 8.

Obstetrics literature review



Overview

Context

The previous three chapters have shown that there is the potential to improve the application of stroke care indicators and, in doing so, to make meaningful inferences about quality and safety. This chapter – the first of three to focus on acute obstetric care – looks at how administrative data have been used to evaluate perinatal care – a very distinct area of hospital care to stroke treatment – and sets out how the indicators have been applied and validated to date.

Methods

A systematic review of original studies that applied, evaluated or validated obstetric-specific indicators based on administrative data. Indicator and study details were collated.

Findings

The initial search identified 1,670 unique citations of which 80 studies had used administrative data to evaluate perinatal care. There has been a steady growth in the use of such indicators; however, few studies have described the strategy used for identifying perinatal episodes from the data and there is a lack of consistency in the use of risk-adjustment of validation techniques. Many countries have used obstetrics-specific indicators although each study has only focused on a very small subset of measures. There is potential to make more meaningful inferences about the quality and safety of acute obstetric care by applying a more comprehensive set of indicators and being more explicit about the assumptions used.

What this chapter adds

The first review on obstetrics care focusing on: the use of administrative data; how these indicators have been used; and, importantly, how they have been applied and validated.

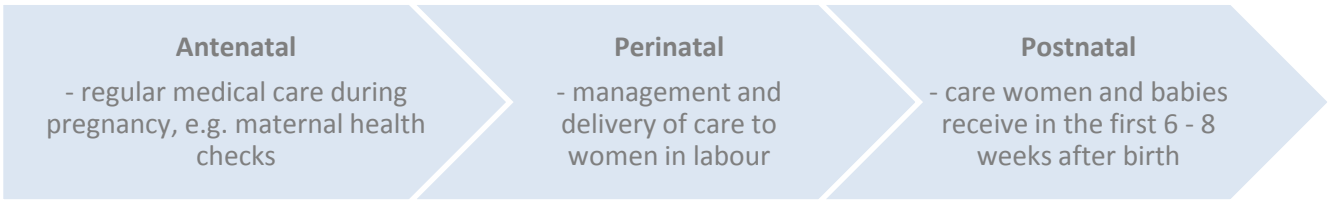
8.1. Background

8.1.1. Background on obstetrics

8.1.1.1. About obstetrics

Obstetrics, from *obstare* the Latin word for "to stand by", is the specialty dealing with care of all women and their children during pregnancy, child birth and postnatal period. The delivery of obstetric care varies considerably from one nation to another; however, Figure 18 outlines key stages. Obstetrics care is delivered by a range of different healthcare professionals from country-to-country and may include hospital physicians and surgeons, community/family doctors, nurses, healthcare assistants or midwives.

Figure 18. Overview of the obstetric care pathway



8.1.2. Importance

8.1.2.1. Impact: global burden

Globally, there has been a general trend in improvement in obstetric outcomes with an average annual decline of maternal mortality (3.1%) and neonatal mortality (1.6%) in the 20 years from 1990.[1] However, annual mortality rates remain at colossal levels, with at least: 3 million stillborn babies, 4 million neonatal deaths and half a million maternal deaths. Moreover, it has been suggested that the majority of these deaths are avoidable. [367] Whilst poor outcomes are more prevalent in developing countries, it is far from limited to these nations. A study from the USA, published in 1992, found that 8% of the mothers or infants had serious complications in pregnancy, with twice this number being transferred for a higher-level of care either during or after delivery.[368]

8.1.2.2. Impact: increasing risk

In these developed countries, in particular, there have been changes in the burden of obstetrics care with increasing numbers and complexity of deliveries. For instance, in England and Wales, overall fertility rates (which account for changes in over population numbers) increased by 21.5% in the decade to 2011 and by around 50% for women aged over 30.[369]

8.1.2.3. Amenable to healthcare

There is good evidence to suggest that obstetrics outcomes can be improved through delivering better healthcare. For instance, one study found that problems of quality of care are the main cause of obstetric readmissions,[370] while another study showed that continuous improvement efforts have shown to be effective for obstetric care in the UK. [64]

8.1.3. Obstetrics measurement

8.1.3.1. Demand for indicators

A Cochrane review on maternal mortality and morbidity, which did not identify any RCTs on critical incident audit in any of the forty years preceding its publication, concluded that there is an essential need for feedback, directed to the relevant person.[371] In England, a national audit report recently concluded that the Department has only limited assurance on the performance of maternity services due to a lack of comprehensive data on key outcomes and activity. The report recommended that a framework for gaining assurance should be developed.[152]

8.1.3.2. Existing measures

The measurement of maternity outcomes is not new. For instance, in the UK, the current Confidential Enquiry into Maternal and Child Health can be traced to 1952, just four years after the inception of the NHS.[372] A number of different methods have been introduced and, in 2006, Simpson and colleagues wrote a review of current methods for measuring perinatal patient safety. While flawed – for instance, there was no indication of how these current methods were identified – the paper does usefully set out the range of measurement techniques used to date, such as: safety attitude and climate surveys, focus groups, storytelling, executive walk rounds, and external review.[373]

Indeed there are many existing quality measures of obstetrics. A recent review of guidance from the RCOG found 290 quality indicators covering 96 clinical categories, with a range of different definitions for each category.[374] This exercise yielded a large number than a rapid literature review of the indicators on processes and outcomes of maternity services in Europe, Australia, New Zealand and the USA which identified 194 different indicators from 30 sources.[375]

8.1.3.3. Using administrative data

In their review, Simpson et al suggest that the use of inexpensive, readily available administrative data to identify shortcomings in safety may provide a useful tool for identifying issues and monitoring trends in perinatal care and, went further by saying, “an on-going quality improvement program based on analysis of data from selected PSIs is essential”. [373] Yet, in the UK, there seems to

be little use of such indicators; of the 50 indicators relating to childbirth[5] identified from the NHS indicator portal,[245] none appear to use routinely collected hospital data as source. Results have been presented on a range using, for instance, radar plots and CUSUM charts to monitor obstetrician's performance. [376-378]

8.2. Objectives

To produce a comprehensive list of indicators of the quality and safety of perinatal care that can be based on administrative hospital data, and provide a critical assessment of how these have been applied previously. This latter objective is designed to inform the quantitative analyses presented in the following two chapters.

8.3. Methods

8.3.1. Search resources and validation

An outline of the systematic methodology was given in a previous chapter (paragraphs 3.4.2 - 3.4.3, p72). The method used to search for peer-reviewed journal articles and other research is described earlier (paragraph 3.4.4, p74). An initial search of the databases – designed to identify peer-reviewed journal articles – was performed in June and July 2012, with a further search in February 2013.

8.3.2. Definitions

8.3.2.1. Definition for administrative data

Recall from paragraph 2.1.1.1 that 'administrative data' are defined, for the purpose of this review and to meet the ambition of the study, using the inclusion criteria: routinely collected data covering more than one hospital; used for administrative purposes (e.g. for reimbursement); and can be replicated using the internationally-recognised ICD9 or ICD10 diagnoses coding framework.

As a principle, where a study was based on administrative data linked to another supplementary source, the individual measures described in the paper were reviewed to assess whether they could be replicated with administrative data alone. This ensured that no potentially beneficial indicator was excluded and that validation studies where results from administrative data are compared to an augmented or different dataset were included.

[5] Searched using the string: neonatal OR birth OR matern*

Studies based on registers of births and/or deaths without linkage to hospital episode data (such as diagnoses and procedures) were excluded.

8.3.2.1. Definition of quality and safety

As discussed earlier (paragraph 1.1.4), providing an exact definition of quality and safety is also difficult and its application for the purpose of this review on obstetric care is again best defined by the treatment of contentious indicators:

- ✓ Length of stay is included. There is an argument that length of stay is a poor measure of quality and safety: one study used length of stay instead as a measure of efficiency [379], another found that it is not be statistically correlated with satisfaction[380] and Oberer and Auckerman found that it is not inversely correlated with readmission rates.[381] However, some variations – such as a dichotomous indicator of long length of stay – are more likely to be considered a proxy for aspects of quality.

- ✓ Emergency caesarean section rates are included, although they could be considered either a process measure (as caesarean section is an intervention initiated by a clinician) or an outcome of the care received earlier in labour (for example, emergency caesarean section following induced labour). [375]

- × Hospital charges is not considered as a quality indicator since this is a compound measure of patient characteristics, length of stay, and treatments provided – which can be picked up as separate indicators and risk adjustments – and is probably more obviously interpreted as a measure of hospital efficiency.

- × Clinically associated complications of childbirth that are not amenable to healthcare – such as congenital malformations – are not included.

- × Elective caesarean section (CS) rates are excluded as this is not related to a decision or intervention during the delivery admission. [382]

8.3.2.2. Limitation to delivery admission

During the research it became evident that clinical care for the antenatal and postnatal aspects of the pathway have few agreed indicators of quality that can be applied to administrative data – due to limitations in the coding frameworks and relative lack of clinical guidelines – and clinical care provided is poorly recorded within English HES data. As a result, the obstetrics chapters focus on the delivery admission (or equivalent for home births) within the perinatal (intra-partum) stage.

Restricting the research to this aspect of the pathway does not affect the ability to meet the objectives of the project and, as well as being pragmatic, is grounded in evidence on:

- the influence of this aspect of the pathway on outcomes – an Italian review of neonatal deaths found that the great majority of deaths (87%) were due to perinatal causes;[383] and
- activity and outcomes in perinatal care are less likely to be influenced by culturally-driven differences in decision making, which are likely to be more significant in ante- and post-natal care such as access to antenatal care.

8.3.3. Search strategy

8.3.3.1. Electronic databases

The literature review was primarily based on searching the following databases:

- Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) 1946 to Present
- Ovid Embase Classic+Embase 1947 to 2013 February 15
- Web of Science (to present)

8.3.4. Search terms

The search strategy was informed by previous literature reviews relating to obstetric indicators. [384-386] The first stage was to identify text words, synonyms and index terms for the studies on obstetrics care that applied, evaluated or validated quality and safety indicators based on routinely-collected administrative data. The synonyms were identified using the existing literature and MESH entry terms, a process suggested by the Cochrane review organisation (Table 35).[387]

Table 35. Search terms

Key text words	Obstetrics	Quality Indicators
Medline index terms (MESH)	Delivery, Obstetric/ Parturition/	Quality Indicators, Health Care/
Embase index terms (EMTREE)	Birth/ Delivery /	Health care quality/

Key text words	Obstetrics	Quality Indicators
Synonyms for the text words (from MESH entry terms)	Parturition* Birth* Childbirth* Obstetric* Deliver*	Quality indicator*
Other synonyms	Labour	Quality: quality; safe*; performance; medical error*; iatrogenic*; nosocomial*; adverse event*; sentinel event*; harm* Indicators: administrative data; routine data; routinely collected data; claims data; international classification of diseases; international statistical classification of diseases; ICD*
Other spelling	Labor	International Classification of Diseases

The databases were interrogated using a two-dimensional search string using the 'AND' conjunction to identify papers relating to obstetric care which mention either a quality indicator or routine data. Medline, Embase and Web of Science were interrogated using the search string outlined in Figure 19 which was customised to the specific requirements of the search engines, searching the title, keywords, and abstract. Embase and Medline were also searched using indexing terms (EMTREE and MESH). For completeness non-English articles were included in the review.

Figure 19. Search strategy

Birth terms:

parturition* OR birth* OR childbirth* OR obstetric*

OR [(deliver* OR labo?r*) **AND** (bab* OR mother* OR matern* OR neonat*)]

Quality indicators:

[(quality OR safe* OR performance OR medical error* OR iatrogenic* OR nosocomial* OR adverse event* OR sentinel event* OR harm*) **AND** (administrative data OR routine data OR routinely collected data OR claims data OR international classification of diseases OR international statistical classification of diseases OR ICD-6* or ICD6* or ICD-7* or ICD7* or ICD-8* or ICD8* or ICD-9* or ICD9* or ICD-10* or ICD10*)]

OR [quality indicator* OR safety indicator*]

8.3.4.1. Duplicates

Duplicates citations were identified using Endnote reference management software by identifying entries with identical authors, date, journal, and title.

8.3.5. Review process

8.3.5.1. First stage: Inclusion and exclusion criteria for review of indicator use

Recall that literature on quality and safety measures were collated if both: (1) based on routinely collected hospital administrative data; and (2) applied, evaluated or validated a potential or actual indicator of the quality or safety of hospital obstetric care. To meet this latter criterion, an included study had to provide numerical results on performance against the indicator.

Specifically, as a first stage, abstracts were reviewed and candidate papers were excluded if:

- None of the described indicators related to an perinatal care;
- None of the described indicators explicitly relate to the quality and safety of health care provided, such as measures of cost. As a result, readmission following birth is included but neither number of birth episodes nor those complications of births not associated with quality are included;

- The study was not based on routinely-collected hospital administrative data (as defined above); and
- The assessment of face validity of indicators which did not involve applying the algorithms to actual data did not meet the inclusion criteria.[388]

Where there was insufficient detail in the abstracts to ascertain whether the criteria above were met, the study was included in the second stage (full text) review.

8.3.5.2. First stage: Inclusion of additional papers describing indicators

I also separately collated studies which described indicators that could be applied to administrative data but no actual database was used (and so failing the first stage criteria to provide numerical results) to ensure a well-defined comprehensive list of indicators. A key example of this is Dr Foster Intelligence's consultation on obstetric measures (blood transfusion post-delivery rate; elective, emergency and total caesarean section rate; instrumental deliveries rate; maternal length of stay of 2 or 4 nights following normal or instrumental delivery) which described the indicators but did not give results on their application.[389]

8.3.5.3. Second stage review criteria

For the second stage, the remaining articles had the same exclusion criteria applied to the full texts. This stage provided a form of quality assurance by double-reviewing many of the studies whilst also giving the opportunity to collate a range of additional information from the papers. As well as the details of the publication (such as author, title, year of publication) a standard template was used to extract additional information (Table 36). Information on validation techniques applied and any key limitations were also recorded where available. Information about the key results was also extracted. For the limited number of foreign language studies, these were translated by skilled colleagues.

Table 36. Information extracted

Study section	Data field
Design	Country
	Details of cohort
	Longitudinal*, including months of data*
Data	Database(s)
	Year(s) of data

	Diagnosis/procedure coding framework
Variables	Quality and safety indicator(s)
	Key covariate(s)
Case ascertainment	Diagnosis field(s)
	Diagnosis codes
Statistical methods	Risk adjustment
	Validation technique(s)

8.3.5.4. Data extraction and review: analytical tools

I catalogued and reviewed the extracted abstracts using Endnote version X6 (EndNote Carlsbad, CA) reference management software. Statistical analyses of study details, such as trends in number of publications over time, were performed using Microsoft Office Excel 2010 (Microsoft Corp., USA).

8.3.5.5. Data extraction and review: Comparative analyses

In discussing the descriptive statistics on the papers identified, I use comparisons to previous literature reviews on patient safety indicators, predominantly that of Tsang et al [66,67], to provide a benchmark to judge whether the specialty-specific nature of the review, and the specialty in question (obstetrics), have revealed atypical results. A comparison to the stroke literature review (Chapter 4) is included in the discussion (Chapter 11).

The information from the studies was compared as a whole and in subsets since certain purposes, for instance, would require specific risk-adjustment models and strategies for identifying obstetric episodes. For example, where the purpose of the study is to validate the identification of safety issues against medical notes, it is unlikely that you would want to adjustment for case-mix where, for comparisons between providers, such a risk-adjustment would be important.

8.4. Results

8.4.1. Identification of studies that applied administrative data

8.4.1.1. Studies on application of obstetrics QIs

The initial search identified 1,670 articles of which 395 duplicates were removed. Of the remaining 1,275 articles screened, the full text was retrieved for 176 studies with the most common reason for exclusion being: non-relevant (n=283); the study describes an indicator using a non-administrative dataset (n=472); and evaluated only number of births (n=236). During the second review stage, a further 33 potential abstracts were added from the search of grey literature and review of the

bibliographies of key articles. Following the review of full texts, 129 citations were excluded resulting in 80 studies (Figure 20). A summary of the key characteristics is included in Figure 21. The details of a subset of papers are given in Table 37 on page 180.

Figure 20. Inclusion and exclusion of articles in the literature review

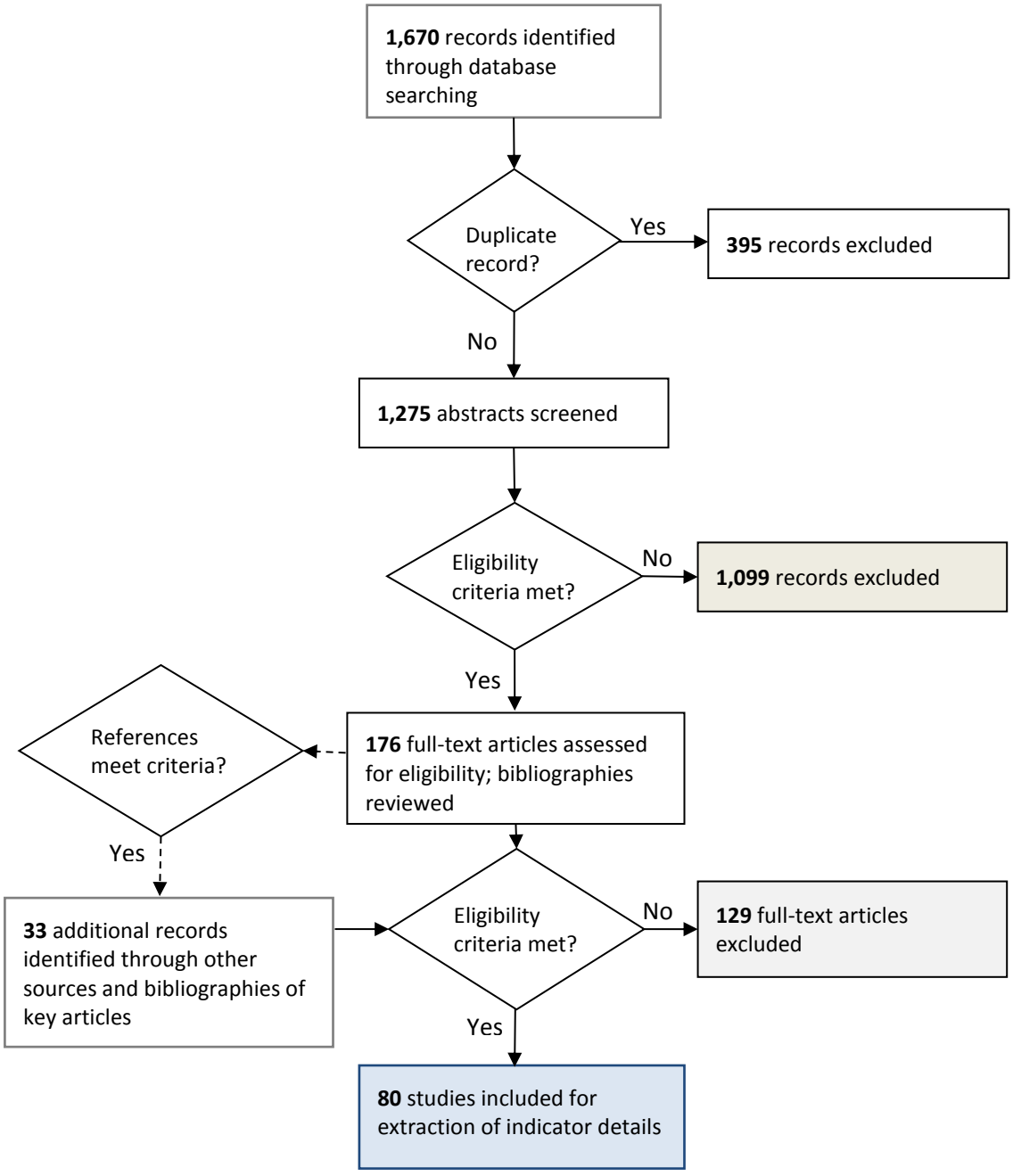


Figure 21. Characteristics of literature

	Number of studies	Proportion of all studies	Example study
Publication year			
Pre-1995	0	-	
1995-1999	1	1.3%	Kotagal (1999)[390]
2000-2004	17	21.3%	Oleske (2000)[391]
2005-2009	31	38.8%	Romano (2005)[392]
Post-2009	30	37.5%	Srinivas (2010)[393]
Not stated / not applicable	1	1.3%	BirthChoiceUK[394]
Country of study			
USA	40	50.0%	Simonson (2007)[395]
Australia	9	11.3%	Baghurst (2012)[396]
England	8	10.0%	Bragg (2010)[397]
France	4	5.0%	Bahrami (2008)[398]
Taiwan	3	3.8%	Tseng (2010)[399]
Other nations with n<3 studies	15	18.8%	e.g. Canada in Thompson (2003)[400]
Multi-national	1	1.3%	Drosler (2009)[401]
Trend analysis			
Longitudinal	19	23.8%	Linton (2004)[402]
Non-longitudinal	60	75.0%	Fantani (2006)[403]
Detail not available/not applicable	1	1.3%	BirthChoiceUK[394]
Months of data, mean (IQR)	40.7 (12-68)	-	
Diagnosis coding			
ICD-9	40	50.0%	Romano (2005)[392]
ICD-10	16	20.0%	Bottle (2009)[404]
ICD-9 and -10	2	2.5%	Lutomski (2012)[405]
Not described	18	22.5%	Keino (2012)[406]
Number of indicators			
1	38	47.5	Baghurst (2012)[396]
>1	42	52.5	Reilly (2004)[407]
TOTAL	**80**	**-**	

8.4.2. Details from studies applying administrative data

8.4.2.1. Country

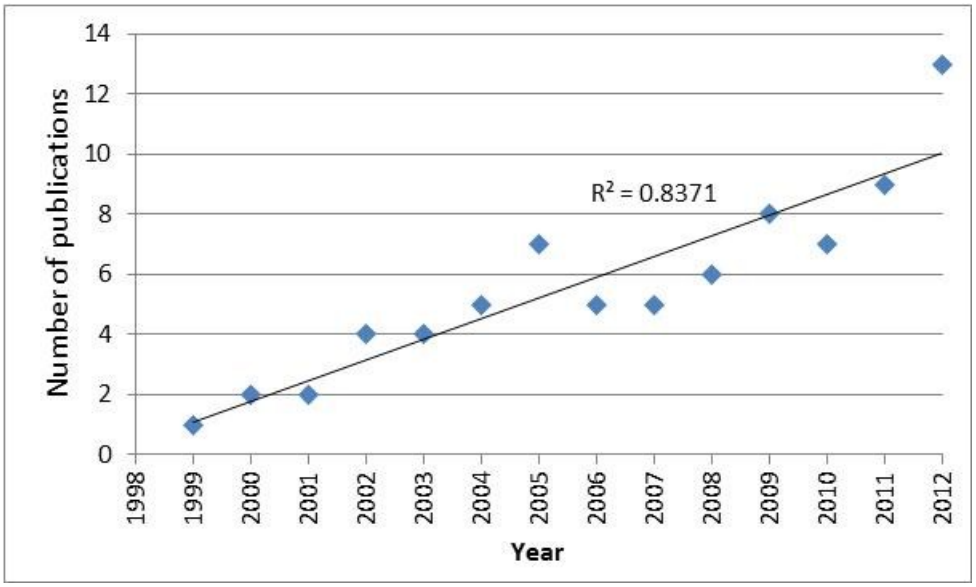
Half of studies (n=40; 50.0%) were from the USA, with only four other nations (Australia, England, France and Taiwan) identified as having 3 or more studies included. There was one international study, which covered seven countries,[401] although a further study compared the results from the authors' country (England) to that of previously published results the USA.[408] The USA predominantly

use ICD-9 diagnosis coding frameworks and, as a result, the majority of studies included used this coding framework, within only 16 explicitly using the ICD-10 framework.

8.4.2.2. Year of study

I did not identify any studies applying administrative data meeting the criteria that were published from 1998 or earlier, even though the databases were searched from their inception to present. There has been a strong approximately linear trend (Figure 22, $\beta = 0.7$) in annual publications, rising from one in 1999 to thirteen in 2012.

Figure 22. Trend in annual publication numbers



8.4.2.3. Study types

The majority of studies were cross-sectional in design, with a limited number of other designs, such as: case-control studies (n = 4) to, for instance, understand the burden of safety incidents;[407-410] and quasi-experimental studies (n=4) designed to evaluate the effect of the implementation of a policy.[411-414] One of these studies showed that the reporting of performance data may have a positive impact on quality improvement.[414] The papers also included studies that could be described as having cohort designs, with hospital type[415] or day of admission as the exposure.[416]

8.4.2.4. Purpose

The studies covered a wide range of uses. Some compared performance between different delivery units, whether: physician;[417] hospitals;[394,397,403,404,418-422] regions;[397,423,424] and countries.[401,408] Other studies, however, used sub-group analysis to evaluate differences in performance, with key variables of:

- hospital characteristics – including location, size, teaching status and ownership[83,406,415,425-429] – and including performance against other measures, such as blood supply at the facility,[430] level of obstetrical specialist support,[431] anaesthetist staffing level,[395] and caesarean section rate[432];
- patient characteristics[83,396,399,400,402,433-437] including type of insurance[391];
- delivery risks, such as gestational age[438], type of delivery [405,439], and length of stay[390,398];
- temporal aspects, such as year-on-year changes[411,440-443], evaluation of the "July" phenomena,[444] and day of admission[445];

Nine studies looked at the overall incident rate [81,370,446-449] and their burden [407-409], with sixteen focusing on validation.[392,393,410,439,449-460] Four studies evaluated change in performance from pre- to post- implementation of a new policy,[412-414] including the impact of IT.[461]

8.4.3. Details from datasets

8.4.3.1. Data source

The studies reviewed used a range of different administrative databases, with the most frequent single dataset (n= 9) being the Nationwide Inpatient Sample in USA, which includes all patients from a stratified random sample of non-federal hospitals in 19-28 states. The State Inpatient Databases (SID), which are also from the AHRQ-sponsored Health Cost and Utilization Project (HCUP), were also used widely.

Table 37. Example papers

Article author (year)	Country, data year	Study design	Sample	Measures	Results	Covariates	Key subject	Related articles
Zhan et al (2003)	USA, 2000	Case-control	720,021 (neonates)	Birth trauma Obstetric trauma (VBwI) Obstetric trauma (VBwoI) Obstetric trauma (CS)	0.7% 22.4% 8.6% 0.7%	Age, sex, case-mix (DRG), SES, insurance, comorbidities, hospital characteristics	Burden of incidence	AHRQ PSIs
Cameron et al (2007)	Australia, 1993-2002	Cross-sectional	775,073	PPH PPH readmissions	5.6% 1.1%	Age, mode of delivery	Year	
Chantry et al (2011)	France, 2006-07	Cross-sectional	30,614	Severe Maternal Morbid Events	1.5%	Age, parity, delivery type, delivery risk (stratified as high- and low-risk)	Prevalence study	
Gould et al (2003)	USA (1995-97)	Cohort	1,615,041	CS rates Neonatal deaths	20.7% 0.3%	Birth-weight, case-mix	Day of admission	
Meara et al (2004)	USA (1991-98)	Quasi-experimental	155,352 (full-term, VB, short-stay neonates)	Include: Readmissions within 10 days Emergency department visits within 21 days	 8.0% 10.4%	Age, marital status, parity, ethnicity, education, Gestational age, birth-weight, location Day of admission, Prenatal care visits	Policy evaluation	
Bottle et al (2009)	England (2005-06)	Cross-sectional	15,298	Obstetric trauma (VBwI) Obstetric trauma (VBwoI) Obstetric trauma (CS)	60.5% 27.9% 2.9%	Age	Hospital comparison	AHRQ PSIs Vincent (2008)[462]
Lutomski et al (2011)	Ireland	Cross-sectional	649,019	PPH Blood transfusion rate	2.6% 1.3%	Age, marital status, medical card status, SES, comorbidities, delivery complications, mode of delivery	Delivery type; Year	

8.4.3.2. Case ascertainment

Only a minority of studies (26.1%) described the strategy for identifying birth/delivery episodes from the administrative data. Of these, a range of methods were used, such as: specific ICD diagnosis codes;[405,459] diagnosis-related group (DRG);[455] procedure codes;[397,411] or a mixture of these strategies.[454 414,424,446] Across just those studies which used the less commonly applied ICD-10 data, the following identifiers were used: procedure codes (OPCS) R17-R25; DRG and diagnosis codes; OPCS R17 (elective CS only); and presence of ICD-10 Z37 (outcome of delivery) flag.

8.4.3.1. Extract validation

A small number of studies evaluated or explicitly discussed coding accuracy, including: purposefully identifying coding errors;[434] accounting for variations in how comprehensively the data is entered (coding depth);[401,455] and making more general descriptive discussion on changes in coding practice. [442] One study explicitly evaluated the influence of the strategy for identifying birth episodes within the administrative data identification strategy.[454]

8.4.4. Risk models

8.4.4.1. Case mix adjustment

When evaluating whether an algorithm based on administrative data is accurate in identifying actual events one might expect that non-risk-adjusted rates would be used but, even when excluding these validation studies, there were still many which did not use risk adjustment.[394,401,407,411,420,422,430,434,442,446]

Where risk-adjustment was used, the maternal factors that were included as covariates ranged from a single factor (age)[404] to applying numerous maternal factors (such as age, ethnicity, proxy for socio-economic status, parity, comorbidities and prior CS[397]). Other studies applied infant risk factors (e.g. birth-weight[390,391,396,403,405,417,418,426,431,438,439,445,460] and congenital malformation[393,403]), delivery risks (e.g. mode of delivery[390,391,405,433,436,440,443,461] and complications during delivery [395,403,425,436,454,461]) and other proxies for complexity (e.g. day of admission[433] or hospital characteristics[391,395,399,401,409,425,436,441,443,444,450,454,461]).

Even for the same indicators, there was little consistency in the use of risk-adjustment models. For instance:

- CS rates – ranged from no adjustment [422] to age, citizenship, insurance, comorbidities, pregnancy risks, birth-weight, congenital malformations, delivery complications [403]

- Obstetric trauma rates – ranged from no adjustment[401] or just age [404] to a combination of age, case-mix (DRG), proxy for socio-economic status, insurer, and comorbidities [409]

8.4.4.2. Participant exclusions

Again there was little consensus in the exclusion of patients from the denominator, even when the same potential bias was being mitigated against. For instance, in trying to exclude high-risk deliveries, some studies used gestational age as a cut off but, even in these instances, they ranged in the cut off mark (from 24 weeks to 37 weeks).[411,415,438] Other approaches included setting a minimum birth-weight[420], excluding deliveries after transfer[429], or excluding non-cephalic deliveries[375]. Some studies also excluded deliveries to hospitals with fewer than a certain threshold of annual cases. [425,430] Many studies excluded certain modes of delivery, such as vaginal or caesarean, with some of these limitations reflecting the scope of the study rather than purposeful efforts to reduce bias.

8.4.5. Details from indicators

8.4.5.1. Range of indicators

The literature review also identified studies, that whilst not applying actual data so not meeting the inclusion criteria of the primary review, described an indicators which is proposed for application to administrative data. As set out in the methods, these studies were retained separately to ensure that the resulting list of indicators was comprehensive and well described. Such studies included the papers recommending the use of AHRQ indicators,[40] SimPatIE,[104] and OECD HCQIs.[85]

A large number of studies were based on AHRQ's indicators – consisting either just AHRQ indicators (using either all or a subset of obstetrics measures)[81,83,401,404,407-409,416,437,439,441,451,455] or using explicitly AHRQ indicators alongside other measures [393,436]. A full list of indicators is included at Appendix E, p306 with some key measures included in Table 38.

Table 38. Key obstetrics indicators

Pathway	Outcome measures	Example study
Maternal	Emergency readmissions	Thompson (2003)[400]
	Puerperal sepsis	Lu (2005)[427]
	Obstetric trauma	Downey (2012)[441]
	Episiotomy	Webb (2002)[418]
	Mortality	Magalhaes (2012)[446]

	Post-partum haemorrhage	Goff (2012)[452]
	Severe maternal morbidity (composite)	Callaghan (2012)[442]
Delivery	Unsuccessful vaginal birth after caesarean (VBAC)	Knight (2013)[375]
	Emergency caesarean section	Roberts (2009)[456]
	Failed vacuum or forceps delivery	Knight (2013)[375]
	Late caesarean section	Knight (2013)[375]
Neonate	Emergency readmissions	Meara (2004)[412]
	Infections	Korst (2005)[415]
	Birth trauma – injury to neonate	Downey (2012)[441]
	Adverse outcome indicators	Walker (2010)[453]

8.4.5.2. Indicator validation

The majority of studies (n=42) studies for which full-text was available did not evaluate the validity of the indicators applied. Where used, the most common techniques were:

1. construct validation, which is based on the hypothesis that measures of similar aspects should be statistically, positively associated; and

2. validation of the risk adjustment model (minimum bias).

The construct validation – which also inherently assesses data accuracy – took the form of comparison to either a "gold standard", such as case notes [392,410,452,453,456-459], or another dataset [449,451]. A similar method involved comparing results derived from the same administrative dataset by applying different algorithms for the same adverse outcome (e.g. trauma[439]). Forms of validation of the risk adjustment model were similarly diverse and included simply reporting model fit performance,[393,421] comparison of different model specifications,[396,397,403,417,425,429,433] or evaluating reliability of the model by applying additional data.[450] For one study, the main focus of the research was to develop a valid model for predicting an adverse outcome.[460]

Less common strategies for validation were to assess reliability by evaluating performance for different cohorts (e.g. including or excluding breech presentations in the denominator [411]) and to follow a multi-faceted validation process (e.g. following the procedure set out by AHRQ for their indicators[415]). In one study, the hospital-level results were sent to the organisations for qualitative feedback, with hospital trusts "generally positive".[404]

8.4.5.3. Indicator interpretation

"An example of a commonly used maternity indicator that is difficult to interpret is the overall caesarean section rate. Lower caesarean section rates are often assumed to reflect better care. However, there is also a threshold below which the caesarean section rate is too low and babies may be harmed. One problem is that there are no established guidelines for determining this threshold. A second problem is that as elective caesarean sections become increasingly popular," (p1) [375] which may increase rates – irrespective of the intervention of the health services – where there is a policy of maternal choice.

8.5. Discussion

8.5.1. Principle findings

8.5.1.1. Descriptive data on studies

I identified 80 studies that described the application, evaluation or validation of indicators of obstetrics care using routinely collected hospital data. While there were no studies meeting the inclusion criteria published prior to 1999, there has been an approximately linear increase since then. The majority of studies were based on data from the USA and only a third (n=21, from 61 where diagnosis framework stated) used the ICD-10 coding framework. Approximately a quarter of studies used the longitudinal nature of the administrative data to make inferences about quality and safety of care. Around half of studies (38 of 80) used only one indicator. The studies took many forms, from descriptive cross-sectional studies of differences in performance for various hospital and patient characteristics to quasi-experimental designs to evaluate the effect of the implementation of an initiative to improve care.

Some studies have used composite indicators, such as the Maternity Unit Performance Index that include both mother and child outcomes as used by Anastasakis.[463] There are some potential advantages in bringing together different facets of quality into a concise quantified measure; however, within the literature I found little explanation over whether these indicators are instead heavily influenced by the sub-measure with the largest range in performance.

8.5.1.2. Case ascertainment strategies for births

The review of studies identified a lack of consistency – both in terms of which fields, and codes within the chosen fields, are used – in the identification of birth episodes within the administrative data. These studies are also not necessarily consistent with other purposes with, for instance:

- a study led by York University on obstetrics efficiency - using length of stay as a proxy – that used a combination (ICD10: any code beginning "Z37"; OPCS4.5: any code beginning "R17", "R18", "R19", "R20", "R21", "R22", "R23", "R24", "R25");[379] and
- a recent publication on coding accuracy used admission classification: admimeth 82 (other: babies born in healthcare provider) and 83 (other: babies born outside the healthcare providers, except when born at home as intended).[464]

This could represent an issue if the episodes that are not consistent to two different strategies are not wholly representative of the shared cases within the cohort. Taking the latter example,[464] which used admission method as a flag, the strategy excludes 'other birth events' which captures NHS-funded home births and all other birth events which are not NHS funded, either directly or under an NHS service agreement. Previous research[465] has shown that results for patient characteristics and outcomes are not necessarily the same for home births and, therefore, differing treatment (in terms of inclusion/exclusion in study population) of this group would potentially introduce bias. Some excluded transfers as proxy for high-risk; however increases in midwifery-led care, which have high transfer rates, may invalidate this assumption.[429]

8.5.2. Indicator validity

8.5.2.1. Risk-adjustment

Achieving good outcomes for the increasingly complex caseload is problematic and sometimes will not be amenable to the quality of healthcare provided. For instance, in the UK between 1985 and 1993, 55% of maternal deaths occurred despite being judged to have received high quality medical care.[466] Given this evolving complexity in obstetrics care, adjusting performance measures for case-mix is increasingly important. Yet the literature review revealed a lack of transparency and consistency. Disaggregating the results further highlights this problem. In the 2007 indicator set described by the SimPatIE project, there were four obstetrics indicators, with no stratification of results, and yet different depth of risk adjustment, from none to accounting for age and comorbidities (Table 39). No rationale for the differences in sophistication between the different risk models was presented.

Table 39. Risk adjustment in the SimPatIE obstetrics measures

Indicator	Risk adjustment
Obstetric trauma - Vaginal Delivery without instrument	Age and comorbidities
Obstetric trauma - Vaginal Delivery with instrument	Age
Obstetric trauma – Caesarean section	None
Problems with Child Birth	None

Source: SimPatIE. Catalogue of Patient Safety Indicators. Work Package 4. March 2007.

Other studies have mentioned the importance and current issues with risk models but then fail to address these. For instance, Janakiraman and Ecker[467] mention that risk adjustment of birth trauma is not well established; however, the authors do not take the next logical step to recommend that this should be empirically tested – as is possible using the administrative data. Indeed such analysis has been done elsewhere with one study – although using a neonatal specific data source – evaluating the risk adjustment explicitly using a systematic review of literature to identify potential risk factors.[468] Whilst having a good predictive risk model is important for the purpose of evaluating quality and safety, it is also has wider importance such as to ensure that activity-based payment schemes are fair. [469]

8.5.2.2. Indicator validation

Whilst some indicators have been accepted for over a decade – such as the AHRQ PSIs, including obstetrics measures, that were accepted in 2002[82] – the validation is embryonic. As discussed earlier there is also a lack of consistency in the techniques and comprehensiveness in the validation of the obstetrics indicators presented here. The European-wide SimPatIE project only applied an assessment of face validity, involving reaching a consensus amongst a small group of clinicians, before recommending their implementation.[388] Despite the relative ease with which it can be done using administrative data, few studies evaluated construct validity (by comparing results from similar measures). Indeed, one study evaluated the AHRQ PSIs by comparing performance (using Spearman's rank) across different indicators but did not include the obstetrics measures. [470]

There are also examples of recommendations being made with only limited knowledge of the validity of the indicators. In recommending indicators for future use, the SimPatIE project repeated a conclusion that "the AHRQ PSIs are a broad screen for potential safety events that point to needed improvement in the quality of care for specific population"; however, this was based on a study that looked only at patient characteristics rather than, say, hospital or temporal covariates.[104]

While this reliance on using clinical consensus over definitions goes some way to meeting two of the three aspects in Freidman's model on what makes a good indicator – proxy power (meaningful to health) and communication power (transparent) – it fails to address data power (ensuring high quality and available data). [471] Moreover, as discussed earlier, in reality the application of the data often lacks the required transparency to be judged as having reasonable communication power.

8.5.2.3. Performance

The heterogeneity of the studies – for instance in their use of risk-adjustment and strategy to identifying birth episodes – means that formal meta-analyses of results were not possible. As well as synthesising results and comparing against peers it is also possible to compare indicators performance to a predetermined benchmark and, whilst not an a priori aim of this part of the study, reviewing the literature revealed that some acceptable standards for these indicators have been stated (Table 40). This adds to the evidence that perinatal care is amenable (as certain performance can/should be achieved).

Table 40. Selected benchmarks for performance

Measure	Acceptable range	Cohort	Organisation
Caesarean section rates	5 – 15%	All births in the population	UNICEF 1997[472]
Post-Partum Haemorrhage	≤4%	Numerator: post-partum haemorrhage ≥1000ml within 2 hours post-partum	Danish National Indicator Project [473]
Maternal mortality	≤1%	Proportion of women with obstetric complications admitted to a facility who die	UNICEF 1997[472]
Third or fourth degree lacerations	≤6%	All first time vaginal deliveries	Danish National Indicator Project [473]

8.5.3. Strengths and limitations

8.5.3.1. Strengths

Previous studies – including one in 2013 – had brought together clinically recommended indicators from medical associations, such as the Nordic obstetric and gynaecological association (NOGA) and

American College for Obstetricians and Gynaecologists (ACOG), RCOG (UK), Royal Australian & New Zealand College Obstetrics and Gynaecology (RANZCOG), Society for Obstetricians and Gynaecologists of Canada (SOGC).[374,474,475] However, this current review represents the most comprehensive assessment of how indicators have been applied using administrative data and adds important transparency to the issues over their applications, such as on risk-adjustment models. The comprehensive nature from this study stems from its scope and also the process followed such as the inclusion of extracts from all languages.

