Record linkage study to discriminate between community-acquired, healthcare-associated and hospital-acquired bloodstream infection in children in England

Thesis presented for the degree of

Doctor of Philosophy

Katherine Louise Henderson

Institute of Child Health, University College London

2017

Declaration

I, Katherine Louise Henderson, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signature: Date:

Abstract

Current national guidance for empiric antibiotic therapy of children with bloodstream infections (BSIs) indicates that treatment options are dependent on whether a BSI is community-acquired (CA) or hospital-acquired (HA). Recent changes in healthcare delivery have seen an emergence of patients from the community with healthcare-associated (HCA) BSIs who require different antibiotics compared to 'classic' CA BSIs; research in this field has predominantly concentrated on adults. This thesis aims to predict the relative proportions of CA, HCA and HA BSIs in children aged 1 month to 5 years by identifying demographic and clinical characteristics associated with these three groups using two datasets.

This thesis introduces the structure of the microbiology laboratory surveillance (LabBase2) and clinical (Hospital Episode Statistics) datasets and describes the probabilistic methodology used to link them, encompassing the match-weight calculation, threshold selection and evaluation of the final linked dataset. Three statistical approaches were used to predict CA, HCA and HA BSI using indicators of child susceptibility to infection, hospital exposure and timing of BSI.

The combination of CA and HCA BSI accounted for 74% or more of all BSIs, but differences in the ranges of individual CA and HCA BSI estimates from each of the three statistical approaches were too wide (difference of 48%) to discern a consistent proportion of either CA or HCA BSI; conversely, the estimated proportion of HA BSI was consistently predicted (26% or less of all BSI). The cross-tabulation of child characteristics and invasive pathogens by predicted CA, HCA and HA BSI group identified only a few low prevalence risk factors (more than two chronic conditions, indwelling devices, hospital discharge in the month prior to the BSI) that had a high predictive value for HCA as opposed to CA BSI. Pathogens historically associated with causing HA BSI, particularly Gram-negative pathogens that are harder to treat with antibiotics, were frequently isolated on or before the day of hospital admission.

This thesis demonstrated the wide range of pathogens isolated before and on admission to hospital, reflecting the growing mix of children acquiring CA and HCA BSIs in England in the community setting. A few, low-frequency clinical characteristics were predictive of CA, HCA or HA BSI, however, additional data from other sources (e.g. outpatient, primary care), may help to increase the accuracy of the prediction models. The mix of pathogens isolated from children is likely to become more heterogeneous as healthcare provision outside of hospital increases, suggesting that the concept of using clinical characteristics to identify the source of BSI is becoming less relevant. This highlights the importance of developing rapid indicators, such as early bedside testing, to inform appropriate treatment strategies.

Acknowledgements

My very sincere thanks to my supervisors Ruth Gilbert, Angie Wade, Berit Muller-Pebody and Alan Johnson for their continued support, guidance and inspiration throughout my PhD. It has been a privilege and an incredible experience to have been supervised by you. My thanks also to Mike Sharland (St. George's Hospital), for both his clinical advice and contagious enthusiasm in advancing the field of paediatric epidemiology, and to Mehdi Minaji, for helping with the technical IT aspects of the data linkage. Thank you also to Katie Harron for introducing me to probabilistic linkage.

Thank you to the members of the Improving Child Antimicrobial Prescribing group, the Institute of Child Health Early Career Research group and the St. George's Hospital Paediatric Department, who have listened, challenged and provided useful suggestions in response to various practice presentations of my PhD work over the past five years.

I would also like to thank my colleagues in the HCAI & AMR Department at PHE who have been very supportive throughout my PhD, making it possible for me to balance my job with my thesis, and for that I am extremely grateful. Thank you as well to my fellow PhD students and colleagues at ICH, it has been a pleasure to meet and work amongst such a talented and friendly team.

Finally, my immeasurable thanks to my family and friends for their endless support, humour, encouragement and provision of tea; to Harry for his musical inspiration; and to Antonio for his patience and faith in me, thank you. This thesis would not have been possible without you all.

Funding

This work was partially-funded by PHE's Education and Training Committee and the HCAI & AMR Department, and partially self-funded. The views and opinions expressed in this thesis are those of the author and do not necessarily reflect those of PHE or UCL ICH.

Contents

Declaration 2

Abstract 3

Acknowledgements 4

Funding 4

Contents 5

Figures 9

Tables 13

Appendices 15

Abbreviations 16

Chapter 1. Background, aims and objectives 18

1.1 Introduction 18
1.1.1 Structure and content of Chapter 1 19
1.2 Bloodstream infections (BSIs) 19
1.2.1 Community-acquired (CA) and hospital-acquired (HA) BSI 20
1.2.2 Healthcare-associated (HCA) BSIs 21
1.2.3 BSI in children 22
1.2.4 Previous Public Health England work on children with BSI 24
1.2.5 Existing Evidence 24
1.3 Systematic literature review 25
1.3.1 Methods 25
1.3.2 Results 26
1.3.3 Key findings 33
1.4 Professional opinion survey: what pathogens causing paediatric BSI are judged to be (or not be) HA BSI in the UK? 35
1.4.1 Methods 35
1.4.2 Results 36
1.4.3 Key findings 40
1.5 Discussion 40
1.6 Rationale for PhD 41
1.6.1 Aim 42
1.6.2 Objectives 42
1.6.3 Thesis structure 43

Chapter 2. Defining the cohort of children with BSIs for analysis 44

2.1 Introduction: data sources 44
2.2 Structure and Content 44
2.3 LabBase2: Public Health England's surveillance system 46
2.3.1 Purpose 46
2.3.2 Background 46
2.3.3 What data are reported to LabBase2? 46
2.3.4 Capturing data on positive blood cultures 47
2.3.5 Coverage 48
2.3.6 Data structure 49
2.3.7 Identification of infection episodes 49
2.3.8 Matching BSI episode reports to the same patient episode 50

2.3.9 Query design for LabBase2 data extraction 52
2.3.10 Patient identifiers in LabBase2 53
2.4 Hospital Episode Statistics (HES) 56
2.4.1 Purpose 56
2.4.2 Background 56
2.4.3 Data format 57
2.4.4 Patient identifiers in HES 58
2.4.5 HESid – identifying patients over time 59
2.4.6 HES three-step search and matching method 59
2.4.7 Episodes and Spells 60
2.4.8 Data structure 61
2.4.9 Query design for HES data extraction 61
2.5 Data sources: considerations for linkage and analysis 62
2.6 Information governance: permissions to access data 63
2.7 Cohort selection 63
2.7.1 Cohort time period 63
2.7.2 Exposure period 64
2.7.3 Definition of a child with a BSI 64
2.7.4 Definition of a BSI index case 64
2.7.5 Treatment of discharge and re-admission on the same day 65
2.7.6 Age 66
2.8 Summary 66
2.9 Conclusion 67

Chapter 3. Calculating match-weights to probabilistically link LabBase2 and HES datasets 68

3.1 Introduction 68
3.2 Data linkage 68
3.2.1 Growing demand 68
3.2.2 Methods of data linkage 70
3.2.3 Data linkage in the UK 74
3.2.4 Chosen linkage method 74
3.3 Methods: determining probabilistic match-weights 75
3.3.1 Patient identifiers 75
3.3.2 PHE match-weights for all ages (MW-All) 76
3.3.3 Calculation of child match-weights (MW-Child) 77
3.4 Results: determining probabilistic match-weights 79
3.5 Methods: probabilistic linkage of LabBase2 and HES 83
3.5.1 Threshold selection 85
3.6 Results: probabilistic linkage of LabBase2 and HES 85
3.6.1 Comparison of linked and unlinked OPIEs 88
3.6.2 Sensitivity analysis 89
3.7 Results: Association between MW-All and MW-Child 91
3.8 Final match-weight set selection 93
3.8.1 Data management and preparation of linked cohort for analysis 93
3.8.2 Final distribution of links by MW-Child 97
3.9 Discussion 97
3.9.1 Limitations 97
3.9.2 Summary 98

Chapter 4. Clinical and demographic characteristics of children with a BSI 99

4.1 Introduction 99
4.1.1 Mechanisms for acquisition of a BSI 101
4.2 Methods 105

4.2.1 Child susceptibility 105
4.2.2 Direct exposure to invasive BSI from inpatient healthcare 107
4.2.3 Timing of positive blood samples in relation to hospital admission 110
4.3 Results 111
4.3.1 Child susceptibility 111
4.3.2 Direct inpatient exposure to infection 115
4.3.3 Timing of positive blood samples in relation to hospital admission 117
4.4 Discussion 120
4.4.1 Child susceptibility to BSI 120
4.4.2 Direct exposure to invasive BSI from inpatient healthcare 120
4.4.3 Timing of positive blood samples in relation to hospital admission 121
4.4.4 Limitations of cohort 121

Chapter 5. Methods and Results: Predicting the proportion of CA, HCA and HA BSI 123

5.1 Introduction 123
5.2 Definitions and assumptions about the linked cohort 125
5.2.1 Defining children with HA BSI 125
5.2.2 Defining children with *a priori* CA BSI 126
5.2.3 Results 128
5.2.4 Summary and discussion 133
5.3 Methods of the statistical approaches A, B and C 134
5.3.1 Methods: approach A 137
5.3.2 Methods: approach B 139
5.3.3 Methods: approach C 141
5.3.4 Sensitivity analyses 144
5.4 Results of approaches A, B and C 144
5.4.1 Results: approach A 144
5.4.2 Results: approach B 157
5.4.3 Results: approach C 162
5.5 Chapter summary 171
5.5.1 Discussion and limitations of the methodology 171

Chapter 6. Distribution of risk factors and pathogens by predicted CA, HCA and HA BSI category 174

6.1 Introduction 174
6.1.1 Chapter structure 175
6.2 Comparing the estimated proportion of CA, HCA and HA BSI according to each method (A, B and C) 177
6.2.1 Methods 177
6.2.2 Results 177
6.2.3 Comparing the estimated proportions of CA, HCA and HA BSI with the commonly used CA/HA BSI threshold 180
6.2.4 Section summary 181
6.3 Distribution of risk factors by CA, HCA and HA BSI category 182
6.3.1 Methods 182
6.3.2 Results 183
6.3.3 Section summary 187
6.4 Distribution of bacterial species by time and BSI category 187
6.4.1 Methods 188
6.4.2 Results 188
6.4.3 Section summary 194
6.5 Pathogen distribution by predicted CA, HCA and HA BSI category 194
6.5.1 Methods 194

6.5.2 Results 194
6.5.3 Section summary 202
6.6 Discussion 202

Chapter 7. Discussion 204

7.1 Aim of the thesis 204
7.1.1 Objectives 204
7.2 Key findings 204
7.3 Strengths and limitations 205
7.3.1 Objective 1: Link microbiology surveillance and clinical data and evaluate the effect of match-weight choice on the final linked dataset 205
7.3.2 Objective 2: Develop statistical methods to identify patient and clinical characteristics associated with CA, HCA and HA BSI in children 207
7.3.3 Objective 3: Predict the proportion of CA, HCA and HA BSI occurring in children and the distribution of causative pathogens in each group 208
7.4 What this thesis adds 210
7.4.1 Implications for clinical practice 210
7.4.2 Implications for Public Health 210
7.4.3 Implications for Research 211
7.5 Closing remarks 213

References 214

Appendices 234

Figures

Figure 1.1 Flow diagram to illustrate the process of study selection using the inclusion criteria 23
Figure 1.2. Flow diagram to illustrate the process of study selection using the inclusion criteria 28
Figure 1.3. Screenshot example of the online professional opinion survey 38
Figure 1.4a-m. Professional opinion survey results from UK Paediatrician and Microbiologists for 12 pathogens for the percentage of BSIs caused by each pathogen, in their opinion, is likely to be HA 39
Figure 1.5. Flow diagram of steps to select appropriate antibiotic therapy for children with BSIs 42
Figure 2.1. Flow diagram of the steps already addressed in previous chapters (grey), those that will be addressed in this Chapter (red) and the following chapters 45
Figure 2.2. Flow diagram showing the path of the specimen results data from the local laboratories to the national LabBase2 database 48
Figure 2.3. Example demonstrating that OPIEs are assigned based on the organism species 50
Figure 2.4. Example of a LabBase2 query (fictitious data) 52
Figure 2.5. Proportion of identifier records with a data entry for each of the six patient identifiers in LabBase2 between 2007/08 and 2010/11 for children aged 0-5 years and patients with no date of birth 54
Figure 2.6. Proportion of identifier records with a data entry for each of the five patient identifiers in HES between 2007/08 and 2010/11 for children aged 0-5 year 58
Figure 2.7. Episodes and spells in HES 60
Figure 2.8. Example of the first restricted, patient identifier-only HES query (fictitious data) 61
Figure 2.9. Example of the second more comprehensive HES query (fictitious data) 62
Figure 2.10. Diagram indicating cohort time-period with exposure prior to the BSI 64
Figure 2.11. Examples of cohort inclusion for OPIE (1, 2, and 3) based on date of the positive blood isolate 65
Figure 2.12. Treatment of a discharge and a re-admission episodes occurring on the same day 65
Figure 3.1. Flow diagram of the steps already addressed in previous chapters (grey), those that will be addressed in this Chapter (red) and the following chapters 69
Figure 3.2. Data linkage error 70
Figure 3.3. Diagram illustrating that when full patient identifiers are known and complete in both datasets, there is very high confidence in the match being true 71
Figure 3.4. Diagram illustrating that when patient identifiers are missing or there are data entry errors, there is less confidence in the match being true 72
Figure 3.5. Range of possible probabilistic linkage thresholds 73
Figure 3.6. Example dataset containing true-matches 77
Figure 3.7. Dataset containing true-matches (grey) and false-matches (white) 78
Figure 3.8. Agree match-weights: Radar plot overlaying the sets of match-weights and comparing the proportion that each of the eight individual patient identifiers contributes towards the sum of the total match-weight (100%). DD = 2-digit day of birth; MM – 2-digit month of birth; YYYY = 4-digit year of birth; MW = match-weight 81
Figure 3.9. Disagree match-weights: Radar plot overlaying the sets of match-weights and comparing the proportion that each of the eight individual patient identifiers contributes towards the sum of the total match-weight (100%); DD = 2-digit day of birth; MM – 2-digit month of birth; YYYY = 4-digit year of birth; MW = match-weight 81

Figure 3.10. Missing data match-weights: Radar plot overlaying the sets of match-weights and comparing the proportion that each of the eight individual patient identifiers contributes towards the sum of the total match-weight (100%); DD = 2-digit day of birth; MM – 2-digit month of birth; YYYY = 4-digit year of birth; MW = match-weight....82
Figure 3.11. Linkage of all the LabBase2 data to each individual year of HES data....83
Figure 3.12. Computational challenges involved with probabilistic linkage of LabBase2 and HES data....84
Figure 3.13. Distribution of the total number of linked pairs per match-weight considered as true, uncertain or false matches using the MW-All in relation to the chosen threshold (selected via manual review)....86
Figure 3.14. Distribution of the total number of linked pairs per match-weight considered as true, uncertain or false matches using the MW-Child in relation to the chosen threshold (selected via manual review)....87
Figure 3.15. Number of unique and mutual OPIEs and HESid linked pairs for the MW-All and MW-Child....88
Figure 3.16. The number of linked pairs (denoted by the size of each circle) for MW-All compared to MW-Child and their relative assigned match-weights before the selection of the highest match-weight per OPIE....92
Figure 3.17. Flow diagram of LabBase2, good-links patient identifiers and HES data linkage, processing, manipulation and cleaning.....94
Figure 3.17 *(cont.)* Flow diagram of LabBase2, good-links patient identifiers and HES data linkage, processing, manipulation and cleaning. 95
Figure 3.17 *(cont.)* Flow diagram of LabBase2, good-links patient identifiers and HES data linkage, processing, manipulation and cleaning. 96
Figure 3.18. MW-Child distribution of the highest match-weight assigned to each linked LabBase2 and HES pair of records for children aged 1 month – 5 years, April 2007 – March 2011....97
Figure 4.1. Flow diagram of the steps already addressed in previous chapters (grey), those that will be addressed in this Chapter (red) and the following chapters....100
Figure 4.2. Summary of mechanisms of BSI epidemiology in children admitted to hospital....104
Figure 4.3. Frequency distribution of children with a BSI by quarter year of age....111
Figure 4.4. Frequency distribution of inpatient postnatal length of stay (no. of nights) for children aged 1 month < 1 year (0.5 symbolises day admissions only) 115
Figure 4.5. Frequency and timing when positive bacterial blood specimens were taken in relation to hospital admission (day 0) for children aged 1 month < 1 year, 2007-2011. Use of log-scale in main graph, and normal-scale in the excerpt 118
Figure 4.6. Frequency and timing when positive bacterial blood specimens were taken in relation to hospital admission (day 0) for children aged 1 - 5 years, 2007-2011. Use of log-scale in main graph, and normal-scale in the excerpt 119
Figure 5.1. Flow diagram of the steps taken (grey) and those that will be addressed in this chapter (red) to predict the proportion of CA, HCA and HA BSI in children 124
Figure 5.2. Expected timing of specimen date sample in relation to hospital admission: Varying HA1-HA9 cut-off dates for *a priori* HA BSI selection 126
Figure 5.3. Expected timing of specimen date sample in relation to hospital admission for *a priori* CA BSI 127
Figure 5.4. Proportional distribution of *a priori* CA pathogens in children aged 1 month < 1 year in relation to hospital admission 130
Figure 5.5 Proportional distribution of *a priori* CA pathogens in children aged 1-5 years in relation to hospital admission 131
Figure 5.6. Expected timing of specimen date sample in relation to hospital admission in approach A: Mixed CA, HCA and HA BSI....137

Figure 5.7. Expected timing of specimen date sample in relation to hospital admission in approach A: predicted HA BSI selection 138
Figure 5.8. Expected timing of specimen date sample in relation to hospital admission in approach A: Differentiating between predicted CA BSI and HCA BSI selection 139
Figure 5.9. Expected timing of specimen sample date in relation to hospital admission in approach B: Using the *a priori* CA and *defined* HA BSI cases for the modelling ... 139
Figure 5.10. Expected timing of specimen sample date in relation to hospital admission in approach B: predicted CA, HCA and HA BSI selection 140
Figure 5.11. K-medians three-group output 143
Figure 5.12. Distribution of children in the three K-medians output groups compared to *defined* HA BSI, *a priori* CA BSI and the remaining mixed BSI 143
Figure 5.13. Example of optimum outcome when comparing the three K-medians groups 143
Figure 5.14a-i. Distribution of the predicted HA cases (≥0.5) in relation to all the other cases of BSI between -5 and +30 of hospital admission in children aged 1 month < 1 year (jittered data points) 149
Figure 5.15a-i. Distribution of the predicted HA cases (≥0.5) in relation to all the other cases of BSI between -5 and +30 of hospital admission in children aged 1-5 years (jittered data points) 150
Figure 5.16. Distribution of predicted CA cases in relation to predicted HCA and *a priori* CA cases between -5 and +30 of hospital admission in children aged 1 month<1 year (jittered data points) 154
Figure 5.17. Distribution of predicted CA cases in relation to predicted HCA and *a priori* CA cases between -5 and +30 of hospital admission in children aged 1-5 years (jittered data points) 155
Figure 5.18. Distribution of predicted HCA, predicted HCA/HA, predicted CA, *a priori* CA and *defined* HA BSI between -5 and +30 of hospital admission in children aged 1 month < 1 year (jittered data points) 160
Figure 5.19. Distribution of the predicted HCA, predicted HCA/HA, predicted CA, *a priori* CA and *defined* HA BSI between -5 and +30 of hospital admission in children aged 1-5 years (jittered data points) 161
Figure 5.20. Plot of the multiple correspondence analysis results between dimension 1 and 2 for all the explanatory variables; children aged 1 month < 1year 165
Figure 5.21. Plot of the multiple correspondence analysis results between dimension 1 and 2 for all the explanatory variables; children aged 1-5 years 166
Figure 5.22. Scatter plot illustrating the values for dimension 1 and 2 assigned to each child, aged 1 month < 1 year, in the multiple correspondence analysis 167
Figure 5.23. Scatter plot illustrating the values for dimension 1 and 2 assigned to each child, aged 1-5 years, in the multiple correspondence analysis 168
Figure 6.1. Steps taken in this thesis (in the box) to inform preventive and treatment strategies for children with BSI dependent on the prediction capability of the models 174
Figure 6.2. Flow diagram of the steps taken (grey) and those that will be addressed in this chapter (red) to predict the proportion of CA, HCA and HA BSI in children 176
Figure 6.3. Proportion of BSI categorised as CA, HCA or HA for approaches A, B and C (including sensitivity analyses) by age-group and HA threshold 178
Figure 6.4. Comparing model classification of children with CA, HCA or HA BSI (using day 2 of hospital admission as the threshold for *defined* HA BSI) subdivided according to specimens taken before 2 days after admission and 2 or more days after admission for children aged (i) 1 month < 1 year, (ii) 1-5 years 181
Figure 6.5. Clinical/demographic characteristics and time used to predict the source of infection 182
Figure 6.6. The predicted source of infection used to predict the causative pathogen 187

Figure 6.7. Distribution of BSI by pathogen and timing of the positive specimen in relation to hospital admission for children aged 1 month < 1 year; triangles and circles distinguish adjacent pathogens from each other; the dashed line indicates the HA2 threshold.....192
Figure 6.8. Distribution of BSI by pathogen and timing of the positive specimen in relation to hospital admission for children aged 1 – 5 years; triangles and circles distinguish adjacent pathogens from each other; the dashed line indicates the HA2 threshold.....193

Tables

Table 1.1. Criteria for CA, HCA and HA BSI in children 29
Table 1.2. Proportions of BSI that are categorised as CA, HCA and HA in children 32
Table 1.3. Proportions of BSI due to specific organisms that are categorised as CA, HCA and HA BSI in children 34
Table 2.1. Core and notifiable organisms to be reported to LabBase2 47
Table 3.1. Match-weights for each patient identifier for MW-All and MW-Child 80
Table 3.2. Number and proportion of linked OPIEs for MW-All and MW-Child per year for children aged 1 month – 5 years in England 88
Table 3.3. Number of OPIEs successfully linked to HES data using MW-All and MW-Child out of the total number OPIEs extracted from LabBase2 for children aged 1 month–5 years with a BSI, England 90
Table 4.1. Categories for clinical and demographic exposure variables 109
Table 4.2. Ratio of BSI for children aged 1 month < 1 year compared to children aged 1-5 years by year of age 112
Table 4.3 Rate of BSI by 1,000 bed-days and by 100,000 population, adjusted for the approximate 80% voluntary reporting of laboratory data to LabBase2 112
Table 4.4. Susceptibility of child: Study population demographic and clinical characteristics by age-group 114
Table 4.5. In-patient exposure to infection: study population demographic and clinical characteristics by age-group 116
Table 5.1. Number of children with *defined* HA BSI for each of the nine threshold scenarios by age-group 128
Table 5.2. Tabulation of *defined* HA1 and HA9 cases by demographic and clinical characteristics from HES in the 12 months prior to the BSI 129
Table 5.3. Distribution of the number of children with *a priori* CA BSI by day of bacterial isolation in relation to hospital admission (day 0) 130
Table 5.4. Tabulation of *a priori* CA cases by demographic and clinical characteristics 132
Table 5.5. Summary of methodological steps and description of approaches A, B and C 136
Table 5.6 Additional descriptive breakdown of children with the additional categorical variables for approach C 142
Table 5.7 Unadjusted OR results for explanatory variables used in models A and B with defined HA BSI (2 days) as the outcome in children aged 1 month < 1 year 145
Table 5.8 Unadjusted OR results for explanatory variables used in models A and B with defined HA BSI (2 days) as the outcome in children aged 1-5 years 146
Table 5.9. *Defined* HA BSI and predicted HA BSI for the range of cut-off days (HA1-HA9) 148
Table 5.10. Approach A: Predicted proportions of CA BSI, HCA BSI and HA BSI in children aged 1 month < 1 year 153
Table 5.11. Approach A: Predicted proportions of CA BSI, HCA BSI and HA BSI in children aged 1-5 years 153
Table 5.12. Approach B: Predicted proportions of CA BSI, HCA BSI and HA BSI in children aged 1 month < 1 year, n=4,415 159
Table 5.13. Approach B: Predicted proportions of CA BSI, HCA BSI and HA BSI in children aged 1-5 years, n=6,172 159
Table 5.14 The MCA dimensions and the proportional inertia that they account for 163
Table 5.15. Cross-tabulation of the K-medians groups with the *defined* HA, *a priori* CA and the remaining mixed BSI-cases by age-group 170
Table 6.1. Final distribution of CA, HCA and HA BSI by approach A, B and C for children aged 1 month < 1 year using the 2-day *defined* HA threshold 179
Table 6.2. Final distribution of CA, HCA and HA BSI by approach A, B and C for children aged 1-5 years using the 2-day *defined* HA threshold 179

Table 6.3. Distribution of demographic and clinical characteristics by approach A, B and C for children aged 1 month < 1 year (HA2 threshold) 184
Table 6.4. Distribution of demographic and clinical characteristics by approach A, B and C for children aged 1 - 5 years (HA2 threshold) 186
Table 6.5. Proportional distribution of BSI by pathogen and timing of the positive specimen in relation to hospital admission for children aged 1 month < 1 year 189
Table 6.6. Proportional distribution of BSI by pathogen and timing of the positive specimen in relation to hospital admission for children aged 1-5 years.....190
Table 6.7. BSI causing pathogen breakdown by approach A, B and C for children aged 1 month < 1 year (percentage calculated by pathogen between CA, HCA and HA groups).....196
Table 6.8 BSI causing pathogen breakdown by approach A, B and C for children aged 1 month < 1 year (percentage calculated for all pathogens within each individual CA, HCA and HA group) 197
Table 6.9. BSI causing pathogen breakdown by approach A, B and C for children aged 1-5years (percentage calculated by pathogen between CA, HCA and HA groups)..199
Table 6.10 BSI causing pathogen breakdown by approach A, B and C for children aged 1-5 years (percentage calculated for all pathogens within each individual CA, HCA and HA group) 200
Table 6.11 Results of proportion of CA, HCA and HA BSI using the day 2 *defined* HA threshold when CoNS, *Micrococcus* spp. and diphtheroids were removed for children aged 1 month < 1 year.....202
Table 6.12 Results of proportion of CA, HCA and HA BSI using the day 2 *defined* HA threshold when CoNS, *Micrococcus* spp. and diphtheroids were removed for children aged 1-5 years 202

Appendices

Appendix Chapter 1a: Peer-reviewed literature review published in Journal of Hospital Infection 234

Appendix Chapter 1b: PubMed search strategy 246

Appendix Chapter 4a: ICD-10 codes identifying chronic conditions in children..... 247

Appendix Chapter 4b: OPCS codes identifying indwelling devices..... 249

Appendix Chapter 5: Summary figures depicting all the AIC and BIC values from the approach A and approach B model..... 250

Appendix Chapter 6: Peer-reviewed article in Epidemiology and Infection (2015) 262

Appendix Chapter 7: Conference presentations 268

Abbreviations

A&E	Accident and Emergency
AIC	Akaike's Information Criterion
BIC	Baysian Information Criterion
BNF-C	British National Formulary for Children
BPAIIG	British Paediatric Allergy, Immunology and Infectious Diseases Group
BPSU	British Paediatric Surveillance Unit
BSI	Bloodstream Infection
CA	Community-Acquired
CDC	Center for Disease Control and Prevention
CoNS	Coagulase-negative staphylococcus
CVC	Central Venous Catheter
CHI	Community Health Index
CI	Confidence Interval
DBS	Demographic Batch Service
DoH	Department of Health
ECOSS	Electronic Communication of Surveillance in Scotland
GAS	Group A streptococcus
GBS	Group B streptococcus
GP	General Practitioner
HA	Hospital-Acquired
HCA	Healthcare-Associated
HCAI & AMR	Healthcare Associated Infections and Antimicrobial Resistance
HES	Hospital Episode Statistics
HPA	Health Protection Agency
HSCIC	Health and Social Care Information Centre
ICD-9	International Classification of Disease version 9
ICD-10	International Classification of Disease version 10
IgG	Immunoglobulin G (antibody)
IP	Inpatient
MRSA	Methicillin-Resistant *Staphylococcus aureus*
MW	Match-weight
MW-All	Match-weight all ages
MW-Child	Match-weight child

NEQAS	National External Quality Assessment Service
NHS	National Health Service
NIGB	National Information Governance Board
ONS	Office for National Statistics
OP	Outpatient
OPCS-4	Office of Population, Censuses and Surveys: Classification of Interventions and Procedures, 4th Revision
OPIE	Organism Patient Illness Episode
OR	Odds ratio
PDS	Personal Demographics Service
PHE	Public Health England
PHLS	Public Health Laboratory Service
PICU	Paediatric Intensive Care Unit
PMG	Paediatric Microbiologists' Group
UK	United Kingdom
US	United States

Chapter 1. Background, aims and objectives

1.1 Introduction

Bloodstream infections (BSIs) in children are serious, clinical conditions associated with high morbidity and mortality, that require prompt initiation of antibiotic therapy.(1) However, when a patient presents with a clinical diagnosis of BSI, the causative pathogen and its antibiotic susceptibility will not be known for approximately the first two days of treatment due to the time required to grow cultures from the blood samples, identify the bacterial species and test their antibiotic susceptibilities.(2;3) Hence, empirical selection of the most likely effective antibiotic treatment is dependent on knowledge of the relative frequency of the main invasive bacterial pathogens and their likely antibiotic susceptibility. To this end, clinicians have used definitions such as community-acquired (CA) and hospital-acquired (HA) BSIs to guide initial antibiotic treatment due to the inherent differences in pathogens isolated from these two patient groups. These definitions have been based on the timing of infection in relation to hospital admission: HA usually being defined as more than 48 hours after admission and CA as an infection within 48 hours of admission.(4-6) However, changes in the way healthcare is delivered outside of the hospital setting mean that an investigation into whether these definitions are still appropriate to adequately predict and distinguish the bacterial specimens or the antibiotic sensitivities of the bacteria causing BSIs in children is needed.(7-9) In September 2013, the UK Government released a five-year (2013-2018) strategy in to address the increasing threat of antimicrobial resistance and promote prudent prescribing. As children consume a large proportion of the total antibiotics administered in England, the strategy highlights the importance and timeliness of this thesis to investigate how best to differentiate invasive pathogens for the purposes of ensuring that the most appropriate empiric antibiotic therapy is administered.(10;11)

The work presented herein to address the aims and objectives of this thesis (described later in this chapter) was undertaken between October 2011 and June 2016.

1.1.1 Structure and content of Chapter 1

This chapter describes the historical context of the treatment of BSIs in children and aims to investigate what information is available to distinguish the locations where BSIs are acquired to make inferences about likely patterns of antibiotic susceptibility. The current literature is reviewed and discussed, and results of a survey of professional opinions from clinicians and microbiologists are reported. The findings are used to underpin the rationale for the aims and objectives of my thesis, which are presented at the end of this chapter.

1.2 Bloodstream infections (BSIs)

Local infections develop when bacteria are able to penetrate the body's natural skin and mucosal barriers and evade immune defences, sometimes via accidental or medical intervention routes (invasive procedures). BSIs can subsequently develop when the bacteria causing a local infection spread into the bloodstream. Transient, asymptomatic BSIs are common, for example, after tooth brushing.(12) The spread of bacteria via the bloodstream can infect other organs in the body and thereby increase the severity of infection due to the greater bacteria-dependent tissue damage, and the broader and stronger immune response required to combat the infection. The earlier that effective antibiotic treatment is given, the lower the risk of bacterial spread, organ damage and death due to the infection. According to the Surviving Sepsis Campaign recommendations, patients should receive antibiotic therapy within one hour of a severe sepsis diagnosis.(13) Clinically, there is a difference between the definition of sepsis and BSIs. Sepsis is the presence of clinical signs and symptoms of bacterial bloodstream infection requiring medical attention, whereas BSIs (bacteraemia) are defined by microbiologic methods, i.e. detection of bacteria in the blood.(13) Sepsis is theoretically a subset of bacteraemia, but in practice in some cases of sepsis, no bacteria can be cultured from the blood. The clinical symptoms observed in patients with a BSI can range from no clinical symptoms to symptoms of sepsis such as high fever, organ failure and septic shock.(1)

Antibiotics to treat bacterial infections have been available for medical use since 1932.(14) Although increases in antibiotic resistance due to widespread antibiotic use driving the development of resistance have been noted for several decades, it is only in recent years that resistance has become a major threat due to the paucity of new antibiotics in development.(10;11) The continued use of antibiotics can result in selection of resistant pathogens, resulting in the emergence and spread of antibiotic resistant infections and loss of treatment effectiveness. This is why appropriate empirical treatment given between the time of the blood sample and the laboratory results is crucial, and why the proxy indicator of the

timing of the infection in relation to hospital admission has been used in the past to predict the most likely causative pathogens.

This thesis focuses on BSIs for several reasons. Firstly, patients with BSIs deteriorate rapidly and require quick and appropriate treatment. When a clinician clinically assesses a patient, they need to decide whether or not their patient has symptoms consistent with suspected BSI; if they do suspect a BSI, they will then have to take a blood sample and make the next decision about what antibiotics to treat with whilst awaiting the laboratory results of the blood culture. To guide their choice of empiric antibiotic treatment before the laboratory results are available, clinicians need a simple way to predict the most likely invasive pathogen(s).(1;11)

Secondly, blood culture samples should be taken from all patients who have a suspected serious bacterial infection, acting as an indicator of severity, and blood samples are rarely taken as a routine from patients without symptoms.

Thirdly, apart from when a patient has a BSI, bacteria should not be in the bloodstream as it is a sterile environment. Whereas bacteria can be found at many other sites (e.g. urine, bronchial lavage, urine) because of high rates of contamination ((e.g. Coagulase-negative staphylococci (CoNS), *Corynebacterium* spp., Viridans group streptococci, *Bacillus* spp.) or because of ongoing colonisation.(15-19) For these reasons, BSIs are widely used for national monitoring of serious bacterial infections.(20)

1.2.1 Community-acquired (CA) and hospital-acquired (HA) BSI

The time interval used for distinguishing CA from HA infections is based on pathophysiological reasoning, namely the average time when bacteria will become detectable in a blood sample after the onset of infection and represents a pragmatic cut-off, although observations allowing a distinct transition between CA and HA BSI are limited.(5;21;22) The original CA and HA categories reflected a natural differentiation between patient groups in terms of the pathogens acquired (and their antibiotic susceptibility) in the community or in hospital.(23) In the past, patients needing intensive healthcare procedures were often managed as hospital inpatients (IP), often for prolonged periods as there was little provision of appropriate care outside the hospital environment. CA bacterial infections were therefore predominantly due to organisms that were relatively virulent pathogens capable of infecting otherwise healthy people in the community, although patients in hospital were also susceptible to these infections.(24) In contrast, HA infections were predominantly due to organisms that flourished among patients prone to infection due to a number of pre-disposing factors including:

underlying co-morbidities, compromised immunity or invasive procedures that breached the body's external defences allowing pathogens direct entry. Pathogens circulating in the hospital are also influenced by the selective pressure from intensive antibiotic use that encourages the emergence of resistant strains.(25-27)

1.2.2 Healthcare-associated (HCA) BSIs

Since the initial introduction of the CA and HA categories by the US Center for Disease Control and Prevention (CDC) in the 1970s, advances in healthcare and delivery have changed considerably.(4) For example, congenital and chronic conditions in infants are being diagnosed earlier, extending the life expectancy due to the improvement in treatment.(28) This change in the delivery of care has resulted in the reduction in the length of hospital stay due to the introduction of clinical pathways that increasingly allow patients with complex conditions requiring invasive devices to be managed outside hospitals, either at out-patient (OP) clinics or at home. For example, the OP parenteral antibiotic therapy service enables patients to administer their own antibiotic treatment intravenously at home.(29)

The type of care for children has changed. More complex conditions are increasingly being managed in the community rather than in hospital. The type of care that can be provided in the home, is also expanding in range and complexity.(30) The national charity for children's health care, now named Action for Sick Children, changed their name in 1991 from the National Association for the Welfare of Children in Hospital due to the recognition of the growing delivery of healthcare in the community.(31) One example of healthcare delivery in the community is hospital at home services that are accessible to children with life-limiting conditions.(32) In addition, the Children's Community Nursing teams are active in England and provide short and long term care for all children requiring healthcare in a community setting. Nursing is provided for conditions that include invasive devices, such as intravenous therapy, gastrostomy tube feeding and tracheostomy care and therefore present a potential risk for HCA infections.(33) To quantify the increase in delivery of community healthcare in England, by the 1993 there were 62 generic and 124 specialised services, of which 15% had been established in the previous year, increasing slightly to 192 services (proportion of generic to specialised unknown) by 2007.(34-36) Patients from the community presenting to hospital with a BSI now include a mix of previously healthy individuals and patients with long-term illnesses who may have had frequent hospital contact, OP visits or care at home.

The changing distribution of bacterial pathogens causing BSIs in different settings is well illustrated by the example of methicillin-resistant *Staphylococcus aureus* (MRSA). After the discovery of MRSA in 1961, which was regarded as causing HA BSI, this pathogen later emerged in the community setting causing CA BSI.(37) CA-MRSA infections began occurring in the 1990s in people with little or no hospital contact (e.g. skin infections in people playing close contact sports) or commonly associated risk factors for MRSA, and were found to exhibit different molecular and antibiotic susceptibility characteristics.(38) Today, CA-MRSA has gone "full circle" and is now responsible for a proportion of HCA and HA infections as well as CA infections in children and adults.(39) However, attempts to identify patients who acquire their infection from previous healthcare contact (healthcare-associated [HCA] infections) have concentrated on adult patients.(9;23;40-43) In adults, the main risk factor used to differentiate between CA and HA BSI has been the time interval between hospital admission and a blood culture sample that yields a positive bacterial culture. However, variable combinations of risk factors have been used *a priori* to define HCA BSI, for example, the presence of previous invasive devices, recent discharge from hospital, and underlying chronic conditions requiring invasive treatment.(9;23;40-43)

1.2.3 BSI in children

Children are particularly susceptible to infections due to their immature immune systems, especially in the first few years of life. A proportion will also be born premature, or born with a congenital condition requiring additional medical care, or both.(1;44;45) In the UK, the incidence of BSIs (for both uncommon and recurrent BSIs) in children aged 1 month-1 year was 10-fold higher (362/100,000 population) than in children aged 5-15 years (36/100,000 population) between 1998 and 2007.(44) Similar differences in rates between age-groups have also been reported in other countries.(8;46-48)

Measures have been taken to reduce the rates of BSI by targeting certain pathogens using effective vaccines in the UK.(44;45;49-51) The *Haemophilus influenzae* type B vaccine was introduced in 1992,(52) the meningococcal C vaccine was introduced in 1999,(53) and the 7-valent conjugate pneumococcal vaccine was introduced in 2006, subsequently replaced by the 13-valent conjugate pneumococcal vaccine in 2010.(54) However, an increase in infections by serotypes not covered by the pneumococcal vaccine has been observed, and BSIs caused by other pathogens are still an important burden.(55)

In the current British National Formulary for Children (BNF-C) aged between 1 month and 18 years, guidance for the empirical treatment of BSIs is given for CA and HA BSI, but there is no mention of HCA BSI.(56) There are, however, a number of additional antibiotics suggested to be given in addition to the first line treatment if certain pathogens are suspected, but there is no guidance to suggest how clinicians should suspect different invasive pathogen species that might require a different antibiotic combination. To ensure that national guidance is up to date, it is crucial that there is evidence to support this guidance to reflect the current epidemiology of BSI in children.

Figure 1.1 illustrates the patient pathway and the decision to treat a BSI both for a patient admitted from the community with a suspected infection, and for a patient who develops an infection in hospital. Previous recent healthcare contact, known co-morbidities, surgery or catheter insertions in a child who is in hospital (or had previous recent healthcare contact) may affect the decision to use different antibiotics compared to a patient admitted from the community without previous healthcare contact. At a local, clinical level, other factors such as case-mix, population demographics, travel history and local antibiotic-resistance patterns may also influence a clinician's choice of antibiotics.

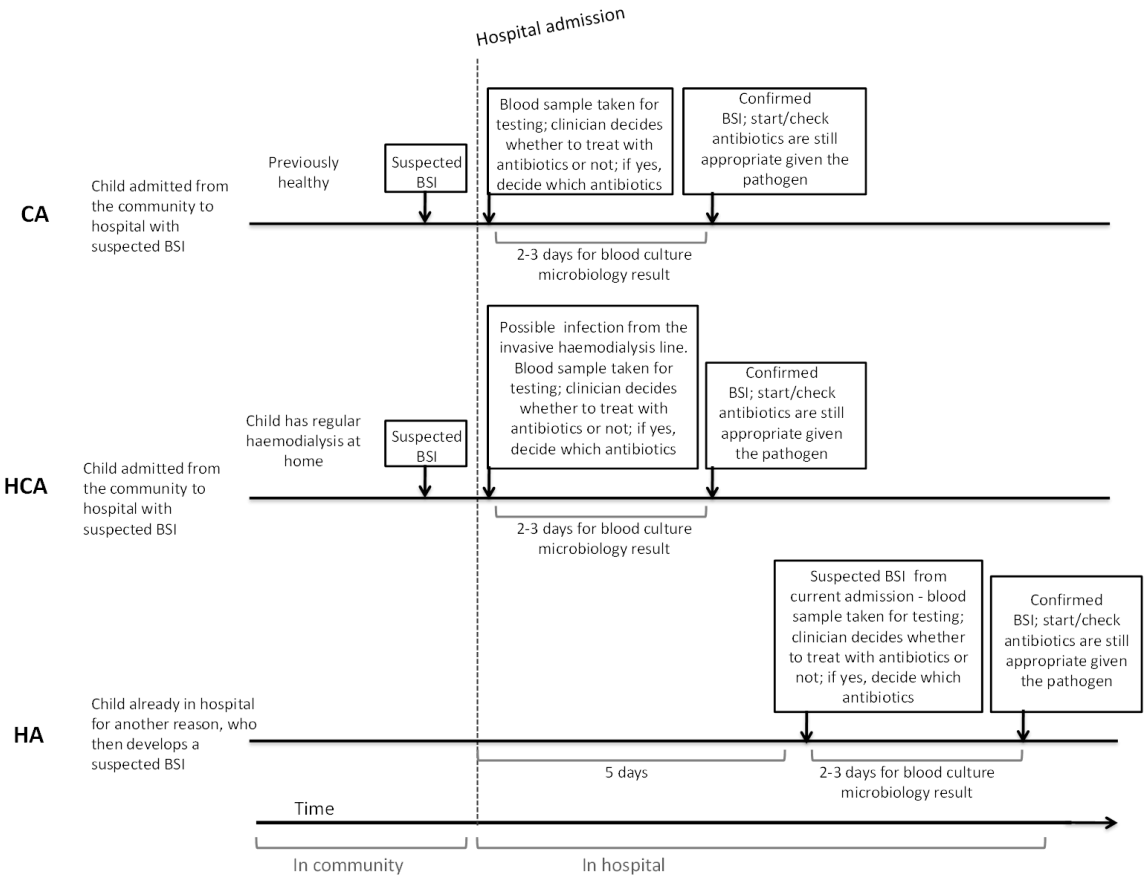


Figure 1.1 Flow diagram to illustrate the process of study selection using the inclusion criteria

1.2.4 Previous Public Health England work on children with BSI

Previous surveillance of bacterial infections in children, undertaken by Public Health England (PHE), identified national changes in the most frequently isolated pathogens over time (children aged 1 month – 18 years), and from a one-hospital study, the small proportion of children (aged 0-28 days) with a positive blood culture from all blood cultures tested.(44;57) Major limitations of these analyses, however, include the lack of clinical data to supplement the microbiology data, a lack of information as to where these changes in aetiology were taking place (i.e. the community or hospital setting) and whether there was a particular clinical focus, for example a specific group of patients who represented the majority of those with BSI. Until 2011, the PHE Department of Healthcare-Associated Infections and Antimicrobial Resistance (HCAI & AMR) did not hold the relevant data permissions to link routinely collected microbiology data voluntarily submitted by laboratories across England, Wales and Northern Ireland to a national clinical dataset for England, which is why this work has not previously been undertaken.

In an attempt to address this lack of clinical data, previous projects have used point prevalence surveys or the British Paediatric Surveillance Unit (BPSU) to enhance the microbiology data with clinical information.(58;59) However, these studies have either focused on a specific short period of time that does not allow for longer term surveillance, on specific bacterial groups, such as pathogens causing bacterial meningitis or on a small geographic region.

However, from 2011, the Department of HCAI & AMR obtained the relevant permissions to enable the access to and linkage of Hospital Episode Statistics (HES) data to the national microbiology dataset (described in Chapter 2), opening up the possibility for investigating BSI in children from a combined microbiology and clinical standpoint.

1.2.5 Existing Evidence

To investigate whether the CA and HA BSI categories are still appropriate for differentiating the likely source of pathogens causing BSI and thereby, to guide antibiotic treatment, I performed a systematic literature review. The aim was to evaluate the consistency of criteria used for CA and HA BSI, and whether a category of HCA BSI was considered in describing paediatric BSI. If identified, different studies were compared to determine if there was general agreement about the type of pathogens, the prevalence of pathogens isolated and whether they fit into one or multiple categories, based on timing.

1.3 Systematic literature review

I conducted a systematic review to answer three key questions:

(1) What are the criteria for CA, HCA and HA BSIs in children?
(2) What proportion of BSIs are categorised as CA, HCA or HA in children?
(3) What proportion of BSIs due to specific organisms are categorised as CA, HCA or HA in children?

A copy of the peer-reviewed literature review manuscript, published in the Journal of Hospital Infection, can be found in Appendix 1a.

1.3.1 Methods

1.3.1.1 Selection criteria: Q1. What are the criteria for CA, HCA and HA BSIs in children?

A systematic review was conducted to extract the criteria for CA, HCA and HA BSIs reported in the published research literature and on websites for the UK Department of Health (DoH) and the CDC.

Inclusion criteria comprised any study or review article that defined criteria for both CA and HA in children (<19 years old) intended for high income countries (based on the World Bank classifications)(60); low income settings were excluded because susceptibility of children to infection, the range of causative pathogens and thresholds for access to healthcare differ. We restricted studies to those published after 1988, the date when the widely accepted CDC guidelines for CA and HA were first published.(5) Included studies had to report criteria for a mixed population of children attending hospital with a possible BSI, not criteria restricted to children with a specific condition (e.g. cancer). Multi-organism studies reporting CA and HA BSIs were selected, and both single and multi-organism studies reporting HCA BSI were selected. I excluded studies that were not written in English to minimise time and translation costs.

I searched PubMed from 1988 to 6th March 2012 using synonyms for four concepts: CA infection, and HA or HCA infection, and BSI, and child (see Appendix 1b). I (KLH) screened titles and abstracts of potentially eligible studies. All potential studies were scrutinised by me and a second reviewer (Ruth Gilbert), and full papers were retrieved for studies considered eligible by both reviewers. Studies reported in multiple papers were counted as a single study. I

extracted data, which were checked by the second reviewer (Ruth Gilbert) against the source article.

1.3.1.2 Selection criteria: Q2. What proportion of BSIs are categorised as CA, HCA or HA in children?

A subset of cross-sectional or cohort studies identified by the systematic review (Figure 1.2) that reported CA and HA categories for more than 50 BSI episodes in children were analysed to ensure sufficient numbers for sub-group analysis. Data were extracted on the type of hospital(s) and age-group in each study and on the number of BSI episodes by CA, HA and HCA categories.

1.3.1.3 Selection criteria: Q3. What proportion of BSIs due to specific organisms are categorised as CA, HCA or HA in children?

A subset of studies identified by the systematic review (Figure 1.2) that reported organisms for at least two of the CA, HA and HCA categories in children were compared and analysed.

1.3.2 Results

1.3.2.1 Q1. What are the criteria for CA, HCA and HA BSIs in children?

The search strategy retrieved 303 papers from PubMed (Appendix 1b), of which 23 were unique studies from 12 countries (Figure 1.2).(8;61-82) Two of these studies were restricted to a single pathogen (MRSA) but reported results for HCA BSI.(81;82) Table 1.1 shows the criteria used by the 23 studies to distinguish CA and HA BSI (and HCA where mentioned) according to date of publication.

Thirteen variations of CA BSI and 15 variations of HA BSI were defined by the 23 studies, of which four focused exclusively on paediatric patients; these results are summarised in Table 1.1. Fourteen studies based their criteria on those reported in seven different sources (CDC counted as one source).(4-6;9;83-89) The two oldest studies defined CA and HA BSI by whether the infection was present or not on admission. Over time, the criteria for HA BSI shifted to be based on a positive blood-culture taken at either ≥48 hours, >48 hours or ≥72 hours after hospital admission, of which the two former were most frequently applied (14 studies). Several variations to these time factors included post-discharge infections and prior hospital contact. One study used the number of days since admission instead of the number of hours.(82) Three studies used criteria that focused on clinical symptoms of BSI only; timing of the BSI was not

defined in their criteria.(63;64;79) From 1999 onwards, five studies defined criteria for HCA and included criteria for the timing and type of previous healthcare contact received prior to the admission relevant to the BSI sample.

England's DoH defined BSI as HA if the positive blood sample was taken on or after day 3 of admission to hospital (with day of admission being day 1),(90) which is a proxy for the 48 hour threshold (Table 1.1). In contrast, the most recent definition from CDC replaces the previous "nosocomial" term for HA infection with "health care-associated infection" in the acute care setting.(91) The CDC definition states that infection must not be present or incubating at time of admission, but no time-specific cut-off was defined, unlike its 1988 predecessor. It also states that the population for which the term clinical sepsis is used is restricted to patients ≤1 year old, indicating that paediatric patients were considered and that the guidelines hence apply to all ages.

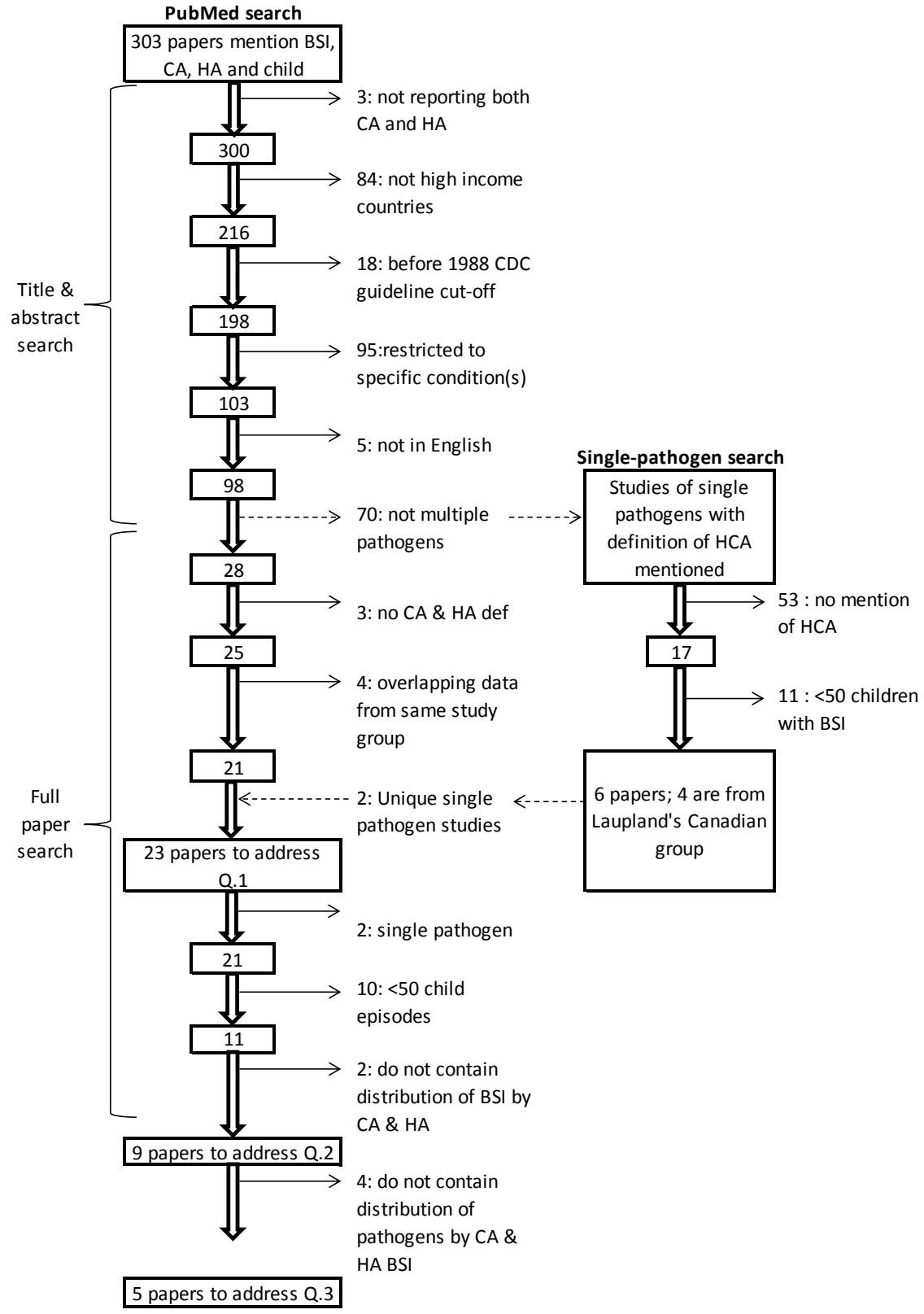


Figure 1.2. Flow diagram to illustrate the process of study selection using the inclusion criteria

Table 1.1. Criteria for CA, HCA and HA BSI in children

Authors	Study period	Country	Pop[n]	Definition of the presumed location where infection acquired			References directly relating to CA, HCA, HA criteria
				CA	HCA	HA	
Hounsom *et al.*(61)	2007-2008	UK	C&A	• BSI from sample taken ≤48hrs after admission	• BSI from sample taken ≤48hrs after admission and • intravenous therapy, or • haemodialysis within 30 days, or • hospitalised in prior 90 days	• BSI from sample taken>48hrs after admission (assumed)	Friedman *et al.* 2002(9)
Bloomfield &Briant (62)	2006-2007	New Zealand	C&A	• BSI from sample taken ≤48hrs after admission, unless organism with long incubation period (e.g. *Salmonella Typhi*) and •not HCA	• BSI from sample taken ≤48hrs after admission, and • acquired in hospital & BSI not present/ incubating on admission, or • complication of indwelling device, or • BSI <30 days after surgery, or • incision related to BSI performed <48hrs before onset of infection, or • associated with neutropenia by cytotoxic therapy	• BSI from sample taken >48hrs after admission or •BSI<48 hrs after discharge	Australian Infection Control Association, modified from CDC(89) & PHLS, UK(84)
Ang *et al.*(63)	2006	UK	C	•Symptom onset ≤72hrs after admission		•Symptom onset >72hrs after admission	None
Nicoletti *et al.*(64)	2002-2003	Italy	C&A	•Symptom onset <72 hrs after admission		•Symptom onset ≥72hrs after admission	None
Raymond *et al.*(65)	2001	New Zealand	C&A	• BSI from sample taken ≤48hrs after admission, unless organism with long incubation period (e.g. *Salmonella Typhi*) is isolated, and • not HCA	• BSI from sample taken ≤48hrs after admission, and • acquired during hospitalisation & BSI not present/ incubating on admission, or • complication of indwelling device, or occurring <30 days after procedure, or • incision related to BSI performed <48hrs before onset of infection, or •associated with neutropenia by cytotoxic therapy	• BSI from sample taken>48hrs after admission, or •BSI<48 hrs after discharge	Australian Infection Control Association, modified from CDC(89) & PHLS, UK(84)
Laupland *et al.*(8)	2000-2006	Canada	C	• BSI from sample taken ≤48hrs after admission or BSI from sample taken >48 hrs after discharge, or • not HCA	• BSI from sample taken ≤48hrs after admission, and •attended hospital-based clinic or A&E dept. in prior 2-30days, or • admitted to acute care hospital for 2≤ days in prior 2-90 days	• BSI from sample taken>48hrs after admission or • BSI from sample taken<48 hrs after discharge	CDC, 1996(6)

Note: C=children; A=adult; BSI=bloodstream infection; OP=outpatient; IP=in-patient; CVC=central venous catheter

Authors	Study period	Country	Pop[n]	Definition of the presumed location where infection acquired			References directly relating to CA, HCA, HA criteria
				CA	HCA	HA	
Douglas *et al.*(66)	2000	Australia	C&A	• BSI from sample taken ≤48hrs after admission		• BSI from sample taken>48hrs after admission, or • BSI<7 days after discharge, or • Intensive OP therapy with regular hospital contact	adapted from CDC, 1996(6)
Wyllie *et al.*(67)	1999-2005	UK	C&A	• BSI from sample taken ≤48hrs after admission and • no hospital contact in last year	• BSI from sample taken ≤48hrs after admission and • previous IP hospital contact in last year	• BSI from sample taken>48hrs after admission (assumed)	CDC 1988(5) Friedman *et al*. 2002(9)
Marchaim *et al.*(68)	1997-2004	Israel	C&A	• BSI from sample taken ≤48hrs after admission, and • not hospitalized in preceding month		• BSI from sample taken >48hrs after admission	None
Luzzaro *et al.*(69)	1999-2000	Italy	C&A	• BSI from sample taken<72 hrs after admission		• BSI from sample taken≥72hrs after admission	None
Raveh *et al.*(70)	1998	Israel	C&A	• BSI from sample taken<72 hrs after admission		• BSI from sample taken ≥72hrs after admission, and • no relation between infection and original reason for admission	CDC1988(5) ; 1992 (88)
Khan *et al.*(71)	1997-1998	Qatar	C&A	• BSI from sample taken ≤48hrs after admission, and • not hospitalised in preceding month		• BSI from sample taken >48hrs after admission, or •BSI from sample taken<48hrs after discharge, or •preceded by invasive procedure, or related to CVC	None
Gray *et al.*(72)	1996-1999	UK	C	• BSI from sample taken ≤48hrs after admission		• BSI from sample taken>48hrs after admission	None
Jensen *et al.*(73)	1992-1993	Denmark	C&A	• BSI from sample taken ≤48hrs after admission		• BSI from sample taken>48hrs after admission, or • infections from invasive device	Weinstein *et al.* 1983(4)
McGregor &Collignon (74)	1990	Australia	C&A	• BSI from sample taken<48hrs after admission		• BSI from sample taken ≥48hrs after admission	CDC,1988(5)
Ashkenazi *et al.*(75)	1988-1991	Israel	C&A	• BSI from sample taken ≤48hrsafter admission		• BSI from sample taken>48hrs after admission	CDC, 1972(83)
Gomez *et al.*(76)	1988-1992	Spain	C&A	•"BSI using 72-hr rule"		•"BSI using 72-hr rule"	None
Iversen & Scheel(77)	1985, 1989	Norway	C&A	• BSI from sample taken ≤48hrs after admission		• BSI from sample taken >48hrs after admission	None
Cheng *et al.*(78)	1984-1989	Hong Kong	C&A	• BSI from sample taken ≤48hrs after admission		• BSI from sample taken >48hrsafter admission	Weinstein *et al.*1983(4)
McLaws *et al.*(79)	1984	Australia	C&A	• BSI from clinical evidence on admission		• BSI from hospitalised patient & no clinical evidence of BSI on admission	None
Eykyn *et al.*(80)	1969-1988	UK	C&A	• BSI before admission		• BSI >0hrs after admission	Gransden *et al.*1990(85)

Note: C=children; A=adult; BSI=bloodstream infection; OP=outpatient; IP=in-patient; CVC=central venous catheter

Single pathogen studies

Authors	Study period	Country	Pop[n]	Definition of the presumed location where infection acquired			References directly relating to CA, HCA, HA criteria
				CA	HCA	HA	
Kempker *et al.*(81)	2005-2008	US	C&A	• <48hrs after admission, and •no documented healthcare risk-factors	• <48hrs after admission, and • 1 or more: CVC, or •history of MRSA infection/colonization, or • surgery/dialysis, or • long-term care facility residence in prior year	• >48hrs after admission, and • 1 or more: CVC, or • history of MRSA infection/colonization, or • surgery/dialysis, or • long-term care facility residence in prior year	Klevens *et al.* 2006(86)
Kallen *et al.*(82)	2005-2008	US	C&A	• BSI<4days after admission(day 1), or BSI from OP, and no recent healthcare exposure	• BSI acquired from OP, or <4days after admission (day 1), and •presence of CVC at admission, or • 1 of following in prior year: • overnight stay in acute care/long-term care facility, or •dialysis or surgery (MRSA colonisation does not count)	• BSI ≥4days	Cohen *et al.* 2008(87)

Websites

Authors	Study period	Country	Pop[n]	Definition of the presumed location where infection acquired			References directly relating to CA, HCA, HA criteria
				CA	HCA	HA	
Dept. of Health(90)		UK		Trust-apportioned, BSI on day 3 ≤ of hospital admission(admission=day 1)	-	Non-trust apportioned, BSI on 1st or 2nd day of hospital admission;	
CDC (Horan et al. 2008)(91)		US			Health care-associated infection (previously termed 'nosocomial'): •No evidence that the infection was present or incubating at time of hospital admission and •occurs during hospital admission		

Note: C=children; A=adult; BSI=bloodstream infection; OP=outpatient; IP=inpatient; CVC=central venous catheter

1.3.2.2 Q2: What proportion of BSIs are categorised as CA, HCA and HA in children?

Nine studies (Table 1.2) met the selection criteria outlined in Figure 1.2. All nine studies included hospitals with specialised paediatric units (e.g. cancer unit). The mean proportion of BSIs in children considered to be CA was 50% (95% CI: 48%-52%) when including neonates. The two extremes were from Gray *et al.* who reported <20% CA BSI in a study that focused only on children in paediatric intensive care units (PICU),(72) and from the study from Qatar by Khan *et al.*, where the epidemiology may have been different and potentially influenced by the small sample size (n=63).

The mean proportion of BSIs in children that were CA increased to 58% (CI: 55%-61%) when neonates were excluded from three studies (where age-group was differentiated in the study).(8;65;78) Laupland *et al.* was the only study to report all three categories of CA, HCA and HA BSIs, and defined 26% of their reported paediatric BSIs as HCA (when both including and excluding neonates).(8)

Table 1.2. Proportions of BSI that are categorised as CA, HCA and HA in children

Authors	Study period	Country	Hospital type or region	Age	No. of BSI	No. of BSI by category (%)		
						CA	HCA	HA
Raymond *et al.*(65)	2001	New Zealand	2 hospitals (secondary & tertiary care)	0<20yrs *30d<20yrs*	120 episodes *70 episodes*	40 (33%) *40 (57%)*	¥ 80 (67%) ¥ *30 (43%)*	¥ ¥
Laupland *et al.*(8)	2000-2006	Canada	Calgary region	0<18yrs *30d<18yrs*	995 episodes *815 episodes*	475 (48%) *451 (55%)*	256 (26%) *209 (26%)*	264 (26%) *155 (19%)*
Wyllie *et al.* (67)	1999-2005	UK	2 hospitals	2m<24yrs	112 episodes	55 (49%)	57 (51%)	-
Khan *et al.*(71)	1997-1998	Qatar	3 hospitals (tertiary referral centre)	0<15yrs	63 patients	51 (81%)	-	12 (19%)
Gray *et al.*(72)	1996-1999	UK	1 Children's hospital	0<18yrs	143 isolates	27 (19%)	-	116 (81%)
Ashkenazi *et al.*(75)	1988-1991	Israel	1 university hospital & tertiary centre	0<19yrs	339 episodes	191 (56%)	-	148 (44%)
Iversen & Scheel(77)	1985, 1989	Norway	1 general teaching hospital	0<16yrs	92 episodes	45 (49%)	-	47 (51%)
Cheng *et al.*(78)	1984-1989	Hong Kong	1 regional university hospital & tertiary local hospitals	0<16yrs *29d<16yrs*	344 episodes *243 episodes*	188 (55%) *168 (69%)*	-	156 (45%) *75 (31%)*
Eykyn *et al.*	1969-1988	UK	1 general teaching hospital	0<16yrs	367* episodes	199 (54%)	-	168 (46%)

¥ combined HCA and HA results - given same definition

* approximate number extracted from graphical data

1.3.2.3 Q3: What proportion of BSIs due to specific organisms are categorised as CA, HCA, HA BSI in children?

Five studies (Table 1.3) met the selection criteria outlined in Figure 1.2. Overall, the majority of BSIs due to *S. pneumoniae* (mean CA=78%, 95% CI: 72%-82%) and *Salmonella* spp. (mean CA=93%, 95% CI 0.85%-0.98%) were categorised as CA BSI, and the majority of infections due to CoNS (mean HA=90%, 95% CI: 82-94%) and *Enterococcus* spp. (mean HA=84%, 95% CI: 72-92%) BSI were classified as HA. *S. aureus*, *E. coli*, *Pseudomonas* spp., *Klebsiella* spp. and *Enterobacter* spp. did not clearly fit into either the CA or HA BSI categories. These five pathogens represented a substantial proportion of paediatric HCA infections in the Laupland e*t al.* study, which was the only study to report HCA.

1.3.3 Key findings

This is the first overview of the published literature on the criteria used to differentiate between CA, HCA and HA BSI in children. The review revealed variable use of sampling after a 48-hour cut-off following hospital admission to categorise HA BSI and the increasing but inconsistent use of criteria to define HCA and re-define CA in recent years. Few pathogens were clearly categorised into CA or HA categories in the review and where defined, HCA accounted for a substantial proportion of BSIs in children.

Of the 23 studies in the review that included children in the study population, just under half of these addressed the distribution of CA and HA (question 2) with only one study addressing CA, HCA and HA BSI, and only five studies addressed the distribution of organisms in CA and HA BSI (question 3) where the numbers of children per study were also small. The problems of distinguishing between CA, HCA and HA are well illustrated by studies of MRSA in adults and children. A meta-analysis by Salgado *et al.* (2003) investigated criteria used to define CA-MRSA in adults and children. They found eight different groups of criteria for CA-MRSA, distributed over 31 studies. The range of criteria included positive cultures in samples taken <24 hours (2 studies), <48 hours (16 studies) and <72 hours (11 studies) after admission, each with and without other exclusion criteria including healthcare-associated risk factors. Eight studies did not report CA-MRSA criteria.(92)

Table 1.3. Proportions of BSI due to specific organisms that are categorised as CA, HCA and HA BSI in children

Authors	Study Period	Organism	Age	BSI (%)			No. of BSI	Total no. of BSI in study
				CA	HCA	HA		
Laupland *et al.*(8)	2000-2006	*S. pneumoniae*	0<18yrs	76%	21%	3%	159	1,033
Wyllie *et al.*(67)	1999-2005	*S. pneumoniae*	2m<24yrs	*50%*	*50%*	-	40	112
Gray *et al.*(72)	1996-1999	*S. pneumoniae*	0<18yrs	100%	-	0%	3	143
Ashkenazi *et al.*(75)	1988-1991	*S. pneumoniae*	0<19yrs	91%	-	9%	22	355
Cheng *et al.*(78)	1984-1989	*S. pneumoniae*	0<16yrs	100%	-	0%	43	371
Laupland *et al.*(8)	2000-2006	*Salmonella* spp.	0<18yrs	87%	4%	9%	23	1,033
Cheng *et al.*(78)	1984-1989	*Salmonella* spp.	0<16yrs	96%	-	4%	53	371
Laupland *et al.*(8)	2000-2006	*S. aureus*	0<18yrs	56%	17%	27%	120	1,033
Wyllie *et al.*(67)	1999-2005	*S. aureus*	2m<24yrs	*38%**	*63%**	-	32	112
Gray *et al.*(72)	1996-1999	*S. aureus*	0<18yrs	10%	-	90%	20	143
Ashkenazi *et al.*(75)	1988-1991	*S. aureus*	0<19yrs	51%	-	49%	37	355
Cheng *et al.*(78)	1984-1989	*S. aureus*	0<16yrs	47%	-	53%	34	371
Laupland *et al.*(8)	2000-2006	*E.coli*	0<18yrs	39%	40%	21%	119	1,033
Wyllie *et al.*(67)	1999-2005	*E.coli*	2m<24yrs	*57%*	*43%*	-	40	112
Gray *et al.*(72)	1996-1999	*E.coli*	0<18yrs	25%	-	75%	8	143
Ashkenazi *et al.*(75)	1988-1991	*E.coli*	0<19yrs	73%	-	27%	30	355
Cheng *et al.*(78)	1984-1989	*E.coli*	0<16yrs	78%	-	22%	27	371
Laupland *et al.*(8)	2000-2006	*Pseudomonas* spp.	0<18yrs	11%	39%	50%	18	1,033
Gray *et al.*(72)	1996-1999	*Pseudomonas* spp.	0<18yrs	38%	-	62%	8	143
Ashkenazi *et al.*(75)	1988-1991	*Pseudomonas* spp.	0<19yrs	52%	-	48%	31	355
Cheng *et al.*(78)	1984-1989	*Pseudomonas* spp.	0<16yrs	25%	-	75%	20	371
Laupland *et al.*(8)	2000-2006	*Enterobacter* spp.	0<18yrs	10%	10%	80%	20	1,033
Gray *et al.*(72)	1996-1999	*Enterobacter* spp.	0<18yrs	0%	-	100%	3	143
Ashkenazi *et al.*(75)	1988-1991	*Enterobacter* spp.	0<19yrs	54%	-	46%	24	355
Laupland *et al.*(8)	2000-2006	*Klebsiella* spp.	0<18yrs	15%	39%	46%	26	1,033
Gray *et al.*(72)	1996-1999	*Klebsiella* spp.	0<18yrs	0%	-	100%	9	143
Ashkenazi *et al.*(75)	1988-1991	*Klebsiella* spp.	0<19yrs	29%	-	71%	41	355
Cheng *et al.*(78)	1984-1989	*Klebsiella* spp.	0<16yrs	50%	-	50%	8	371
Laupland *et al.*(8)	2000-2006	CoNS	0<18yrs	8%	9%	83%	52	1,033
Gray *et al.*(72)	1996-1999	CoNS	0<18yrs	0%	-	100%	45	143
Cheng *et al.*(78)	1984-1989	CoNS	0<16yrs	17%	-	83%	18	371
Laupland *et al.*(8)	2000-2006	*Enterococcus* spp.	0<18yrs	13%	23%	64%	31	1,033
Gray *et al.*(72)	1996-1999	*Enterococcus* spp.	0<18yrs	0%	-	100%	15	143
Ashkenazi *et al.*(75)	1988-1991	*Enterococcus* spp.	0<19yrs	25%	-	75%	12	355
Cheng *et al.*(78)	1984-1989	*Enterococcus* spp.	0<16yrs	0%	-	100%	3	371

*CA=12 MSSA; HA=3 MRSA & 17 MSSA

Similarly, other studies in adults focusing on infections acquired before hospital admission have reported a wide variety of clinical criteria, in addition to the timing of the infection, used to distinguish HCA BSI from CA BSI.(9;23;40;41) There is a lack of national data to support

changing the BNF-C guidelines, as observed by the systematic literature review, from a focus of CA and HA BSIs only, to including the changes in paediatric BSI epidemiology and the increasing role of HCA BSI.(56) A subsequent paper by Gray, which was excluded from the systematic review because the data overlapped with Gray's earlier article, reports the overall proportion of CA BSI (51.8%) and HA BSI (48.2%).(7) However, as opposed to the first paper, Gray chose to differentiate the organisms causing BSI according to whether they were from children with serious underlying chronic medical conditions versus those without, rather than as CA or HA BSI, because the majority of BSIs occurred in patients with regular hospital contact and serious underlying chronic conditions. For example, of all the CA BSI identified in his study (n=1,224), only 11.9% were in children who had previously been healthy.(7)

Clinicians deciding on empirical treatment, before blood culture results are known, can use their clinical judgement and additional information, such as co-morbidity, recent healthcare contacts or a history of invasive procedures, to make an informed decision about appropriate antibiotic treatment.

To enhance the information obtained in the systematic review and gain more information from front line staff treating children with BSI, I investigated what pathogens clinicians and microbiologists perceived to be causing HA BSI in the UK.

1.4 Professional opinion survey: what pathogens causing paediatric BSI are judged to be (or not be) HA BSI in the UK?

To capture professional opinions from across the UK, I surveyed frontline paediatricians and paediatric microbiologists, via an online survey (screenshot example is shown in Figure 1.3), to elicit their opinions on the proportion of BSI episodes they believed to be HA for 12 frequently isolated BSI causing pathogens.

1.4.1 Methods

The aim of the questionnaire was to ascertain clinicians' views on what proportion of BSIs in children, by type of bacteria, they judged to be HA. I surveyed clinicians by sending them a brief questionnaire via email to ask them, based on their experience, what proportion of BSI due to 12 different pathogens that commonly cause BSI would be HA in children aged 3 months – 5 years. Neonates aged <3 months were excluded to avoid vertically transmitted BSI and to maximise the likelihood of having time at home to enable exposure to CA infections. The 12 pathogens included in the survey were: *S. aureus* (split into MRSA and methicillin-

susceptible *S. aureus* [MSSA] where a methicillin susceptibility result was reported), non-pyogenic streptococci, *Enterococcus* spp., group B streptococcus (GBS; *Streptococcus agalactiae*), *Streptococcus pneumoniae*, group A streptococcus (GAS; *Streptococcus pyogenes*), *Escherichia coli*, *Klebsiella* spp., *Enterobacter* spp., *Pseudomonas aeruginosa*, *Salmonella* spp., and *Neisseria meningitidis*.

I developed the online survey using the survey facilities of 'Eval & Go' (www.evalandgo.com), one of the very few companies to offer a sliding scale of 21 points (0%-100% in 5% increments) within a survey design. Each of the 12 pathogens had a sliding scale, positioned at 50%, where the clinician or microbiologist could slide the scale up or down depending on their chosen estimate. Survey questions included the geographic region of work and whether the respondents were trainees or consultants. The survey was open for completion from February 2012 onwards and was closed in May 2012 once two consecutive weeks of no new responses had elapsed. A screenshot of the survey can be seen in Figure 1.3. No definition of HA BSI was given so that individual professionals could report their estimates according to their definition.

Permission to distribute this survey to the national paediatrician and paediatric microbiologist professionals was sought from the British Paediatric Allergy, Immunology and Infectious Diseases Group (BPAIIG) and the Paediatric Microbiologists' Group (PMG) respectively. A sample survey was sent to a contact from each group for checking. Agreement was reached with both parties, after having piloted the survey on three microbiologists and two paediatricians (who did not go on to complete the final survey) with the understanding that the surveys were anonymous and no follow up would be possible. These central contacts emailed the link to the survey with a brief introduction about the study to their respective members: 286 members (107 consultants and 179 trainees) of BPAIIG and 25 members of PMG.

1.4.2 Results

Between February and May 2012, 34 responses to the web-based survey were received. The respondents comprised 9 paediatric microbiologists (8 consultants; 1 trainee) from five English regions (East Midlands, London, South-East, West Midlands, and Yorkshire and the Humber) and 25 paediatricians (7 consultants, 17 trainee paediatricians, 1 not disclosed) from Scotland and eight English regions (London, North-East, North-West, South-East, South-West, Wales, West Midlands, Yorkshire and the Humber). Frequency distributions of the estimates of the

proportion of patients seen with these BSIs considered by the respondents to be HA are shown in Figure 1.4a-m.

Microbiologists agreed that few children (5% or less) with *S. pneumoniae* (median: 5%), GAS (median: 5%), *N. meningitidis* (median: 0%), or *Salmonella* spp. (median: 0%) BSI had a HA infection. Paediatricians showed similar agreement for these pathogens but the range of estimated proportions of BSIs due to HA were more variable (range: 0-50%). The majority of microbiologists and paediatricians considered GBS (Figure 1.4e) to cause BSIs that were not HA (median: 5%, range: 0-75%). For the remaining pathogens (Figure 1.4a-d,h-k), there were no consistent responses in the estimated proportions of BSI that were likely to be HA (range 0-100%).

A few of the respondents added extra comments at the end of the survey. From the paediatrician responses, as no definition for HA BSI was given at the beginning of the survey, one trainee reported that they were using a time threshold for any infection occurring >72-hours after admission as HA BSI. Two other trainee clinicians reported that their percentages for HA BSI were very rough estimates, where one had 'no good guess' for either *Enterobacter* spp. or *Enterococcus* spp., and the other commented that these percentages would vary from hospital to hospital depending on the classification as a tertiary or specialist centre for example.

From the microbiologist responses, one consultant discussed how a large proportion (unquantified) of patients with BSIs comprise regular hospital attenders; from this I inferred that many of the cases that this consultant witnessed are most likely HCA BSI cases. They also reported that as the numbers of BSI cases within their own hospitals are relatively small, it is hard to estimate an overall percentage. Another microbiologist gave either a 0% or 5% contribution to HA BSI for all 12 pathogens, justifying their results by explaining that they had observed very few HA BSIs in this age-group over the past 2 years.

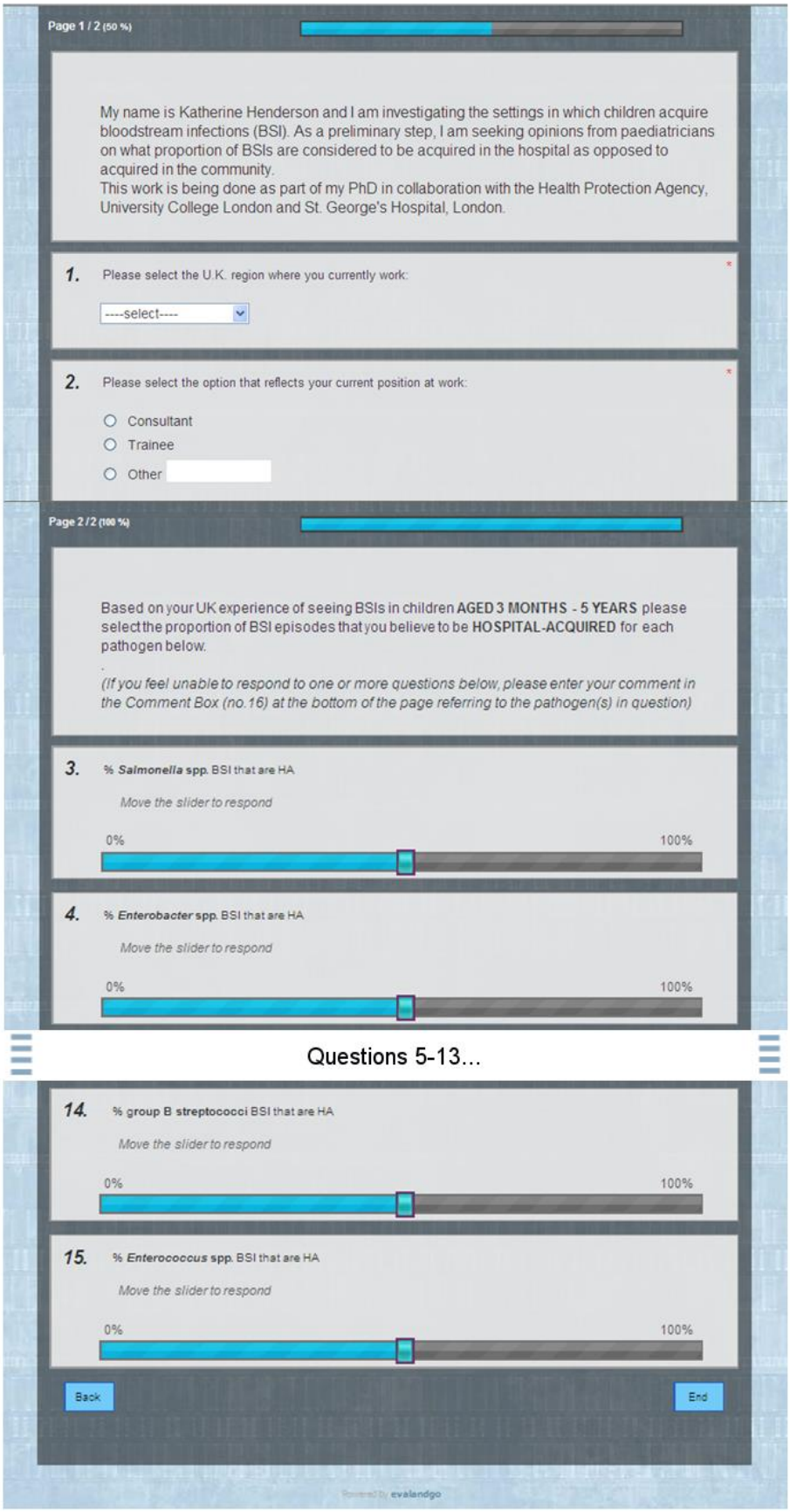
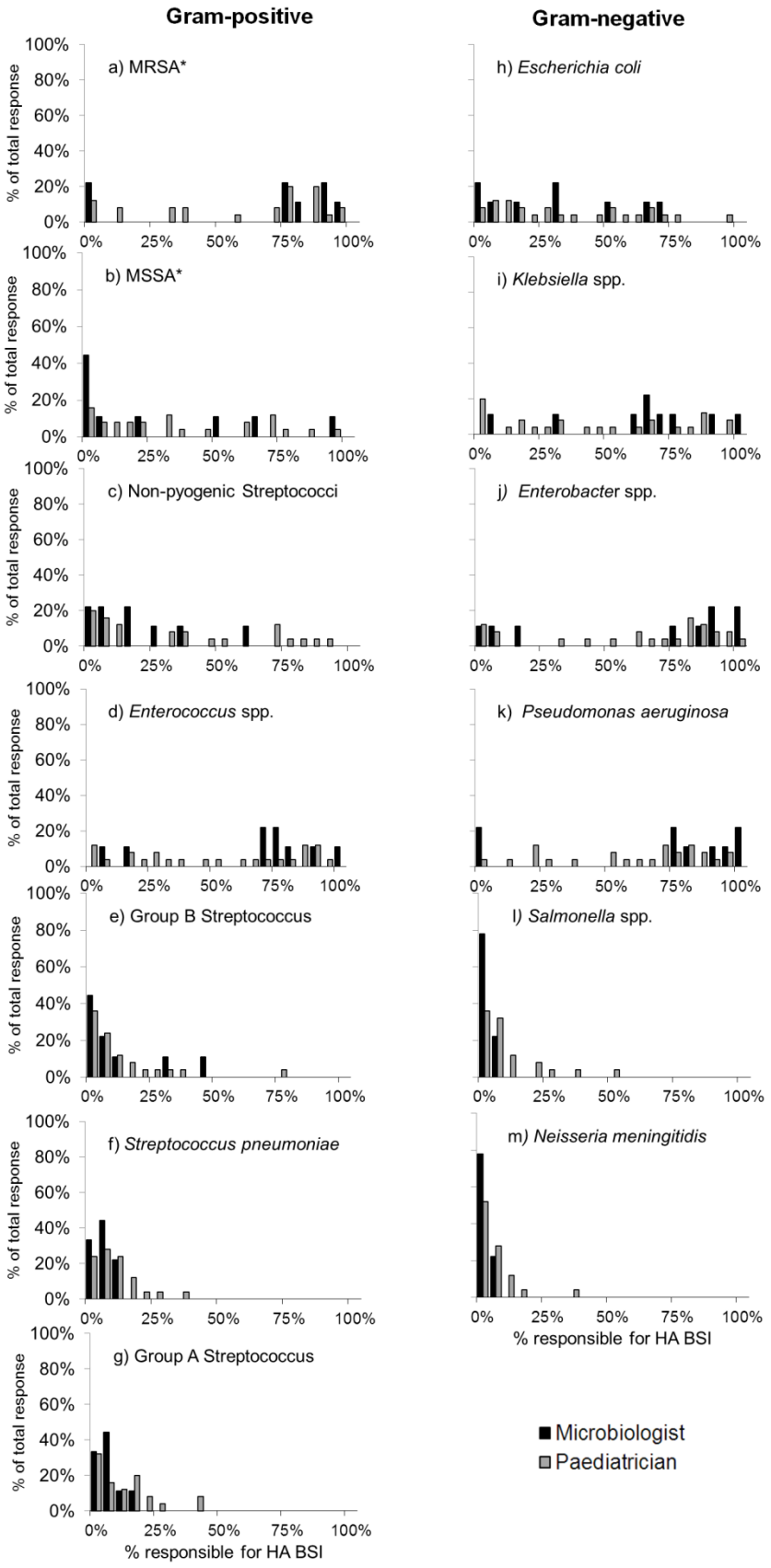


Figure 1.3. Screenshot example of the online professional opinion survey



*MRSA = methicillin-resistant *Staphylococcus aureus*; MSSA = methicillin-susceptible *Staphylococcus aureus*; HA-BSI = Hospital-acquired bloodstream infection

Figure 1.4a-m. Professional opinion survey results from UK Paediatrician and Microbiologists for 12 pathogens for the percentage of BSIs caused by each pathogen, in their opinion, is likely to be HA

1.4.3 Key findings

The professional opinion survey demonstrated consistency in identifying those pathogens causing BSI that were likely not to be HA (*S. pneumoniae*, GAS and GBS, *Salmonella* spp. and *N. meningitidis)*. The wide range of estimated prevalence rates for pathogens historically considered to be HA, for example *Klebsiella* spp. and *Enterococcus* spp., may reflect the increasing uncertainty in the contribution of these pathogens to HA versus HCA BSI. From the responders' additional comments, there was awareness of HCA BSI as an emerging group of BSI. The general low numbers of BSIs at a hospital-level increased the level of uncertainty over the professionals' perceived levels of HA BSI causing pathogens.

A limitation of the professional opinion survey was the low response rate of paediatric microbiologists (36%) and paediatricians (9%). However, no follow-up emails to potential responders were permitted in the original agreement with the two groups. With the aim of keeping the survey short to increase response rates, we did not ask about different cut-off definitions for HA BSI.

1.5 Discussion

In both the literature review and the opinion survey, *S. pneumoniae*, GAS, *N. meningitidis* and *Salmonella* spp. were most associated with causing CA BSI, but for the majority of the other pathogens there was no clear distinction between their contribution to CA, (HCA in the literature review) and HA BSI.

The inclusion of HCA BSI in the more recent literature suggests that using timing of positive blood culture samples to differentiate between CA and HA BSI for paediatric patients is no longer valid, as the pathogens being isolated on admission to hospital include those previously associated with HA BSI.(8) The high prevalence of isolation of organisms previously considered to be HA is also reported by studies in the adult population.(9;42;93) The results of the systematic review highlight that there is no universal definition for CA, HCA or HA BSI in children. The variable definitions used have become more complex over time, alluding to the awareness of the increasing complexity of the epidemiology of paediatric BSI.

The change in the epidemiology of MRSA exemplified the shift of HA pathogens into the community. MRSA was previously considered a HA pathogen but more recent evidence has shown its isolation from patients on admission to hospital.(94) MRSA is transmitted from person to person predominantly via colonisation or from direct healthcare worker contact inadvertently spreading the pathogen between susceptible patients receiving healthcare

treatment. This is why there is a cycle of transmission from the hospital to the community setting and back to the hospital again, enabled by the movement of people in and out and hospital. Of the patients being admitted to hospital with MRSA, currently only a small fraction have true CA BSI, where the majority of the burden is in patients with HCA BSI.(92;94-97)

To further distinguish between CA and HA BSI, many authors have expressed the need for the exploration of the third HCA BSI category and the use of demographic and clinical characteristics to identify other factors other than time that can be used, which could be more clinically meaningful for guiding antibiotic therapy. (7;9;42;66-68;93)

Classification of CA, HCA and HA infection requires reassessment and a stronger evidence base to understand the current epidemiology of BSIs in children. A large, population-based study that includes sufficient numbers of infections and clinical characteristics of patients and data on recent contact with healthcare would offer a means of evaluating associations between the timing and type of infection, patient characteristics, co-morbidities and timing of invasive procedures. Due to the high costs of primary data collection and to avoid data duplication, the linkage of routinely collected laboratory and clinical data offers an efficient solution.

1.6 Rationale for PhD

In this PhD, I link HES IP data for England with national microbiology surveillance data to determine associations between clinical characteristics in the patient's hospitalisation record with CA, HCA and HA BSI. My analyses of the linked data aim to estimate the proportion of BSI attributable to CA, HCA and HA. The aim is to use the results to make inferences about the likely sources of invasive pathogens and patterns of antibiotic resistance. The information can be used to target BSI reducing interventions and to develop nationally representative algorithms to help guide empirical treatment for children with BSI. Figure 1.5 illustrates the data informing the empirical treatment, from the raw data where only timing of the BSI is used, to the laboratory confirmed antibiotic susceptibility data. The black boxes encapsulate what data I used in this thesis and the grey boxes highlight the data that I either did not have permissions to link to (Accident & Emergency [A&E] or OP data), did not have permission to access (prescribing data), does not currently exist (national collection of negative culture data) or data that was incomplete/insufficient (antimicrobial susceptibility) to analyse, which were not included in this thesis.

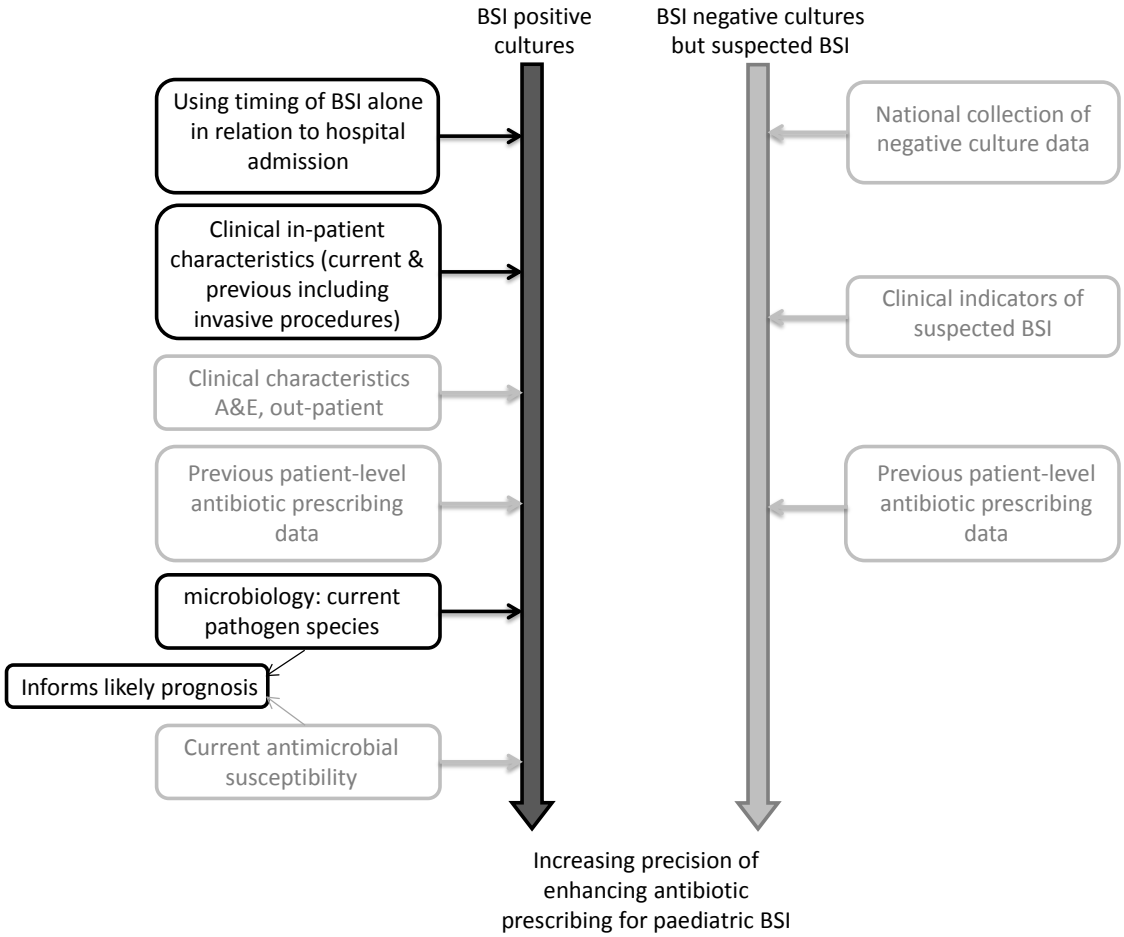


Figure 1.5. Flow diagram of steps to select appropriate antibiotic therapy for children with BSIs

1.6.1 Aim

To predict the relative proportion of CA, HCA and HA BSI in children aged 1 month to 5 years old, by identifying demographic and clinical characteristics associated with these three groups, using linked microbiology surveillance and clinical data for children admitted to National Health Service (NHS) hospitals in England.

1.6.2 Objectives

1. Link microbiology surveillance and clinical data and evaluate the effect of match-weight choice on the final linked dataset.

2. Develop statistical models to identify patient and clinical characteristics associated with CA, HCA and HA BSI in children.

3. Predict the proportion of CA, HCA and HA BSI occurring in children in England and the distribution of causative pathogens in each group.

1.6.3 Thesis structure

In **Chapter 2**, I describe the data sources available to address the aim of the thesis and define the cohort of children with BSI for analysis

In **Chapter 3**, I describe the attributes of different data linkage methods and identify the most appropriate method (probabilistic linkage) to link the data sources. I describe the process involved in linking the data sources, including the creation of match-weights, and subsequently create the cohort for analysis (Objective 1).

In **Chapter 4**, I report the indirect markers of child susceptibility and exposure to invasive BSI and the direct exposures to invasive BSI from the linked data and introduce the rationale for the statistical analysis in Chapter 5.

In **Chapter 5**, I propose three different statistical modelling approaches to predict the proportion of CA, HCA and HA BSI (Objective 2) and describe the assumptions made concerning the linked data. The results of each approach are reported separately, which are then compared in Chapter 6.

In **Chapter 6**, I compare and evaluate the predicted sources (CA, HCA, HA) of infection from the three statistical approaches (Objective 3). I then identify if the clinical and demographic characteristics and timing of blood specimen adequately predict the source of infection, and reveal the distribution of pathogens for CA, HCA and HA BSI for each of the three approaches.

Finally, **Chapter 7** summarises the key findings from the results, review the rationale for the thesis and discusses the limitations and implications of the results on clinical, public health and research practice in the future.

Chapter 2. Defining the cohort of children with BSIs for analysis

2.1 Introduction: data sources

The aim of this thesis is to estimate the relative proportion of CA, HCA and HA BSI in children in England with positive blood culture results reported to PHE. The data required to address this aim are located in two separate databases: the national voluntary microbiology surveillance database (LabBase2) and the NHS hospital administrative IP database (HES).

By linking these two datasets together, the resulting linked dataset will contain the timing of when a positive blood specimen was isolated from a patient in relation to the timing of admission to hospital and the identity of the causative pathogen. Additional information, including the longitudinal hospitalisation record detailing patients' underlying medical conditions, history of previous invasive procedures and hospital admissions are also collected.

I used a retrospective four-year cohort (April 2007 to March 2011) of linked LabBase2 and HES data of children with a BSI. The aspects considered in the selection of the cohort are described individually.

2.2 Structure and Content

This chapter will describe and summarise the individual strengths and limitations of the LabBase2 and HES data that may influence the data linkage, the data fields collected, the criteria used to determine what data are included into each database, and the coverage and quality of the data. The issues surrounding information governance and permissions to access the data will then be outlined. Finally, the selection of the patient cohort and the explanatory variables that were used to develop the statistical models to distinguish between CA, HCA and HA BSI will be described. Figure 2.1 illustrates Chapter 2 (in red) in context within the framework of the thesis.

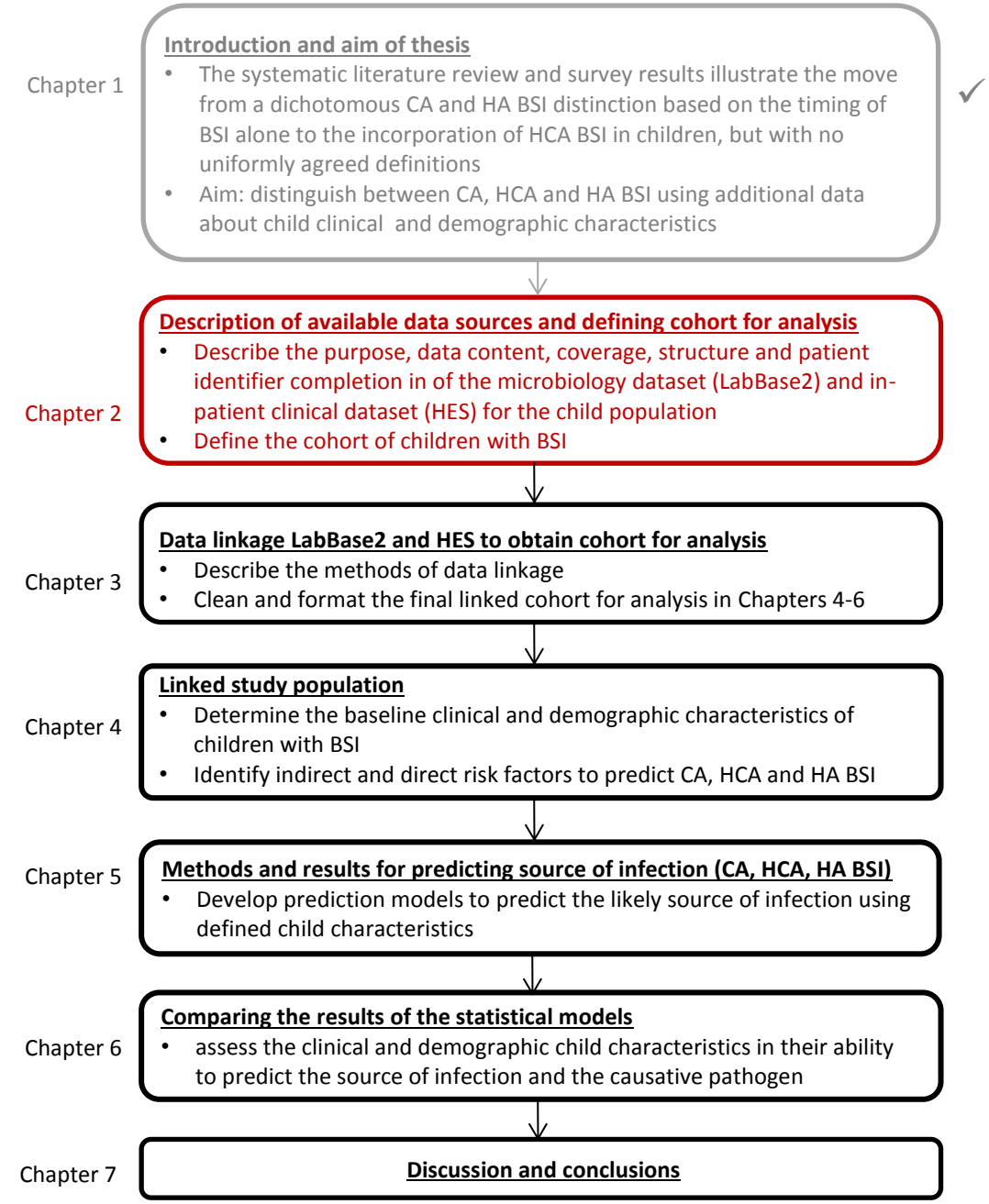


Figure 2.1. Flow diagram of the steps already addressed in previous chapters (grey), those that will be addressed in this Chapter (red) and the following chapters

2.3 LabBase2: Public Health England's surveillance system

2.3.1 Purpose

National microbiology data are voluntarily reported by participating laboratories to PHE (formally known as the Health Protection Agency [HPA]), and prior to that the Public Health Laboratory Service [PHLS]) and stored in a centralised database (LabBase2), which is held at the PHE site in Colindale, London.(98) This microbiology data forms the cornerstone of communicable disease surveillance in England, Wales and Northern Ireland; Scotland has its own microbiology surveillance database ECOSS (Electronic Communication of Surveillance in Scotland), which was not included in this thesis.(99)

LabBase2 contains reports of clinically significant isolates from different specimen types (e.g. blood, urine, respiratory isolates), organism identification to species level, antimicrobial susceptibility results (where tested and reported), the date the specimen was taken, the location where the specimen was taken (e.g. in hospital, at a general practitioners), the source laboratory, and patient demographic data (including patient identifiers).(100) There are no routinely recorded clinical data (e.g. contact with healthcare, diagnoses, procedures) in the voluntary reports to LabBase2. In addition, there are no data on the date of the patient's admission to hospital and therefore this database alone cannot be used to identify the timing of a BSI in relation to the date of hospital admission.(44)

2.3.2 Background

PHE began collecting results of positive clinical isolates via paper reports in the 1970s as PHLS from England, Wales and Northern Ireland for epidemiological surveillance and aiding outbreak investigation. In 1989, the first national microbiology database, LabBase1, was introduced and direct electronic reporting to this database from the testing laboratories began in 1993, first piloted in Wales.(98) This electronic reporting was expanded to a few regions in England in 1995, and then to all regions by 1997. This system developed and increased in capacity until 2002 when LabBase2 was introduced as a replacement to LabBase1.(101)

2.3.3 What data are reported to LabBase2?

The laboratory reporting to LabBase2 is voluntary. To ensure the continuity and representativeness of the data, a minimum standard of core pathogens isolated from

specimens is requested by the laboratory guidance document for participating laboratories.(102) These specimens must be deemed to be clinically significant, although the laboratory guidelines do not stipulate the criteria to be used for determining what is and is not clinically significant. However, there is no active validation of this process. LabBase2 does not collect any results for negative specimens (specimens that do not grow any bacteria).

In addition to these guidelines, the Health Protection (Notification) Regulations (introduced in October 2010) require a specified subset of core causative bacterial pathogens from human samples to be reported as notifiable organisms, because these bacteria were considered to cause clinically serious infections. The core and notifiable pathogens include the frequently isolated bacterial pathogens that this thesis investigated. These core and notifiable organisms are listed in Table 2.1.

Table 2.1. Core and notifiable organisms to be reported to LabBase2

Core Organisms
Acinetobacter spp., *Actinomyces* spp., *Aeromonas* spp., *Bartonella* spp., *Chlamydia pneuomniae, Chlamydia trachomatis, Citrobacter* spp., *Clostridium* spp., *Coxiella* spp., *Enterobacter* spp., *Enterococcus* spp., *Erysipelothrix* spp., *Escherichia coli, Klebsiella* spp., *Leptospira* spp., *Listeria* spp., *Morganella morganii, Mycobacterium* spp., *Mycoplasma* spp., *Neisseria gonorrhoeae, Nocardia* spp., *Pasteurella* spp., *Plesiomonas* spp., *Proteus* spp., *Providencia* spp., *Pseudomonas* spp., *Serratia* spp., *Staphylococcus aureus, Staphylococcus* spp., *Stenotrophomonas maltophilia, Streptobacillus moniliformis, Streptococcus* spp., *Treponema* spp., *Vibrio* spp., *Yersimina* spp. **Of which the following are notifiable organisms** *Bacillus anthracis, B. cereus, Bordetella pertussis, Berrelia* spp., *Brucella* spp., *Burkholderia mallei/pseudomallei, Campylobacter* spp., *Chlamydophila psittaci, Clostridium* spp., *Corynebacterium* spp., *Coxiella burnetii, Francisella tularenis, Haemophilus influenzae, Legionella* spp., *Leptospira interrogans, Listeria monocytogenes, Mycobacterium tuberculosis, Neisseria meningitidis, Rickettsia* spp., *Salmonella* spp., *Shigella* spp., *Streptococcus pneumoniae, S. pyogenes, Vibrio cholera, Yersinia pestis*

2.3.4 Capturing data on positive blood cultures

Local laboratory data are reported to LabBase2 through a national electronic network (Figure 2.2). The local laboratory reports of clinically significant positive specimen results are transferred from the local laboratory information system via CoSurv (a module system used for extracting data) using a LabLink+ interface, to the regional centres. The data are cleaned and

checked at the regional level before being submitted to the central LabBase2 database in Colindale, London.(102) When local laboratories need further testing of isolates, for example for speciation or antibiotic susceptibility profile, they send the isolate to one of the corresponding national reference laboratories. Results from isolates tested at the national reference laboratories feed straight into the LabBase2 system, bypassing CoSurv.

Data can be uploaded daily into LabBase2, which allows for timely analysis of current data. In a recent analysis of LabBase2 data, although the median reporting delay of isolate reports to LabBase2 was nine days (Inter-Quartile Range: 6-16 days) across all trusts, the variation by trust was substantial.(100) It was found that 45% of laboratories (75/167) reported isolates in a timely manner, whereas 12 trusts were responsible for the largest delays in the 90th centile of reporting (>100 days delay). The timeliness in reporting did improve after October 2010 when the additional Health Protection (Notification) Regulations were brought in.(102)

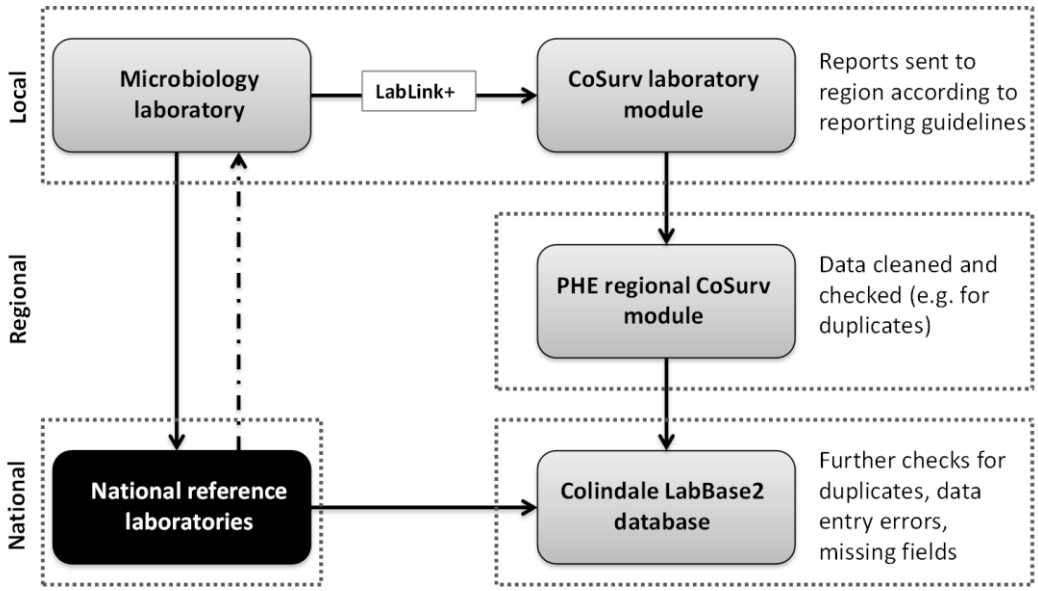


Figure 2.2. Flow diagram showing the path of the specimen results data from the local laboratories to the national LabBase2 database.

2.3.5 Coverage

As a voluntary system, LabBase2 does not receive all the positive bacterial specimen results. However, a parallel PHE mandatory reporting system for *S. aureus* and *E. coli* BSI can be used as a proxy to monitor the difference between the numbers of positive isolates reported to the two systems.(98;103) The trends of voluntary and mandatory reporting have been similar over time where the voluntary system received approximately 80% (*S. aureus*) and 84% (*E. coli*) of

the reports submitted to the mandatory surveillance system in 2011.(104;105) Whilst the proportion of completion for other pathogens in LabBase2 is not routinely investigated by PHE, these results are used as a proxy for the level of reporting in LabBase2.

To evaluate the estimated ascertainment of records in LabBase2, part of the work that Harron *et al.* undertook compared microbiology results from children (aged <16 years) that were eligible for reporting to LabBase2 and those that were retrievable, in reality, from LabBase2 between 2003-2010 for two gold-standard hospital laboratories (Birmingham Children's Hospital and Royal London Hospital). The estimated ascertainment was 81.5% and 79.5% for each laboratory respectively.(106) In the wider study of the 22 laboratories that could report PICU microbiology data to LabBase2, 548/2,596 (21%) laboratory-months were found to have incomplete reporting.(106) In her PhD thesis, Harron reports the difficulties associated with determining the process of data collection at the various stages of the data flow for LabBase2. The reasons for unreported and/or inconsistent reporting of data, which included data upload issues, information technology problems and variable staff support, indicates that the nature of the under-reporting is a system-level issue rather than a pathogen-level issue. The assumption was therefore made that the representativeness of the data by pathogen species should not be negatively impacted.(107)

2.3.6 Data structure

The LabBase2 data are formatted and stored as a line-list. Each row contains data specific to each blood specimen reported. A patient may have one or more rows for each infection episode depending on how many reported blood specimens (with different specimen dates) isolated the same pathogen.