As with the stroke review, the specialty-specific literature review process was found to be a sensitive strategy for identifying relevant research and, as such, provides more assurance that the resultant citations are comprehensive. For instance, a literature review of all patient safety measure based on routinely collected hospital data, which followed a similar process, identified a similar magnitude of studies (n = 124) despite its far broader scope.[67]

8.5.3.2. Limitations of the review process

No formal meta-analysis could be undertaken because of the heterogeneity in study methodologies and underlying data frameworks. I did not use a template to assess the quality of the papers since there was a very diverse range of purposes across the studies and, as already discussed, validity needs to be proportionate to the purpose. Given the strong evidence on the existence of publication bias,[316] studies using indicators which did not identify significant results when assessing the quality and safety of perinatal care may have been missed by the search strategy in this study. However, the design of the study – to retain articles which described indicator sets, but did not apply those to administrative data – mitigates against this bias to some extent.

The inclusion criteria that the indicators should measure performance of perinatal care excluded studies that looked at general obstetric performance, such as measures of obstetric nosocomial infections[457] and obstetric readmissions[370]. However, these indicators were applied to perinatal care elsewhere so are included in the list of potential indicators.

8.5.3.3. Limitations of results

Maternity care differs from other areas of health care in a number of ways – such as by predominantly dealing with healthy people – and, therefore, the findings might not be directly generalisable to other specialties. Also maternity services deal with the culturally and emotionally sensitive area of childbirth and so non-biomedical outcomes, which are not captured in the administrative datasets used in the studies presented here, would also need to be considered to have a full understanding of the overall performance of obstetric care.

The results do not give a comprehensive set of indicators of obstetric care nor results on this. The scope of the review, which was necessitated by a number of factors described above (paragraph 8.3.2.2, p170), also meant that the measures were focused on the perinatal aspect of the obstetrics care pathway and, as such, the indicators described could not give a holistic picture on the overall performance of obstetric services. For instance, neonatal indicators (which in many health services will be the responsibility of a separate specialty: neonatalogy) including nosocomial infections from neonatal units [476] and ambulatory care sensitive conditions [477] which are inherently important aspects of quality and safety are not included. A specific reason for the potential incompleteness of the list of indicators is that some sets of measures are not based on routinely collected data[475,478-482] such as the commonly applied trigger tools based on chart reviews. [483]

8.5.4. Conclusions

8.5.4.1. Contribution to knowledge of indicators

As with the literature review of stroke indicators (Chapter 4) looking at a specialty in detail, had advantages over a broader scope; not least, to better identify existing indicators. The study has highlighted a paucity of use of such indicators and a lack of transparency and consensus on key assumptions. The importance of the transparency added by this review was reiterated by previous calls for consistent definition and reporting practices for hospital complications across countries. [66]

8.5.4.2. Purpose

A set of quality and safety measures would be valuable in improving obstetrics care. [373] However, such an indicator set should be meaningful, actionable, timely, feasible and presented data to caregivers; [373] which are all plausible requirements within the constraints of routinely-collected hospital data. Recommendations on exactly how these indicators should be used vary. Many applications have been suggested for use by organisations such as AHRQ, JACHO, OECD, Australian Council for Safety and Quality in Health Care (ACSQ) and Canadian Institute for Health Information (CIHI); however, these recommendations have not necessarily been echoed by the academic community who have conducted much of the research to date. For instance, Kristensen and colleagues recommended the use of obstetric trauma-vaginal delivery without instrument (AHRQ, JCAHO), with instrument (AHRQ, CIHI, OECD) and birth trauma-injury to neonate (AHRQ, CIHI, OECD) for use, whereas the same study suggested that obstetric trauma following caesarean delivery (AHRQ, OECD) and problems with childbirth (ACSQ, OECD) should not be used. [388]

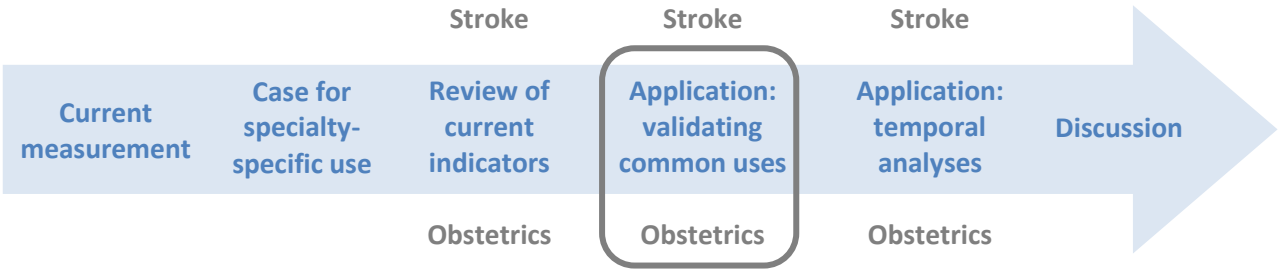
Whilst administrative data have been used in published literature to identify nosocomial infections, there is not universal agreement that this is an appropriately sensitive surveillance tool.[457] Table 41 sets out key recommendations for future obstetrics research based on the findings of the review.

Table 41. Summary of implications of study design

Finding	Implication	Recommendation
1. Obstetrics indicators based on administrative data have been used infrequently outside the USA and with ICD-10 coding framework	Barriers to generalising findings across borders mean that little is known about the use of such indicators.	*Key study designs should be replicated using data from outside USA or Canada to identify any **potential outliers**.*
2. There is a lack of consensus and transparency about the assumptions used in applying the indicators to administrative data	Barrier to generalising, validating or benchmarking results.	*To explicitly set out the **assumptions** used in studies, such as the choice of diagnosis codes and fields.*
3. Around half of studies used only one indicator	Content validity of many existing studies questionable	*Apply **bundle** of indicators*
4. Only two studies evaluated the effect of coding practice or accuracy on performance and these focused solely on coding depth	Concerns remain over the validity of results given the potential influence of coding bias.	*To analyse the potential effect of **coding variation**.*
5. Only one study – USA-based with limited scope – evaluated the effect of day of admission on performance.	Potentially important quality/safety issue is being ignored.	*A non-USA study on **day of admission** should be conducted to identify whether there is a potential issue.*

Chapter 9.

Obstetrics variation and validation



Overview

Context

The previous chapter reviewed the current use of administrative data to evaluate the quality and safety of obstetric care in hospitals. This chapter sets out how a subset of the indicators identified can be applied, using data from hospitals in England. As well as evaluating whether outliers can be identified (precision), and to address the gap in the literature identified in the previous chapter, the following study will also measure the effect of differences in coding and investigate other aspects of validity.

Methods

Eight indicators, spanning process and outcomes measures for both the mother and neonate, were applied to English data in 2011/12. Logistic regression analyses were used to adjust for case-mix, and funnel plots to look at potential outliers. I also investigated the effect of variation in hospitals' coding practice using a range of proxies and multi-level regression models.

Findings

The indicators were sufficiently sensitive to identify variations in performance as recorded in the data, with 441 occurrences of hospital trusts – or 550 outliers at the site-level – performing statistically differently than expected at the 99.8% significant level across the eight indicators. Differences in coding practice appeared to only partially explain the variation in performance. The work showed that administrative data can be used to indicate potential concerns around performance across some of the different aspects of quality and safety of obstetric care.

What this chapter adds

This chapter represents the most comprehensive review of obstetric care in England using administrative data. This study is also the first to evaluate the effect of variation in hospital- and site-level coding of such measures.

Acknowledgements

Professor Susan Bewley (King's College, London) commented on potential indicators.

9.1. Background

9.1.1. Obstetrics care in England

9.1.1.1. Policy development in maternity care

In England, having a baby is the most common reason for admission to hospital,[152] with some 700,000 births currently each year.[484] In total, maternity care costs around £2.6 billion annually which accounts for nearly 3% of the NHS budget.[152] Maternity services are relatively unique in that they are focused on supporting predominantly healthy people through a natural life event.

Maternity care has changed dramatically over time. In the 1950s, around a third of births took place at home; however, following a trend of increased hospitalisation during the 1960s and 1970s, this proportion currently stands at around 2% nationally.[485] There have been improvements in outcomes with, for example, the number of babies dying within 28 days of births falling from 6.7 to 2.9 per 1,000 live births in the 30 years from 1981.[486] However, there is evidence to suggest that outcomes in obstetrics in England can be improved further, including:

- that some outcomes are consistently worse than in the other UK nations and wider international comparisons; and
- the failure to fully address the common causes for litigation claims, such as errors during caesarean section, with the cost to cover against maternity claims totalling for around a fifth of the total cost of maternity services.[152]

9.1.1.2. Existing measures of obstetrics care

The policy for the commissioning and delivery of maternity services was defined by the Department of Health in Maternity Matters (2007),[487] which recommended the development of a range of indicators, that will enable users, trusts and commissioners to compare performance against others. However, as recently as 2013, a publication from the Royal College of Obstetricians and Gynaecologists highlighted that “there has been an astonishing lack of robust information on even simple clinical outcomes on a national basis.” (p.viii)[375]

9.1.2. Revisiting existing literature

9.1.2.1. Previous hospital-level comparisons

To date, there have been few studies[394,397,403,404,418-422] that have used administrative data to compare obstetric performance at a hospital level, with the existing studies focusing on caesarean section or obstetric trauma rates (Figure 23). These studies do cover a range of countries’ health services,

including developed and developing countries. However, there have also been comparative studies focusing on a different delivery unit, whether physicians,[417] regions,[397,423,424] or countries.[401,408] Other studies have also used sub-group analysis to focus on specific hospital characteristics, such as:

- location, size, teaching status and ownership,[83,406,415,425-429] including performance against other measures, such as bloody supply at the facility;[430]
- level of obstetrical specialist support;[431]
- anaesthetist staffing level;[395] and
- caesarean section rate.[432]

A recent report by the Royal College of Obstetricians and Gynaecologists introduced a number of indicators of maternity indicators for English providers. However, the work did not look at neonatal outcomes – rather focusing solely on data from the maternal patient record – and did not empirically test the validity of the measures. [375]

9.1.2.2. Data and indicator validation revisited

As discussed in the previous chapter, there have been a small number of studies investigating coding accuracy. For instance, on identifying coding errors by reviewing medical records and coding depth,[434] with Drosler et al using the mean number of secondary codes (although not presenting the data on the obstetrics measures) as a proxy for coding practice,[401] and Grobman et al using quartile of ICD-9 coding intensity (although this is metric is poorly defined in the paper).[455] Other studies on data validation include a more descriptive discussion on changes in coding practice [442] and an investigation into the strategy for identifying birth episodes within the administrative data identification strategy.[454]

Few studies have explicitly validated the indicator algorithms, with the techniques used being primarily comparison to other measures of performance and validation of the risk adjustment model. The former primarily took the form of criterion validation; namely comparison to a validated measure, such as medical records.[392,410,449,451-453,456-459] Forms of validation of the risk adjustment model included simply reporting model fit performance,[393,421] comparison of different model specifications,[396,397,403,417,425,429,433] or evaluating reliability of the model by applying to additional data.[450] For one study, the main focus of the research was to develop a valid model for predicting an adverse outcome.[460]

Figure 23. Studies comparing hospital performance using obstetric QIs

	Cohort	Country	Measures	Risk-adjustment	Main results
Caesarean section rates					
Turner (2011)[420]	19 hospitals (72,400 births, 2009)	Irish	CS rates (elective and emergency)	None	Overall CS rates ranged from 18.7 – 35.6%
Bragg et al (2010)[397]	146 hospital trusts (620 604 births, 2008)	England	CS rate (elective and emergency)	Age, ethnicity, SES, and clinical risk factors for CS	Adjusted CS rates ranged from 13.6 – 31.9%
Fantini et al (2006)[403]	29 hospitals (62,836 births, 2003-04)	Italy	CS rates (primary)	24 risk factors in the full model; 7 (marital status, maternal age, infant weight, and clinical risk factors) in the parsimonious model.	Correlation between hospital rankings of two models was 0.92
Hsu (2006) [421]	6 hospitals (27,693 births, 1999-2001)	Taiwan	CS rates	Maternal and paternal age and education, gestational age, multiple births, parity, infant gender, clinical risk factors	33.7% rate. Substantial change in rankings for crude vs risk adjusted models.
Gichangi (2001) [422]	39 hospital districts (86,105 births, 1983-97)	Kenya	CS rates	None	Hospital-based rates of 6.3% (range 0.3 – 37.7%)
Obstetric trauma					
Baghurst (2013) [419]	18 hospitals (65,598 births, 2002-08)	Australia	Third and fourth degree lacerations	Maternal age, parity, ethnicity, assistance with instruments and episiotomy, shoulder dystocia, infant birth-weight	8 hospitals with systematically more/fewer tears
Bottle et al (2009)[404]	170 hospital trusts (572,180 births, 2005-06)	England	Third and fourth degree lacerations (for deliveries with instrument; without instruments; and CS)	Age	60.5% (with instrument), 29.7% (without instrument), 2.9% (CS). Part of wider study using AHRQ PSIs.
Webb et al (2002)[418]	18 hospitals (49,692 births, 1994-98)	USA	Episiotomy rates Third and fourth degree lacerations	"Case mix" factors including maternal age, race/ethnicity, education and infant birth-weight	Hospital episiotomy rates ranged from 20 to 73%, and lacerations from 4 to 13%. Rates of two indicators correlated ($p<0.01$)

9.1.2.3. Revisiting limitations in existing literature

The review of literature described in the previous chapter suggested that future research should focus on:

1. applying key study designs for identifying any potential outliers outside the USA or Canada;
2. using a more comprehensive set of indicators;
3. being explicit about the assumptions used in studies, such as the choice of codes and fields for identifying delivery episodes;
4. analysing the potential effect of coding variation; and
5. investigating the association between quality and day/month of delivery outside the USA to identify whether there is a potential issue.

The latter area is covered in the next chapter while the other areas are covered in this section.

9.2. Objectives

In this chapter I investigate the feasibility of using HES data to evaluate the quality and safety of obstetric care at a provider level, focusing on the ability of measures to identify important effects in a robust fashion.

9.3. Methods

9.3.1. Data

There are some noteworthy differences in the data collected for maternity care in comparison to the other areas of healthcare contained in the HES dataset, which are outlined here and discussed in further detail at the end of the chapter. HES contains data for all births in England, including those at home and in non-NHS hospitals, irrespective of whether a consultant or midwifery-led episode. Pregnant women are admitted as general admissions and only once they have given birth does their record become a maternity record.

There are two types of maternity record – (1) a delivery record for the mother, and (2) the birth record*(s)* for the baby *(or babies in the case of multiple deliveries)* – with each containing an additional 19 fields covering, for instance, parity of mother, method of onset of labour, place of delivery, gestational age and birth-weight. The data in these records comes from birth notification

records and requires only a limited dataset to be completed; however, these additional fields are not as complete as the rest of the HES data.[488,489] Despite including both maternal and baby records, the HES dataset do not link the baby's record to that of the mothers, or vice versa.

9.3.2. Selection of indicators

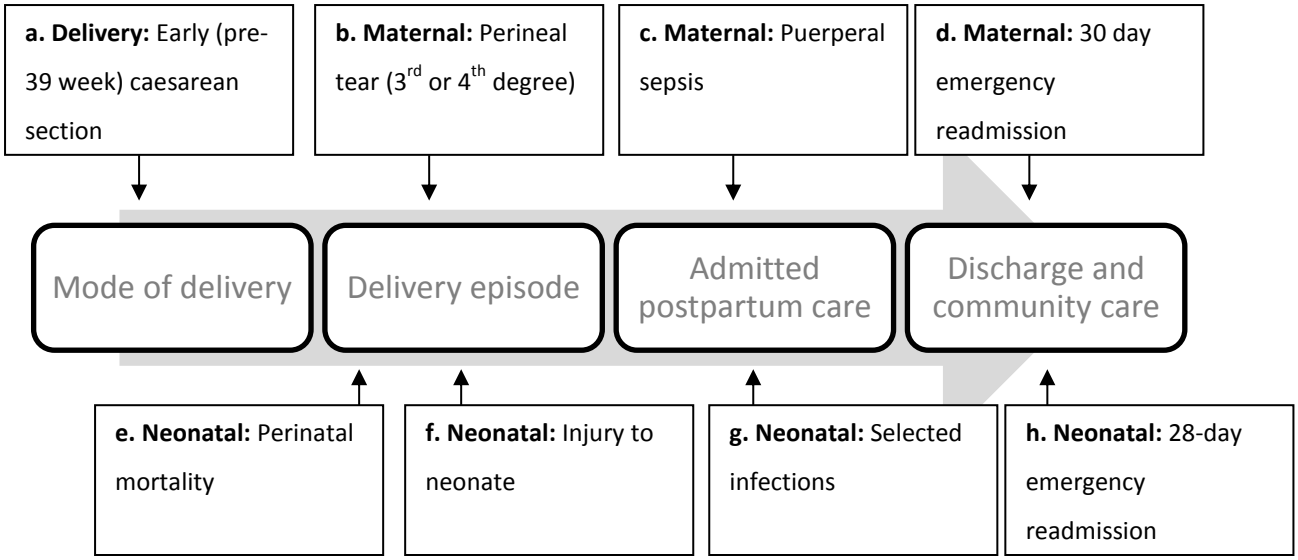
9.3.2.1. Selection criteria revisited

Recall the criteria listed earlier (paragraph 3.5.1.2, p76) for the selection of indicators. For this obstetric element to the project, the experts involved in getting a consensus on the indicators are listed in paragraph 9.3.3.1 below.

9.3.2.2. Selection of indicators

The literature review described in the previous chapter identified 24 maternal, 9 delivery, 11 neonatal categories of indicators with the potential to be applied to administrative data in England. Following the selection criteria, a subset of eight indicators was chosen for application here (Figure 24). The indicators were chosen to cover both process (e.g. early caesarean section rates) and outcome (e.g. infection rates). These two examples also demonstrate that the chosen indicators span care given during the delivery admission, from decision on type of delivery through to postpartum care.

Figure 24. Obstetric indicators across the care pathway



Note: the positioning of the complications is for indicative purposes only

9.3.3. Developing indicator definitions

9.3.3.1. Development process

As with the stroke measures, the selected indicators were then refined using a multi-faceted approach. In the first instance, the indicator definitions were implied from any previous studies that applied these measures:

1. I derived draft definitions using details from algorithms applied in existing literature, identified through the systematic review outlined in the previous chapter.

2. I used an iterative consensual approach to ascertain feedback from a multidisciplinary panel of experts, consisting of a senior obstetrician, a public health doctor, and administrative data experts (from Imperial's Dr Foster Unit).

3. The definitions were also reviewed by a multi-disciplinary team, including people experienced in the management of health services and clinicians, as part of the review of the process of publishing a national report on maternity services.[152]

9.3.3.2. Final definitions

The finalised indicator definitions are included below in Table 42.

Table 42. Details of obstetric indicators used in analysis

	Indicator	Rationale	**Definition** (discharges meeting the inclusion/exclusion criteria for the denominator with...)	**Exclusions**
Delivery				
a	Early (pre-39 week) caesarean section	NICE recommends[1] that planned caesarean section should not routinely be carried out before 39 weeks due to increased risk of respiratory morbidity.	Elective caesarean sections (OPCS code: R17) where gestational age < 39 weeks	Diagnoses possibly justifying early caesarean section (ICD-10 codes: O100-9, O11, O140-9, O16, O244/9, O266/8, O356, O360/1/5, O410/1, O421, O430, O440, O694, O991/4/8, Z352)[2] Missing gestational age
Maternal				
b	Perineal tear	Third and fourth degree tears during vaginal delivery are not all preventable, but the risk can be reduced through appropriate labour management and care standards.	Admission includes both diagnosis (ICD-10: O702-3) and procedure code (OPCS: R322, R325)	Caesarean sections
c	Puerperal infection	Puerperal sepsis is associated with poor care (especially hygiene) at the time of birth	Admission includes diagnosis code for puerperal sepsis (ICD-10: O85-6) within 42 days of birth	None
d	30-day emergency readmission (all cause)	Readmission rates have previously been shown to be associated with the quality of hospital care	Emergency inpatient admission within 30 days, where 'admission source' field does not suggest transfer from other provider	Births where the discharge date is in final month of year covered in data year (March)
Neonatal				
e	In-hospital perinatal mortality	Still births plus in-hospital deaths within 7 days	Value suggesting still born in birth status (2, 3, 4) or discharge fields (5) fields, and discharged within 7 days with discharge method (4) suggesting death	None
f	Injury to neonate	Although rare, birth trauma to the neonate is often preventable[3]	Injury diagnosis code within admission (ICD-10: P102-4, P108-12, P114-5, P119, P122, P130-1, P138-9, P142, P148-9, P15)	Premature births, as identified through diagnosis code (ICD-10: P070-3), gestational age (<28 weeks) or birth-weight (< 2.5kg). Injury of skeleton (Osteogenesis imperfecta diagnosis, ICD-10: Q780) Stillborn babies
g	Selected neonatal infections	Neonatal infection is associated with poor care (especially hygiene) at the time of birth	Diagnosis procedure code (ICD-10: P36, P372/5/8-9, O753, O85-6, A41, A32, A49)	Premature births, as identified through diagnosis code (ICD-10: P070-3), gestational age (<28 weeks) or birth-weight (< 2.5kg). Stillborn babies
h	28-day emergency readmissions (all cause)	Readmission rates have previously been shown to be associated with the quality of hospital care	Emergency inpatient admission within 28 days, where 'admission source' field does not suggest transfer from other provider	Births where the discharge date is in final month of year covered in data year (March) Missing birth discharge date (c. 1% of births) Death during birth admission

Notes:

1. NICE guidelines on caesarean sections: http://www.nice.org.uk/nicemedia/live/13620/57163/57163.pdf

2. Specifications Manual for Joint Commission National Quality Measures (v2013 A1). Appendix A Table 11.07, available at: https://manual.jointcommission.org/releases/TJC2013A/

3. AHRQ, details available at: http://www.innovations.ahrq.gov/content.aspx?id=2081

9.4. Application and validation of indicators

9.4.1. Assumptions

9.4.1.1. Case ascertainment considerations

The next stage was to extract both baby and maternal records from 1 April 2011 to 31 March 2012. There are many potential fields and codes within HES to identify records relating to the delivery admission (Figure 25).

Figure 25. Key diagnosis fields and codes for identifying delivery episode

Admission methods (admimeth)	Episode type (Epitype)	Diagnosis codes (ICD-10)	Procedure codes (OPCS-4)	Delivery method
31 Maternity: where the baby was delivered after the mother's admission **32** Maternity: where the baby was delivered before the mother's admission **82** Other: babies born in health care provider **83** Other: babies born outside the healthcare provider, except when born at home as intended **98** Not applicable (eg other maternity event)	**2** Delivery episode **3** Birth episode **5** Other delivery event **6** Other birth event	**Z37** Outcome of delivery **Z38** Liveborn infants according to place of birth **O80** Single spontaneous delivery **O81** Single delivery by forceps and vacuum extractor **O82** Single delivery by CS **O83** Other assisted single delivery **O84** Multiple delivery	**R17** Elective caesarean delivery **R18** Other caesarean delivery **R19** Breech extraction delivery **R20** Other breech delivery **R21** Forceps cephalic delivery **R22** Vacuum delivery **R23** Cephalic vaginal delivery with abnormal presentation of head at delivery without instrument **R24** Normal delivery **R25** Other methods of delivery	**0** Spontaneous vertex *(OPCS R24)* **1** Spontaneous other cephalic *(R23)* **2** Low forceps, not breech *(R21.5,8-9)* **3** Other forceps, not breech *(R21.1 -4)* **4** Ventouse, vacuum extraction *(R22)* **5** Breech *(R20)* **6** Breech extraction *(R19)* **7** Elective caesarean section *(R17)* **8** Emergency caesarean section *(R18)* **9** Other than those specified above *(R25)*

To evaluate the importance of the case ascertainment algorithm, I checked the consistency of coding across the different fields. Some of these are governed by coding rules, with examples given in Figure 26 and whilst these rules were adhered to in 100% of cases in 2010-11 (used for exploratory purposes), across both birth are delivery records, I identified there were potential issues for using the 'epitype' field for case ascertainment.

Figure 26. Coding rule examples

Rule	Detail	Example issue
Rule 150	Where date of birth and start of episode are not missing, and date of birth equals epistart, then the first episode in the spell (epiorder = 1) with admission method 82 (*Other: babies born in health care provider*) is set as epitype = 3 (birth episode)	Does not apply for admimeth = 83 (*Other: babies born outside the healthcare provider, except when born at home as intended*) which were coded as: epitype = 3 (birth episode) for 57.7% (2709) of episodes; with 0.0% (2) as epitype = 2 (delivery episode) and 42.2% (1982) as epitype = 1 (general episode).
Rule 160	Where certain treatment codes (first 3 chars of oper_01 = R17 - R25) are recorded then episode set as epitype = 2 (delivery episode)	Does not cover secondary diagnoses (fields oper_02 to oper_20).

Source: HES Data Dictionary

9.4.1.2. Case ascertainment algorithm

I decided that a more detailed algorithm would be required than using the 'epitype' as a flag for a delivery or birth episode. After comparing the potential sensitivity and specificity of the possible case ascertainment algorithms and reviewing similar algorithms used in the existing literature, such as Knight et al (2013) and Murray et al (2012),[375,464] I proceeded with the codes outlined below. I did not use an algorithm based on procedure codes for identifying birth episodes since this strategy lacked sensitivity as the completeness of coding of these fields has been questionable, particularly historically.

Delivery episodes: Valid delivery method in either procedure (oper_01 to oper_20 = R17 to R25) OR maternity tail (delmeth = 0 to 9).

Birth episode: epitype = 3, 6 OR admimeth = 82, 83.

As with the stroke analysis, episodes were grouped into superspells – representing the continuous period that a patient receives care and includes transfers between hospitals. Where analyses were undertaken at the hospital/unit level, performance against the quality and safety measure was attributed to the first provider of care.

9.4.1.1. Unit of analysis: trust- and unit-level

To investigate the variation in performance across England, an appropriate level of analysis is required. In 2007, the Department of Health reiterated the policy that women should have a choice of where to give birth[487] and, as a result, providers are expected to offer multiple different maternity units. For instance, many NHS trusts have both a hospital-based consultant-led obstetric unit and a separate midwifery-led unit on a different site. Taking this example of an organisational set up, analysis at the unit-level will be better able to investigate the effect of organisational and structural decisions at this unit-level, whereas it is more likely to be affected by bias due unmeasured differences in case-mix resulting from different in the admission criteria to different units (a freestanding midwifery unit is typically intended for low risk mothers only). As such, I compared the two provider fields ('procode5' and 'sitetret') to an existing list of providers from a national audit of maternity services.[152] To validate the matching, I checked that the annual number of births recorded in the two datasets were similar. I report the analyses at both the trust and unit-levels.

9.4.2. Risk adjustment

9.4.2.1. Overview of case-mix adjustment

Once an extract had been obtained using the above criteria, the first step was to identify outliers, which involved applying the indicator definitions to obtain denominators and numerators for each hospital. A logistic regression was used to calculate an expected number of numerator events based on the case-mix for each trust/unit to account for case mix factors.[f] In addition, some of the measures have specific case-mix factor considerations as outlined in Table 42. The methods used are described in more detail in paragraphs 3.6.2 to 3.6.6.

9.4.3. Validation

9.4.3.1. Statistical testing and testing construct validity

Crude and standardised rates were plotted using funnel plots with 95% and 99.8% control limits and identified outliers (see paragraph 3.6.8, p81). Hospitals' performance across the different indicators was compared – at both trust and unit-levels – to evaluate the hypothesis that certain indicators would be correlated. This involved 28 pairwise comparisons for each of the institutional levels, with further detail on this method included in the discussion below. Given the lack of process measures, I supported the evaluation of construct validity by also exploring the association between

[f] Age of mother, sex of baby, parity (maternal indicators only), multiple deliveries, socio-economic deprivation (carstairs quintile), previous caesarean section (maternal only), ethnic group, gestational age, birthweight, delivery method and other maternal conditions (Pre-existing diabetes, gestational diabetes, pre-existing hypertension, pre-eclampsia or eclampsia, placenta praevia or abruption, polyhdramnios, oligohydramios.)

performance and intervention rates (using standardised instrumental rates). Since these were hypothesised as being correlated and not an un-hypothesised statistical trawling exercise, Bonferroni correction or other adjustment to p-values for multiple comparisons was not used, although I highlight where relationships are significant at the 99.8%, as well as 95%, levels.

9.4.3.2. Sensitivity analysis for coding practice

There is a risk that hospitals' performance for these indicators might be largely affected by variation in the way they code their data rather than due to differences in quality and safety. As such, the final stage of analysis was to investigate, at a site level, the consistency of coding practice and evaluate the relationship between any coding bias and hospital performance. Coding completeness was investigated, by comparing to performance in a relevant indicator, through:

> [1] Including a measure of 'coding depth', with the coding practice in some sites increasing the likelihood that they record secondary diagnoses and, therefore, identify complications and comorbidities.
>
> [2] Excluding sites with poor coding practice (defined as those sites with fewer than 90% of cases with both birth-weight and gestational age fields completed), and evaluating whether a similar proportion are still identified as outliers.

The effect of the first of these variations in the modelling [1], involved evaluating the effect of hospital-level variable to account for coding completeness, and so I fitted multi-level generalised linear models and again plotting on a funnel chart to identify outliers.

9.4.4. Amenable: associations between organisational factors and performance

9.4.4.1. Amenable: size and performance

To test whether the quality of care being measured by the indicators is amenable to how care is delivered, I undertook some exploratory analysis to investigate whether organisational factors were associated with variations in performance. In the first instance, I compared unit size (number of births) with performance against the indicators. This should be interpreted with caution as tertiary or midwifery-led units are likely to be different sizes to general maternity units and will treat a different case-mix.

9.4.4.2. Amenable: midwifery staffing levels

I also used data requested from the Health and Social Care Information Centre (HSCIC) on midwifery staffing levels, by trust and month. The number of births and maternities each month were calculated and a trust-level indicator of relative staffing levels, 'S', was defined as:

$$S_{i,t} = \frac{\frac{D_{i,t}}{M_{i,t}} - \sum_m \frac{D_{i,t}}{M_{i,t}}}{\sum_m \frac{D_{i,t}}{M_{i,t}}}$$

where i = trust i; t = month ϵ {April 2011,…, March 2012};

D = number of deliveries (births or maternities); M = number of midwives

I used this approach rather than comparing birth-to-midwife ratios between providers, since the latter is likely to reflect differences in type of provider (indeed staff-to-bed ratio have previously been used as a proxy to differentiate between major teaching, minor teaching, and nonteaching hospitals[257]).

9.4.5. Alignment with analytical framework

The table below summarises how this chapter aligns with the overall analytical framework (Table 43).

Table 43. Analytical framework

	Domain	Study design	Method described	Results described
Importanc	1. Impact	**Descriptive, national rates**	-	9.5.1.1
	2. Amenable	**Descriptive, range**	-	9.5.3
		Staffing levels	9.4.4.2	9.5.2.1
Scientific soundness of indicators	3. Face and content validity	**Consultation**	9.3.3.1	-
	4. Precision	**Hospital results** displayed on **funnel plots**	9.4.3.1	9.5.3
	5. Construct validity	**Correlations** between the set of indicators, including **outcome v. process** indicators	9.4.3.1	9.5.4
	6. Minimum bias	**Risk-adjustment model**	9.4.2	9.5.5.1
		Sensitivity analyses of influence of coding error and practice.	9.4.3.2	9.5.5.2, 9.5.5.3
Applicability	7. Data availability	**Application** of indicators to show ICD-10 based data is sufficient.	By implication	
	8. Reporting burden	**Application** of indicators.	By implication	
	9. External and ecological validity	Explicit statement of **assumptions** used.	9.4.1	

9.4.6. Analytical packages

All regression analyses were conducted using SAS version 9.2 using either the PROC LOGISTIC or PROC GLIMMIX procedures. Funnel plots were created using the template provided by the Association of Public Health Observatories (APHO, http://www.apho.org.uk).

9.5. Results

9.5.1. Importance

9.5.1.1. Maternal characteristics

I identified 669,617 maternities and 678,785 live and still births in the period April 2011 to March 2012. Some 61.5% of maternities were delivered by spontaneous vertex, with the characteristics of the study population in Table 44. The table also includes details of the cohort when excluding sites with poor coding, as described in the methods section (paragraph 9.4.3.2, p203). In addition other maternal complications included: pre-existing hypertension (0.5% of women), pre-existing diabetes

(0.6%), gestational diabetes (3.2%), pre-eclampsia/eclampsia (1.9%), placenta praevia/abruption (1.0%), polyhdramnios (1.0%), oligohydramnios (0.9%).

Table 44. Characteristics of mothers in the study population

Characteristic	Categories	Maternities, %	Restricted cohort	p-value (restricted vs excluded cohort)
		n = 669,616	*n = 516,919*	
Delivery method				p < 0.0001
	Spontaneous vertex	61.5	61.7	
	Spontaneous, other cephalic	0.5	0.5	
	Low forceps, non-breach	4.2	4.4	
	Other forceps, non-breach	2.2	2.0	
	Ventouse, vacuum extraction	6.3	6.2	
	Breech	0.4	0.5	
	Breech extraction not otherwise specified	0.1	0.1	
	Elective caesarean	10.1	9.9	
	Emergency caesarean	14.7	14.8	
	Other than those specified above	0.1	0.1	
Maternal age				p < 0.0001
	<19	5.0	5.1	
	20-24	18.5	18.8	
	25-29	27.8	27.9	
	30-34	28.9	28.7	
	35-39	15.8	15.5	
	40 and over	4.0	4.0	
Ethnicity				p < 0.0001
	White	72.1	72.4	
	Asian	10.5	10.5	
	Black (inc. Black British)Afro-Caribbean	5.0	5.2	
	Mixed	1.5	1.5	
	Other (inc. Chinese)	3.3	3.4	
	Unknown / not stated	7.6	7.0	
Level of deprivation				p < 0.0001
	1	14.9	14.2	
	2	16.1	15.9	
	3	18.9	18.8	
	4	22.1	22.2	

	5	26.6	27.6	
	'6' (unknown)	1.3	1.4	
Parity				p < 0.0001
	Primiparous	43.2	42.3	
	Multiparous	56.8	57.7	
Gestational age				p < 0.0001
	< 37	7.5	8.6	
	37-39	34.9	38.7	
	40-41	42.0	46.7	
	>=42	3.7	4.0	
	Unknown	11.9	1.9	
Birth-weight				p < 0.0001
	<2500g	5.8	6.4	
	2500-4000g	74.3	80.1	
	>4000g	10.2	11.0	
	Unknown	9.7	1.7	
Multip				p = 0.164
	Yes	1.5	1.5	
	No	91.7	97.1	
	Unknown	6.9	1.4	

Notes: $\chi 2$ test for association, excluding 'unknown' categories

9.5.1.1. National level results

Of those women meeting the inclusion criteria for the maternal complications: 3.1% had a perineal tear, 0.9% had an infection, and 1.0% were readmitted within 30 days of discharge. Of those neonates meeting the inclusion criteria for the relevant complications, 0.7% were stillborn or died in-hospital within 7 days, 1.5% were injured, 2.1% had an infection, and 6.1% were readmitted within 28 days of discharge. Around a third (33.9%) of the 67,370 elective caesarean sections were recorded as being conducted prior to 39 weeks gestation; after excluding deliveries for which there was no recorded gestational age (15.4%) and/or a recorded indication for early caesarean section (20.7%) this equates to 35.2% of cases.

9.5.2. Amenable

9.5.2.1. Association with organisational factors

The range in performance is shown in Table 46, p211. The association between staffing level and complication rates suggested that lower staffing levels were associated with higher perinatal mortality (p = 0.05) while there were not significant correlations with the other measures (Table 45).

Table 45: Association between birth-to-midwife ratio and performance measures

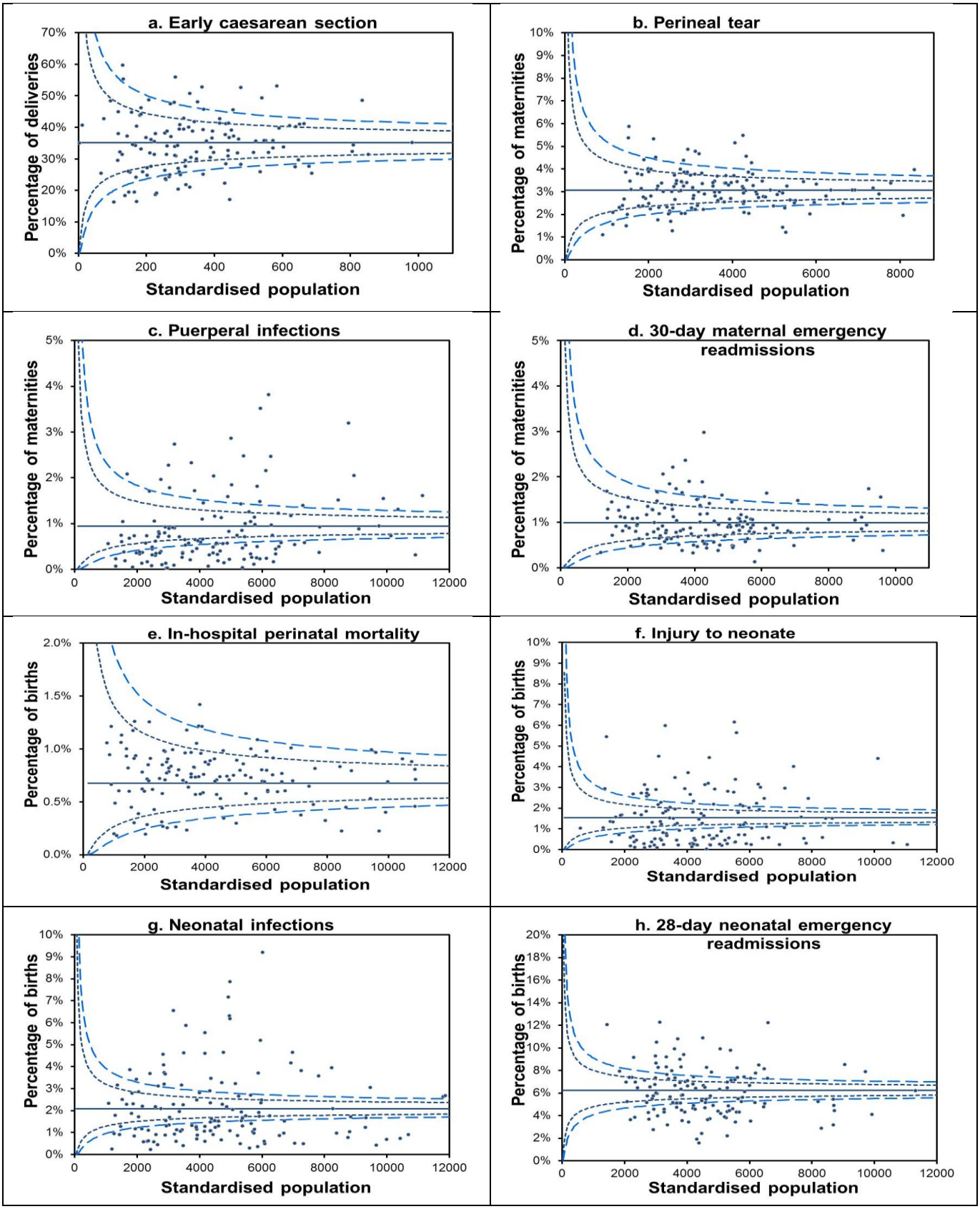
Label	Indicator	Effect	p-value
B	Perineal tear	-0.109	0.336
C	Puerperal infection	0.083	0.603
E	In-hospital perinatal mortality	0.397	0.054
F	Injury to neonate	0.228	0.107
G	Selected neonatal infections	0.134	0.277

Notes Readmission rates were not calculated as the performance could be affected by staffing levels in subsequent month. Early caesarean section was not calculated as this was not assumed to be affected by midwifery staffing levels.