2.3.7 Identification of infection episodes

Each record in LabBase2 is identified using an Organism Patient Illness Episode (OPIE). An OPIE represents an infection episode for a patient where a pathogen is isolated from a specimen and is considered clinically significant; it does not identify the patient uniquely within the database, where one patient may have multiple OPIEs. Every positive isolate record is assigned its own row in the database. Figure 2.3 demonstrates that the same patient (identified by the matching NHS number) is assigned two different OPIEs if the pathogen species isolated is different, even if they are from the same specimen type and isolated on the same date. LabBase2 has a standard time-period of 14 days for a bacterial infection episode (isolated from

blood and flagged as a bacteraemia), where any subsequent positive bacterial specimen taken within 14 days of the first specimen for the same pathogen for that patient is considered as part of the same infection, retaining the earliest specimen date in addition to the new specimen date for the new specimen.(102)

NHS no.	OPIE	Organism species	Earliest specimen date	Specimen date	Type
1111111111	AAAAA	*S. pneumoniae*	03/02/2010	03/02/2010	Sputum
1111111111	AAAAA	*S. pneumoniae*	03/02/2010	05/02/2010	Blood
1111111111	BBBBB	*H. influenzae*	03/02/2010	03/02/2010	Sputum
1111111111	BBBBB	*H. influenzae*	03/02/2010	05/02/2010	Blood

Figure 2.3. Example demonstrating that OPIEs are assigned based on the organism species

2.3.8 Matching BSI episode reports to the same patient episode

For an electronic isolate result to be accepted into the LabBase2 system and be assigned an OPIE, it must pass through a three-stage process. These three stages comprise the loading, de-duplication and merging of potential records.

2.3.8.1 Stage 1: Data Loading

An isolate report passes through a loading check process in order to be submitted from the local laboratory to the national LabBase2 database. It is presented to the system to be either accepted or rejected based on whether the report is deemed to be valid or not (e.g. the correct format). If invalid, the report may be corrected for a re-load attempt into the system, otherwise this is one stage where data are lost.(102) Once validated, the report is then tagged to be passed through to the duplicate identification stage.

2.3.8.2 Stage 2: Data de-duplication

Stage 1 tagged reports are individually compared against already processed records to identify duplicates that have to match on three levels:

a) Organism name
b) The same episode (within the 14-day period)
c) The same patient

For Levels a) and b), duplicates of an OPIE may occur when more than one positive specimen of the same species sub-group is isolated from any source (e.g. blood, urine etc.) from the same patient within a 14-day window. In contrast, if two different organisms, including different sub-types of a species, are isolated during the same 14-day episode, then two different OPIEs are assigned to the same patient to identify each organism as demonstrated in Figure 2.3. Equally, more than one OPIE is assigned if a patient has one or more re-infections with the same species outside of the primary infection time-period, implying that they have recovered after the first infection before getting the subsequent infection and so on.

For level c), to identify the same patient, the probability that two records match is assessed using the available patient identifiable data: patient name, soundex (phonetic algorithm of patient name), sex, date of birth, hospital number, NHS number, and postcode. This is performed by comparing every possible combinational pairing between the tagged reports and LabBase2 reports. A total match-weight is then given to each pair of records depending on how well (high weight) or poorly (low weight) they matched. Although this process maximises the potential matches by using a combination of patient identifiers to identify duplicates, those reports that have limited or no patient identifiers will not be merged with a current LabBase2 record, even if they are a real duplicate, due to a lack of information. This is one of the issues with regards to the voluntary nature of the dataset.

2.3.8.3 Stage 3: Data merging

Subsequent new records that are found to be duplicates of already existing records in the database are then merged together and assigned a new 'merged' OPIE, removing the original OPIE from the LabBase2 confirmed model. The original OPIEs are kept in a separate table should the records need to be unmerged at a later point.

To merge records successfully, the laboratory regulations state that at least one of the following patient identifiers must be present:(102)

- Patient's surname and initial, or soundex code and initial
- Patient's Hospital number
- Patient's NHS number

The regulations do state that although the patient identification details are passed onto the regional information managers by the reporting laboratories, the laboratories are able to suppress some identifiers before onward submission of the data. This is another step where data may be lost before reaching LabBase2.

2.3.9 Query design for LabBase2 data extraction

To ensure all the possible LabBase2 OPIEs were captured, a comprehensive query to extract data between April 2007 and March 2011 was created. The extraction criteria for the query and a fictitious data output example are outlined in Figure 2.4. This query extracted a total of 39,297 reports, accounting for 35,497 unique OPIEs. Within this extract, environmental isolates and National External Quality Assessment Service (NEQAS) test isolates were identified and removed, leaving a total of 35,555 reports accounting for 32,536 unique OPIEs.

Variables	Selected for	Example output
OPIE		01AAA11111/01
Organism name		HAEMOPHILUS INFLUENZAE
Earliest specimen date		DD/MM/YYYY
Specimen date	01/04/2007 - 31/03/2011	DD/MM/YYYY
Patient initial		R
Patient surname		RICKON
Patient forename		STARK
Patient NHS number		1111111111
Patient hospital number		ST22222
Sex		M
Date of birth	02/04/2001-31/03/2011	DD/MM/YYYY
Patient postcode prefix		WT1
Patient postcode suffix		3FL
Hospital name		NORTH HOSPITAL
Organism classification description	BACTERIA, FUNGI	BACTERIA
Age Group	0-9 YEARS OLD, UNKNOWN	0-9 YEARS OLD
Laboratory name		NORTH MICROBIOLOGY
Laboratory country	ENGLAND, WALES, SCOTLAND, ISLE OF MAN, CHANNEL ISLANDS, UNKNOWN	ENGLAND
Method of id description		CULTURE
Parent specimen type code	BLOOD, CSF	BLOOD
Specimen source type description		INPATIENT

Figure 2.4. Example of a LabBase2 query (fictitious data)

2.3.9.1 OPIE anomalies

Of the 32,536 unique OPIEs in the pre-linked LabBase2 dataset, 17 OPIEs (0.05%) had a >14 day difference between the first and last specimen date indicating that there may have been a data entry error. There is no way to check from the LabBase2-user access level, as the earliest specimen date and date loaded are all automatically duplicated from the first specimen in the OPIE to all subsequent specimens reported.

2.3.10 Patient identifiers in LabBase2

The patient identifiers reported in LabBase2 are patient name, soundex, NHS number, hospital number, sex, date of birth and postcode. To assess the completion of the identifier fields in LabBase2 that are also found in the HES dataset, the number of unique patient identifiers from positive blood specimen records from children aged 0 – 5 years old, as well as patients with an unknown date of birth (32,536 OPIEs) reported between April 2007 and March 2011, are plotted in Figure 2.5.

The level of completion was lowest for NHS number and patient postcode, ranging from 51%-70% and 65%-75% respectively over the four-year period. Date of birth and sex are the most complete (average of 95%), followed by hospital number (88-91%). Although the proportion of completion improves over time, none of the identifiers were 100% complete.

The completion does not necessarily reflect fully qualified data entries as some may contain errors. To try and overcome this potential error problem, additional data was retrieved and included as a result of tracing all the patient data through the Demographic Batch Service (DBS) to optimise NHS number and sex for the data linkage process described in more detail in the following paragraph.

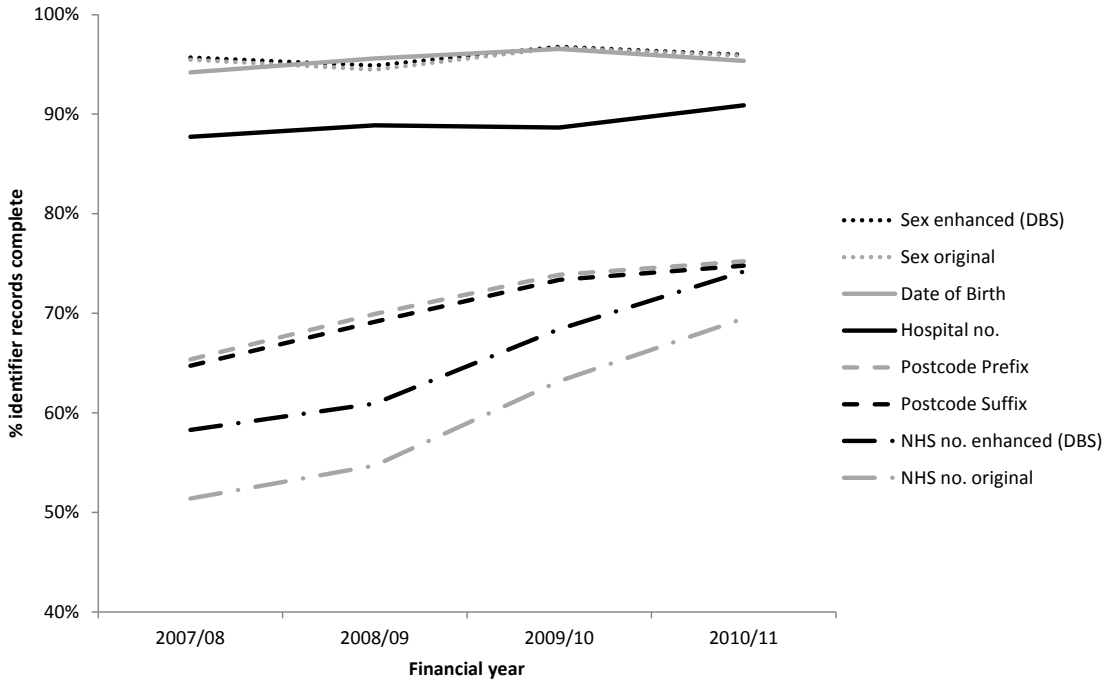


Figure 2.5. Proportion of identifier records with a data entry for each of the six patient identifiers in LabBase2 between 2007/08 and 2010/11 for children aged 0-5 years and patients with no date of birth

2.3.10.1 Demographic Batch Service (DBS) and data pre-processing

An additional method employed to enhance the data completion of the patient identifiers in the LabBase2 data is the DBS system. The DBS is used by the NHS and related organisations to trace files of patient data using the Personal Demographics Service (PDS).(108) One purpose is to obtain additional patient data for entries that may be missing, for example to find out the NHS number of a patient if one was not completed in the original file. The fields that are submitted include date of birth (mandatory), NHS number, surname, forename, gender and postcode. In order for a positive match to be returned, the uploaded file must contain one of the following:

1) a valid NHS number and date of birth; or
2) a valid NHS number, date of birth (2/3 of the day, month and year items need to match), surname (first three characters) and forename initial; or
3) surname, forename, date of birth and gender must match if no NHS number is given.

The patient identifier data for the 32,536 OPIEs (35,555 records) were then sent to the DBS for tracing to enhance the completion of the NHS number and sex fields.

Date of birth is crucial at all three stages of the trace; therefore no additional information was available for those without a date of birth (5%) in the LabBase2 data. No changes were made to any LabBase2 postcodes whether missing or different to the DBS patient postcode. This was to avoid changing any historical data that would be potentially incorrect at the time of the positive specimen, as patients can frequently move residence. Between LabBase2 and the DBS, 23,180 (65%) records matched perfectly out of the total 35,555 records (32,536 OPIEs) that were traced and did not require any further enhancement. Of the remaining 35% (12,375 records), I accepted new NHS numbers from the DBS trace where the LabBase2 entry was missing if they matched on:

- forename, surname and date of birth = 1,904 (15%) NHS numbers accepted
- surname, date of birth and postcode = 94 (0.8%) NHS numbers accepted
- forename, surname and date of birth (accounting for minor spelling errors) = 33 (0.3%) NHS numbers accepted

Sex was reassigned for 29 records where forename, surname, and date of birth matched.

2.3.10.2 Other patient identifier considerations

All patients who are registered with the NHS in England and Wales are assigned a unique, 10-digit identification NHS number. Patients admitted to hospital are assigned a hospital number for that hospital spell, which is unique to the hospital, but may not be unique between hospitals.

In LabBase2, the local General Practitioner (GP) postcode is used as a proxy if the patient's postcode is not entered, and if there is no GP postcode, then the system will default to the hospital postcode. How often a patient's family moves address may also provide more than one linkable postcode. Patients who are not current residents in the UK are assigned a postcode where the prefix is always "ZZ99" and the 3-digit suffix will denote the country or geographic area of residence.(109)

2.4 Hospital Episode Statistics (HES)

2.4.1 Purpose

The HES database holds routinely collected IP, OP and A&E patient data for England. NHS providers supply data for every patient (including private and foreign patients) attending an NHS hospital or an independent sector hospital funded by the NHS in England.(110) The IP database can capture over 100 pieces of data per patient episode that include data on hospital admission and discharge, patient identifiers, patient demographics, clinical diagnoses, operations, length of stay, socio-economic status and geographical location of care. HES does not contain microbiology data or specific information as to whether infections were acquired in the community or from hospital, although there may be indications from some treatment codes.(110) The original purpose of HES was to capture detailed patient records for episodes of care following an investigation into how hospital activity data were gathered and used.(110)

2.4.2 Background

The Health and Social Care Information Centre (HSCIC) is responsible for the HES national statistical data warehouse (http://www.hesonline.nhs.uk). HES was first established in 1987 to collect national IP data. Prior to this, only a small subset (10%) of patient admission data had been collected. The first complete year of data was the 1989-1990 HES dataset.(110)

The methods of data collection have been adapted over time to account for organisational change. Originally, the collation of data was at the regional level compared to the most recent change in 2006 to the National Programme for Information Technology's Secondary Uses Service.(111) Additional data fields have also been added over time which has offered flexibility to the amount of detail recorded [examples: expanding number of diagnoses fields to 20 and procedures to 24). OP and A&E datasets were introduced as additional experimental datasets reporting data from 2003-04 for OP and 2007-08 for A&E. The OP dataset became accredited as a national statistic in 2008; these datasets will not be considered in this study because both datasets were still experimental at the start of the cohort period, and had variable data completion.(110)

2.4.3 Data format

There are over 100 cleaning and validation rules that IP data are subject to in HES.(112) A certain proportion of the fields that undergo the validation rules are populated by data derived from other data sources within the IP dataset. For example, episode duration and patient age at start of episode are derived from the start and end episode dates and the difference between the patient's date of birth and the episode start date respectively.

The majority of HES fields are coded. Diagnoses fields are populated by ICD-10 (International Classification of Diseases, 10th Revision) codes, which replaced the ICD-9 codes in 1995. ICD-10 codes are a method of categorising clinical diagnoses within the healthcare environment for patient management, for surveillance purposes and can be used as a resource allocation tool. There are approximately 69,000 different ICD-10 codes that are categorised into 22 chapters. Of these chapters, 14 are dedicated to diseases affecting body systems including neoplasms, infectious diseases, parasitic diseases, mental disorders, circulatory and nervous system disorders. Three chapters are for pregnancy, perinatal conditions and malformations. The remaining five chapters cover external causes of harm such as poisoning or accidents, external causes of morbidity and mortality, factors affecting health status and special additional codes that come about from emerging health related diseases.(113)

Procedures and intervention fields are populated by OPCS-4 (Office of Population, Censuses and Surveys: Classification of Interventions and Procedures, 4th Revision) codes. These codes are entered by professionally trained coders who refer to the available patient hospital notes recorded by the medical staff in the hospitals.(114)

2.4.3.1 Coding error in HES

As a national dataset collecting millions of records each year, there is inevitably a margin of error. Although coding in HES has improved with time, coding errors can occur in the form of completely random errors, or vary between different hospitals.(115-117) A way of measuring error was introduced by the Audit Commission using the Payment by Results Data Assurance Framework that was introduced within the NHS in England between 2003 and 2005.(33) For example, the average clinical coding error in 2007/08 for healthcare resource groups was 9.4% (ranging from 0.3-52.0%), which had a subsequent economic effect of £3.5 million.(118) HES data has improved significantly over time, although the data completion and accuracy do vary by variable field.(119) Payment by results has incentivized coding for comorbidities as these attract additional payments for hospitals. This is unlikely to be related to BSI results and as PBr

has been in place since 2003, coding practices would not have changed substantially during the study period. There are also quality control systems in place to ensure adherence to coding rules. This has led to improvements in coding quality over time in HES and LabBase2.

2.4.4 Patient identifiers in HES

The patient identifiers reported in HES are sex, date of birth, hospital number, NHS number and postcode. The completion of the identifier fields in common with the LabBase2 dataset, for records from children aged 0 – 5 years old and patients with an unknown date of birth (4.8 million records) reported between 1st April 2007 and 31st March 2011, is illustrated in Figure 2.6. Sex, date of birth and hospital number were 100% complete. NHS number completion varied between 91% and 97%. Patient postcode completion decreased over the four-year period from 65% to 51%. On closer inspection, 99% of patients with missing postcode were babies admitted on their date of birth; although this does not explain the decrease in completion over the four years, these postcodes were subsequently completed if the children were re-admitted to hospital after birth. As with LabBase2, patients who are not current residents in the UK are assigned a postcode where the prefix is "ZZ99" and the 3-digit suffix will denote the country or geographic area of residence; these form part of the completed postcode proportion in Figure 2.6.(109) Data completion does not necessarily reflect valid data entry as some may contain errors.

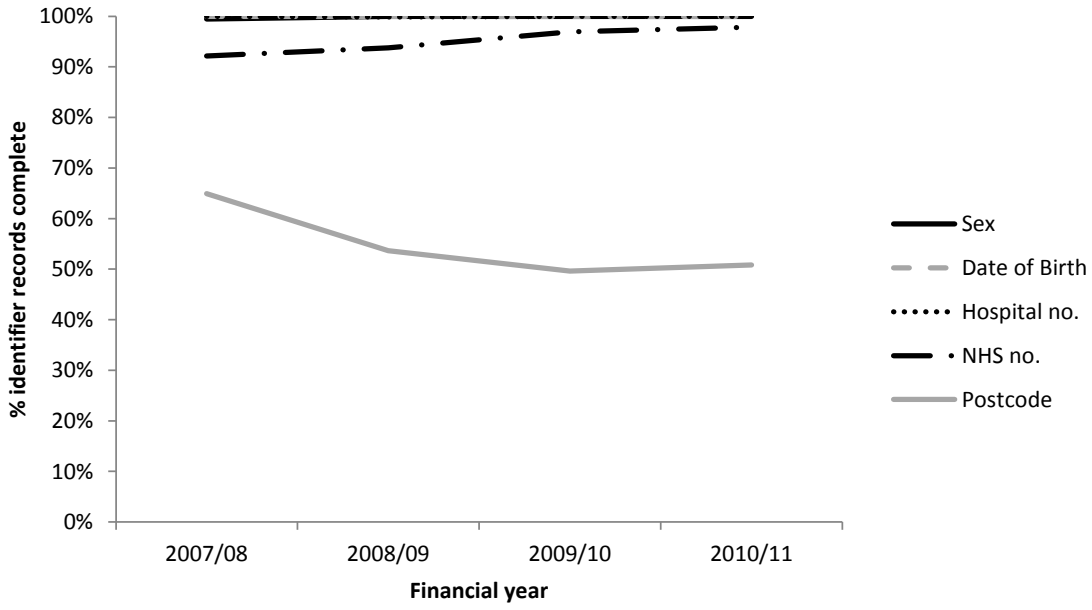


Figure 2.6. Proportion of identifier records with a data entry for each of the five patient identifiers in HES between 2007/08 and 2010/11 for children aged 0-5 year

2.4.5 HESid – identifying patients over time

To identify and track individual patients during the course of their IP admission, each patient is assigned a unique 32-character alphanumeric HES identification code (HESid) that is derived from the patient's identifier information. This HESid is subsequently used to identify any future episodes of IP care that the patient receives within the NHS system. This identification process searches for a match on patient identifiers of the new admission in the HES database to see whether or not there has already been a HESid assigned to that patient. This is achieved by an integrated three-step search and matching method.

2.4.6 HES three-step search and matching method

For one of the processing stages, a three-pass algorithm is used in the HES database to identify a potential HESid match:(120)

1. Sex AND Date of birth (partial) AND NHS number
2. Sex AND Date of birth (partial) AND Postcode AND Hospital number within Provider
3. Sex AND Date of birth (exact) AND Postcode*

** Postcodes on an exclusion list are exempt from this step. This includes densely populated communal facilities such as prisons (Prisons for women - mothers are allowed to keep their babies with them for up to 18 months in a Mother and Baby Unit)(121) and hospital to avoid mismatching patients.*

In addition, patients (including foreign patients) with no permanent address in the UK are assigned a special postcode depending on their country of residence or their present situation. These postcodes begin with a prefix of 'ZZ99', and the suffix will indicate their resident location. For example, patients from the Republic of Ireland are assigned the postcode ZZ99 3AZ.

The HES system is programmed to recognise when there are sufficient patient identifiers to achieve a match. Therefore it categorises new episodes into two groups, those that have sufficient patient identifiers but do not match a prior HESid, and those that do not match due to inadequate patient identifiers in the system. The HESid allows higher probability of matching patients instead of solely relying one identifier, which has had varying rates of completion over time. There are, however, negative aspects to this process. For example, the algorithm uses sex for each stage of the process to assess any matches, therefore if sex is misclassified by mistake, a patient is assigned a new HESid regardless of the fact that the NHS numbers matched.

2.4.7 Episodes and Spells

The HES data are made up of individual records per period of care for each patient, known as 'episodes'.(110) When a patient is admitted to hospital, their first HES episode is their 'admission episode', which represents the majority of cases who are only in hospital for one episode under the care of an assigned consultant. If this consultant should change, the patient moves into their second episode under a different consultant. This continues until the patient is discharged, whereupon their last episode is their 'discharge episode'. This whole journey from one hospital admission to discharge is called a 'spell' (Figure 2.7).

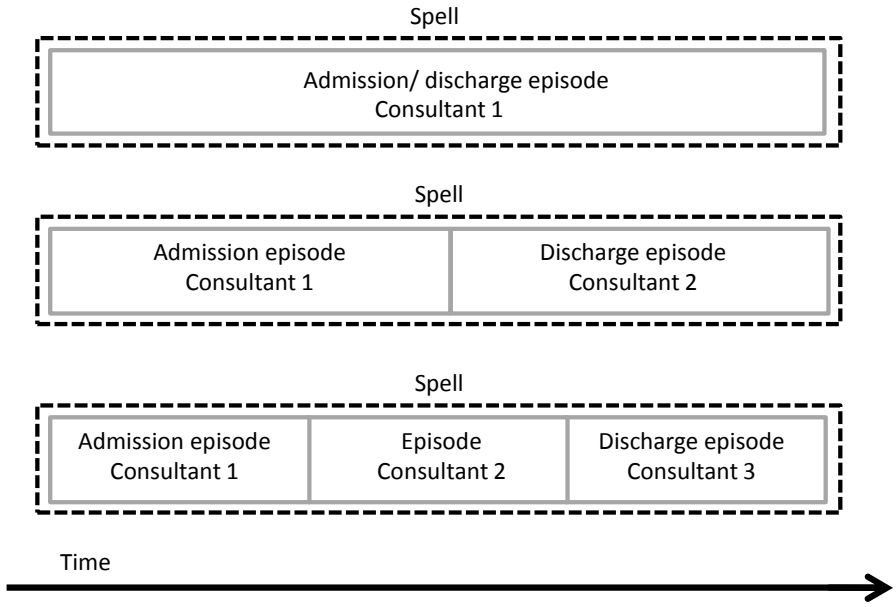


Figure 2.7. Episodes and spells in HES

2.4.8 Data structure

The HES data are formatted and stored as a line-list. Each row contains unique information about each hospital consultant episode for a particular patient, which describes the pathway of care a patient receives during their hospital stay (spell). A hospital spell can be made up of one or more consultant episodes depending on the number of different consultants a patient is assigned during the course of their treatment. For example, a patient can be in hospital for 20 days under the care of one consultant resulting in just one row of data, or a patient can be in hospital for 10 days under the care of five different consultants equating to five rows of data that are all linked together by the HESid, admission and discharge dates.

2.4.9 Query design for HES data extraction

In the first instance, only the essential components required for the linkage (patient identifiers and HESid) were extracted from the HES database to restrict the size and processing time required to link the HES and LabBase2 data. Each HES IP financial year was queried in turn (2007/08, 2008/09, 2009/10, 2010/11) using the example query shown in Figure 2.8.

Variables	Selected for	Example output
HESid		AAA00AA0000AAAAAA0
Patient NHS number		1111111111
Local Patient Identifier (hospital no.)		ST22222
Sex		M
Date of birth	02/04/2001-31/03/2011	DD/MM/YYYY
Patient postcode prefix		WT1
Patient postcode suffix		3FL

Figure 2.8. Example of the first restricted, patient identifier-only HES query (fictitious data)

Once linked, the HES database was queried for the additional demographic and clinical information using the linked HESids to extract the remaining data for the analysis, and no additional selection were applied (Figure 2.9).

Variables	Example output
HESid	AAA00AA0000AAAAAA0
Patient NHS number	1111111111
Patient hospital number	ST22222
Sex	M
Date of birth	DD/MM/YYYY
Patient postcode prefix	WT1
Patient postcode suffix	3FL
Hospital name	NORTH HOSPITAL
Admission date	DD/MM/YYYY
Ethnicity	BRITISH (WHITE)
Episode start date	DD/MM/YYYY
Episode end date	DD/MM/YYYY
Discharge date	DD/MM/YYYY
Method of admission	EMERGENCY A&E
Source of admission	USUAL RESIDENCE
Method of discharge	ON CLINICAL ADVICE/CONSENT
Discharge destination	USUAL RESIDENCE
Died on discharge?	NO
Main specialty	PAEDIATRICS
Treatment specialty	A&E
Diagnoses 1-20	A413, R05
Procedures 1-24	X355
Dates of procedures	DD/MM/YYYY

Figure 2.9. Example of the second more comprehensive HES query (fictitious data)

2.5 Data sources: considerations for linkage and analysis

As described above, LabBase2 captures on average 80% of all positive blood specimens in England, with some variability by month, whereas HES captures all cases of patients admitted to an NHS hospital, therefore any analyses produced from using the LabBase2 data will be an under-representation of the true number of BSI cases. This may be due to a variety of issues from laboratories supressing results, to data being rejected at the data loading stage. Patient identifiers are more complete in HES compared to LabBase2, suggesting that not all the LabBase2 cases will have sufficient information to successfully link to the corresponding HES IP data. The comparison of linked versus unlinked LabBase2 cases was analysed to report any potential bias between the two groups of patients.

In comparison, there is some recording of clinical sepsis (unconfirmed BSIs) based on the clinician's medical notes in the HES database, however indicators of suspected BSI are not collected in any routine linkable records to create a definition for a BSI case, unless they are linked to a microbiology report. Clinical reports of sepsis will not be quantified in this thesis as

it falls outside the parameters of the analysis of the microbiology-confirmed BSI when a specific specimen date can be used and pathogen detected.

2.6 Information governance: permissions to access data

PHE is the custodian of the LabBase2 data and is required to keep all patient identifiable information within the confines of the PHE IT system for patient confidentiality assuredness. This means that no data linkage by a third party is permitted and any data linkage has to be performed within the PHE IT system.

When I joined the HPA (now PHE) HCAI & AMR Department in 2008, the Department already had access to unidentifiable HES data via a standalone computer set up with a remote connection to the querying tool, Business Objects. To comply with the HPA data handling regulations at the time, I wrote an application to the HSCIC (formally known as the NHS Information Centre) and the National Information Governance Board (NIGB) on behalf of the HCAI & AMR department to extend our HES data handling permissions to include access to patient identifiable data (NHS number, hospital number, date of birth and postcode) for the purposes of linkage. This involved upgrading the HPA IT security system to incorporate a secure drive with encrypted back-up drive in-line with HSCIC standards of data protection and produce a security level system policy to outline and critique the handling and storage of patient identifiable HES data. The process took one year to complete.

2.7 Cohort selection

I used a retrospective four-year cohort of linked LabBase2 and HES data of children with a BSI. The aspects considered in the selection of the cohort and case definition are described individually.

2.7.1 Cohort time period

The cohort time period contained paediatric data where a positive blood isolate was detected between 1st April 2007 and 31st March 2011. This time period ensured that the most recent data was obtained at the time of data extraction in December 2011 (the annual HES data release for a particular year occurs in late October/early November of the following financial year) and sufficient numbers of paediatric BSI records were captured during this time (>10,000 BSI). This time-frame allowed sufficient time to capture patients' exposures in relation to the

timing of hospital admission whether they spent a short or long period in hospital. It also enabled the identification of previous IP exposures up to one year previously.

The 4-year time period was also selected as the quality of LabBase2 and HES data improved in the most recent years, therefore restricting the potential quantity and quality of data for analysis. To ensure the majority of laboratory reports had been submitted to LabBase2, I observed a six-week lag before extracting the required LabBase2 data.

2.7.2 Exposure period

To capture any exposure data including chronic conditions, previous NHS hospital healthcare contact and invasive procedures, the exposure period prior to the first BSI per patient in the cohort spanned 1 year per child, with a total range of 1st April 2006 to 31st March 2011; Figure 2.10 illustrates this timeline. The clinical and demographic exposure variables that will be considered in the analyses will be described in further detail in Chapter 4.

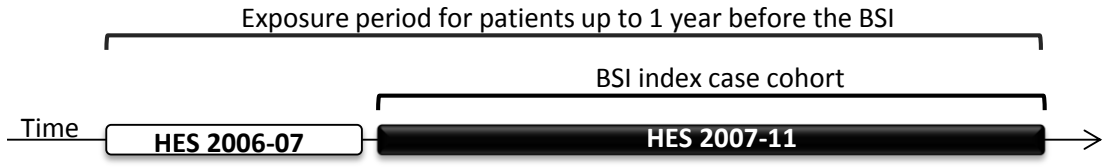


Figure 2.10. Diagram indicating cohort time-period with exposure prior to the BSI

2.7.3 Definition of a child with a BSI

Children with a BSI were defined as having a positive bacterial blood isolate in LabBase2.

2.7.4 Definition of a BSI index case

A BSI index case was identified as the first positive blood culture per patient within an OPIE episode (14-day period) identified on or after April 1st 2007 up until and including March 31st 2011. Any positive bacterial blood isolate found on or after the 1st April 2007, where a positive blood isolate for the same OPIE was also cultured before this date was not included in the cohort, as the infection episode would have started before the defined cohort start point. Examples of the inclusion and exclusion criteria for defining the BSI index case are shown in Figure 2.11. OPIE 1) would be accepted into the cohort, OPIE 2) would be excluded as first positive blood culture for that episode is outside the cohort time period, and OPIE 3) shows

the first being excluded and the second OPIE being included as the first (index) occurrence of a BSI within the cohort time period.

Within the cohort time period, BSIs occurring between 5 days before hospital admission until the date of hospital discharge were considered in the analysis. A liberal 5-day cut-off prior to an admission was chosen to maximise the number of records in the study. Post-discharge BSIs were not considered if there was not a subsequent hospital admission related to the infection as no current admission data would be available, indicating a potential miss-match between the datasets.

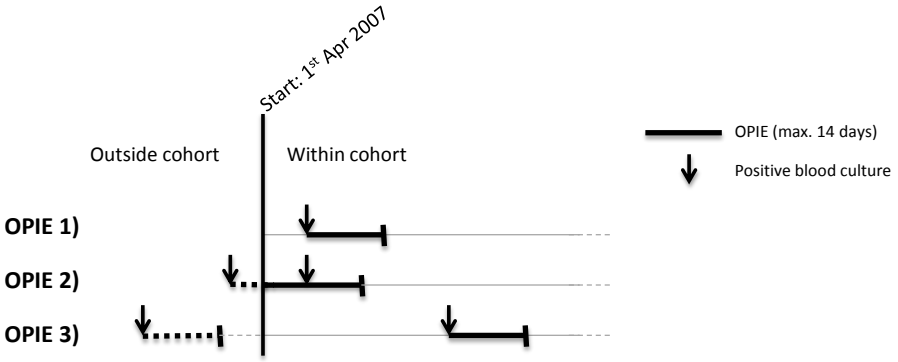


Figure 2.11. Examples of cohort inclusion for OPIE (1, 2, and 3) based on date of the positive blood isolate

2.7.5 Treatment of discharge and re-admission on the same day

In HES, there are occurrences of patients who are discharged and re-admitted on the same day. I have treated these re-admissions as the continuation of the spell because I am unable to quantify the time spent in the community as the time in hours is not recorded in HES. Therefore, I recoded the discharge date to be the last date where there is not a subsequent admission on the same day as that discharge date. An example of this is depicted in Figure 2.12.

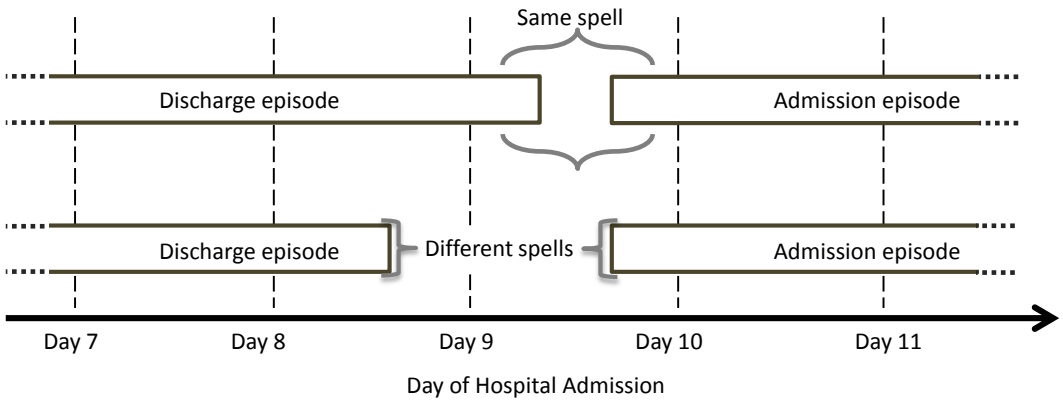


Figure 2.12. Treatment of a discharge and a re-admission episodes occurring on the same day

2.7.6 Age

Only patients aged between 1 month and 5 years 364 days old on the date of the positive blood culture were included in the cohort. I excluded children aged 0-28 days to avoid analysing a high proportion of neonates with vertically transmitted BSI from their mothers (e.g. GBS).(122) Collecting a sufficient volume of blood per sample is also technically more challenging in neonates as the veins are physically tiny and the quantity of circulating blood is small, leading to high rates of contaminated cultures.(123)

Children with a BSI acquired in hospital (e.g. neonatal intensive care) who had never been discharged from the hospital since birth were also excluded. For the purposes of the PhD, children need to have been exposed to potential risks within the community setting in order to be included; otherwise the focus of the analysis may be skewed. This young age-group of children aged 1 month – 5 years has the largest number of reports of BSI in the paediatric population.(8;44;46-48)

The upper age cut-off at 5 years and 364 days was chosen because most paediatric BSI and emergency hospital admissions occur in this younger age-group compared to older children.(44;124) Social situations (e.g. school), co-morbidities and age-related complications (e.g. alcohol-related, pregnancy, sexually transmitted disease) in older children differ to the larger concentration of congenital diseases and illnesses linked to immature immune systems in younger children. Limiting the cohort to this age-group also made the dataset size technically more manageable for the linkage process.

2.8 Summary

This chapter describes the content and limitations of both the LabBase2 and HES IP datasets, and proposes the inclusion criteria for the analysis cohort.

Completeness of identifiers in LabBase2 is lower than that of the HES data for all patient identifiers, with the exception of patient postcode, especially for NHS number which is the most patient discriminating variable due to its unique assignment. This data completeness will impact on the internal de-duplication of each dataset and will also affect the ability to link the datasets together.

The collection of laboratory results in LabBase2 is voluntary, and although the analysis is limited to reported isolate results, these reports still account for, on average, 80% of all BSI. From Harron *et al*.'s(106) findings in addition to my own research into the structure and

reporting of LabBase2, involving discussions with the LabBase2 database managers, I will assume that the non-reported data will not be biased according to pathogen but rather unreported at random. Additionally, I am not trying to determine overall rates of infection, but rather the proportion of infection according to the location where the infection was acquired. The cohort will capture sufficient data of children with BSIs (>10,000) over a four-year period and the additional explanatory variables using data from both LabBase2 and HES will be used to address the aim of this PhD.

2.9 Conclusion

This chapter has explored the strengths and weaknesses of both the LabBase2 microbiology and HES clinical data, which are the two national datasets that are available, accessible and with permission to link them within PHE, to address the aim of this thesis. Based on these observations, the criteria for child inclusion into the BSI cohort have been outlined and the explanatory variables defined for analysis. This chapter has highlighted some of the key issues that will need to be considered when interpreting the results of both the data linkage and the analysis in the following chapters.

Chapter 3. Calculating match-weights to probabilistically link LabBase2 and HES datasets

3.1 Introduction

The linkage of LabBase2 and HES datasets is necessary to create the cohort for analysis, described in Chapter 2, as no other national dataset containing both microbiology and clinical information exists. This chapter introduces the concept of data linkage and describes the method of the chosen data linkage approach. The results of the linkage and the proportion of BSI records successfully linked will be reported, followed by a description of the subsequent data management that was carried out to prepare the cohort of children with BSI for the descriptive data analysis in Chapter 4. Figure 3.1 illustrates the context of this chapter within the whole thesis.

3.2 Data linkage

3.2.1 Growing demand

Data linkage has progressed significantly since it was first discussed in the 19th century.(125-130) In the past few decades, data linkage has become increasingly more popular due to the dramatic expansion of electronic data production and the rapid advances in technology that have enabled the linkage process to take place. The numbers of global publications that mention 'data' and 'linkage' in PubMed has increased steadily between 1990 and 2013 from 20,000 to more than 100,000 publications per year.

Health and social records, among others, have migrated (or are in the process of migrating) from paper-based to electronic recording. Electronic records enable users to locate specific records more efficiently and quickly on a centralised database rather than sifting through thousands of paper copies. In terms of healthcare, children with chronic conditions who require long-term treatment, benefit from linked systems that capture all their historical medical events so that the medical practitioner treating the present day episode of care is aware of the full patient history.

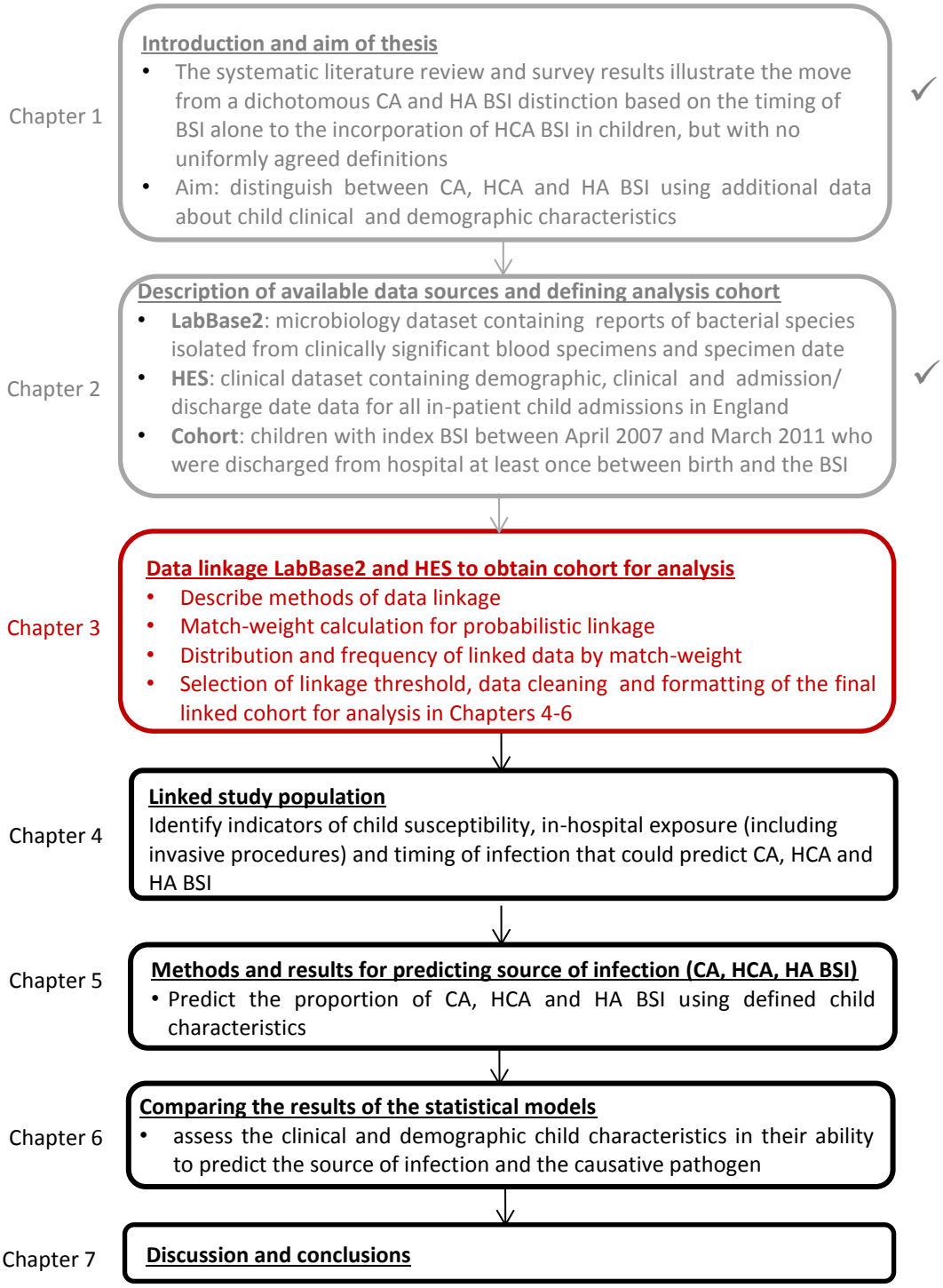


Figure 3.1. Flow diagram of the steps already addressed in previous chapters (grey), those that will be addressed in this Chapter (red) and the following chapters

However, as medical systems and delivery of care are also expanding to deliver further specialised services, a larger number of record systems for each individual or group of services and the integration of this data with established data systems is becoming more

important.(131) In addition, the increasing ease with which people can move and relocate nationally and internationally highlights the importance of patient record tracing.(127-129)

Data linkage offers a more time and cost effective approach to collecting data than the establishment of new data collection systems.(132;133) It is used for both establishing long-term linkage systems and ad-hoc linkage datasets for specific research projects; however, it depends on the information technology available, the data collection infrastructure, data management skills and funding available to support the linkage work.

3.2.2 Methods of data linkage

In its simplest form, the desired outcome from data linkage is to successfully link two records from two different datasets that are a true match, meaning that a patient from dataset A corresponds to the same patient in dataset B. There are four main outcomes when two records are linked: a true-match, a false match, a false non-match, or a non-match. This is summarised in Figure 3.2. The aim of the linkage process is to reduce the potential error of false true-matches (type I error) and false non-matches (type II error) to a minimum to increase the sensitivity (proportion of true matches correctly identified) and specificity (proportion of non-matches correctly identified) of the linkage in order to avoid analysing incorrect data.

	Match	**Non-match**
Linked	True match	False true-match (Type I error)
Not-linked	False non-match (Type II error)	Non-match

Figure 3.2. Data linkage error

Unique or partial identifiers common to both datasets have to be present for linkage to be possible. There are broadly three types of methods that can be employed to link two or more datasets are manual, deterministic, and probabilistic linkage.

3.2.2.1 Manual linkage

Manual linkage can be used on small datasets and relies on human interpretation of the data. The benefit of this method is that the human eye can identify potential data entry errors (e.g. misspelt names) that computer-based only linkage may miss, depending on the level of

programming and incorporated algorithms. However, human error may also be introduced if a systematic approach is not taken or checks are not in place and the reproducibility of the linkage is often challenging. The technical (and practical) limiting factor of manual linkage is dependent on the size of the datasets, due to the time consuming nature of this process. The method will become more inefficient, and prone to human error as the size of the datasets increase, and therefore the manual method is not the most appropriate for linking the LabBase2 and HES records.

3.2.2.2 Deterministic linkage

Deterministic linkage is where the linkage process tests all the possible match-pair combinations using the selected common identifier(s) and returns only those where the identifier(s) match exactly between the data records (Figure 3.3).(134) This method has been adapted to allow for some variability in the data, where the matching can be done over several stages and is known as step-wise deterministic linkage, where if two records do not match in the first round, but are only different by one or two identifiers, they could be considered as a true match depending on the agreed threshold.

Deterministic linkage works well for large datasets with good data completion and results in high sensitivity and specificity, however, this method is conservative in its matching when recording errors and data completion are poor, resulting in a large number of missed-matches. Taking this into consideration, deterministic linkage would not be the best approach to take with LabBase2 data considering the large proportion of missing patient identifiers (e.g. NHS number).

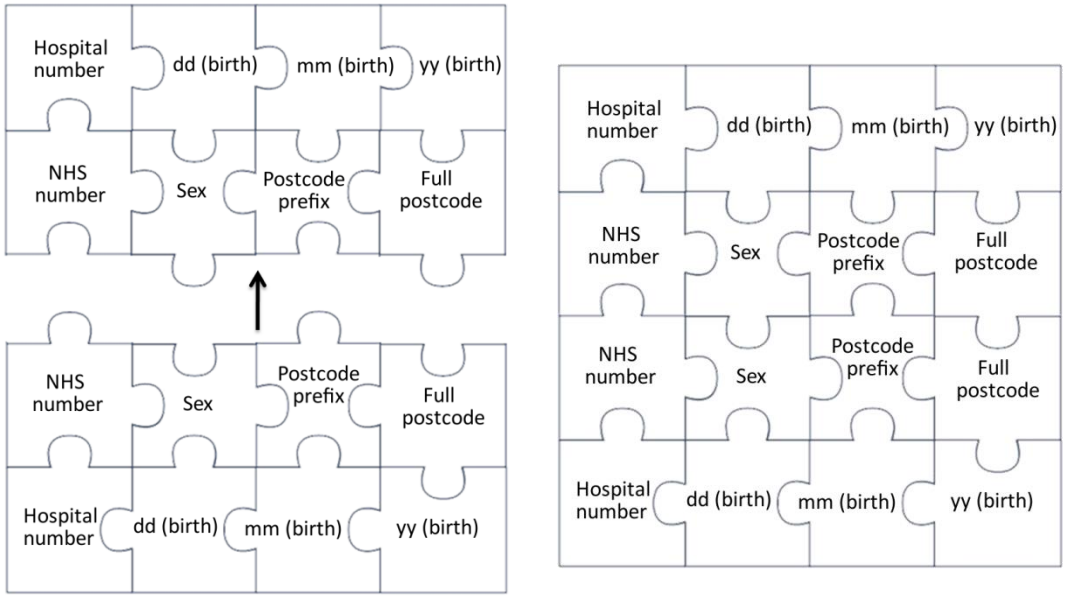


Figure 3.3. Diagram illustrating that when full patient identifiers are known and complete in both datasets, there is very high confidence in the match being true

3.2.2.3 Probabilistic linkage

Probabilistic linkage uses a combination of patient identifiers to test the strength of each match-pair by calculating the probability of two records matching by chance by estimating the error and accounting for the identifier completion variability of both datasets. This allows for any variation in the values assigned to each patient identifier between the datasets, for example data entry errors or missing entries (Figure 3.4).(134;135) When two datasets are probabilistically linked, each record in the first dataset is matched to all potential candidate records in the second dataset (i.e. a single record may be matched multiple times).

The likelihood of a true or false match is quantified by match-weights. Match-weights are calculated values that are assigned to each identifiable variable in the linkage process, based on the probability that two records should match by chance. Probabilistic linkage match-weights are calculated from a training dataset, using the equation: match-weight=$\log_2$ (m/u), where 'm' is the probability that an identifier agrees/disagrees/is missing for records that are a true match and 'u' is the probability that an identifier agrees/ disagrees/is missing by chance.(134;136)

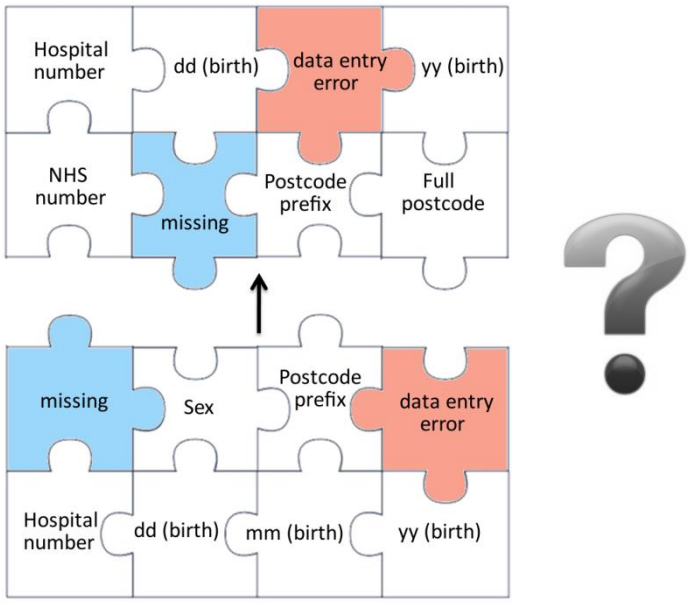


Figure 3.4. Diagram illustrating that when patient identifiers are missing or there are data entry errors, there is less confidence in the match being true

The training dataset is created from a set of true-match (m-value) pairs of patient identifiers. The u-values are subsequently calculated by taking the true matched records from each dataset and cross-matching them with each other to artificially generate false-match

combinations. The assumption is that the data completion of these artificially created false-matches reflects the composition of false-matches in the real dataset. The relationship between the datasets used to calculate the match-weights and the datasets linked using those match-weights is not well reported in the literature. When the linkage occurs, each patient identifier (e.g. sex, date of birth) in every linked pair of records is assigned a match-weight depending on the type of match (e.g. agree, missing or disagree). The total match-weight assigned to each matched-pair is then calculated by summing the individual weight scores of each patient identifier:

Match-weight for Sex + **Match-weight for NHS no.** + … = **Total match-weight**

These match-weights form a spectrum of linked pairs where high match-weight scores are deemed to be true above a high-threshold and low match-weight scores are deemed to be false below the lower threshold.(134) Inevitably, due to the nature of the data completeness, there is often an overlap between records that are true links and those that are false, creating a grey area of uncertainty between the true and false links (Figure 3.5). It is then up to the data analyst to decide on a threshold, where they accept all the matches equal to and above a particular match-weight into the final dataset. This can be done by manually reviewing the match-pairs between a conservative threshold (accepting only definite true links without including potential false links) that increases the specificity, to a more liberal threshold (capturing more of the true links but more false links) that increases the sensitivity but lowers the specificity. The threshold selection is subjective and may therefore differ between persons linking the data. Upper and lower thresholds for sensitivity analyses can also be defined.

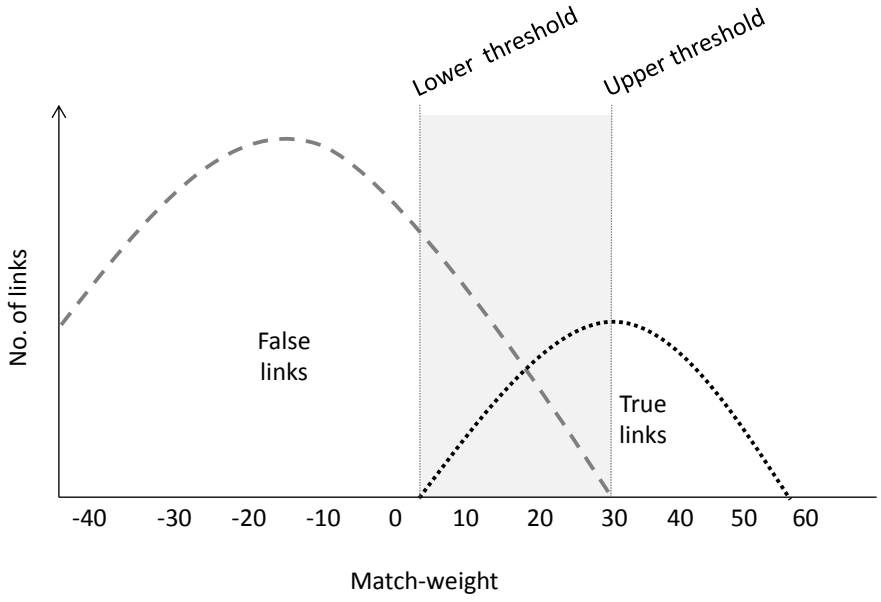


Figure 3.5. Range of possible probabilistic linkage thresholds

3.2.3 Data linkage in the UK

Unlike its neighbouring countries, England currently does not have an established national data linkage system that routinely links multiple datasets. Both Scotland and Wales have their own national linkage systems. Since the 1970s, the Scottish Health Informatics Programme has been Scotland's comprehensive linkage platform [www.scot-ship.ac.uk] made possible by the high data completion (93%) of Scotland's Community Health Index (CHI) number, similar to the NHS number. All linkage processes are based on using deterministic linking for patient data with a CHI number because data completion is very high and probabilistic linkage for those who do not possess a CHI number.(137) The Welsh Secure Anonymised Information Linkage databank uses a trusted third party to anonymise and encrypt the datasets on their behalf before any data linkage occurs.(138)

An example of one of the largest regional linkage programmes in England is the Oxford Record Linkage Study.(139) Although no longer functional today, the original aim was to link mortality and hospital discharge data using NHS number alone; however, the inconsistent completion of NHS numbers did not provide a sufficient linkage rate. Instead, probabilistic linkage was used to link a wider range if identifying characteristics.(140) Administrative data are very likely to contain incomplete or erroneous identifiers indicating a need for probabilistic linkage to support epidemiological surveillance and public health policy to inform current and new policies.

3.2.4 Chosen linkage method

I chose to link the LabBase2 and HES datasets using probabilistic linkage given the sub-optimal patient identifier completion found in LabBase2 in Chapter 2 as well as potential data entry errors that may be present in such large surveillance and administrative datasets.

There was no established and validated training dataset for child cases in LabBase2 and HES and no external validation dataset is readily available. There was, however, an established and validated dataset for calculating match-weights for the probabilistic de-duplication of new microbiology records that are submitted to LabBase2. These match-weights were calculated from patients of all ages with a positive microbiology result in LabBase2, referred to as MW-All from now on.(141)

Initially, I was only going to use the established match-weights derived from the LabBase2 database to link the 4-year LabBase2 and HES datasets because the match-weights had been

validated, were readily available and would be a quick solution. I decided to investigate the creation of new match-weights that would take into account the patient identifier completion in HES, as well as calculating child specific match-weights. Adapting the match-weights to the specific datasets was inspired from other HPA data linkage work to a different external dataset, although a slightly different methodology was undertaken in this thesis.(142) A new set of match-weights was therefore created from a linked LabBase2 and HES training dataset and will be referred to as MW-Child from now on.