9.5.3. Precision

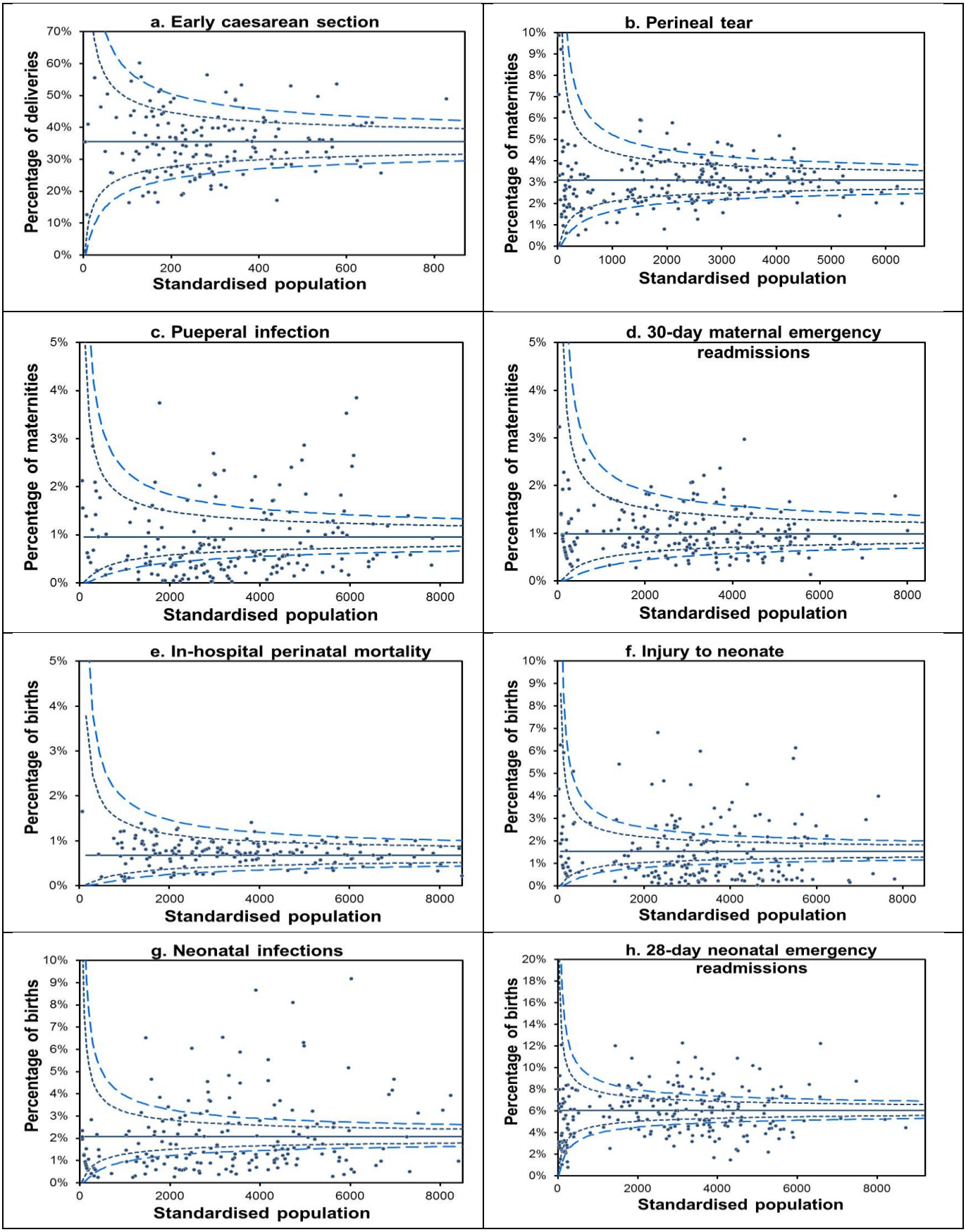
Displaying the performance at the provider level on funnel plots, across either the 149 trusts or 229 units, highlighted the variation in performance (Figure 27 and Figure 28). All the indicators identified at least trusts/sites with performance outside the 99.8% control limits (Table 46). The number of births per trust ranged from 1,291 to 10,782, and for sites from 10 to 8,516.

Figure 27: Funnel plots of trust-level performance across the obstetric indicators, 2011-12



Note: Each dot represents a trust. The horizontal line refers to national average; short-gauge dotted line refers to p<0.025 significance level; long-gauge dotted line refers to p<0.001 significance level.

Figure 28: Funnel plots of site-level performance across the obstetric indicators, 2011-12



Note: Each dot represents a site. The horizontal line refers to national average; short-gauge dotted line refers to p<0.025 significance level; long-gauge dotted line refers to p<0.001 significance level.

Table 46: Number of trusts identified as having statistically significant above or below average performance

Label	Indicator	National rate	Variation	Outliers: number of providers with different performance to average (p<0.001)[1]			
			Range, by trust[2]	Trust-level (n = 149)		*Site-level (n = 229)*	
			[min, max]	Higher[3]	Lower[3]	*Higher[3]*	*Lower[3]*
a	Early caesarean section	35.2%	[1.1% – 60.7%]	12	21	*11*	*21*
b	Perineal tear	3.1%	[1.0% – 5.3%]	13	17	*17*	*21*
c	Puerperal infection	0.9%	[0.2% – 1.9%]	22	25	*26*	*70*
d	30-day maternal emergency readmissions	1.0%	[0.2% – 3.0%]	15	21	*15*	*23*
e	In-hospital perinatal mortality	0.7%	[0.3% – 1.0%]	5	15	*3*	*17*
f	Injury to neonate	1.5%	[0.1% – 6.9%]	37	63	*41*	*73*
g	Selected neonatal infections	2.1%	[0.2% – 11.1%]	34	63	*36*	*77*
h	28-day neonatal emergency readmissions	6.1%	[1.6% – 14.1%]	31	47	*42*	*57*

Note 1: Some trusts/sites were removed for specific indicators (the total number given represents the maximum)

2: Range represents minimum and maximum performance for providers, for unadjusted rates

3: Better care is indicated by lower than average performance for these measures.

9.5.4. Construct validity

Across the trust-level indicator results, there were five pairs of indicators which had a statistically significant correlation at the 95% level, of which one was significant at the 99.8% level (Table 47).

Table 47: Trust-level Coefficient of correlation between pairs of indicators

	Perineal tear	Puerperal infection	30-day maternal emergency readmissions	In-hospital perinatal mortality	Injury to neonate	Selected neonatal infections	28-day neonatal emergency readmissions
Early caesarean section	-0.122	-0.076	-0.049	0.005	-0.189*	-0.077	-0.135
Perineal tear		0.121	0.198**	-0.119	-0.107	-0.174*	-0.227*
Puerperal infection			-0.067	0.014	0.218*	0.190*	-0.121
30-day maternal readmissions				0.002	0.080	-0.022	0.048
In-hospital perinatal mortality					0.038	0.008	-0.069
Injury to neonate						-0.062	0.074
Selected neonatal infections							-0.012

Note 1: Correlation significant at the 95% level marked with '*'; those at 99.8% level with '**'

In further exploratory analysis, I found that higher standardised instrumental rates are associated with higher rates of: emergency caesarean section (0.451, $p < 0.0001$); maternal infection rates (0.225, $p = 0.007$); and tears (0.38, $p< 0.0001$). Across the site-level indicator results, there were seven pairs of indicators which had a statistically significant correlation at the 95% level, of which three were significant at the 99.8% level (Table 48).

Table 48: Site-level coefficient of correlation between pairs of indicators

	Perineal tear	Puerperal infection	30-day maternal emergency readmissions	In-hospital perinatal mortality	Injury to neonate	Selected neonatal infections	28-day neonatal emergency readmissions
Early caesarean section	-0.107	-0.014	-0.030	0.006	-0.076	0.030	-0.004
Perineal tear		0.036	0.012	0.081	0.153*	-0.045	-0.202*
Puerperal infection			-0.036	0.129	0.067	0.259**	-0.051
30-day maternal readmissions				0.218**	-0.024	0.179*	0.074
In-hospital perinatal mortality					0.047	0.355**	0.153*
Injury to neonate						0.017	-0.064
Selected neonatal infections							0.097

Note 1: Correlation significant at the 95% level marked with '*'; those at 99.8% level with '**'

9.5.5. Minimum bias: coding practice

9.5.5.1. Effect of case-mix adjustment

The analysis of the association between adjusted and unadjusted rates, suggested that the case-mix adjustment had the least effect on the measure of maternal readmissions (Pearson correlation coefficient, r = 0.969) and greatest effect on perinatal mortality (r = 0.540) (Table 49). These results tally with the model fit statistics, which show that the case-mix factors are better at predicting perinatal mortality (c = 0.853, r^2 = 0.268) than maternal readmissions (c = 0.605, r^2 = 0.014).

Table 49: Correlation between adjusted and unadjusted rates, and model fit statistics

Label	Indicator	Correlation coefficient		Model fit statistics	
		Pearson	Spearman's rank	c	r^2
a	Early caesarean section	0.849**	0.935**	0.688	0.235
b	Perineal tear	0.903**	0.851**	0.721	0.077
c	Puerperal infection	0.868**	0.971**	0.714	0.057
d	30-day maternal emergency readmissions	0.969**	0.972**	0.605	0.014
e	In-hospital perinatal mortality	0.540**	0.576**	0.853	0.278
f	Injury to neonate	0.948**	0.965**	0.727	0.097
g	Selected neonatal infections	0.923**	0.913**	0.765	0.134
h	28-day neonatal emergency readmissions	0.936**	0.945**	0.586	0.016

Notes ** $p < 0.0001$;

r^2: the max-rescaled r-square score reported from PROC LOGISTIC

9.5.5.2. Coding depth (diagnoses)

Between sites the average number of distinct diagnosis codes used (excluding those relating to perineal tear) per admission ranges from 2.8 to 6.6. Similarly, between sites the average number of distinct diagnosis codes used (excluding those relating to injury to neonate) per admission ranges from 1.1 to 4.4. Compared to the standard logistic model, including a multi-level model with a variable to account for coding practice made little difference to trusts performance in the perineal tear measure (Pearson correlation coefficient, r = 0.986) although there was a more substantial effect for the injury to neonate measure (r = 0.794) (Table 50).

Table 50: Effect of including coding practice in case-mix model

Indicator	Correlation coefficient	
	Pearson	Spearman's rank
Perineal tear (v. adjusted rates 'standard' model)		
Multi-level model (including site identifier)	0.991**	0.997**
Multi-level model, including site coding variable	0.986**	0.983**
Injury to neonate (v. adjusted rates 'standard' model)		
Multi-level model (including site identifier)	0.992**	0.989**
Multi-level model, including site coding variable	0.794**	0.809**

Note ** $p < 0.0001$

For the perineal tear measure, at this site-level there was no correlation ($r = -0.029$, $p = 0.666$) between coding depth (measured as average number of unique diagnosis codes in an admission) and performance. Of the 38 sites identified at the 99.8% level in the original regression, all (100.0%) were again flagged as outliers (and in the same direction) at this significance level when coding practice was included in the regression.

For the injury to neonate measure, at this site-level there was a significant correlation ($r = 0.260$, $p = 0.0002$) between coding depth (measured as average number of unique diagnosis codes in an admission) and performance. Of the 114 sites identified at the 99.8% level in the original regression, 77 (67.5%) were again flagged as outliers (and in the same direction) at this significance level when coding practice was included in the regression. In addition, three of the original outliers (2.6%) were flagged as outliers in the opposite direction.

9.5.5.3. Coding depth (characteristics)

As shown above (**Table 44**, p206), the maternal characteristics of the cohort treated by sites with good coding depth on birth-weight and gestational age, were statistically different to those at poor coding sites, although the actual differences are small.

Across poor coders, 44 of 71 sites were outliers at the 99.8% level for the injury measure whereas 70 of the 121 good coders were. The distribution of performance (Higher 0.001, High 0.025, Within control, Low 0.025, Low 0.001) was not statistically different between good and bad coders ($p = 0.438$).

Across poor coders, 8 of 55 were outliers for perineal tear measure whereas 30 of 161 of good coders were. The distribution of performance ('Higher 0.001', 'High 0.025', 'Within control', 'Low 0.025', 'Low 0.001') was not statistically different between good and bad coders (p = 0.485).

9.6. Discussion

9.6.1. Summary of findings

The results show the potential for using hospital administrative data, meeting the Government's call to develop measures of performance and, moreover, to highlight potentially significant variations in the quality and safety across the delivery admission. Eight measures of quality and safety, covering the perinatal care pathway, were applied to English hospital administrative data identifying almost 700,000 births. All eight indicators identified trusts and units with statistically outlying performance at the 99.8% level. Further analysis suggested that differences in coding practice appeared to only partially explain the variation. The analysis reiterated the high impact of maternity care with large numbers of potential adverse events.

9.6.2. Amenable: scope for improvement

For most of the measures, there is no clinical consensus or guidelines for what actual levels are acceptable. In these cases, hospitals should strive to operate at the level of the best performers, after accounting for random variation and variation in case-mix. Indeed, regulators and guideline-setters could use this analysis to suggest levels of performance that are attainable. The exception to this rule is on caesarean sections prior to 39 weeks, for which the extant guidelines state that "the risk of respiratory morbidity is increased in babies born by CS before labour, but this risk decreases significantly after 39 weeks [and] therefore planned CS should not routinely be carried out before 39 weeks" (p18). [490] As such, the current rate of around a third of elective caesarean sections being in contravention to these guidelines seems unacceptable.

9.6.3. Precision

By using funnel plots and control limits, I was able to account for random variation; if the deviation were entirely due to random variation, you would only expect to identify around one outlying maternity unit every 2 measures (given there are 229 units, with an even lower likelihood for across the lower number of trusts). Further to chance, differences in performance may be due to case mix, how the data were collected, or quality of care. [337] Given the ambition was to compare providers on the basis of the last of these factors, the previous two also need to be accounted for and are, therefore, discussed below.

The panels of funnel plots above look like there is less over-dispersion when analysed at a site level. Indeed, there seems to be a smaller proportion of sites than trusts identified as outliers with, for instance, 33 trusts highlighted as outliers at the 99.8% confident level and only 32 sites. However, this may be explained to some extent by the large number of very small units (those clustered on the left of the funnel plots) where numbers are insufficient to give appropriate power to identify variations in performance.

9.6.4. Minimum bias

9.6.4.1. Adjusting for case-mix

The influence of maternal and foetal risk factors on outcomes is well-recognised. [491] I accounted for case-mix using mother- or neonate-level logistic regression to calculate the expected number of events for outcome measures. However at an individual-level some significant case-mix factors, such as maternal obesity and smoking,[491] are not recorded within the data and, therefore, some of the variation may still be caused by differences in case-mix. If we hypothesise that there are small areas of extreme case-mix but, at a larger geographical level, case-mix is more homogenous coupled with units and trusts predominantly treating patients from their local area then it holds that any bias from this potential 'categorisation error' is less likely to have a substantial effect within the trust-level analysis given that at least 1,000 deliveries occurred at each provider, in comparison to the unit-level results whereby one unit only had 10 deliveries.

I report, in detail, the model fit statistics and effect of the case-mix adjustment of hospitals' performance. This showed that the case-mix models were better at explaining performance in the perinatal mortality score to the readmission measures. This information is not for making conclusion on the appropriateness of the case-mix adjustment – since it is not possible to disaggregated the effect of unmeasured differences in case-mix to differences in quality of care provided – but should be considered when interpreting the results. One specific issues relating to case-mix originates from the difference between obstetrician-led and midwifery-led care. Midwifery-led units are recommended only for low-risk mothers and, therefore, obstetric units are more likely to a higher proportion of high-risk cases.

9.6.4.2. Accounting for variation in coding practice

The variation in performance due to how data were collected is harder to disaggregate since some of the possible variation in coding practice, such as depth of diagnosis coding, might mimic differences in case-mix. However, I showed that two key coding issues that were central to the assumptions for identifying complications and adjusting for case-mix explained only a minority of the differences in

performance against the measures. Previous research has raised questions about differences in performance between hospitals with good and poor data recording – for instance, a higher postoperative mortality rate following resection of colorectal cancer was found in trusts that do not voluntarily report data to a national audit[492] –the findings here suggest that outliers were in both good and bad recording trusts.

9.6.5. Errors in underlying data

Records of births outside of NHS hospitals are significantly affected by data quality issues and do not contain information for fields such as admission/discharge date (admidate/disdate) or patient classification (classpat). Historically there has been poorer coverage of home deliveries, and for this reason, only records for births in an NHS setting are included in the HES Annual Maternity publication.[493] Yet there are wide geographical and demographic variations in the proportion of home births,[152] which could cause bias if creating performance indicators without these home births. Moreover, some of this missing data can be estimated from the date of birth and, as such, the research presented here did not actively exclude births outside hospitals.

The additional fields have their own data quality issues and the coverage is not as complete as for the rest of HES data. For instance, a research project suggested that key data items were missing in over 20% of records overall, with some NHS trusts not submitting usable maternity tail data for any deliveries. [375] However, I investigated these issues by disaggregating the analysis between good and poor coding hospitals, as reported above.

9.6.5.1. Case-ascertainment strategy

If the total number of cases attained from the algorithm applied to the HES extract was either substantially higher or lower than that derived from the national birth register then this would suggest either poor sensitivity or specificity of the case-ascertainment strategy. Reassuringly, however, the number of maternities and births from the extract were similar to the register. The numbers of maternities and births from the HES data closely matches the register data – differing by only 0.3 and 0.2%, respectively, in 2011 – and has improved over time (Table 51).

Table 51. Comparison of numbers of maternities and births from the HES extract and ONS births register

	Maternities			Live and still births		
	Register	HES Extract	*Variation*	Register	HES Extract	*Variation*
2008	665,779	643,658	*-3.3%*	676,236	649,741	*-3.9%*
2009	645,802	640,957	*-0.8%*	656,880	651,824	*-0.8%*
2010	662,752	661,162	*-0.2%*	673,641	670,565	*-0.5%*
2011	664,468	666,352	*+0.3%*	675,646	673,961	*-0.2%*

Notes: ONS data from Characteristics of Birth 1

9.6.6. Face and content validity

Whilst previous studies have used hospital administrative data to measure the performance of aspects of obstetric care, none have brought together such a broad range of measures covering both maternal and neonatal care and, instead, have tended to focus on a single area such as perineal tears. A combination of process and outcome measures (both short- and longer-term) was used in this study, so benefiting from the advantages of process measures (which tend to be more sensitive to differences in the quality of care and offer a clear action for improvement) and outcome measures (greater intrinsic interest, high face validity, and can reflect all aspects of care, including those that are otherwise difficult to measure such as technical expertise and operator skill).[337,342] Face validity can be derived from the fact that the indicators appeared in existing literature or, for one of the measures, clinical guidelines.

The overall level of quality of care at a provider-level cannot be deduced from the results of only a small set of indicators; however almost half of studies within the literature review applied only one indicator. This study included eight indicators covering many aspects of the care provided during the delivery admission and, therefore, is likely to have better content validity (extent to which a measure represents all facets of the subject being investigated) over these previous more limited evaluations.

9.6.7. Criterion validity

As well as these considerations on whether the indicators' algorithms are functioning as intended, further assurance needs to be taken about whether the results of the indicators reflect actual performance. The most common process for validating the results is to compare against a gold standard. Previous research has focused on the perineal tear (obstetric trauma) indicator, including: the California Obstetric Validity Study which found that the indicator had sensitivity of 90% (95% CI, 82-96%) and PPV of 90-95%;[392] a study using a clinical research dataset as comparison which

suggested sensitivity of 77% (72-81%) and PPV of approximately 93%;[494] and, within England, research suggested a PPV of 85% based a review of 995 case notes.[112]

9.6.8. Construct validity

In the absence of a gold standard dataset to compare results, some assurance on validity can be derived by comparing with other indicators of performance. I investigated whether there were associations between trusts' and units' performance across the different measures. One might hypothesise there would be correlation between good performance across all the indicators with, for example, good management of maternity care likely to affect all the quality and safety measures to some extent. More specifically, you would expect a hospital with good hygiene levels to have good performance against both the maternal and neonatal infection measure. Ideally there would be a range of outcomes and process measures focusing on the same event within the delivery to support this validation; however, the literature review revealed only limited process measures and these were not directly linked to the outcome measures.

There were 6 statistically significant correlations at the 95% confidence level at the trust-level, and 7 at the unit-level. However, only 2 of these were common to both trust- and unit-level, which could suggest that there are different unit and trust-level effects that affect performance.

At the trust-level, the only correlation significant at the 99.8% level was a positive correlation ($r = 0.198$) in the direction of good performance in both maternal readmissions and perineal tears, which might be expected as poor care during the delivery, as indicated by obstetric trauma, is likely to result in readmissions after discharge.

As hypothesised, at the unit-level there is a strong positive association between neonatal and maternal infections ($r = 0.259$, $p < 0.002$). Across the units and at the 99.8% confidence level, there was also positive statistically significant correlations between perinatal mortality and both neonatal infections ($r = 0.355$) and maternal readmissions (0.218).

There were further statistically significant correlations at the 95% confidence level. Some of these (such as injury to neonate and puerperal infection at both trust and unit levels) showed a positive association between complication rates, as expected. However, at this significance level, there were also unexpected negative associations (neonatal readmissions with perineal tear at the site and trust levels; and at a trust-level only for early caesarean sections and injury to neonate and between neonatal infections and perineal tear). It is unclear why these correlations are found; however, across the 56 comparisons we might expect around 3 significant correlations by chance alone at this level which might provide the explanation.

9.6.9. Unit of analysis

9.6.9.1. Provider identification

While I was able to identify the trust-level 'NHS code' for every birth and maternity, the attribution to specific maternity unit was more complex. The matching of fields for identifying the unit of provider identified, for 2011-12, 224 distinct geographic locations – either obstetric unit, obstetric unit and alongside freestanding midwifery unit, or freestanding midwifery unit – and 5 providers where I was unable to disaggregate the distinct locations, although for 2 of these providers[g] the separate units were of the same type. In comparison a national audit of maternity surveys identified 229 distinct geographical locations, as at June 2013. The closeness of these numbers suggests that the HES fields can be used to identify the vast majority of maternity units.

Performance might be affected by both trust-level factors – e.g. overall staffing levels given that midwives are often used flexibly between units – and unit factors – e.g. suitability of the estate. As discussed above, analysis at the trust level will better control for unmeasured differences in case-mix. However this more aggregated level might not be sensitive to some of the local differences in quality. As such, there seems to be value to analyse the results at both levels.

Whereas the report by the Royal College of Obstetricians and Gynaecologists was limited to those units with greater than 1000 deliveries, I included the full set here. This inclusion criterion by the Royal College seems unnecessary as they too apply a statistical approach that accounts for the random variation which is likely to have a greater effect on smaller units.

9.6.10. Strengths of this study to other measurement efforts

The Confidential Inquiry into maternal mortality has been commonly cited and linked to many of the national standards that had been used in the scheme for providing litigation cover from trusts in England (Clinical Negligence Scheme for Trusts). However, the old confidential inquiry was also widely criticised for being "anecdotal", expensive and without evidence of benefit.[495,496] The RCOG have made advances in measuring performance;[375] however, the set of indicators used considered only maternal process and outcomes of care, and did not compare sites performance across the different indicators. Indeed, the RCOG recommended that "there is also a need to better understand the relationships between different process and outcome indicators. Priority areas should be neonatal outcomes and measures of user experience." [375]

[g] County Durham and Darlington (RXP) includes two obstetric units (RXP00 and RXPBA). Doncaster and Bassetlaw (RP5) includes two obstetric units.

HES has the advantage of being: longitudinal; timely; covering all hospital admissions; and being relatively cheap, costing £1 per record to collect compared with around £10 - £60 per record for clinical registers.[89] The introduction of a National Maternity Dataset for England, mandated from April 2014, should provide a rich and accurate source of information on the care of pregnant women.[375]

9.6.11. Conclusion

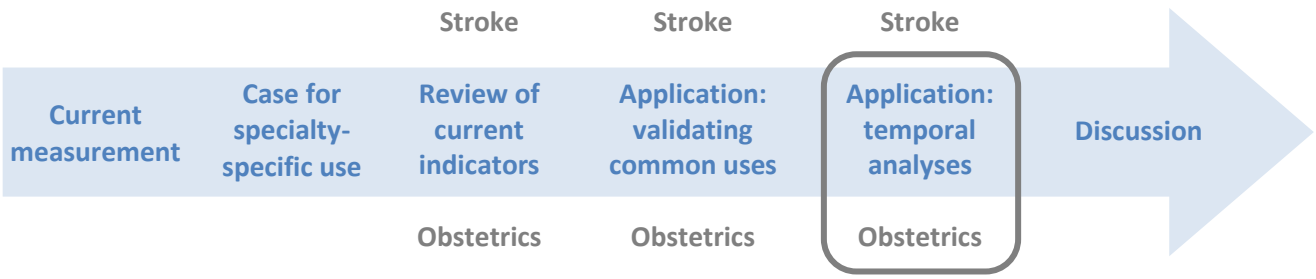
This study shows that hospital administrative data have the potential for benchmarking the standard of hospitals performance over the care provided in the delivery admission, in a timely way and without increasing the bureaucratic burden on hospitals. While the study benefits from drawing together both neonatal and maternal indicators being able to link mother and baby records – for instance to create a compound measure of 'no complication form delivery' – would give a more complete picture of performance [375]

As has been previously noted above and reiterated by others, "the data on which our analyses are based are not as accurate as we would have liked; however, until these data are used to provide information to allow for meaningful benchmarking, encouraging clinicians to take ownership of their own hospital data and attempts to drive up quality will be difficult." (p. viii) [375] There are a number of reasons for the coverage and data quality issues, such as different capabilities of local systems to record the additional birth details.[497] Indeed, a recent audit found that nearly a fifth of maternity units' information systems were either heavily reliant on paper-based notes or not linked to the patient administration system, which will negatively affect the ability to provide accurate, timely data. [152] However, the government recently stated that they were ensuring that all maternity units would be required to have appropriate IT.[498]

Whilst neither the lack of credible gold standard nor evidence of bias from coding practice render the indicators redundant, it does temper the extent to which they could be used, such as for regulating hospitals. This study shows that HES provides the facility to record some key process and outcome measures across key aspects of maternity care in a cheap and timely manner. These results could be linked to structural measures, such as unit-level data on staffing levels, to investigate their effectiveness. [375] With suitable modification and further development, these data could also be used routinely as currently measurement efforts.

Chapter 10.

Obstetrics temporal trends



Overview

Context

The previous chapter showed how obstetric indicators can be used to highlight variations in the quality and safety of care and demonstrated that, while the findings are relatively robust to differences in coding practice, this does represent a limitation. This chapter shows how these limitations can be overcome through different application of the indicators. The literature review revealed that less than a quarter of studies on obstetric care had used the longitudinal nature of administrative data, and none had comprehensively assessed weekend care.

Methods

Application of indicators covering key aspects of the delivery admission to investigate changes in performance over time and association between day of delivery and adverse events. Results were adjusted for case-mix using logistic regression.

Findings

The longitudinal analysis highlighted both seasonal and longitudinal trends. The day of the week analysis showed performance across four of the seven measures was significantly lower at weekends. In particularly, the perinatal mortality rate was 7.3 per 1,000 at weekends , some 0.9 per 1,000 higher than at weekdays (adjusted odds ratio 1.071; 95% Confidence Interval 1.019 – 1.126).

What this chapter adds

This chapter presents the most comprehensive evaluation of the “weekend effect” in obstetric care. The study design mitigates for common limitations cited on the use of such indicators and so is able to produce robust results about the extent of this important issue.

10.1. Background

10.1.1. Existing obstetric studies on administrative data

10.1.1.1. Topics, designs and results

The literature review (Chapter 8) revealed that few studies have taken advantage of the longitudinal nature of administrative data to evaluate temporal trends in the quality and safety of obstetrics care. Where studies have existed, the temporal trends investigated have included longitudinal annual trends, seasonal trends (a so called "July" effect) and association with day of admission (Table 52).

Only one of these studies was based on English data, and this evaluation included only timing of caesarean sections as a measure.[411] The treatment of changes in case-mix and patient characteristics varied amongst this small set of studies with, for instance, Callaghan and colleagues (2012) not making any risk-adjustment to the evaluation in changes in maternal mortality and severe morbidity.[442] Most of these studies fail to stand up to the criticism of primarily focusing on a single outcome and therefore not capturing wider aspects of the quality and safety of care.

The study designs and aims were heterogeneous and so a formal meta-analysis is not possible but, in summary, they present a mixed picture on longitudinal trends with some countries showing improvement in certain indicators and worsening performance in others. Neither the study of the "July" effect [444] nor "weekend effect" [445] showed statistically significant variations in quality and safety.

10.1.2. Other literature

10.1.2.1. Longitudinal and temporal trends

A recent national audit suggested that there had been improvement across some key aspects of maternity care over time, such as early access and consultant presence.[152]

10.1.2.2. Association between performance and day of the week

Previous studies, across a range of countries, have identified higher mortality in patients admitted on weekends across a range of medical conditions; a phenomenon termed the 'weekend effect'.[253,350-353,499] This calls into question the idea that quality of care is equal irrespective of when you present at hospital. However, not all studies have identified an association between poor outcomes and out-of-hours periods.[500-502]

In 1978, Alison MacFarlane published a paper showing a seven-day cycle in birth numbers across England (and Wales) and that perinatal mortality was higher among babies born at weekends.[503] Similar studies in the 1970s and 1980s found this phenomenon in other developed countries.[504-506] The delivery of obstetrics care has changed dramatically since this time; however, where the "weekend effect" has been evaluated it has predominantly been based on mortality as the sole indicator. Indeed, in setting out key challenges to address in obstetrics care, a paper by the World Health Organisation (WHO) highlighted ineffective referral to and inadequate availability of 24-hour quality services to emergency obstetric care services.[367]

10.2. Objectives

In this chapter I investigate the potential to use the longitudinal nature of administrative data to identify important shortcomings in quality and safety in a way that overcomes some of the known limitations to such indicators.

Table 52. Studies using obstetric indicators to investigate temporal issues

Study	Cohort	Country	Variable	Measures	Risk-adjustment	Main results
Cameron et al (2006)[440]	52,151 women with PPH (1994 – 2002)	Australia	Year	Primary measures: perinatal PPH and readmissions from PPH	Age and mode of delivery Adjusted for under-reporting	From 1994-2002, PPH increased from 8.3 to 10.7% of deliveries. Hospital readmission for PPH declined from 1.2% to 0.9%.
Gurol-Urgancil et al (2011)[411]	507,410 low risk singletons (2000 – 09)	England	Year	Elective CS after 39 weeks	None – excluded women who had elective CS before 24 weeks or contra-indication for delaying an elective CS	Elective CS deliveries after 39 weeks steadily increased from 39% in 2000/01 to 63% in 2008/09.
Downey et al (2012)[441]	Not stated (1998 – 2007)	USA	Year	Obstetric trauma – vaginal with and without instrument Birth trauma – injury to neonate	Age	Average annual decrease in events: birth trauma -17.8%; obstetric trauma without -5.7%; obstetric trauma with -4.1%.
Callaghan et al (2012)[442]	49,346,974 deliveries (1998 – 2009)	USA	Two-year periods	Severe maternal morbidity and related in-hospital mortality	None	Severe maternal morbidity increased by 75% (and 114% for delivery and postpartum hospitalisation) from 1998/99 to 2008/99.
Roberts et al (2012)[443]	500,603 women (1999 – 2004)	Australia	Year	Maternal morbidity outcome indicator (MMOI) PPH	Age, parity, mode of delivery, smoking, delivery hospital, multiple pregnancy, previous CS, and other clinical factors	MMOI increased annually by an average 3.8% (95% CI: 2.3-5.3%).
Ford et al (2007)[444]	299,130 singleton births (1998 – 2002)	USA	Month (July vs non-July)	Include: third and fourth degree laceration; PPH; transfusion; anaesthesia related complication; birth asphyxia	Age, ethnicity, socio-economic status, hospital size and location	No increased rate of operator-dependent complications of delivery at teaching hospitals nationwide in the month of July.
Gould et al (2003)[445]	1,615,041 live births (1995 – 97)	USA	Day of birth	Neonatal mortality rates	Method of delivery, birth weight and excluding congenital anomalies	Weekend neonatal mortality OR 1.12 (95% CI 1.05-1.19) compared to weekdays. Difference no longer significant after adjusting for birth weight.

10.3. Methods

10.3.1. General methodological considerations

10.3.1.1. Indicators

The indicators used for this part of the study are similar to those used in the previous chapter (paragraph 9.3.3.2, p198) and were identified in the literature review (Chapter 8). In summary, the indicators used in this chapter are set out in the below (Table 53), although some amendments were used in the weekend analysis, with the rationale for this discussed below.

Table 53. Obstetric indicators

Delivery	Maternal	Neonatal
a. Early (pre-39 week) caesarean section)	**b**. Perineal tear	**e.** In-hospital perinatal mortality
	c. Puerperal infection	**f.** injury to neonate
	d. 30-day emergency readmission rates	**g.** selected neonatal infections
		h. 28-day emergency readmission rates

10.3.1.2. Data extract

The details of deliveries from 1 April 2009 to 31 March 2012 were extracted from the HES database. The indicator definitions were applied to the extract to obtain denominator and numerators, categorised by either month or day of admission (for maternal indicators) or birth (for neonatal records). The weekend analysis was restricted to 1 April 2010 to 31 March 2012 to reduce any bias from poorer coding in the preceding year.

10.3.1.3. Case-mix adjustment

Variables expected to influence the outcome of the association between the quality and safety indicators and day of admission/birth were extracted, including gestational age, birth-weight, and maternal age. A full set of case-mix factors and the method for risk-adjustment is covered in paragraph 9.4.2, p202.

10.3.1.4. Data analysis and software

National performance across the measures was calculated, disaggregated by either month or day of admission/birth. Multiple logistic regressions were used to adjust to the effect of the covariates. I did not adjust for the clustering of patients within hospital as the hospital-level effects were found to be small (as shown in Chapter 9) except when including site level data (such as hours of consultant presence) as described explicitly below. Results were also displayed by plotting – using either April 2009 (or Tuesday) as a reference – odds ratios and 95% confidence intervals, by month (or day) of delivery.

Analyses were carried out using SAS Version 9.2, using the PROC LOGISTIC and PROC GLIMMIX procedure for regression analyses.

10.3.2. Additional 'day of the week' methodological considerations

10.3.2.1. Defining 'weekend'

Where results are presented as weekend versus weekday performance, the former was defined as the period from midnight on Friday to midnight on Sunday, with all other times defined as weekdays (the time of admission is not captured in HES).

10.3.2.2. Estimating impact

Using regression analysis on just Tuesdays, probabilities of in-hospital perinatal death and puerperal infections (maternal) were derived. By matching these probabilities for each Tuesday admission, based on the mother's or neonate's characteristics, indirectly standardised estimates for the outcomes as if those non-Tuesday cases had had similar rates as their Tuesday counterparts were calculated, with these particular results reported in the discussion. I used Tuesday (rather than Monday, as for the stroke analysis) as the reference since the neonatal indicators use date of birth to assign the day even through the outcome might be affected by the standard of care given to the mother during labour and, as such, Monday's performance might be affected by quality of care over the weekend for non-same-day births.

10.3.2.3. Amendments to indicators

The amendments made to these indicators, for the purpose of this day-of-the-week study, were to:

- use 3-day readmission rates for both maternal and neonatal indicators in keeping with the suggestion that it is a more appropriate timeframe to evaluate the association between day of admission and mortality.[354,499]

- drop the indicator for early caesarean sections as such deliveries are not scheduled during out-of-hours periods.

10.3.2.4. Consultant presence

Data on hours of consultant presence on the labour ward were collected obtained from two previous audits, by the National Audit Office[152] and Royal College of Obstetricians and Gynaecologists (RCOG). [375] Data were linked using unit names with 100% completeness. A binary indicator of compliance with was created based on recommendations RCOG as set out in the table below (Table 54). The analysis was restricted to obstetric units and sites where data were reported for the obstetric and alongside midwifery unit combined.

Table 54. Recommended levels of consultant presence

Births per year	Recommended minimum consultant presence per week
2,500 – 4,000	60 hours
4,000 – 5,000	98 hours
5,000 +	168 hours

Note: Source table quoted in National Audit Office report.[152]

10.3.2.5. Association between levels of activity and performance

The level of activity, A, was defined using the following equation:

$$A_{u,y} = \begin{cases} 0 \; if \; \frac{n_{u,d,y}}{\frac{\sum_d n_{u,d,y}}{365}} \leq 0.8 \\ 1 \; if \; \frac{n_{u,d,y}}{\frac{\sum_d n_{u,d,y}}{365}} > 0.2 \end{cases}$$

where u = unit u;

d = day ϵ {1 April,…, 31 March};

y = year ϵ {2011, 2012}

10.4. Results

10.4.1. Longitudinal analysis

10.4.1.1. Overview of results

Between April 2009 and March 2012, I identified around 4 million births and maternities. The most common adverse event was early caesarean section (37.50%) with the least common being in-hospital perinatal mortality (0.67%) (Table 55).

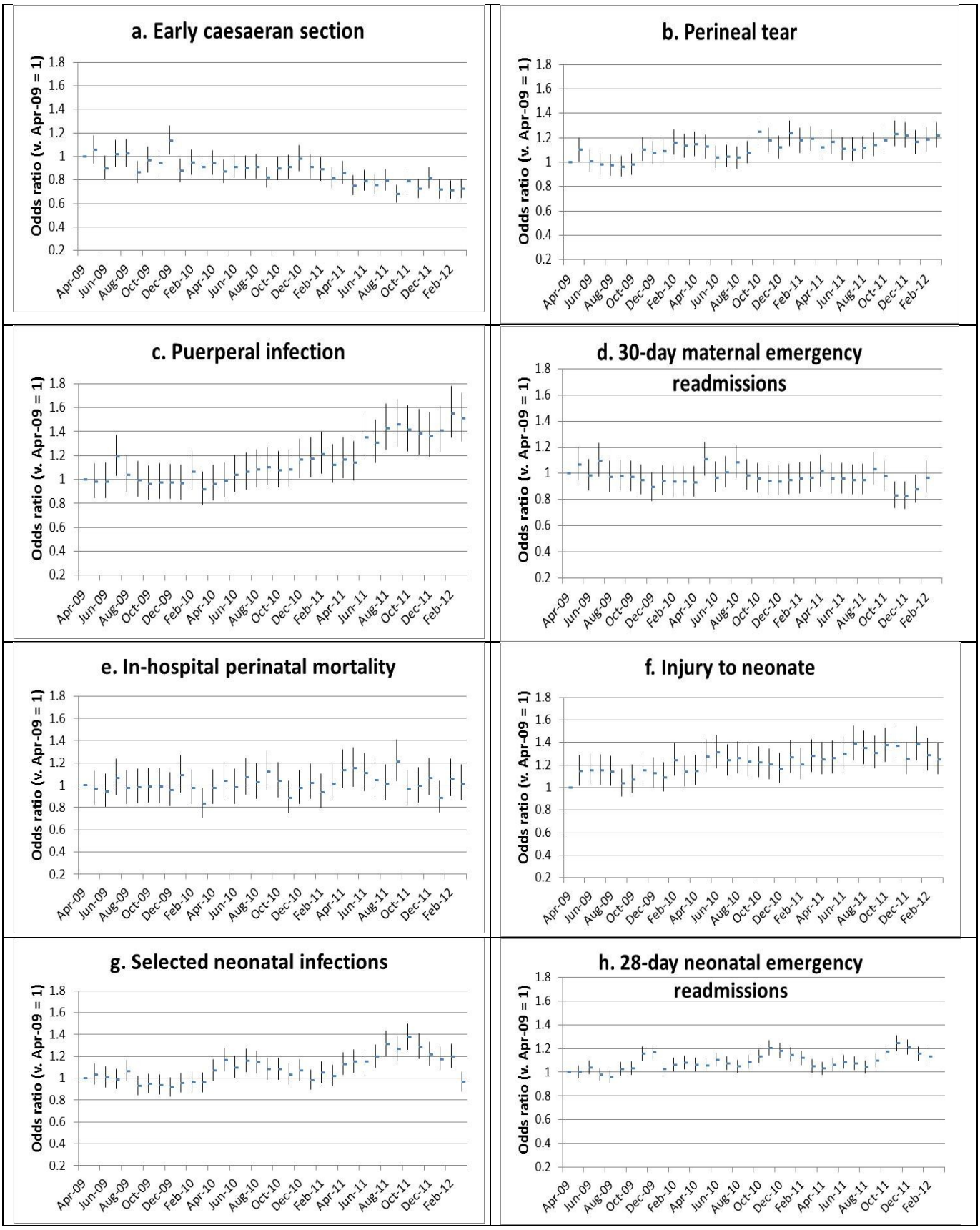
Table 55. Number of births and maternities and complication rates

	Measure	**Total**		2009-10	2010-11	2011-12
		Events, n	%	%	%	%
	Maternities	*1,979,360*	-	*n = 646,525*	*n = 663,218*	*n = 669,617*
	Births	*2,008,144*	-	*n = 658,545*	*n = 670,814*	*n = 678,785*
a	Early caesarean section	49,074	37.50	39.5	38.0	35.2
b	Perineal tear	43,890	2.94	2.73	3.01	3.07
c	Puerperal infection	15,521	0.78	0.68	0.74	0.93
d	30-day maternal emergency readmissions	18,423	1.02	1.04	1.04	0.99
e	In-hospital perinatal mortality	13,402	0.67	0.67	0.66	0.67
f	Injury to neonate	26,005	1.40	1.26	1.41	1.53
g	Selected neonatal infections	38,156	1.91	1.74	1.90	2.08
h	28-day neonatal emergency readmissions	106,799	5.85	5.53	5.98	6.04

Note: Percentage of cases meeting inclusion criteria. Readmission figures exclude final month of data year to allow comparability over years

In the first instance, I plotted the rates per month to better understand what trends might be present (Figure 29). Some of the measures, such as early caesarean sections showed a relatively consistent trend over time and others, for example neonatal infections, showed a seasonal trend.

Figure 29: Trends in obstetric indicators, by month



Note: Data points represent odds ratios, with April 2007 used as a reference (1.00); vertical ranges, 95% confidence intervals.

10.4.1.2. Analysis of longitudinal trends and seasonal effect

The more detail analysis revealed that all indicators showed a statistically significant (at the 95% confidence level) linear longitudinal trend, with six of the eight measures showing a worsening in performance. There were 8 instances of months with statistically different performance to the reference (April) at the 99.8% level, covering early caesarean sections, neonatal infections and neonatal readmissions (Table 56).

Table 56. Longitudinal trend and seasonal effect[1]

	a. Early caesarean section	**b.** Perineal tear	**c.** Puerperal infection	**d.** 30-day maternal emergency readmissions	**e.** In-hospital perinatal mortality	**f.** Injury to neonate	**g.** Selected neonatal infections	**h.** 28-day neonatal emergency readmissions	**No. indicators significantly better[1]**	**No. indicators significantly worse[1]**
Longitudinal trend	*0.990***	*1.004***	*1.014***	*0.998**	*1.003***	*1.007***	*1.009***	*1.003***	***2***	***6***
January	0.975	1.027	1.004	0.953	0.942	1.036	0.899**	1.049**	**1**	**1**
February	1.001	1.032	1.067	0.985	0.927	1.026	0.920*	1.041*	**1**	**1**
March	0.975	1.033	0.977	0.987	0.889*	1.002	0.837**	1.025	**2**	**0**
April	1	1	1	1	1	1	1	1	**-**	**-**
May	0.957	1.033	0.978	1.063	1.011	1.076*	1.039	1.006	**0**	**1**
June	0.948	0.957	1.049	0.99	0.968	1.092*	1.002	1.038*	**0**	**2**
July	0.982	0.946*	1.090*	1.039	1.013	1.089*	1.019	1.003	**1**	**2**
August	1.011	0.940*	1.073	1.026	0.957	1.074*	1.064*	0.975	**1**	**2**
September	0.884**	0.955	1.06	1.024	1.049	1.018	0.98	1.023	**1**	**0**
October	1.003	1.019	1.017	0.997	0.948	1.038	1.008	1.062**	**0**	**1**
November	0.982	1.043	1.001	0.932	0.901*	1.047	0.957	1.141**	**1**	**1**
December	1.125**	1.01	1.004	0.914*	0.939	0.99	0.934*	1.123**	**2**	**2**

Note: 1. Longitudinal trend measured as linear, with positive odds ratio suggesting increasing complication rates. Seasonal trend measured as performance in month in comparison to April.

2: Correlation significant at the 95% level marked with '*'; those at 99.8% level with '**'

10.4.2. Association with day of the week

10.4.2.1. Descriptive overview

The distribution of births is not even across the days of the week (Figure 30). The most common day for giving birth is a Thursday (15% of births and maternities) and the least common is Sunday (12%). Much of the difference is explained by number of elective caesarean sections performed during weekdays.

Figure 30. Proportion of births and maternities and breakdown in number of elective caesarean sections, by day of week

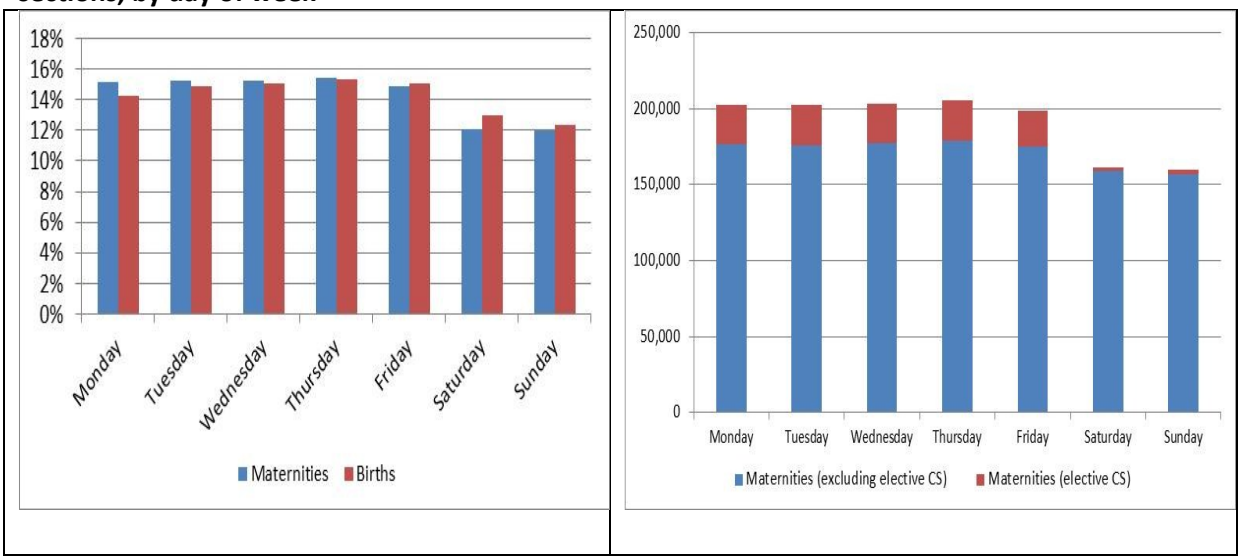


Table 57 gives a description of the characteristics of patients included in the study.

Table 57. Characteristics of the study population

Characteristic	Categories	Weekdays	Weekends	*p-value*
Delivery method	Spontaneous vertex	59.7	68.6	*<0.001*
	Spontaneous, other cephalic	0.5	0.5	
	Low forceps, non-breach	3.9	4.6	
	Other forceps, non-breach	2.1	2.5	
	Ventouse, vacuum extraction	6.1	6.9	
	Breech	0.4	0.4	
	Breech extraction not otherwise specified	0.1	0.1	
	Elective caesarean	12.7	1.5	
	Emergency caesarean	14.6	14.9	
	Other than those specified above	0.1	0.1	
Maternal age	<19	5.1	5.8	

	20-24	18.4	19.9	
	25-29	27.5	28.2	
	30-34	28.7	28.2	
	35-39	16.2	14.6	
	40 and over	4.1	3.3	
	Unknown			
Ethnicity	White	72.3	71.8	*<0.001*
	Asian	10.3	10.5	
	Black (inc. Black British)Afro-Caribbean	5.1	5.0	
	Mixed	1.5	1.5	
	Other (inc. Chinese)	3.3	3.4	
	Unknown / not stated	7.5	8.0	
Level of deprivations	1	15.2	14.9	*<0.001*
	2	16.3	16.0	
	3	19.0	18.9	
	4	22.1	22.3	
	5	26.5	27.0	
	'6'	0.9	0.8	
Parity	Primiparous	43.5	46.4	*<0.001*
	Multiparous	56.5	53.6	
Gestational age	< 37	7.6	7.4	*<0.001*
	37-39	35.7	30.3	
	40-41	39.7	45.2	
	>=42	3.6	4.1	
	Unknown	13.5	13.2	
Birth-weight	<2500g	5.9	5.4	<0.001
	2500-4000g	74.1	74.7	
	>4000g	10.1	10.2	
	Unknown	9.9	9.7	
Multip	Yes	1.6	0.9	*<0.001*
	No	91.0	91.8	
	Unknown	7.4	7.3	
Previous caesarean section		12.8	6.9	*<0.001*

Note: Data are for maternities

10.4.2.1. Association between day of delivery and performance

Table 58 shows the results of the association between weekday/weekend admission and performance in the seven measures of quality and safety. There were statistically significant

associations in four of the indicators, all of which were consistent with a lower standard of care at weekends. The largest effects were seen in the higher rates perinatal mortality rates (adjusted odds ratio [OR] 1.07; 95% Confidence Interval 1.02 – 1.13) and puerperal infections (OR 1.06; 1.01 – 1.11).

Table 58. Association between weekday/weekend admission (or birth) and indicators of quality and safety of care

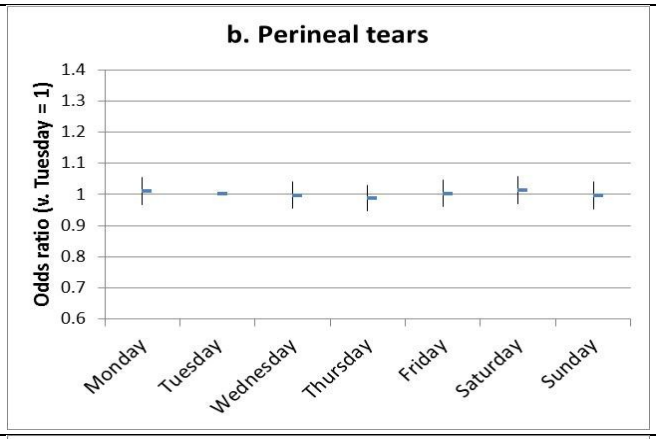
		Weekday admission		Weekend admission			
	Indicator	Unadjusted % (number)	Adjusted rate, %	Unadjusted % (number)	Adjusted rate, %	p-value	OR (95% CI) (weekday as reference)
b	Perineal tear	3.03 (22,299)	3.04	3.07 (8,249)	3.05	0.812	1.003 (0.977 – 1.030)
c	Puerperal infection	0.83 (8,358)	0.82	0.86 (2,770)	0.87	0.010	1.059 (1.014 – 1.107)
d	3-day maternal readmissions	0.20 (1,926)	0.20	0.18 (545)	0.18	0.135	0.929 (0.843 – 1.023)
e	In-hospital perinatal mortality	0.64 (6,481)	0.65	0.73 (2,518)	0.71	0.007	1.071 (1.019 – 1.126)
f	Injury to neonate	1.43 (13,278)	1.45	1.59 (5,038)	1.53	0.002	1.055 (1.020 – 1.091)
g	Selected neonatal infections	1.99 (19,900)	1.99	2.01 (6,852)	2.00	0.542	1.009 (0.980 – 1.038)
h	3-day neonatal readmissions	1.19 (11,323)	1.18	1.22 (3,976)	1.23	0.044	1.039 (1.001 – 1.078)

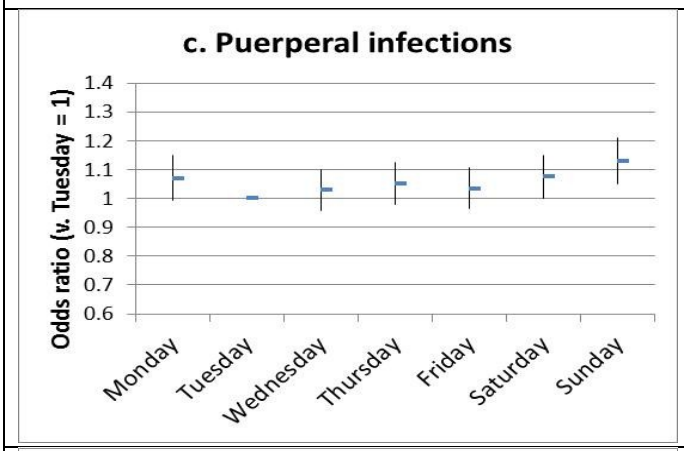
Notes: CI denotes confidence interval.

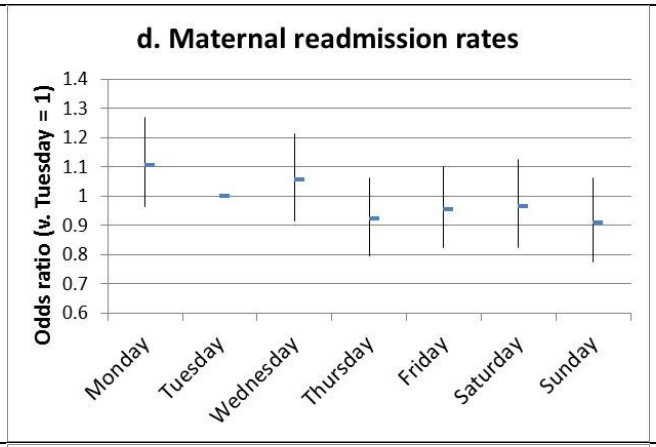
Over the seven indicators, there were 11 examples of statistically significant differences performance of days in comparison to Tuesday. Compared to this reference day, all days bar Wednesday had statistically higher rates of perinatal mortality (Figure 31).

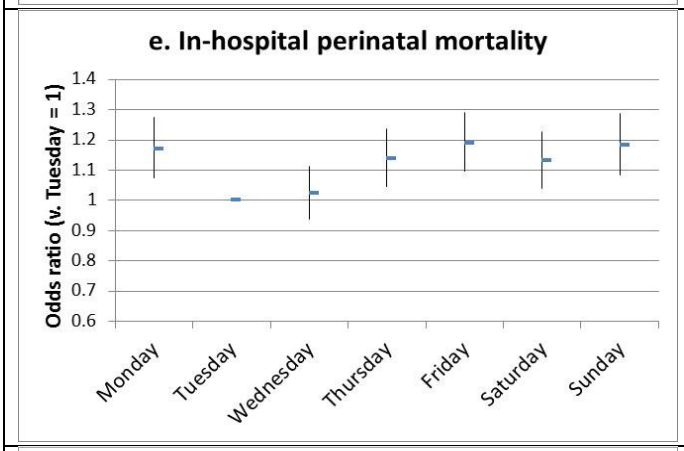
Figure 31: Association between performance and day of admission/birth

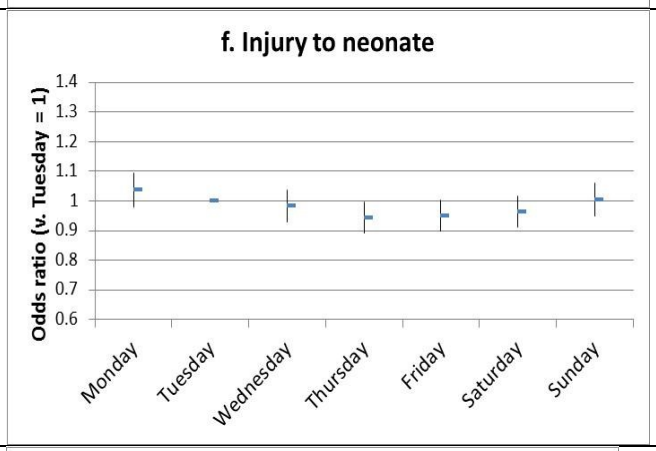
Indicator 'a. Early caesarean sections' not used in this part of the study

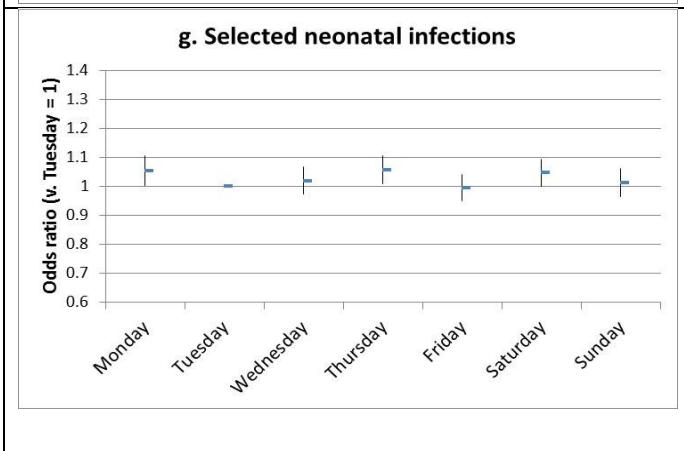


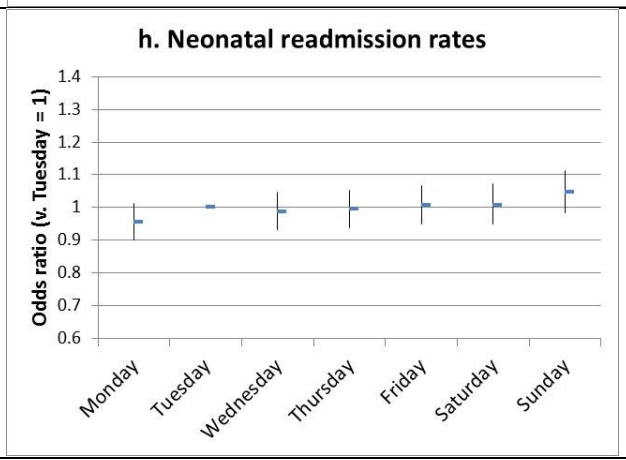












Note: Data points represent odds ratios, with Tuesday used as a reference (1.00); vertical ranges, 95% confidence intervals.

The weekend effect varied by delivery method (Table 59).

Table 59 Perinatal rates by delivery rates for common weekend delivery methods

	Weekday		Weekend		
Delivery method	Unadjusted % (number)	Adjusted rate, %	Unadjusted % (number)	Adjusted rate, %	OR (95% CI) (weekday as reference)
Spontaneous vertex	0.57 (484,998)	0.76	0.61 (189,210)	0.80	1.059 (0.983 – 1.140)
Unknown	1.12 (133,424)	0.64	1.36 (45,527)	0.76	1.202 (1.093 – 1.322)
Emergency caesarean	0.60 (116,280)	0.40	0.58 (42,513)	0.41	1.035 (0.890 – 1.204)
Ventouse, vacuum extraction	0.16 (48,887)	0.34	0.11 (19,161)	0.23	0.676 (0.415 – 1.103)
Spontaneous, other cephalic	0.81 (36,496)	0.96	0.79 (14,094)	0.92	0.946 (0.746 – 1.199)
Low forceps, non-breach	0.20 (29,509)	0.39	0.17 (11,704)	0.33	0.836 (0.501 – 1.397)

Note: Common weekend delivery methods defined as > 5,000 deliveries per year at weekends

10.4.2.1. Association between activity and performance

For sites, high activity times accounted for 9.67% of maternities and 10.09% of births. For trusts, the corresponding figures were 11.10% and 11.37% respectively. The logistic non-hierarchical regression model suggested higher perinatal mortality rates at busy times (Table 60). However, when provider-level effects are added, this raised mortality is no longer statistically significant at the 95% level.

Table 60. Association between business and complication rates, for weekdays April 2010 – March 2012

	Trust level[1] (busy vs not busy)		**Trust level**[1] **with trust effects** (busy vs not busy)		**Unit level with unit level effects**[2] (busy vs not busy)	
Measure	**OR** (95% CI)[3]	p-value	**OR** (95% CI)[3]	p-value	**OR** (95% CI)[3]	p-value
Perineal tear	**0.986** (0.943 – 1.031)	0.55	**0.981** (0.938 - 1.027)	0.41	**0.999** (0.951 - 1.049)	0.98
Puerperal infection	**0.999** (0.930 – 1.073)	0.98	**0.953** (0.887 - 1.026)	0.20	**0.957** (0.884 - 1.036)	0.28
Maternal readmissions	**0.992** (0.929 – 1.060)	0.82	**1.028** (0.933 - 1.131)	0.58	**0.998** (0.902 - 1.105)	0.97
In-hospital perinatal mortality	**1.101** (1.014 – 1.195)	**0.02**	**1.072** (0.985 - 1.166)	0.11	**1.048** (0.956 - 1.148)	0.32
Injury to neonate	**1.022** (0.965 – 1.082)	0.46	**0.957** (0.903 - 1.015)	0.14	**0.948** (0.89 - 1.01)	0.10
Selected neonatal infections	**1.008** (0.962 – 1.056)	0.74	**0.977** (0.93 - 1.025)	0.33	**0.972** (0.923 - 1.025)	0.29
Neonatal readmissions	**1.025** (0.997 – 1.053)	0.08	**1.002** (0.975 - 1.03)	0.90	**1.004** (0.975 - 1.034)	0.79

Notes: 1. Trusts with > 1000 births per year

2. Sites with > 100 births per year

3. OR denotes odds ratio of busy periods (quintile 5) v. other times (quintiles 1 – 4) as reference; CI denotes Confidence interval

4. Accounting for trust/site effects using multi-level model

10.4.2.2. Consultant presence

Within the maternity extract, 51 of the 128 units (39.8%) with a recommended level of consultant presence were compliant. There were statistically significant differences in rates of perineal tears between compliant (2.95%) and non-compliant (3.29%) sites (Table 61). There were no statistically significant differences across the other measures and, likewise, further analysis against the revised

3-day neonatal readmissions (p = 0.83) and 3-day maternal readmissions (p = 0.89) showed no significant effects.

Table 61. Association between compliance with consultant staffing levels and indicators of quality and safety of care

		Unadjusted, % (number)		**Adjusted**	
	Measure	Compliant	Non-compliant	OR (95% CI) (non-compliant vs compliant)[1]	p-value
b	Perineal tear	2.95 (3,916)	3.29 (9,443)	1.205 (1.001 – 1.450)	**0.05**
c	Puerperal infection	0.76 (1,345)	1.07 (4,149)	0.913 (0.535 – 1.557)	0.73
d	30-day maternal readmissions	1.10 (1,779)	0.95 (3,336)	1.006 (0.770 – 1.315)	0.97
e	In-hospital perinatal mortality	0.60 (1,026)	0.75 (3,076)	1.209 (0.852 – 1.716)	0.29
f	Injury to neonate	1.59 (2,513)	1.46 (5,481)	0.829 (0.460 – 1.494)	0.53
g	Selected neonatal infections	2.04 (3,468)	2.28 (9,342)	1.311 (0.789 – 2.177)	0.30
h	28-day neonatal readmissions	6.48(10,022)	5.76 (21,400)	0.936 (0.730 – 1.200)	0.27

Notes: 1. OR denotes odds ratio of compliant v non-compliant performance; CI denotes confidence interval.

10.5. Discussion

10.5.1. Summary of findings on seasonality

We examined over 2 million deliveries across a three-year period. The longitudinal analysis showed that administrative data could be used to identify potentially important trends in obstetric care quality and safety. The analysis suggests that there are associations between month of delivery and outcomes. Notably, there are raised rates of neonatal readmission during the winter months (October to January). Interestingly, there is a substantial increase in early caesarean section rates in December, with one possible hypothesis being that NICE guidelines are not followed due to a lack of available slots to conduct elective caesareans during holiday season. There is no consistently poorer

month but I would suggest aiming for September and March would put the odds in your favour. The analysis provided some evidence to suggest there could be some poorer outcomes in July and August, a hypothesis that was rejected by a previous study on small preterm infants in USA.[507]

10.5.2. Day of week

10.5.2.1. Summary of findings

The study demonstrated that babies born at the weekend have an increased risk being still born or dying in hospital within the first 7 days. Moreover, the study also suggests that other increased complication rates at weekends, with higher rates of puerperal infection, injury to neonate and 3-day neonatal emergency readmissions.

10.5.2.2. Impact: magnitude of weekend effect

I calculated, as described in the methods section, that the study suggests there are some 750 potentially avoidable perinatal deaths each year and an excess of 450 maternal infections if performance seen at the reference day, Tuesday, was replicated on other days.

There are also reasons to suggest that the inequality of care is more pronounced than identified in this study. If the effect is indeed caused by a staff deficiency and a lack of resources, then you would expect poorer quality and safety at all out-of-hours periods during the week, including bank holidays and weekday evenings and nights. If this is the case, then the out-of-hours periods during weekdays are masking some of the effect.

10.5.2.3. Amenable: resource levels

There are several possible explanations for these findings, including a lack of consultant obstetrician presence. This study provided some evidence to support the theory that one of the contributing factors to the 'weekend effect' might be failure to meet recommended levels of consultant presence. The association between consultant availability and performance also provides some evidence that the 'weekend effect' is amenable to the provision of healthcare.

On the analysis of the effect of activity on performance, there were not statistically significant variations, although the raised mortality at busy period might warrant further investigation. While based on a similar approach in the existing literature on measuring the association between levels of activity and outcomes, which concluded that hospitals that operate at or over capacity may experience heightened rates of patient safety events,[508] a more sophisticated indicator for relative activity could be developed further.

10.5.2.4. Precision

The indicators identified a wide range of performance, by day of admission/birth, with statistically significant differences in four of the seven indicators, with further differences identified by disaggregating to day of the week.

10.5.2.5. Minimum bias

HES data do not include information on the time of admission and therefore I was not able to investigate the wider issue of the quality of out-of-hours care in this study. Other limitations should also be noted. Firstly, the administrative database used gives only limited information on the complexity of the delivery. Whilst I used a number of variables, such as Charlson index and number of previous admissions, to mitigate for any bias in case mix between admissions on different days of the week, there is no information on, for example, maternal obesity or smoking (as discussed in paragraph 9.6.4.1, p217). This may be important since a Canadian study identifying higher rates of early neonatal deaths from 1985-99 in weekends (OR 1.11, 95% CI 1.07 – 1.16) found these differences were no longer significant after risk-adjustment (0.96, 0.91 – 1.01),[509] albeit another study found that adjusting for maternal characteristics had no material effect.[510] A further limitation is that the administrative data can contain errors; however, you would not expect this to be different in the coding of weekday or weekend admissions.

I stratified one of the key results – in-hospital mortality – by mode of delivery to further investigate potential bias. For the most common delivery method, spontaneous vertex, the likelihood of perinatal mortality was higher at weekends than for weekdays, although this was not significant at the 95% confidence level (odds ratio: 1.059, 95% confidence interval: 0.983 – 1.140).

10.5.2.6. Construct validity

That many of the indicators identified evidence of an association between performance and day of delivery, and further analyses revealing lower level of specialist staffing also being associated a poorer outcome, provides evidence of construct validity. However, in lieu of having multiple comparable units – as with the cross-sectional analysis in the previous chapter – I have provided some assurance on validity by comparing consistency with previous literature. The results from this study are consistent with some previous results, including:

- Elsewhere in the UK, where a study in Scotland found adjusted weekend neonatal death of 1.3 (1.0 – 1.6, compared to weekday in-hours) which is similar for **all out-of-hours** 1.3 (1.1 – 1.6).[510] Another study from the same country, while also finding the same effect, identified that a primary reason for the out of hours deaths was intrapartum anoxia.[510]

- Further afield, in Australia early neonatal deaths were much higher at weekends (by 29%, p <0.001)[504], and with a similar affect size (27%) found for neonatal deaths for infants born on Sundays.[509] However, other studies, including those from Canada and USA rejected the hypothesis of greater complications at weekends.[507,509]

10.5.2.7. Further work

Further work is needed to understand what organisational factors might influence the 'weekend effect' and to investigate centres that have reduced the disparities in access and outcome in out-of-hours care. A starting point for this is to allow hospitals to compare the extent of the 'weekend effect' in their organisation to that in their peers. Other future work to add to this analysis could include:

- Further sensitivity analysis by removing difficult cases and therefore reducing potential bias in case-mix. For instance, the analysis could be repeated excluding: cases with gestation age either missing or outside 37-43 weeks; and perinatal deaths ascribed to congenital abnormality or rhesus isoimmunisation, stillbirths and non-cephalic deliveries.[510]

- Analysing the time of onset of the complication as well as the time of delivery. For instance, previous work has looked at the date of death as well as the date of birth [505] but there is difficulty in interpreting date of death from antepartum and intrapartum still births.[509]

- Looking at all out-of-hours periods, with inconsistencies in the existing literature on how to account for holiday periods. [445, 503,505]

10.5.3. Conclusion

There is also scope to extend this analysis to other specialties, with similar results having also been found in a limited number of other clinical areas, such as pulmonary embolism, hip fractures and upper gastro-intestinal haemorrhage.[354] A greater understanding of the issue will also require better data, and the inclusion of an out-of-hours admission flag for hospital administrative data should be considered.

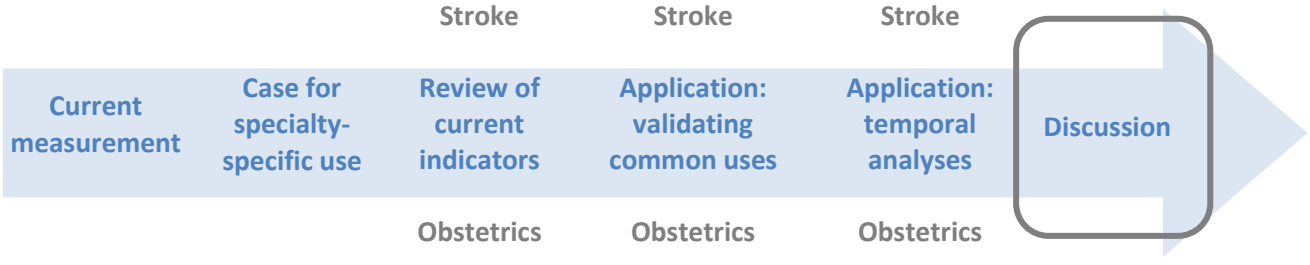
This chapter shows that it is possible to address some of the potential bias from case-mix and coding practice using longitudinal analyses; however, given the changes in coding practice over time which coincides with potential changes in service provision (and therefore hard to detect) this application is still susceptible to bias. The analysis by day of the week addresses most of the coding and risk-adjustment issues raised in the previous chapters and, therefore, permits more robust conclusions in what is, coincidentally, as a very high impact area.

Part IV:

Discussion, recommendations and appendices

Chapter 11.

Discussion



11.1. Summary

'While most people are in favour of measurement,
few are comfortable being measured.' (Loeb 2004, p i6)

11.1.1. Overview

11.1.1.1. Quality measurement revisited

Using administrative data to measure the quality and safety of hospital care offers many opportunities. However, progress has been limited to few countries and predominantly to a small subset of broad measures, such as Hospital Standardised Mortality Rates. In this thesis, I investigate the potential advantages and feasibility – in terms of validity and applicability – of specialty-specific indicators.

Measurement of hospital outcomes is longstanding; at least dating back to the standardised mortality data published by Florence Nightingale in the early 1860s.[511] Yet there remain calls for further development of indicators. The project described here addressed this, in part, through developing two sets of specialty-specific indicators (Chapters 4 and 8). As exemplified in the conclusions to a report on the failed Mid-Staffordshire NHS Foundation Trust, a greater consensus on how they are applied is needed.[512] To address this, I outlined the key implications regarding the current use of indicators and summarised important methodological considerations in Chapters 1 to 3 and, subsequently, I tested and explicitly outlined key assumptions in applying the chosen indicators (Chapters 5 and 9).

11.1.1.2. Structure of this chapter

Given the shortcomings in quality measurement, coupled with the inexorable rise in the use of administrative data, it is important that researchers, practitioners and policy-makers focus on making best use of such datasets and related studies. The implications of the studies described here for researchers (see section 11.3), the collection of the data (11.4), clinicians (11.5) and policy-makers (11.6), along with some more general commentary from existing literature, are all covered in this discussion. These implications are preceded by a summary of the results in relation to the study hypotheses (11.1.2) and an overview of the methodological implications (11.2).

11.1.2. Summary of findings, by hypothesis

11.1.2.1. Hypotheses revisited

The project was designed to test three hypotheses (paragraph 2.7.1, p59) regarding measuring the quality and safety of hospital care using specialty-specific indicators based on routinely collected

administrative data. Two of these hypotheses focused on the advantages of a specialty-specific approach (to create a bundle of indicators and better validate indicators) with the third, more generally, on the value of the temporal nature of administrative data. These propositions (in bold), along with a summary on how they were covered in the studies, are given below.

11.1.2.2. Hypothesis 1: Indicator bundles

A specialty-specific approach to developing indicators can result in a more useful and valid bundle of indicators to identify quality and safety issues. There has been a long-standing recognition that most efforts to measure performance have concentrated on mortality rates[50] and, consequently, there is a need to develop – for the purpose of monitoring hospital performance – additional non-mortality outcomes or process measures.[49] Four particular studies in the thesis supported this hypothesis, with the literature reviews (Chapters 4 and 8) showing high level of sensitivity for identifying indicators, while the experimental studies applying a sample of the indicators (Chapters 5 and 9) highlighted benefits in terms of additional validation opportunities.

The numbers of papers identified in the two literature reviews were also similar to those identified from an exercise to identify the broader category of all patient safety indicators (some 124 papers),[67] giving support to the assertion that the specialty-specific approach to identifying literature has good sensitivity (ability to identify relevant research). This is important since some of the prominent quality and safety initiatives, such as AHRQs patient safety indicators, have been developed using general (non-specialty) literature reviews and, as such, have not been based on a full assessment of existing evidence on, say, validity of existing measures.

It is briefly worth noting that as a rudimentary check, it was reassuring that the magnitude of abstracts identified through the search of electronic databases (1,999 for stroke; 1,670 for obstetrics) and full texts extracted (99 for stroke or 5.0% of abstracts identified; 80 for obstetrics or 4.8% of abstracts) were similar. This provides some evidence to suggest that the sensitivity and specificity of the specialty-specific literature search reviews were reasonable.

This ability to identify a more comprehensive range of indicators meant that I was able to apply broad indicator sets. I identified 20 categories of stroke indicators and 44 for obstetrics, and I applied 6 and 8 indicators, respectively, in the experimental studies. The advantage of applying bundles of indicators means that this project had higher content validity (the extent to which the measurement represents all facets of the area being investigated) than many existing studies identified through the literature reviews, which often applied just a single indicator, such as mortality rates, capturing only a very small proportion of the total possible complications. As health

services continue to improve, the focus on quality and safety is likely to continue to broaden – so moving further away from simple mortality – which lends support for the more holistic approach to measuring performance that I present here.

11.1.2.3. Hypothesis 2: Validation

Hospital-level comparisons could be improved through more robust validation techniques (coding practice, indicator comparison and organisational factors). Through the work, I demonstrated that investigating correlations between indicators, association with organisational factors, and the effect of coding practice, can all provide further assurance on levels of bias, construct validity, and amenability of quality and safety measures. The following results, in particular, are noteworthy:

- I was able to demonstrate that coding depth explained only some of the variation in performance, even across measures that are dependent on the recording of secondary diagnoses (i.e. aspiration pneumonia within stroke patients, injury to neonate and perineal tear) and, therefore, most susceptible to such bias.

- As well as providing more assurance on content validity (as discussed above), the study design – applying a bundle of indicators – had the advantage of allowing the evaluation of construct validity by comparing hospitals' performance across all the measures. The vast majority of existing literature has failed to take advantage of this validation method, even when more than one indicator is applied. I was able to identify interesting associations; for example, trusts with higher scanning rates have lower stroke mortality ($p < 0.05$) and, for obstetrics, sites with higher maternal readmission rates also had higher levels of perinatal mortality ($p < 0.001$).

- There was some evidence to suggest that staffing is important to quality and safety. In particular, I found that lower midwifery staffing levels (i.e. higher birth-to-midwife ratios) were associated with higher perinatal mortality, and lower consultant presence within delivery wards associated with higher rates of perineal tears (both at $p = 0.05$).

While I showed that other applications (i.e. not benchmarking hospitals) hold advantages, as discussed in the following paragraphs, the investigatory work on validity of hospital-level comparisons is important since this remains a dominant use of these indicators.

11.1.2.4. Hypothesis 3: Temporal nature

The temporal nature of administrative data can be used to overcome some of the limitations of such indicators and highlight important quality and safety issues. Three studies presented here

(Chapters 6, 7, and 10) demonstrated how significant issues in healthcare can be identified using the longitudinal nature of the dataset in a robust way. This application of the indicators also has the advantage of controlling for the variations in coding practice over time and between hospitals. The evaluation of the association between day of admission (or delivery, for obstetric measures) and complication rates was the first comprehensive assessments of this issue in England for these two specialties by applying multiple indicators so having higher content validity.

As well as controlling for some of the possible bias, the temporal application also suggested significant shortcomings in care. For example, 7-day stroke in-hospital mortality was 10.6% at weekends compared to 8.9% at weekdays (adjusted OR 1.178, 95% CI 1.120 – 1.240). Similarly, the perinatal mortality rate was 7.3 per 1,000 at weekends, some 0.9 per 1,000 higher than at weekdays (1.071, 1.019 – 1.126). By matching cases between out-of-hours and in-hours cases, I was also able to estimate the scale of avoidable harm. For instance, in stroke care, the study suggested there are, each year, some 250 potentially avoidable in-hospital deaths within 7 days and an additional 650 people could be discharged to their usual place of residence within 56 days if the performance seen at weekdays was replicated at weekends.

I was also able to use the temporal nature of administrative data to evaluate the existence of seasonal trends in obstetric care and the effect of a regional reform of stroke services. In both cases, I found potentially important differences in the quality and safety of care with, for example: a significant reduction in 7-day stroke mortality within the region that had undergone a service restructuring, in comparison to other areas; and raised rates of neonatal readmissions during winter months.

11.2. Methodological issues

11.2.1.1. Introduction to study limitations

Methodological issues related to the specific studies were raised in the previous chapters' discussions. However, some general limitations about the approach taken are worth noting when considering the implication of the work as a whole.

11.2.2. Limitations to development of indicator sets

11.2.2.1. Addressing limitations in review process

The limitations of this review process include the potential bias introduced from the use of a single reviewer. No formal meta-analysis could be undertaken because of the heterogeneity in study methodologies and underlying data frameworks and I did not use a template to assess the quality of

the papers since there was a very diverse range of purposes across the studies. Despite the limitations, the literature reviews shared the features of a systematic, rather than narrative, review (paragraph 3.4.2.3, p72). However, there could be benefits to future research in using a more resource-intensive review process which, at least, could improve face validity of results.

11.2.2.2. Selection of indicators

One limitation of the framework was in the bias towards measuring what has previously been measured. Whilst having peer-reviewed articles on particular indicators may support face validity, it may bias future work towards such indicators. There would be value in complementing such literature reviews with further exercises to develop indicators in areas which are not covered by the existing measures.

The design of the study was, in the first instance, based on a prospective analytical framework, and supplemented reactively – formed on the basis of the gaps in the literature. I used a range of experts, from coding specialists to medical consultants to help develop the indicator set; however, I did not use any formal approach to reach consensus. This could be considered in future research; for instance, if there are sufficient resources available an approach such as RAND's Delphi method could be applied.

11.2.3. Indicator validation

11.2.3.1. Lack of 'gold standard'

Ensuring that indicators are not just easy to apply, but also have appropriate validity, is important: "something that's measurable may not be worth measuring, and maybe you can't measure the things that are worth measuring. What damage do you do by releasing information just because you can measure it?"(GP quote, p 1280) [513] A review, published in 1998, by Fitzpatrick and colleagues outlined the key aspects of validity of patient-based outcome measures albeit in the context of use in clinical trials.[142] The approach used in this study tallies well against this and other such frameworks. However, many aspects and tests of validity (e.g. concurrent, convergent and discriminative) depend on the existence of a 'gold standard' to provide a basis for comparison; if no gold standard exists, they represent a form of construct validity in which the relationship to another measure is hypothesised. It is, therefore, an inherent weakness that no absolute, incontestable statement on validity of indicators can be made here.

11.2.3.2. Minimum bias

This study provided the most comprehensive assessment to date on the effect of coding practice. While the analysis on both specialties suggested that key issues around variations in coding practice

only partially explain variation in performance against the indicators. However this was not comprehensive, with other potential variations in coding practice remaining untested.

The application showed that you can adjust for the measured case-mix factors by building converging regression models. However, the extent to whether this is sufficient remains debated with some suggesting that good risk adjustment is possible[514] whereas others, for instance, highlighted that there is bias in the coding of medical records which has implications on the administrative data derived from it.[106] The lack of some key clinical factors within administrative data is an inherent limitation when comparing units with different case-mixes.

11.2.4. Limitations to findings on shortcomings in care

11.2.4.1. Reference day

As highlighted in some of the exploratory work on stroke care trends (paragraph 6.5.2.2, p146), the choice of reference period when comparing performance has a significant effect on the derived odds ratios. One inconsistency in my approach to the stroke and obstetric out-of-hours analysis should be noted. For the latter I used Tuesday (rather than Monday, as for the stroke analysis) as the reference day. This decision was made a priori as, on Mondays, large numbers of babies are born to women who have received care during the weekend and, as such, this day does not represent a typical in-hours day. Ideally, the analysis would be able to disaggregate between all in-hours and out-of-hours periods but the data used does not include a time field (rather only date) and this is covered later (11.4.2.6, p260).

11.3. Future research

11.3.1.1. Introduction to future research

Some specific areas of future research are covered in early chapter discussions. A few notable areas are outlined here, covering: the development of indicator sets within other specialties; areas for further research within stroke and obstetric specialties; the development of reporting standards for similar research; and an outline of which study designs for such indicators might be fruitful.