3.3 Methods: determining probabilistic match-weights

3.3.1 Patient identifiers

Five patient identifiers (deconstructed into eight individual parts) common to both the LabBase2 and HES data were used to calculate the MW-Child and MW-All match-weights: NHS number, hospital number, sex, patient postcode (divided into prefix and full postcode) and date of birth (divided into three parts: day, month, and year).

3.3.1.1 NHS number

An NHS number is uniquely assigned to a patient receiving NHS treatment for their whole lifetime. A pair-wise match on NHS number should therefore indicate a high confidence in the probability that two records match. There are of course some caveats to consider such as twins or siblings NHS numbers that may be switched in error when attending hospital.

3.3.1.2 Hospital number

Hospital number is considered a strong patient identifier, although not to the same extent as an NHS number due to the ability to uniquely distinguish patients within one hospital, not between hospitals. Therefore, patients may also be assigned multiple hospital numbers.

3.3.1.3 Postcode

The average residential postcode in England encompasses 15 households.(143) The probability of matching on a less specific area is greatly increased if the suffix has been incorrectly entered or is missing, which is why the postcode is split into the prefix only and full postcode for matching, to account for any data entry errors of the suffix code.

3.3.1.4 Sex

There are three possible entries for sex where the patient is either male, female, or the sex is unknown/not entered and is therefore a weak identifier in isolation.

3.3.1.5 Date of Birth

The date of birth is a very important variable. This does not uniquely distinguish patients, but is generally well completed and helps to increase the odds of a true match between two records. However, data entry errors can occur where the month and date may be accidentally switched, for example, from the 05/06/2005 to the 06/05/2005.

I included laboratory records of patients with a missing date of birth to reduce selection bias due to missing data. This inevitably introduced a proportion of adult microbiology records, but these records were removed at the linkage stage when no matches with the HES data were identified.

3.3.2 PHE match-weights for all ages (MW-All)

The MW-All were taken directly from the probabilistic methodology written by A. Grant and were published internally within the HPA (now PHE), as described in Chapter 2.(141) The MW-All were calculated and refined over several years by counting the observations in the data for a 12-month period (e.g. 901,046 OPIE records in 2009) after blocking on organism species using data held in LabBase2. There were several limitations and the potential introduction of bias when using this method:

1. The MW-All set would not take into account the completion and data quality found within the HES dataset.
2. The MW-All would not necessarily be sensitive to linking child-only records, which may require different weighting priorities compared to the whole population, for example, a smaller number of possible date of birth combinations.
3. Soundex, initials, forename or surname are additional variables collected in LabBase2 that are not collected by the HES database. The exclusion of these variables from the total match-weight value may have an impact on the final match-weight to link LabBase2 and HES.

3.3.3 Calculation of child match-weights (MW-Child)

3.3.3.1 Training dataset

The true matches (x=15,046; confirmed by manual review) were selected from a previously linked LabBase2 and HES dataset (children aged 1 month – 18 years with a BSI between January 2009 and March 2010).(144) Match-weights were calculated from this dataset using the following equation:(134;136)

$$match - weight = \log_2\left(\frac{m}{u}\right)$$

$\boldsymbol{m}$ *= Probability that the patient identifier Agrees, Disagrees, or is Missing in at least one dataset given that the record is a <u>true</u> match* (x)

$\boldsymbol{u}$ *= Probability that the patient identifier Agrees, Disagrees, or is Missing in at least one dataset by chance given that the record is a <u>false</u> match* ($N - x$)

3.3.3.2 Worked example of calculating match-weights

The following scenario (using a dummy dataset) illustrates the method of match-weight calculation for a variable. In this example, the true-match dataset has six records (Figure 3.6).

Lab_ID	HES_ID	Lab_variable	HES_variable	match
1	1	F	F	✓
2	2	M	F	x
3	3	.	M	missing
4	4	M	M	✓
5	5	F	F	✓
6	6	M	M	✓

Figure 3.6. Example dataset containing true-matches

The $'\boldsymbol{m}'$ value was calculated for outcome:

$$\begin{aligned} Agree &= 4/6 = 0.67 \\ Disagree &= 1/6 = 0.17 \\ Missing &= 1/6 = 0.17 \end{aligned}$$

The $'\boldsymbol{u}'$ values were calculated by cross-matching the Lab_ID and HES_ID variables to obtain the total 30 false-matches (in white, Figure 3.7).

Lab_ID	HES_ID	Lab_variable	HES_variable	match
1	1	F	F	✓
1	2	F	F	✓
1	3	F	M	x
1	4	F	M	x
1	5	F	F	✓
1	6	F	M	x
2	1	M	F	x
2	2	M	F	x
2	3	M	M	✓
2	4	M	M	✓
2	5	M	F	x
2	6	M	M	✓
3	1	.	F	missing
3	2	.	F	missing
3	3	.	M	missing
3	4	.	M	missing
3	5	.	F	missing
3	6	.	M	missing
4	1	M	F	x
4	2	M	F	x
4	3	M	M	✓
4	4	M	M	✓
4	5	M	F	x
4	6	M	M	✓
5	1	F	F	✓
5	2	F	F	✓
5	3	F	M	x
5	4	F	M	x
5	5	F	F	✓
5	6	F	M	x
6	1	M	F	x
6	2	M	F	x
6	3	M	M	✓
6	4	M	M	✓
6	5	M	F	x
6	6	M	M	✓

Figure 3.7. Dataset containing true-matches (grey) and false-matches (white)

The $'\boldsymbol{u}'$ value was calculated for each outcome:

$$Agree\ by\ chance = 11/30 = 0.37$$
$$Disagree\ by\ chance = 14/30 = 0.47$$
$$Missing\ by\ chance = 5/30 = 0.17$$

The $'\boldsymbol{m}'$ and $'\boldsymbol{u}'$ values were then used to calculate the match-weight for pairs of the linked variable:

$$Agree = \log_2\left(\frac{0.67}{0.37}\right) = 0.86$$

$$Disagree = \log_2\left(\frac{0.17}{0.47}\right) = -1.47$$

$$Missing \quad = \log_2\left(\frac{0.17}{0.17}\right) = 0$$

3.4 Results: determining probabilistic match-weights

The results of the true-matches, false-matches and missing data matches for MW-All and MW-Child are shown in Table 3.1. The MW-All values for missing data were all negative or neutral, whereas MW-Child contained a few positive values apart from NHS number, although these were very small values.

The radar plots in Figure 3.8, Figure 3.9 and Figure 3.10 depict the proportional contribution of each patient identifier's match-weight to the total match-weight for patient identifier pairs that agree, disagree or have missing data. Each octagon represents the total match-weight (100%); the position of each of the octagon's vertices on the radar's scale indicates the proportional weight that each patient identifier contributes to the total match-weight.

For the 'agree' match-weights (Figure 3.8), the octagon distribution on each point was very similar for both MW-Child and MW-All: NHS number (21% vs. 28%), full postcode (20% vs. 23%) and hospital number (19% vs. 22%) were the most important patient identifiers when there was agreement between two linked records, each receiving a large, positive match-weighting. More proportional weighting was given to date of birth and postcode prefix for the MW-Child than MW-All. Overall, sex had a very low contribution to the 'agree' weighting for both match-weight sets (1-2%).

More variance was observed for the 'disagree' irregular MW-Child octagon (Figure 3.9) compared to the MW-All octagon for all eight variable parts. Disagreement on hospital number had the largest disparity between MW-Child (3%) and MW-All (14%), followed by date of birth (MW-Child: 18-20%; MW-All: 14%), NHS number (MW-Child: 18%; MW-All: 14%) and sex.

Table 3.1. Match-weights for each patient identifier for MW-All and MW-Child

	Patient identifier	MW-All	MW-Child
Agree	**NHS number.**	25.8355709	11.4268567
	Sex	0.9985551	0.9719587
	Hospital number	20.2779856	10.1814387
	Date (DD) of birth	4.9040985	4.9072601
	Month (MM) of birth	3.5823147	3.5718076
	Year (YYYY) of birth	4.3895666	3.3975976
	Postcode prefix	11.5586405	8.6782742
	Full postcode	21.1731814	10.6009714

		MW-All	MW-Child
Disagree	**NHS number**	-7.962896	-9.7328087
	Sex	-6.962896	-7.5634034
	Hospital number	-7.9628949	-1.4093752
	Date (DD) of birth	-7.9139864	-9.7375195
	Month (MM) of birth	-7.8373651	-10.9430663
	Year (YYYY) of birth	-7.8925067	-10.7333376
	Postcode prefix	-4.6142411	-2.6771329
	Full postcode	-4.6147092	-1.8666664

		MW-All	MW-Child
Missing	**NHS number**	-0.6311791	-0.3494677
	Sex	-0.0377842	0
	Hospital number	-8.450181	0.0003828
	Date (DD) of birth	0	0
	Month (MM) of birth	0	0
	Year (YYYY) of birth	0	0
	Postcode prefix	-0.4586327	0.01486
	Full postcode	-0.4586327	0.0141547

MW = Match-weight

The proportional weight distribution assigned to linked pairs with missing data was different for both sets of match-weights (Figure 3.10). The majority for MW-All (84%) was focused on the hospital number, receiving a large penalty (-8.45; Table 3.1) to the total match-weight. Although a small combined reduction (-0.34), the proportionally largest (92%) penalty for a missing value for MW-Child was NHS number (Figure 3.10).

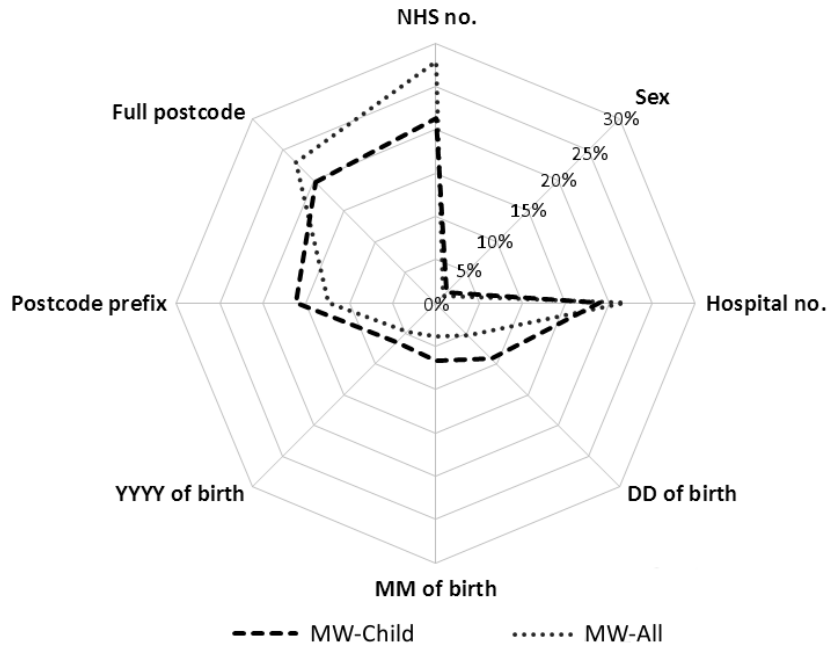


Figure 3.8. Agree match-weights: Radar plot overlaying the sets of match-weights and comparing the proportion that each of the eight individual patient identifiers contributes towards the sum of the total match-weight (100%). DD = 2-digit day of birth; MM – 2-digit month of birth; YYYY = 4-digit year of birth; MW = match-weight

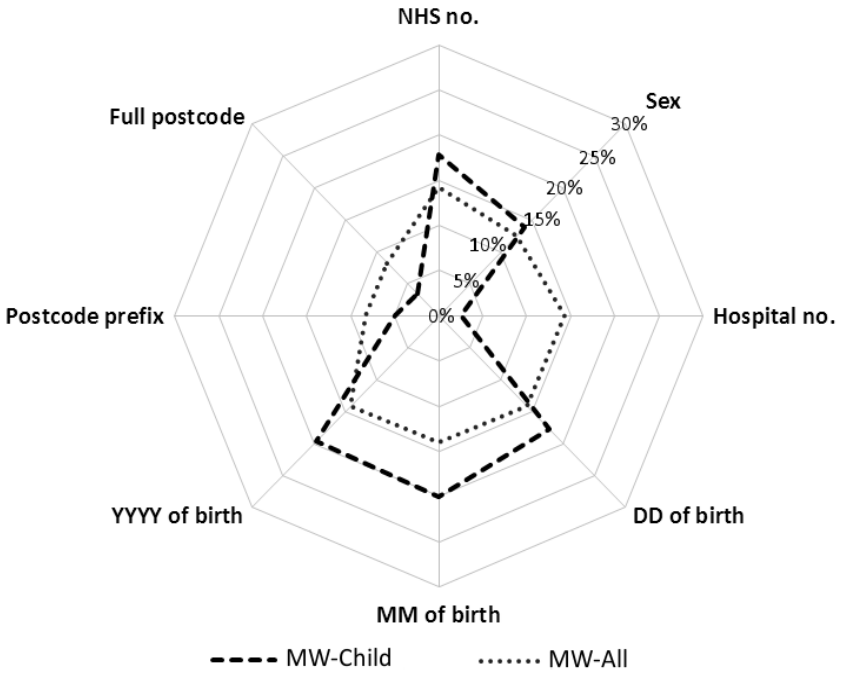


Figure 3.9. Disagree match-weights: Radar plot overlaying the sets of match-weights and comparing the proportion that each of the eight individual patient identifiers contributes towards the sum of the total match-weight (100%); DD = 2-digit day of birth; MM – 2-digit month of birth; YYYY = 4-digit year of birth; MW = match-weight

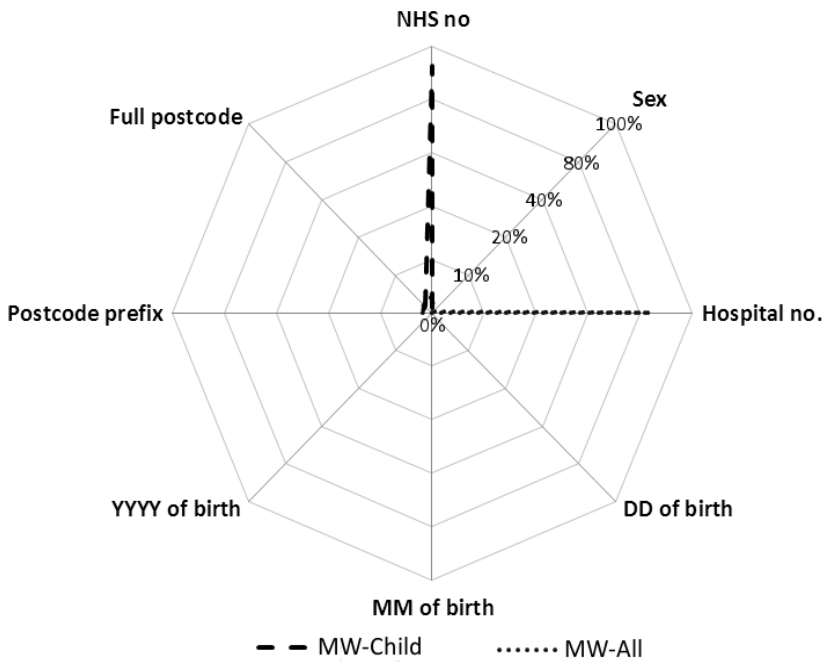


Figure 3.10. Missing data match-weights: Radar plot overlaying the sets of match-weights and comparing the proportion that each of the eight individual patient identifiers contributes towards the sum of the total match-weight (100%); DD = 2-digit day of birth; MM – 2-digit month of birth; YYYY = 4-digit year of birth; MW = match-weight

3.5 Methods: probabilistic linkage of LabBase2 and HES

The extraction of HES data was based on child's age at date of admission between birth and 5 years of age. Although the cohort inclusion criteria only included BSIs occurring between 1 month and 5 years of life, children less than 1 month old were included in the HES extract because the child's date of admission may have been within the first four weeks of life prior to the BSI.

All the LabBase2 patient identifiers were linked twice using MW-All and MW-Child to each of the four financial HES years, as show in Figure 3.11. In-house software to link the datasets was developed using Python (version 2.7.2). Figure 3.12 illustrates the computational procedure of the full probabilistic linkage methodology which was executed within PHE's Department of HCAI & AMR, Colindale, by the database manager (Dr. Mehdi Minaji). Due to the vast size of the data being linked, extra server space and computational power was required for the linkage process to be a success. The linkage for each of the two sets of match-weights was performed separately, taking three days to run each match-weight set.

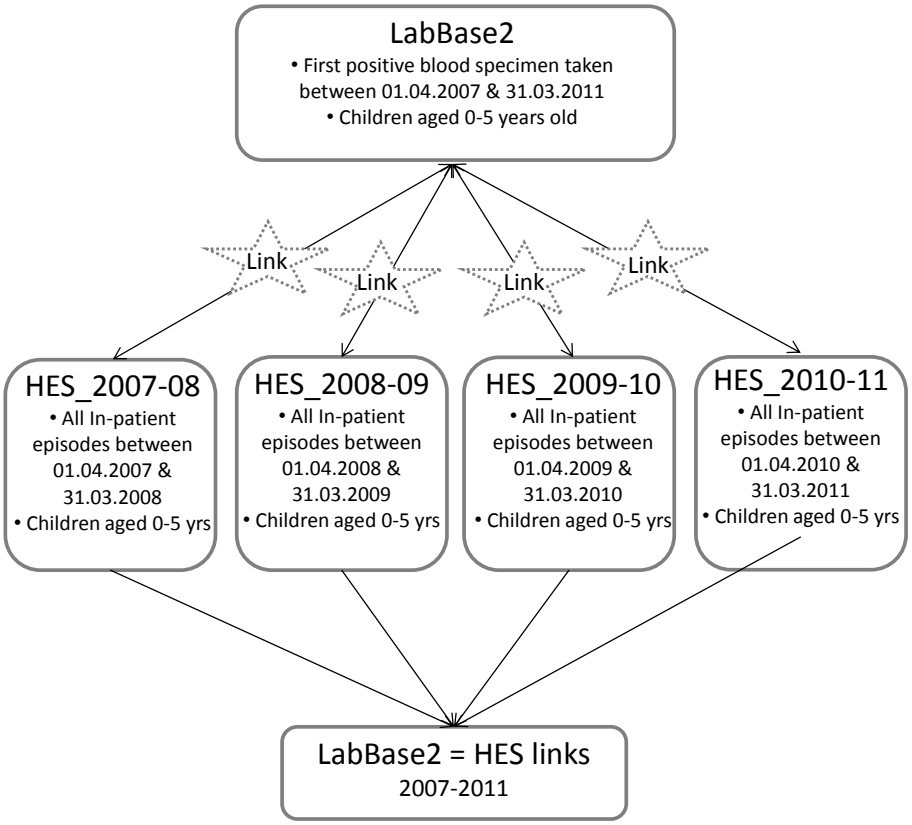


Figure 3.11. Linkage of all the LabBase2 data to each individual year of HES data

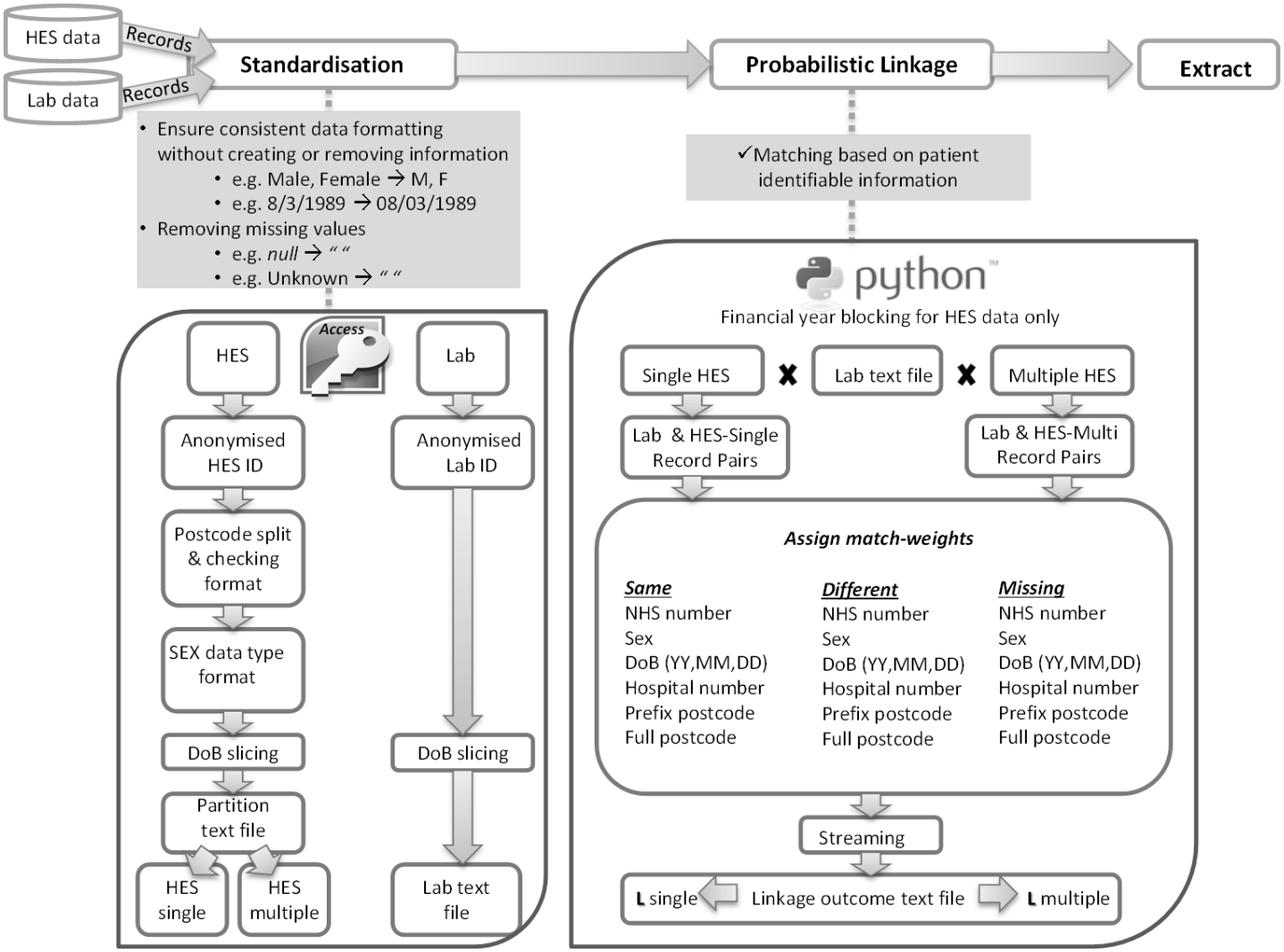


Figure 3.12. Computational challenges involved with probabilistic linkage of LabBase2 and HES data

3.5.1 Threshold selection

The match-weight threshold for MW-All and MW-Child were selected by manual review. All pairs were reviewed if there were ≤100 linked pairs, or a random sample of 100 linked pairs was selected if there were >100 linked pairs for a match-weight (range: 101 pairs to 79 million pairs). The reviewed match-weights ranged from the highest positive match-weight to a match-weight of -10, where negative match-weights strongly suggest the linked pairs are not a true match. To ensure high specificity, the threshold was selected just above the highest match-weight for which definite false matches were identified. The thresholds were selected separately for MW-All and MW-Child [match-weight >12.65], without any cross-comparison. The highest match-weight for each OPIE above the respective thresholds was retained. The proportion of linked (above threshold) versus unlinked (below threshold) OPIEs and the characteristics of these links by year, region and patient identifier completion were compared using chi-square tests and according to age (continuous) using a t-test for both MW-All and MW-Child.

As the distribution of linked pairs is heavily dependent on the selected threshold, a sensitivity analysis using a higher threshold was selected for both sets of match-weights at the point where all matches were considered true. STATA (v13) was used in the manipulation of the data.(145)

3.6 Results: probabilistic linkage of LabBase2 and HES

The linkage process resulted in the creation of millions of linked pairs from the fixed number of LabBase2 and HES records, with an exponential increase in the number of links where the match-weight value decreased (suggesting less evidence of a true match), a large proportion resulting from links where identifiers were either missing or did not match. The manual review classification of linked pairs per match-weight (considered true, uncertain or false) are summarised in Figure 3.13 and Figure 3.14. The match-weight thresholds were selected where all match-weights >14.17 for MW-All and >12.65 for MW-Child were retained. There was a significant increase in the number of OPIEs linked by both MWs between 2007 and 2011 ($P<0.001$).

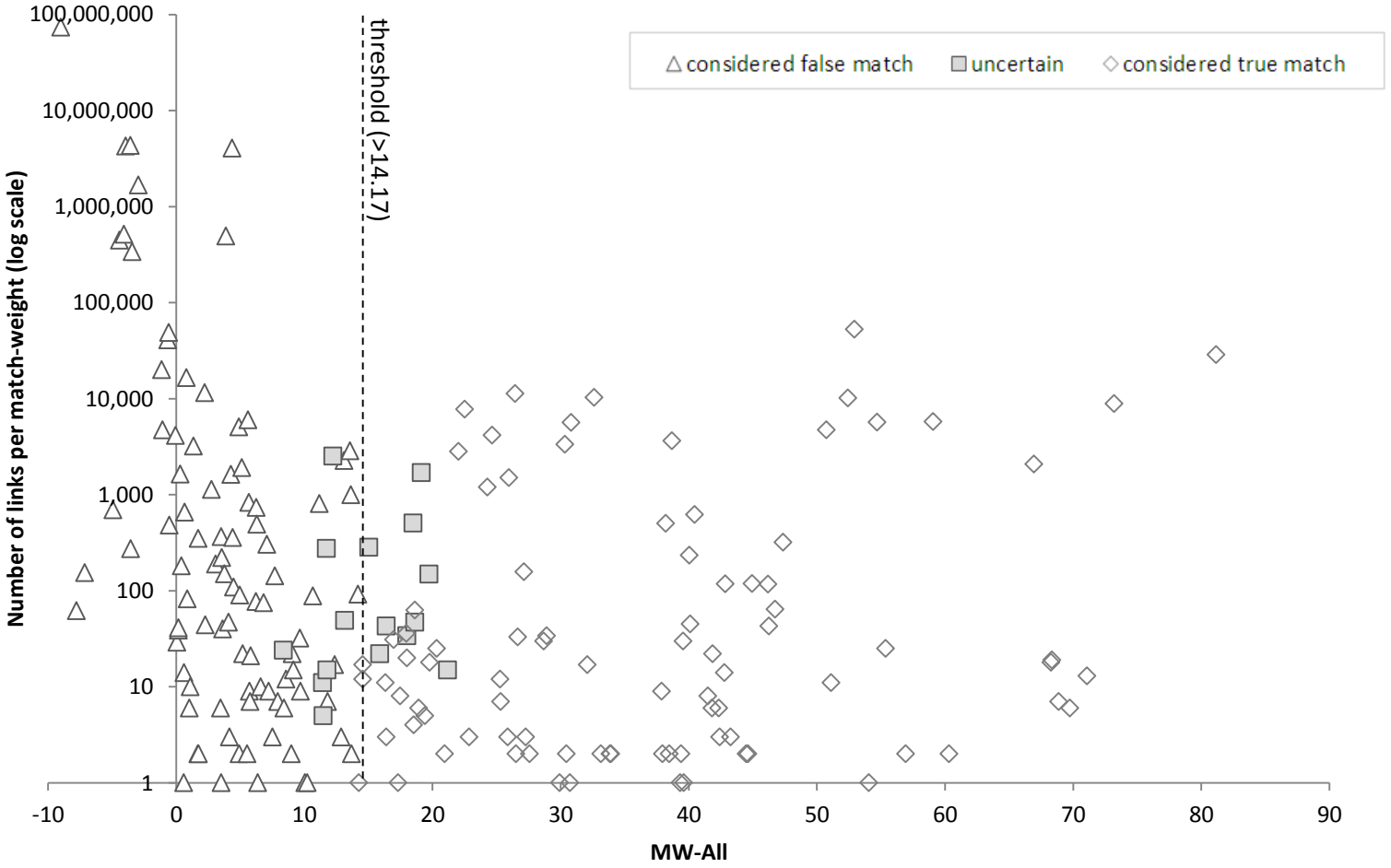


Figure 3.13. Distribution of the total number of linked pairs per match-weight considered as true, uncertain or false matches using the MW-All in relation to the chosen threshold (selected via manual review)

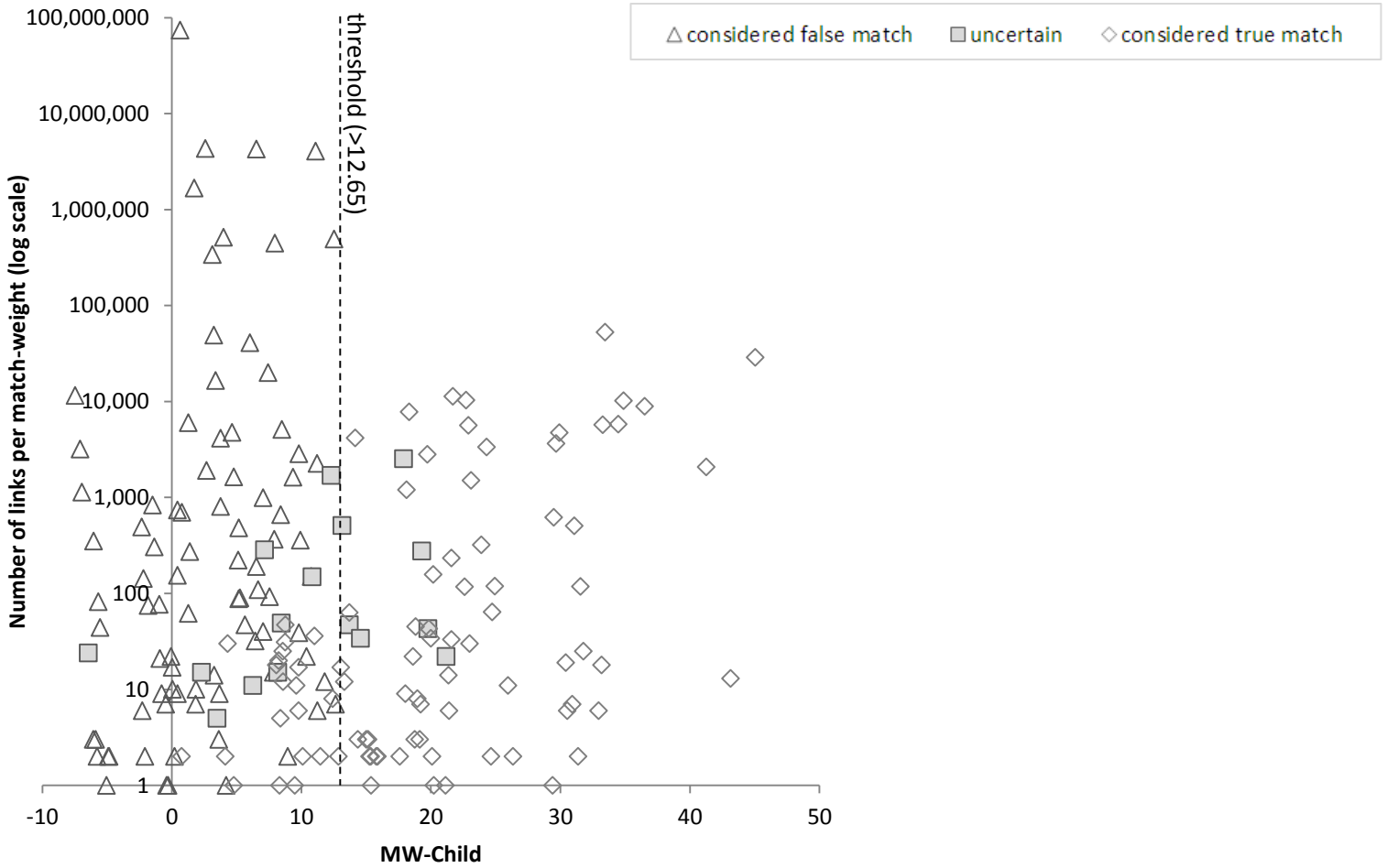


Figure 3.14. Distribution of the total number of linked pairs per match-weight considered as true, uncertain or false matches using the MW-Child in relation to the chosen threshold (selected via manual review)

Of the total 21,074 OPIEs that could have been linked, MW-Child linked a total of 17,141 OPIEs (81.4%) and MW-All linked a total of 17,166 OPIEs (81.6%; Table 3.2). Of the 17,243 linked OPIEs, MW-Child and MW-All both linked the same 17,064 OPIEs (99%); MW-All linked 102 unique OPIEs that were not linked using the MW-Child method and the MW-Child linked 77 unique OPIEs that were not linked using the MW-All method (Figure 3.15).

Table 3.2. Number and proportion of linked OPIEs for MW-All and MW-Child per year for children aged 1 month – 5 years in England

	No. of OPIEs (%)				Total
	2007/08	2008/09	2009/10	2010/11	
No. of OPIEs linked using MW-All	4,337	4,123	4,316	4,390	17,166
(% linked)	(76.7%)	(80.7%)	(84.2%)	(84.7%)	(81.6%)
No. of OPIEs linked using MW-Child	4,336	4,122	4,307	4,376	17,141
% linked	(76.7%)	(80.6%)	(84.1%)	(84.4%)	(81.4%)
Total no. of OPIEs prior to linkage	5,654	5,111	5,124	5,185	21,074

* OPIE = Organism Patient Illness Episode

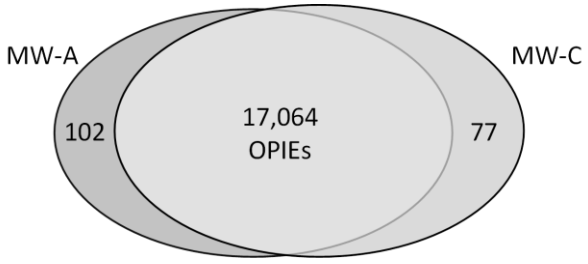


Figure 3.15. Number of unique and mutual OPIEs and HESid linked pairs for the MW-All and MW-Child

3.6.1 Comparison of linked and unlinked OPIEs

Table 3.3 summarises the total number of OPIEs from LabBase2 that were not linked, compared to those that were linked by both MW-Child and MW-All, MW-Child only or MW-All only by year, age, region and patient identifiers. Of the patient identifiers, the completion of NHS number, hospital number, date of birth and postcode in the linked records was significantly higher than in the unlinked records (P<0.001). There was also regional variation between the linked and unlinked OPIEs (P<0.001), although among the linked OPIEs, 59 OPIEs for MW-Child and 58 OPIEs for MW-All were reported from Welsh laboratories.

Of the unlinked OPIEs using both MW-Child and MW-All (n=3831), 15% (n=562) were reported from laboratories in Wales, the Channel Islands and the Isle of Man, which are geographical regions where clinical hospital data is not captured by HES. Of the remaining unlinked OPIEs (n=3,269), 34% (n=1,313) were missing date of birth and 87% (n=2,858) were missing NHS

number. The missing date of birth records may represent records from a mixture of children and adults because no other age proxy was available. This suggests that the actual proportion of unlinked child OPIEs is smaller than the total reported in Table 3.3, however, this cannot be validated. Where data were reported, there were no statistically significant differences by age (t-test, P=0.281) or sex ($Chi^{2,}$ P= 0.054) between the linked and unlinked OPIEs.

3.6.2 Sensitivity analysis

For the sensitivity analysis where the threshold was selected at the match-weight above the highest uncertain square linked pairs (Figure 3.13 and Figure 3.14), a total of 16,398 (77.8%) of the 21,074 OPIEs were linked for MW-Child, and 17,102 (81.1%) for MW-All. Of these, both methods linked the same 16,398 OPIEs, highlighting the agreement of the true-match weighting between both methods; however, 704 unique OPIEs were linked only by MW-All. To avoid reducing the numbers of linked pairs, the original thresholds were maintained for the rest of the analysis.

To compare deterministic linkage (using NHS number only) to probabilistic linkage, 25% more OPIE records were linked using the probabilistic linkage MW-Child method (81%, n=17,141/21,074 OPIEs) than would have been linked if deterministic linkage had been used (58%, n=12,212/21,074 OPIEs).

Table 3.3. Number of OPIEs successfully linked to HES data using MW-All and MW-Child out of the total number OPIEs extracted from LabBase2 for children aged 1 month–5 years with a BSI, England

Variables	Total OPIEs	No. not linked		MW-Child & MW-All		MW-Child		MW-All	
		n	%	n	%	n	%	n	%
Total	21,074	3,831	18	17,064	81	77	0	102	0
Year									
2007/08	5,654	1,295	23	4,314	76	22	0	23	0
2008/09	5,111	966	19	4,100	80	22	0	23	0
2009/10	5,124	784	15	4,283	84	24	0	33	1
2010/11	5,185	786	15	4,367	84	9	0	23	0
Age									
1m<1y	9,196	1,122	12	8,013	87	44	0	17	0
1<2y	3,715	472	13	3,204	86	15	0	24	1
2<3y	2,298	310	13	1,973	86	9	0	6	0
3<4y	1,898	222	12	1,666	88	7	0	3	0
4<5y	1,450	195	13	1,251	86	0	0	4	0
5<6y	1,028	137	13	884	86	2	0	5	0
missing	1,489	1,373	92	73	5	0	0	43	3
Region									
East Midlands	1,162	121	10	1,034	89	7	1	0	0
East of England	1,991	357	18	1,584	80	15	1	35	2
London	5,072	749	15	4,288	85	21	0	14	0
North East	968	70	7	895	92	0	0	3	0
North West	2,611	363	14	2,230	85	9	0	9	0
South East	3,203	638	20	2,555	80	1	0	9	0
South West	1,233	152	12	1,056	86	7	1	18	1
West midlands	1,990	519	26	1,457	73	8	0	6	0
Yorkshire & the Humber	2,223	300	13	1,907	86	8	0	8	0
Wales	609	550	90	58	10	1	0	0	0
CI & IOM*	12	12	100	0	0	0	0	0	0
NHS number									
complete	12,843	614	5	12,210	95	2	0	17	0
missing	8,231	3,217	39	4,854	59	75	1	85	1
Hospital number									
complete	18,556	2,977	16	15,416	83	66	0	97	1
missing	2,518	854	34	1,648	65	11	0	5	0
Date of Birth									
complete	19,585	2,458	13	16,991	87	77	0	59	0
missing	1,489	1,373	92	73	5	0	0	43	3
Postcode prefix									
complete	14,953	996	7	13,857	93	77	1	23	0
missing	6,121	2,835	46	3,207	52	0	0	79	1
Full postcode									
complete	14,834	952	6	13,782	93	77	1	23	0
missing	6,240	2,879	46	3,282	53	0	0	79	1
Sex									
Female	8,569	1,266	15	7,231	84	28	0	44	1
Male	11,286	1,557	14	9,622	85	49	0	58	1
Unknown	1,219	1,008	83	211	17	0	0	0	0

3.7 Results: Association between MW-All and MW-Child

The bubble chart in Figure 3.16 summarises the relationship between linked pairs of records using MW-All and MW-Child and the number of links (size of the circle is proportional to the number of links: the smallest circle = 1 linked pair; the largest circle = 79 million linked pairs) assigned to each respective match-weight combination. The selected thresholds for each set of match-weights are also depicted. Only links where at least one match-weight was above zero are shown in the bubble chart. I considered any linked pair with a match-weight of zero or above to have some positive attribute that should be investigated.

For linked pairs that received a negative match-weight in both MW-Child and MW-All, they were considered to indicate poor agreement with insufficient completion of patient identifier data or disagreeing linked pairs. There was a positive correlation between the distributions of linked pairs and assigned match-weights although some variability in the assignment of match-weights was visible. Of particular interest were those links above the threshold for one MW and below for the other. For example, linked pairs with a MW-All of 12.2 (below the MW-All threshold) and MW-Child of 17.9 (above the MW-Child threshold) accounted for 75 linked OPIEs, meaning that MW-All would not incorporate these 75 OPIEs that MW-Child considers as true-matches into the final dataset. There were few examples where the match-weights had very different results. For example, the highest MW-All value assigned a negative MW-Child value (-9.82) had a match-weight value of 16.4. The three linked pairs with a match-weight of 16.4 would be considered true matches if the MW-All weighting methodology was chosen, whereas the reverse would be true if the MW-Child weighting was chosen.

The core differences between the match-weights are reflected by the different characteristics of the linked MW-All only or MW-Child only records. Forty-three (3%, 95% CI: 2.1-3.8%) OPIEs for patients with missing age (i.e. no date of birth) were linked using the MW-All, whereas MW-Child did not link any additional OPIEs where patients were missing date of birth. The 77 MW-Child only links had completed postcode but had a very low NHS number completion (n=2, 3%) compared to MW-All (n=17, 17%). Hospital number completion in both was high (84%), which was particularly noticeable for MW-All where the negative weight of -8.45 (Table 3.1) for missing hospital number would lower the overall weight of any linked pairs with this weighting. The majority (57%) of linked records using the MW-Child were from children aged 1 month < 1 year. There was also a significant variation between linked pairs resulting from either MW-Child or MW-All linkage according to region where twice as many cases from the East of England and the South-West were linked by MW-Child (P=0.001).

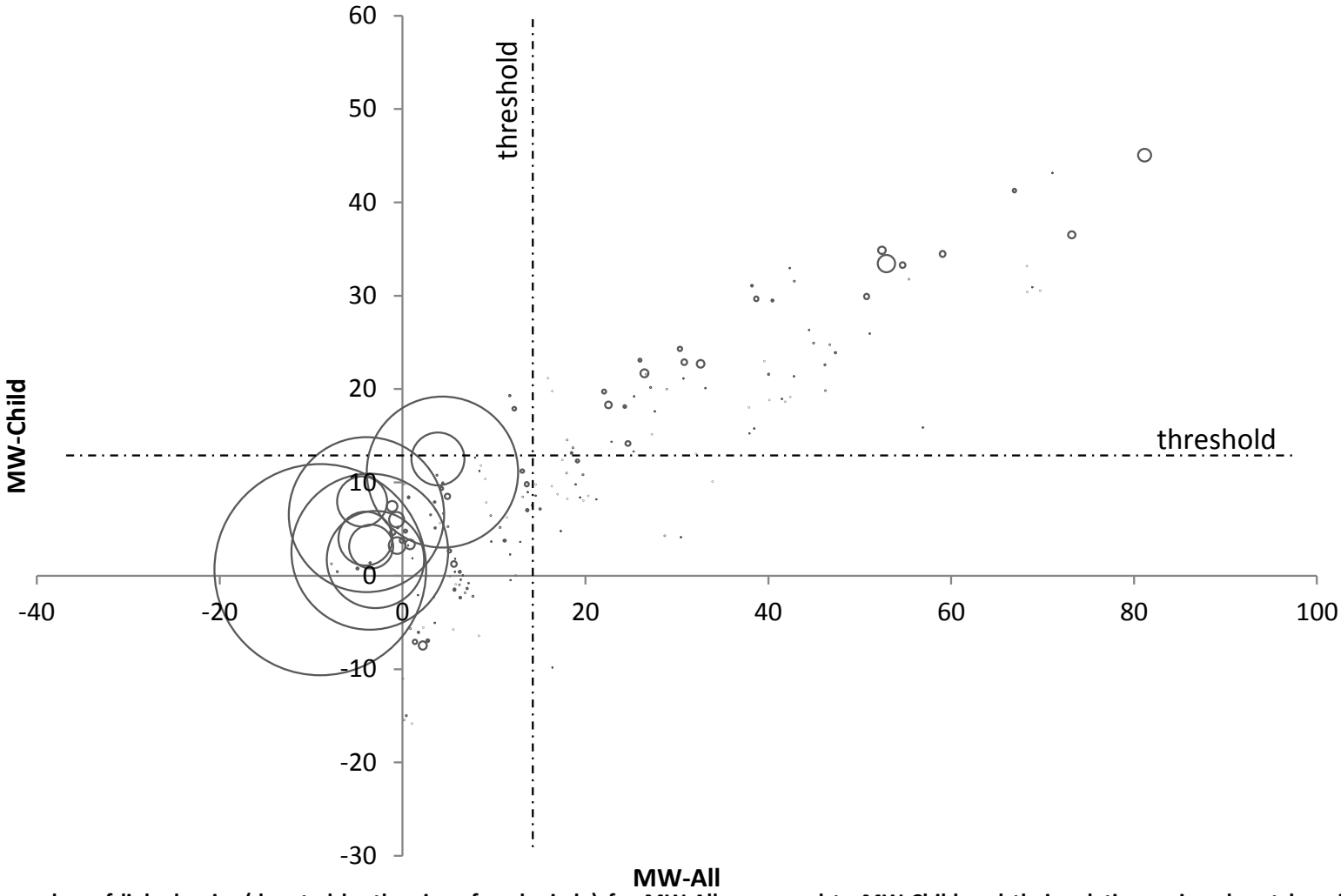


Figure 3.16. The number of linked pairs (denoted by the size of each circle) for MW-All compared to MW-Child and their relative assigned match-weights before the selection of the highest match-weight per OPIE

3.8 Final match-weight set selection

After comparing MW-All and MW-Child, I decided to select the MW-Child results. Firstly, the MW-Child acknowledged and incorporated the data qualities of the HES data, accounting for data completion that the MW-All would not. Secondly, although there were some differences between the MW-All and MW-Child outputs, there was very little difference in the numbers of total linked record pairs, which in this instance would not vastly impact on the results. Thirdly, the MW-Child was focused on the child population and incorporating the data completion values in HES, which MW-All did not.

3.8.1 Data management and preparation of linked cohort for analysis

The highest match-weight link for each OPIE was selected to be taken forward for the analysis. The flow diagram in Figure 3.17 outlines the numbers of observations, OPIEs and HESids at each stage of the data selection, cleaning, linking and manipulation process and the reasons why some observations were removed from the final analysis dataset; the aim was to get a final dataset with one row per paediatric patient, where only the index BSI per patient during the four year period was recorded.

I excluded children (n=1,618) who had been in hospital since birth and who had not been discharged from hospital before developing a BSI. The rationale for this was that they had no possibility of acquiring a BSI in a setting other than in hospital. I also excluded patients who had a positive blood specimen after discharge from hospital but who were not subsequently re-admitted to hospital (n=200).

This resulted in a total of 10,606 children with BSI due to 11,319 bacterial pathogens.

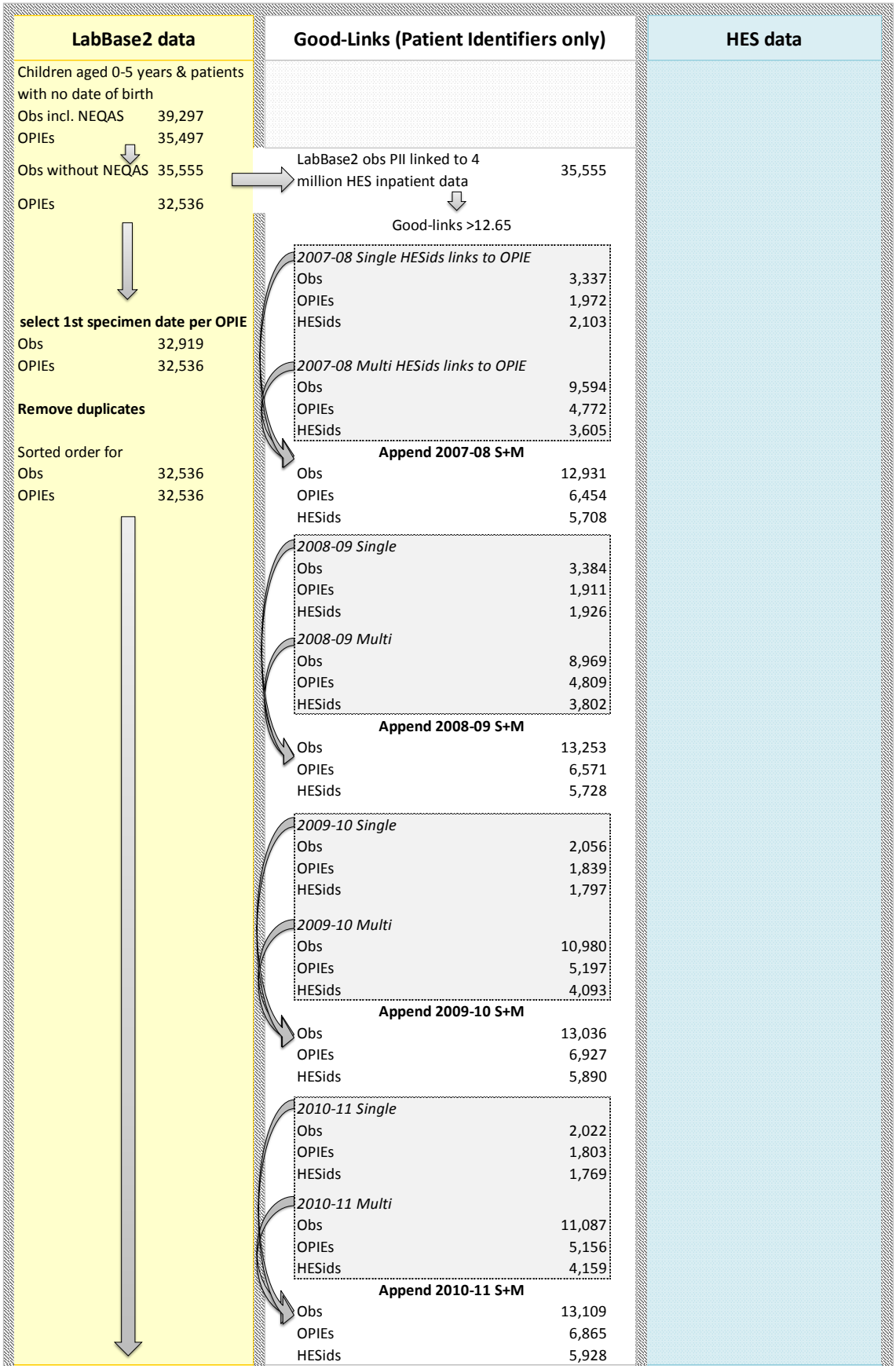


Figure 3.17. Flow diagram of LabBase2, good-links patient identifiers and HES data linkage, processing, manipulation and cleaning.

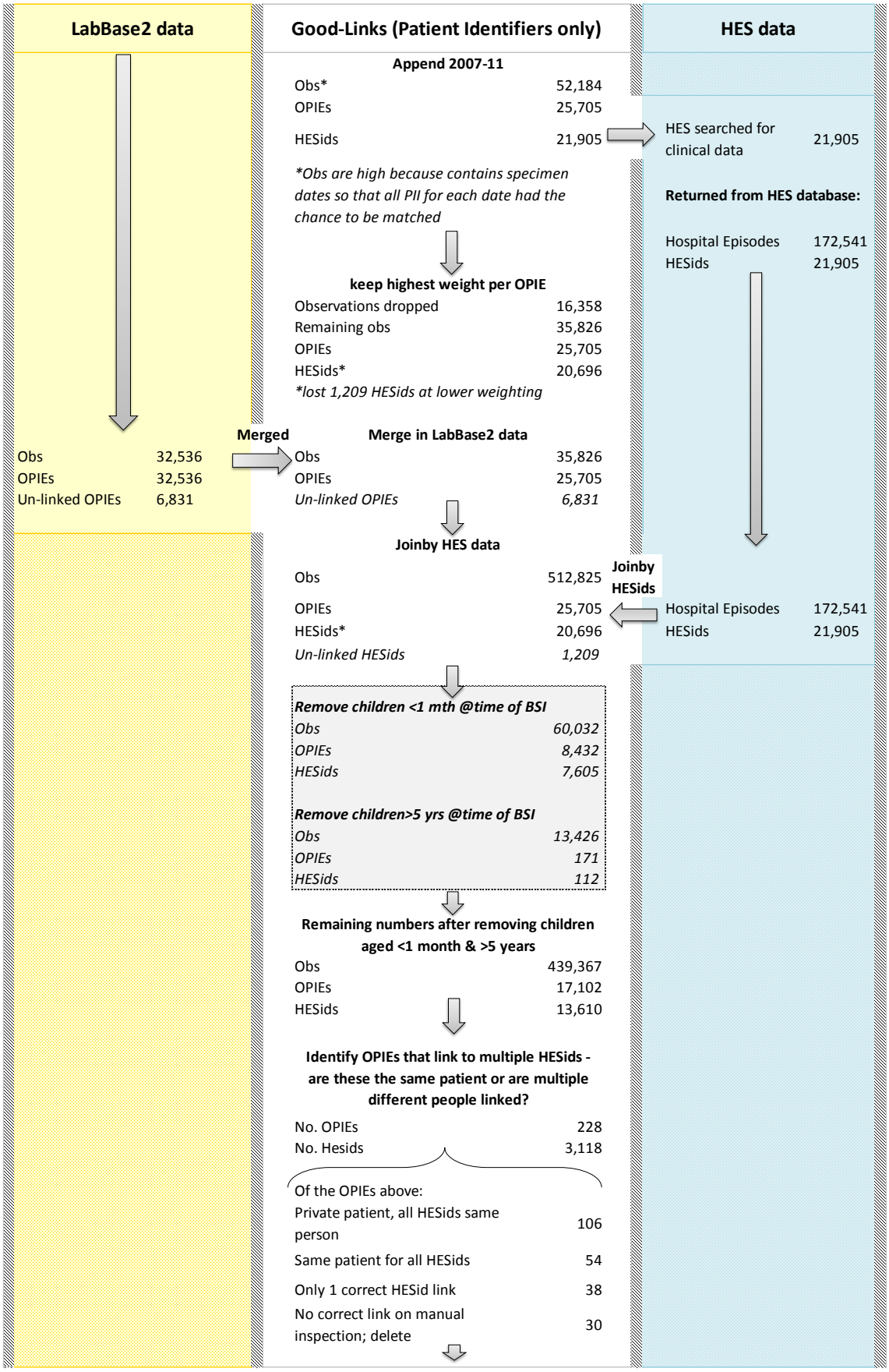


Figure 3.17 *(cont.)* Flow diagram of LabBase2, good-links patient identifiers and HES data linkage, processing, manipulation and cleaning.