11.3.2. Developing indicator sets for other specialties

11.3.2.1. Introduction

An obvious area for expanding this project is to transfer the approach for identifying indicators to other specialties. This section covers why the specialty-specific approach was worthwhile, then sets out a framework for considering what specialties to choose for similar interrogation, and finishes by setting out the importance of publishing the resultant indicator set.

11.3.2.2. Advantage of the specialty-specific approach

The specialty-specific reviews proved useful to developing an indicator set. For both specialties, the increase in research has been similar (50% per annum for stroke, 60% for obstetrics) with few studies prior to 1990 (as expected since few large administrative datasets were available before then, as highlighted in paragraph 2.1.1.2, p39. This increase is likely to reflect both the growth in agreed indicator sets, which even if not directly relevant to the specialties supports a culture of using administrative data to monitor performance, and increasing access to these databases. Both specialties yielded a substantial number of indicators (as discussed below), with many of these appearing regularly in a range of the literature, including peer-reviewed articles. The implication is that future research on developing indicator sets to measure performance should not rely solely on generic searches of indicators and instead consider using speciality-specific terms.

Similarly, I found there are advantages of looking at more than one specialty. For example, I was able to learn from different maturation of each specialty's research, and applied a technique for evaluating coding practice that had been used in one of the specialties (stroke) to the other specialty (obstetrics).[263] I focused on two specialties, both within acute care, and while this scope was shown to be appropriate – with proven importance and difference between specialties – not all lessons will be generalisable. The implication is that there is value in expanding the application of indicators to other specialties.

11.3.2.3. Choosing other specialties

Making the assertion that there is value in expanding this approach to other specialties also raises the question of which would be the most fertile area to expand into. Table 62 shows that comparing the two specialities does not give a clear direction on whether one specialty outperformed the other as a subject for such reviews. The high number of obstetrics indicators could suggest high content validity by representing more facets of the subject of evaluation. Yet for the obstetrics study, indicators are limited to perinatal care whereas the stroke analysis included indicators across the pathway. For the former, this suggests reasonable content validity if perinatal care is the subject but not necessarily the case if all obstetric care. While not giving a prioritisation for which specialties to apply this approach to next, this work does provide a framework to compare facets of specialties (Table 9, p69) and shows that relatively disparate specialties can both provide interesting insight though the literature-review process. That said, development should be directed to account for both clinical importance (e.g. impact and amenability) and methodological feasibility (e.g. scientific soundness and applicability).

Table 62. Systematic review findings in relation to analytical framework

	Domain	Both	Stroke	Obstetrics
Importance	1. Impact	(+) Clearly articulated policy importance and known issue		
	2. Amenable		(+) Evidence of effect of organisational differences	(+) More benchmarks of acceptable performance
Scientific soundness of indicators	3. Face and content validity	(+) Range of (peer-reviewed) indicators	(-) Fewer measures	(-) Indicators focus only on perinatal part of the pathway
	4. Precision	(+) Known variation		
	5. Construct validity	(-) No "gold standard" (-) Few bundles used	(+) More process indicators to supplement outcome measures	(+) More indicators focused on the same construct (i.e. more opportunity for convergent validation)
	6. Minimum bias	(+) Model fit often described (-) No consistency or transparency in risk model (+) Many indicators applied to administrative data		
Applicability	7. Data availability	(+) Examples of administrative data being applied (-) Few studies on ICD-10 data		
	8. Reporting burden	(+) Many examples of previous reporting		
	9. External and ecological validity	(-) Predominantly based on USA data and ICD-9 coding framework (-) Lack of transparency		

Notes: The positives (+) for one of the specialties can be interpreted as negatives (-) for the other specialty in this comparison. These converse arguments are not presented for brevity.

11.3.2.4. Publishing indicator sets

The literature reviews also provide some evidence to suggest that areas with published indicator sets have benefited accordingly from increased volume of, consistency across, and evaluation within, research. Whilst comparisons between the numbers should be treated with caution, a crude cross-check of the categories of indicators identifies more within obstetrics (42) than in stroke (20). The difference may be merely the natural result of the coding frameworks being more appropriate for identifying obstetrics complications and the fact that for this specialty there are two sets of data (the mother and baby records). However, this could also reflect the influence of the widely recognised indicator sets, with a number of obstetrics indicators appearing, for instance, the AHRQ PSIs. This latter hypothesis is supported by the fact that there was more regularity in the use of indicators in these sets even outside their country of creation. If true, the implication is that developing and publishing indicator sets can create a culture of measurement so proliferating indicator numbers and

improve the use of administrative data, increasing the amount and consistency of research and, resultantly, allowing better benchmarking of performance (indeed, I found that there were more published indicative benchmarks for acceptable performance against the obstetrics indicators than for the stroke research).

11.3.3. Further research within these specialties

11.3.3.1. Introduction to further research within stroke and obstetrics care

The stroke and obstetrics studies highlighted some areas for further research within these specialties. In particular, within the two specialties, there would be value in: conducting further work to improve the indicators; applying the indicator set used here in other countries; and better understanding some of the associations between organisational factors and performance.

11.3.3.2. Improving the stroke and obstetrics indicators

As stated above, the studies in this thesis suggested need for further research on validity, which is echoed in the existing literature. The utilisation of such measures – whether, for example, for internal benchmarking by hospitals or for regulation – must be in proportion to the confidence over the validity of the individual indicators. So, as well as tempering current reporting to the present validity of indicators, efforts should also be made to improve the validity of the indicators. Indeed, in addition to developing new indicators (as per section 11.3.2), there have been calls to further develop those that already exist. In fact, the WHO stated that "existing measures should be refined and further validated."(p11)[515]

In the editorial introduction to the first major set of quality indicators – CSP, Iezzoni et al (1992) – the commentator states that they "believe the algorithms will prove difficult to improve upon."(p361)[50] History has proved this statement to be wrong with incremental improvements in the approaches to using administrative data (as well as some retrograde steps) having been made, and this monograph continues the important development. In terms of validating the indicators applied to the data, again the existing literature across the two specialties was limited with the most common techniques using different risk-adjustment models and comparing results to other sources. Further work should continue across all the aspects of validity highlighted in this study.

11.3.3.3. Application of the two indicator sets to other healthcare services

For both specialties, the majority of studies were based on data from the USA and using ICD-9 data, which is consistent with previous literature on measures derived from routinely-collected datasets.[137] The geographical coverage of the obstetrics' studies was greater than in for the stroke research; however, this may reflect differences in the search strategies, with the obstetrics review

able to include foreign language articles. Some of the indicators and findings presented here might be specific to the NHS and its datasets and so further research should focus on less well-researched health services.

11.3.3.4. Addressing specific research gaps

Across stroke and obstetrics research, many studies have focused on the association between performance against the indicators and different hospital characteristics. However, while stroke indicators were used to evaluate different service types, such as the use of stroke units, no studies based on administrative data had looked explicitly at different modalities for obstetrics delivery. This may reflect homogeneity in the organisation of services in some countries, although this explanation does not hold everywhere with, for instance, the proliferation in midwifery units – as an alternative to hospital obstetrics units – in England. The challenge for such research – as highlighted in particular on the discussion on evaluating stroke services in London and effect of size of stroke units – is to fully recognise that different modalities might have heterogeneous case-mixes and not all such factors might be measurable within the data.

More generally, the research presented in this thesis found some evidence to suggest performance is associated with staffing levels. Further work is required to confirm these relationships, including an exploration of the cause.

11.3.4. Developing reporting standards

11.3.4.1. Transparency as a methodological strength of this study

Citing possible bias due to differences in hospital-level coding has become almost a default, generic limitation to include when writing up research on the use of administrative data to compare between providers performance. However, this project aimed to quantify this potential bias. More widely, this study benefited from being explicit about assumptions, methods and model fit.

In comparison, across the existing literature, there was a similar lack of consistency and transparency in the strategies used for case ascertainment for both the obstetrics and stroke research. Likewise there were a plethora of strategies used for accounting of potential patient confounders, ranging from risk-adjustment models including only age to others combining diagnosis, procedure, demographic and case-mix factors.

11.3.4.2. Recommendation to develop reporting standards

Given the lack of consistency in the application and reporting of studies using administrative data, a natural conclusion would be to introduce a checklist or similar process to standardise the research

and, therefore, improve interpretability and generalisability. Indeed, in 2002 a study by the Agency for Healthcare Research and Quality called for more work on identifying and resolving quality rating issues for observational studies.[162] Generic existing scoring system on quality of studies – such as Grading of Recommendations, Assessment, Development and Evaluations (GRADE) [512] – lack the specific requirements needed to add transparency to studies using administrative data for which non-randomness places an importance on risk-adjustment techniques. In fact, even checklists on observational studies such as Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines fail in this same respect.[516] Such research might start with the most recent version of the Cochrane Handbook which contains guidance on dealing with non-randomised studies in systematic reviews of interventions, from the protocol to synthesis stages.[517] In particular, such reporting should include details on: data fields interrogated; diagnosis and procedure codes used in the indicator algorithms; the case-mix model fit; and validation techniques applied.

11.3.5. Study designs

11.3.5.1. Introduction

I applied the indicators using a range of different experimental designs to evaluate the specialty-specific indicators based on routinely-collected hospital administrative data. In all these applications, and across both specialties, the studies revealed that the area was important (e.g. through estimating potentially avoidable deaths) and that performance could be discriminated (e.g. the high numbers of statistical outliers); however, each design has inherent the strengths and weaknesses as summarised below (Table 63, with a more comprehensive assessment of performance of each approach against the analytical framework given in the respective chapters). A few key study designs which could be fruitful areas for further research are described in more detail below.

Table 63. Summary of strengths and weaknesses of the hospital and patient-level approaches

	Hospital-level analysis	Longitudinal	Regional reform	Day of week
Study design	Cross-sectional	Cross-sectional	Quasi-experimental	Cohort
Key strengths	Discriminates even after accounting for some coding variations	Hospital-level trends will control for some hospital level influences on coding	Ability to compare with both historical performance and that of control group	Unlikely to have variation in coding by day of the week
Key limitations	a. Bias from unaccounted variations in coding by hospitals b. Bias within regional networks with added heterogeneity of patients between hospitals	a. Evidence that coding practice is changing over time b. Variation in case-mix over time	a. Case-mix changes due to reorganisation b. Uncertainty of counter-factual	a. Possibility of unmeasured variation in case-mix by day of the week

11.3.5.2. Epidemiological studies

There is a longstanding understanding that the patient factors (e.g. co-morbidity and age) and hospital factors (e.g. coding and numbers of caesareans performed) associated with adverse events are important.[47] Yet epidemiological studies are difficult, with challenges to interpret findings when there are biological explanations for risk factors affecting outcomes and, specifically on using administrative data, there are weaknesses with, for instance, poor recording of some key risk factors.

A literature review conducted in 2003 concluded that analysis of administrative data was not relevant to the purpose of understanding the latent causes and contributory factors behind adverse events.[59] However, since this review a number of patient factors have been analysed using administrative data, including significantly higher rates for some adverse events for: children;[24] schizophrenic patients;[132] certain races and ethnicities;[437] and females.[120] Similarly, significant associations with hospital factors have been identified, including lower rates of some adverse events

for providers with: more sophisticated and mature IT infrastructures performed significantly better on the largest number of PSIs;[123] access to critical care;[124] and holding daily ward rounds.[518] For better understanding of the risks and to permit future benchmarking, further work should focus on gaining better understanding of hospital factors.[519]

11.3.5.3. Quasi-experimental

Such indicators offer the potential for undertaking trend analysis.[62] As a result, it is possible to assess whether an initiative, such as implementing an IT solution,[122,123] has had an effect on patient safety. And there are examples which show that rates of adverse events can be influenced by initiatives; one project found that improved staffing reduced errors in a 24 hour paediatric critical care satellite pharmacy with unit dose drug distribution.[14] However, such positive findings might be the result of the Hawthorne (observer) effect whereby people modify their behaviour when new initiatives are introduced. A weakness of such quasi-experimental approach is that coding changes over time, thus making disaggregating underlying effects more difficult.

Despite limitations in such study designs, researchers should take advantage of the longitudinal and comprehensive data which allow for time- and cohort-based comparisons. This thesis has already presented examples of evaluating interventions, such as those existing studies on implementing stroke unit care and the original work undertaken on investigating London's regional reorganisation. However, there are many other techniques – such as synthetic control matching – and numerous other possible foci for further studies.

11.4. Data implications

11.4.1. Background

11.4.1.1. Study implications on data

While the studies here demonstrated that key coding issues could only explain some of the variation in performance, it did highlight the extent of variation in coding practice. For example, providers' average number of distinct diagnoses codes per admission ranged from 5.0 to 10.7 within stroke care (by hospital trust) and from 2.8 to 6.6 for maternal admissions (by hospital site). Similarly, the use of the 'stroke unspecified' diagnosis code ranged from 0.2 to 42.6% of admissions across hospital trusts. Some of this variation may be explained by variations in patient case-mix but the huge differences suggest that there are also differences in coding practice.

11.4.1.2. Framework for improving data

There is a need to gain understanding of why data errors are created and, to this end, Benin and colleagues have provided a useful framework by further disaggregating the potential errors as: categorising error (non-recordable differences in case-mix); entry error (improper inputting of the data); and query-error (misunderstanding of the data schema).[315] Further work is required to ensure coding within HES is sufficiently accurate, including improved and more accessible coding guidance, and mandatory recording of specific procedures and diagnoses. Where discrepancies are found in the recording of data, there is a need for further research for both quality and safety auditors and regulators (to ensure performance is known) and hospitals (given the implications on their remuneration regarding, for instance, the coding of thrombolysis and scanning).

11.4.2. Recommended interventions

11.4.2.1. Peer review of coding

To improve consistency in application of the coding rules, there should be a degree of peer review on practice. Given the changes in the governance arrangements around the quality of data given the forthcoming abolishment of the Audit Commission (who previously had a role in auditing this data) such a peer review process could be incorporated into any new data validation process.

11.4.2.2. Adjust the coding framework

The Government has proposed that the coding framework for administrative data may be updated.[191] This study suggests that, if administrative data are going to be used for monitoring healthcare performance, consideration should be given to ensuring the framework allows for the recording of a wider range of quality and safety issues.

11.4.2.3. Reassess coding rules

There should be a review of whether some of the coding rules, such as the guidance to not record procedures undertaken before the decision to admit has been taken, should be amended. In particular, this current rule could introduce some bias where hospitals have differing procedures for admitting patients and might also result in an underestimate in, for example, scanning rates.

11.4.2.4. Present on admission

A commonly cited issue with using administrative data, is attribution with, for instance, potential errors when adverse events are present on admission.[520] One solution is to enrich the administrative data, such as adjusting for conditions present on admission can be useful.[101] The identification of diagnoses that are present on admission can be addressed through linking to other datasets, such as laboratory results. This endeavour of linking datasets also has the advantage of potentially allowing

a greater number and detail of indicators, so improving many aspects of validity. Alternatively, a present-on-admission flag could be introduced as used in some systems in the USA and Australia.

11.4.2.5. Do not resuscitate

As with 'present on admission', there are potential advantages in being able to flag up, within the data, patients whose status is Do Not Resuscitate (DNR).[520] For instance, a previous publication suggested that removing DNR cases from the denominator might improve indicator performance.[149] However, this particular amendment to the coding framework should be done with consideration to both the fact that a DNR flag could reflect an adverse outcomes from poor care (and so such cases shouldn't be excluded from quality measures) and the risk of incentivising gaming.

11.4.2.6. Out-of-hours identifier

The research I conducted in both specialties produced findings that were consistent with poorer standard of care at weekends. However, the true extent of inequality in outcomes by time of admission might be masked by a lack of time-, rather than just day-, stamp. Such a flag should be considered since it would allow researchers to identify the effect of different access to physical resources and staff during out-of-hours periods.

11.5. Clinical implications

11.5.1. Clinical guidelines and quality improvement

For most of the measures applied in this thesis, there is no clinical consensus or guidelines for what are acceptable levels. In some instances this may be appropriate but in others this information is potentially useful to clinical teams and services because it supports the monitoring of performance. The two exceptions (where measures had consensus/guidelines) – with both coincidentally having deficiencies in performance – are:

- access to a scan for people who have had a stroke, for which extant guidelines recommend that all such patients should receive brain imaging "within a maximum of 24 hours after onset of symptoms".[334] The actual performance is that 69.7% of patients receiving a scan within one day of admission.

- on caesarean sections prior to 39 weeks, for which the extant guidelines state that "the risk of respiratory morbidity is increased in babies born by CS before labour, but this risk decreases significantly after 39 weeks [and] therefore planned CS should not routinely be carried out before 39 weeks"(p18).[490] The current rate is around a third of elective caesarean sections being carried out before 39 weeks. While some of this apparent

shortcoming may be due to poor coding, the seasonal differences (with particularly high rates of elective caesarean sections during December) suggest there are clinical shortcomings.

These two cases, with large proportions of people being treated in contravention to the clinical guidelines, seem unacceptable. The implication is that the clinical guidelines need to be revisited, checked to ensure they are deliverable and appropriate, then re-communicated.

11.5.2. Advantages of reporting to a 'specialty team'

11.5.2.1. Introduction

The studies clearly identified findings of clinical importance and, therefore, the implication is that clinicians need to have these findings reported to them.

Health services should consider how these specialty-specific indicators could be reported routinely and directly to the specialty team. This reporting style would represent a hybrid of the administrative data identification method and 'morbidity and mortality conference reporting' format (where the clinical team formally discuss adverse events) and, in doing so, could harness the benefits of both. The motivation for reporting to clinicians is implicit in the arguments set out for reporting to specialties directly (section 2.6, p54). Further to these, the literature does set out further motivation for feeding back the results from indicators directly to specialty teams, as summarised below. Most refer to clinicians but hospital managers or even coders in that specialty should not be overlooked for risk of entrenching any interdisciplinary divide.

11.5.2.2. Learning and culture

Huber's review of organisational learning research makes a distinction between experiential learning and "learning from searching and noticing".[521] The former type of learning is the category under which clinicians learn from their mistakes,[73,117] whilst the second category would include the use of indicators to help clinicians prospectively improve through understanding the risk factors associated with patient safety.[522] Certainly, there is recognition that reducing adverse events requires in-depth involvement of clinicians. Leggat and colleagues (2008), for instance, point out the importance of the skill and capacity of healthcare workers in delivering safer patient care.[135] Similarly, one of the keys to improving patient care is to ensure that the health workforce develops the will to address the issues,[135] and clinicians are well placed to change cultures.[137]

11.5.2.3. Develop new and existing indicators

The specialty team members are in a good position to propose indicators. The increased demand for indicators that has precipitated from the raised profile of quality and patient safety could be addressed, at least in part, by developing and validating more indicators.[73,523] Further to meeting the need for additional information, developing additional indicators would also:

- reduce the focus on particular diagnoses and whether they are coded. This would diminish any ability and incentive to 'game' (as discussed in §11.6.1.2) by not coding accurately or comprehensively;
- would create the opportunity to test correlations and identify factors from among indicators. This in turn would enable the indicators to depict quality and safety on a number of distinct dimensions;[73] and
- reduce reliance on existing research, which I highlighted earlier as a methodological issue.

The specialty team can also focus new indicators on areas where they believe there is scope for improvement and so meet calls that measurement tools should focus on whether any harm caused was preventable.[71]

11.6. Policy implications

11.6.1. Variations in performance

11.6.1.1. Scale of variation

There were large variations between providers in their performance across the indicators. For example, same-day scanning rates varied from 20.4% to 79.3% between hospital trusts and early caesarean section rates varied from 1.1% to 60.7%. Even after accounting for random chance, there was a strong indication of substantial underlying differences. There were 181 occurrences of hospital trusts performing statistically significantly differently to the average at the 99.8% confidence level across the 6 stroke measures. For the 8 obstetrics indicators there were 441 such outliers. Moreover, differences in coding practice appeared to only partially explain the variation in performance.

11.6.1.2. Use of benchmarking

These findings raise questions on how such research should be used. Benchmarking – or comparing performance against peers – can be done either confidentially or publicly. There are cases of specific indicators being used for public reporting of hospital performance, by benchmarking against similar

providers. In the NHS, a number of indicators of quality and safety, derived from administrative data, are now used to benchmark providers and commissioners of care and indeed the publication of results from this study builds on this. However, a number of problems associated with this possible utilisation have been pointed out, including that:

- if indicators are used to evaluate individual hospitals and in the absence of an effective system for auditing the recording of adverse events in the administrative data, there is a risk that hospitals may stop coding incidents that could implicate the occurrence of a patient safety event;[71] and
- due to the limitations on validity, a report could show an unduly poor rate and, if not confidential, a hospital might be pressurised into making a knee-jerk reaction to correct this reported problem.[119]

Given these unresolved issues, inter-institutional comparisons should be "approached with caution" (p788),[101] although Marshall and Romano argue that the concept of publicly available data is becoming more acceptable to both clinicians and managers.[513] A study based in the NHS, which involved sending out benchmarking reports to 170 hospital trusts, reported, however, that "trusts were generally positive about receiving the reports"(p13).[112] The pressure to publish benchmarking data to support policies such as choice of provider, which exists in many health services, means that benchmarking is likely to remain a common application and, as such, I committed a large proportion of my research into investigating some of the issues around this application. As discussed earlier, this study suggests that some of the commonly cited limitations of benchmarking – such as coding depth – can be addressed to some extent.

11.6.1.3. Acceptability of variation

The providers of health services in England are increasing autonomous, although there remains a quality regulator (the Care Quality Commission, CQC) with responsibility for registering, monitoring and inspecting providers. Some level of variation may be inevitable but it is unclear what level policy makers deem acceptable. The scale of the variation is sufficient to suggest that providers and the CQC should consider how to address some of the more significant variations. One action could be to use such indicators to attempt to reduce such variations and this is covered in the next section.

11.6.2. Addressing issue on association between complications and day of the week

11.6.2.1. Scale and causation

This project identified significant variations in out-of-hours performance, as outlined earlier in the discussion. The stroke analysis benefited from reporting both process and outcome measures for this weekend analysis which provides greater construct validity and also from certain process measures being less susceptible to bias from case-mix. As such, policy makers should ensure that further studies on out-of-hours care should endeavour to include process measures. Further research should also continue to look at this issue through a specialty-specific lens, with previous research suggesting a heterogeneous effect across specialties, which could result in significant shortcomings of care being masked by a broader scope.[524]

11.6.2.2. Addressing the 'weekend effect'

In England, in response to findings on quality of care at weekends, a number of initiatives have been developed to reduce mortality, increase efficiency, ease access, and ensure patients receive the same standard of care regardless of the day of the week. Further research suggested that achieving a seven day service would: require local solutions; cost up to 2% of total patient care income; and unlikely to be cost-neutral.[525] In a resource constrained health service, this equation has clear policy implications, requiring health service leaders to evaluate how much value they place on providing an equal level of care at all times.

11.6.3. Application of indicators by policy makers

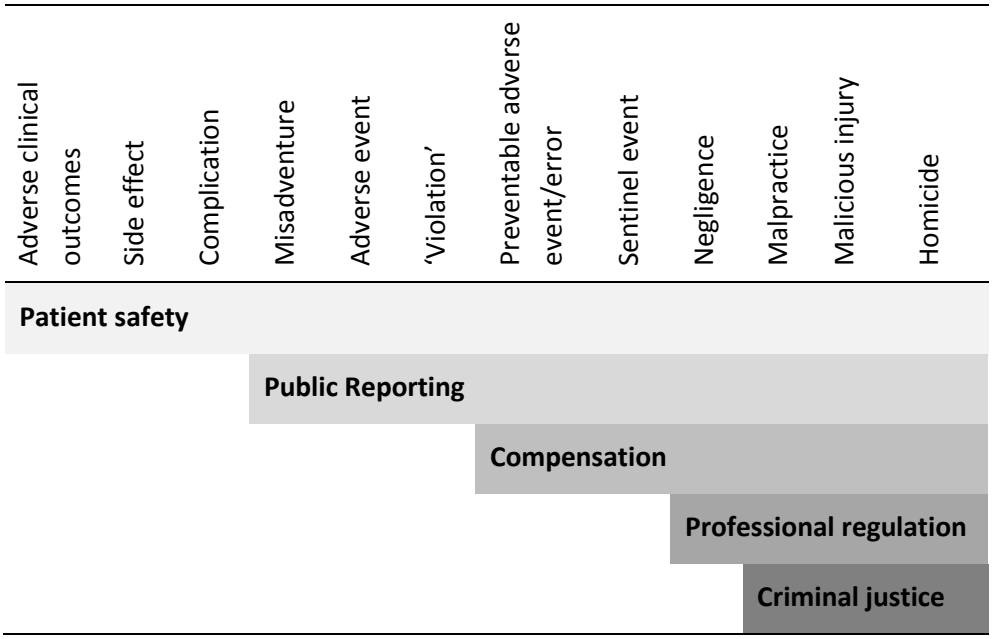
11.6.3.1. Scope of discussion on application of indicators

The previous ten chapters have set out the rationale, methods and considerations relating to the application of specialty-specific indicators. Coupled with the proliferation of indicator use, it would therefore be remiss not to set out some key issues regarding the use of these indicators to improve quality and safety. The work represents only a narrative summary of the multitude of considerations and views expressed elsewhere by a range of disciplines, including that of Benn and colleagues which draw on psychology,[126] but aims to give an overview of the considerations for policy-makers regarding how such indicators could be applied. The similar discussion for researchers – on experimental design (e.g. quasi-experimental or epidemiological) – was covered earlier.

11.6.3.2. Revisiting reporting

The application and interpretation of indicators based on administrative data depends on a multitude of often dependent factors, such as the purpose and type of event being identified (Figure 32).

Figure 32: Potential uses of patient safety indicators



Source: Based on Australian Centre for Economic Research on Health (ACERH)

Of these broad categories of purposes, the following specific examples of applications have been suggested: document the quality of care; make comparisons and benchmarking over time between places (e.g. units, hospitals); make judgments and set priorities (e.g. choosing a hospital or surgery or organising medical care); support accountability, regulation, and accreditation; support quality improvement; and support patients' choice of providers.[65]

The WHO has recommended that reporting systems should be blame-free, concluding that successful systems are: non-punitive; confidential; independent; expert analysis; credible; timely; systems-orientated; responsive.[97] This conclusion was echoed by the team constructing a national set of indicators in Denmark which called for reporting systems to be: confidential, sanction-free, learning oriented. [117] Both of these recommendations include 'confidential' but this may be at odds with policies to improve transparency, such as the Freedom of Information act in the UK, which highlights the conflict between theoretically-appealing and practical solutions.

11.6.3.3. Types of application

The potential application of indicators was covered briefly in the section 2.5 (p51-) to explain the context with which the indicators are used. I give greater detail, including some commonly cited

strengths and weaknesses (summarised in Table 64), of some of these applications below drawing predominantly on the literature on generic (non-specialty-specific) indicators. The limitations and strengths of benchmarking between providers were covered earlier (paragraph 11.6.1.2, p262) and are not repeated here.

As a WHO report identified, whilst "they [PSIs] have not yet been validated extensively,… they have been found useful in several studies for estimating rates, identifying risk factors and tracking trends in inpatient safety events." [515] More generally, a review of patient safety recommended that indicators "can serve as timely indicators for latent problems that the organization is not otherwise tracking"(p1644).[73] My view would be to use more robust applications for benchmarking adverse events and perhaps even accountability, with other less-valid applications for case finding, with any use of such indicators within a wider portfolio of measures. I did not identify sufficient evidence within the literature that using such indicators for reimbursement or policy-making would necessarily be effective.

Table 64. Key strengths and weakness of common applications of administrative data

Application		Strength	Weakness
Accountability	11.6.3.4	Indicators can be aggregated to whichever level of accountability is being evaluated	Majority of commentators warn against 'hard' application due to, for instance, gaming
Benchmarking	11.6.1.2	Administrative data cover many providers (or other units) and are longstanding (so can compare over time)	Evidence of variation in coding practice
Reimbursement	11.6.3.5	Easy to implement as data often forms part of funding framework	Strong incentive to bias coding
Priorities	11.6.3.6	Linked to activity and cost	Lack of comparability between specialties due to coding framework
Case finding	11.6.3.7	Promotes openness	

11.6.3.4. Accountability

To assess how indicators could be utilised to hold people to account, one has to consider to whom the assessment will be made. [73] An advantage of administrative data is that the indicators can be aggregated to different levels to meet these difference reporting needs. In England, the levels of aggregation have focused on organisations or physicians.[526] Such transparency has, for example, been called for in reports on adverse care at Bristol Royal Infirmary and Mid-Staffordshire NHS Foundation Trust. However, as also suggested earlier, a review of the literature suggested that any patient safety measure should be non-punitive,[97] so this application remains contentious.

If health systems start to employ indicators for collecting 'hard' evidence to apportion accountability and responsibility then there will be a perverse incentive to under-report adverse events and, therefore, reduce the validity of the underlying dataset for the indicator. There is a need for a reporting culture [527] and, to catalyse a move to such a culture, there will need to be an end to any perceived or actual blame culture. This argument leads to the conclusion that indicators should not be used to hold people accountable. This is particularly important since many adverse events may be blameless. Certainly, "in aviation maintenance – a hands-on activity similar to medical practice in many respects – some 90% of quality lapses were judged as blameless" (Marx 1997 cited in [527]). Given the level of specificity and sensitivity of current indicators, they could only be applicable for compensation, professional regulation, or criminal justice if used in conjunction with other methods. Certainly, many academics have warned that current indicators should not be used as embarrassment, let alone blame or sanction. [73]

11.6.3.5. Reimbursement

Another potential use for these indicators is to determine levels of reimbursement for healthcare providers. In fact, in the USA if the administrative data reveals that one of a specific list of sentinel events has occurred during an admission the hospital will not be reimbursed for the costs of the healthcare for the index event or for treating the harm relating to the adverse event. However, rate-based indicators have rarely been used to determine levels of reimbursements. The issues relating to benchmarking that were outlined above also hold for this use. Bahl and colleagues have concluded that indicators based on administrative data are not appropriate to profile hospital performance or determine levels of reimbursement.[99] Indeed, Davies, in explaining the refinement of the HCUP quality indicators, recommended that indicators should not create incentives to improve measured performance without truly improving the quality of care provided, which could be an unintended consequence of linking with reimbursement.[80]

11.6.3.6. Make judgements about priorities

Priority setting – deciding about the areas to invest resources and focus - can be made using a myriad of evidence to make informed decisions and, therefore, these indicators can offer information relevant to assessing strategies. There is also scope for indicators to be developed and analysed further to increase their utility in this respect, for example, by calculating the associated costs. One of the advantages offered by analysing hospital administrative data – that it is linked, in many countries, to funding frameworks – remains largely unexploited. Further, despite countries assessing the value of technology by its cost-effectiveness, such analysis is limited. "There is no full economic evaluation of the burden related to the implementation of various methods… Obtaining such information is usually a secondary objective of the published studies, and no details are available on how the calculations were performed" [59]. The WHO noted that the cost-benefit analysis of patient safety measures "could provide an impetus and help countries make the necessary investments"(p86). [515]

11.6.3.7. Case findings

The most consistently recommended use of administrative data are for case finding, perhaps representing a 'safe option' as it does not introduce any significant incentives to game the indicators. In a paper outlining how the AHRQ PSIs could be applied to the NHS, Bottle and Aylin suggested that the indicators would be a useful tool to assist clinical audit towards potential issues.[112] Similarly, Miller and Zhan have highlighted the appropriate use of the PSIs as institutional case-finding tools aimed at internal quality improvement as opposed to use for directly comparing individual institutions, especially in public reports that identify individual hospitals.[24] In a paper setting out the limitations of hospital mortality rates, Lilford and Pronovost cautioned against the use of such indicators as the basis for sanctions or rewards but did concede that there was value in using such indicators to flag where further investigation might be needed.[49]

11.6.3.8. As part of measurement portfolio

A full picture of quality and safety would not be possible from any one data source.[528] In particular, for many uses, indicators are not detailed enough when used alone [73] and so "caution should be exercised when using administrative data, exclusively, to report quality and safety outcomes".[111] As well as a plurality of indicators, there are advantages in having a range of types of indicators; recall the discussion from earlier that, for instance, process of care measures are more positively influenced by feedback than outcome of care measures but the latter more intrinsically interesting to most stakeholders (with different data sources being more adept to being used for either outcomes or process measures).[529] As such, rather than conclude spuriously on which is the best

application, I instead recommend that any use of such indicators by policy-makers is, where possible, as part of a broader measurement portfolio. This should also involve patient input and experiences, perhaps in the form of patient reported outcome measures.

11.7. Conclusion

This project showed that specialty-specific standardised mortality rates can reveal significant differences in performance both between providers and over time. Moreover, I was able to show that routinely-collected hospital administrative data can offer a far broader understanding of quality and safety than solely mortality rates.

While I set out a range of limitations to both this and existing research that has used administrative data to evaluate quality and safety of healthcare, researchers in this area should not be too bashful about limitations in terms of validity. A paper setting out the limitations of HSMR concluded that "quality of care should remain innocent until proved guilty", although this view has been subsequently criticised.[512] Indeed there is no perfect measure of quality and so, if this maxim was held, then indolence would prevail.

This project itself has shown that using indicators based on administrative data, albeit in this speciality-specific context and applied in a way that overcomes key limitations, can focus government, senior doctors and the public alike on areas of likely substandard care.[530] To return to the similarly-themed fable and idiom introduced at the start of this thesis, if such research continues and some of the areas for further work are addressed then the blind men will have a better chance of identifying their subject and the issue of patient safety will no longer remain the elephant in the room.

References

1. World Health Organisation. World Health Statistics 2012. Geneva; 2012.
2. World Health Organisation. World Health Statistics 2008. Geneva; 2008.
3. Institute of Medicine. To err is human: Building a safer health system. . Washington, DC: National Academy Press; 1999.
4. McDonald KM, Romano PS, Geppert J, et al. Measures of Patient Safety Based on Hospital Administrative Data. The Patient Safety Indicators. Technical Review 5: Prepared by the University of California San Francisco*B*Stanford Evidence-based Practice Center under Contract No. 290-97-0013; 2002 August 2002.
5. Department of Health. Departmental Report 2008: The Health and Personal Social Services Programme. Annual report; 2008 19 May 2008. Report No.: Cm 7393.
6. . (Accessed 8 October, 2008, at http://www.npsa.nhs.uk/nrls/reporting/what-is-a-patient-safety-incident/.)
7. Medical Errors: The Scope of the Problem. Publication No. AHRQ 00-P037. (Accessed 8 May, 2010, at https://www.ahrq.gov/qual/errback.htm.)
8. World Health Organisation. World Alliance for Patient Safety: forward programme; 2004 October 2004.
9. Vincent C, Aylin P, Franklin B, et al. Is health care getting safer? BMJ British medical journal (Clinical research ed) 2008;337:a2426.
10. Lohr KN (ed.). Medicare: A Strategy for Quality Assurance. Vols I and II. Washington, DC: National Academy Press. cited in Mainz Defining and classificying clinical indicators for quality improvement IJQHC 2003;15:523-30 1990.
11. Brown C, Hofer T, Johal A, et al. An epistemology of patient safety research: a framework for study design and interpretation. Part 1. Conceptualising and developing interventions. Qual Saf Health Care 2008;17:158-62.
12. Kazandjian VA, Wicker KG, Matthes N, Ogunbo S. Safety is part of quality: a proposal for a continuum in performance measurement. J Eval Clin Pract 2008;14:354-9.
13. Donabedian A. Evaluating the quality of medical care. 1966. The Milbank quarterly 2005;83:691-729.
14. Wheeler SJ, Wheeler DW. Medication errors in anaesthesia and critical care. Anaesthesia 2005;60:257-73.
15. Reason J. Human error. New York: Cambridge University Press; 1990.
16. World Alliance for Patient Safety. Research for Patient Safety: Better Knowledge for Safer Care. Geneva: World Health Organisation; 2008.
17. Brennan TA, Leape LL, Laird NM, et al. INCIDENCE OF ADVERSE EVENTS AND NEGLIGENCE IN HOSPITALIZED-PATIENTS - RESULTS OF THE HARVARD MEDICAL-PRACTICE STUDY-I. New England Journal of Medicine 1991;324:370-6.
18. Wilson RM, Runciman WB, Gibberd RW, Harrison BT, Newby L, Hamilton JD. The Quality in Australian health care study. Medical Journal of Australia 1995;163:458-71.
19. Weingart SN, Wilson RM, Gibberd RW, Harrison B. Epidemiology of medical error. British Medical Journal 2000;320:774-7.
20. Vincent C. Patient Safety. First ed: Elsevier; 2006.
21. Medical Errors: The Scope of the Problem. An Epidemic of Errors. (Accessed 1 February, 2009, at http://www.ahrq.gov/qual/errback.htm.)
22. Leape LL, Lawthers AG, Brennan TA, Johnson WG. Preventing medical injury. Quality review bulletin 1993;19:144-9.
23. Vincent C, Neale G, Woloshynowych M. Adverse events in British hospitals: preliminary retrospective record review. BMJ 2001;322:517-9.

24. Miller, Zhan. Pediatric patient safety in hospitals: A national picture in 2000 (vol 113, pg 1741, 2004). Pediatrics 2004;114:907-.
25. Hayward RA, Hofer TP. Estimating hospital deaths due to medical errors - Preventability is in the eye of the reviewer. JAMA-J Am Med Assoc 2001;286:415-20.
26. Miller MR, Zhan C. Pediatric patient safety in hospitals: A national picture in 2000. Pediatrics 2004;113:pp 1741-6.
27. Kohn LT, Corrigan JM, Donaldson MS. To Err Is Human: Building a Safer health System. Washington, DC: National Academy Press; 1999.
28. CNN. CNN Headline News, Factoid, January 16, 1999. In; 1999.
29. McDonald CJ, Weiner M, Hui SL. Deaths due to medical errors are exaggerated in Institute of Medicine report. Jama 2000:93-5.
30. Landrigan CP. The safety of inpatient pediatrics: preventing medical errors and injuries among hospitalized children. Pediatr Clin North Am 2005;52:979-93.
31. Ehsani JP, Jackson T, Duckett SJ. The incidence and cost of adverse events in Victorian hospitals 2003-04. Medical Journal of Australia 2006;184:pp 551-5.
32. Arah OA, Westert GP, Hurst J, Klazinga NS. A conceptual framework for the OECD Health Care Quality Indicators Project. International Journal for Quality in Health Care 2006;18:pp 5-13.
33. Mainz J. Defining and classifying clinical indicators for quality improvement. International Journal for Quality in Health Care 2003;15:523-30.
34. Garcia Armesto S, Kelley E, Wei L. Patient Safety Data Systems in the OECD: A Report of a Joint Irish Department of Health - OECD Conference; 2007 22-Mar-2007. Report No.: DELSA/HEA/HCQ(2007)1.
35. Peterson AMB, Walker PH. Hospital-acquired infections as patient safety indicators. Annu Rev Nurs Res 2006;24:75-99.
36. Cook RI, Woods DD, Miller C. A Tale of Two Stories: Contrasting Views of Patient Safety: National Health Care Safety Council of the National Patietn Safety Foundation at the AMA; 1998.
37. Department of Health. Building a safer NHS for patients - implementing an organisation with a memory. London; 2001.
38. Zhan C, Friedman B, Mosso A, Pronovost P. Medicare payment for selected adverse events: building the business case for investing in patient safety. Health affairs (Project Hope) 2006;25:1386-93.
39. Hayashida K, Imanaka Y, Fukuda H. Measuring hospital-wide activity volume for patient safety and infection control: a multi-centre study in Japan. BMC Health Services Research 2007;7:7.
40. Agency for Healthcare Research and Quality. AHRQ Quality Indicators: Guide to Patient Safety Indicators (version 3.1): Agency for Healthcare Research and Quality; 2007.
41. Department of Health. Annual Report and Accounts, 2012-13. HC46. London: The Stationery Office; 2013.
42. National Audit Office. Reducing Healthcare Associated Infections in Hospitals in England. HC560. London: The Stationery Office; 1999.
43. Blendon RJ, DesRoches CM, Brodie M, et al. Views of practicing physicians and the public on medical errors. The New England journal of medicine 2002;347:1933-40.
44. cited in Medical Errors: The Scope of the Problem. An Epidemic of Errors. (Accessed 12 April 2009, at http://www.ahrq.gov/qual/errback.htm.)
45. Snow J. Edinburgh Medical Journal 1849;72:75-87.
46. Anonymous. Fatal application of chloroform. Lancet 1848;i:161-2.
47. Stichler JF. Is Your Hospital Hospitable?: How Physical Environment Influences Patient Safety. Nursing for Women's Health 2007;11:506-11.
48. Joint Commission on Accreditation of Healthcare Organizations. Characteristics of clinical indicators. QRB Quality review bulletin 1989;15:330-9.
49. Mannion R, Davies H, Marshall M. Impact of star performance ratings in English acute hospital trusts. J Health Serv Res Policy 2005;10:18-24.

50. Iezzoni LI, Foley SM, Heeren T, et al. A method for screening the quality of hospital care using administrative data: preliminary validation results. QRB Quality review bulletin 1992;18:361-71.
51. Berenholtz SM, Pronovost PJ. Monitoring patient safety. Crit Care Clin 2007;23:659-73.
52. Thomas EJ, Petersen LA. Measuring errors and adverse events in health care. Journal of General Internal Medicine 2003;18:61-7.
53. McIntyre N, Popper K. The critical attitude in medicine: the need for a new ethics. British Medical Journal Clinical Research Ed 1983;287:1919-23.
54. Sari AB-A, Sheldon TA, Cracknell A, Turnbull A. Sensitivity of routine system for reporting patient safety incidents in an NHS hospital: retrospective patient case note review. BMJ 2007;334:79.
55. Noble DJ, Pronovost PJ. Underreporting of patient safety incidents reduces health care's ability to quantify and accurately measure harm reduction. Journal of patient safety 2010;6:247-50.
56. Mainz J, Krog BR, Bjornshave B, Bartels P. Nationwide continuous quality improvement using clinical indicators: The Danish National Indicator Project. Published 2004. In: 5th International Summit on Indicators and Performance Measures; 2002 Nov; Paris, FRANCE; 2002. p. I45-I50.
57. McLoughlin V, Leatherman S, Fletcher M, Owen JW. Improving performance using indicators. Recent experiences in the United States, the United Kingdom, and Australia. International Journal for Quality in Health Care 2001;13:455-62.
58. Bankauskaite V, Dargent G. Health systems performance indicators: Methodological issues. Presupuesto y Gasto Público 2007;49:125-37.
59. Michel P. Strengths and weaknesses of available methods for assessing the nature and scale of harm caused by the health system: literature review; 2003.
60. Mainz J. Developing evidence-based clinical indicators: a state of the art methods primer. In: 18th International Conference of the International-Society-for-Quality-in-Health-Care; 2001 Oct 04; Buenos Aires, Argentina; 2001. p. I5-I11.
61. About NICE guidance. 2013. (Accessed 19 March, 2013, at http://guidance.nice.org.uk/.)
62. Rosen AK, Zhao S, Rivard P, et al. Tracking rates of Patient Safety Indicators over time: lessons from the Veterans Administration. Medical Care 2006;44:850-61.
63. Zhan C, Miller MR. Excess Length of Stay, Charges, and Mortality Attributable to Medical Injuries during Hospitalization. Journal of the American Medical Association 2003;290:pp 1868-74.
64. Department of Health. An organisation with a memory. Report of an expert group on learning from adverse events in the NHS chaired by the Chief Medical Officer. London: The Stationery Office; 2000.
65. Kristensen S, Mainz J, Bartels P. Patient Safety: Establishing a set of European patient safety indicators; 2007.
66. Tsang C, Aylin P, Palmer WL. Patient safety indicators: A systematic review of the literature: Imperial College London; 2008.
67. Tsang C, Palmer W, Bottle A, Majeed A, Aylin P. A Review of Patient Safety Measures Based on Routinely Collected Hospital Data. Am J Med Qual 2012;27:154-69.
68. Zhan C, Miller MR. Administrative data based patient safety research: a critical review. Qual Saf Health Care 2003;12 Suppl 2:ii58-63.
69. WHO working group on patient safety. Patient safety : rapid assessment methods for assessing hazards : report of the WHO working group meetings. Published 2003. In; 2002 17-19 December; Geneva, Switzerland: World Health Organisation; 2002.
70. World Health Organisation. Quality of care: patient safety; 2002 23 March 2002.
71. Weingart SN, Lezzoni LI. Looking for medical injuries where the light is bright. JAMA-J Am Med Assoc 2003;290:1917-9.
72. Databases and Related Tools from the Healthcare Cost and Utilization Project (HCUP): Fact Sheet. 2011. (Accessed at http://www.ahrq.gov/legacy/data/hcup/datahcup.htm.)

73. Rivard PE, Rosen AK, Carroll JS. Enhancing patient safety through organizational learning: Are patient safety indicators a step in the right direction? Health Services Research 2006;41:1633-53.
74. Davies SM, Geppert J, McClellan M, McDonald KM, Romano PS, Shojania KG. Refinement of the HCUP Quality Indicators. Technical Review Number 4: Prepared by University of California San Francisco-Stanford Evidence-based Practice Center for Agency for Healthcare Research and Quality; 2001.
75. Informational Brochure. (Accessed 15 February, 2013, at https://www.google.co.uk/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CC8QFjAA&url=http%3A%2F%2Fwww.ospfe.it%2Fil-professionista%2Fprogetto-iqip%2Fallegati%2Fiqip_brochure.pdf&ei=iOpCUf_9IIXHPLaogZAG&usg=AFQjCNEboHop698Q602SkVOkBNf7VKbY8w&bvm=bv.43828540,d.d2k.)
76. Iezzoni LI. Assessing quality using administrative data. Ann Intern Med 1997;127:666-74.
77. Lawthers AG, McCarthy EP, Davis RB, Peterson LE, Palmer RH, Iezzoni LI. Identification of in-hospital complications from claims data. Is it valid? Med Care 2000;38:785-95.
78. Ball J, Elixhauser A, M. J, et al. HCUP Quality Indicators, 1 Methods, Version 1.1: Outcome, Utilization, and Access Measures for Quality Improvement. Rockville, MD: Agency for Health Care Policy and Research; 1998.
79. Johantgen M, Elixhauser A, Bali JK, Goldfarb M, Harris DR. Quality indicators using hospital discharge data: state and national applications. Jt Comm J Qual Improv 1998;24:88-105.
80. Davies SM, Geppert J, McClellan M, McDonald KM, Romano PS, Shojania KG. Refinment of the HCUP Quality Indicators. Technical Review, No 4 2001.
81. Miller MR, Elixhauser A, Zhan C, Meyer GS. Patient Safety Indicators: using administrative data to identify potential patient safety concerns. Health Services Research 2001;36:110-32.
82. McDonald KM, Romano PS, Geppert JD, et al. Measures of Patient Safety Based on Hospital Administrative Data - The Patient Safety Indicators. (Technical Reviews, No. 5.). Rockville (MD): Agency for Healthcare Research and Quality (USA); 2002.
83. Romano PS, Geppert JJ, Davies S, Miller MR, Elixhauser A, McDonald KM. A national profile of patient safety in U.S. hospitals. Health affairs (Project Hope) 2003;22:154-66.
84. Zhan C, Miller MR. Administrative data based patient safety research: a critical review. Qual Saf Health Care 2003;12 Suppl 2:ii58-63.
85. Millar J, Mattke S, Members of the OECD Patient Safety Panel. Selecting Indicators for the Quality of Patient Safety at the Health Systems Level in OECD Countries. Health Technical Papers No. 18.; 2004.
86. McLoughlin V, Millar J, Mattke S, et al. Selecting indicators for patient safety at the health system level in OECD countries. International Journal for Quality in Health Care 2006;18:14-20.
87. How did the HCUP QIs differ from the new AHRQ QIs? , 2001. (Accessed 11 May 2009, at http://www.qualityindicators.ahrq.gov/hcup_archive.htm#2.)
88. Tsang C, Aylin P, Palmer W. Patient safety indicators: A systematic review of the literature: Centre for Patient Safety and Service Quality. Imperial College, London; 2008 October 2008.
89. Raftery J, Roderick P, Stevens A. Potential use of routine databases in health technology assessment. Health Technology Assessment 2005;9.
90. Weingart SN, Iezzoni LI, Davis RB, et al. Use of administrative data to find substandard care: validation of the complications screening program. Medical Care 2000;38:796-806.
91. Zhan C, Miller MR. Excess length of stay, charges, and mortality attributable to medical injuries during hospitalization.[see comment]. Jama 2003;290:1868-74.
92. Johnson CE, Handberg E, Dobalian A, Gurol N, Pearson V. Improving perinatal and neonatal patient safety: The AHRQ patient safety indicators. J Perinat Neonatal Nurs 2005;19:15-23.
93. Jackson T, Duckett S, Shepheard J, Baxter K. Measurement of adverse events using "incidence flagged" diagnosis codes. J Health Serv Res Policy 2006;11:21-6.
94. HESonline. (Accessed 23 March, 2009, at www.hesonline.nhs.uk.)

95. International Classification of Diseases (ICD) Information Sheet. (Accessed 15 March, 2013, at http://www.who.int/classifications/icd/factsheet/en/index.html.)
96. Burns EM, Rigby E, Mamidanna R, et al. Systematic review of discharge coding accuracy. Journal of Public Health 2012;34:138-48.
97. World Health Organisation: World Alliance for Patient Safety. The WHO World Alliance for Patient Safety: WHO draft guidelines for adverse event reporting and learning systems: From information to action; 2005. Report No.: WHO/EIP/SPO/QPS/05.3.
98. Johnson CE, Handberg E, Dobalian A, Gurol N, Pearson V. Improving perinatal and neonatal patient safety - The AHRQ patient safety indicators. J Perinat Neonatal Nurs 2005;19:15-23.
99. Bahl V, Thompson MA, Kau TY, Hu HM, Campbell DA. Do the AHRQ patient safety indicators flag conditions that are present at the time of hospital admission? Medical Care 2008;46:516-22.
100. Lawthers AG, McCarthy EP, Davis RB, Peterson LE, Palmer RH, Iezzoni LI. Identification of in-hospital complications from claims data. Is it valid? Medical Care 2000;38:785-95.
101. Naessens JM, Campbell CR, Berg B, Williams AR, Culbertson R. Impact of diagnosis-timing indicators on measures of safety, comorbidity, and case mix groupings from administrative data sources. Medical Care 2007;45:781-8.
102. Naessens JM, Huschka TR. Distinguishing hospital complications of care from pre-existing conditions. International Journal for Quality in Health Care 2004;16 Suppl 1:i27-35.
103. International Classification of Diseases 10th Revision (ICD-10). 2001. (Accessed 26 March, 2013, at http://www.cdc.gov/nchs/data/dvs/icd10fct.pdf.)
104. Kristensen S, Mainz J, Bartels P. Catalogue of Patient Safety Indicators: Safety Improvement for Patients in Europe - SimPatIE - Work Package 4; 2007 March.
105. Mainz J. Developing evidence-based clinical indicators: a state of the art methods primer. Published 2003. In: 18th International Conference of the International-Society-for-Quality-in-Health-Care; 2001 Oct 04; Buenos Aires, Argentina; 2001. p. I5-I11.
106. Romano PS, Mark DH. Bias in the coding of hospital discharge data and its implications for quality assessment. Med Care 1994;32:81-90.
107. Houchens RL, Elixhauser A, Romano PS. How often are potential patient safety events present on admission? Jt Comm J Qual Patient Saf 2008;34:154-63.
108. Audit Commission. PbR Data Assurance Framework 2007/08: Findings from the first year of the national clinical coding audit programme; 2008 August 2008.
109. De Coster C, Quan H, Finlayson A, et al. Identifying priorities in methodological research using ICD-9-CM and ICD-10 administrative data: report from an international consortium. BMC Health Serv Res 2006;6:77.
110. Stone PW, Horan TC, Shih HC, Mooney-Kane C, Larson E. Comparisons of health care-associated infections identification using two mechanisms for public reporting. American Journal of Infection Control 2007;35:pp 145-9.
111. Polancich S, Restrepo E, Prosser J. Cautious use of administrative data for decubitus ulcer outcome reporting. Am J Med Qual 2006;21:262-8.
112. Bottle A, Aylin P. How NHS trusts could use patient safety indicators to improve care. Health Care Risk Report 2008 2008;May:12-4.
113. Rosen AK, Rivard P, Zhao SB, et al. Evaluating the Patient Safety Indicators - How well do they perform on Veterans Health Administration data? Medical Care 2005;43:873-84.
114. Rivard PE, Elwy AR, Loveland S, et al. Applying Patient Safety Indicators across Healthcare Systems: Achieving Data Comparability. In: Henriksen K, Battles J, Lewin DI, Marks E, eds. Advances in Patient Safety: From Research to Implementation. Rockville, MD: Agency for Healthcare Research and Quality; 2005:pp. 7-24.
115. Rosen AK, Rivard PE, Loveland S, et al. Identification of Patient Safety Events from VA Administrative Data: Is It Valid? In: VA Health Services Research and Development National Meeting; 2006; Arlington, VA.; 2006.

116. Goddard M, Mannion R, Smith PC. Assessing the performance of NHS Hospital Trusts: the role of 'hard' and 'soft' information. Health policy (Amsterdam, Netherlands) 1999;48:119-34.
117. Danish Society for Patient Safety. Patient Safety in Denmark - Past, current and future activities; 2005 November 2005.
118. Perneger TV. A research agenda for patient safety. International Journal for Quality in Health Care 2006;18:1-3.
119. Weeks WB, West AN, Rosen AK, Bagian JP. Comparing measures of patient safety for inpatient care provided to veterans within and outside the VA system in New York. Quality and Safety in Health Care 2008;17:pp 58-64.
120. Pulcins I, Wan E. CIHI Survey: In-Hospital Hip Fractures in Canada: Using Information to Improve Patient Safety. Healthcare Quarterly 2004;7:25-7.
121. Coffey RM, Andrews RM, Moy E. Racial, ethnic, and socioeconomic disparities in estimates of AHRQ patient safety indicators. Medical Care 2005;43:I48-57.
122. Culler SD, Hawley JN, Naylor V, Rask KJ. Is the availability of hospital IT applications associated with a hospital's risk adjusted incidence rate for patient safety indicators: results from 66 Georgia hospitals. J Med Syst 2007;31:319-27.
123. Menachemi N, Saunders C, Chukmaitov A, Matthews MC, Brooks RG. Hospital adoption of information technologies and improved patient safety: A study of 98 hospitals in Florida. J Healthc Manag 2007;52:398-409.
124. Li PX, Schneider JE, Ward MM. Effect of critical access hospital conversion on patient safety. Health Services Research 2007;42:2089-108.
125. Jarman B, Bottle A, Aylin P, Browne M. Monitoring changes in hospital standardised mortality ratios. BMJ 2005;330:329.
126. Benn J, Arnold G, Wei I, Riley C, Aleva F. Using quality indicators in anaesthesia: feeding back data to improve care. British Journal of Anaesthesia 2012.
127. Wikipedia. Specialty (medicine).
128. Wu AW, Pronovost P, Morlock L. ICU incident reporting systems. Journal of Critical Care 2002;17:86-94.
129. Leape LL. Reporting of Adverse Events. New England Journal of Medicine 2002;347:1633-8.
130. Soeren Mattke, Edward Kelley, Peter Scherer, Jeremy Hurst, Maria Luisa G. Lapetra, HCQI Expert Group Members. Health Care Quality Indicators Project: Initial Indicators Report. OECD Health Working Papers 2006;No. 22, OECD Publishing.
131. Roberts V, Perryman MM. Creating a culture for health care quality and safety. Health Care Manag (Frederick) 2007;26:155-8.
132. Daumit GL, Pronovost PJ, Anthony CB, Guallar E, Steinwachs DM, Ford DE. Adverse events during medical and surgical hospitalizations for persons with schizophrenia. Archives of General Psychiatry 2006;63:267-72.
133. Rothschild JM, Landrigan CP, Cronin JW, et al. The Critical Care Safety Study: The incidence and nature of adverse events and serious medical errors in intensive care. Critical Care Medicine 2005;33:1694-700.
134. Duckett S, Daniels S, Kamp M, Stockwell A, Walker G, Ward M. Pay for performance in Australia: Queensland's new Clinical Practice Improvement Payment. J Health Serv Res Policy 2008;13:174-7.
135. Leggat S, Bartram T, Stanton P. Exploring the lack of progress in improving patient safety in Australian hospitals. Health Serv Manage Res 2008;21:32-9.
136. Kline TJ, Willness C, Ghali WA. Determinants of adverse events in hospitals--the potential role of patient safety culture. Journal for healthcare quality : official publication of the National Association for Healthcare Quality 2008;30:11-7.
137. Davies H, Powell A, Rushmet R. Healthcare professionals' views on clinician engagement in quality improvement: A literature review. London: The Health Foundation; 2007 April 2007.

138. Howard R, Avery A, Bissell P. Causes of preventable drug-related hospital admissions: a qualitative study. Qual Saf Health Care 2008;17:109-16.
139. Weller WE, Gallagher BK, Cen L, Hannan EL. Readmissions for venous thromboembolism: expanding the definition of patient safety indicators. Jt Comm J Qual Saf 2004;30:497-504.
140. Mayer-Oakes SA, Barnes C. Developing indicators for the Medicare Quality Indicator System (MQIS): challenges and lessons learned. Jt Comm J Qual Improv 1997;23:381-90.
141. Weiner BJ, Alexander JA, Baker LC, Shortell SM, Becker M. Quality improvement implementation and hospital performance on patient safety indicators. Med Care Res Rev 2006;63:29-57.
142. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. Health Technology Assessment 1998;2.
143. Read JL. The new era of quality of life assessment. In: Walker SR, Rosser RM, eds. Quality of life assessment. Dordrecht: Kluwer Academic Publishers; 1993:3-10.
144. Mainz J. Developing evidence-based clinical indicators: a state of the art methods primer. International Journal for Quality in Health Care 2003;15:i5-i11.
145. Romano P. Selecting Indicators for Patient Safety at the Health Systems Level in OECD Countries; 2007.
146. Brewer M. Research Design and Issues of Validity. In: Reis H, Judd C, eds. Handbook of Research Methods in Social and Personality Psychology. Cambridge: Cambridge University Press; 2000.
147. Romano PS, Mull HJ, Rivard PE, et al. Validity of selected AHRQ patient safety indicators based on VA National Surgical Quality Improvement Program data. Health Serv Res 2009;44:182-204.
148. Bernal-Delgado E, Garcia-Armesto S, Martinez-Lizaga N, Abadia-Taira B, Beltran-Peribanez J, Peiro S. Should policy-makers and managers trust PSI? An empirical validation study of five patient safety indicators in a national health service. BMC Med Res Methodol 2012;12:19.
149. Rosen A, Weeks W. Validating the Patient Safety Indicators (PSIs) in the VA: a Multi-Faceted Approach (01/10/07 - 30/9/2011); 2007 28 September 2007.
150. Anaesthesia: Improvement through partnership; Safer practice in neonatal care; Improving patient safety in intrapartum care; A collaborative approach to cancer care (Accessed 22 September, 2008, at http://www.npsa.nhs.uk/nrls/improvingpatientsafety/patientsafetyinitiatives/.)
151. National Audit Office. Progress in Improving Stroke Care. London; 2010 3 February 2010.
152. National Audit Office. Maternity Services in England; 2013 8 November 2013.
153. Cochrane Reviews. 2013. (Accessed 20 January, 2013, at http://www.cochrane.org/cochrane-reviews.)
154. West S, King V, Cary T. Systems to Rate the Strength of Scientific Evidence. Evidence Report/Technology Assessment No. 47. AHRQ Publication No. 02-E016. Rockville, MD: Agency for Healthcare Research and Quality; 2002 April 2002.
155. Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. Ann Intern Med 1997;126:376-80.
156. Green S, Higgins J, Alderson P, Clarke M, Mulrow C, Oxman A. Chapter 1: Introduction. . In: Higgins J, Green S, eds. *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.0.1 [updated September 2008]. ed: The Cochrane Collaboration 2008
157. Centre for Reviews and Dissemination. Systematic Reviews: CRD's guidance for undertaking reviews in health care. York: York Publishing Services; 2009.
158. Agency for Healthcare Research and Quality. National Quality Measures Clearinghouse. . Disease/Condition: Stroke, and Measures in Progress with 'Stroke' or 'Transient Ischaemic Attack' or 'TIA' in title 2011;2011.
159. Euphoric. Cariovascular disease and surgery. 2008;2011.
160. Queensland Health: Clinical Practice Improvement Centre. Acute Stroke Care Indicators 30/06/2007 (Version 1.7); 2007.

161. Rinere O'Brien S, Rinere O'Brien S. Trends in inpatient rehabilitation stroke outcomes before and after advent of the prospective payment system: a systematic review. Journal of Neurologic Physical Therapy 2010;34:17-23.
162. Agency for Healthcare Research and Quality. Systems to rate the strength of scientific evidence. Evidence Report/Technology Assessment: number 47. Rockville, MD: Agency for Healthcare Research and Quality (AHRQ); 2002.
163. Reker DM, Rosen AK, Hoenig H, et al. The hazards of stroke case selection using administrative data. Medical Care 2002;40:96-104.
164. Breslow N, NE D. Statistical methods in cancer, volume II: The design and analysis of cohort studies. Lyon: International Agency for Research on Cancer, World Health Organisation; 1987.
165. Kramer AA, Zimmerman JE. Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited. Crit Care Med 2007;35:2052-6.
166. Iezzoni LI, Shwartz M, Ash AS, Mackiernan YD. Predicting In-hospital Mortality for Stroke Patients: Results Differ across Severity-measurement Methods. Medical Decision Making 1996;16:348-56.
167. Ayres D. Commonly used public health statistics and their confidence intervals: Association of Public Health Observations (APHO); 2008 March 2008.
168. Spiegelhalter DJ. Funnel plots for comparing institutional performance. Stat Med 2005;24:1185-202.
169. National Audit Office. Departmental Overview: A summary of the NAO's work on the Department of Health 2011-12; 2013.
170. About us. 2013. (Accessed 15 March, 2013, at http://www.cqc.org.uk/public/about-us.)
171. About reporting patient safety incidents. (Accessed 15 March, 2013, at http://www.nrls.npsa.nhs.uk/report-a-patient-safety-incident/about-reporting-patient-safety-incidents/.)
172. Department of Health. The NHS Outcomes Framework 2012/13; 2011.
173. NHS Executive. Quality & Performance in the NHS: Clinical Indicators. Leeds: NHS Executive; 1999.
174. Aylin P, Tanna S, Bottle A, Jarman B. How often are adverse events reported in English hospital statistics? BMJ 2004;329:369.
175. Fraenkel DJ, Cowie M, Daley P. Quality benefits of an intensive care clinical information system. Crit Care Med 2003;31:120-5.
176. Evans RS, Pestotnik SL, Classen DC, Horn SD, Bass SB, Burke JP. Preventing adverse drug events in hospitalized patients. The Annals of pharmacotherapy 1994;28:523-7.
177. Silverman J, Stapinski C, Churchill W, Neppl C, Bates D, Gandhi T. Multifaceted approach to reducing preventable adverse drug events. American Journal of Health-System Pharmacy 2003;60:582-6.
178. Singer S, Meterko M, Baker L, Gaba D, Falwell A, Rosen A. Workforce perceptions of hospital safety culture: development and validation of the patient safety climate in healthcare organizations survey. Health Serv Res 2007;42:1999-2021.
179. The hospital guide. 2013. (Accessed 20 January, 2014, at http://drfosterintelligence.co.uk/thought-leadership/hospital-guide/.)
180. Hospital Episode Statistics, Admitted Patient Care, England - 2012-13. 2013. (Accessed 23 March, 2014, at http://www.hscic.gov.uk/searchcatalogue?productid=13264&q=title%3a%22Hospital+Episode+Statistics%2c+Admitted+patient+care+-+England%22&sort=Relevance&size=10&page=1#top.)
181. Hospital Episode Statistics. 2014. (Accessed 20 January, 2014, at http://www.hscic.gov.uk/hes.)
182. Stroke, Cerebrovascular accident. 2013. (Accessed 12 February, 2013, at http://www.who.int/topics/cerebrovascular_accident/en/.)

183. Stroke Medicine (Sub Specialty). (Accessed at www.jrcptb.org.uk/trainingandcert/ST3-SpR/Pages/Stroke-Medicine.aspx.)
184. Collaboration SUT. Collaborative systematic review of the randomised trials of organised inpatient (stroke unit) care after stroke. . BMJ 1997;314:1151-9.
185. Stroke Unit Trialists' Collaboration. How Do Stroke Units Improve Patient Outcomes?: A Collaborative Systematic Review of the Randomized Trials. Stroke 1997;28:2139-44.
186. National Institute for Health and Clinical Excellence. Final Appraisal Determination - alteplase for the treatment of acute ischaemic stroke; 2007.
187. Mooney H. Quality of stroke care varies widely across England. BMJ (Clinical research ed) 2010;340:c1816.
188. Schwamm L, Reeves MJ, Frankel M, Schwamm L, Reeves MJ, Frankel M. Designing a sustainable national registry for stroke quality improvement. American Journal of Preventive Medicine 2006;31:S251-7.
189. World Health Organisation. Global Priorities for Research in Patient Safety (first edition); 2008.
190. Wise J. Goodbye to the "hello nurse" in casualty departments. BMJ 1997;315.
191. Department of Health. Equity and excellence: Liberating the NHS. London; 2010 12 July 2010.
192. Department of Health. Transparency in outcomes: a framework for the NHS. London; 2010 19 July 2010.
193. Faxon DP, Schwamm LH, Pasternak RC, et al. Improving quality of care through disease management: Principles and recommendations from the American Heart Association's expert panel on disease management. Stroke 2004;35:1527-30.
194. Jones DW, Peterson ED, Bonow RO, et al. Translating research into practice for healthcare providers - The American heart association's strategy for building healthier lives, free of cardiovascular diseases and stroke. Circulation 2008;118:687-96.
195. Rudd AG, Williams LS. Advances in health policy and outcomes. Stroke 2009;40:e301-e4.
196. Core, Hospital-based Outcome Indicators. 2013. (Accessed 20 January, 2014, at
197. Mitka M, Mitka M. Researchers hope measuring performance will improve stroke care and prevention. Jama 2010;303:17-8.
198. Smith DB, Murphy P, Jones WJ, et al. Adherence to Stroke Quality Indicators in Colorado: Monitoring Temporal Trends by Sharing a Common Dataset. Stroke 2010;41:E255-E.
199. Weir NU, Dennis MS. A triumph of hope and expediency over experience and reason? J Neurol Neurosurg Psychiatry 2008;79:852.
200. American Academy of Neurology ACoR, Physician Consortium for Performance Improvement, National Committee for Quality Assurance. Stroke and Stroke Rehabilitation: Performance Measurement Set. Chicago (IL): American Medical Association (AMA); 2010.
201. Katzan IL, Furlan AJ, Lloyd LE, et al. Use of tissue-type plasminogen activator for acute ischemic stroke - The Cleveland area experience. JAMA-J Am Med Assoc 2000;283:1151-8.
202. Svendsen ML, Ehlers LH, Andersen G, Johnsen SP. Quality of care and length of hospital stay among patients with stroke. Medical Care 2009;47:575-82.
203. Brown DL, Morgenstern LB, Majersik JJ, Kleerekoper M, Lisabeth LD. Risk of fractures after stroke. Cerebrovascular Diseases 2008;25:95-9.
204. McKevitt C, Redfern J, Mold F, Wolfe C. Qualitative studies of stroke: A systematic review. Stroke 2004;35:1499-505.
205. Luengo-Fernandez R, Gray AM, Rothwell PM. Costs of Stroke Using Patient-Level Data. A Critical Review of the Literature. Stroke 2008.
206. Allen NB, Holford TR, Bracken MB, et al. Geographic Variation in One-Year Recurrent Ischemic Stroke Rates for Elderly Medicare Beneficiaries in the USA. Neuroepidemiology 2010;34:123-9.

207. Bateman BT, Claassen J, Willey JZ, et al. Convulsive status epilepticus after ischemic stroke and intracerebral hemorrhage: frequency, predictors, and impact on outcome in a large administrative dataset. Neurocrit Care 2007;7:187-93.
208. Ostbye T, Levy AR, Mayo NE. Hospitalization and case-fatality rates for subarachnoid hemorrhage in Canada from 1982 through 1991: The Canadian Collaborative Study Group of Stroke Hospitalizations. Stroke 1997;28:793-8.
209. Shapiro MF, Park RE, Keesey J, Brook RH. The effect of alternative case-mix adjustments on mortality differences between municipal and voluntary hospitals in New York City. Health Services Research 1994;29:95-112.
210. Mosimaneotsile B, Braun KL, Tokishi C. Stroke patient outcomes: Does an integrated delivery model of care make a difference? Physical and Occupational Therapy in Geriatrics 2000;17:67-82.
211. Polanczyk CA, Lane A, Coburn M, Philbin EF, Dec GW, DiSalvo TG. Hospital outcomes in major teaching, minor teaching, and nonteaching hospitals in New York state. American Journal of Medicine 2002;112:255-61.
212. Kuwabara K, Imanaka Y, Matsuda S, et al. Profiling of resource use variation among six diseases treated at 82 Japanese special functioning hospitals, based on administrative data. Health Policy 2006;78:306-18.
213. Lee HC, Chang KC, Huang YC, Lan CF, Chen JJ, Wei SH. Inpatient Rehabilitation Utilization for Acute Stroke Under a Universal Health Insurance System. American Journal of Managed Care 2010;16:E67-E74.
214. Kleindorfer D, Xu YY, Moomaw CJ, Khatri P, Adeoye O, Hornung R. US Geographic Distribution of rt-PA Utilization by Hospital for Acute Ischemic Stroke. Stroke 2009;40:3580-4.
215. Lichtman JH, Jones SB, Leifheit-Limson EC, Wang Y, Goldstein LB. 30-Day Mortality and Readmission After Hemorrhagic Stroke Among Medicare Beneficiaries in Joint Commission Primary Stroke Center-Certified and Noncertified Hospitals. Stroke 2011;42:3387-91.
216. Tirschwell DL, Longstreth WT, Kukull WA. Administrative data, comorbidity, and survival after ischemic stroke. Annals of Neurology 1999;46:501-.
217. Mitchell JB, Ballard DJ, Whisnant JP, Ammering CJ, Samsa GP, Matchar DB. What role do neurologists play in determining the costs and outcomes of stroke patients? Stroke 1996;27:1937-43.
218. Borzecki AM, Christiansen CL, Loveland S, Chew P, Rosen AK. Trends in the inpatient quality indicators: the Veterans Health Administration experience. Medical Care 2010;48:694-702.
219. Field TS, Green TL, Roy K, Pedersen J, Hill MD. Trends in hospital admission for stroke in Calgary. Canadian Journal of Neurological Sciences 2004;31:387-93.
220. Johansen HL, Wielgosz AT, Nguyen K, Fry RN. Incidence, comorbidity, case fatality and readmission of hospitalized stroke patients in Canada. Canadian Journal of Cardiology 2006;22:65-71.
221. Laudicella M, Street A. Quality of Care for TIA & Stroke Patients: Analysis of routine data; 2008.
222. Lazzarino AI, Palmer W, Bottle A, Aylin P. Inequalities in Stroke Patients' Management in English Public Hospitals: A Survey on 200,000 Patients. PLoS ONE 2011;6:e17219.
223. Scott I, Youlden D, Coory M. Are diagnosis specific outcome indicators based on administrative data useful in assessing quality of hospital care? Qual Saf Health Care 2004;13:32-9.
224. Slobbe LCJ, Arah OA, de Bruin A, Westert GP. Mortality in Dutch hospitals: Trends in time, place and cause of death after admission for myocardial infarction and stroke. An observational study. BMC Health Services Research 2008;8:11.
225. Tonarelli SB, Tibbs M, Qureshi MN, et al. Accuracy of the New ICD-9 Code for "Drip-and-Ship" Thrombolytic Treatment in Patients with Acute Ischemic Stroke. Neurology 2010;74:A515-A.
226. Tseng MC, Lin HJ. Readmission after hospitalization for stroke in Taiwan: Results from a national sample. J Neurol Sci 2009;284:52-5.

227. Votruba ME, Cebul RD. Redirecting patients to improve stroke outcomes - Implications of a volume-based approach in one urban market. Medical Care 2006;44:1129-36.
228. Williams LS, Ghose SS, Swindle RW. Depression and other mental health diagnoses increase mortality risk after ischernic stroke. Am J Psychiat 2004;161:1090-5.
229. Smith MA, Liou JI, Frytak JR, Finch MD. 30-day survival and rehospitalization for stroke patients according to physician specialty. Cerebrovascular Diseases 2006;22:21-6.
230. Tu JV, Gong Y. Trends in Treatment and Outcomes for Acute Stroke Patients in Ontario, 1992-1998. Arch Intern Med 2003;163:293-7.
231. Zhu HF, Newcommon NN, Cooper ME, et al. Impact of a Stroke Unit on Length of Hospital Stay and In-Hospital Case Fatality. Stroke 2009;40:18-23.
232. Volpp KGM, Epstein AJ, Williams SV. The effect of market reform on racial differences in hospital mortality. Journal of General Internal Medicine 2006;21:1198-202.
233. Zhu HF, Hill MD. Stroke - The Elixhauser Index for comorbidity adjustment of in-hospital case fatality. Neurology 2008;71:283-7.
234. Adeoye O, Hornung R, Khatri P, Kleindorfer D. Recombinant Tissue-Type Plasminogen Activator Use for Ischemic Stroke in the United States A Doubling of Treatment Rates Over the Course of 5 Years. Stroke 2011;42:1952-5.
235. Andaluz N, Zuccarello M. Recent trends in the treatment of cerebral aneurysms: analysis of a nationwide inpatient database. Journal of Neurosurgery 2008;108:1163-9.
236. Cho SH, Yun SC. Bed-to-nurse ratios, provision of basic nursing care, and in-hospital and 30-day mortality among acute stroke patients admitted to an intensive care unit: Cross-sectional analysis of survey and administrative data. Int J Nurs Stud 2009;46:1092-101.
237. Gattellari M, Worthington J, Jalaludin B, Mohsin M. Stroke Unit Care in a Real-Life Setting Can Results From Randomized Controlled Trials Be Translated Into Every-Day Clinical Practice? An Observational Study of Hospital Data in a Large Australian Population. Stroke 2009;40:10-7.
238. Westert GP, Lagoe RJ, Keskimaki I, et al. An international study of hospital readmissions and related utilization in Europe and the USA. Health Policy 2002;61:269-78.
239. Gregory PC, Han E. Disparities in postacute stroke rehabilitation disposition to acute inpatient rehabilitation vs. home findings from the North Carolina Hospital Discharge Database. American Journal of Physical Medicine and Rehabilitation 2009;88:100-7.
240. Meyer K, Simmet A, Arnold M, Mattle H, Nedeltchev K. Stroke events and case fatalities in Switzerland based on hospital statistics and cause of death statistics. Swiss Med Wkly 2009;139:65-9.
241. Lanska DJ, Kryscio R. Geographic distribution of hospitalization rates, case fatality, and mortality from stroke in the United States. Neurology 1994;44:1541-50.
242. Cleves MA, Sanchez N, Draheim M. Evaluation of two competing methods for calculating Charlson's comorbidity index when analyzing short-term mortality using administrative data. Journal of Clinical Epidemiology 1997;50:903-8.
243. Censullo J, Chiu D. Comprehensive stroke center quality metrics. Critical Pathways in Cardiology 2008;7:178-84.
244. Crowley RW, Yeoh HK, Stukenborg GJ, Medel R, Kassell NF, Dumont AS. Influence of Weekend Hospital Admission on Short-Term Mortality After Intracerebral Hemorrhage. Stroke 2009;40:2387-92.
245. NHS Information Centre. Indicator Portal. 2011;2012.
246. Health Care Quality Indicators Project 2006 Data Collection Update Report. [Online]. Organisation for Economic Cooperation and Development, 2007. (Accessed 7 May, 2012, at http://www.oecd.org/.)
247. Bardach NS, Zhao S, Gress DR, Lawton MT, Johnston SC. Association Between Subarachnoid Hemorrhage Outcomes and Number of Cases Treated at California Hospitals. Stroke 2002;33:1851-6.
248. Palmieri L, Barchielli A, Cesana G, et al. The Italian register of cardiovascular diseases: Attack rates and case fatality for cerebrovascular events. Cerebrovascular Diseases 2007;24:530-9.

249. Volpp KGM, Ketcham JD, Epstein AJ, Williams SV. The effects of price competition and reduced subsidies for uncompensated care on hospital mortality. Health Services Research 2005;40:1056-77.
250. Camilo O, Goldstein LB. Lower stroke-related mortality in counties with stroke centers: North Carolina Stroke Facilities Survey. Neurology 2005;64:762-3.
251. Bian J, Dow WH, Matchar DB, Bian J, Dow WH, Matchar DB. Medicare HMO penetration and mortality outcomes of ischemic stroke. American Journal of Managed Care 2006;12:58-64.
252. Carinci F, Roti L, Francesconi P, et al. The impact of different rehabilitation strategies after major events in the elderly: the case of stroke and hip fracture in the Tuscany region. BMC Health Services Research 2007;7:10.
253. Bell CM, Redelmeier DA. Mortality among patients admitted to hospitals on weekends as compared with weekdays. New England Journal of Medicine 2001;345:663-8.
254. Rosenthal GE, Harper DL, Quinn LM, Cooper GS. Severity-adjusted mortality and length of stay in teaching and nonteaching hospitals - Results of a regional study. JAMA-J Am Med Assoc 1997;278:485-90.
255. Saposnik G, Hill MD, O'Donnell M, Fang J, Hachinski V, Kapral MK. Variables associated with 7-day, 30-day, and 1-year fatality after ischemic stroke. Stroke 2008;39:2318-24.
256. Taylor Jr DH, Whellan DJ, Sloan FA. Effects of admission to a teaching hospital on the cost and quality of care for medicare beneficiaries. New England Journal of Medicine 1999;340:293-9.
257. Volpp KG, Rosen AK, Rosenbaum PR, et al. Mortality among patients in VA hospitals in the first 2 years following ACGME resident duty hour reform. JAMA-J Am Med Assoc 2007;298:984-92.
258. Escarce JJ, Jain AK, Rogowski J. Hospital competition, managed care, and mortality after hospitalization for medical conditions: Evidence from three states. Medical Care Research and Review 2006;63:112S-40S.
259. Qureshi A, Suri MFK, Nasar A, et al. Thrombolysis for ischemic stroke in the United States: data from National Hospital Discharge Survey 1999-2001. Neurosurgery 2005;57:647-54; discussion
260. Saposnik G, Baibergenova A, O'Donnell M, Hill MD, Kapral MK, Hachinski V. Hospital volume and stroke outcome: Does it matter? Neurology 2007;69:1142-51.
261. Saposnik G, Jeerakathil T, Selchen D, et al. Socioeconomic Status, Hospital Volume, and Stroke Fatality in Canada. Stroke 2008;39:3360-6.
262. Cross DT, Tirschwell DL, Clark MA, et al. Mortality rates after subarachnoid hemorrhage: variations according to hospital case volume in 18 states. Journal of Neurosurgery 2003;99:810-7.
263. Ogbu UC, Slobbe LCJ, Arah OA, De Bruin A, Stronks K, Westert GP. Hospital stroke volume and case-fatality revisited. Medical Care 2010;48:149-56.
264. Smith MA, Frytak JR, Liou JI, Finch MD. Rehospitalization and survival for stroke patients in managed care and traditional Medicare plans. Medical Care 2005;43:902-10.
265. Lichtman JH, Jones SB, Wang Y, Watanabe E, Leifheit-Limson E, Goldstein LB. Outcomes after ischemic stroke for hospitals with and without Joint Commission–certified primary stroke centers. Neurology 2011;76:1976-82.
266. Di Matteo M, Anderson C, Ratnasabapathy Y, Green G, Tryon K. The Acute Stroke Unit at Middlemore Hospital: An evaluation in its first year of operation. New Zealand Medical Journal 1190;117.
267. Sekimoto M, Kakutani C, Inoue I, Ishizaki T, Hayashida K, Imanaka Y. Management patterns and healthcare costs for hospitalized patients with cerebral infarction. Health Policy 2008;88:100-9.
268. Tung Y-C, Chang G-M, Chen Y-H. Associations of Physician Volume and Weekend Admissions With Ischemic Stroke Outcome in Taiwan: A Nationwide Population-Based Study. Medical Care 2009;47:1018-25 10.97/MLR.0b013e3181a81144.
269. Gillum LA, Johnston SC. Influence of physician specialty on outcomes after acute ischemic stroke. Journal of Hospital Medicine 2008;3:184-92.