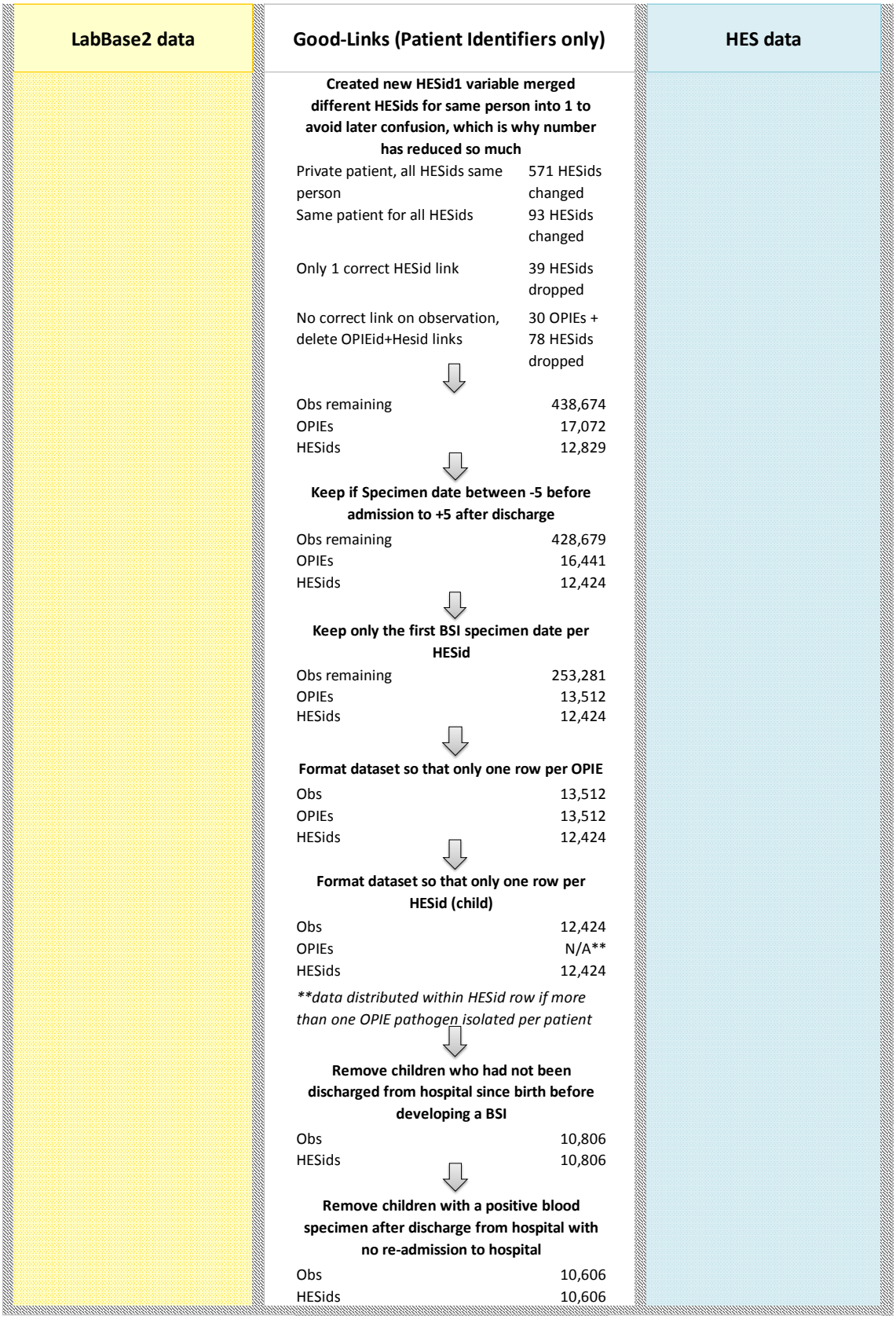


Figure 3.17 *(cont.)* Flow diagram of LabBase2, good-links patient identifiers and HES data linkage, processing, manipulation and cleaning.

3.8.2 Final distribution of links by MW-Child

Figure 3.18 illustrates the final highest match-weight assigned to each linked pair of the 12,424 de-duplicated LabBase2 and HES observations for children aged 1 month – 5 years that was reached in Figure 3.17. The majority (97%) of the match-weights are above the weight of 20, which is comfortably above the >12.65 threshold.

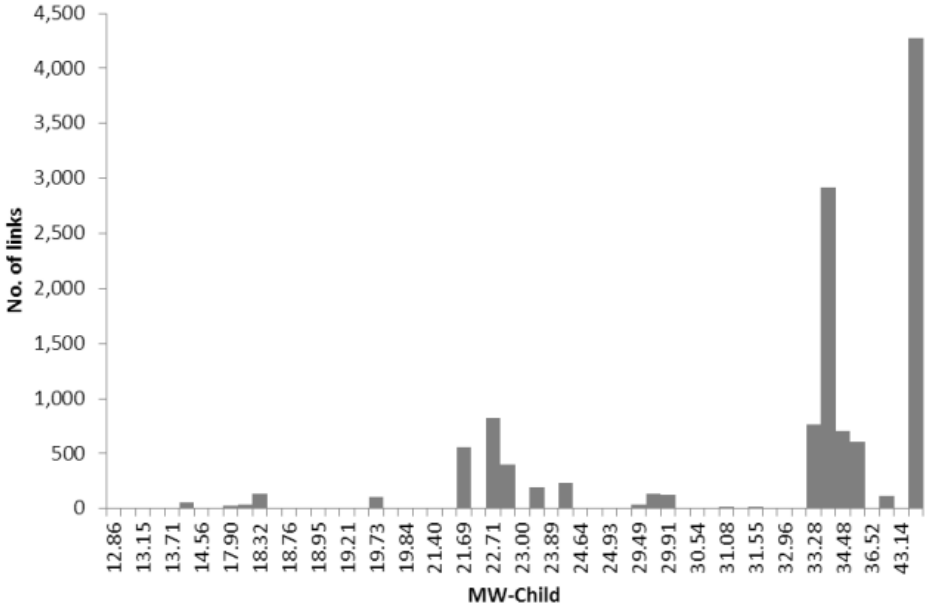


Figure 3.18. MW-Child distribution of the highest match-weight assigned to each linked LabBase2 and HES pair of records for children aged 1 month – 5 years, April 2007 – March 2011

3.9 Discussion

3.9.1 Limitations

Data linkage errors (type I and type II errors) can occur and introduce bias into the results.(146) The selected threshold was not expected to contain any false matches (even though some true or uncertain matches would also have not been included) was used to strike a balance between having high specificity and having the largest possible cohort.

Using the highest match-weight pair per OPIE combination may also have some limitations in the analysis of the linked data. The analysis of not just the highest match-pair, but the subsequent pairs with lower match-weights above the threshold may offer more information as to the probable true matches. Alternatively, the enhancement of patient identifiers using multiple imputation may help to improve the accuracy of the data linkage.(147;148) However,

due to the enormity of the separate analyses required, these approaches were not employed for this thesis.

To enhance the linkage process of MW-Child, field-specific weightings could have been included, which would have taken into account the relative frequency of each of the observations within a field. Zhu *et al*. argue that a match-weight score assigned to an agree-match should be different depending on whether those matching values are common or not.(149) For example, the frequency of the 'ZZ99' postcode prefix, denoting a patient without current UK residence, occurs more frequently than a specific geographical postcode (i.e. 'ME14' prefix) and therefore the 'ZZ99' should be awarded a lower match-weight score for two agreeing linked-pairs than in the case of the 'ME14' prefix.

3.9.2 Summary

Over 80% of all the OPIEs of children with a BSI between April 2007 and March 2011 were successfully linked, and 99% of those OPIEs were linked by both MW-Child and MW-All, demonstrating a high level of similarity in the final cohort selection. This is the first time that match-weights have been calculated to probabilistically link paediatric LabBase2 and HES data. The majority of the disparity between MW-Child and MW-All stemmed from the 'disagree' and 'missing' match-weights, which was reflected by the linked pairs that were either linked by MW-Child or MW-All. Date of birth was more important for the linkage using the child-only population (MW-Child) than the whole population (MW-All), which may be because the age-group in the whole population with the majority of reported BSIs are those aged 65 years and over and these match-weights are calculated from a dataset that is less sensitive for linking child-only records.

This chapter has explained and demonstrated the probabilistic linkage employed to link the LabBase2 and HES data. Although there was only a small difference between the linkage of MW-All and MW-Child, MW-Child linked data best represented the LabBase2 and HES datasets as well as the child population and was therefore selected as the chosen set of match-weights to probabilistically link the datasets.

Chapter 4. Clinical and demographic characteristics of children with a BSI

4.1 Introduction

The literature review in Chapter 1 demonstrated that the timing of BSI alone no longer provides sufficient information to predict the likely invasive pathogen as a result of the increasing mix of pathogens isolated from children on admission to hospital with CA and HCA BSI.(8;61;62;65-68;71) The aim of this PhD is to determine whether clinical characteristics from hospital records in England, in addition to the timing of blood sampling in relation to admission, can be used to predict the likely source of infection and the pathogen to inform strategies to prevent and treat BSI. This chapter aims to quantify the baseline characteristics of children with BSI and identify potential risk factors captured in administrative data that might discriminate between CA, HCA and HA BSI. The selection of potential risk factors needs to be based on understanding the mechanisms influencing the acquisition of BSI in children.

Figure 4.1 illustrates Chapter 4 (in red) in context within the framework of the thesis. The risk factors identified in this chapter will inform the development of the prediction models in chapter 5 to estimate the proportion of CA, HCA and HA BSI.

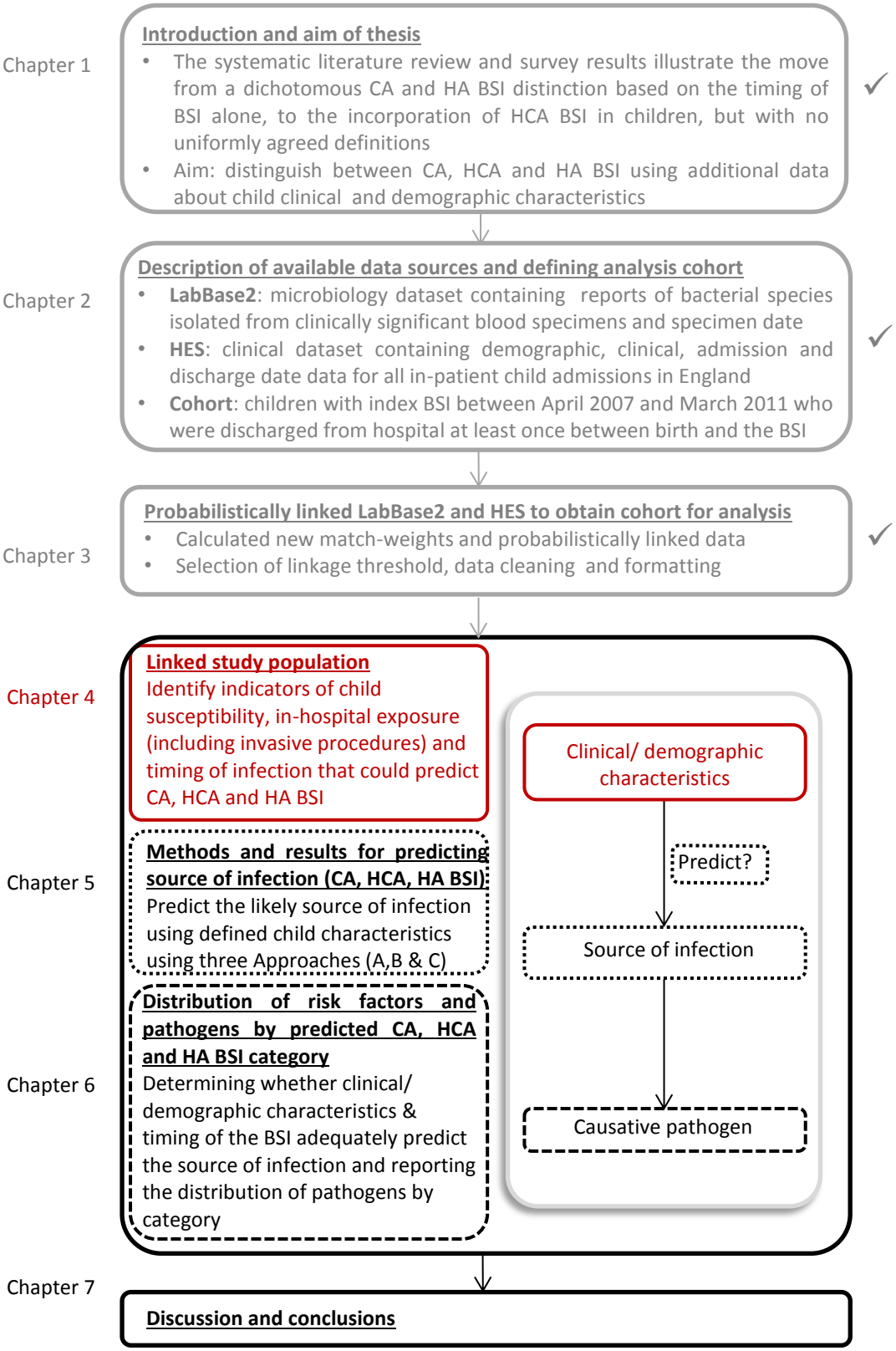


Figure 4.1. Flow diagram of the steps already addressed in previous chapters (grey), those that will be addressed in this Chapter (red) and the following chapters

4.1.1 Mechanisms for acquisition of a BSI

The development of a BSI is associated with children's individual susceptibility to serious infection and their exposure to external factors that enable invasive pathogens to invade the body and multiply. These risk factors can be grouped into three categories: 1) child susceptibility (indirect markers of susceptibility and exposure to invasive BSI), 2) direct exposure to invasive BSI (including breach of skin barrier), and 3) timing of the BSI. These categories are described in more detail in the following three sections and are summarised in Figure 4.2.

4.1.1.1 Child susceptibility to infection

Two factors influencing children's susceptibility to infection are the maturity of the immune system and underlying chronic conditions.

Maturation of immune system

A child's susceptibility to BSIs is shaped by the development of their immune system over time, which has limited functionality at birth. Neonates acquire most protection via vertical transmission from the mother (passive immunity via IgG antibodies) *in utero* and for the first few months of life via breast milk.(150;151) This is important as the child's adaptive immune system will have little or no immunological memory due to lack of direct immune response exposure to the external environment whilst *in utero*. The maternally acquired immunoglobulin G (IgG) protection, transmitted via the placenta, declines over time and breast milk transmitted protection will decline once breast feeding ceases. The component cells forming the innate immune system develop at different rates from birth, affecting the functionality of the immune system compared to a fully grown adult. For example, neonate's neutrophils are immature at birth, which increases their susceptibility to bacterial infections compared to adults with a fully mature immune system. This early-life impairment is amplified in neonates who are born prematurely or with low birth-weight. Other cells, such as monocytes and macrophages, are also immature at birth, resulting in a reduced cytokine response and impaired antibacterial effector functions (such as phagocytosis and oxidative burst) with consequent lack of clearance or killing of potential pathogens.(151;152)

Child susceptibility indicators in HES include codes for pre-term babies, low birth weight and length of postnatal stay. However, vaccination history, as an indicator of child susceptibility to bacterial infection, is not collected in HES.

Chronic conditions

Congenital and other chronic conditions may increase a child's risk of developing a BSI because their immune system is compromised and they are more physically susceptible to CA, HCA and HA infection from different pathogens:

- **HA**: children with chronic conditions will often receive treatment for their condition(s) in hospital, including invasive operations, medical tests and the insertion of indwelling devices for treatment delivery. This increases their exposure to opportunistic pathogens circulating in the hospital.
- **HCA**: children with chronic conditions requiring frequent healthcare contact in the community (e.g. OP or GP clinics), may undergo invasive procedures and the use of indwelling devices for the long-term management of their condition. This increases their exposure to pathogens related to healthcare that are in the community environment.
- **CA**: in the community setting, children with chronic conditions are equally exposed to pathogens causing CA BSI and may be more susceptible than healthy children.

The chronic conditions are measured from the longitudinal health record in HES using ICD-10 diagnostic codes. Children with chronic and congenital conditions will also have more exposure to antibiotics, which can induce the development of resistant bacterial strains that are more commonly associated with HCA and HA BSI; however, this is not an indicator that is captured in HES.

4.1.1.2 Direct inpatient exposure to invasive BSI

Invasive procedures and the insertion of indwelling devices predominantly occur in healthcare settings, especially in IP care. They are indicators for HCA and HA BSI as they disrupt the natural, protective skin barrier and provide easy access for bacteria to enter the body and replicate.(8;61;62;65-68;71) Indwelling devices, in particular, provide a continually open pathway for bacteria naturally found on a patient's skin or from contaminated ward surfaces, to enter directly into the bloodstream. Patients are also exposed to potential pathogens via contact with healthcare workers who may transmit organisms between patients, from direct contact with other patients, and from fomites (e.g. bedding, bed frames, door handles, monitors).(9;95;153;154)

4.1.1.3 Timing of BSI

Time is a key determinant of CA, HCA and HA BSI on three levels.

1) The timing of BSI in relation to the current hospital admission, previous invasive procedure, previous indwelling device and previous hospital discharge are all important indicators of heightened exposure to invasive pathogens. Children with recent healthcare contact have a higher likelihood that the BSI is HCA, whereas a BSI occurring during a current admission indicates a higher likelihood that the BSI is HA. Conversely, the lack of these indicators suggests that the BSI is more likely to be CA BSI. However, this distinction is further complicated by pathogens (e.g. MRSA) that were associated with the hospital, causing infections in people in the community without any indicators for recent healthcare contact (i.e. previously healthy).
2) The timing of blood sample is important when determining the proximity of the BSI to the indicators mentioned above. Blood samples are routinely taken from children on admission to hospital with suspected BSI to microbiologically confirm the pathogen causing the (CA or HCA) BSI. Blood samples are also taken at any point during a hospital admission when a BSI is suspected. Of these, a proportion of children will have a positive blood culture result, a proportion will have a negative blood culture result but with clinically diagnosed BSI, and a proportion will have a negative blood culture with no clinically diagnosed BSI. Antibiotic treatment is also relevant to the timing of when the blood sample is taken, where if the blood sample is taken after antibiotic treatment has been administered, the bacteria may not be detectable in the blood culture. Only positive blood cultures are reported to LabBase2.
3) The bacteria circulating in a given environment is influenced by the time the bacteria are able to survive outside a human host. For example, *N. meningitidis* spreads most successfully from human to human but can only survive for approximately one day outside of a human host on glass or fabric, which is why *N. meningitidis* predominantly causes infection in the community. Conversely, *Enterococcus* spp., MRSA and *Acinetobacter* spp. are resilient and can survive for weeks on hospital surfaces following patient contamination.(95;153;155)

This chapter describes the selection of risk factors that can be measured in LabBase2 and HES.

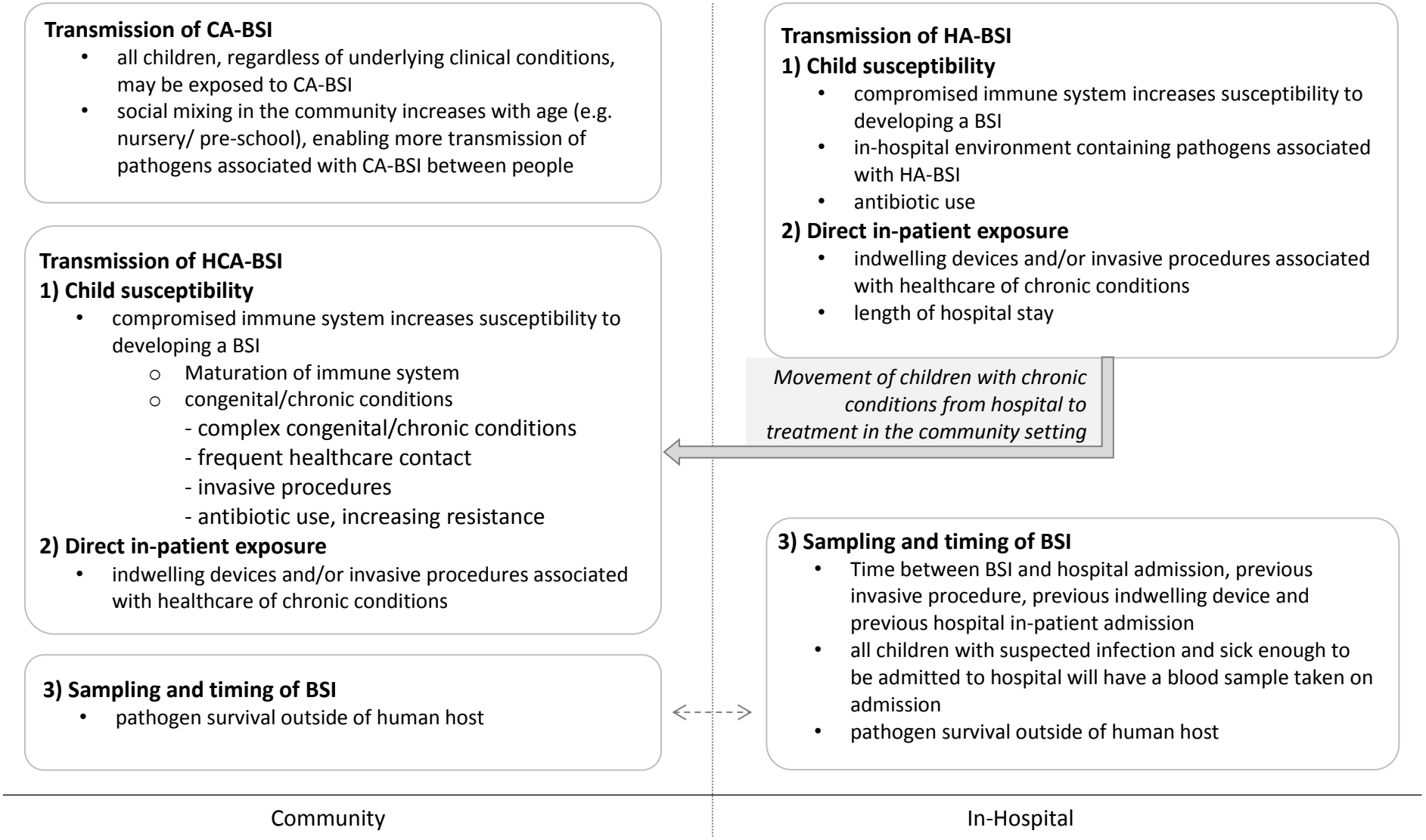


Figure 4.2. Summary of mechanisms of BSI epidemiology in children admitted to hospital

4.2 Methods

This chapter describes a set of risk factors in the LabBase2 and HES linked dataset containing children with a BSI who have all been exposed to the community setting prior to acquiring their first (index) episode of BSI within the 4-year cohort period. The risk factors were detected from IP hospital records within a one-year (1-365 days) look back prior to the date of positive blood sample (day 0). Each risk factor is summarised in Table 4.1.

4.2.1 Child susceptibility

4.2.1.1 Age

The linked cohort was divided into two age-groups, 1 month to < 1 year and aged 1-5 years to reflect the differences in the maturation of the immune system in the first few months of life compared to older children with more adaptive immunity exposure and vaccination cover, as well as differences in social mixing.(51;151) Infants less than 1 year old will have a short look-back period (to birth) to identify risk factors than for children aged 1 year or more.

BSI rates per 100,000 population were calculated using ONS mid-year population data.(156) The rates have been adjusted to reflect the estimated 80% case ascertainment by LabBase2 as discussed in Chapter 2, section 2.3.5 (by increasing the numerator by an additional 20% of the original total). However, rates have not been adjusted for the level of linkage error and serves only to indicate the approximate rate of BSI.

4.2.1.2 Ethnicity

To account for any differences in the distribution of BSI by ethnicity, the HES coding for ethnicity was collapsed into seven groups: White, Mixed, Indian subcontinent, any other Asian, Black, other ethnic group, and not known.(157;158)

4.2.1.3 Evidence of prematurity at birth or low birth-weight

Prematurity at birth or low birth-weight of any child were identified by the ICD-10 diagnostic codes P05 or P07. The baby-tail birth-weight data recorded in the HES baby-tail was not used as the data are incomplete.(159)

4.2.1.4 Evidence of chronic conditions

Chronic conditions were selected from a pre-defined list of ICD-10 codes, grouped into eight clinical categories according to the clinical focus:

- Mental health/behavioural
- Cancer/blood disorders
- Chronic infections
- Respiratory
- Metabolic/endocrine/digestive/renal/genitourinary
- Musculoskeletal/skin
- Neurological
- Codes indicating a non-specific chronic condition

The categorised list of ICD-10 codes was used in a recent study of child deaths in England and can be found in Appendix 4a.(160) All 20 ICD-10 code HES fields were scanned for a matching pre-defined ICD-10 code. A child who had multiple ICD-10 codes from the same clinical focus was coded as having one chronic condition. If multiple ICD-10 codes were identified from two different clinical focus groups, then a child would be coded as having two chronic conditions, and so on. I used these codes because there is no internationally agreed coding system for grouping chronic conditions in children, unlike the Charlson Comorbidity Index for adults.(161)

4.2.1.5 Evidence of congenital conditions

Congenital conditions recorded in any on the 20 ICD-10 code fields (Q01-Q99) were identified using the same ICD-10 groupings from the child death study mentioned above.(160)

4.2.1.6 Length of postnatal stay

The length of postnatal stay (number of nights in hospital since birth or length of an admission within the first seven days of life) was calculated as a proxy indicator for the health status of children aged <1 year. Date of birth was not counted as part of the postnatal stay. Children who were re-admitted for only one day (within the first seven days of life) without staying overnight were recorded as having stayed 0.5 days.

4.2.2 Direct exposure to invasive BSI from inpatient healthcare

To distinguish between prolonged or acute exposure to infection from invasive medical interventions, I divided medical interventions into two groups: indwelling devices (e.g. catheter insertion that can remain inserted for several weeks) and invasive procedures (e.g. surgical procedure lasting a specific number of hours). A previous hospital discharge was used as an indicator of prior healthcare contact.

4.2.2.1 Insertion of last indwelling device

I used procedure codes to identify the most recent insertion date of an indwelling device in the year before the blood specimen date and calculated the number of days between the two dates. Indwelling devices included catheters, shunts, and devices associated with the gastrointestinal area using the codes outlined in Appendix 4b. As these are operative codes, these indwelling devices may represent only those inserted during an operation. Indwelling devices inserted on the wards might not have received a procedure code.

4.2.2.2 Previous invasive procedure

I identified the date of the most recent invasive procedure in the year before the blood specimen date and calculated the number of days between the two dates. To avoid overlap with the indwelling device exposure variable, I excluded any indwelling device codes from the last invasive procedure variable. I corresponded with the lead author to request the list of OPCS invasive procedure codes (Appendix 4b) that were used in the Hardelid *et al*.'s report to distinguish between invasive and non-invasive/low risk for an adverse event procedure.(160) Antibiotic prophylaxis prior to surgery was not captured in HES.

4.2.2.3 Previous hospital discharge

A more recent discharge indicates that the BSI is more likely to be associated with the previous healthcare spell. I calculated the number of days between the last discharge and the current admission date in the year before the blood specimen date.

The measurement of the most recent indwelling device, invasive procedure and hospital discharge before the BSI was categorised into four time groupings: 1-6 days before, 7-30 days before, 31-90 days before and 91-365 days before the BSI (Table 4.1). Lookback thresholds of 30-day, 90-day and 1-year have been commonly used in the literature.(8;61;62;65;67) England's Surgical Site Infection Surveillance Service (SSISS) uses a 30-day threshold for the

monitoring of surgical site infections (SSIs) for superficial, deep incisional and organ space infections where there was no implant involved, or a 1-year threshold if an implant is used.(162) I included an additional threshold of day 1-6 before the BSI to add more detail to capture any events in very close proximity to the BSI.

Table 4.1. Categories for clinical and demographic exposure variables

Category	Variable	Type	Description
susceptibility of child	**Age**	Continuous	Between 29 – 2191 days old (1 month - 5 years)
	Gender	Binary	0 = male 1 = female
	Ethnicity	Categorical	0 = White 1 = Mixed 2 = Indian subcontinent 3 = Any other Asian 4 = Black 5 = Other ethnic group 6 = Not known
	Premature birth/low birth-weight	Binary	0 = babies weighing ≥2500g at birth and/or born ≥37 weeks gestation 1 = babies weighing <2500g at birth and/or born <37 weeks gestation
	Chronic condition(s)	Categorical	0 = no chronic conditions in year prior to the BSI* 1 = at least 1 ICD-10 code for any chronic condition within the same chronic condition group* in year prior to the BSI 2 = at least 1 ICD-10 code for any chronic condition* from two different chronic condition groups* in year prior to the BSI 3 = at least 1 ICD-10 code for any chronic condition* from three or more chronic condition groups* in year prior to the BSI
	Congenital condition(s)	Binary	0 = no congenital conditions in year prior to the BSI 1 = at least 1 ICD-10 code for any congenital condition in year prior to the BSI
	Postnatal stay	Continuous	Number of nights in hospital after birth, or number of nights in hospital when admitted in the first 7 days of life
In-patient exposure to infection	**Indwelling device**	Categorical	0 = no indwelling devices in year before the BSI specimen date 1 = date of last indwelling device insertion between 91 days<1 year before the BSI specimen date 2 = date of last indwelling device insertion between 31-90 days before the BSI specimen date 3 = date of last indwelling device insertion between 7-30 days before the BSI specimen date 4 = date of last indwelling device insertion between 1-6 days before the BSI specimen date

Category	Variable	Type	Description
In-patient exposure to infection	**Time since last invasive procedure**	Categorical	0 = no invasive procedure in year before the BSI specimen date 1 = last invasive procedure between 91 days<1 year before the BSI specimen date 2 = last invasive procedure between 31-90 days before the BSI specimen date 3 = last invasive procedure between 7-30 days before the BSI specimen date 4 = last invasive procedure between 1-6 days before the BSI specimen date
	Time between current admission date and last hospital discharge date	Categorical	0 = no previous hospital discharge in the year before the BSI specimen date 1 = last discharge between 91 days<1 year before the current admission date 2 = last discharge between 31-90 days before the current admission date 3 = last discharge between 7-30 days before the current admission date 4 = last discharge between 1-6 days before the current admission date

*outlined in appendix 4a

4.2.3 Timing of positive blood samples in relation to hospital admission

The timing of the blood sampling in relation to hospital admission was determined by identifying the hospital spell where the blood specimen date fell between the admission and discharge dates, or within the five days prior to the admission date to capture GP samples, OP samples or A&E samples prior to being admitted as an IP for their infection.

4.3 Results

4.3.1 Child susceptibility

Between 1st April 2007 and 31st March 2011, 4,422 children aged 1 month < 1 year and 6,184 children aged 1-5 years had an index case of BSI. Figure 4.3 illustrates the frequency distribution of BSI by child age (categorised into 3-month blocks). The highest number of positive blood specimens were taken in the youngest children aged 1 month to < 3 months (n = 1,595). The frequency of blood specimens subsequently decreased rapidly in the first year of life, having halved for children aged 6 months to <9 months (n=816), and continued to decrease with increasing age. The lowest number of positive specimens were for children aged 5 years or more (range: 147-166 isolates/3-months).

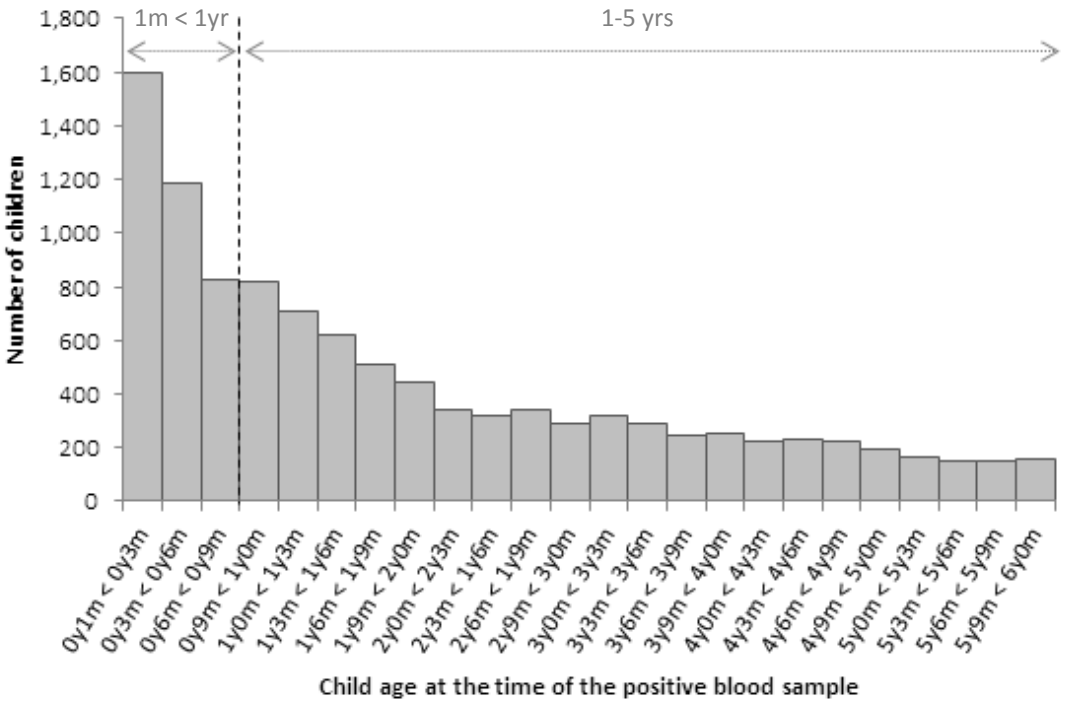


Figure 4.3. Frequency distribution of children with a BSI by quarter year of age

The average ratio of BSI in children aged 1 month < 1 year was 3.9 BSI to every 1 BSI in children aged 1-5 years (Table 4.2). By year, for every child aged 1< 2 years with a BSI (n = 2,283), there were nearly twice as many children (n = 4,422) aged 1 month < 1 year with a BSI. The comparative ratio increased to 7.8 BSI in children aged 1 month < 1 year (n = 4,422) for each BSI in children aged 5 < 6 years (n = 619).

Table 4.2. Ratio of BSI for children aged 1 month < 1 year compared to children aged 1-5 years by year of age

1m < 1 yr age-group		Ratio of BSI by age category	1-5 yrs age-group	
category	No. of children with BSI		No. of children with BSI	category
1m < 1yr	4,422	2.1 : 1	2,283	1 < 2 yrs
1m < 1yr	4,422	3.7 : 1	1,288	2 < 3 yrs
1m < 1yr	4,422	4.3 : 1	1,113	3 < 4 yrs
1m < 1yr	4,422	5.5 : 1	881	4 < 5 yrs
1m < 1yr	4,422	7.8 : 1	619	5 < 6 yrs
1m < 1yr	**4,422**	**3.9 : 1**	**1,237**	**year average**

Between 1st April 2007 and 31st March 2011, the average rate of BSI in children aged 1 month < 1 year was 209.0/100,000 population compared to an average rate of 62.4/100,000 population in children aged 1-5 years. The annual rates are shown in Table 4.3. The rate by population is four times higher in younger children (aged 1 month < 1 year) compared to the older children (1-5 years).

Table 4.3 Rate of BSI by 1,000 bed-days and by 100,000 population, adjusted for the approximate 80% voluntary reporting of laboratory data to LabBase2

Age-group	HES year	No. of children with BSI (estimated 80% captured by LabBase2)	Estimated no. of children with BSI (adjusted to 100%)	Estimated BSI rate/100,000 population	
1m < 1 yr					
	2007/08	1,050	1,312	204.2	*
	2008/09	1,069	1,336	200.1	*
	2009/10	1,100	1,375	207.5	*
	2010/11	1,203	1,504	223.8	*
1-5 yrs					
	2007/08	1,622	2,027	68.3	
	2008/09	1,472	1,840	60.3	
	2009/10	1,509	1,886	59.9	
	2010/11	1,581	1,976	61.4	

*ONS population based on all children aged 0<1 year; could not differentiate child age in months

The characteristics of children with a BSI are summarized by age-group in Table 4.4. More BSIs were isolated from boys than girls in both the 1 month > 1 year and 1-5 years age-groups (59.2% and 56.4% respectively). There were similar ethnic group proportions in both age-groups, with White being the predominant ethnic group (67.9% and 67.5% respectively).

Nearly 20% of the younger children were either premature or had a low birth-weight, compared to 9% in the older age-group.

A total of 52% (n = 2,279) of children aged 1 month < 1 year and 56% (n = 3,467) of children aged 1-5 years had either a chronic or congenital condition (or both) regardless of the blood sample date. Children with at least one chronic condition and congenital condition accounted for 18% (n = 799/4,422) of children aged 1 month < 1 year and 10% (600/6,184) of children aged 1-5 years. Of children with blood samples taken prior to or on the day of admission, 39% (n = 1,139/2,886) of children aged 1 month < 1 year and 50% (n = 2,225/4,420) of children aged 1-5 years had at least one chronic condition.

Eighty-three per cent (n = 3,671) of children aged 1 month < 1 year had a birth record in HES and a postnatal stay (range= day-case to 331 nights; mean= 11.5 nights; median= 2 nights). The frequency and distribution of the length of postnatal stay for the 3,671 children is illustrated in Figure 4.4. The majority (58%; n = 2,121) of postnatal stays were two-nights or less, where 13% (n = 493) were day cases only, 26% (n = 945) were one-night hospital spells and 19% (n = 483) were two-night hospital spells.

Table 4.4. Susceptibility of child: Study population demographic and clinical characteristics by age-group

Category	Variable	1m<1yr	%	1-5yrs	%	Total	%
susceptibility of child	**Age group**	4,422	41.7%	6,184	58.3%	**10,606**	100%
	Sex						
	Boy	2,618	59.2%	3,488	56.4%	**6,106**	57.6%
	Girl	1,804	40.8%	2,696	43.6%	**4,500**	42.4%
	Ethnicity						
	White	2,999	67.8%	4,173	67.5%	**7,172**	67.6%
	Mixed	143	3.2%	198	3.2%	**341**	3.2%
	Indian subcontinent	411	9.3%	555	9.0%	**966**	9.1%
	Any other Asian	125	2.8%	174	2.8%	**299**	2.8%
	Black	245	5.5%	398	6.4%	**643**	6.1%
	other ethnic group	163	3.7%	201	3.3%	**364**	3.4%
	NK	336	7.6%	485	7.8%	**821**	7.7%
	Premature birth/low birth-weight						
	no	3,546	80.2%	5,629	91.0%	**9,175**	86.5%
	yes	876	19.8%	555	9.0%	**1,431**	13.5%
	Chronic and/or congenital condition						
	none	2,139	48.4%	2,708	43.8%	**4,847**	45.7%
	yes	2,283	51.6%	3,476	56.2%	**5,759**	54.3%
	Chronic conditions						
	none	2,311	52.3%	2,782	45.0%	**5,093**	48.0%
	1 chronic condition	1,123	25.4%	1,703	27.5%	**2,826**	26.6%
	2 chronic conditions	437	9.9%	777	12.6%	**1,214**	11.4%
	>2 chronic conditions	551	12.5%	922	14.9%	**1,473**	13.9%
	Congenital conditions						
	none	3,451	78.0%	5,510	89.1%	**8,961**	84.5%
	1 or more congenital conditions	971	22.0%	674	10.9%	**1,645**	15.5%
	Postnatal stay						
	No postnatal stay	751	17.0%	-	-	**751**	17.0%
	Children with postnatal length of stay	3,671	83.0%	-	-	**3,671**	83.0%

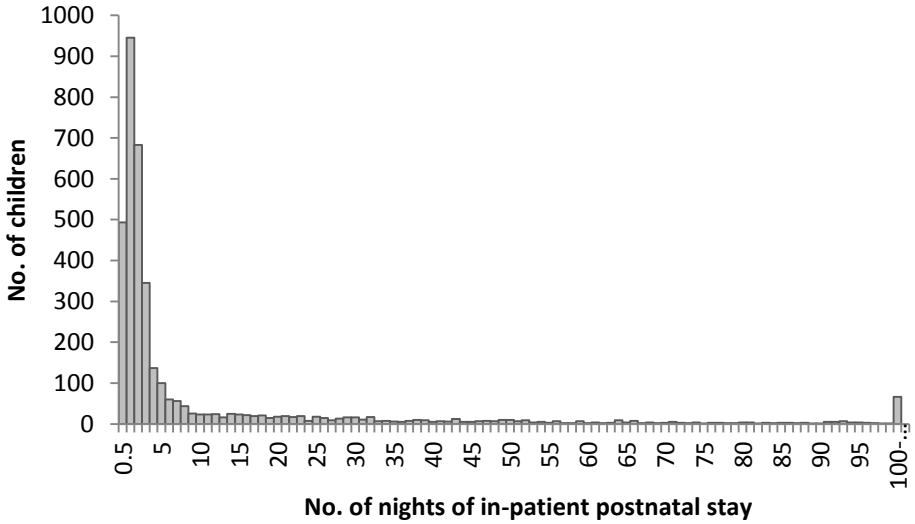


Figure 4.4. Frequency distribution of inpatient postnatal length of stay (no. of nights) for children aged 1 month < 1 year (0.5 symbolises day admissions only)

4.3.2 Direct inpatient exposure to infection

Children with both an invasive procedure and an indwelling device in the past year accounted for 12% (n = 527/4,422) of the children aged 1 month < 1 year and 20% (n = 886/6,184) of children aged 1-5 years (Table 4.5). The proportion of children with a record of the most recent invasive procedure was similar in both age-groups, where 33.5% aged 1 month < 1 year and 29.7% aged 1-5 years did not have any evidence of an invasive procedure in the year prior to the BSI. A total of 13.0% of children aged 1 month < 1 year and 15.6% of children aged 1-5 years had a record of having an indwelling device inserted in the past year before the BSI. Forty-three per cent (n = 2,643) of children aged 1-5 years had a previous discharge from hospital in the year before the BSI. In comparison, 87.1% of children aged 1 month < 1 year had at least one recorded discharge from hospital in the year prior to the BSI (includes discharge after birth delivery).

Table 4.5. In-patient exposure to infection: study population demographic and clinical characteristics by age-group

Category	Variable	1m<1yr	%	1-5yrs	%	Total	%
In-patient exposure to infection	**Indwelling device and/or invasive procedure in the year before the BSI specimen date**						
	no	2,933	66.3%	4,338	70.1%	7,271	68.6%
	yes	1,489	33.7%	1,846	29.9%	3,335	31.4%
	No. days since last indwelling device insertion						
	no indwelling device in the year before the BSI specimen date	3,846	87.0%	5,217	84.4%	**9,063**	85.5%
	last insertion 91 days < 1 yr	151	3.4%	346	5.6%	**497**	4.7%
	last insertion 31-90 days	192	4.3%	262	4.2%	**454**	4.3%
	last insertion 7-30 days	160	3.6%	246	4.0%	**406**	3.8%
	last insertion 1-6 days	73	1.7%	113	1.8%	**186**	1.8%
	No. of days since last invasive procedure						
	no invasive procedure in the year before the BSI specimen date	2,941	66.5%	4,347	70.3%	**7,288**	68.7%
	91 days<1 yr	289	6.5%	401	6.5%	**690**	6.5%
	31-90 days	399	9.0%	361	5.8%	**760**	7.2%
	7-30 days	410	9.3%	550	8.9%	**960**	9.1%
	1-6 days	383	8.7%	525	8.5%	**908**	8.6%
	No. of days between last discharge and current admission						
	no previous discharge in the year before the BSI specimen date	570	12.9%	3,541	57.3%	**6,111**	57.6%
	91 days<1 yr	1,384	31.3%	734	11.9%	**2,118**	20.0%
	31-90 days	1,222	27.6%	460	7.4%	**1,682**	15.9%
	7-30 days	692	15.6%	616	10.0%	**1,308**	12.3%
	1-6 days	554	12.5%	833	13.5%	**1,387**	13.1%

4.3.3 Timing of positive blood samples in relation to hospital admission

The frequency distribution of children with BSIs is shown by the index blood culture sample date in relation to day of hospital admission (day 0 is day admitted to hospital) by age-group in Figure 4.5 and Figure 4.6 (log scale on the large graph, and normal scale on the sub-set graph). The majority of BSIs from all children (85%; n=8,983/10,606) were isolated <2 days after hospital admission. The highest frequency of BSI causing pathogens was observed on day 0 for children aged 1 month < 1 year (61%; n=2,673/4,415) and children aged 1-5 years (67%; n=4,138/6,184), followed by day 1 for children aged 1 month < 1 year (16%; n=686/4,422) and children aged 1-5 years (16%; n=991/6,184) of hospital admission.

The timing between hospital admission and the BSI sample date was more skewed (longer tail) in children aged 1 month < 1 year (range=0-763 days; mean=6.3 days; median=0 days) compared to children aged 1-5 years (range=0-268 days; mean=3.7 days; median=0 days). The mean length of stay in hospital for the admission containing the BSI was 20.8 days (median=5 days, range=0-928 days) in children aged 1 month < 1 year and 12.2 days (median=4 days; range=0-1,028 days) in children aged 1-5 years.

Children from both age-groups with a positive blood specimen taken 2-5 days (0.8%; n=80/10,606) before hospital admission were from a range of sources. The majority (75%; n=60) were emergency admissions from a combination of A&E units, GP or OP admissions. Thirteen admissions were elective planned admissions, three were elective booked admissions and four were samples taken at another hospital prior to being transferred, however, no record of the respective transferring hospital reports were found in HES, which is why the admission to the current hospital was considered a new admission.

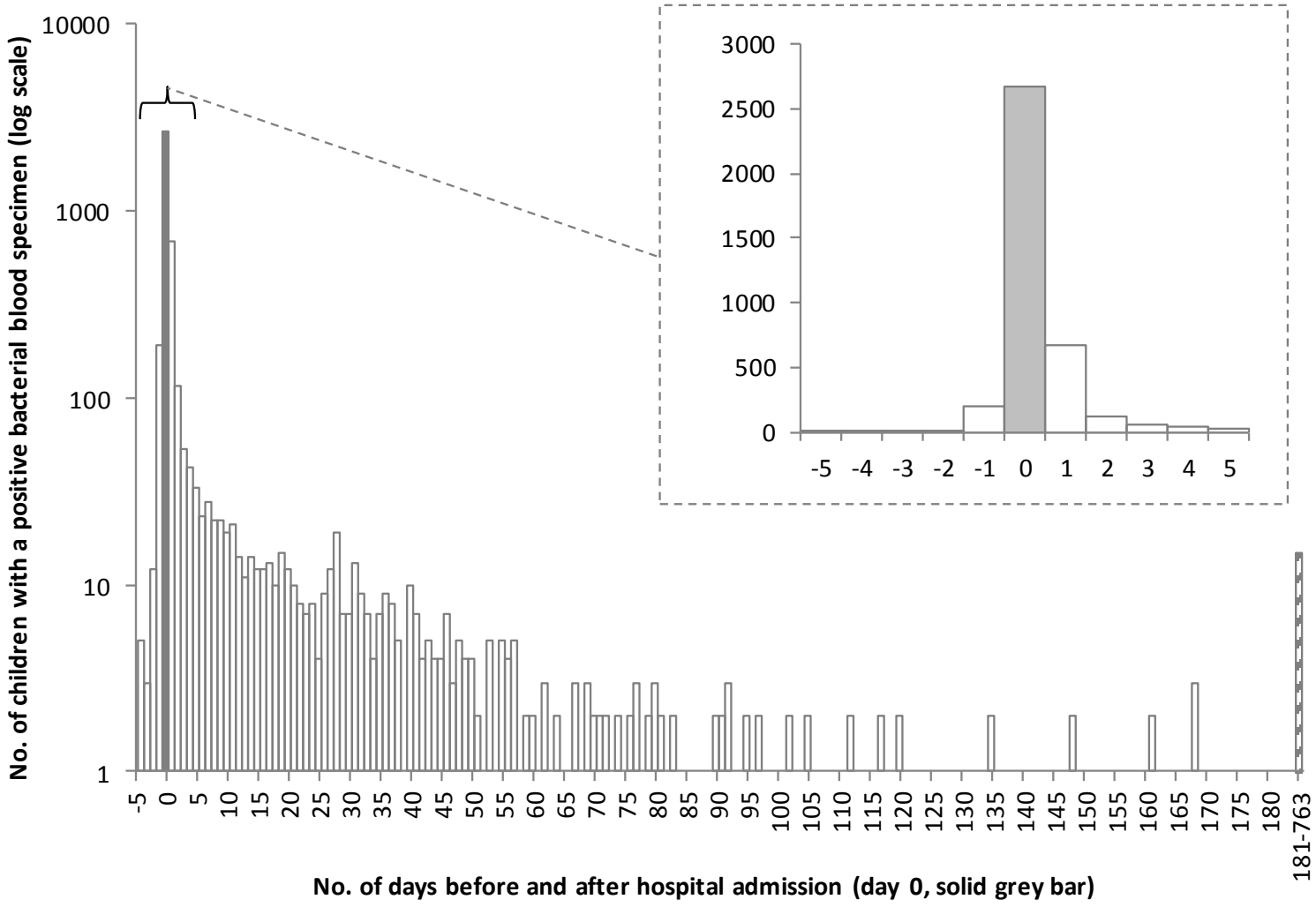


Figure 4.5. Frequency and timing when positive bacterial blood specimens were taken in relation to hospital admission (day 0) for children aged 1 month < 1 year, 2007-2011. Use of log-scale in main graph, and normal-scale in the excerpt

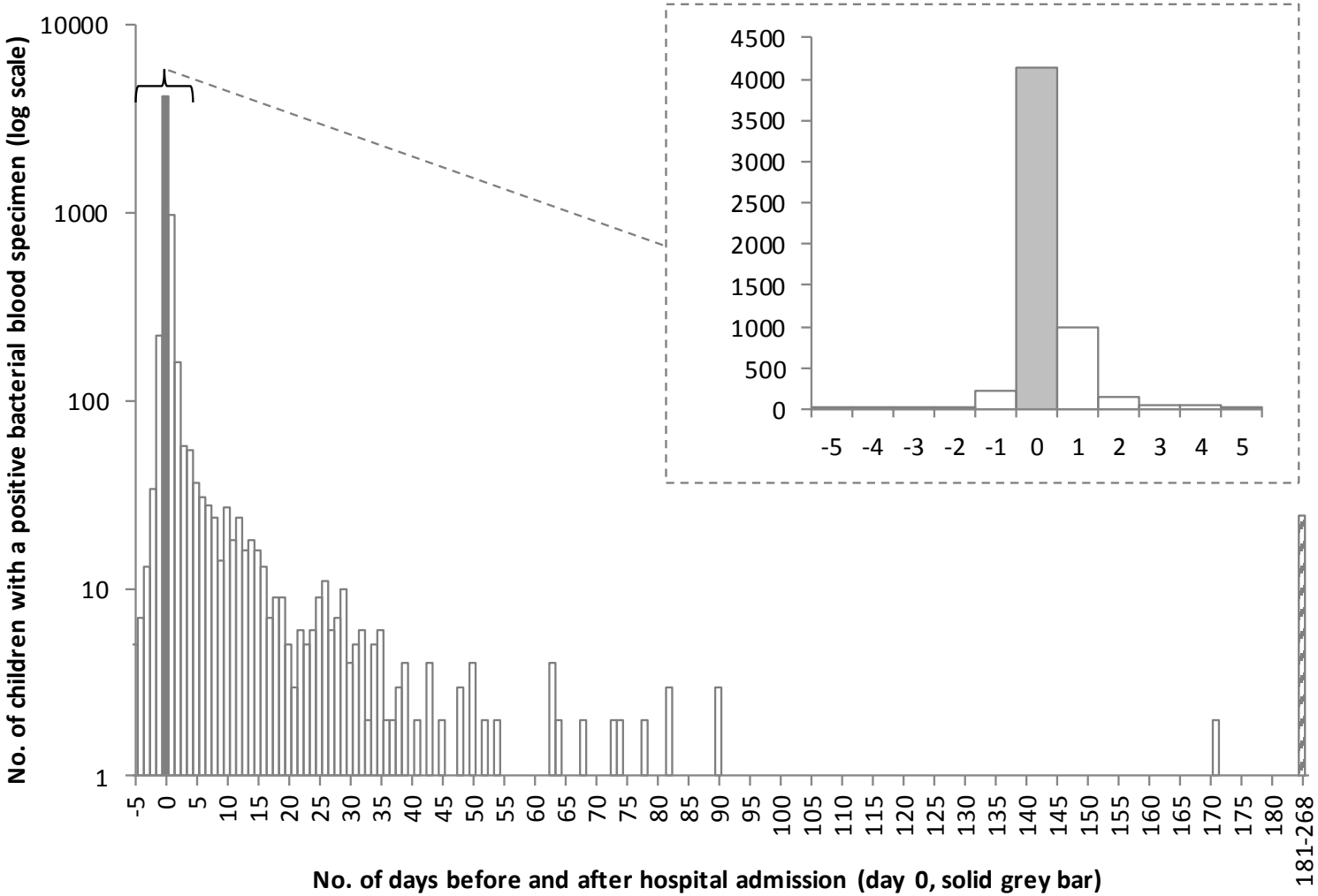


Figure 4.6. Frequency and timing when positive bacterial blood specimens were taken in relation to hospital admission (day 0) for children aged 1 - 5 years, 2007-2011. Use of log-scale in main graph, and normal-scale in the excerpt

4.4 Discussion

4.4.1 Child susceptibility to BSI

There were proportionally more BSIs among children aged 1 month < 1 year than any other year-group of children, with the largest difference observed (ratio 7:1 BSI) between the youngest and oldest year-group (5<6 years old) respectively. The rate of BSI was higher in the younger age-group per 100,000 child population. Proportionally more male than female children had BSIs in both age-groups, most noticeably in children aged 1 month < 1 year. Half of children, regardless of the timing of their BSI, had complex morbidity from either a chronic or congenital condition and were therefore potentially more susceptible to infections from a wider range of pathogens than children who were previously healthy (according to the reported data) prior to the BSI. In comparison to the literature, the proportion of children with chronic conditions was lower than the 88% of children with a BSI in a single-centre children's hospital in England (1995-2002) although this was a tertiary referral centre that would receive proportionally more complex cases than the average hospital.(7) A more recent study reported similar (58%) proportion of children presenting from the community-setting with chronic conditions in a South-West London study (2009-11).(163)

4.4.2 Direct exposure to invasive BSI from inpatient healthcare

A third of all children had exposure to previous invasive procedures and recorded indwelling devices within the past year. The majority of BSIs in both age-groups were isolated less than two days after hospital admission. For in-hospital exposure, children aged 1 month < 1 year had longer lengths of hospital stay compared to children aged 1-5 years, giving them greater exposure to HA BSI. A recent study reported that more children aged <1 year were admitted (and re-admitted) as an emergency admission to hospital in England compared to children aged 1-4 years and 5-9 years.(124)

The majority of children aged 1 month < 1 year had IP hospital contact within the year prior to the BSI and postnatal stay. There was, however, a higher than expected proportion of children aged 1 month < 1 year who had no birth delivery record (17%). This may be due children born at home or born abroad. According to the Office for National Statistics (ONS), only 2-3% of births take place outside of an NHS hospital.(164) It may also reflect linkage error. In an ONS study of birth registrations occurring in 2007, linked birth records and NHS Numbers for Babies

(NN4B) records were linked to baby HES records (not the mothers' records), where 16% were un-linkable.(164)

4.4.3 Timing of positive blood samples in relation to hospital admission

The largest proportion of positive blood specimens in both age-groups (60% in children aged 1 month < 1 year; 67% in children aged 1-5 years) were taken on the day of hospital admission compared to any other day, highlighting the large burden of BSI (mixture of CA and HCA) coming from the community setting in both age-groups.