270. Shen JJ, Washington EL, Aponte-Soto L, Shen JJ, Washington EL, Aponte-Soto L. Racial disparities in the pathogenesis and outcomes for patients with ischemic stroke. Managed Care Interface 2004;17:28-34.
271. Howrey BT, Kuo Y-F, Goodwin JS. Association of Care by Hospitalists on Discharge Destination and 30-day Outcomes After Acute Ischemic Stroke. Medical Care 2011;49:701-7 10.1097/MLR.0b013e3182166cb6.
272. Kind AJH, Smith MA, Frytak JR, Finch MD. Bouncing back: Patterns and predictors of complicated transitions 30 days after hospitalization for acute ischemic stroke. Journal of the American Geriatrics Society 2007;55:365-73.
273. Kind AJH, Smith MA, Liou JI, Pandhi N, Frytak JR, Finch MD. The price of bouncing back: One-year mortality and payments for acute stroke patients with 30-day bounce-backs. Journal of the American Geriatrics Society 2008;56:999-1005.
274. Polsky D, Lave J, Klusaritz H, et al. Is lower 30-day mortality posthospital admission among blacks unique to the veterans affairs health care system? Medical Care 2007;45:1083-9.
275. Reed SD, Blough DK, Meyer K, Jarvik JG. Inpatient costs, length of stay, and mortality for cerebrovascular events in community hospitals. Neurology 2001;57:305-14.
276. Reed SD, Cramer SC, Blough DK, Meyer K, Jarvik JG. Treatment With Tissue Plasminogen Activator and Inpatient Mortality Rates for Patients With Ischemic Stroke Treated in Community Hospitals. Stroke 2001;32:1832-40.
277. Lee WC, Joshi AV, Wang Q, Pashos CL, Christensen MC. Morbidity and mortality among elderly Americans with different stroke subtypes. Adv Ther 2007;24:258-68.
278. Lee AJ, Baker CS, Gehlbach S, Hosmer DW. Do Black elderly Medicare patients receive fewer services? An analysis of procedure use for selected patient conditions. Medical Care Research and Review 1998;55:314-33.
279. Gonzalez-Fernandez M, Kuhlemeier KV, Palmer JB. Racial disparities in the development of dysphagia after stroke: Analysis of the California (MIRCal) and New York (SPARCS) inpatient databases. Archives of Physical Medicine and Rehabilitation 2008;89:1358-65.
280. Janjua N, Nasar A, Lynch JK, Qureshi AI. Thrombolysis for ischemic stroke in children - Data from the nationwide inpatient sample. Stroke 2007;38:1850-4.
281. Kind AJH, Smith MA, Liou JI, Pandhi N, Frytak JR, Finch MD. Discharge Destination's Effect on Bounce-Back Risk in Black, White, and Hispanic Acute Ischemic Stroke Patients. Archives of Physical Medicine and Rehabilitation 2010;91:189-95.
282. Pathak EB, Sloan MA. Recent Racial/Ethnic Disparities in Stroke Hospitalizations and Outcomes for Young Adults in Florida, 2001-2006. Neuroepidemiology 2009;32:302-11.
283. Volpp KG, Stone R, Lave JR, et al. Is thirty-day hospital mortality really lower for black veterans compared with white veterans? Health Services Research 2007;42:1613-31.
284. Zacharia BE, Grobelny BT, Komotar RJ, Connolly ES, Mocco J. The influence of race on outcome following subarachnoid hemorrhage. J Clin Neurosci 2010;17:34-7.
285. Sheikh K, Bullock CM. Effect of measurement on sex difference in stroke mortality. Stroke 2007;38:1085-7.
286. Kapral MK, Wang H, Mamdani M, Tu JV. Effect of Socioeconomic Status on Treatment and Mortality After Stroke. Stroke 2002;33:268-75.
287. Shen JJ, Washington EL. Disparities in outcomes among patients with stroke associated with insurance status. Stroke 2007;38:1010-6.
288. Hoh BL, Chi YY, Lawson MF, Mocco J, Barker FG. Length of Stay and Total Hospital Charges of Clipping Versus Coiling for Ruptured and Unruptured Adult Cerebral Aneurysms in the Nationwide Inpatient Sample Database 2002 to 2006. Stroke;41:337-42.
289. Crowley RW, Yeoh HK, Stukenborg GJ, Ionescu AA, Kassell NF, Dumont AS. Influence of weekend versus weekday hospital admission on mortality following subarachnoid hemorrhage Clinical article. Journal of Neurosurgery 2009;111:60-6.

290. Hoh BL, Chi YY, Waters MF, Mocco J, Barker FG. Effect of Weekend Compared With Weekday Stroke Admission on Thrombolytic Use, In-Hospital Mortality, Discharge Disposition, Hospital Charges, and Length of Stay in the Nationwide Inpatient Sample Database, 2002 to 2007. Stroke 2010;41:2323-8.
291. Ho V, Hamilton BH. Hospital mergers and acquisitions: Does market consolidation harm patients? Journal of Health Economics 2000;19:767-91.
292. Medicare Provider Analysis and Review (MEDPAR) File. 2012. (Accessed 17 March, 2013, at http://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/IdentifiableDataFiles/MedicareProviderAnalysisandReviewFile.html.)
293. Katzenellenbogen JM, Vos T, Somerford P, Begg S, Semmens JB, Codde JP. Excess Mortality Rates for Estimating the Non-Fatal Burden of Stroke in Western Australia: A Data Linkage Study. Cerebrovascular Diseases 2010;30:57-64.
294. Pajunen P, Paakkonen R, Hamalainen H, et al. Trends in fatal and nonfatal strokes among persons aged 35 to >= 85 years during 1991-2002 in Finland. Stroke 2005;36:244-8.
295. Harris-Lane P, Siddiqui F, Kirmani JF. Validity of ICD-9 procedure code and CPT codes for thrombolysis in patients with ischemic stroke. Stroke 2005;36:439-.
296. Qureshi AI, Harris-Lane P, Siddiqi F, et al. International classification of diseases and current procedural terminology codes underestimated thrombolytic use for ischemic stroke. Journal of Clinical Epidemiology 2006;59:856-8.
297. Kleindorfer D, Lindsell CJ, Brass L, Koroshetz W, Broderick JP. National US estimates of recombinant tissue plasminogen activator use - ICD-9 codes substantially underestimate. Stroke 2008;39:924-8.
298. Glance LG, Osler TM, Mukamel DB, Dick AW. Impact of the present-on-admission indicator on hospital quality measurement - Experience with the Agency for Healthcare Research and Quality (AHRQ) Inpatient Quality Indicators. Medical Care 2008;46:112-9.
299. Tu JV, Nardi L, Fang J, et al. National trends in rates of death and hospital admissions related to acute myocardial infarction, heart failure and stroke, 1994–2004. Canadian Medical Association Journal 2009;180:E118-E25.
300. Mitchell JB, Ballard DJ, Matchar DB, Whisnant JP, Samsa GP. Racial variation in treatment for transient ischemic attacks: Impact of participation by neurologists. Health Services Research 2000;34:1413-28.
301. Samsa GP, Bian J, Lipscomb J, Matchar DB. Epidemiology of recurrent cerebral infarction - A Medicare claims-based comparison of first and recurrent strokes on 2-year survival and cost. Stroke 1999;30:338-49.
302. Hannerz H, Nielsen ML. Life expectancies among survivors of acute cerebrovascular disease. Stroke 2001;32:1739-44.
303. Goldacre MJ, Roberts SE, Griffith M. Place, time and certified cause of death in people who die after hospital admission for myocardial infarction or stroke. European Journal of Public Health 2004;14:338-42.
304. Department of Health. National Stroke Strategy; 2007 5 December.
305. Kamalesh M, Shen JZ, Tierney WM. Stroke mortality and race: Does access to care influence outcomes? American Journal of the Medical Sciences 2007;333:327-32.
306. Kirkman MA, Albert AF. Accuracy of hospital discharge coding is important in assessing trends in stroke rates. Swiss Med Wkly 2009;139:463-4.
307. Benesch C, Witter DM, Wilder AL, Duncan PW, Samsa GP, Matchar DB. Inaccuracy of the International Classification of Diseases (ICD-9-CM) in identifying the diagnosis of ischemic cerebrovascular disease. Neurology 1997;49:660-4.
308. Lakshminarayan K, Anderson DC, Jacobs DR, Barber CA, Luepker RV. Stroke rates: 1980-2000. American Journal of Epidemiology 2009;169:1070-8.
309. Coding Clinic. March 2002. Connecting for Health, 2002. (Accessed 7 Jun, 2011, at www.connectingforhealth.nhs.uk/systemsandservices/data/clinicalcoding/docs/ccc_mar02.pdf.)

310. Kokotailo RA, Hill MD. Coding of stroke and stroke risk factors using International Classification of Diseases, revisions 9 and 10. Stroke 2005;36:1776-81.
311. Weir N, Dennis MS. Towards a national system for monitoring the quality of hospital-based stroke services. Stroke 2001;32:1415-21.
312. Williams JY. Inventory of studies on the accuracy of Canadian health administrative databases. Toronto, Ontario; 1996.
313. Williams JY. A summary of studies on the quality of health care administrative databases in Canada. Ottowa: Institute of Clinical Effectiveness Sciences; 1996.
314. Canadian Institute for Health Information. Data quality of the discharge abstract database following the first-year implmementation of ICD-10-CA/CCI. Final report.; 2004 September.
315. Benin AL, Fenick A, Herrin J, Vitkauskas G, Chen J, Brandt C. How Good Are the Data? Feasible Approach to Validation of Metrics of Quality Derived From an Outpatient Electronic Health Record. Am J Med Qual 2011;26:441-51.
316. Dickensin K. Publication bias: recognizing the problem, understanding its origins and scope, and preventing harm. In: Rothstein H, Sutton AJ, Borenstein M, ed. Publication nias in meta-analysis - prevention, assessment and asjustments. Chichester: John Wiley; 2005:11-33.
317. Hamilton BB, Granger CV. Disability outcomes following inpatient rehabilitation for stroke. Physical Therapy 1994;74:494-503.
318. Deutsch A, Granger CV, Heinemann AW, et al. Poststroke Rehabilitation. Stroke 2006;37:1477-82.
319. Freburger JK. An analysis of the relationship between the utilization of physical therapy services and outcomes for patients with acute stroke. Physical Therapy 1999;79:906-18.
320. Schwamm LH, Reeves MJ, Pan W, et al. Race/Ethnicity, Quality of Care, and Outcomes in Ischemic Stroke. Circulation 2010;121:1492-501.
321. Bosworth HB. Trends in stroke mortality: The impact of the year 2000 age standard and the International Statistical Classification of Diseases and Related Health Problems, 10th Revision. Stroke 2002;33:1722-.
322. Gillum RF. New considerations in analyzing stroke and heart disease mortality trends - The year 2000 age standard and the International Statistical Classification of Diseases and Related Health Problems, 10th Revision. Stroke 2002;33:1717-21.
323. Kokotailo RA, Hill MD, Kokotailo RA, Hill MD. Coding of stroke and stroke risk factors using international classification of diseases, revisions 9 and 10. Stroke 2005;36:1776-81.
324. Salter K, Jutai JW, Teasell R, Foley NC, Bitensky J. Issues for selection of outcome measures in stroke rehabilitation: ICF Body Functions. Disability and Rehabilitation 2005;27:191-207.
325. Louis DZ, Yuen EJ, Braga M, et al. Impact of a DRG-based hospital financing system on quality and outcomes of care in Italy. Health Services Research 1999;34:405-15.
326. Rosenthal GE, Quinn L, Harper DL. Declines in Hospital Mortality Associated with a Regional Initiative to Measure Hospital Performance. Am J Med Qual 1997;12:103-12.
327. Department of Health. The Operating Framework for the NHS in England 2011/12. London; 2010 15 December 2010.
328. Department of Health Payment by Results Team. Payment by Results Guidance for 2009-10. London; 2009 5 February 2009.
329. Department of Health Payment by Results Team. Payment by Results Guidance for 2010-11. London; 2010 22 February 2010.
330. Department of Health. Operational plans 2008/09 - 2010/11. London; 2008 31 January 2008.
331. Lakhani A, Olearnik, H., Eayres, D. (eds). Compendium of Clinical and Health Indicators. London; 2008.
332. Dr Foster Hospital Guide Consultation: Stroke Indicators. Closed 3 September 2010. 2010. (Accessed 21 Jun, 2011, at http://www.hsj.co.uk/topics/dr-foster-hospital-guide-consultation/stroke-indicators/5018756.article.)

333. Connecting for Health. ICD-10 Clinical Coding Instruction Manual, Version 2.0. London; 2007 Febrary 2007.
334. National Collaborationg Centre for Chronic Conditions. Stroke: Diagnosis and initial management of acute stroke and transient ischaemic attack (TIA). London; 2008 July 2008. Report No.: NICE clinical guideline 68.
335. The Intercollegiate Stroke Working Party. National Sentinel Stroke Audit Clinical Audit 2010. London; 2011 May 2011.
336. The Danish National Indicator Project. Manual for the Development of Disease-specific Quality Indicators, Standards and Prognostic Factors under The Danish National Indicator Project, June 2009; 2009.
337. Mant J. Process versus outcome indicators in the assessment of quality of health care. International Journal for Quality in Health Care 2001;13:475-80.
338. Davenport RJ, Dennis MS, Warlow CP. Effect Of Correcting Outcome Data For Case Mix: An Example From Stroke Medicine. BMJ: British Medical Journal 1996;312:1503-5.
339. Irwin P, Rudd A. Case mix and process indicators of outcome in stroke. Journal of the Royal College of Physicians London 1998;32:442-4.
340. Kirkman MA, Mahattanakul W, Gregson BA, Mendelow AD. The Accuracy of Hospital Discharge Coding for Hemorrhagic Stroke. Acta Neurol Belg 2009;109:114-9.
341. Audit Commission. PbR data assurance framework 2008/09. London; 2009 27 August 2009.
342. Walsh K, Gompertz PH, Rudd AG. Stroke care: how do we measure quality? Postgraduate Medical Journal 2002;78:322-6.
343. The Intercollegiate Stroke Working Party. National Sentinel Stroke Audit Organisational Audit 2010. London; 2011 August 2011.
344. London Stroke Strategy. 2008. (Accessed 15 March, 2013, at http://www.londonprogrammes.nhs.uk/wp-content/uploads/2011/03/London-Stroke-Strategy.pdf.)
345. Bottle A, Aylin P. Comorbidity scores for administrative data benefited from adaptation to local coding and diagnostic practices. J Clin Epidemiol 2011;Jul 15.
346. Feldstein AC, Smith DH, Perrin N, et al. Reducing warfarin medication interactions: An interrupted time series evaluation. Archives of Internal Medicine 2006;166:1009-15.
347. Reiss SK, Ross-Degnan D, Zhang F, Soumerai SB, Zaslavsky AM, Wharam JF. Effect of Switching to a High-Deductible Health Plan on Use of Chronic Medications. Health Services Research 2011;46:1382-401.
348. Noyez L. Control charts, Cusum techniques and funnel plots. A review of methods for monitoring performance in healthcare. Interactive CardioVascular and Thoracic Surgery 2009;9:494-9.
349. Benneyan JC. Use and interpretation of statistical quality control charts. International journal for quality in health care : journal of the International Society for Quality in Health Care / ISQua 1998;10:69-73.
350. Aylin P, Yunus A, Bottle A, Majeed A, Bell D. Weekend mortality for emergency admissions. A large, multicentre study. Quality and Safety in Health Care 2010;19:213-7.
351. Barba R, Losa JE, Velasco M, Guijarro C, García de Casasola G, Zapatero A. Mortality among adult patients admitted to the hospital on weekends. European Journal of Internal Medicine 2006;17:322-4.
352. Cram P, Hillis SL, Barnett M, Rosenthal GE. Effects of weekend admission and hospital teaching status on in-hospital mortality. The American Journal of Medicine 2004;117:151-7.
353. Schilling PL, Campbell DA, Englesbe MJ, Davis MM. A Comparison of In-hospital Mortality Risk Conferred by High Hospital Occupancy, Differences in Nurse Staffing Levels, Weekend Admission, and Seasonal Influenza. Medical Care 2010;48:224-32.
354. Ogbu UC, Westert GP, Slobbe LCJ, Stronks K, Arah OA. A multifaceted look at time of admission and its impact on case-fatality among a cohort of ischaemic stroke patients. J Neurol Neurosurg Psychiatry 2011;82:8-13.

355. Saposnik G, Baibergenova A, Bayer N, Hachinski V. Weekends: A Dangerous Time for Having a Stroke? Stroke 2007;38:1211-5.
356. Reeves MJ, Smith E, Fonarow G, Hernandez A, Pan W, Schwamm LH. Off-hour admission and in-hospital stroke case fatality in the get with the guidelines-stroke program. Stroke 2009;40:569-76.
357. Hasegawa Y. The effect of weekends and holidays on stroke outcome in acute stroke units. Cerebrovascular Diseases 2005;20:325-31.
358. Janszky I, Ahnve S, Ljung R. Weekend Versus Weekday Admission and Stroke Outcome in Sweden From 1968 to 2005. Stroke 2007;38:e94.
359. Albright KC, Raman R, Ernstrom K, et al. Can Comprehensive Stroke Centers Erase the 'Weekend Effect'? Cerebrovascular Diseases 2009;27:107-13.
360. Kazley AS, Hillman DG, Johnston KC, Simpson KN. Hospital Care for Patients Experiencing Weekend vs Weekday Stroke: A Comparison of Quality and Aggressiveness of Care. Arch Neurol 2010;67:39-44.
361. Heros RC. Weekend versus weekday admission mortality. Journal of Neurosurgery 2009;111:57-8.
362. Haapaniemi H, Hillbom M, Juvela S. Weekend and Holiday Increase in the Onset of Ischemic Stroke in Young Women. Stroke 1996;27:1023-7.
363. Manfredini R, Boari B, Salmi R. Higher Stroke Mortality on Weekends: Are All Strokes the Same? Stroke 2007;38:e112.
364. Rudd AG, Hoffman A, Down C, Pearson M, Lowe D. Access to stroke care in England, Wales and Northern Ireland: the effect of age, gender and weekend admission. Age and Ageing 2007;36:247-55.
365. Care Quality Commission. Supporting life after stroke: A review of services for people who have had a stroke and their carers. London; 2011 January 2011.
366. Jakovljevic D, on behalf of the FINSTROKE Register Group. Day of the Week and Ischemic Stroke: Is It Monday High or Sunday Low? Stroke 2004;35:2089-93.
367. Islam M. The Safe Motherhood Initiative and Beyond. Bulletin of the World Health Organization 2007;85.
368. Rooks JP, Weatherby NL, Ernst EK. The National Birth Center Study. Part III--Intrapartum and immediate postpartum and neonatal complications and transfers, postpartum and neonatal care, outcomes, and client satisfaction. J Nurse Midwifery 1992;37:361-97.
369. Births and Deaths in England and Wales, 2011 (Final). 2012. (Accessed 28 February, 2013, at http://www.ons.gov.uk/ons/rel/vsob1/birth-summary-tables--england-and-wales/2011--final-/sb-births-and-deaths-in-england-and-wales--2011--final-.html.)
370. Martinez Marco EA, Aranaz Andres J. Is there a relationship between hospital readmission and quality of care?. [Spanish]

?Existe relacion entre el reingreso hospitalario y la calidad asistencial? Revista de Calidad Asistencial 2002;17:79-86.
371. Pattinson R, Say L, J. M, MH. B. Critical incident audit and feedback to improve preinatal and maternal mortality and morbidity. Cochrane Database of Systematic Reviews 2005;Issue 4.
372. Weindling AM. The confidential enquiry into maternal and child health (CEMACH). Archives of Disease in Childhood 2003;88:1034-7.
373. Simpson KR. Measuring Perinatal Patient Safety: Review of Current Methods. Journal of Obstetric, Gynecologic, & Neonatal Nursing 2006;35:432-42.
374. Sibanda T, Fox R, Draycott TJ, Mahmood T, Richmond D, Simms RA. Intrapartum care quality indicators: a systematic approach for achieving consensus. European Journal of Obstetrics, Gynecology, & Reproductive Biology 2013;166:23-9.
375. Knight H, Cromwell D, Meulen Jvd, et al. Patterns of Maternity Care in English NHS Hospitals 2011/12. London: The Royal College of Obstetricians and Gynaecologists; 2013.
376. Stafoggia M, Lallo A, Fusco D, et al. Spie charts, target plots, and radar plots for displaying comparative outcomes of health care. Journal of Clinical Epidemiology 2011;64:770-8.

377. Boulkedid R, Sibony O, Bossu-Salvador C, Oury JF, Alberti C. Monitoring healthcare quality in an obstetrics and gynaecology department using a CUSUM chart. Bjog-an International Journal of Obstetrics and Gynaecology 2010;117:1225-35.
378. Lane S, Weeks A, Scholefield H, Alfirevic Z. Monitoring obstetricians' performance with statistical process control charts. Bjog-an International Journal of Obstetrics and Gynaecology 2007;114:614-8.
379. Gaughan JM, Mason A, Street AD, et al. English hospitals can improve their use of resources: analysis of costs and length of stay for ten treatments. In: CHE Research paper 78. York: University of York, Centre for Health Economics;; 2012:1-71.
380. Borghans I, Kleefstra SM, Kool RB, Westert GP. Is the length of stay in hospital correlated with patient satisfaction? International Journal for Quality in Health Care 2012;24:443-51.
381. Oberer D, Auckerman L. Best practice: clinical pathways for uncomplicated births. Best practices and benchmarking in healthcare : a practical journal for clinical and management application 1996;1:43-50.
382. Korst LM, Gregory KD, Gornbein JA. Elective primary caesarean delivery: accuracy of administrative data. Paediatric and Perinatal Epidemiology 2004;18:112-9.
383. Carvalho ML, Silver LD. Reliability of the certification of the basic causes of neonatal deaths: implications for the study of preventable mortality. [Portuguese]

Confiabilidade de declaracao da causa basica de obitos neonatais: implicacoes para o estudo da mortalidade prevenivel. Revista de saude publica 1995;29:342-8.
384. Dogba M, Fournier P. Human resources and the quality of emergency obstetric care in developing countries: a systematic review of the literature. Human Resources for Health 2009;7:7.
385. Hussein J, Kanguru L, Astin M, Munjanja S. The Effectiveness of Emergency Obstetric Referral Interventions in Developing Country Settings: A Systematic Review. PLoS Med 2012;9:e1001264.
386. Hofmeyr GJ, Kulier R. Operative versus conservative management for 'fetal distress' in labour. Cochrane database of systematic reviews (Online) 2000:CD001065.
387. How to develop a search strategy for a Cochrane Review. 2007. (Accessed 20 February, 2013, at http://chmg.cochrane.org/sites/chmg.cochrane.org/files/uploads/How%20to%20develop%20a%20search%20strategy-support-manual.pdf.)
388. Kristensen S, Mainz J, Bartels P. Selection of indicators for continuous monitoring of patient safety: recommendations of the project 'safety improvement for patients in Europe'. International Journal for Quality in Health Care 2009;21:169-75.
389. Hospital Guide Consultation. 2011. (Accessed 12 March, 2013, at http://drfosterintelligence.co.uk/thought-leadership/hospital-guide-consultation/obstetrics-blood-transfusion-post-delivery-rate.)
390. Kotagal UR, Atherton HD, Eshett R, Schoettker PJ, Perlstein PH. Safety of early discharge for Medicaid newborns. JAMA-J Am Med Assoc 1999;282:1150-6.
391. Oleske DM, Linn ES, Nachman KL, Marder RJ, Sangl JA, Smith T. Effect of medicaid managed care on pregnancy complications. Obstet Gynecol 2000;95:6-13.
392. Romano PS, Yasmeen S, Schembri ME, Keyzer JM, Gilbert WM. Coding of perineal lacerations and other complications of obstetric care in hospital discharge data. Obstet Gynecol 2005;106:717-25.
393. Srinivas SK, Fager C, Lorch SA. Evaluating risk-adjusted cesarean delivery rate as a measure of obstetric quality. Obstet Gynecol 2010;115:1007-13.
394. The Definition of Normal Birth Rate. 2013. (Accessed 20 February, 2013, at http://www.birthchoiceuk.com/BirthChoiceUKFrame.htm?http://www.birthchoiceuk.com/NormalExplain.htm.)
395. Simonson DC, Ahern MM, Hendryx MS. Anesthesia staffing and anesthetic complications during cesarean delivery: a retrospective analysis. Nursing Research 2007;56:9-17.

396. Baghurst PA, Antoniou G. Risk Models for Benchmarking Severe Perineal Tears during Vaginal Childbirth: a Cross-sectional Study of Public Hospitals in South Australia, 2002-08. Paediatric and Perinatal Epidemiology 2012;26:430-7.
397. Bragg F, Cromwell DA, Edozien LC, et al. Variation in rates of caesarean section among English NHS trusts after accounting for maternal and clinical risk: cross sectional study. British Medical Journal 2010;341.
398. Bahrami S, Holstein J, Chatellier G, Le Roux YE, Dormont B. Using administrative data to assess the impact of length of stay on readmissions: Study of two procedures in surgery and obstetrics. [French]

Lien entre duree de sejour et readmissions en chirurgie et en obstetrique : etude de deux procedures a partir des donnees du PMSI. Revue d'Epidemiologie et de Sante Publique 2008;56:79-85.
399. Tseng YH, Chen CW, Huang HL, et al. Incidence of and predictors for short-term readmission among preterm low-birthweight infants. Pediatrics International 2010;52:711-7.
400. Thompson AH, Alibhai A, Saunders LD, Cumming DC, Thanigasalam N. Post-maternity outcomes following health care reform in Alberta: 1992-1996. Canadian Journal of Public Health 2003;94:104-8.
401. Drosler SE, Klazinga NS, Romano PS, et al. Application of patient safety indicators internationally: a pilot study among seven countries. International Journal for Quality in Health Care 2009;21:272-8.
402. Linton A, Peterson MR, Williams TV. Effects of maternal characteristics on cesarean delivery rates among U.S. Department of Defense healthcare beneficiaries, 1996-2002. Birth 2004;31:3-11.
403. Fantini MP, Stivanello E, Frammartino B, et al. Risk adjustment for inter-hospital comparison of primary cesarean section rates: need, validity and parsimony. Bmc Health Services Research 2006;6.
404. Bottle A, Aylin P. Application of AHRQ patient safety indicators to English hospital data. Qual Saf Health Care 2009;18:303-8.
405. Lutomski JE, Byrne BM, Devane D, Greene RA. Increasing trends in atonic postpartum haemorrhage in Ireland: an 11-year population-based cohort study. Bjog-an International Journal of Obstetrics and Gynaecology 2012;119:306-14.
406. Keino A, Gissler M, Jakobsson M, Lehtonen L, Tapper AM. The rate of the 3rd and 4th degree vaginal lacerations in the Finnish obstetric units as a patient safety indicator. Acta Obstetricia et Gynecologica Scandinavica 2012;91:43.
407. Reilly AF, Reilly PM. Medical injuries can increase mortality risk and incur additional costs. Evidence-Based Healthcare 2004;8:60-2.
408. Raleigh VS, Cooper J, Bremner SA, Scobie S. Patient safety indicators for England from hospital administrative data: case-control analysis and comparison with US data. BMJ 2008;337:a1702.
409. Zhan C, Miller MR. Excess length of stay, charges, and mortality attributable to medical injuries during hospitalization. Jama 2003;290:1868-74.
410. Roberts CL, Cameron CA, Bell JC, Algert CS, Morris JM. Measuring maternal morbidity in routinely collected health data - Development and validation of a maternal morbidity outcome indicator. Medical Care 2008;46:786-94.
411. Gurol-Urganci I, Cromwell DA, Edozien LC, Onwere C, Mahmood TA, van der Meulen JH. The timing of elective caesarean delivery between 2000 and 2009 in England. BMC pregnancy and childbirth 2011;11.
412. Meara E, Kotagal UR, Atherton HD, Lieu TA. Impact of early newborn discharge legislation and early follow-up visits on infant outcomes in a state medicaid population. Pediatrics 2004;113:1619-27.
413. Einarsdottir K, Kemp A, Haggar FA, et al. Increase in Caesarean Deliveries after the Australian Private Health Insurance Incentive Policy Reforms. Plos One 2012;7.

414. Renzi C, Sorge C, Fusco D, Agabiti N, Davoli M, Perucci CA. Reporting of quality indicators and improvement in hospital performance: the P.Re.Val.E. Regional Outcome Evaluation Program. Health Services Research 2012;47:1880-901.
415. Korst LM, Fridman M, Friedlich PS, et al. Hospital rates of maternal and neonatal infection in a low-risk population. Maternal and Child Health Journal 2005;9:307-16.
416. Bendavid E, Kaganova Y, Needleman J, Gruenberg L, Weissman JS. Complication rates on weekends and weekdays in US hospitals. American Journal of Medicine 2007;120:422-8.
417. Tang CH, Wang HI, Hsu CS, Su HW, Chen MJ, Lin HC. Risk-adjusted cesarean section rates for the assessment of physician performance in Taiwan: a population-based study. BMC Public Health 2006;6.
418. Webb DA, Culhane J. Hospital variation in episiotomy use and the risk of perineal trauma during childbirth. Birth-Issues in Perinatal Care 2002;29:132-6.
419. Baghurst PA. The case for retaining severe perineal tears as an indicator of the quality of obstetric care. Aust N Z J Obstet Gynaecol 2013;53:3-8.
420. Turner MJ. The use of quality control performance charts to analyze cesarean delivery rates nationally. International Journal of Gynecology & Obstetrics 2011;113:175-7.
421. Hsu CC, Shieh GR, Wu CS, Shen HC, Tang CH. Risk adjustment for inter-hospital comparisons of caesarean section rates in Taipei municipal hospitals. European Journal of Obstetrics, Gynecology, & Reproductive Biology 2006;127:190-7.
422. Gichangi P, Apers L, Temmerman M. Rate of caesarean section as a process indicator of safe-motherhood programmes: The case of Kenya. Journal of Health Population and Nutrition 2001;19:52-8.
423. Canadian Institute for Health Information. Health Indicators 2012. 2012.
424. Sousa MH, Cecatti JG, Hardy EE, Serruya SJ. Severe maternal morbidity (near miss) as a sentinel event of maternal death. An attempt to use routine data for surveillance. Reproductive Health 2008;5:6.
425. Korst LM, Gornbein JA, Gregory KD. Rethinking the cesarean rate - How pregnancy complications may affect interhospital comparisons. Medical Care 2005;43:237-45.
426. Kanter DE, Turenne W, Slonim AD. Hospital-reported medical errors in premature neonates. Pediatric Critical Care Medicine 2004;5:119-23.
427. Lu MC, Fridman M, Korst LM, et al. Variations in the incidence of postpartum hemorrhage across hospitals in California. Maternal and Child Health Journal 2005;9:297-306.
428. Srinivas S, Fager C, Lorch S. Variations in maternal outcomes by hospital teaching status. Am J Obstet Gynecol 2009;1):S165.
429. Kyser KL, Lu X, Santillan DA, et al. The association between hospital obstetrical volume and maternal postpartum complications. Am J Obstet Gynecol 2012;207.
430. Srinivas SK, Parrot M, Fager C, Simhan H, Lorch S. Evaluating the association between hospital blood availability and maternal and neonatal outcomes. Reproductive Sciences 2011;1):382A.
431. Sanderson M, Sappenfield WM, Jespersen KM, Liu Q, Baker SL. Association between level of delivery hospital and neonatal outcomes among South Carolina medicaid recipients. Am J Obstet Gynecol 2000;183:1504-11.
432. Gould JB, Danielsen B, Korst LM, et al. Cesarean delivery rates and neonatal morbidity in a low-risk population. Obstet Gynecol 2004;104:11-9.
433. Cheesman K, Brady JE, Flood P, Li G. Epidemiology of anesthesia-related complications in labor and delivery, New York State, 2002-2005. Anesth Analg 2009;109:1174-81.
434. Benedetti T, Walker S. Inappropriate identification of MFM practice as low quality by a NQF/LEAPFROG indicator. Am J Obstet Gynecol 2009;1):S69.
435. Guendelman S, Thornton D, Gould J, Hosang N. Mexican women in California: Differentials in maternal morbidity between foreign and US-born populations. Paediatric and Perinatal Epidemiology 2006;20:471-81.

436. Roberts LL, Ely JW, Ward MM. Factors contributing to maternal birth-related trauma. Am J Med Qual 2007;22:334-43.
437. Coffey RM, Andrews RM, Moy E. Racial, ethnic, and socioeconomic disparities in estimates of AHRQ patient safety indicators. Med Care 2005;43:I48-57.
438. Lain SJ, Algert CS, Nassar N, Bowen JR, Roberts CL. Incidence of Severe Adverse Neonatal Outcomes: Use of a Composite Indicator in a Population Cohort. Maternal and Child Health Journal 2012;16:600-8.
439. Moczygemba CK, Paramsothy P, Meikle S, et al. Route of delivery and neonatal birth trauma. Am J Obstet Gynecol 2010;202:361.e1-6.
440. Cameron CA, Roberts CL, Olive EC, Ford JB, Fischer WE. Trends in postpartum haemorrhage. Aust N Z J Public Health 2006;30:151-6.
441. Downey JR, Hernandez-Boussard T, Banka G, Morton JM. Is Patient Safety Improving? National Trends in Patient Safety Indicators: 1998-2007. Health Services Research 2012;47:414-30.
442. Callaghan WM, Creanga AA, Kuklina EV. Severe Maternal Morbidity Among Delivery and Postpartum Hospitalizations in the United States. Obstet Gynecol 2012;120:1029-36.
443. Roberts CL, Ford JB, Algert CS, Bell JC, Simpson JM, Morris JM. Trends in adverse maternal outcomes during childbirth: a population-based study of severe maternal morbidity. BMC pregnancy and childbirth 2009;9.
444. Ford AA, Bateman BT, Simpson LL, Ratan RB. Nationwide data confirms absence of 'July phenomenon' in obstetrics: it's safe to deliver in July. Journal of Perinatology 2007;27:73-6.
445. Gould JB, Qin C, Marks AR, Chavez G. Neonatal Mortality in Weekend vs Weekday Births. Journal of the American Medical Association 2003;289:2958-62.
446. Magalhaes MD, Bustamante-Teixeira MT. Severe acute maternal morbidity: use of the Brazilian Hospital Information System. Revista De Saude Publica 2012;46:472-8.
447. Gutierrez Saucedo ME, Hernandez Herrera RJ, Luna Garcia SA, Flores Santos R, Alcala Galvan LG, Martinez Gaytan V. Perinatal mortality at Hospital de Ginecoobstetricia no. 23 of Monterrey, Nuevo Leon, 2002-2006 period. [Spanish]

Mortalidad perinatal en el Hospital de Ginecoobstetricia num. 23 de Monterrey, Nuevo Leon (2002 a 2006). Ginecologia y Obstetricia de Mexico 2008;76:243-8.
448. Gregory KD, Fridman M, Shah S, Korst LM. Global measures of quality- and patient safety-related childbirth outcomes: should we monitor adverse or ideal rates? Am J Obstet Gynecol 2009;200:681.e1-.e7.
449. Mann S, Pratt S, Gluck P, et al. Assessing quality obstetrical care: development of standardized measures. Jt Comm J Qual Patient Saf 2006;32:497-505.
450. Lee KS, Lee SI, Seo K, Do YM. Impact of risk adjustment with insurance claims data on cesarean delivery rates of healthcare organizations in Korea. [Korean]. Journal of preventive medicine and public health = Yebang Uihakhoe chi 2005;38:132-40.
451. Maass C, Schleiz W, Weyermann M, Drosler SE. Are hospital administrative data suitable for external quality assurance? Comparison of quality indicators based on separate statutory data collections (BQS) and hospital administrative data. [German]

Krankenhaus-routinedaten zur externen qualitatssicherung? Vergleich von qualitatsindikatoren anhand der daten der gesetzlichen externen qualitatssicherung (BQS) und routinedaten. Deutsche Medizinische Wochenschrift 2011;136:409-14.
452. Goff SL, Pekow PS, Markenson G, Knee A, Chasan-Taber L, Lindenauer PK. Validity of Using ICD-9-CM Codes to Identify Selected Categories of Obstetric Complications, Procedures and Co-morbidities. Paediatric and Perinatal Epidemiology 2012;26:421-9.
453. Walker S, Strandjord TP, Benedetti TJ. In search of perinatal quality outcome measures: 1 hospital's in-depth analysis of the Adverse Outcomes Index. Am J Obstet Gynecol 2010;203:336.e1-.e7.

454. Kuklina EV, Whiteman MK, Hillis SD, et al. An enhanced method for identifying obstetric deliveries: Implications for estimating maternal morbidity. Maternal and Child Health Journal 2008;12:469-77.
455. Grobman W, Feinglass J, Murthy S. Are the agency for healthcare research and quality obstetric trauma indicators valid measures of hospital quality? Am J Obstet Gynecol 2005;193:S179-S.
456. Roberts CL, Bell JC, Ford JB, Morris JM. Monitoring the quality of maternity care: how well are labour and delivery events reported in population health data? Paediatric and Perinatal Epidemiology 2009;23:144-52.
457. Gerbier S, Bouzbid S, Pradat E, et al. Use of the French medico-administrative database (PMSI) to detect nosocomial infections in the University hospital of Lyon. [French]

Interet de l'utilisation des donnees du Programme medicalise des systemes d'information (PMSI) pour la surveillance des infections nosocomiales aux Hospices Civils de Lyon. Revue d'Epidemiologie et de Sante Publique 2011;59:3-14.
458. Chantry AA, Deneux-Tharaux C, Bal G, et al. French hospital discharge database: Data production, validity, and origins of errors in the field of severe maternal morbidity. Revue D Epidemiologie Et De Sante Publique 2012;60:177-88.
459. Chantry AA, Deneux-Tharaux C, Cans C, et al. Hospital discharge data can be used for monitoring procedures and intensive care related to severe maternal morbidity. Journal of Clinical Epidemiology 2011;64:1014-22.
460. Peaceman AM, Feinglass J, Manheim LM. Risk-adjustment of cesarean delivery rates: A practical method for use in quality improvement. Am J Med Qual 2002;17:113-7.
461. Deily ME, Hu TY, Terrizzi S, Chou SY, Meyerhoefer CD. The Impact of Health Information Technology Adoption by Outpatient Facilities on Pregnancy Outcomes. Health Services Research 2013;48:70-94.
462. Vincent C, Aylin P, Franklin BD, et al. Is health care getting safer? BMJ 2008;337.
463. Anastasakis E, Antsaklis A. Maternity unit performance index. A novel approach for evaluation of the changing obstetric practice in a single maternity unit. Archives of Gynecology and Obstetrics 2008;277:121-6.
464. Murray J, Saxena S, Modi N, et al. Quality of routine hospital birth records and the feasibility of their use for creating birth cohorts. Journal of Public Health 2012.
465. Group BiEC. Perinatal and maternal outcomes by planned place of birth for healthy women with low risk pregnancies: the Birthplace in England national prospective cohort study. BMJ 2011;343.
466. Department of Health. Report on confidential enquiries into maternal deaths in the United Kingdom 1991-1993. London; 1996 1 January.
467. Janakiraman V, Ecker J. Quality in obstetric care: Measuring what matters. Obstet Gynecol 2010;116:728-32.
468. Kutschmann M, Bungard S, Kotting J, Trumner A, Fusch C, Veit C. The Care of Preterm Infants With Birth Weight Below 1250 g Risk-Adjusted Quality Benchmarking as Part of Validating a Caseload-Based Management System. Deutsches Arzteblatt International 2012;109:519-U28.
469. Fleschler RG, Knight SA, Ray G. Severity and risk adjusting relating to obstetric outcomes, DRG assignment, and reimbursement. J Obstet Gynecol Neonatal Nurs 2001;30:98-109.
470. Rosen AK, Rivard P, Zhao S, et al. Evaluating the patient safety indicators: how well do they perform on Veterans Health Administration data? Med Care 2005;43:873-84.
471. Friedman M. Trying Hard Is Not Good Enough: BookSurge Publishing; 2009.
472. Pathak LR, Kwast BE, Malla DS, Pradhan AS, Rajlawat R, Campbell BB. Process indicators for safe motherhood programmes: their application and implications as derived from hospital data in Nepal. Tropical Medicine & International Health 2000;5:882-90.

473. Kesmodel US, Jalving LR. Measuring and improving quality in obstetrics - The implementation of national indicators in Denmark. Acta Obstetricia et Gynecologica Scandinavica 2011;90:295-304.
474. Zeitlin J, Wildman K, Breart G, et al. PERISTAT: Indicators for monitoring and evaluating perinatal health in Europe. European Journal of Public Health 2003;13:29-37.
475. Korst LM, Gregory KD, Lu MC, Reyes C, Hobel CJ, Chavez GF. A framework for the development of maternal quality of care indicators. Maternal and Child Health Journal 2005;9:317-41.
476. Thwaites R, Piercy J. Nosocomial respiratory syncytial virus infection in neonatal units in the United Kingdom. Acta Paediatr Suppl 2004;93:23-5.
477. Korst LM, Reyes C, Fridman M, Lu MC, Hobel CJ, Gregory KD. Gestational pyelonephritis as an indicator of the quality of ambulatory maternal health care services. Obstet Gynecol 2006;107:632-40.
478. de Bruin-Kooistra M, Amelink-Verburg MP, Buitendijk SE, Westert GP. Finding the right indicators for assessing quality midwifery care. International Journal for Quality in Health Care 2012;24:301-10.
479. Haller G, Kern C, Morales M, Pfister R, Clergue F. Development and validation of clinical indicators to measure iatrogenic complications in obstetrics: A Delphi-based study. European Journal of Anaesthesiology 2009;26:151.
480. Melman S, Hermens R, Nijhuis J, Scheepers L. Development of indicators to determine the quality of cesarean section care. Am J Obstet Gynecol 2012;1):S285.
481. Sandin Bojo AK, Hall-Lord ML, Axelsson O, Uden G, Wilde Larsson B. Midwifery care: development of an instrument to measure quality based on the World Health Organization's classification of care in normal birth. J Clin Nurs 2004;13:75-83.
482. Haller G, Camparini-Righini N, Kern C, et al. Patient safety indicators for obstetrics: A Delphi based study. [French]

Indicateurs securite en obstetrique : une etude Delphi. Journal de Gynecologie Obstetrique et Biologie de la Reproduction 2010;39:371-8.
483. IHI Perinatal Trigger Tool. 2012. (Accessed 15 February, 2013, at http://www.ihi.org/knowledge/Pages/Tools/PerinatalTriggerTool.aspx.)
484. Office for National Statistics. Births Summary Tables, England and Wales, 2012; 2013 10 July 2013.
485. Macfarlane A, Mugford M, Henderson J, Furtado A, Stevens J, A D. Birth counts: statistics of pregnancy and childbirth. . London: The Stationery Office; 2000.
486. Infant Mortality rate in England and Wales hits record low. Part of Child Mortality Statistics: Childhood, Infant and Perinatal, 2011 Release. 2013. (Accessed 20 December, 2013, at http://www.ons.gov.uk/ons/rel/vsob1/child-mortality-statistics--childhood--infant-and-perinatal/2011/sty-infant-mortality.html.)
487. Department of Health. Maternity Matters: Choice, access and continuity of care in a safe service; 2007.
488. Admitted Patient Care CDS Type - Other Birth event. 2013. (Accessed 5 December, 2013, at http://www.datadictionary.nhs.uk/version3/retired/data_dictionary/messages/commissioning_data_set_v5/admitted_patient_care_cds_type_-_other_birth_event_fr.asp?shownav=1.)
489. Dattani N, Datta-Nemdharry P, Macfarlane A. Linking maternity data for England, 2005-06: methods and data quality. Health statistics quarterly / Office for National Statistics 2011:53-79.
490. National Institute for Health and Clinical Excellence. Caesarean section: NICE clinical guideline 132; 2011 November 2011.
491. Gardosi J, Madurasinghe V, Williams M, Malik A, Francis A. Maternal and fetal risk factors for stillbirth: population based study. BMJ 2013;346.

492. Almoudaris AM, Burns EM, Bottle A, Aylin P, Darzi A, Faiz O. A colorectal perspective on voluntary submission of outcome data to clinical registries. The British journal of surgery 2011;98:132-9.
493. Maternity data is HES. (Accessed at http://www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=1937&categoryID=925.)
494. Brubaker L, Bradley CS, Handa VL, et al. Anal sphincter laceration at vaginal delivery - Is this event coded accurately? Obstet Gynecol 2007;109:1141-5.
495. Shennan A, Bewley S. What has happened to the UK Confidential Enquiry into Maternal Deaths? BMJ 2012;344.
496. Angelow A, Black N. The use and impact of national confidential enquiries in high-income countries. BMJ Qual Saf 2011;20:38-45.
497. Hospital Episode Statistics. 2013. (Accessed 5 December, 2013, at http://www.hscic.gov.uk/hes.)
498. House of Commons Public Accounts Committee. Maternity Service in England. Uncorrected Transcript of Oral Evidence.; 2013.
499. Palmer WL, Bottle A, Davie C, Vincent CA, Aylin P. Dying for the weekend: A retrospective cohort study on the association between day of hospital presentation and the quality and safety of stroke care. Archives of Neurology 2012;69:1296-302.
500. Bailit J, Landon M, Thom E, et al. The MFMU Cesarean Registry: impact of time of day on cesarean complications. Am J Obstet Gynecol 2006;195:1132 - 7.
501. Chalmers J, Shanks E, Paterson S, McInneny K, Baird D, Penney G. Scottish data on intrapartum related deaths are in same direction as Welsh data. Bmj 1998;317:539 - 40.
502. Rautava L, Hakkinen U, Korvenranta E, et al. Health-related quality of life in 5-year-old very low birth weight infants. J Pediatr 2009;155:338 - 43.
503. MacFarlane A. Variations in number of births and perinatal mortality by day of week in England and Wales. British Medical Journal 1978;2:1670-3.
504. Mathers CD. Births and perinatal deaths in Australia: variations by day of week. J Epidemiol Community Health 1983;37:57-62.
505. WD M. Neonatal mortality by the day of the week in the 1974-75 Arkansas live birth cohort. Am J Public Health 1981;71:601-5.
506. Hamilton P, Restrepo E. Weekend birth and higher neonatal mortality: A problem of patient acuity or quality of care? Jognn-Journal of Obstetric Gynecologic and Neonatal Nursing 2003;32:724-33.
507. Bell EF, Hansen NI, Morriss FH, Jr., et al. Impact of timing of birth and resident duty-hour restrictions on outcomes for small preterm infants. Pediatrics 2010;126:222-31.
508. Weissman JS, Rothschild JM, Bendavid E, et al. Hospital workload and adverse events. Med Care 2007;45:448-55.
509. Luo ZC, Liu S, Wilkins R, Kramer MS. Risks of stillbirth and early neonatal death by day of week. CMAJ 2004;170:337-41.
510. Pasupathy D, Wood AM, Pell JP, Fleming M, Smith GCS. Time of birth and risk of neonatal death at term: retrospective cohort study. Bmj 2010;341.
511. Street A. The resurrection of hospital mortality statistics in England. J Health Serv Res Policy 2002;7:104-10.
512. Clinical Evidence: What is GRADE? , 2012. (Accessed 28 March, 2013, at http://clinicalevidence.bmj.com/x/set/static/ebm/learn/665072.html.)
513. Marshall MN, Romano PS. Impact of reporting hospital performance. Qual Saf Health Care 2005;14:77-8.
514. Austin PC, van Walraven C, Wodchis WP, Newman A, Anderson GM. Using the Johns Hopkins Aggregated Diagnosis Groups (ADGs) to predict mortality in a general adult population cohort in Ontario, Canada. Medical Care 2011;49:932-9.

515. World Alliance for Patient Safety. Summary of the evidence on patient safety: Implications for research; 2008.
516. The STROBE Initiative: Strengthening the Reporting of Observational Studies in Epidemiology. STROKE Checklist for Cohort, Case-Control, and Cross-Sectional Studies. (Accessed 13 February, 2013, at http://www.strobe-statement.org/index.php?id=available-checklists.)
517. Higgins JPT, Green S. Cochrane handbook for systematic review of interventions, Version 5.0.0: The Cochrane Collaboration; 2008.
518. Pronovost PJ, Jenckes MW, Dorman T, et al. Organizational characteristics of intensive care units related to outcomes of abdominal aortic surgery. Jama 1999;281:1310-7.
519. Grobman WA, Feinglass J, Murthy S. Are the Agency for Healthcare Research and Quality obstetric trauma indicators valid measures of hospital safety? Am J Obstet Gynecol 2006;195:868-74.
520. Rosen A, Weeks W. Validating the Patient Safety Indicators (PSIs) in the VA: a multi-faceted approach, available at: http://www.qualityindicators.ahrq.gov/Downloads/Resources/Presentations/2007/rosen_validation_study_9_21.pdf.
521. Huber GP. Organizational Learning: The contributing processes and the literatures. Organization Science 1991;2.
522. Thompson R, Luettel D, Healey F, Scobie S. Safer care for the acutely ill patient: learning from serious incidents. The fifth report from the Patient Safety Observatory (PSO/5). 2007.
523. Mattke S, Epstein AM, Leatherman S. The OECD Health Care Quality Indicators Project: history and background. International Journal for Quality in Health Care 2006;18 Suppl 1:1-4.
524. Concha OP, Gallego B, Hillman K, Delaney GP, Coiera E. Do variations in hospital mortality patterns after weekend admission reflect reduced quality of care or different patient cohorts? A population-based study. BMJ Qual Saf 2013.
525. HFMA. NHS Services, Seven Days a Week Forum: Costing secen day services - The financial implications of seven day services for acute emergency and urgent services and supporting diagnostics. 2014.
526. Major breakthrough in NHS Transparency as consultant mortality data goes online for first time. 2013. (Accessed 20 January, 2014, at http://www.england.nhs.uk/2013/06/28/mjr-brkthgh-nhs-transp-cons/.)
527. Reason J. Human error: models and management. British Medical Journal 2000;320:768-70.
528. Thomson R. Adverse events reporting in English hospital statistics: Data relevant to patient safety should not be presented alone and out of context. BMJ 2004;329:857-.
529. van der Veer SN, de Keizer NF, Ravelli AC, Tenkink S, Jager KJ. Improving quality of care. A systematic review on how medical registries provide information feedback to health care providers. International journal of medical informatics 2010;79:305-23.
530. Smith R, Adams S. Hundreds die from strokes at weekends because of poorer NHS care: study. The Telegraph 2012 9 July 2012.
531. Ball J, Elixhauser A, M. J, et al. HCUP Quality Indicators, Software User's Guide, Version 1.1: Outcome, Utilization, and Access Measures for Quality Improvement. Rockville, MD: Agency for Health Care Policy and Research; 1998.
532. Collaborative CPIC-S. Acute Stroke Care Indicators. Version 1.7. 2007.
533. National Quality Measures Clearinghouse. 2013. (Accessed 27 February, 2013, at http://www.qualitymeasures.ahrq.gov/.)
534. Rosato S, Manno V, Mariotti S, Badoni G, D'Errigo P, Torre M. Deliverable N. 6: Detailed sheets of the colelcted outcome indicators (long list); 2008.
535. Gerova V, Griffiths P, Jones S, Bick D. The association between midwifery staffing and outcomes in maternity services in England: observational study using routinely collected data. Preliminary report and feasibility assessment. London: King's College London; 2010.

536. Jimenez-Puente A, Garcia-Alegria J, Gomez-Aracena J, Hidalgo-Rojas L, Lorenzo-Nogueiras L, Fernandez-Crehuet-Navajas J. Analysis of the causes and potential avoidability of readmissions in an acute patients' hospital. Medicina Clinica 2002;118:500-5.
537. NHS Maternity Statistics: Provider-level analysis. 2013. (Accessed 15 March, 2013, at http://www.ic.nhs.uk/catalogue/PUB03071/nhs-mate-eng-2010-2011-tab-v2.xls.)

Appendices

Appendix A. Outputs from the project

Publications related to the project

1. Palmer WL, Bottle A, Davie C, Vincent CA, Aylin P. Meeting the ambition of measuring the quality of hospitals' stroke care using routinely collected administrative data: a feasibility study. *Int J for Quality in Health Care*. 2013: 1-8

2. Palmer WL, Bottle A, Davie C, Vincent CA, Aylin P. Dying for the weekend: a retrospective cohort study on the association between day of hospital presentation and the quality and safety of stroke care. *Arch Neurol*. 2012; 69(10):1296-302.

3. Tsang C, Palmer W, Bottle A, Majeed A, Aylin P. A Review of Patient Safety Measures Based on Routinely Collected Hospital Data. *Am J Med Qual*. 2012; 27(2):154-69.

4. Lazzarino AI, Palmer W, Bottle A, Aylin P. Inequalities in Stroke Patients' Management in English Public Hospitals: A Survey on 200,000 Patients. *PLoS ONE* 2011;6:e17219.

Other outputs

1. Tsang C, Aylin P, Palmer W. Patient safety indicators: A systematic review of the literature: Centre for Patient Safety and Service Quality. Imperial College, London; 2008 October 2008

Conference abstracts related to the project

1. Palmer W, Lazzarino A, Tsang C, Aylin P. (2010) Monitoring safe stroke care: the development of indicators based on administrative data. International Forum on Quality and Safety in Health Care, Nice, 20th-24th April (Poster).

2. Ward D, Palmer W. (2011) Measuring Up – Showing Ambulance Trusts How They Compare as a Tool for Improvement. International Forum on Quality and Safety in Health Care 2011, Amsterdam, 5th-8th April (Poster)

3. Palmer W, Bottle A, Aylin P. (2011) Evaluating the quality and safety of hospital stroke care using administrative data. International Forum on Quality and Safety in Health Care 2011, Amsterdam, 5th-8th April (Poster)

Appendix B. Details of previous literature reviews of quality and safety indicators

Review	Pub Date	Description / Aims	Methods	Search terms	Resources	Inclusion / Exclusion criteria	Research identified	Indicator shortlisting	Final indicators
International Quality Improvement Programme (IQIP)	Since 1985 in US and 1991 internationally http://www.internationalqip.com/documents/brochure.pdf	Publish topic specific indicators that can be used for benchmarking performance	The measures were reviewed and modified following, firstly, feedback from subject-matter experts and then feedback and data from the first 6 to 12 months of use.	Tailored to the specific topic/procedure[8]	PubMed. National sites (e.g., NPSA, Joint Commission) and Departments and Ministries of Health in the US and other countries as well as any of the accrediting agencies.				Nearly 700 indicators, covering four care settings: acute, psychiatric (behaviour health) care, long term care, and home care.
HCUP QIs	1994	Meet the short-term needs for information on health care quality using standardised, user-friendly and existing sources of data	Described in Ball et al (1998)[531]	Not available	Not available	Not available			HCUP I consisted of 33 quality measures
Joint Commission's National Hospital Quality Measures.	April 2009 (Version 2.6b) (based on work since 1999)	To construct quality measure set for commissioners and providers of hospital care to publicly report their activities.		Not stated (and unavailable on request)	-		The ORYX® initiative, which became operational in March 1999, has over 8,000 disparate measures.		The Centers for Medicare & Medicaid Services (CMS) chose 10 starter set measures and 11 additional measures from a consensus-derived set of 39 measures.

[8] IQIP information from KarolW@mhaonline.org

Review	Pub Date	Description / Aims	Methods	Search terms	Resources	Inclusion / Exclusion criteria	Research identified	Indicator shortlisting	Final indicators
AHRQ QIs (refinement of HCUP QIs)	2001	To identify Patient Safety Indicators that could be used using ICD-9. They are based on the HCUP QIs.	Steps: 1. Defined the concepts and the evaluation framework; 2. Search the Literature; 3. Develop a Candidate List of PSIs; 4. Review PSIs; 5. Evaluate the PSIs	Medical Subject Heading (MeSH) terms "hospital, statistic, and methods" and "quality indicators"	In addition, electronic searches for articles published before February 2002, followed by hand search and Author search for Iezzoni, given her work on CSP	Inclusion criteria: quality indicator; novel indicator; and based on administrative data	PQI: 2,600 articles since 1994 of which 181 articles provided information on potential quality indicators based on administrative data. 27 met full criteria	Semi-structured interviews, evaluation framework (7 aspects: face validity, precision, minimum bias, construct-validity, fosters real quality improvement, application)	28 provider level and 4 area level Inpatient Quality Indicators (IQIs); Patient Safety Indicators; Paediatric Quality Indicators; and Prevention Quality Indicators.
Millar et al 2004, [OECD Health Care Quality Indicator (HCQI) project]	2004	Develop PSIs to cover: hospital-acquired infections; sentinel events; operative and postoperative complications; obstetrics; and other care-related AEs.	Review of existing routine data-based PSIs in OECD countries. PSIs selected according to indicator importance, scientific soundness and feasibility.	n/a	Previous indicator searches		7 indicator sets: AHRQ PSIs; AHRQ/CIHI PSIs; ACSQ; Complication Screening Programme (BIH, Israel), JCAHO Infection Control; JCAHO Sentinel Events.	Expert panel	21 indicators, from total of 59
SIMPATIE	2006	Described in Kristensen et al (2007)[104]	The development of indicators was based on an adjusted method described in Mainz (2003)[144] The literature search was started in the beginning of 2006, and repeated and extended in the process of the work, it was finalised by the end of 2006. "Stepwise Assessment Framework Approach"	An extensive literature search was initiated using the search terms: "Patient safety", "Indicator", "Risk" and "Harm".	PubMed and http://scholar.google.dk/ were searched. [9]	Search terms: "Test*", "Usage"/"Use", "Apply" and "Valid*" were used to assess if indicators previously evaluated in clinical setting.	Indicators or indicator programmes of the organisations mentioned in 'resources' were reviewed to determine whether they were suitable for characterisation.	3 aspects: relevance; validity and reliability; feasibility	42 indicators with 24 recommended in all or part of Europe

[9] Indicator programmes were reviewed Australian Council for Safety and Quality in Health Care (ACSQHC); Agency for Healthcare Research and Quality (AHRQ); The Good Medical Department, Denmark (DGMA); International Compendium of Health Indicators (WHO, OECD, Eurostat and ECHIM; Institute of Healthcare Improvements (IHI); Joint Commission on accreditation in Health Care (JCAHO); Nordic Indicators (NI); Performance Assessment Tool for Quality improvement in Hospitals (PATH); The Danish National Indicator Project (NIP); Performance Indicators on Patient Safety and effectiveness for Dutch Hospitals.

Review	Pub Date	Description / Aims	Methods	Search terms	Resources	Inclusion / Exclusion criteria	Research identified	Indicator shortlisting	Final indicators
NIP (Denmark)			Perform focused literature searches concerning the existence and experience with measurable parameters. Simultaneously search relevant international indicator libraries	^	e.g. the AHRQ, National Quality Measures Clearinghouse, indicators from professional society websites (mainly US and UK) and indicators from international agencies, such as OECD, Joint Commission etc.)			The indicators were developed, tailored to the chosen disease area, based on a search of scientific evidence or, in its absence, the decided by expert consensus.	For each disease about 6-10 indicators were determined.
National Indicators of Safety and Quality Project (Australia)	The Australian Council for Safety and Quality in Health Care was established in January 2000 to improve the safety and quality of health care provision in Australia.	Report on: suitability of OECD indicators for Australian data; measuring and reporting mortality; possible primary care indicators recommending indicators to support monitoring and improvement in safety and quality of health care.	General search of patient safety literature				230 papers from both health (152 papers) and non-health literature (78 papers) that are case studies, literature reviews, research studies or guideline documents containing data.		
Patient safety indicators: A systematic review of the literature Carmen Tsang, Paul Aylin, William Palmer	Oct 2008 (11/07 & 03/08)	review the literature on patient safety indicators developed from routinely collected hospital data	Systematic Literature review	patient safety OR hospital safety OR patient safety indicator$ OR safety improvement$ AND adverse event$ OR medica$ error$ OR sentinel event$ OR healthcare associated injury OR healthcare associated injuries OR iatrogenic disease$ OR preventable complication$	Medline (1950 to present) and Embase (1980 to present). Inc. grey literature & Government websites (Aus, Can, Den, Uk, US)	Exclusion: - Studies outside of secondary care; and - Studies primarily on specific diseases, diagnoses or treatments; or - Studies evaluating teaching or research tools; and - Articles not in English.	Total 1517 citations (1295 databases, 158 from reference lists, 43 from websites) 726 non-relevant, 244 duplications, 75 non-English, 2 not available, 2 non-human Of 468 fully abstracted, 277 either Strategies or Background and 91 articles on PSIs		

Notes: ^ The search terms for health related literature were based on the standard medical subject headings (MeSH) and include any or all of the following terms:

• Manpower, doctors, medical staff, nursing staff, health personnel, staffing, personnel staffing, scheduling, personnel administration, workload# • Nursing staff, hospital/supply, distribution • Quality assurance • Hospital safety management • Patient safety, patient incidents, patient adverse health effects, clinical risk, quality management, quality care, risk management, quality of healthcare, quality indicators, outcome assessment/methods, hospital/standards, accident prevention, safety management, adverse events, adverse drug reactions • Staff* rostering, staff fatigue, staff supervision, work place harassment, staff dissatisfaction, clinical governance, staff competency. # The term 'out-of-hours' was not searchable in the majority of databases used for this review. As a general rule, the term 'of' is not searchable in a database as

it is considered a ‘stop’ word that occurs too frequently to make search results relevant. ‘Out-of-hours’ is found in the searching related to workload, scheduling and working hours. * The term “staff” will include doctors, nurses, midwives, managers, surgeons, anaesthetists, physicians, allied health professionals.

Appendix C: Additional websites searched

The list below was adapted from a list used by Tsang and colleagues in their review of general patient safety indicators.[66]

Organisation	URL
International	
International Compendium of Health Indicators	http://www.healthindicators.org/ICHI/general/startmenu.aspx
European Commission Health Indicators Monitoring	http://www.echim.org/index.html
European Observatory on Health Systems and Policies	http://www.euro.who.int/observatory
Organization for Economic Co-operation and Development	http://www.oecd.org/home/0,2987,en_2649_201185_1_1_1_1_1,00.html
Safety Improvement for Patients In Europe Project	http://www.simpatie.org/Main
The Commonwealth Fund	www.commonwealthfund.org/
The Joint Commission International Center for Patient Safety	http://www.jointcommissioninternational.org/
WHO Collaborating Centre for Patient Safety Solutions	http://ccforpatientsafety.org/
World Health Organization Regional Office for Europe	http://www.euro.who.int/
World Health Organization World Alliance on Patient Safety	https://www.who.int/patientsafety/en/
Australia	
Australian Commission for Safety and Quality in Health Care	http://www.safetyandquality.gov.au/
Australian Council of Healthcare Standards	http://www.achs.org.au/
Australian Government	http://www.australia.gov.au/
Australian Healthcare Association	http://www.aushealthcare.com.au/
Australian Institute of Health & Welfare	http://www.aihw.gov.au/
Australian Medical Association	http://www.ama.com.au/
Australian Medical Council	http://www.amc.org.au/
Australian Patient Safety Foundation	http://www.apsf.net.au/
Commonwealth Department of Health and Ageing	http://www.health.gov.au/
Department of Health and Ageing	http://www.health.gov.au/internet/main/publishing.nsf/Content/Home
Queensland Health	http://www.health.qld.gov.au/
Royal Australian College of General Practice	http://www.racgp.org.au/
Royal Australian College of Physicians	http://www.racp.edu.au/
Royal College of Nursing	http://www.rcna.org.au/

Victorian Government Health Information - Department of Human Services, Victoria	http://www.health.vic.gov.au/
Canada	
Accreditation Canada	http://www.accreditationcanada.ca/default.aspx
Canadian Institute for Health Information	http://secure.cihi.ca/cihiweb/dispPage.jsp?cw_page=home_e
Canadian Patient Safety Institute	http://www.patientsafetyinstitute.ca/index.html
Institute for Safe Medication Practices	http://www.ismp-canada.org/
Denmark	
Ministry of Health and Protection	http://www.sum.dk/sum/site.aspx?p=34
The Danish Society for Patient Safety	http://www.patientsikkerhed.dk/en/about_the_danish_society_for_patient_safety/about_the_society/
The National Board of Health and Welfare	http://www.socialstyrelsen.se/en/
United Kingdom	
Centre for Reviews and Dissemination, University of York	http://www.crd.york.ac.uk/
Clinical and Health Outcomes Knowledge Base, National Centre for Health Outcomes Development	http://www.nchod.nhs.uk/
Department of Health	http://www.dh.gov.uk/en/index.htm
National Audit Office	http://www.nao.org.uk/
National Patient Safety Agency	http://www.npsa.nhs.uk/
Patient Safety Research Portfolio, University of Birmingham	http://www.pcpoh.bham.ac.uk/publichealth/psrp/index.shtml
Royal College of Obstetricians and Gynaecologists	http://www.rcog.org.uk
Royal College of Midwives	http://www.rcm.org.uk
United States of America	
Agency for Healthcare Research and Quality	http://www.ahrq.gov/
American Society for Healthcare Risk Management	http://www.ashrm.org/asp/home/home.asp
Center on Health Policy/Center on Primary Care and Outcomes Research, Stanford University	http://healthpolicy.stanford.edu/
Centers for Disease Control and Prevention	http://www.cdc.gov/ncidod/hip/default.htm
Healthcare Cost and Utilization Project	http://hcupnet.ahrq.gov/
Institute for Healthcare Improvement	http://www.ihi.org/
Institute for Safe Medication Practices	http://www.ismp.org/
Institute of Medicine	http://www.iom.edu/
National Coalition on Health Care	http://www.nchc.org/
National Coordinating Council for Medication Error Reporting and Prevention	http://www.nccmerp.org/
National Institutes of Health	http://www.nih.gov/
National Committee for Quality Assurance	http://www.ncqa.org/

National Patient Safety Foundation	http://www.npsf.org/
Quality Interagency Coordinating Task Force	http://www.quic.gov/
The Joint Commission	http://www.jointcommission.org/
National Center for Patient Safety, Department of Veterans Affairs	http://www.patientsafety.gov/

Appendix D: Stroke indicators

Indicator	Definition (example)	Reference (example)
Urgent brain scan for stroke patients	Proportion of stroke patients who have a brain scan (CT or MRI) performed on the day of (or by following day after) admission	Queensland Government[532]
Thrombolysis	Proportion of patients receiving thrombolysis treatment	American Academy of Neurology (2010)[200]
Emergency hospital transfers	All patients with suspected acute stroke are immediately transferred by ambulance to a receiving hospital providing hyper-acute stroke services	Kind (2007)[272]
Urinary Tract Infections (UTIs)	Proportion of stroke admissions with secondary diagnoses of urinary tract infection	Laudicella (2008)[221]
Pneumonia due to swallowing problems	Proportion of stroke admissions with pneumonitis due to food or vomiting (a consequence of swallowing problems)	Laudicella (2008)[221]
Lower respiratory tract infection (LRTI)	Proportion of stroke admissions with unspecified acute lower respiratory tract infection	Laudicella (2008)[221]
Deep Vein Thrombosis	Proportion of stroke admissions with secondary diagnoses of deep vein thrombosis	Cnesullo (2008)[243]
Pulmonary Embolism	Proportion of stroke patients with documented pulmonary embolism occurring during admission	Field (2004)[219]
Clostridium difficile	Proportion of stroke admissions with Enterocolitis due to Clostridium difficile	Laudicella (2008)[221]
Depression and other mental health problems following stroke	Diagnosis of depression and other mental health diagnosis in the first 3 years post-stroke	Williams (2004)[228]

Inpatient rehabilitation services	Proportion of stroke admissions with recorded use of rehabilitation services, defined as reimbursement claims for physical therapy, occupational therapy, speech therapy, or a combination of these therapies	Kapral (2002)[286]
Wafarin use	Warfarin use as proxied by outpatient claims for prothrombin time tests	Smith (2005)[264]
Length of stay	Length of hospitalisation	Zhu (2009)[231]
In-hospital mortality within 30 days of admission	The standardised proportion of stroke patients who die in-hospital within 30 days of admission	Queensland Government[532]
All-location mortality within 28 days of admission	Proportion of stroke patients who die within 28 days of admission	Pajunen (2005)[294]
Craniotomy: mortality rate	Number of deaths per 100 discharges with a code for craniotomy, with and without comorbidities and complications.	AHRQ (2013)[533]
Return to usual place of residence	Proportion of stroke patients returning to their usual place of residence following hospital treatment within 56 days	Qureshi (2005)[259]
Emergency readmission (all cause) to hospital	Percentage of patients of all ages with emergency readmission to any hospital within 27 days (inclusive) of the last, previous discharge from hospital after admission with a stroke	Rosato (2006)[534]
Recurrent stroke within 28 days	Percentage of patients of all ages with emergency readmission for stroke to any hospital within 27 days (inclusive) of the last, previous discharge from hospital after admission with a stroke	Johansen (2006)[220]
Readmission for incontinence	Percentage of patients readmitted to any hospital within 27 days with diagnosis for incontinence	Queensland Government[532]

Appendix E: Obstetrics indicators

Indicator	Indicator type	Detail	Inclusions or exclusions	Alternatives	Indicator sets, primary (other)	Key articles	Validation
For mother							
M1. Non-elective readmissions	Outcome	Within 42 days of the start of a delivery episode	Excludes readmissions of less than a day	1. Different time periods: any post-natal admission of mother (NCHOD); 6 months (Jimenez-Puente 2002); 90 day [400]; 30 day [423](Tseng 2010); 15 days (Tseng 2010); 7 or 14 days (Kotagal 1999) 2. Readmissions for PPH [440] 3. Readmissions within 28 days [535]	CQC, NPIC	390,398-400,423,536	
M2. Puerperal sepsis	Outcome	ICD-10 diagnosis code of O85 at any point during delivery spell within 42 days of the start of a delivery spell	Includes readmission within 42 days of the start of the delivery spell.	1. Puerperal sepsis and other puerperal infection (ICD-10 O85 and O86) or specified puerperal infection (excludes O86.4) 2. Chorioamnionitis (JCAHO) 3. Uterine infection (Korst 2005) 4. Wound complications (AHRQ)	CQC	427 392 415,428,435,444,452	Chorioamnionitis and morbidities recommended and detailed evaluation of infections indicator exists [415]
M3. Obstetric Trauma	Outcome /Proxy outcome (procedures)	Stratified by delivery type (VB with instrument; VB without instrument; CS) Third or fourth degree tear in dx or procedure codes	Includes uterine rupture, fracture of pelvis, including coccyx, laceration or haematoma of cervix, vagina, vulva, perineum and anus.	1. Trauma to perineal (PERISTAT) 2. Cases of obstetric trauma (third or fourth degree lacerations) (JCAHO) 3. Selected primipara with an intact perineum or unsutured perineal tear 4. Surgical repair of perineum (ACHS) 5. Aggregation of w/ and w/o	AHRQ, NPCI, JCAHO, NOGA, RANZCOG/ACHS, SimPatIE, DNBH, RCOG	83,375,391,393,401,404,407-409,416,427,435-437,441,444,451,452,455,456,461 3rd/4th degree tears 396,406,418,419	Korst et al (2005) recommended – validation of coding needed Extensive testing by AHRQ

Indicator	Indicator type	Detail	Inclusions or exclusions	Alternatives	Indicator sets, primary (other)	Key articles	Validation
				instrument (OECD)		Other [392]	
M4. Episiotomy	Proxy outcome			1. Perineal trauma and episiotomy rates (NCHOD) 2. Separated into selected primipara undergoing episiotomy AND sustaining a perineal tear during VB or with NO tear	JCAHO (future measures), NOGA	[418]	
M5. Uterine rupture	Outcome				NPIC		
M6. Incontinence	Outcome	Post-natal urinary or faecal incontinence		1. Faecal incontinence (PERISTAT)	NCHOD		
M7. Mortality	Outcome	Limited to only delivery admissions	Discharges with disposition of "deceased"		AHRQ (death in low-mortality DRGs), OBCQID, NCHOD, OECD, PERISTAT	[391,442,446,454]	Core measure in PERISTAT
M8. Long Length of stay	Outcome			1. Average LOS (OECD) 2. LOS>1 day>local standard (ACOG) 3. LOS > 2 (or 4) days for normal or instrumental delivery (Dr Foster)	NPIC	[389,390,398,461]	
M9. Complications of anaesthesia	Outcome	Discharges with dx code for anaesthesia complications in any dx field			AHRQ	[391,395,433,444]	
M10. Foreign body left during procedure	Outcome	Discharges with dx code for foreign body left in during procedure in any dx field			AHRQ		

Indicator	Indicator type	Detail	Inclusions or exclusions	Alternatives	Indicator sets, primary (other)	Key articles	Validation
M11. Post-partum haemorrhage (PPH)	Outcome	Diagnosis for postoperative haemorrhage/haematoma or related procedures	Procedure codes for control of haemorrhage must occur within one day after birth	1. Obstetric haemorrhage (ACOG) 2. PPH or transfusion (RANZCOG/ACHS) 3. Wider measure of PPH morbidity [440]	AHRQ (postoperative haemorrhage or hematoma) , NCHOD	391,392,405,427,435,440,452	Korst et al (2005) recommended
M12. Transfusion	Proxy outcome	Women who give birth vaginally who receive a blood transfusion during the same admission.		1. Transfusions (Korst et al 2005) PPH or transfusion (RANZCOG/ACHS) 2. Transfusions for CS (ACHS)	ACHS	389,405,427,444	
M13. Transfusion reaction	Outcome				AHRQ-derived		
M14. Episiotomy	Proxy outcome			1. Perineal trauma and episiotomy rates (NCHOD) 2. Separated into selected primipara undergoing episiotomy AND sustaining a perineal tear during VB/NO tear	JCAHO (future measures), NOGA	418,436,452,456	
M15. Maternal transfer to perinatal centre	Outcome			1. Transfer to ICU (NCHOD)	JCAHO (future measures)		
M16. Eclampsia	Process	ICD-10 O15			OBCQID, NCHOD, ACOG	391	
M17. Severe maternal morbidity	Composite			1. Adverse Outcome Index (Walker 2010) 2. Maternal composite (Srinivas 2010) 3. Weighted adverse outcome score	PERISTAT	393,410,424,429,430,442,443,446,449,453,454,458,459	Recommended for further development (PERISTAT) Requires extensive data collected and

Indicator	Indicator type	Detail	Inclusions or exclusions	Alternatives	Indicator sets, primary (other)	Key articles	Validation
				(WAOS, Mann 2006) 4. Problems with Child Birth (ACSQ/OECD)			risk adjustment (Janakiraman 2010)
M18. Selected complications				1. Other complications (Romano 2005) 2. Selected complications (Oleske 2000) 3. Preventable complications (Deily 2012) 4. Obstetric misadventure [ref!]		81,391,392,461	
M19. Cardiac arrest					Goff 2000	391	
M20. General anaesthesia for CS	Process				RANZCOG/ACHS		
M21. Pharmacological thromboprophylaxis & CS	Process	Total number of high risk women undergoing caesarean section who receive appropriate pharmacological thromboprophylaxis		1. Thrombosis (Srinivas 2009)	RANZCOG/ACHS	428	
M22. Timely antenatal assessment	Process	Within 10-14 weeks or by 12 weeks			NHS Information Centre	537	
For delivery method							
D1. Primary caesarean birth rate	Proxy outcome		Selected primipara only	1. Caesarean section rates (NPIC, NOGA, OECD, CIHI).	ACHS, JCAHO (future measures), ACOG	389,393,402,403,413,414,417,421-423,425,436,445,460	Korst et al (2005) recommended
D2. VBAC	Outcome	vaginal birth after	Any dx of abnormal		AHRQ, JCAHO,	402	Korst et al (2005)

Indicator	Indicator type	Detail	Inclusions or exclusions	Alternatives	Indicator sets, primary (other)	Key articles	Validation
		Cesarean (VBAC) delivery rate, uncomplicated	presentation, preterm, foetal death, multiple gestation diagnosis code. Breech procedures codes.		NPIC, RANZCOG/ACHS , RANZCOG/ACHS		recommended
D3. Unsuccessful VBAC	Outcome				JCAHO (future measures), ACOG		
D4. Failed vacuum or forcep delivery	Outcome			1. CS after failed instrumental delivery (RCOG)	Roberts 2009 RCOG	375,456	
D5. Emergency caesarean section	Outcome				RCOG	389,397,420,456	
D6. Late elective VB or CS	Process	At greater or equal to 37 and less than 39 weeks of gestation completed		1. Late CS (Gurol-Urganci 2011) (Benedetti 2009)	JCAHO	411,434	
D7. Normal births	Composite	A normal delivery is one without induction, without the use of instruments, not by caesarean section and without general, spinal or epidural anaesthetic before or during delivery.	Excludes: induction of labour; epidural or spinal; general anaesthetic; forceps or ventouse; CS; episiotomy	1. Operative VB rate (NPIC) 2. Instrumental delivery (OBCQID, Dr Foster) 3. Forceps or ventouse deliveries (NOGA) 4. Births without medical intervention (PERISTAT) 5. Outcome of selected primapara (RANZCOG/ACHS) 6. Indications and/or rate of elective labour (JCAHO future measures) 7. Spontaneous vs instrument VB	BirthChoiceUK, PERISTAT,RCOG	375,389,394,436	Further development recommended (PERISTAT) Instrumental VB recommended by Sibanda 2013

Indicator	Indicator type	Detail	Inclusions or exclusions	Alternatives	Indicator sets, primary (other)	Key articles	Validation
				(ACHS) 8. Mode of delivery rates (NCHOD) Vacuum: forcep rates (RCOG)			
D8. Induction rates	Proxy outcome	Deliveries with surgical and/or medical induction of labour	Selected primipara only		ACHS		
D9. Induction of labour resulting in emergency CS	Outcome				RCOG	[375]	
For neonate							
N1. Non-elective readmissions	Outcome	Within 28 days of birth	Excludes readmissions of less than a day	1. Readmission for jaundice (Meara 2004) 2. Emergency department visits (Meara 2004)	CQC	[412]	
N2. Infections	Outcome	Live babies of greater than or equal to 37 weeks gestational age (GA) born at the reporting hospital who develop a blood and/or cerebrospinal fluid (CSF) infection within 48 hours		1. Babies admitted to the neonatal intensive care unit (NICU) who have a significant blood infection occurring more than 48 hours after birth at any time during their whole admission - disaggregated for those above or below 1000 grams birth weight. (ACHS also) 2. Chorioamnionitis (JCAHO) 3. Neonatal sepsis (VON) 4. Part of wider category of system problems (Gould 2004)	AHRQ (ACHS)	[412,415,432]	Korst et al (2005) recommended (chorioamnionitis and morbidities) Detailed evaluation of validity and feasibility [415]
N3. Birth Trauma –	Outcome	Discharges with code for	Preterms with	1. Brachial plexus injury (Ford 2007)	AHRQ, OECD,	[81,83,393,401,407,]	Janakiraman et al

Indicator	Indicator type	Detail	Inclusions or exclusions	Alternatives	Indicator sets, primary (other)	Key articles	Validation
Injury to Neonate		birth trauma in any diagnosis field per 1000 live births	subdural or cerebral haemorrhage; infant dx of injury to skeleton or osteogenesis imperfecta	2. Wider definition of trauma (Moczygemba 2010) 3. Brachial plexus injury excluded (SimPatIE)	CIHI, SimPatIE, DNBH	409,416,437,439,441,444	(2010) suggest risk adjustment not well established. Generally performed well in AHRQ validation.[40]
N4. Neonatal mortality					Gould 2004	431,432,445,447	
N5. Iatrogenic pneumothorax	Outcome	dx code for iatrogenic pneumothorax in principle field; any dx code of chest trauma or pleural effusion or selected procedures	Less than 28 days and Birth-weight <500g;				
N6. Term NICU admissions	Proxy outcome	Percentage of inborn term babies transferred/admitted to a neonatal intensive care nursery or special care nursery facility	or reasons of congenital abnormality	1. Neonatal admission to intensive or special care (NCHOD) 2. Neonatal transfer to perinatal centre (JCAHO future measures)	ACHS (JCAHO)		Korst et al (2005) recommended Sibanda (2013) recommended
N7. Long length of stay	Proxy outcome	Infant stay over 3 days for VB and 5 days for CS			Gould et al 2004	432	
N8. Conditions related to deprivation of oxygen to brain	Outcome			1. Prevalence of hypoxic-ischaemic encephalopathy and cerebral palsy [ref!] 2. Birth asphyxia (Ford 2007)	PERISTAT	432,444	Recommended further development (PERISTAT)
N9. Dehydration	Outcome				Meara 2004	412	
N10. Adverse Outcome	Composite	Selected dx and procedure codes		1. Medical errors (Kanter 2004)	Lain et al (2011)	393,426,430,438	

Indicator	Indicator type	Detail	Inclusions or exclusions	Alternatives	Indicator sets, primary (other)	Key articles	Validation
Indicator		indicating adverse outcome		2. Neonatal composite (Srinivas 2010)			
N11. Selected complications	Composite			1. Complications of labour (Gould 2004) 2. Gastrointestinal complications (Gould 2004)		[432]	
Mother and neonate							
Proportion of births with no trauma, infection or readmissions	Composite	A "combination" of the above		Include/exclude the individual measures (e.g. mortality) to make a different composite measure 1. Ideal delivery (Gregory 2009)		[448]	

Notes:

National Perinatal Information Centre (NPIC); National Centre for Health Outcomes Development (NCHOD), UK; Nordic obstetric and gynaecological association (NOGA) – Knut Dalaker, Einar J Berle (Norwegian Society of Obstetrics and Gynecology): Clinical Guidelines in Obstetrics 1999. Norwegian Medical Association, Oslo,1999; OBSQID. European Consensus Conference on Quality Indicators for Perinatal Care. Annex II: The 21 Essential Indicators and their Definitions. Nov 1994; OECD measures described in PERISTAT report; American College of Obstetricians and Gynecologists (ACOG) as cited in Korst et al (2005) A Framework for the Development of Maternal Quality of Care Indicators; Vermont Oxford Network (VON) as cited in Korst et al (2005) A Framework for the Development of Maternal Quality of Care Indicators; Danish National Board of Health's (DNBH) Obstetrics Indicators; Australian Council for Safety and Quality (ACSQ)

*Selected primipara is defined as: A woman who is 20-34 years of age at the time of giving birth; Giving birth for the first time at greater than 20 weeks of gestation; Singleton pregnancy; Cephalic presentation; At 37 weeks to 41 weeks gestation

Clearing house: http://www.qualitymeasures.ahrq.gov/search/results.aspx?3062=1001%2c1002%2c&term=birth