4.4.4 Limitations of cohort

A potential limitation of the inclusion criteria for the cohort was whether children with a minimum exposure (1 day) in the community had sufficient exposure to pathogens circulating in the community. Only a very small proportion of children (n=34/4,415, 0.8%) aged 1 month < 1 year were re-admitted to hospital the day after birth before subsequently developing a BSI during that hospital admission where 91% (n=31) had a chronic or congenital condition and 56% (n=19) had an indwelling device inserted in the past year. They had a very similar distribution of clinical characteristics to the 1,618 children who were excluded from the cohort for being in hospital since birth when they acquired their BSI, where 89% (n=1,444) had at least one chronic or congenital condition and 46% (n=745) had an indwelling device inserted within the past year before the BSI. The majority (86%; n=1,385/1,618) of excluded children were aged 1<3 months old. Although a small minority of children whose true exposure to the community setting is debatable, all of these children developed BSI after having spent at least four weeks in hospital and therefore would have been classified as having HA BSI regardless. The aim of the thesis was to examine risk factors in all children admitted with BSI to determine whether clinical characteristics could be partitioned according to likely source of infection: CA, HCA or HA BSI. The study was not designed to answer questions about how to prevent BSI because it did not include children without BSI. Therefore, I did not attempt to determine risk factors for BSI.

The categorisation of chronic diseases into different body systems (0, 1, 2, >2 chronic conditions) is a crude indicator of patients with chronic conditions and multi-morbidity. This coding is inevitably crude because important clinical information is missing from the HES data, such as immune-suppression, antibiotic treatment and test results. The chronic condition categorisation is used to determine infection risk and could be used to target certain groups

for treatment or not (similar to targeting of vaccines to those with at risk conditions). The chronic conditions were grouped together because small numbers in specific chronic disease categories would reduce the power of the regression models. In practice, clinicians will be using all of the available information (including previous antibiotic prescribing, underlying chronic conditions and test results from other clinical samples) to inform their clinical decision about treatment.

Chapter 5. Methods and Results: Predicting the proportion of CA, HCA and HA BSI

5.1 Introduction

Predicting the proportion of CA, HCA and HA BSI in children is a difficult problem to resolve due to the complex mechanisms of BSI that were described in Chapter 4. In the absence of universally accepted definitions of HCA BSI, it is reasonable to use the linked cohort of children with BSI (linked using the MW-Child described in Chapter 4) to determine whether there are subgroups of children with distinct clinical and demographic characteristics in addition to the timing of infection that might infer different sources of BSI (CA, HCA and HA).

The rationale for trying to predict the proportions of CA, HCA and HA BSI is that this information will enable more efficient and appropriate targeting of prevention BSI strategies in children; determining whether the source of infection leads to a clear separation in the causative pathogens will be investigated in Chapter 6.

In this chapter, I present three models (denoted A, B and C) that were developed, using the linked dataset described in Chapter 4, to estimate the proportion of CA, HCA and HA BSI in children. These models incorporate different assumptions about the relationships between the three BSI categories. This is an exploratory method where the aim is not to select one model over another, but to obtain a range of likely estimates of the proportions.

I begin by introducing the definitions and assumptions made about CA, HCA and HA BSI that were employed for the analytic approaches. The methodologies of the three statistical approaches are subsequently described, followed by the results of each approach. The results from the three approaches are then compared and discussed in Chapter 6. Figure 5.1 illustrates the content of this chapter in relation to the thesis.

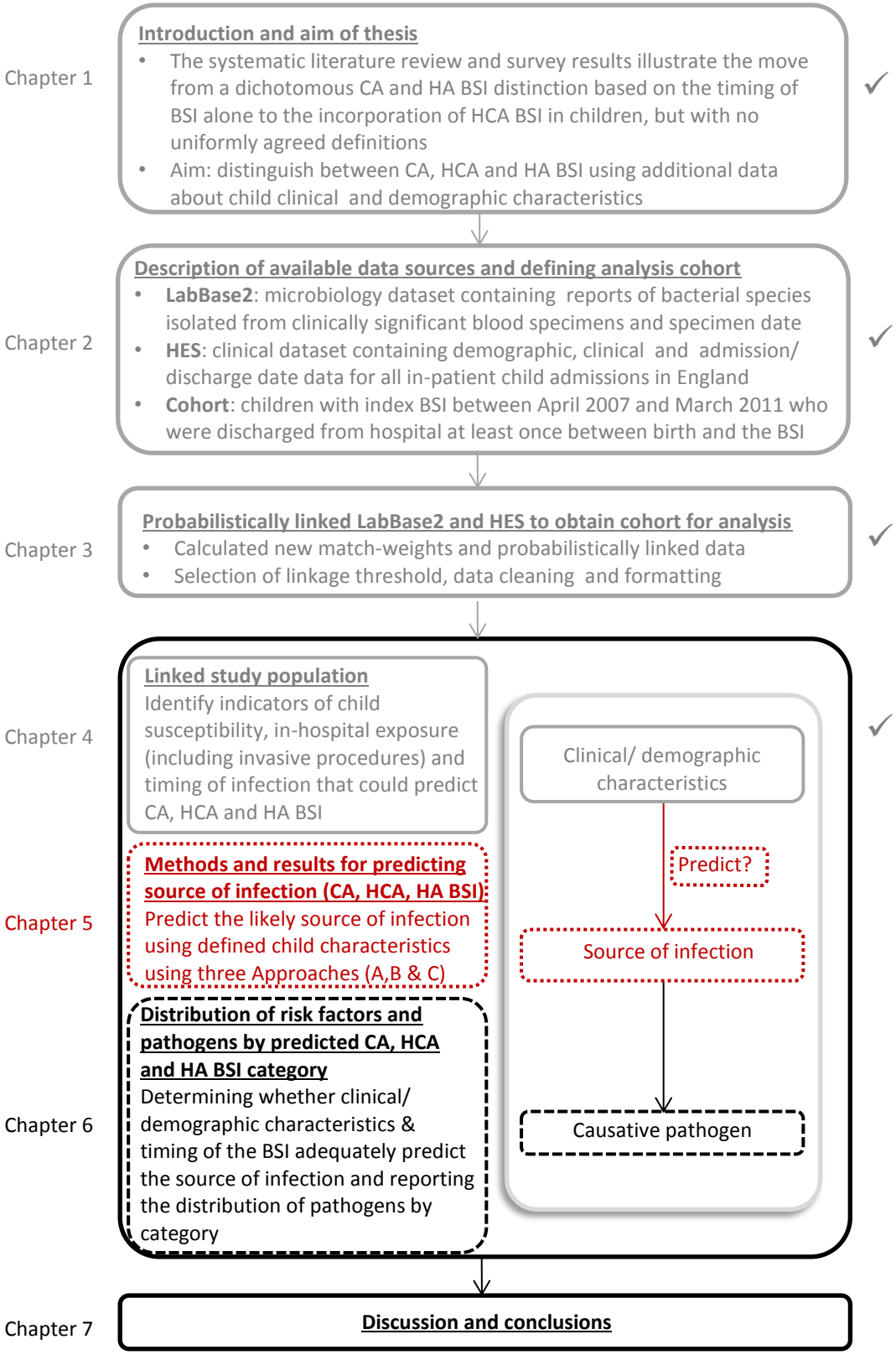


Figure 5.1. Flow diagram of the steps taken (grey) and those that will be addressed in this chapter (red) to predict the proportion of CA, HCA and HA BSI in children

5.2 Definitions and assumptions about the linked cohort

In this chapter, assumptions were made about some BSIs in relation to timing or the causative pathogen prior to looking at the actual data obtained (described in Chapter 6). The Figures in sections 5.2.1, 5.2.2, 5.3.1 and 5.3.2 are for illustrative purposes only and do not reflect real data distributions.

For two of the statistical approaches (A and B), it was necessary to identify both a group of children with '*defined* HA' BSI and a group who were '*a priori* CA' to give 2 binary baseline variables (*defined* HA or not, *a priori* CA or not).

I excluded child records that detected two or more isolated pathogens where one pathogen was classified as *defined* HA BSI and another pathogen was classified as *a priori* CA to avoid any double counting of BSI episodes in the regression models. This was applied from the first threshold HA1 (defined below) for consistency across all the models because the number was low.

5.2.1 Defining children with HA BSI

HA BSIs were assumed to occur between day 1 of hospital admission and hospital discharge, to encapsulate both before and after the commonly used 2-day threshold. To test different time thresholds relating to the development of HA versus that of CA or HCA BSI, the time threshold was varied from 1 day up to 9 days post admission, subsequently referred to as HA1-HA9 and illustrated in Figure 5.2. The HA9 threshold was chosen to ensure the majority, if not all, pathogens causing CA or HCA BSI would not be included in the subgroup treated as *defined* HA. The HA1-HA9 range covers a likely over-estimate of *defined* HA (HA1) through to probable under-estimation (HA9) via this initial starting dichotomy. BSI on the day of hospital admission (day 0) were not included in the possible HA range because the time (in hours) of admission to hospital was not recorded in HES, making it impossible to determine the length of stay on the first day of admission and therefore whether or not it was possible for a patient to develop a HA BSI in that time.

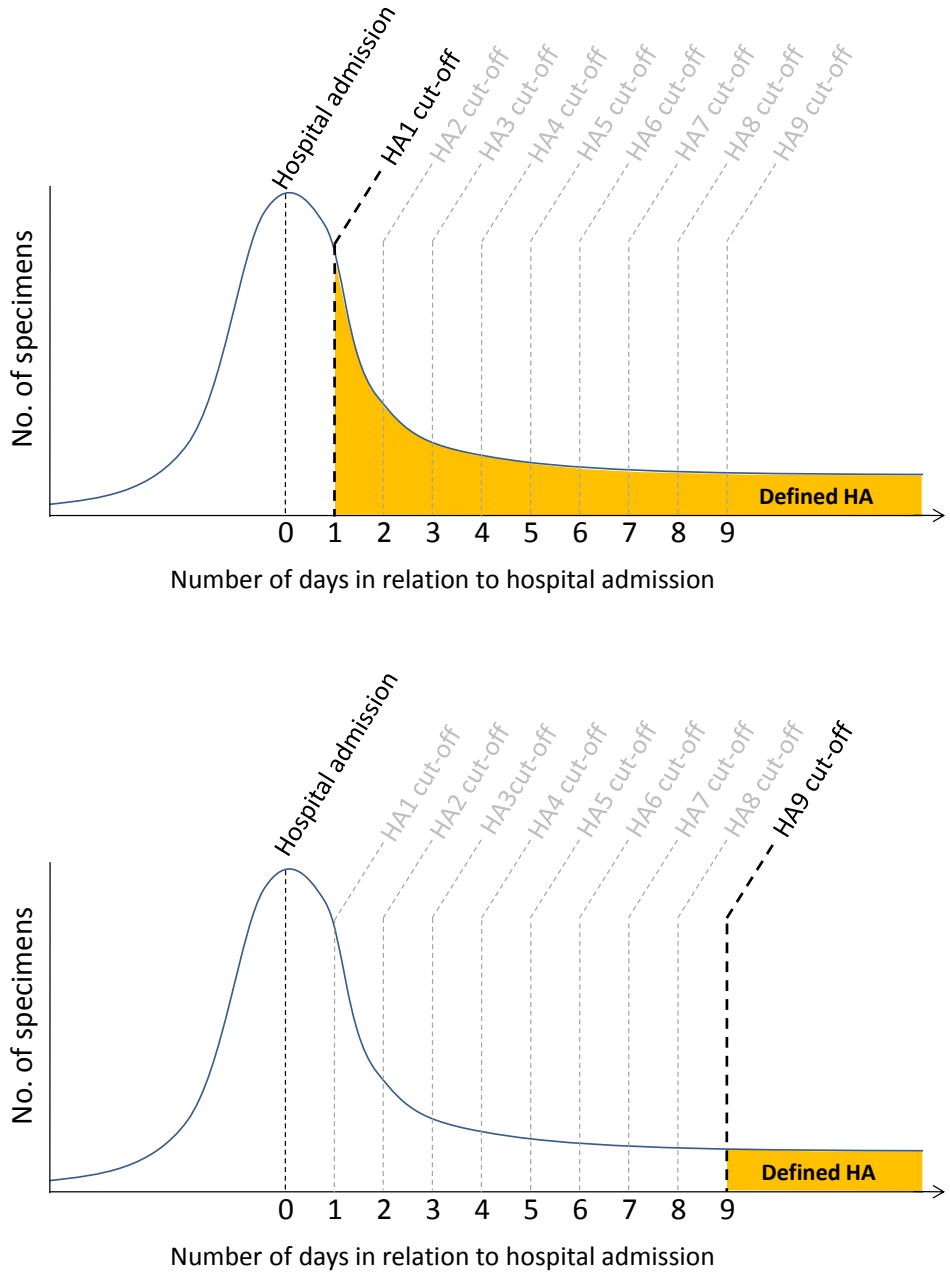


Figure 5.2. Expected timing of specimen date sample in relation to hospital admission: Varying HA1-HA9 cut-off dates for *a priori* HA BSI selection

5.2.2 Defining children with *a priori* CA BSI

In the literature review and professional opinion survey in Chapter 1, *Neisseria meningitidis, Salmonella* spp., *S. pneumoniae* and GAS were identified as predominantly causing CA BSI, with little or no evidence of causing HA BSI.(8;72;75;78;165) This is supported by other published literature.(165) The survival time outside of a human host is short for both *N. meningitidis* bacteria (<1.25 days on glass or fabric) and *Salmonella* spp. (1-3 days on inanimate

surfaces).(155;166-168) Streptococcal bacteria, however, are more resilient, where *S. pneumoniae* can survive from <1-15 days on glass or fabric, and GAS can survive between 3 days and 6.5 months on environmental surfaces if they remain dry and untouched by cleaning products.(155;168)

Although GBS in children is predominantly perceived as either vertically transmitted pathogen from mother to baby at birth or causing CA BSI, there is some evidence that GBS can also be acquired in hospital.(169;170) It was not included as one of the *a priori* CA BSI pathogens due to the occurrence of late-onset sepsis, which can occur up to 90 days after exposure at birth.

Children with BSI caused by *S. pneumoniae, N. meningitidis,* GAS or *Salmonella* spp. were labelled as having *a priori* CA BSI, regardless of the day the positive blood specimen was detected in relation to admission, including BSI occurring after the definite HA BSI threshold because I assumed delayed detection of an underlying CA BSI.

Figure 5.3 demonstrates the expected timing of *a priori* CA BSI (blue) in relation to *defined* HA (yellow), where day x varied from 1 to 9 (HA1-HA9), and the remaining mixed CA and HCA BSI prior to day 1 of hospital admission (dark grey) and mixed CA, HCA and HA BSI between day 1 or more of hospital admission and day x (light grey).

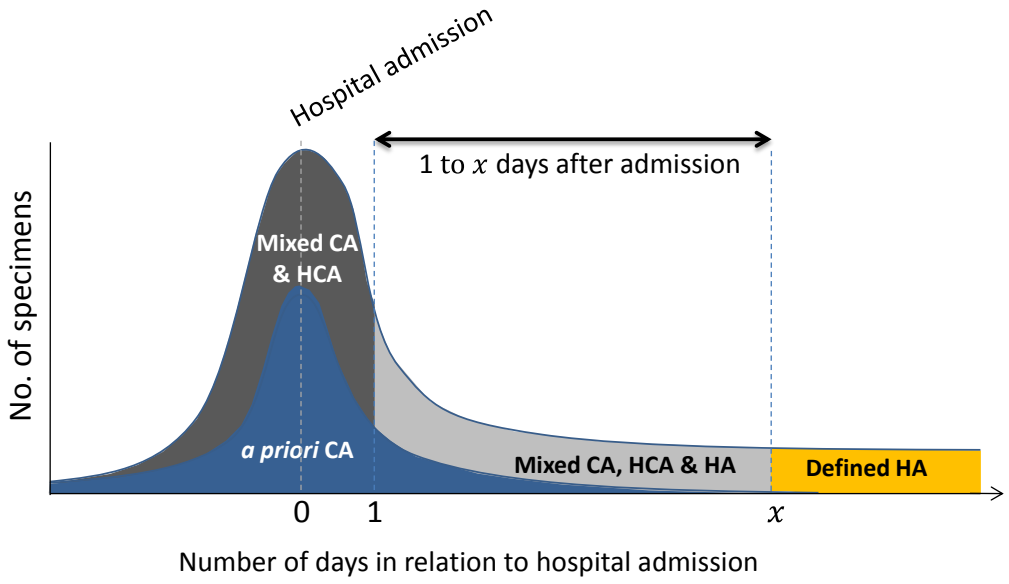


Figure 5.3. Expected timing of specimen date sample in relation to hospital admission for *a priori* CA BSI

5.2.3 Results

A total of 19 children with BSI were classified as both *defined* HA (using the HA1 threshold) and *a priori* CA and were removed from the linked dataset.

5.2.3.1 Children with *defined* HA BSI

The number of children with *defined* HA BSI (that were not *a priori* CA) for each of the nine scenarios is tabulated by age-group in Table 5.1. Proportions varied from 30% in children aged 1 month < 1 year and 22% in children aged 1-5 years in the lowest HA-threshold (HA1) down to 12% and 6% respectively for the highest (most conservative) threshold (HA9). The largest reduction between consecutive thresholds was between HA1 and HA2 (12% decrease in children aged 1 month < 1 year; 11% decrease in children aged 1-5 years).

Table 5.1. Number of children with *defined* HA BSI for each of the nine threshold scenarios by age-group

	Age-group					
***Defined* HA-threshold**	**1m < 1 y**	**%**	**1-5y**	**%**	**Total**	**%**
HA1 (day 1 or more)	1,325	30%	1,379	22%	2,704	26%
HA2 (day 2 or more)	803	18%	688	11%	1,491	14%
HA3 (day 3 or more)	708	16%	569	9%	1,277	12%
HA4 (day 4 or more)	661	15%	523	8%	1,184	11%
HA5 (day 5 or more)	625	14%	478	8%	1,103	10%
HA6 (day 6 or more)	598	14%	449	7%	1,047	10%
HA7 (day 7 or more)	578	13%	419	7%	997	9%
HA8 (day 8 or more)	552	13%	394	6%	946	9%
HA9 (day 9 or more)	530	12%	373	6%	903	9%
Total population	4,415		6,172		10,587	

The distributions of children with *defined* HA BSI for thresholds HA1 and HA9 by clinical and demographic characteristics are shown in Table 5.2. The proportion of children with any of the clinical characteristics increased between day 1 and day 9, highlighting that the children with a positive blood specimen occurring later into a hospital admission had more complicated medical conditions that required longer medical attention due to the nature of their illness or condition.

Table 5.2. Tabulation of *defined* HA1 and HA9 cases by demographic and clinical characteristics from HES in the 12 months prior to the BSI

Characteristic	Aged 1m<1y		Aged 1-5y		Aged 1m<1y		Aged 1-5y	
	HA1	%	HA1	%	HA9	%	HA9	%
Sex								
Boy	759	57%	762	55%	295	56%	216	58%
Girl	566	43%	617	45%	235	44%	157	42%
Ethnicity								
White	821	62%	916	66%	290	55%	232	62%
Mixed	44	3%	61	4%	16	3%	8	2%
Indian subcontinent	147	11%	136	10%	56	11%	39	10%
Any other Asian	47	4%	44	3%	19	4%	10	3%
Black	79	6%	78	6%	37	7%	18	5%
other ethnic group	59	4%	46	3%	35	7%	23	6%
Not Known	128	10%	128	9%	77	15%	43	12%
Premature birth/low birth-weight								
no	959	72%	1,216	88%	319	60%	314	84%
yes	366	28%	163	12%	211	40%	59	16%
Chronic conditions								
No chronic conditions	462	35%	397	29%	42	8%	7	2%
1 chronic condition	269	20%	353	26%	89	17%	53	14%
2 chronic conditions	210	16%	222	16%	121	23%	72	19%
>2 chronic conditions	384	29%	407	30%	278	52%	241	65%
Congenital conditions								
No congenital conditions	756	57%	1,093	79%	174	33%	240	64%
1 or more congenital	569	43%	286	21%	356	67%	133	36%
Days between last discharge and current admission								
no previous discharge in the year before BSI	207	16%	584	42%	121	23%	117	31%
91 days<1 yr	243	18%	171	12%	30	6%	46	12%
31-90 days	302	23%	118	9%	56	11%	37	10%
7-30 days	267	20%	198	14%	119	22%	63	17%
1-6 days	306	23%	308	22%	204	38%	110	29%
Days since last invasive procedure								
no invasive procedure in the year before BSI	526	40%	624	45%	75	14%	42	11%
91 days<1 yr	84	6%	87	6%	25	5%	22	6%
31-90 days	157	12%	102	7%	100	19%	48	13%
7-30 days	274	21%	268	19%	227	43%	171	46%
1-6 days	284	21%	298	22%	103	19%	90	24%
No. days since last indwelling device insertion								
No device in the year before BSI	939	71%	933	68%	251	47%	132	35%
last insertion 91 days<1 yr	74	6%	112	8%	37	7%	45	12%
last insertion 31-90 days	115	9%	109	8%	83	16%	62	17%
last insertion 7-30 days	138	10%	157	11%	123	23%	109	29%
last insertion 1-6 days	59	5%	68	5%	36	7%	25	7%
Postnatal stay in the 12 months prior to the BSI								
No postnatal stay	266	20%	-	-	126	24%	-	-
Postnatal stay	1,059	80%			404	76%		
Total	1,325		1,379		530		373	

5.2.3.2 Children with *a priori* CA BSI

A total of 848 children aged 1 month < 1 year and 1,711 children aged 1-5 years were classified as having *a priori* CA BSI. As demonstrated in Table 5.3, the majority (90-91%) of the BSI that were assumed to be *a priori* CA in both age-groups were isolated on day 0 and day 1 of hospital admission, which correlates with the timing of when the majority of the blood samples were taken from children as see in Chapter 4. Four to 5% of *a priori* CA BSI were identified prior to hospital admission. A further 5% of *a priori* CA were isolated from samples taken between day 2 and day 11 (median = day 3) of hospital admission in children aged 1 month < 1 year, and between day 2 and day 17 (median = day 2) of hospital admission in children aged 1-5 years. The relative proportion of *a priori* CA BSIs by pathogen are shown in Figure 5.4 (children aged 1 month < 1 year) and Figure 5.5 (children aged 1-5 years).

Table 5.3. Distribution of the number of children with *a priori* CA BSI by day of bacterial isolation in relation to hospital admission (day 0)

Day BSI isolated in relation to hospital admission (day 0)	Age-group				Total
	1m<1y	(%)	1-5y	(%)	
-5 to -1 days	38	(4.5%)	74	(4.3%)	112
0 to 1 days	764	(90.1%)	1,554	(90.8%)	2,318
2-9 days	44	(5.2%)	79	(4.6%)	123
10-17 days	2	(0.2%)	4	(0.2%)	6
Subtotal of *a priori* CA	848		1,711		2,559
Total children with BSI	4,415		6,172		10,587

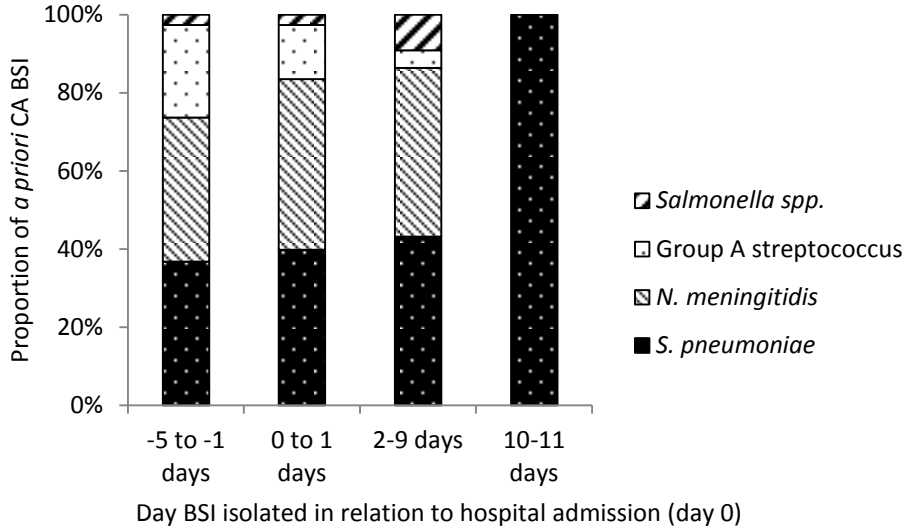


Figure 5.4. Proportional distribution of *a priori* CA pathogens in children aged 1 month < 1 year in relation to hospital admission

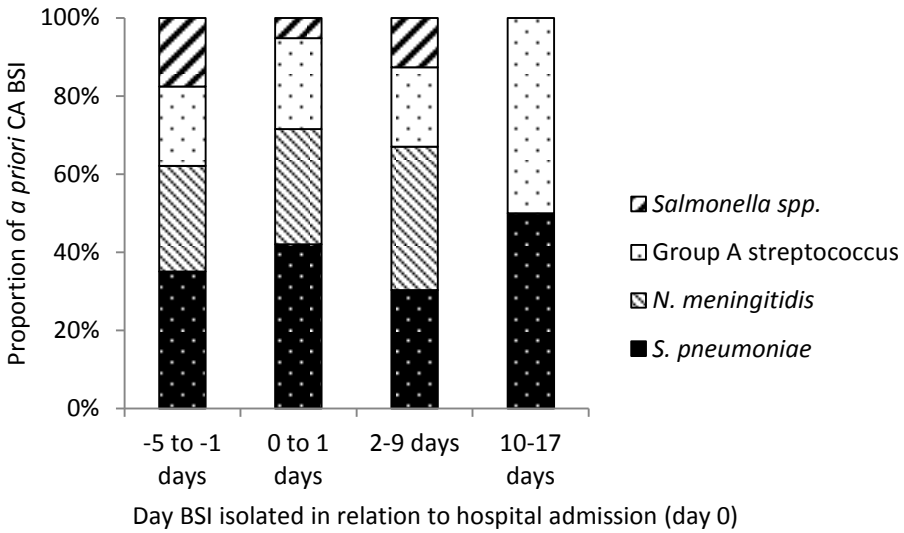


Figure 5.5 Proportional distribution of *a priori* CA pathogens in children aged 1-5 years in relation to hospital admission

The demographic and clinical characteristics of children with *a priori* CA BSI are shown in Table 5.4, both including and excluding the CA BSIs occurring two or more days after hospital admission; this table does not include any duplicate data from the previous *defined* HA BSI table (Table 5.2) because all *a priori* CA BSI occurring after the HA-threshold remained *a priori* CA (as described in section 5.2.2).

The majority (≥96%) of children with *a priori* CA BSI did not have any history of an indwelling device in the year before the BSI, less than 15% were born prematurely or with a low birth-weight, and 83% of children < 1 year old were re-admitted for a certain time period within the first week of discharge from hospital after birth. No previous discharge within the past year was noted for 75% of children aged 1-5 years. Half of children in both age-groups had at least one chronic condition. The frequency of these clinical characteristics in children with *a priori* CA BSI were significantly different when compared to children with *defined* HA BSI (Chi^2 $P<0.005$). There was no significant difference by sex for children with *a priori* CA or *defined* HA BSI in either age-group (Chi^2 $P=0.051$ for children aged 1 month < 1 year; Chi^2 $P=0.699$ for children aged 1-5 years).

Table 5.4. Tabulation of *a priori* CA cases by demographic and clinical characteristics

Characteristic		all *a priori* CA				CA <=2 days after admission			
		1m<1y		1-5y		1m<1y		1-5y	
		CA	%	CA	%	CA	%	CA	%
Sex									
	Boy	480	57%	984	58%	464	56%	956	57%
	Girl	368	43%	727	42%	359	44%	715	43%
Ethnicity									
	White	663	78%	1,184	69%	641	78%	1,158	69%
	Mixed	23	3%	62	4%	22	3%	60	4%
	Indian subcontinent	53	6%	133	8%	52	6%	129	8%
	Any other Asian	7	1%	33	2%	7	1%	33	2%
	Black	23	3%	98	6%	22	3%	96	6%
	other ethnic group	22	3%	44	3%	22	3%	42	3%
	NK	57	7%	157	9%	57	7%	153	9%
Premature/low birth-weight									
	No	734	87%	1,585	93%	711	86%	1,552	93%
	Yes	114	13%	126	7%	112	14%	119	7%
Chronic conditions									
	None	436	51%	881	51%	425	52%	864	52%
	1 chronic condition	300	35%	561	33%	292	35%	550	33%
	2 chronic conditions	75	9%	157	9%	72	9%	153	9%
	>2 chronic conditions	37	4%	112	7%	34	4%	104	6%
Congenital conditions									
	No	777	92%	1636	96%	759	92%	1,599	96%
	Yes	71	8%	75	4%	64	8%	72	4%
Postnatal stay									
	No	148	17%	-	-	139	17%	-	-
	Yes	700	83%	-	-	684	83%	-	-
No. days between last discharge and current admission									
	No reported previous discharge in past year	112	13%	1,277	75%	106	13%	1,251	75%
	91 days<1 yr	446	53%	182	11%	438	53%	178	11%
	31-90 days	163	19%	84	5%	161	20%	83	5%
	7-30 days	60	7%	64	4%	55	7%	62	4%
	1-6 days	67	8%	104	6%	63	8%	97	6%
Days since last invasive procedure									
	No invasive procedure in the past year	696	82%	1,503	88%	686	83%	1,478	88%
	91 days<1 yr	60	7%	80	5%	59	7%	79	5%
	31-90 days	30	4%	40	2%	27	3%	37	2%
	7-30 days	12	1%	31	2%	10	1%	29	2%
	1-6 days	50	6%	57	3%	41	5%	48	3%
No. days since last indwelling device insertion									
	No indwelling device in the past year	815	96%	1,655	97%	797	97%	1,623	97%
	last insertion 91 days < 1 year	16	2%	30	2%	13	2%	29	2%
	last insertion 31-90 days	4	0%	11	1%	3	0%	8	0%

Characteristic	all *a priori* CA				CA <=2 days after admission			
	1m<1y		1-5y		1m<1y		1-5y	
	CA	%	CA	%	CA	%	CA	%
last insertion 7-30 days	1	0%	6	0%	1	0%	5	0%
last insertion 1-6 days	12	1%	9	1%	9	1%	6	0%
Total	848		1,711		823		1,671	

5.2.4 Summary and discussion

I have described the assumptions made when determining a sub-group of children with *defined* HA BSI, using different time thresholds after hospital admission, and *a priori* CA BSI, identified by the causative pathogens *S. pneumoniae, N. meningitidis,* GAS and *Salmonella* spp.

Children with *defined* HA BSI as opposed to *a priori* CA BSI had higher proportions of chronic conditions (65-98% versus 50%, P<0.0005), indwelling devices (32-75% versus 3-4%, P<0.0005) and invasive procedures (55-89% versus 12-18%, P<0.0005). It seems reasonable that other CA or HA BSI will be in children with similar distributions of clinical characteristics, i.e. where children with HA BSI will have proportionally more recent and frequent invasive procedures and chronic conditions. These results are similar to previous studies that found children with CA BSI were generally healthier and had less healthcare contact compared to children with HA BSI.(8;72) However, it was surprising that half of the children with *a priori* CA BSI had at least one chronic condition, emphasising that this group of children were a mixture of healthy and immunocompromised children before the BSI.

The 129 children with *a priori* CA isolates reported two or more days after hospital admission may be due to other circumstances not necessarily related to transmission within the hospital setting. An initial blood sample may have been negative due to antibiotic treatment prior to the blood specimen being taken, making the bacterial load in the blood undetectable. Delayed sampling, where a BSI was not initially suspected by the clinician, may also partly explain the later timing of the blood specimen date. Unfortunately, these theories cannot be verified in the current data.

In the next section, I describe the statistical analysis models that were to attempt to differentiate between children with CA, HCA or HA BSI using clinical and demographic patient characteristics. Nine different predicted proportions of CA BSI, HCA BSI and HA BSI in children, related to the different HA1-HA9 cut-offs, were outputted for each age-group using each of the 3 methods.

5.3 Methods of the statistical approaches A, B and C

I used three different approaches to estimate the proportional burden of CA, HCA and HA BSI in children by using clinical and demographic risk factors in the linked dataset containing 10,,587 children with BSI (described in Chapter 4). The first two methods, A and B, used similar statistical models but differing assumptions about the relationship between HCA BSI and the other categories (CA and HA). The third method, C, used a different statistical approach and did not make any assumptions about the relationship between, CA, HCA and HA BSI. The purpose of the models was to develop an upper and lower estimate of the predicted proportion of CA, HCA and HA BSI. The approaches A, B and C are expanded below.

The first two approaches (A and B) used different assumptions about the relationship of patient characteristics between children with CA, HCA or HA BSI. There have historically been mixed reports of the characteristics of patients with HCA BSI compared to the characteristics of patients with CA or HA BSI.(9;171) Approach A focuses on the assumption that children with HCA BSI have different characteristics compared to children with CA or HA BSI. Approach B assumes that children with HCA BSI have similar characteristics to children with HA BSI.

Both approaches A and B rely on a series of logistic regression models to identify characteristics associated with HA or CA in different subgroups. From these models, predictions are made about others from the dataset and their likely affiliation. In all logistic models, first- and second-degree fractional polynomials were used to characterise changes in log odds of being in a given category (CA, HCA or HA) for the continuous variables (age and length of postnatal stay).(172) The first-degree (p) and second degree (q) power transformation ($x^{p,q}$) of a continuous variable comprised: -2, -1, -0.5, 0, 0.5, 1, 2. If the first-degree transformation was the best fit of the data (including a linear fit where $p = 1$), then then second-degree was not employed. In summary, there were 36 possible combinations.(173)

Using a best subset framework to identify the best fitting regression model from all the possible combinations of explanatory variables specified, a total of 1,023 models (all the possible model combinations using 10 explanatory variables listed in Table 4.1, Chapter 4) were run for children aged 1month < 1 year, and 551 models (all the possible model combinations using 9 explanatory variables listed in Table 4.1, Chapter 4) were run for children aged the 1-5 years.

All the models were evaluated using the Akaike Information Criteria (AIC) and Bayesian Information Criterion (BIC) to identify the most succinct and best fitting models, i.e. the model

with the lowest AIC score and the model with the lowest BIC score. The AIC measured the adequacy of the regression model by estimating the expected information loss considering the variables in the model compared to the true model, where the model with the lowest AIC value was the model with the lowest expected information loss and therefore preferred.(174;175) The BIC is less liberal than the AIC model selection and was more partial to selecting simpler models with the lowest BIC value, measuring the degree to which a model was the true data-generation model.(176;177) In each case, the BIC model was nested within the respective AIC model, and I used a likelihood ratio test to select the best fitting model of the two. Odds ratios (ORs) with a P-value<0.05 were considered statistically significant.

The third approach, C, did not use any prior assumptions but modelled the relationships between children's characteristics to group the children into three distinct groups, which would then be deemed to be CA, HCA and HA BSI via consideration of the numbers of *a priori* CA BSI and *defined* HA BSI in each group.

The method, hypothesised outcome and assumptions for approaches A, B and C are summarised in Table 5.5. Descriptions of approaches A, B and C are each detailed in the following section. All figures in the next methods sections are for illustration purposes only and do not reflect real data distributions.

Table 5.5. Summary of methodological steps and description of approaches A, B and C

Approach	Methodological steps	Description
A	Initial assumptions about BSI	***Defined* HA BSI** all BSI beyond threshold x (x=varying threshold between day 1-9 after hospital admission) ***A priori* CA BSI** isolation of *S. pneumoniae, N. meningitidis*, GAS or *Salmonella* spp.
	Model 1: Logistic regression	**Subset: all** **Outcome:** child with *defined* HA BSI vs not. Used to determine additional subgroup of HA from those not *a priori* CA with BSI 1 + days from admission.
	Model 2: Logistic regression	**Subset: Those not *defined* HA or from the additional HA subgroup given by model 1** **Outcome** child with *a priori* CA BSI vs not. Used to determine additional subgroup of CA (ie. with predicted probability > = 0.5) and the remainder assumed to be HCA
B	Initial assumptions about BSI	***Defined* HA BSI** all BSI beyond threshold x (x=varying threshold between day 1-9 after hospital admission) ***A priori* CA BSI** isolation of *S. pneumoniae, N. meningitidis*, GAS or *Salmonella* spp.
	Model 1: Logistic regression	**Subset: Those *defined* HA or *a priori* CA** **Outcome** children with *defined* HA vs *a priori* CA Used to determine additional subgroup likely CA (predicted probability <0.5) and give upper limit for HCA (predicted probability >0.5)
C	Initial assumptions about BSI	None
	Model 1: Multiple correspondence analysis (MCA)	Summary of relationships between BSI and child related variables condensed into 2 or more numerical summaries (factors)
	Model 2: Discriminate analysis using K-medians	Children classified into one of three groups using the created MCA numerical summaries
	Compare Model 2 results to assumptions about BSI from approach A and B	The group of children with the largest proportion of : • *defined* HA BSI will be considered as having HA BSI; *a priori* CA will be considered as having CA BSI • The remaining group will be considered having HCA BSI

5.3.1 Methods: approach A

Assumption: *children with HCA BSI do not have the same characteristics/risk factors as children with HA BSI or CA BSI*

Two logistic regression models were used to predict children with CA, HCA or HA BSI. The models were run for each of the nine HA BSI thresholds (HA1-HA9) to investigate different definitions of HA BSI (as described in section 5.2.1). The first model predicted children with likely HA BSI between day 1 and day x of hospital admission (proportion of children in the light grey curve, Figure 5.6), and the second model predicted children with CA or HCA BSI (proportion of children in the dark and light grey curves, Figure 5.6).

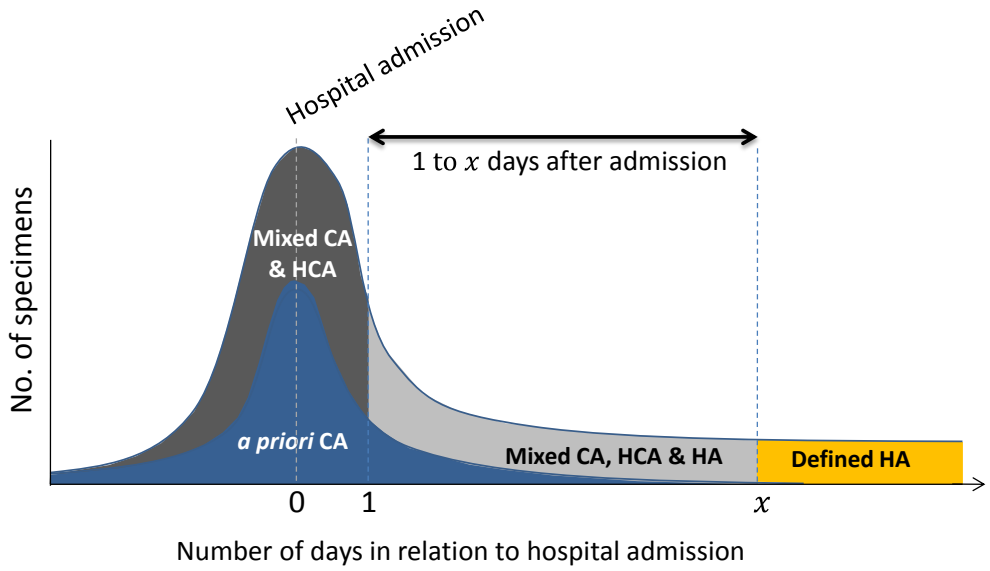


Figure 5.6. Expected timing of specimen date sample in relation to hospital admission in approach A: Mixed CA, HCA and HA BSI

5.3.1.1 Model 1: Predicting HA BSI cases

I performed the first logistic regression to model the odds of the clinical and demographic characteristics of children with *defined* HA BSI versus all other children before day x (including children with *a priori* CA BSI).

The probability of each child having HA BSI was then predicted using the obtained regression model. Children with a BSI 1 or more days after admission and with a predicted probability of 0.5 or more were classified as having predicted HA BSI (with the exception of children with *a priori* CA BSI who remained in their original classification). The logistic models yielded estimated probabilities of the BSI being HA (as opposed to not HA). An estimated probability of 0.5 inferred that, on the basis of the model, each option was equally likely. BSI with estimated

probabilities of 0.5 or greater were classified as HA on the basis that they were more likely to be HA than not (i.e. probability (HA) > probability (not HA). This was applied in both Approach A and Approach B. Children with a predicted probability of less than 0.5, who were not in the *defined* HA group, remained an unknown mix of CA and HCA BSI Figure 5.7.

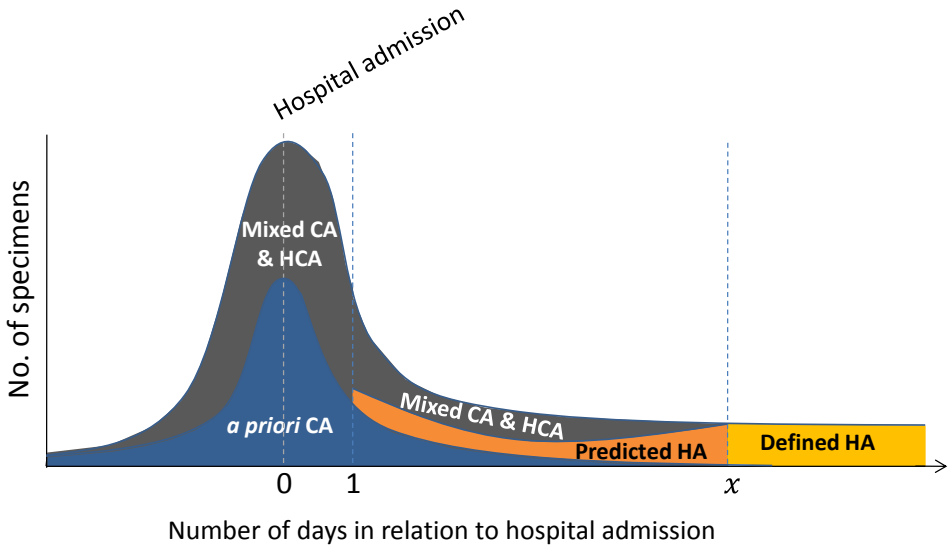


Figure 5.7. Expected timing of specimen date sample in relation to hospital admission in approach A: predicted HA BSI selection

5.3.1.2 Model 2: Predicting predicted CA and HCA BSI cases

I removed all children with predicted and *defined* HA BSI from the dataset before running the logistic regression analysis to model the odds of a child having *a priori* CA or not; i.e. modelling the log odds of *a priori* CA (in blue, Figure 5.7) as opposed to mixed CA and HCA BSI (in grey, Figure 5.7). Children with a predicted probability of 0.5 or more were classified as having CA BSI; those with a predicted probability less than 0.5 were classified as having HCA BSI (Figure 5.8). A child with *a priori* CA BSI was not re-classified as having HCA BSI if the predicted probability was less than 0.5.

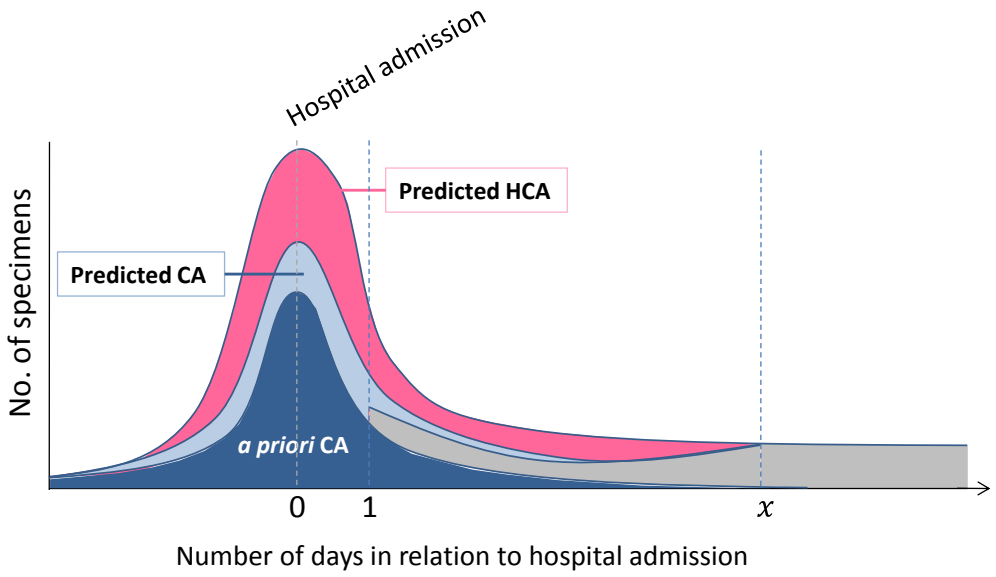


Figure 5.8. Expected timing of specimen date sample in relation to hospital admission in approach A: Differentiating between predicted CA BSI and HCA BSI selection

5.3.2 Methods: approach B

Assumption: *children with HCA BSI have similar characteristics/risk factors to children with HA BSI*

5.3.2.1 Model 1: Predicting CA, HCA or HA BSI cases

Logistic regression was used to model the odds of the clinical and demographic characteristics of children with *defined* HA BSI after day x versus children with *a priori* CA BSI (Figure 5.9).

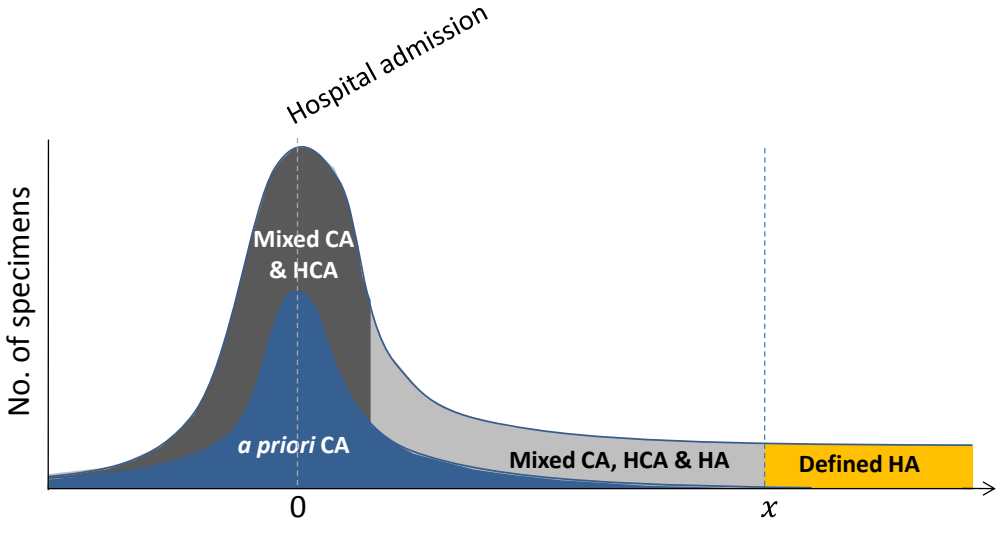


Figure 5.9. Expected timing of specimen sample date in relation to hospital admission in approach B: Using the *a priori* CA and *defined* HA BSI cases for the modelling

The final regression coefficients from the best fitting models for each age-group and HA-threshold were used to predict the probability of a child having CA, HCA or HA BSI for the remaining children in the sample. Children with a predicted probability less than 0.5 were classified as having CA BSI. Children with a 0.5 or more probability in the mixed group (grey section in Figure 5.9) were classified as having HCA BSI if their blood specimen was isolated before or on the day of hospital admission (day 0), or as a mix of HCA and HA BSI if the specimen was isolated between day 1 and day x after hospital admission. An illustrated example of these distinctive groups is depicted in Figure 5.10. This approach gave an upper estimate of those patients considered to have a HCA BSI (where the mixed HCA and HA BSI, in orange in the figure below, were considered to be HCA).

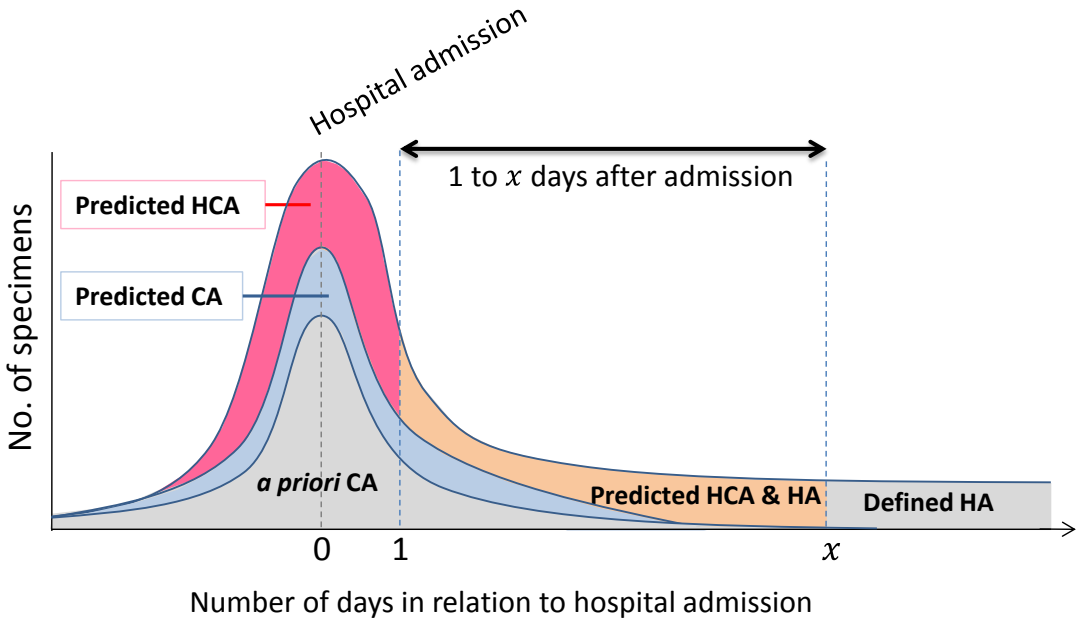


Figure 5.10. Expected timing of specimen sample date in relation to hospital admission in approach B: predicted CA, HCA and HA BSI selection

5.3.3 Methods: approach C

This unsupervised approach partitioned all the children with BSI into three distinct groups based on the relationships between the demographic and clinical risk factors within and between children, as well as the timing of the positive specimen before and after hospital admission. I hypothesised that these three groups would reflect a natural distinction between children with CA, HCA or HA BSI, resulting in an estimated proportion of HCA BSI in children that would be between the lower and upper estimates of HCA BSI identified in approaches A and B.

No prior assumptions or information on the invasive pathogen(s) were used during the statistical analysis. The process yielded three distinct subgroups and these were assigned as CA, HCA and HA by comparison with those previously determined as *a priori* CA and *defined* HA.

5.3.3.1 Step 1: Multiple Correspondence Analysis (MCA)

I used multiple correspondence analysis (MCA) to identify clusters of children according to the associations between their clinical and demographic characteristics. MCA is an exploratory method that creates one or more numeric variables (dimensions) that quantify and simplify the relationships between categorical variables by measuring the multi-dimensional dispersion between them.(178;179) The aim is that a subset of the numeric dimensions account for the majority of the inertia (i.e. variance). These numeric dimensions, known as factor scores, will be depicted in two (or more) dimensional graphs to illustrate the clustering of characteristics according to child similarities across clinical and demographic characteristics.(178;179) The dimensional graphs will be divided into four quadrants to describe the relationships between the variables.

MCA requires that explanatory variables are factors and hence the two continuous variables (age and re-admitted postnatal length of stay) were transformed into categorical variables as described in Table 5.6. The categorical grouping of the number of days between admission and the BSI incorporated into Approach C are also described in Table 5.6. The number of dimensions was selected to represent the majority of the variation between BSI episodes.

Table 5.6 Additional descriptive breakdown of children with the additional categorical variables for approach C

Variable		1m<1yr	%	1-5yrs	%	Total
Age						
	1<2 months; 1<2 years	999	(22.6%)	2,279	(36.9%)	-
	2<4 months; 2<3 years	1,064	(24.1%)	1,286	(20.8%)	-
	4<6 months; 3<4 years	727	(16.5%)	1,110	(18.0%)	-
	6<9 months; 4<5 years	811	(18.4%)	881	(14.3%)	-
	9<12 months; 5<6 years	814	(18.4%)	616	(10.0%)	-
Re-admitted postnatal length of stay in days						
	Not readmitted	749	(17.0%)	-	-	**749**
	0.5-1 day	1,436	(32.5%)	-	-	**1,436**
	2-6 days	1,324	(30.0%)	-	-	**1,324**
	7-30 days	505	(11.4%)	-	-	**505**
	31-90 days	302	(6.8%)	-	-	**302**
	≥91 days	99	(2.2%)	-	-	**99**
Use the number of days between admission and BSI categories instead of *a priori* CA or HA						
	-5 to -1 day	213	(4.8%)	282	(4.6%)	**282**
	0 days (admission date)	2,673	(60.5%)	4,138	(66.9%)	**4,138**
	1 day	570	(15.4%)	981	(15.9%)	**981**
	2 days	116	(2.6%)	162	(2.6%)	**162**
	3 days	53	(1.2%)	57	(0.9%)	**57**
	4 days	42	(1.0%)	54	(0.9%)	**54**
	5 days	33	(0.7%)	37	(0.6%)	**37**
	6 days	23	(0.5%)	31	(0.5%)	**31**
	7 days	28	(0.6%)	28	(0.5%)	**28**
	8 days	22	(0.5%)	24	(0.4%)	**24**
	9+ days	532	(12.1%)	378	(6.1%)	**378**

5.3.3.2 Step 2: K-medians analysis

Step 1 produces a series of factor scores for each child BSI to be grouped according to similarity. Cluster analysis is a process whereby the numeric distances between two children with BSIs on a set of variables are summarised to give a measure of similarity. Over all children in the dataset, these similarity measures are compared to find groups that behave more alike than others. Using this process, subgroups of children who are relatively similar with respect to the factor scores (sets of numeric measures summarising each BSI) are identified. The algorithm that is used to do this can be set so that a specified number of subgroups are obtained. I used a K-medians cluster analysis to separate the BSI according to the factor scores obtained in step 1 into three distinct subgroups as illustrated in Figure 5.11.(180)

Group identified from the model			
1	2	3	
n_1	n_2	n_3	n = children

Figure 5.11. K-medians three-group output

In order to interpret the allocation of children to one of the three K-medians groups, the three groups were compared to the number of children categorised as having *a priori* CA BSI, *defined* HA BSI (defined in section 5.2.1 and 5.2.2) or having an undefined mixed BSI (CA, HCA or HA), as illustrated in Figure 5.12.

	Group identified from the model				
	1	2	3	Total	
defined HA BSI	n_{1a}	n_{2a}	n_{3a}	n_{HA}	n = children
a priori CA BSI	n_{1b}	n_{2b}	n_{3b}	n_{CA}	
remaining mixed BSI	n_{1c}	n_{2c}	n_{3c}	n_{mixed}	

Figure 5.12. Distribution of children in the three K-medians output groups compared to *defined* HA BSI, *a priori* CA BSI and the remaining mixed BSI

Assuming that the classifications of children with *a priori* CA or *defined* HA BSI are true and complete, Figure 5.13 illustrates the optimum outcome for the K-medians model and the model is aiming to get as close to this outcome as possible. All children with *defined* HA BSI are classified into one group (e.g. group n_{3a}), all children with *a priori* CA BSI are classified into another group (e.g. group n_{1b}), and the remainder of children with mixed BSI being distributed between all three groups (e.g. n_{1c}-n_{3c}). It would be assumed that the remaining group (e.g. group 2) are associated with children with HCA BSI. In interpreting the results, I determined the CA, HCA and HA BSI from the three groups that most closely resembled the results in Figure 5.13.

	Group identified from the model				
	1	2	3	Total	
defined HA BSI	0	0	n_{3a}	n_{HA}	n = children
a priori CA BSI	n_{1b}	0	0	n_{CA}	
remaining mixed BSI	n_{1c}	n_{2c}	n_{3c}	n_{mixed}	

Figure 5.13. Example of optimum outcome when comparing the three K-medians groups

For each fitted K-medians model, the generation of the group names (1, 2 and 3) varied. For ease of comparing the results, after the groups had been created by the models, I ensured that the largest proportion of children with *defined* HA BSI was named as group 3 and *a priori* CA BSI was named as group 1 in both age-groups.

5.3.4 Sensitivity analyses

Three sensitivity analyses were undertaken for approaches A, B and C and are described below (the results will be presented in Chapter 6).

1. Hospital trust where the patient was being treated at the time of the BSI was included in the final model to account for any clustering
    - approach A and B: hospital trust was added into the final logistic regression models as a random effect to adjust for any clustering by hospital
    - approach C: hospital trust was included as an additional fixed effect variable and accounted for in the MCA and therefore subsequently in the K-medians models
2. *A priori* CA cases where the BSI was isolated ≥2 days after hospital admission were excluded from the analysis. This was performed to investigate the impact that these few observations may have had on the overall predictions of the proportion of CA, HCA and HA BSI in children.
3. CoNS, *Micrococcus* spp. and diphtheroids were excluded from the analysis. These bacteria are frequently considered to be contaminants due to skin contamination when a blood sample is taken.(181) However, clinically they are also a common cause of BSI associated with indwelling devices. Although all the isolates that are reported by laboratories to LabBase2 are defined as clinically significant, this sensitivity test will determine if there are any differences in the prediction of CA, HCA or HA BSI.

5.4 Results of approaches A, B and C

5.4.1 Results: approach A

The unadjusted odds ratios for children with defined HA BSI (as opposed to non-defined HA BSI) by the demographic and clinical characteristics for both age-groups are shown in Table 5.7. These values are valid for both Approach A and B.

Table 5.7 Unadjusted OR results for explanatory variables used in models A and B with defined HA BSI (2 days) as the outcome in children aged 1 month < 1 year

Characteristic	1 month < 1 year old					
	Total no. of obs.	No. of HA2	Unadjusted OR	P>z	95% CI	
Age in days	4,415	803	0.99	0.008	0.99	0.99
Sex						
boy	2,615	451	1.00			
girl	1,800	352	1.17	0.051	1.00	1.36
Ethnicity						
White	2,997	452	1.00			
Mixed	141	26	1.27	0.279	0.82	1.97
Indian subcontinent	408	92	1.64	<0.001	1.27	2.11
Any other Asian	126	31	1.84	0.004	1.21	2.79
Black	246	53	1.55	0.008	1.12	2.13
other ethnic group	161	50	2.54	<0.001	1.79	3.59
NK	336	99	2.35	<0.001	1.82	3.04
Premature/low birthweight						
no	3,540	523	1.00			
yes	875	280	2.71	<0.001	2.29	3.22
Chronic conditions						
no chronic conditions	2,308	124	1.00			
1 chronic condition	1,119	160	2.94	<0.001	2.30	3.76
2 chronic conditions	437	177	11.99	<0.001	9.22	15.60
>2 chronic conditions	551	342	28.82	<0.001	22.45	37.00
Congenital conditions						
No congenital conditions	3,447	318	1.00			
1 or more congenital condition(s)	968	485	9.88	<0.001	8.33	11.72
Days between last discharge and current admission						
no previous discharge in the year before BSI	569	151	1.00			
91 days<1 yr	1,381	92	0.20	<0.001	0.15	0.26
31-90 days	1,221	125	0.32	<0.001	0.24	0.41
7-30 days	692	181	0.98	0.878	0.76	1.26
1-6 days	552	254	2.36	<0.001	1.84	3.03
Days since last invasive procedure						
no invasive procedure in the year before BSI	2,958	178	1.00			
91 days<1 yr	296	44	2.73	<0.001	1.91	3.89
31-90 days	391	121	7.00	<0.001	5.38	9.10
7-30 days	401	250	25.86	<0.001	20.09	33.29
1-6 days	369	210	20.63	<0.001	15.97	26.64
No. of days since last indwelling device insertion						
no device in the year before BSI	3,840	459	1.00			
last insertion 91 days<1yr	150	54	4.14	<0.001	2.93	5.86
last insertion 31-90 days	192	101	8.18	<0.001	6.06	11.03
last insertion 7-30 days	160	132	34.73	<0.001	22.83	52.81
last insertion 1-6 days	73	57	26.24	<0.001	14.94	46.08
Postnatal stay in days						
no postnatal stay	749	174	1.00			
postnatal stay	3,666	629	1.02	<0.001	1.01	1.02

Table 5.8 Unadjusted OR results for explanatory variables used in models A and B with defined HA BSI (2 days) as the outcome in children aged 1-5 years

Characteristic	1-5 years old					
	Total no. of obs.	No. of HA2	Unadjusted OR	P>z	95% CI	
Age in days	6,172	688	0.99	0.68	0.99	1.00
Sex						
boy	3,483	393	1.00			
girl	2,689	295	0.97	0.699	0.83	1.14
Ethnicity						
White	4,163	431	1.00			
Mixed	198	14	0.66	0.139	0.38	1.14
Indian subcontinent	552	71	1.28	0.073	0.98	1.67
Any other Asian	174	21	1.19	0.469	0.75	1.90
Black	397	41	0.99	0.987	0.71	1.40
other ethnic group	202	33	1.69	0.008	1.15	2.49
NK	486	77	1.63	<0.001	1.25	2.12
Premature/low birthweight						
no	5,618	593	1.00			
yes	554	95	1.75	<0.001	1.38	2.22
Chronic conditions						
no chronic conditions	2,779	65	1.00			
1 chronic condition	1,698	155	4.19	<0.001	3.12	5.64
2 chronic conditions	776	131	8.48	<0.001	6.22	11.56
>2 chronic conditions	919	337	24.18	<0.001	18.27	32.00
Congenital conditions						
No congenital conditions	5,499	481	1.00			
1 or more congenital condition(s)	673	207	4.63	<0.001	3.84	5.60
Days between last discharge and current admission						
no previous discharge in the year before BSI	3,532	230	1.00			
91 days<1 yr	733	74	1.61	0.001	1.22	2.12
31-90 days	460	64	2.32	<0.001	1.73	3.12
7-30 days	615	115	3.30	<0.001	2.59	4.21
1-6 days	832	205	4.69	<0.001	3.82	5.77
Days since last invasive procedure						
no invasive procedure in the year before BSI	4,365	146	1.00			
91 days<1 yr	403	39	3.10	<0.001	2.14	4.48
31-90 days	364	72	7.13	<0.001	5.25	9.68
7-30 days	541	220	19.80	<0.001	15.61	25.13
1-6 days	499	211	21.17	<0.001	16.61	26.98
No. of days since last indwelling device insertion						
no device in the year before BSI	5,207	336	1.00			
last insertion 91 days<1yr	345	78	4.24	<0.001	3.22	5.58
last insertion 31-90 days	262	89	7.46	<0.001	5.64	9.85
last insertion 7-30 days	246	130	16.25	<0.001	12.36	21.36
last insertion 1-6 days	112	55	13.99	<0.001	9.50	20.59
Postnatal stay in days						
no postnatal stay	-	-	-	-	-	-
postnatal stay	-	-	-	-	-	-

Graphs depicting the results of either the determined best fit AIC or BIC models after the likelihood ratio test from 1,023 and 551 logistic regression models by age-group for each of the nine scenarios are in Appendix 5. The results (adjusted ORs) of the best fitting logistic regression model outputs by age-group and *defined* HA-threshold are in Appendix 5.

5.4.1.1 Model 1: Predicting HA BSI cases

Increasing age in days (no transformation of age) was associated with a statistically significant, although very small decrease in the odds of having HA BSI (OR=0.99; 95% CI: 0.99-0.99; P<0.001) as opposed to not in children aged 1 month < 1 year. The models indicated some association that girls were more likely than boys aged 1 month < 1 year to have HA BSI in the HA2 and HA6 models (OR: 1.26; 95% CI: 1.02-1.55, and OR: 1.31; 95% CI: 1.02-1.69 respectively), this relationship was similar but did not attain statistical significance for HA1 and HA3-HA5 models. Gender was not included in the final models for children aged 1-5 years suggesting that there was no significant gender difference in the distribution of BSI by *defined* HA BSI or not.

Children aged 1 month < 1 year with certain clinical characteristics had statistically significant higher odds of having a *defined* HA BSI compared to children without these clinical characteristics in the year prior to the BSI for HA2-HA9 models in both age-group: premature/low birth-weight (HA2 OR: 1.83, 95% CI: 1.38-2.44; HA9 OR: 1.92, 95% CI: 1.38-2.67), one or more chronic conditions with the largest odds for children with more than two chronic conditions (HA2 OR: 5.87, 95% CI: 4.19-8.22; HA9 OR:9.84, 95% CI: 6.32-15.31), indwelling device inserted 1-90 days before the BSI where 7-30 days had the largest OR (HA2 OR: 4.87, 95% CI: 2.94-8.06; HA9 OR: 8.15, 95% CI:4.94-13.44), previous invasive procedure in the past 1-90 days before the BSI where both 1-6 days (HA2 OR: 7.57, 95% CI: 5.42-10.58) and 7-30 days (HA9 OR: 9.40, 95% CI: 6.09-14.51) before BSI had the largest OR depending on the threshold (Appendix 5). The only variable that had significantly lower odds (P<0.001) for children with *defined* HA BSI compared to the rest of the cohort was for a previous discharge in the past year (OR ranging from HA2: 0.28 [95% CI: 0.16-0.47], to HA9: 0.05 [95% CI: 0.03-0.11]). The model using the HA1 threshold returned the least number of statistically significant differences in clinical indicators between children with *defined* HA versus not compared to the HA2-HA9 models.

Children aged 1-5 years with *defined* HA BSI between the threshold HA2 and HA9 had statistically significant higher odds of having one or more chronic conditions with the highest odds for children with more than two chronic conditions (HA2 OR: 8.98, 95% CI: 6.25-12.89;

HA9 OR: 60.07, 95% CI: 26.40-136.67), a congenital condition (HA2 OR: 2.24, 95% CI: 1.76-2.86; HA9 OR: 2.36, 95% CI: 1.74-3.20), an indwelling device previously inserted in the past 1-90 days with the highest odds for 7-30 days (HA2: 2.55, 05% CI: 1.79-3.64; HA9: 5.37, 95% CI: 3.49-8.28) , or an invasive procedure in the past year with the highest odds for 7-30 days (HA2 OR: 15.57, 95% CI: 10.77-22.51; HA9 OR: 30.87, 95% CI: 18.39-51.84). As with the younger age-group, children with *defined* HA BSI had significantly lower odds of having a previous discharge in the past year.

The ranges of predicted probabilities of children having HA BSI in relation to the timing of hospital admission by age-group are depicted in Figure 5.14 and Figure 5.15 (jittered data points).

Although the predicted value for children with *defined* HA BSI did not change their HA category, both age-groups contained children with *defined* HA BSI that had a predicted range of 0.06-0.99, (1 month < 1 year median: HA1= 0.50, HA9=0.56; 1-5 years median: HA1=0.27, HA9=0.33). As expected, the majority of children with *a priori* CA BSI had a low predicted probability of having HA BSI (aged 1 month < 1 year mean: HA1=0.15, HA9=0.01; 1-5 years mean: HA1=0.14, HA9=0.01). The predicted probability of all other children having a HA BSI ranged between 0-0.98 (median: HA1=0.17, HA9=0.01) for children aged 1 month < 1 year and 0-0.85 (mean: HA1=0.15, HA9=0.01) for children aged 1-5 years.

Table 5.9 reports the number of children who were predicted to have HA BSI in addition to children with *defined* HA BSI.

Table 5.9. *Defined* HA BSI and predicted HA BSI for the range of cut-off days (HA1-HA9)

HA-threshold	1 month < 1 year old		1-5 years old	
	***defined* HA BSI**	**predicted HA BSI**	***defined* HA BSI**	**predicted HA BSI**
day 1	1,325	0	1,379	0
day 2	803	36	688	24
day 3	708	46	569	27
day 4	661	50	523	30
day 5	625	54	478	28
day 6	598	58	449	29
day 7	578	56	419	22
day 8	552	55	394	22
day 9	530	58	373	23

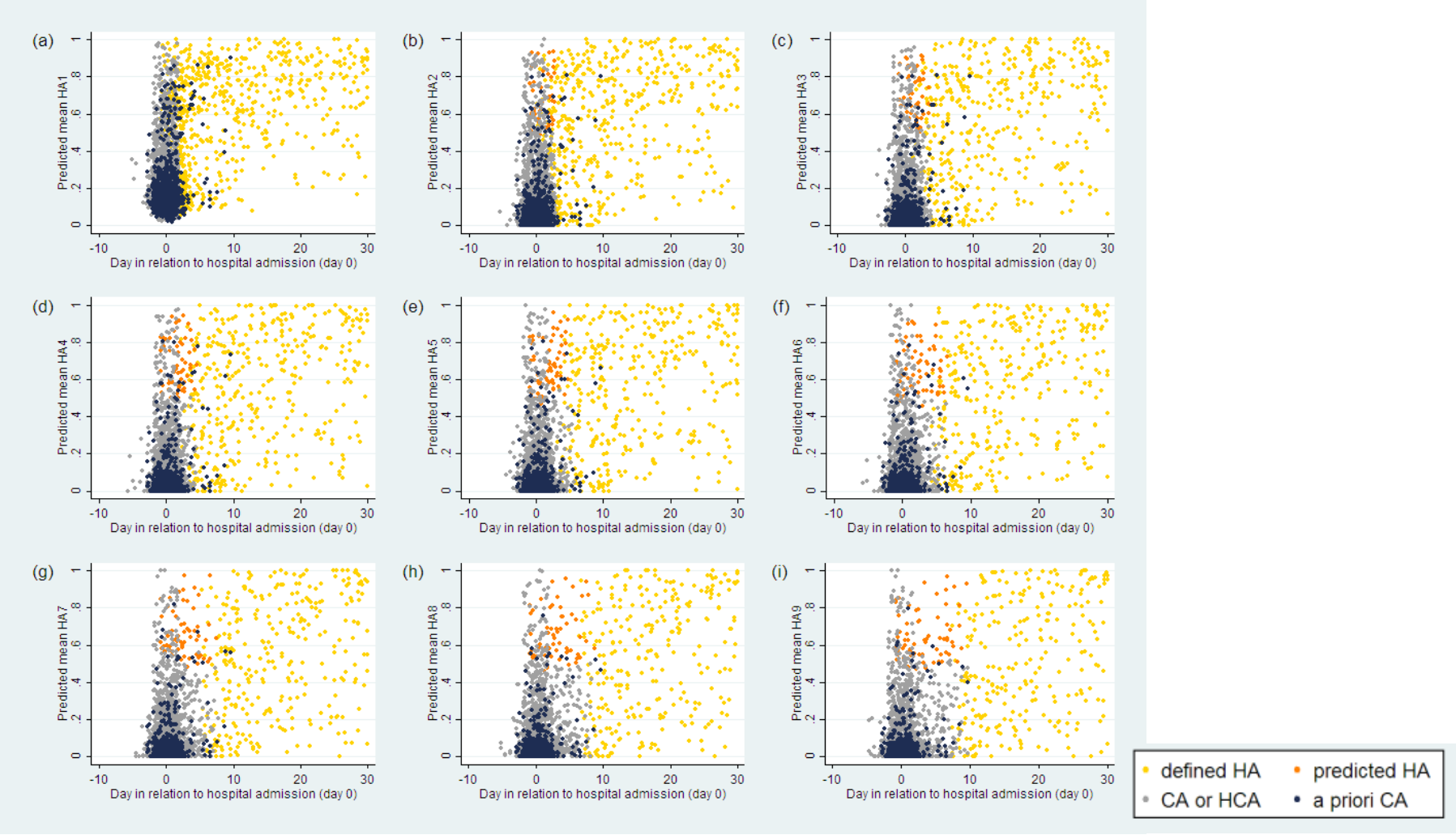


Figure 5.14a-i. Distribution of the predicted HA cases (≥0.5) in relation to all the other cases of BSI between -5 and +30 of hospital admission in children aged 1 month < 1 year (jittered data points)

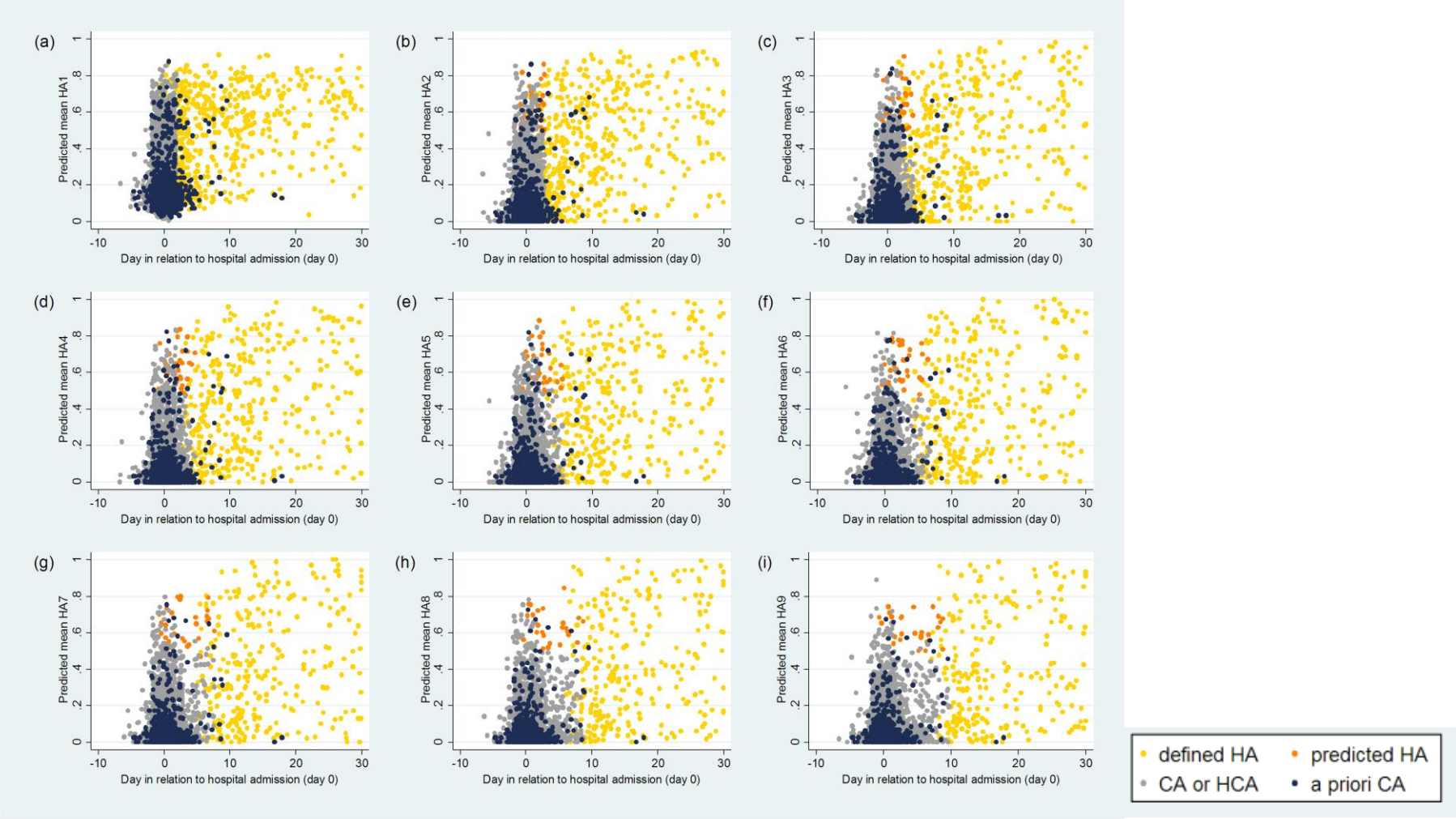


Figure 5.15a-i. Distribution of the predicted HA cases (≥0.5) in relation to all the other cases of BSI between -5 and +30 of hospital admission in children aged 1-5 years (jittered data points)

5.4.1.2 Model 2: Predicting CA and HCA BSI cases

Comparing children with *a priori* CA to the remaining children with a mixture of CA and HCA BSI, children from the following ethnic groups had a lower odds of having *a priori* CA BSI compared to the children from the white ethnicity group for all thresholds in both age-groups (HA1-HA9): Indian subcontinent (e.g. 1 month < 1 year HA2 OR: 0.60, 95% CI: 0.44-0.84), any other Asian (e.g. 1 month < 1 year HA2 OR: 0.23, 95% CI: 0.10-0.52) and Black ethnicities (e.g. 1 month < 1 year HA2 OR: 0.36, 95% CI: 0.23-0.58). For children aged 1-5 years, sex was not included in any of the models.

In the younger age-group, children who were born prematurely or with low birthweight (single variable) were more likely to have *a priori* CA BSI (HA1: OR: 1.56 [95% CI: 1.12-2.18]; HA9 OR: 1.57 [95% CI: 1.14-2.15]) compared to the rest of the population in the HA1-HA9 scenarios. This variable was not included in the final models for children aged 1-5 years.

Children with congenital conditions, in both age-groups, were less likely to have *a priori* CA BSI compared to the remaining population of children (e.g. 1 month < 1 year HA2 OR: 0.66, 95% CI: 0.48-0.92; 1-5 years HA2 OR: 0.60, 95% CI: 0.46-0.80). However, children with one or more chronic conditions in both age-groups had higher odds of having *a priori* CA BSI than other children (e.g. 1 month < 1 year HA2 OR for 1 chronic condition: 2.05, 95% CI: 1.70-2.47; 1-5 years HA2 OR for 1 chronic condition: 1.53, 95%CI: 1.33-1.75).

For the HA2-HA9 models, children in both age-groups with an invasive procedure in the last 7-30 days had statistically significant lower odds of having *a priori* CA BSI (lowest in the younger age-group, for example 1 month < 1 year HA2 OR: 0.34, 95% CI: 0.18-0.66; 1-5 years HA2 OR: 0.59, 95% CI: 0.37-0.94); the odds of having an *a priori* CA BSI were not statistically different for children with or without an invasive procedure in the previous 1-6 days.

Children aged 1 month < 1 year with an indwelling device in the year prior to the BSI had very low odds of having *a priori* CA BSI, except for when the indwelling device was inserted 1-6 days prior to the BSI, where the odds of having *a priori* CA BSI ranged from two- to eight-fold higher (HA2 OR: 8.16, 95% CI: 1.70-39.21; HA9 OR: 2.49, 95% CI: 1.04-5.94). The higher odds in the 1-6 days prior are due to 12 children (eight *N. meningitidis* BSI, four *S. pneumoniae* BSI) with *a priori* CA BSI with an indwelling device during this time period. In children aged 1-5 years, children with an indwelling device in the last 7 days < 1 year had significantly reduced odds of having *a priori* CA BSI compared to other children (e.g. HA2 7-30 days: 0.20 95% CI: 0.08-0.48; 31-<90 days OR: 0.24, 05% CI: 0.12-0.47; 91 days < 1 year OR: 0.46, 05% CI: 0.29-0.71).

The range of predicted values (using the coefficients from the logistic regression models) in relation to the timing of hospital admission by age-group are depicted in Figure 5.16 and Figure 5.17 (jittered data points). The children with BSI are colour-coded into the *a priori* CA BSI (dark blue), predicted CA BSI (light blue) or predicted HCA BSI (pink). The predicted probability of children with BSI that were predicted as CA (≥0.5) ranged from 0.50-0.93 (n=118) for HA1 and 0.50-0.76 (n=62) for HA9 in children aged 1 month < 1 year, and 0.50-0.69 (n=237) for HA1 and 0.50-0.62 (n=77) for HA9 in children aged 1-5 years.

Table 5.10 and Table 5.11 report the number and proportion of children with predicted CA and predicted HCA compared to the *a priori* CA BSI, *defined* HA and predicted HA BSI by age-group. The proportion of predicted CA BSI consistently accounted for 21-22% of all BSI in children 1m<1y; this proportion was nearly 10% higher in the older age-group, from 29-32%. The proportion of predicted HCA BSI was similar in both age-groups accounting for nearly two thirds of all BSI. The proportion of predicted HA BSI was highest in the younger age-group (13-19% between HA2-HA9; 30% on HA1) compared to the 1-5 year olds (7-12%). The largest difference between the nine different scenarios was between HA1 and HA2.

Table 5.10. Approach A: Predicted proportions of CA BSI, HCA BSI and HA BSI in children aged 1 month < 1 year

HA threshold	Total	CA BSI			HCA BSI		HA BSI		
		a priori	predicted	%	predicted	%	*defined*	predicted	%
HA1	4,415	848	124	22%	2,118	48%	1,325	0	30%
HA2	4,415	848	93	21%	2,635	60%	803	36	19%
HA3	4,415	848	83	21%	2,730	62%	708	46	17%
HA4	4,415	848	79	21%	2,777	63%	661	50	16%
HA5	4,415	848	75	21%	2,813	64%	625	54	15%
HA6	4,415	848	82	21%	2,829	64%	598	58	15%
HA7	4,415	848	84	21%	2,849	65%	578	56	14%
HA8	4,415	848	85	21%	2,875	65%	552	55	14%
HA9	4,415	848	75	21%	2,904	66%	530	58	13%

Table 5.11. Approach A: Predicted proportions of CA BSI, HCA BSI and HA BSI in children aged 1-5 years

HA threshold	Total	CA BSI			HCA BSI		HA BSI		
		a priori	predicted	%	predicted	%	*defined*	predicted	%
HA1	6,172	1,711	253	32%	2,829	46%	1,379	0	22%
HA2	6,172	1,711	128	30%	3,621	59%	688	24	12%
HA3	6,172	1,711	103	29%	3,762	61%	569	27	10%
HA4	6,172	1,711	99	29%	3,809	62%	523	30	9%
HA5	6,172	1,711	98	29%	3,857	63%	478	28	8%
HA6	6,172	1,711	89	29%	3,894	63%	449	29	8%
HA7	6,172	1,711	88	29%	3,932	64%	419	22	7%
HA8	6,172	1,711	87	29%	3,958	64%	394	22	7%
HA9	6,172	1,711	86	29%	3,979	64%	373	23	6%

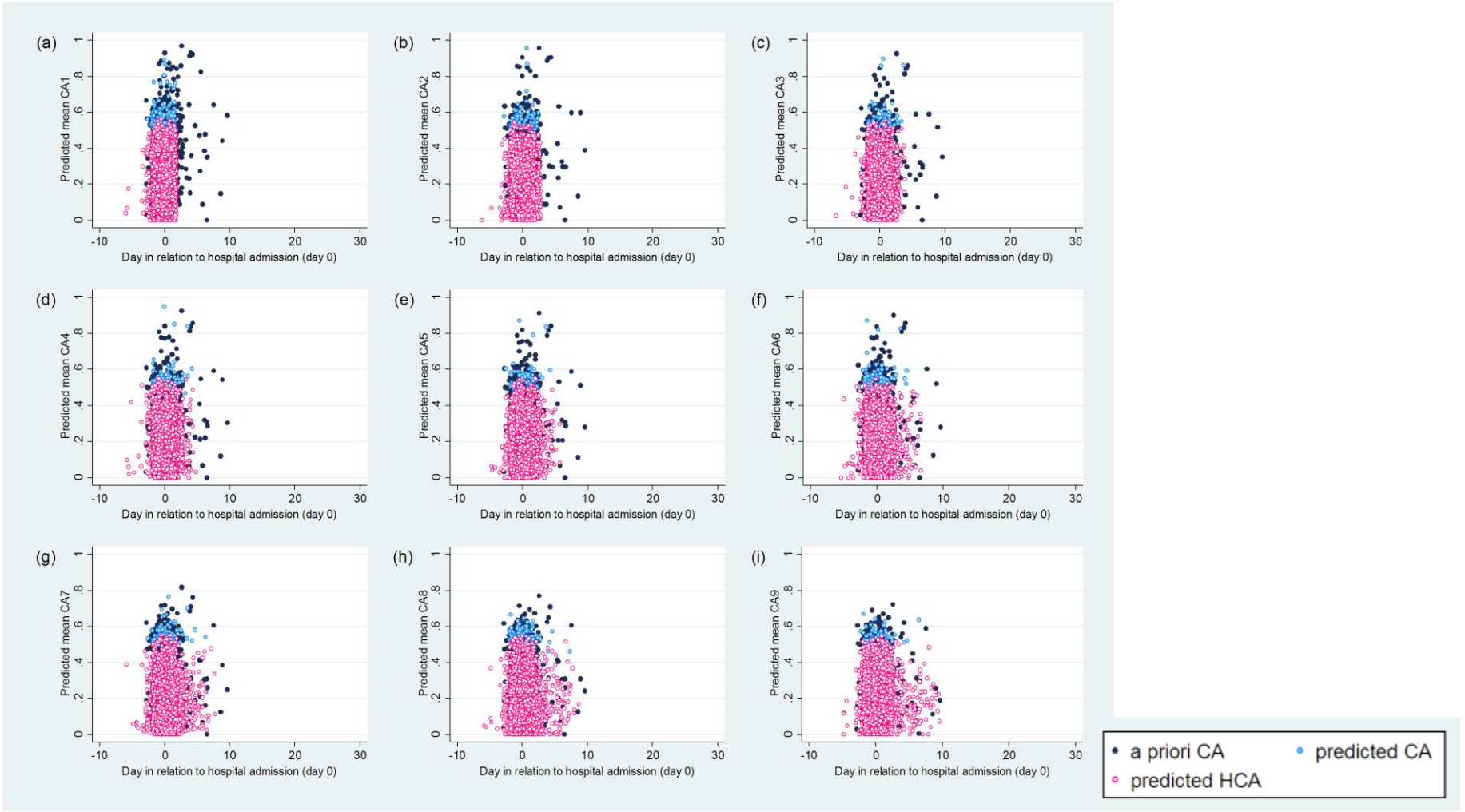


Figure 5.16. Distribution of predicted CA cases in relation to predicted HCA and *a priori* CA cases between -5 and +30 of hospital admission in children aged 1 month < 1 year (jittered data points)

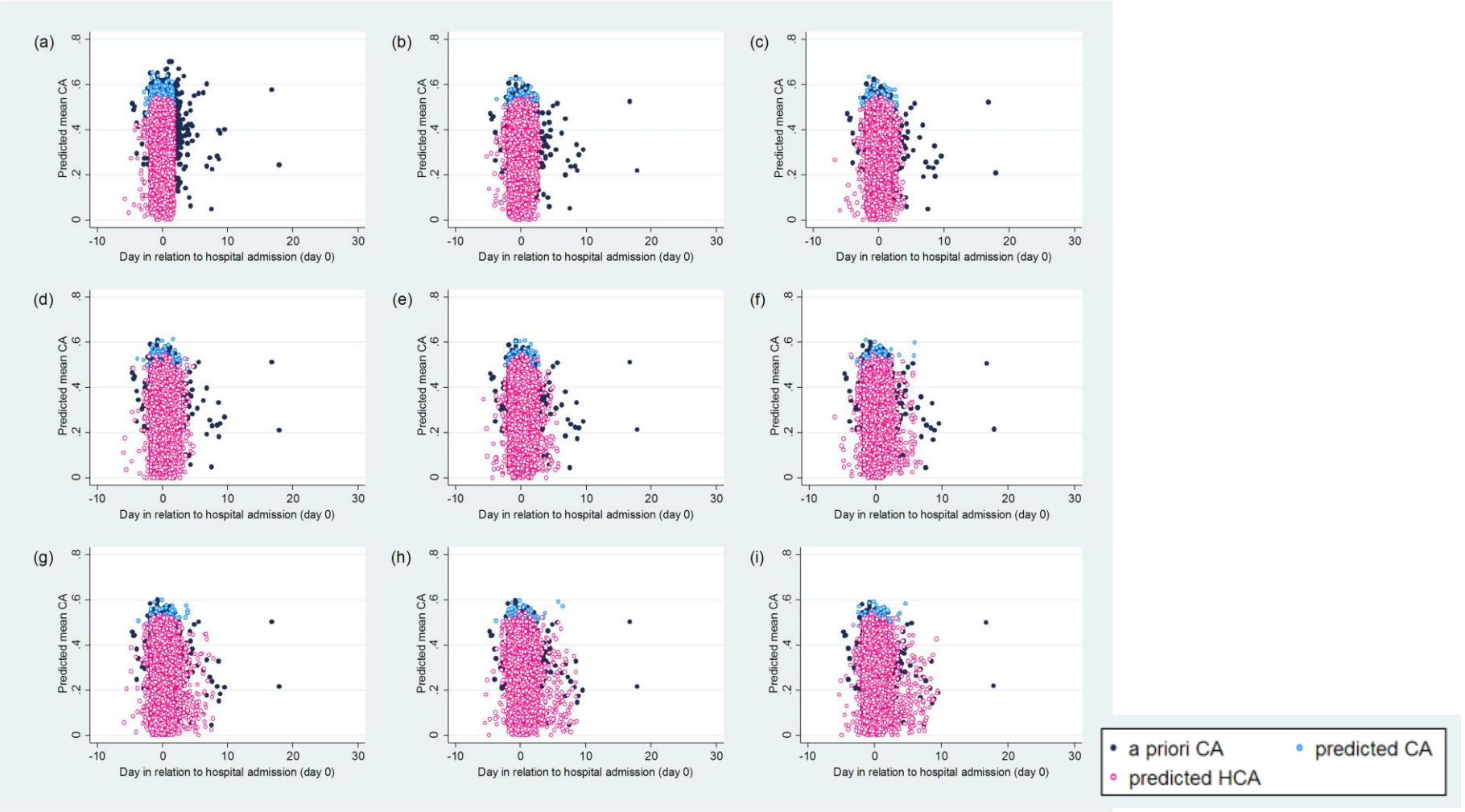


Figure 5.17. Distribution of predicted CA cases in relation to predicted HCA and *a priori* CA cases between -5 and +30 of hospital admission in children aged 1-5 years (jittered data points)

5.4.1.3 Summary of approach A results

Just under a half (HA1) to two-thirds (HA2-HA9) of children in both age-groups with BSI were predicted to have HCA BSI, based on the assumption that children with HCA BSI did not have similar characteristics to children with CA or HA BSI.

There was a clear separation between children with *defined* HA BSI versus not in Model 1, where children in hospital had more severe clinical characteristics than children with CA or HCA BSI. A higher proportion of HA BSI was observed in the children aged 1 month < 1 year. In contrast, there was no clear separation between children with CA and HCA BSI in Model 2 regardless of the HA threshold.

The results of the logistic regression to determine predicted HA BSI supports the hypothesis that the children with HA BSI will have the most clinical risk factors such as chronic conditions, exposure to invasive procedures or indwelling devices compared to other BSI. A child's previous discharge from hospital was the only risk factor that had lower odds than all other BSI, which is most likely explained because HA BSI is due to an infection acquired in the current admission rather than any previous hospital admissions, which is more likely linked to HCA BSI.

It was surprising to observe that children with one or more chronic conditions had statistically significant higher odds of having *a priori* CA BSI compared to children with mixed CA or HCA BSI, particularly because chronic conditions are often more associated with patients who require frequent healthcare contact in the literature. This does support the suggestion that a significant number of children with chronic conditions in the community are at risk of both CA and HCA BSI, whereas children who stay in hospital are at less risk from CA BSI due to restricted contact with the community environment.

Age was an important factor for differentiating between *defined* HA BSI or not. As children got older, the odds of them having a HA BSI decreased as opposed to children with other BSIs. Patients born with complications or who develop them very early on will require more acute hospital treatment at a very young age to treat and manage the medical condition(s) before being discharged, whereas with older children, many more medical conditions can be managed in the community or OP setting.

5.4.2 Results: approach B

5.4.2.1 Predicting CA, HCA or HA BSI cases

All the AIC and BIC values for the model combinations and the Adjusted ORs for the selected final models (HA1-HA9) for each age-group are shown in Appendix 5.

All the models for both age-groups found that children with two or more chronic conditions had the highest statistically significant odds of having *defined* HA BSI (Children aged 1 month < 1 year OR HA1: 2.05, 95% CI: 1.30-3.21; HA9 OR: 18.11, 95% CI: 8.38-39.14. Children aged 1-5 years OR HA1: 1.47, 95% CI: 1.07-2.03; HA9 OR: 64.68, 95% CI: 24.64-169.77). Of the other variables, the models found that children with an invasive procedure in the past 30 days (children aged 1 month < 1 year) or the past 90 days (children aged 1-5 years), had higher odds of having a HA BSI. Children with an indwelling device in the preceding 7-90 days (P<0.01; children aged 1 month < 1 year), or 1-90 days (P<0.01; children aged 1-5 years) had higher odds of having HA BSI, with the highest odds occurring in children with an indwelling device in the 7-30 days prior to the BSI (1m<1y OR: 15.57-56.48; 1-5y OR: 7.82-75.06).

Previous discharge was not included in the model with the lowest AIC or BIC in the HA1-HA3 models in the younger age-group, but was included and found to be a statistically significant indicator for a reduced odds of HA BSI from HA4-HA9 where there was a previous discharge between 7 days and up to a year previously (P<0.01). Similarly, children aged 1-5 years with a previous discharge between 7-90 days prior to the BSI for the HA5-HA9 thresholds had significantly lower odds of having *defined* HA BSI compared to all other children with BSI. The only difference was the HA1 and HA2 threshold in the older age-group, where children with a previous discharge in the 7-30 days for HA1 (OR=2.07, 95% CI: 1.43-3.00; HA2 OR=2.78), and four-fold for HA2 (OR=4.19, 95% CI: 2.95-5.94) had double the odds of having HA BSI.

In children aged 1 month < 1 year, the variable age in the HA1-HA3 models was transformed by a power of (-0.5), inverting the age values so that the youngest age becomes the highest value, therefore the odds (OR = 6.19-9.36) are for $age^{(-0.5)}$, so that increasing age was associated with lower odds of having a HA BSI. For HA4-HA9, the variable age was not transformed, unlike HA1-HA3, and therefore as age increased, the odds of the BSI being HA decreased significantly (OR=0.99, P<0.001).

Information on prematurity or low birth-weight were only included in models HA4-HA9 for children aged 1 month < 1 year, where the odds of being premature or born with low-birth-weight and having *defined* HA BSI as opposed to *a priori* CA BSI was more than double (2.09-

3.11; P<0.01). Sex was either excluded or contributed no statistically significant differences in either age-group.

The range of predicted probabilities (using the coefficients from the logistic regression models) in relation to the timing of hospital admission by age-group are depicted in Figure 5.18 and Figure 5.19 (jittered data points). There was a clear clustering of *defined* HA BSI, particularly in the top fifth (0.8-1.0) of predicted probability in all of the scenarios and both age-groups: 709/1,325 (53%; HA1) and 419/530 (79%; HA9) for children age 1 month < 1 year; 506/1,379 (37%; HA1) and 227/373 (61%) for children aged 1-5 years.

The majority of children with *a priori* CA BSI (60% of children aged 1 month < 1 year; 84% of children aged 1-5 years) were predicted to have a low probability of having HA BSI between 0 and 0.4 (as would be expected). The concentration of children with *a priori* CA BSI and a corresponding low prediction of being HA-BSI (between 0 and 0.2) increased as the HA-threshold moved from HA1 to HA9 (89% of children aged 1 month < 1 year; 92% of children aged 1-5 years). Only a small number of children with *a priori* CA BSI had very similar predicted values to children with *defined* HA BSI.

Table 5.12 and Table 5.13 report the numbers of children with predicted CA, HCA and HA compared to the *a priori* CA BSI and total HA BSI by age-group. The proportion of total CA BSI (*a priori* and predicted) was the dominant category in both age-groups compared to HCA or HA BSI. The proportion of predicted HCA BSI was similar in both age-groups accounting for between 11-13% of BSI (with the exception of HA1 in the younger age-group) across the nine different scenarios. The proportion of HA BSI was highest in the younger age-group (12-18% between HA2-HA9; 30% on HA1) compared to the 1-5 year olds (6-11% between HA2-HA9; 22% for HA1). The largest difference between the nine different scenarios was between HA1 and HA2, which may be a reflection of the large sampling of blood samples taken in day 1 compared to day 2.

Table 5.12. Approach B: Predicted proportions of CA BSI, HCA BSI and HA BSI in children aged 1 month < 1 year, n=4,415

HA BSI threshold	CA BSI			HCA/HA BSI			HA BSI	
	a priori	predicted	%	predicted HCA (<1d after adm)	predicted HCA & HA (1 to x days after adm)	% range	*defined*	%
HA1	848	1,039	43%	1203	0	27%	1,325	30%
HA2	848	2,207	69%	422	135	10-13%	803	18%
HA3	848	2,342	72%	373	144	8-12%	708	16%
HA4	848	2,412	74%	341	153	8-11%	661	15%
HA5	848	2,457	75%	319	166	7-11%	625	14%
HA6	848	2,474	75%	317	178	7-11%	598	14%
HA7	848	2,485	75%	312	192	7-11%	578	13%
HA8	848	2,523	76%	299	193	7-11%	552	13%
HA9	848	2,552	77%	284	201	6-11%	530	12%

Table 5.13. Approach B: Predicted proportions of CA BSI, HCA BSI and HA BSI in children aged 1-5 years, n=6,172

HA BSI threshold	CA BSI			HCA/HA BSI			HA BSI	
	a priori	predicted	%	predicted HCA (<1d after adm)	predicted HCA & HA (1 to x days after adm)	% range	*defined*	%
HA1	1,711	2,292	65%	790	0	13%	1,379	22%
HA2	1,711	3,062	77%	582	129	9-12%	688	11%
HA3	1,711	3,206	80%	517	169	8-11%	569	9%
HA4	1,711	3,242	80%	504	192	8-11%	523	8%
HA5	1,711	3,341	82%	449	193	7-10%	478	8%
HA6	1,711	3,358	82%	445	209	7-11%	449	7%
HA7	1,711	3,409	83%	427	206	7-10%	419	7%
HA8	1,711	3,437	83%	412	218	7-10%	394	6%
HA9	1,711	3,471	84%	402	215	7-10%	373	6%

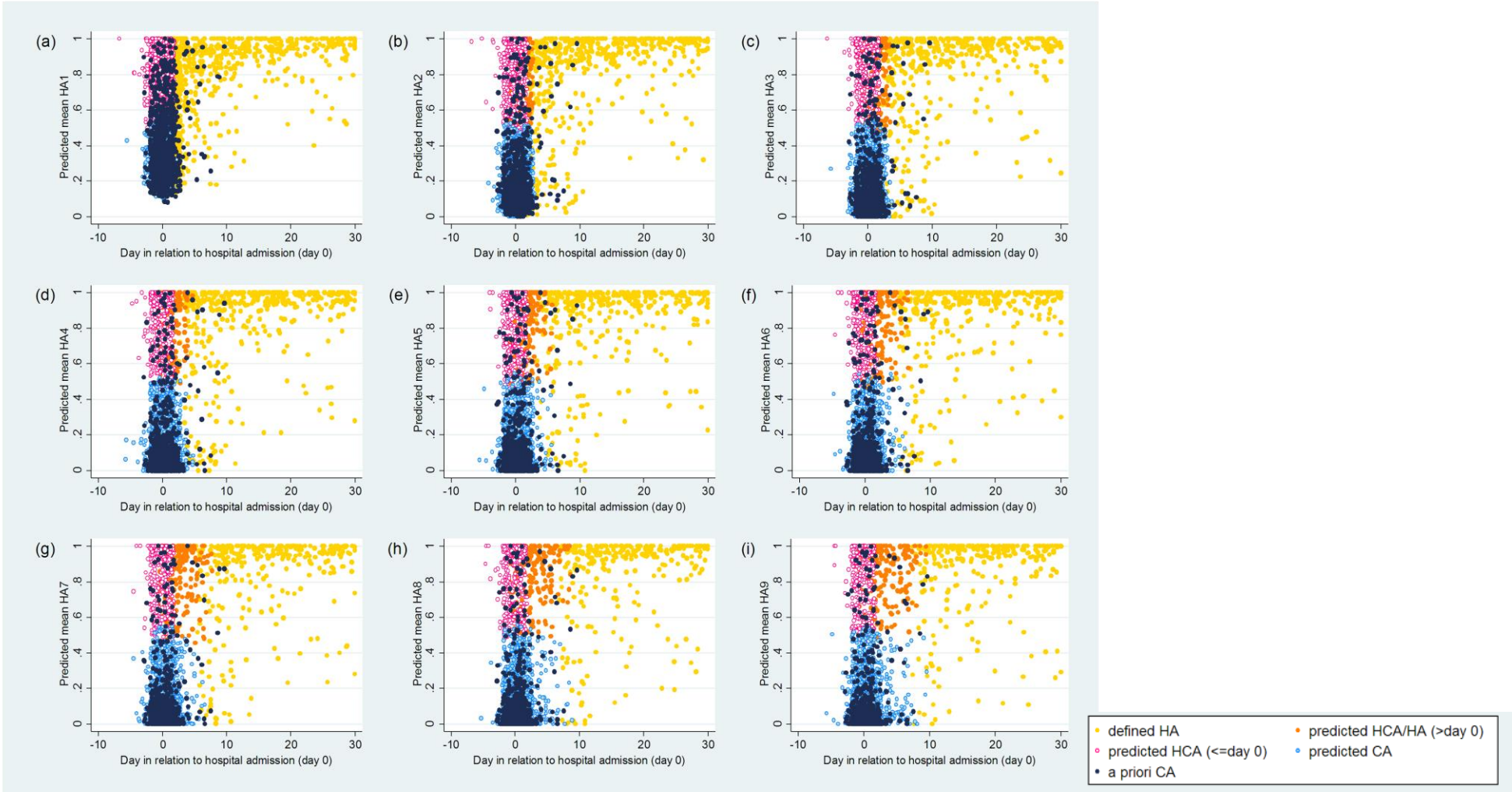


Figure 5.18. Distribution of predicted HCA, predicted HCA/HA, predicted CA, *a priori* CA and *defined* HA BSI between -5 and +30 of hospital admission in children aged 1 month < 1 year (jittered data points)

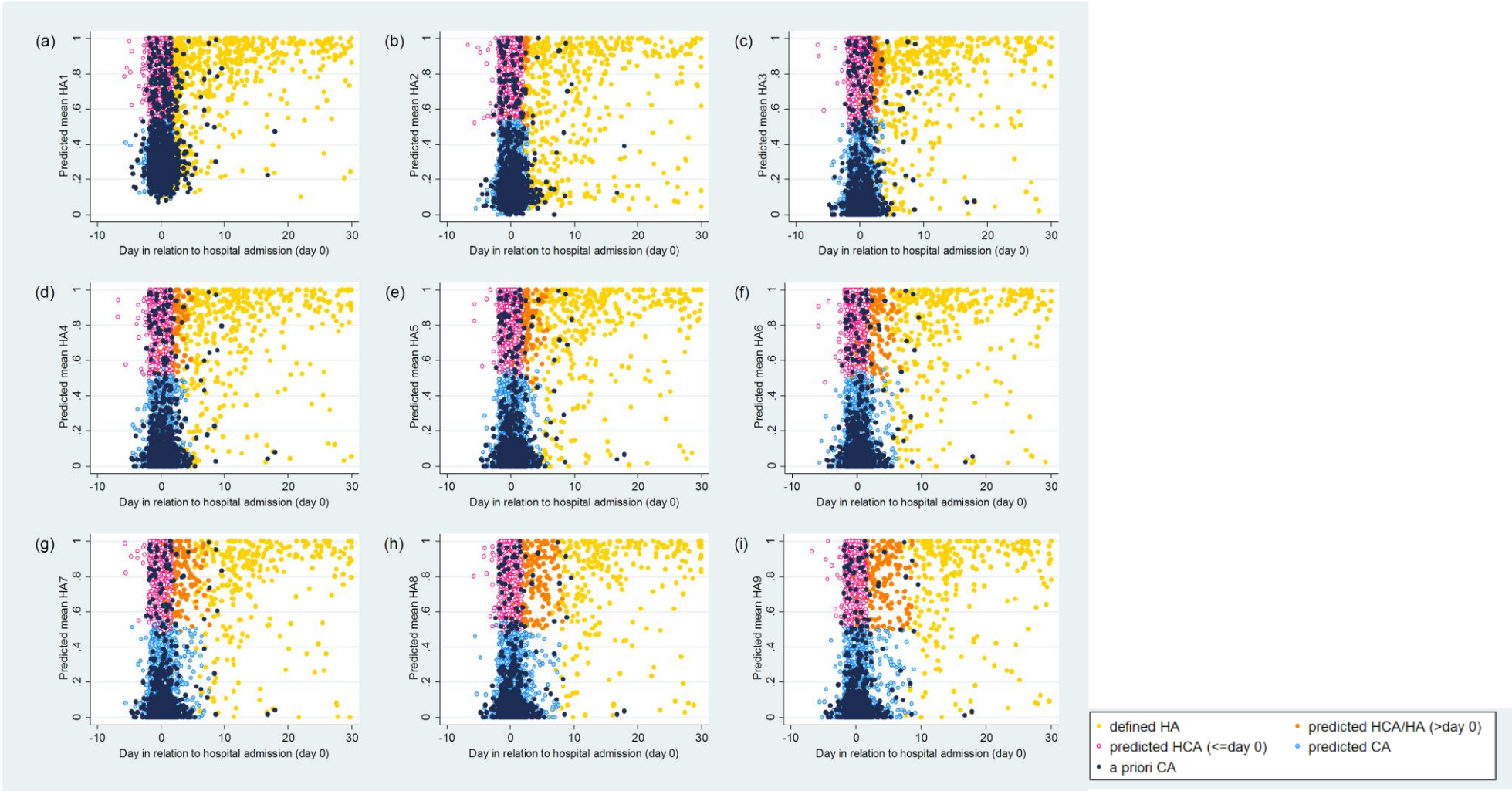


Figure 5.19. Distribution of the predicted HCA, predicted HCA/HA, predicted CA, *a priori* CA and *defined* HA BSI between -5 and +30 of hospital admission in children aged 1-5 years (jittered data points)

5.4.2.2 Summary of approach B results

In this approach, between 6% and a quarter of children with BSI in both age-groups were predicted to have HCA BSI, based on the assumption that children with HCA BSI had similar clinical characteristics to children with HA BSI. The majority of BSIs in both age-groups were predicted to be CA BSI (43-84%).

There was a clear distinction between the predicted values for the majority of *a priori* CA BSI and *defined* HA BSI in the graphs; this was expected since the regression model was based on only these two distinct populations and although the timing of the BSI was not directly included, the timing of events in relation to the BSIs were included and therefore would be highly predictive. The predicted proportions of CA, HCA and HA BSI were consistent when using thresholds HA2-HA9 in comparison to HA1. For example, children with a previous discharge (using HA1) had statistically significant higher odds of having *defined* HA BSI, whereas between HA5 and HA9, children had lower odds of having *defined* HA BSI, potentially as a result of the impact of selecting this low threshold and over-estimating HA rather than HCA BSI.

Unsurprisingly, children with an indwelling device in the previous 7-90 days had significantly higher odds of having a *defined* HA BSI compared to those with *a priori* CA BSI. This may be one of the reasons why so many children were predicted as having CA BSI because they did not have a history of an indwelling device in the year prior to the BSI (data presented in Chapter 6).

5.4.3 Results: approach C

5.4.3.1 Model 1: Multiple Correspondence Analysis (MCA)

The top 15 MCA dimensions and the corresponding proportion of inertia that each dimension accounted for are shown in Table 5.14 by age-group. The top 15 dimensions accounted for 82% (children aged 1 month < 1 year) and 83% (children aged 1-5 years) of all the inertia; the contribution to inertia beyond the 15th dimension diminished with each subsequent dimension (<0.01% per dimension). Since the first two dimensions accounted for the majority of the inertia (72% in children aged 1 month < 1 year; 75% in children aged 1-5 years), I decided to use only the first two dimensions for the rest of the approach C analysis.

Table 5.14 The MCA dimensions and the proportional inertia that they account for

	Children aged 1 month < 1 year			Children aged 1-5 years		
Dimension	**Inertia**	**% of inertia**	**% of inertia (cumulative)**	**Inertia**	**% of inertia**	**% of inertia (cumulative)**
1	0.0751188	62.8	62.8	0.0607750	67.1	67.1
2	0.0112575	9.4	72.2	0.0072357	8.0	75.1
3	0.0050917	4.3	76.4	0.0031687	3.5	78.6
4	0.0029841	2.5	78.9	0.0024434	2.7	81.3
5	0.0024623	2.1	81.0	0.0005901	0.7	81.9
6	0.0005174	0.4	81.4	0.0001592	0.2	82.1
7	0.0004495	0.4	81.8	0.0001415	0.2	82.2
8	0.0001755	0.2	81.9	0.0001133	0.1	82.4
9	0.0001240	0.1	82.0	0.0000675	0.1	82.4
10	0.0000879	0.1	82.1	0.0000416	0.1	82.5
11	0.0000600	0.1	82.2	0.0000263	<0.1	82.5
12	0.0000480	<0.1	82.2	0.0000206	<0.1	82.5
13	0.0000296	<0.1	82.2	0.0000114	<0.1	82.5
14	0.0000248	<0.1	82.3	0.0000088	<0.1	82.6
15	0.0000160	<0.1	82.3	0.0000028	<0.1	82.6
Total	**0.1196671**			**0.0906173**		

The relationships of the clinical and demographic categorical variables from dimension 1 and dimension 2 are plotted in Figure 5.20 for children aged 1 month < 1 year and in Figure 5.21 for children aged 1-5 years.

In children aged 1 month < 1 year, there was a cluster of children without any clinical characteristics (including no indwelling device, no previous procedure, no chronic or congenital conditions) with the exception of those with 1 chronic condition, on the border between quadrants I and II, alongside children who had a BSI isolated before hospital admission (quadrant I). Children with no previous postnatal stay and no previous discharge within the past year were completely separate from all other variables in quadrant II. Children with clinical characteristics based on clinical events (including indwelling devices, last invasive procedure) happening within a week prior to the BSI were grouped in quadrant III, or in quadrant IV if the characteristics were detected further back in time, the majority being for a month or more prior to the BSI. The day on which the BSI was isolated in relation to hospital admission was split between quadrant III and IV, where the increase in days into a hospital admission resulted in a higher negative dimension 1 value away from children who had a positive BSI before hospital admission. The clinical characteristics occurring between 7 days < 1 year before the BSI, and children who were born premature or with low birth-weight, or a congenital condition were predominantly grouped in quadrant IV. There were no large

differences observed by ethnicity, which were grouped at the meeting point of all four quadrants (with the exception of “other” or “not known”).

For children aged 1-5 years, there was a cluster of children with none of the defined clinical characteristics between quadrant I and II of Figure 5.21 (no previous discharge, no chronic or congenital conditions, no invasive procedures, no indwelling devices, term or healthy weight babies). Patients in the categorised age-groups were all clustered at the central value of zero. There was a large spread of children with defined clinical characteristics across quadrants III and IV. These were predominantly split into events occurring close to the time of the BSI (quadrant III) versus events that occurred further in the past (quadrant IV). In this age-group, children with two or more chronic conditions were all grouped in quadrant IV alongside the characteristics of congenital conditions, being premature or having a low birth weight.

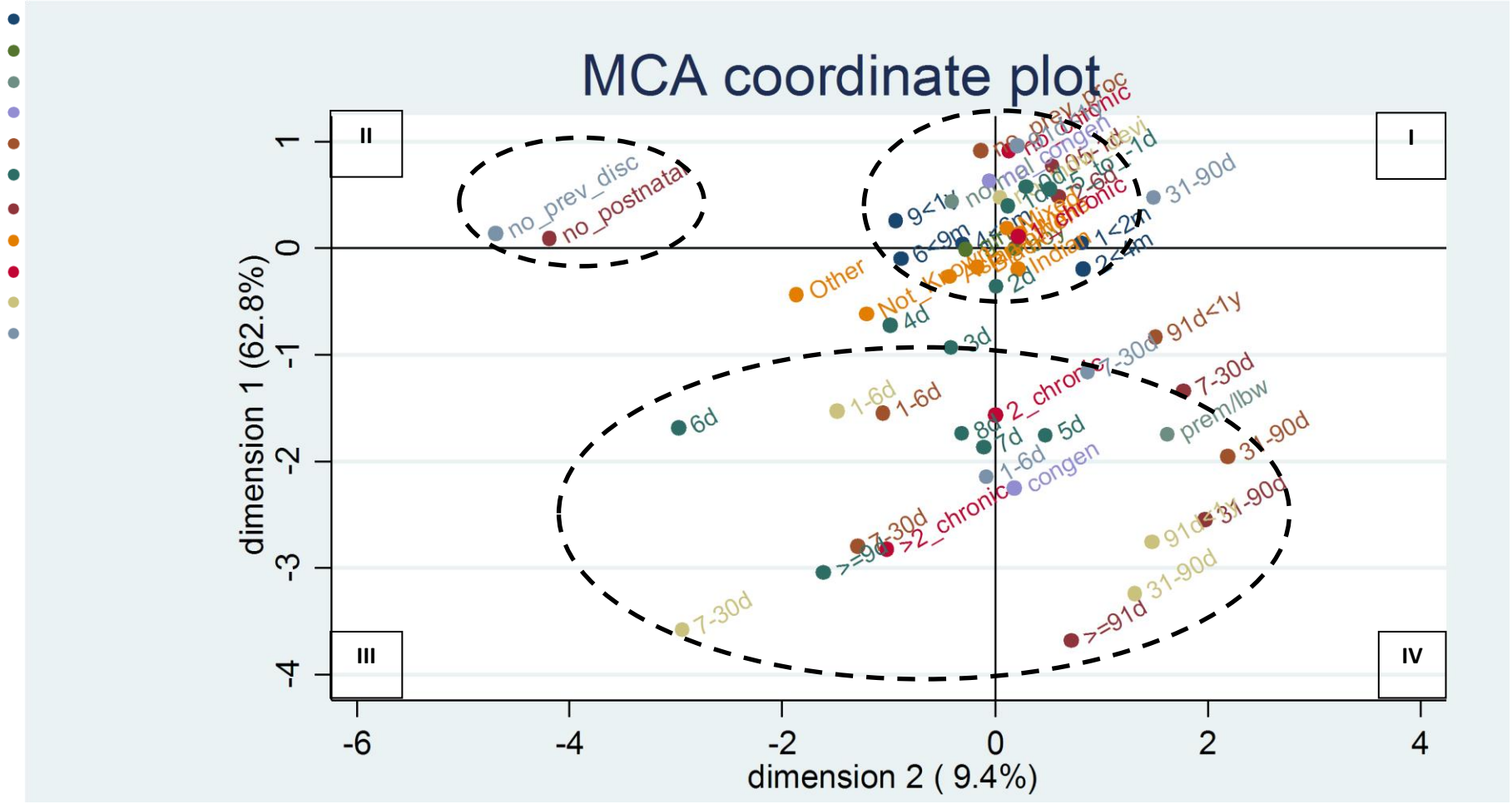


Figure 5.20. Plot of the multiple correspondence analysis results between dimension 1 and 2 for all the explanatory variables; children aged 1 month < 1year

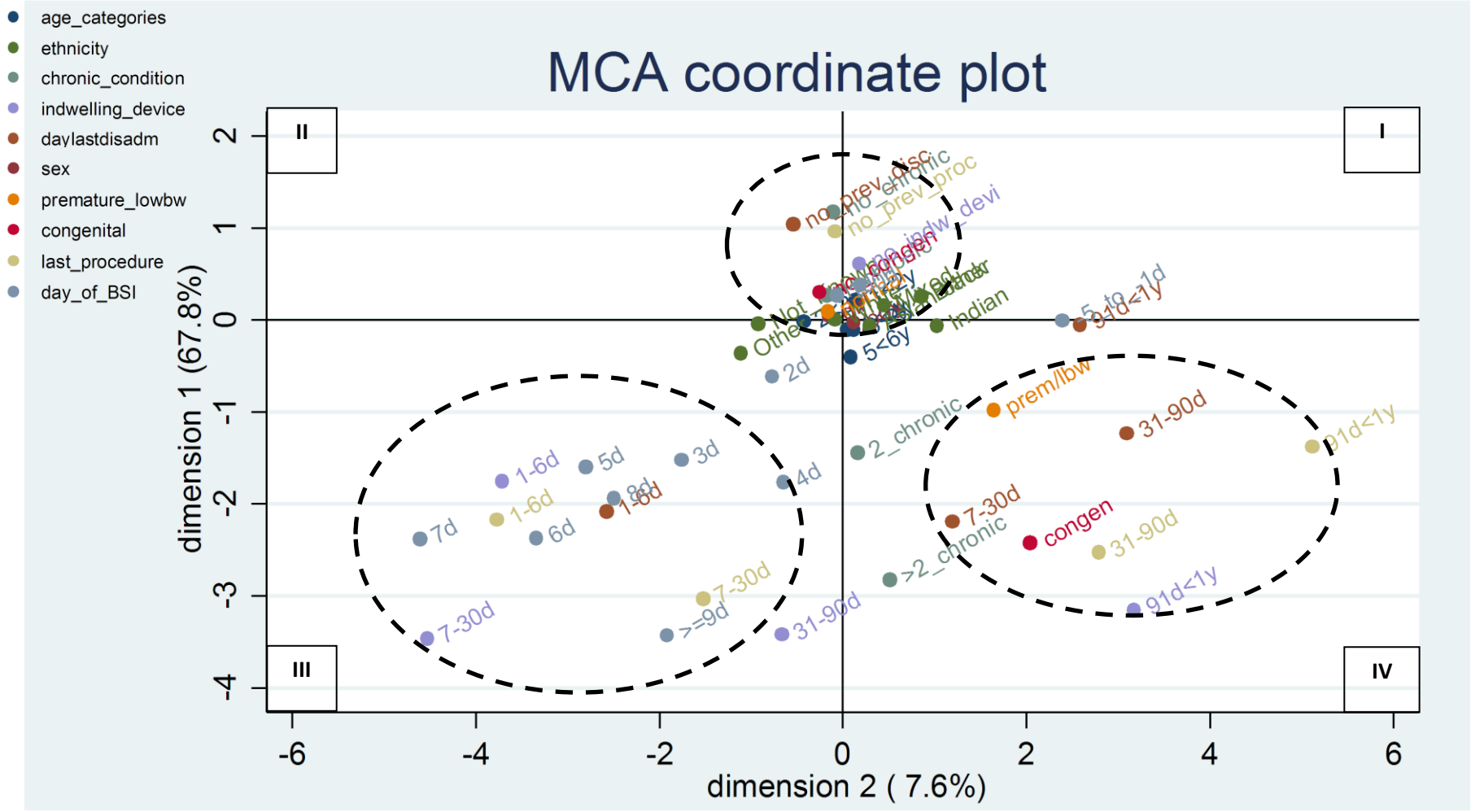


Figure 5.21. Plot of the multiple correspondence analysis results between dimension 1 and 2 for all the explanatory variables; children aged 1-5 years

To illustrate the distribution of the individual MCA dimension values assigned to each child, the scatter plots in Figure 5.22 and Figure 5.23 demonstrate this distribution by age-group. To help interpret the distribution (although the information was not used in the Approach C models), the children with *a priori* CA BSI (blue dots), *defined* HA (yellow dots; using the HA2 threshold), or undefined BSI (grey dots) are highlighted in both figures.

In both age-groups, there is a visible cluster of children with *a priori* CA BSI (between -1 to +1 in Dimension 1; 0 to +1 in Dimension 2) corresponding to children with few clinical characteristics, as seen in Figure 5.20 and Figure 5.21. The values for the remaining children with *defined* HA or undefined BSI were dispersed across the plotted range of values between -4 and +2 in children aged 1 month < 1 year, and from -4 to +4 in children aged 1-5 years.

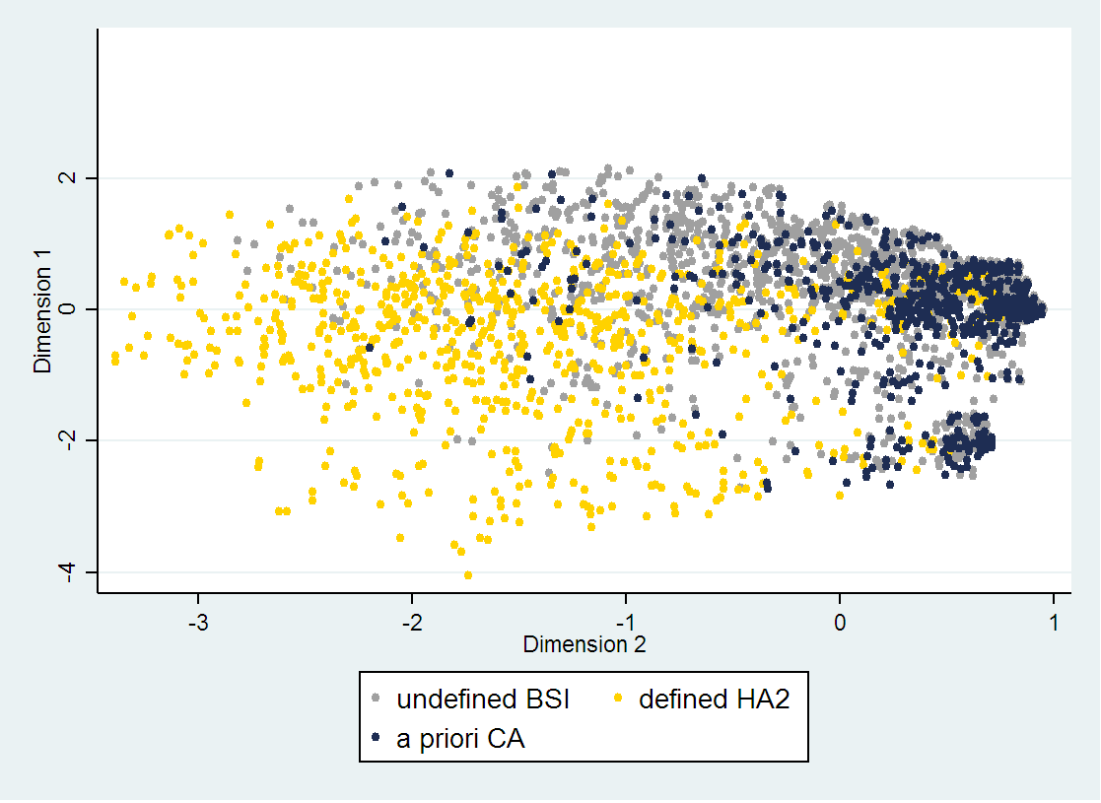


Figure 5.22. Scatter plot illustrating the values for dimension 1 and 2 assigned to each child, aged 1 month < 1 year, in the multiple correspondence analysis

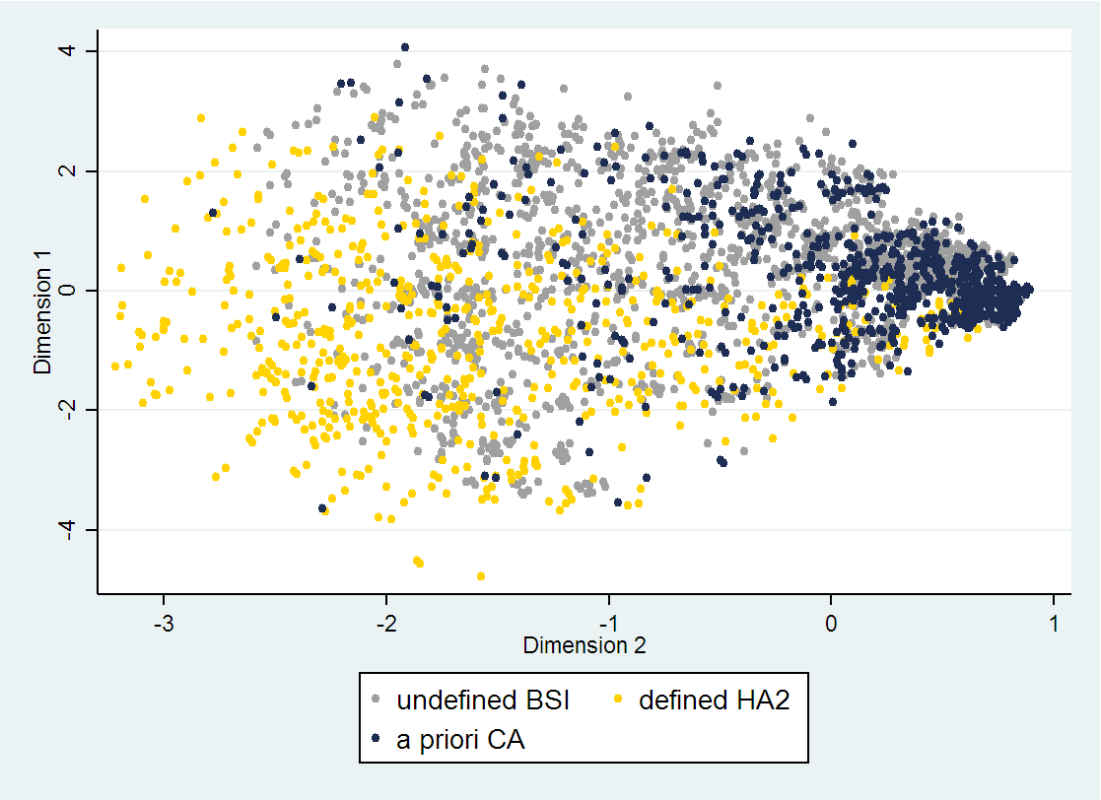


Figure 5.23. Scatter plot illustrating the values for dimension 1 and 2 assigned to each child, aged 1-5 years, in the multiple correspondence analysis

The predicted values from the two dimensions created in the MCA were used to inform the K-medians discriminatory analysis to identify the three subgroups of children described in the methods (section 5.3.3.2).

5.4.3.2 Model 2: K-medians analysis

The frequency of children grouped into the three K-medians groups using the MCA dimensional results are summarised at the top of Table 5.15 and are cross-tabulated with the *a priori* CA BSI, *defined* HA BSI and the remaining mixed BSI cases below. As a reminder, the three groups were compared to the number of children categorised as having *a priori* CA BSI (group 1), undefined mix of BSI (group 2) or *defined* HA BSI (group 3).

The majority of children in both age-groups were allocated to group 1 (60% for 1 month < 1 year olds; 68% in 1-5 year olds). Children in the younger age-group had the lowest number of children allocated to group 2 (15%), whereas the older age-group assigned the least number of children to group 3 (14%).

When compared to the assumptions made in approaches A and B, not all the *a priori* CA or *defined* HA BSI cases fitted perfectly into one particular group in either age-group. This is as expected since the *a priori* CA and *defined* HA groups would not contain all of the CA and HA BSI. Nevertheless, the awarded groupings do follow an interpretable pattern where group 1 accounted for the majority of *a priori* CA (75% for 1m<1yr; 86% for 1-5 years) and group 3 accounted for the majority of *defined* HA BSI (HA1: 50% for 1m<1yr; 39% for 1-5 years; HA9: 80% for 1m<1yr; 72% for 1-5 years). To predict the proportion of HCA BSI, group 2 would most likely represent this group, where between 15-18% of BSI were counted.

The results in Table 5.15 show that for both age-groups, there were low numbers (<20% for HA2, <10% for HA3-HA9) of the *a priori* CA and *defined* HA BSI in groups 1 and 3 respectively. The *a priori* CA observations was also low (<20%) in group 2 for all thresholds. There was more proportional variability of observations in group 2 for children with *defined* HA BSI. All but one (HA6) of the thresholds for the younger age-group showed <20% of the observations occurring in Group 2 for *defined* HA BSI; in the older age-group, only the HA1 threshold observes <20% of the observations occurring in Group 2 for *defined* HA BSI.

Table 5.15. Cross-tabulation of the K-medians groups with the *defined* HA, *a priori* CA and the remaining mixed BSI-cases by age-group

Optimal outcome:	Group identified from the model			
	1	2	3	Total
defined HA BSI	0	0	n_{3a}	n_{HA}
a priori CA BSI	n_{1b}	0	0	n_{CA}
remaining mixed BSI	n_{1c}	n_{2c}	n_{3c}	n_{mixed}

n = children

	Children aged 1 mth < 1 yr Group				Children aged 1 - 5 yrs Group			
	1	2	3	Total	1	2	3	Total
Total	2,636	649	1,130	4,415	4,205	1,102	865	6,172
%	60%	15%	26%		68%	18%	14%	
defined HA1	440	224	661	1,325	591	271	517	1,379
a priori CA	635	141	72	848	1,478	176	57	1,711
mixed	1,561	284	397	2,242	2,136	655	291	3,082
defined HA2	89	149	565	803	110	151	427	688
a priori CA	635	141	72	848	1,478	176	57	1,711
mixed	1,912	359	493	2,764	2,617	775	381	3,773
defined HA3	43	137	528	708	50	127	392	569
a priori CA	635	141	72	848	1,478	176	57	1,711
mixed	1,958	371	530	2,859	2,677	799	416	3,892
defined HA4	29	131	501	661	38	120	365	523
a priori CA	635	141	72	848	1,478	176	57	1,711
mixed	1,972	377	557	2,906	2,689	806	443	3,938
defined HA5	15	125	485	625	27	109	342	478
a priori CA	635	141	72	848	1,478	176	57	1,711
mixed	1,986	383	573	2,942	2,700	817	466	3,983
defined HA6	12	123	463	598	19	108	322	449
a priori CA	635	141	72	848	1,478	176	57	1,711
mixed	1,989	385	595	2,969	2,708	818	486	4,012
defined HA7	12	113	453	578	17	105	297	419
a priori CA	635	141	72	848	1,478	176	57	1,711
mixed	1,989	395	605	2,989	2,710	821	511	4,042
defined HA8	9	108	435	552	17	104	273	394
a priori CA	635	141	72	848	1,478	176	57	1,711
mixed	1,992	400	623	3,015	2,710	822	535	4,067
defined HA9	3	106	421	530	16	102	255	373
a priori CA	635	141	72	848	1,478	176	57	1,711
mixed	1,998	402	637	3,037	2,711	824	553	4,088

Legend <10% 10<20% of total in sub-group (e.g. *defined* HA, or *a priori* CA, or mixed)

5.4.3.3 Summary of approach C results

In the MCA analysis, there was an evident separation in both age-groups between the cluster of children with very few clinical conditions compared to children with multiple clinical conditions occurring in close time proximity to the BSI versus longer term clinical contact or conditions.

The proportion of BSI in groups 1, 2 and 3 remained the same (highlighted in orange in Table 5.15) regardless of the HA threshold used to interpret the frequency of children with CA, HCA and HA BSI. In both age-groups, the majority of BSI was attributed to CA BSI (group 1), with HCA BSI predicted to account for 15-18% of all BSI (group 2). There was a higher proportion of children predicted to have HA BSI in children aged 1 month < 1 year (26%) than in children aged 1-5 years (14%).

5.5 Chapter summary

This chapter has described the methods and results of the three statistical approaches used to predict the proportion of CA, HCA and HA BSI in children (by age-group), including the assumptions that were made in the models.

There was no consistent predicted proportion of CA or HCA BSI in approaches A, B and C. The majority of BSI in approach A were predicted to be HCA BSI (59-60%), whereas the majority of BSI in approach B (69-77%) and C (60-68%) were predicted to be CA BSI. Unlike CA and HCA BSI, the proportion of predicted HA BSI was more consistently predicted in all three models (≤26% in children aged 1 month < 1 year; ≤14% in children aged 1-5 years).

Although a completely different method was employed for approach C, the results most resembled the distributions found by approach B and reflected the clustering of patients with very few clinical characteristics compared to all other children, where there was a spectrum between recent and older time occurrence of clinical characteristics.

5.5.1 Discussion and limitations of the methodology

5.5.1.1 Predicted probability

In approaches A and B, logistic models were used to identify the features most associated with higher probabilities of being HA or CA for various subgroups of the data. These models were then applied in a wider group where the true distributions were unknown and all may be HA,

all may be CA, or more likely there was an intermediate distribution. To avoid super-imposing any prior belief on the percentages expected to be HA (or CA), those with probabilities of 0.5 or greater were allocated to the outcome category (HA or CA depending on the logistic model) on the basis that they were estimated to be more likely to be in that category ($p \geq 0.5$) than the alternative ($p < 0.5$).

Selecting an alternative estimated probability as the cut-point for allocation to categories would have required making some assumption as to the prior likelihood that a BSI was of a particular type (HA or CA). Whilst additional sensitivity models could have been incorporated to test the effect of different probability thresholds in Approach A (and Approach B), given that the purpose was to identify the percentages who belonged in these categories, this was unlikely to be of practical use to interpretation.

5.5.1.2 Best subset framework

Since all potential stepwise models are considered in the best subsets approach (which evaluates all possible models), it would be expected that the best subsets approach would be superior. The technique has long been shown to have advantages over other selection models in terms of selecting the correct model.(182) Berk made both theoretical and empirical comparisons to show that best subsets frequently outperforms other methods and is unlikely to perform poorly.(182) Whilst this is accepted, there is criticism that different subset size may have an effect on the selected subset, which is not a favorable feature. In my analyses, I was aware of the problems of interpreting major fluctuations in the included variable set with increasing subset sizes, but there were none so this was not an issue. The major disadvantage of best subsets is increased computation but this is now mostly negated by availability of computing packages with large capabilities. Hence there has been a growth in the use of best subset model fitting in recent years as computation has become easier.

The AIC and BIC are established means of comparing both nested and non-nested regression models. They are asymptotically equivalent to Mallows Cp.(183)

As regards model validation, these models were exploratory with respect to identification of individuals whose BSI were HA or CA. However, in the analyses, the models being produced are not intended to be used directly by other researchers, so this is likely less important than for those producing prognostic models for clinical application. In addition, I did not have the true outcomes to compare with my models. The precision of the fitted models could be investigated and how this translates into the likely error of predicted percentage distributions shown Figure 6.3. Bootstrap sampling may be used to achieve this investigation of the fitted

model precision in future work.(184-186) A further validation would be an investigation of model outliers using leave-one-out approaches and consideration of beta values for the subset used in the production of the models.

The outcomes of the three approaches will be compared in more detail in Chapter 6 and will include the sensitivity analyses in addition to the distribution of clinical characteristics and pathogen according to whether the BSI was predicted to be CA, HCA, or HA BSI.

Chapter 6. Distribution of risk factors and pathogens by predicted CA, HCA and HA BSI category

6.1 Introduction

In the previous chapter, I developed three statistical approaches (A, B and C) using the linked (MW-Child) dataset described in Chapters 3 and 4 to predict the proportion of CA, HCA and HA BSI using child demographics, clinical characteristics and the timing of positive blood samples in relation to hospital admission. This chapter aims to compare the results from approaches A, B and C and identify if, i) the clinical/demographic characteristics and timing of blood specimen adequately predict the source of infection; ii) the source of infection predicts the pathogen (Figure 6.1). Knowledge about the proportional burden of the source of BSI can inform preventive strategies according to the burden of BSI in each environment (community, or healthcare in the community, or hospital). Ideally, the subsequent identification of the causative pathogen, by source of infection, would help to inform treatment strategies. Although the optimum outcome would be to detect the antibiotic susceptibility of each invasive pathogen to tailor the antibiotic treatment, antibiotic susceptibility data was not included in this thesis due to variable completion of the data because of result suppression at the laboratory level; the implications of additional antimicrobial susceptibility data will be discussed in Chapter 7.

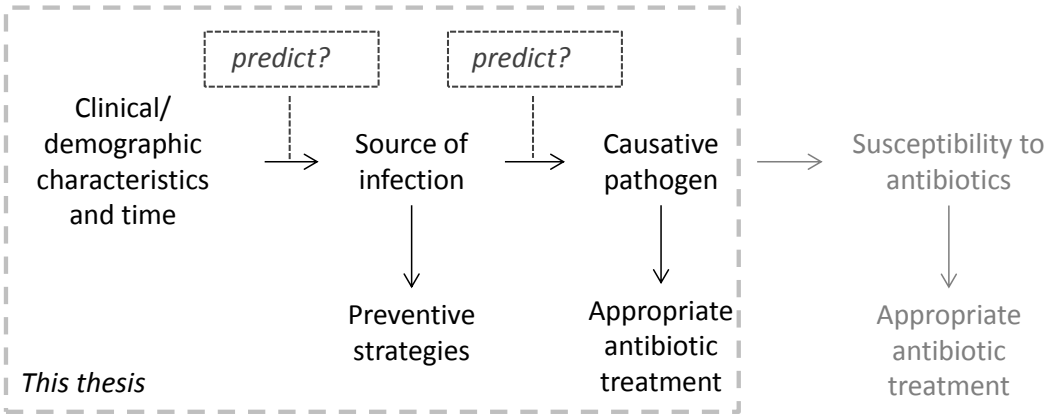


Figure 6.1. Steps taken in this thesis (in the box) to inform preventive and treatment strategies for children with BSI dependent on the prediction capability of the models

6.1.1 Chapter structure

I begin by comparing the predicted proportions of CA, HCA and HA BSI results from approaches A, B and C and their sensitivity analyses according to the nine HA1-HA9 thresholds. I go on to tabulate child demographic and clinical characteristics in relation to the predicted numbers of children with CA, HCA and HA BSI to describe the most prevalent risk factors that defined the groups of children. Subsequently, I reveal the pathogen distribution by specimen date in relation to hospital admission and the model distributions of the pathogen species by age-group for the three approaches. In the discussion, I consider and compare the predicted proportional distribution of CA, HCA and HA BSI from this thesis with the 2-day (proxy for 48-hour) threshold to determine what proportion of BSI are re-classified as HCA using the model predictions from Chapter 5. Figure 6.2 illustrates the context of this chapter (red) in relation to the rest of the thesis.

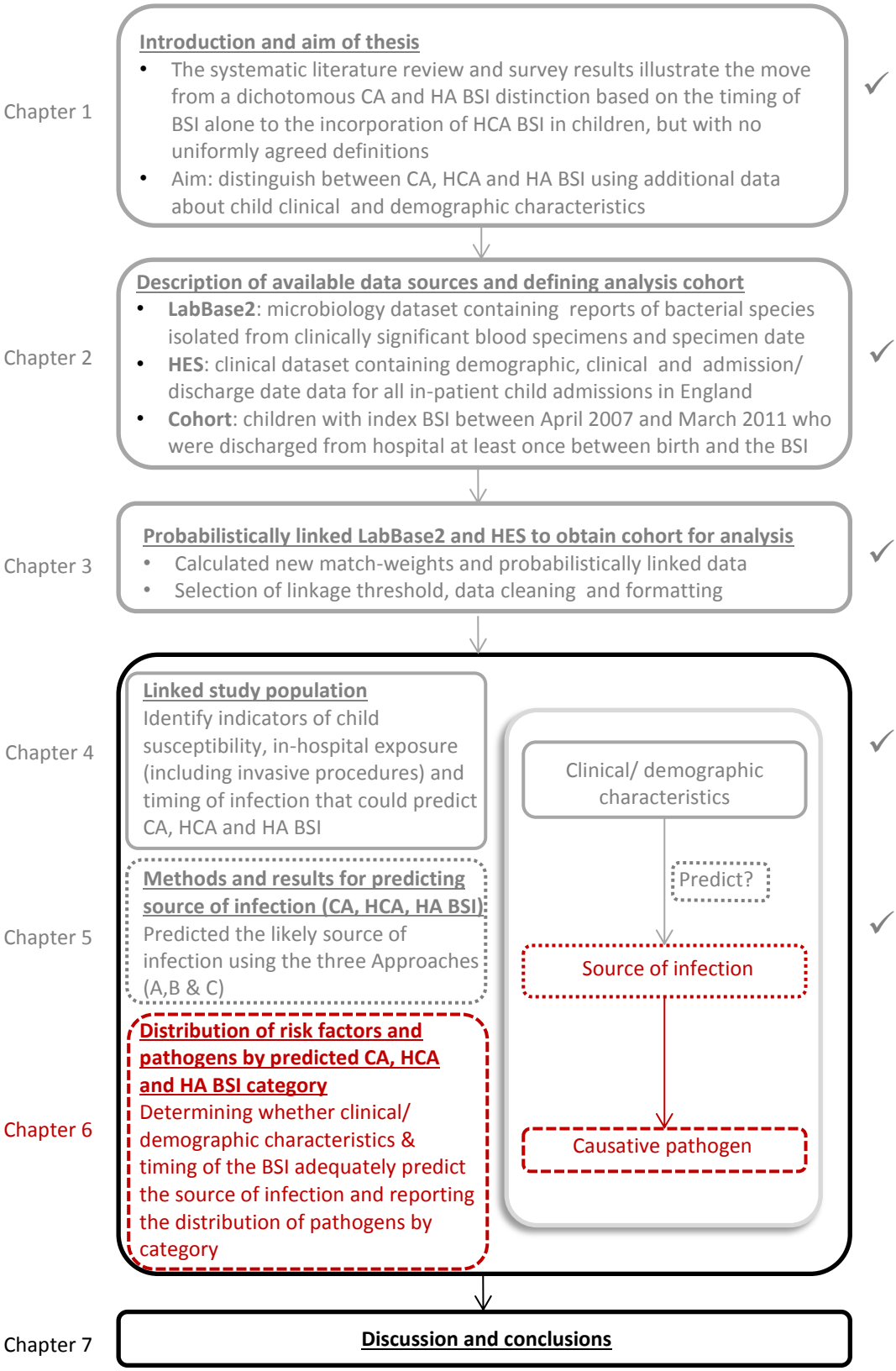


Figure 6.2. Flow diagram of the steps taken (grey) and those that will be addressed in this chapter (red) to predict the proportion of CA, HCA and HA BSI in children

6.2 Comparing the estimated proportion of CA, HCA and HA BSI according to each method (A, B and C)

6.2.1 Methods

Graphical plots were used to compare the predicted proportion of CA, HCA and HA BSI by age-group (using the nine HA BSI thresholds). The predicted proportions from the two sensitivity analyses (described in Chapter 5: 1. including hospital as a random effect; 2. removal of children with *a priori* CA BSI) were also depicted graphically. For comparability, the proportions of CA, HCA and HA BSI for approach C were plotted on the same graph axes as approaches A and B, but the predicted proportions remained the same across all the HA-thresholds as the thresholds were not incorporated into the models. The estimated proportions were compared with the commonly used CA/HA BSI (48-hour/2day) threshold to determine what proportion of BSI would be re-categorised as HCA using the results from this thesis.

6.2.2 Results

The predicted proportions of CA and HCA BSI were very different between approach A compared to approaches B or C, whereas the predicted proportions of HA BSI were similar for all three approaches (Figure 6.3). There was little change in the predicted proportions of CA, HCA and HA BSI by day of HA-threshold from the HA2-day threshold to the HA9-day threshold, with the exception of the predicted proportions on the HA1-day threshold in all models. As the results were predominantly consistent for eight of the nine HA-thresholds, the results using the HA2-day threshold (for approaches A and B) will be reported for the remainder of the results section for consistency.

The predicted proportion of HA BSI was the most consistently predicted category between all approaches A, B and C in children aged 1 month < 1 year (HA2: 18-26%) and in children aged 1-5 years (HA2: 11-14%; Table 6.1 and Table 6.2). The proportional difference between predicted CA and HCA BSI in approaches A, B and C spanned a range of 48%. Approach A predicted the highest proportion of HCA BSI in both age-groups (1 month < 1 year: 60%; 1-5 years: 59%) and approach B predicted the lowest (1 month < 1 year: 13%; 1-5 years: 12%). Inversely, approach A predicted the lowest proportion of CA BSI (1 month < 1 year: 21%; 1-5 years: 30%) and the highest proportion in approach B (1 month < 1 year: 69%; 1-5 years: 77%).

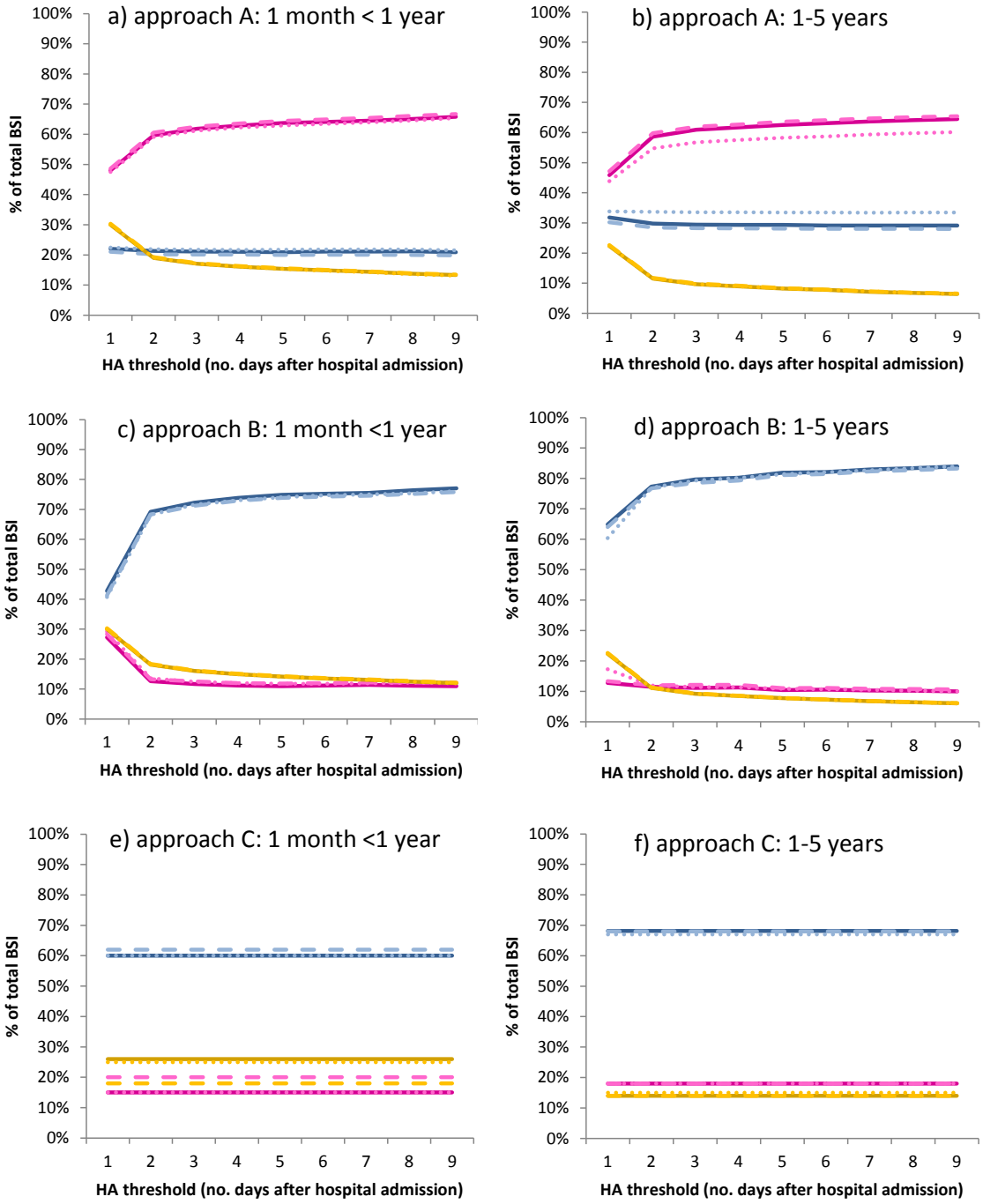


Figure 6.3. Proportion of BSI categorised as CA, HCA or HA for approaches A, B and C (including sensitivity analyses) by age-group and HA threshold

The predicted proportions of CA and HCA BSI for approach C fell between the estimates of approaches A and B, whereas the predicted proportion of HA BSI was 7% higher in children aged 1 month < 1 year and 2% higher in children aged 1-5 years.

Table 6.1. Final distribution of CA, HCA and HA BSI by approach A, B and C for children aged 1 month < 1 year using the 2-day *defined* HA threshold

	Total	CA BSI	%	HCA BSI	%	HA BSI	%
approach A	4,415	941	21%	2,635	60%	839	19%
approach B	4,415	3,055	69%	557	13%	803	18%
approach C	4,415	2,636	60%	649	15%	1,130	26%

Table 6.2. Final distribution of CA, HCA and HA BSI by approach A, B and C for children aged 1-5 years using the 2-day *defined* HA threshold

	Total	CA BSI	%	HCA BSI	%	HA BSI	%
approach A	6,172	1,839	30%	3,621	59%	712	12%
approach B	6,172	4,773	77%	711	12%	688	11%
approach C	6,172	4,205	68%	1,102	18%	865	14%

6.2.2.1 Results of the sensitivity analyses

Hospital as a random effect

There was very little change in the predicted proportions of CA, HCA or HA BSI when hospital was included as a random effect, with the exception of approach A in children aged 1-5 years where the gap between the proportion of CA and HCA BSI was narrowed by 8%, although HCA BSI still accounted for the majority of BSI (Figure 6.3b). The proportion of BSI classified as CA increased from 30% to 34% and the proportion of BSI classified as HCA decreased from 59% to 55%. Overall, I concluded that clustering by hospital did not appreciably affect the predicted distribution of CA, HCA or HA BSI

Removal of CA BSI occurring two or more days after hospital admission

Children with *a priori* CA BSIs were removed (n=129/2,559; 5%) from the dataset where their BSI occurred two or more days after hospital admission (n=46 children aged 1 month < 1 year; n=83 children aged 1-5 years). Of the 129/10,587 excluded children, approach C classified 77 as CA, 22 as HCA and 30 as HA. The only change in the ranking of the CA, HCA and HA categories was observed in approach C in children aged 1 month < 1 year (Figure 6.3) where proportionally more Group 2 BSI (most resembling HCA; 20%) than Group 3 (most resembling HA BSI; 18%) were predicted.

6.2.3 Comparing the estimated proportions of CA, HCA and HA BSI with the commonly used CA/HA BSI threshold

The literature review in Chapter 1 demonstrated that the 2-day (proxy for 48 hours) threshold was commonly employed to distinguish CA from HA BSI, but that timing of BSI alone has become insufficient to distinguish between them. Figure 6.4 illustrates the proportion of children that would have been classified as having CA (81% in children aged 1 month < 1 year; 88% in children aged 1-5 years) or HA BSI (19% in children aged 1 month < 1 year; 12% in children aged 1-5 years) using the 2-day threshold only. Of these CA and HA BSI categories, the breakdown of the proportion of BSI predicted to be CA, HCA or HA using approaches A, B and C are subsequently shown.

In Figure 6.4i, 81% of BSI in children aged 1 month < 1 year using the 2-day threshold only would be classified as CA, but when compared with the different approaches, predicted HCA BSI in approach A accounted for a significant proportion (74%). Although a smaller proportion, both approaches B and C predicted that 16% and 14% respectively were HCA. Of the 19% children classified as having HA BSI using the 2-day threshold only, the majority were also predicted to have HA BSI in all three approaches.

Similarly in Figure 6.4ii, 88% of BSI in children aged 1 month < 1 year using the 2-day threshold only would be classified as CA, but when compared with the different approaches, predicted HCA BSI in approach A accounted for a significant proportion (67%). As with the younger age-group, although a smaller proportion, both Approaches B and C predicted that 13% and 17% respectively of the 2-day threshold CA BSI were HCA BSI. Of the 12% children classified as having HA BSI using the 2-day threshold only, the majority were also predicted to have HA BSI in all three approaches.

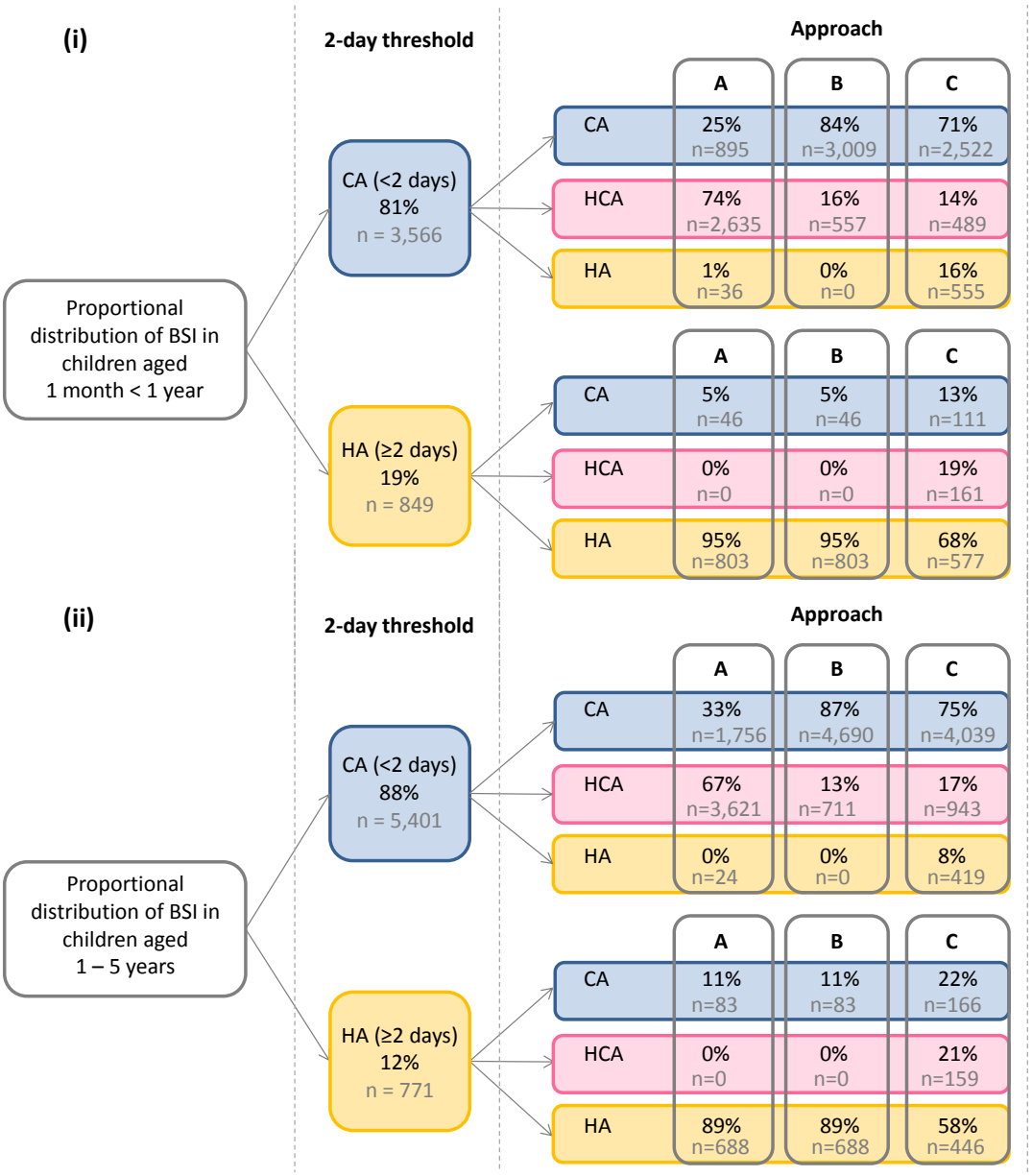


Figure 6.4. Comparing model classification of children with CA, HCA or HA BSI (using day 2 of hospital admission as the threshold for *defined* HA BSI) subdivided according to specimens taken before 2 days after admission and 2 or more days after admission for children aged (i) 1 month < 1 year, (ii) 1-5 years

6.2.4 Section summary

- The proportional estimates of HA BSI were low and consistent in all three approaches, with an estimated quarter of all BSI in children aged 1 month < 1 year and a sixth of all BSI in children aged 1-5 years.
- Although a low and a high estimate of CA and HCA BSI was expected, the predicted range of proportions was very wide (range of 48%), demonstrating that the clinical and demographic indicators do no differentiate sufficiently between CA and HA BSI.

Further investigation is required into the types of children allocated to each group and the pathogens responsible.

- Due to the high numbers of blood samples taken on admission and day 1 after admission in comparison to day 2-9 after admission, there is insufficient evidence to determine whether the 2-day threshold remains the best indicator to identify HA BSI. However, due to the stability of the predicted proportions between the 2-day and 9-day threshold, I decided to report the results of the 2-day HA-threshold for the remainder of the chapter.
- The comparison of the CA/HA threshold only with the predicted CA, HCA and HA BSI illustrates the importance of HCA BSI, distinguishing from CA BSI, which is not currently distinguished by the 2-day (48-hour) threshold. There is evidence to show that HCA BSI are becoming an important source of BSI and should be recognised formally in the considerations of the British empiric therapy guidance.

6.3 Distribution of risk factors by CA, HCA and HA BSI category

This section will present the distribution of risk factors according to the predicted CA, HCA or HA BSI categories (Figure 6.5).

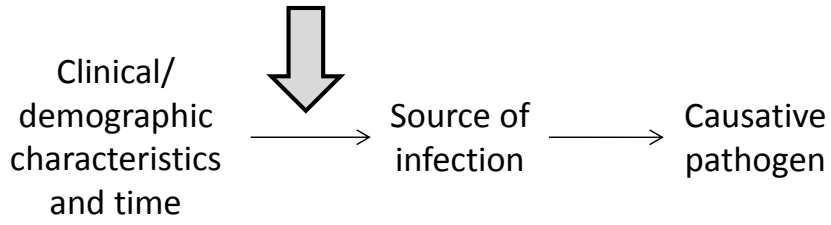


Figure 6.5. Clinical/demographic characteristics and time used to predict the source of infection

6.3.1 Methods

The distributions of demographic and clinical characteristics of children by predicted CA, HCA or HA BSI category were cross-tabulated and compared for approaches A, B and C using the 2-day HA threshold.

6.3.2 Results

6.3.2.1 Children aged 1 month < 1 year

In all three approaches (Table 6.3), children with HA BSI had the highest prevalence of more than two chronic conditions (39-43%), compared to children with HCA BSI (6-28%), or CA BSI (1-4%). There was little proportional variation separating the three categories for children with only one chronic condition.

Children with HA BSI had the highest prevalence of: congenital conditions (60-64%) compared to HCA BSI (14-61%) or CA BSI (5-8%); invasive procedures in the past 1-30 days (19-31%) compared to HCA BSI (3-24%) or CA BSI (1-7%); previous discharge in the past 1-30 days (23-35%) compared to HCA BSI (7-37%) or CA BSI (7-11%); and indwelling devices in the past year (43-45%) where the proportion increased with time proximity to the BSI sample date.

Seven (58%) of the 12 children with an *a priori* CA BSI and an indwelling device had their device inserted the day preceding the positive blood sample, suggesting that the indwelling device was unlikely to be the source of the BSI due to the short time frame. There was little difference between the proportion of CA, HCA or HA BSI for sex or ethnicity.

Table 6.3. Distribution of demographic and clinical characteristics by approach A, B and C for children aged 1 month < 1 year (HA2 threshold)

Clinical and demographic characteristics	Approach A			Approach B			Approach C			Total
	CA %	HCA %	HA2 %	CA %	HCA/ HA %	HA %	Group 1 (most like CA) %	Group 2 (most like HCA) %	Group 3 (most like HA) %	
boy	537 57%	1,605 61%	473 56%	1,805 59%	359 64%	451 56%	1,577 60%	359 55%	679 60%	2,615
girl	404 43%	1,030 39%	366 44%	1,250 41%	198 36%	352 44%	1,059 40%	290 45%	451 40%	1,800
White	746 79%	1,776 67%	475 57%	2,161 71%	384 69%	452 56%	1,891 72%	390 60%	716 63%	2,997
Mixed	26 3%	89 3%	26 3%	103 3%	12 2%	26 3%	91 3%	18 3%	32 3%	141
Indian subcontinent	53 6%	259 10%	96 11%	261 9%	55 10%	92 11%	239 9%	47 7%	122 11%	408
Any other Asian	7 1%	85 3%	34 4%	78 3%	17 3%	31 4%	66 3%	19 3%	41 4%	126
Black	23 2%	169 6%	54 6%	163 5%	30 5%	53 7%	144 5%	35 5%	67 6%	246
other ethnic group	23 2%	87 3%	51 6%	90 3%	21 4%	50 6%	70 3%	51 8%	40 4%	161
NK	63 7%	170 6%	103 12%	199 7%	38 7%	99 12%	135 5%	89 14%	112 10%	336
normal	815 87%	2,176 83%	549 65%	2,638 86%	379 68%	523 65%	2,360 90%	620 96%	560 50%	3,540
premature/low birthweight	126 13%	459 17%	290 35%	417 14%	178 32%	280 35%	276 10%	29 4%	570 50%	875
no chronic condition	437 46%	1,747 66%	124 15%	2,056 67%	128 23%	124 15%	1,819 69%	348 54%	141 12%	2,308
1chronic	369 39%	584 22%	166 20%	812 27%	147 26%	160 20%	691 26%	158 24%	270 24%	1,119
2chronics	95 10%	155 6%	187 22%	134 4%	126 23%	177 22%	93 4%	65 10%	279 25%	437
>2chronic	40 4%	149 6%	362 43%	53 2%	156 28%	342 43%	33 1%	78 12%	440 39%	551
no congenital condition	865 92%	2,257 86%	325 39%	2,910 95%	219 39%	318 40%	2,491 94%	554 85%	402 36%	3,447
congenital condition	76 8%	378 14%	514 61%	145 5%	338 61%	485 60%	145 6%	95 15%	728 64%	968
no invasive procedure	772 82%	2,006 76%	180 21%	2,634 86%	146 26%	178 22%	2,307 88%	518 80%	133 12%	2,958
91days<1yr	63 7%	188 7%	45 5%	196 6%	56 10%	44 5%	140 5%	3 0%	153 14%	296
31-90days	30 3%	235 9%	126 15%	133 4%	137 25%	121 15%	77 3%	8 1%	306 27%	391
7-30days	12 1%	127 5%	262 31%	19 1%	132 24%	250 31%	28 1%	52 8%	321 28%	401
1-6days	64 7%	79 3%	226 27%	73 2%	86 15%	210 26%	84 3%	68 10%	217 19%	369
no previous discharge	115 12%	300 11%	154 18%	373 12%	45 8%	151 19%	0 0%	536 83%	33 3%	569
91days<1yr	510 54%	779 30%	92 11%	1,273 42%	16 3%	92 11%	1,257 48%	34 5%	90 8%	1,381
31-90days	171 18%	922 35%	128 15%	939 31%	157 28%	125 16%	988 37%	10 2%	223 20%	1,221
7-30days	63 7%	437 17%	192 23%	303 10%	208 37%	181 23%	279 11%	25 4%	388 34%	692
1-6days	82 9%	197 7%	273 33%	167 5%	131 24%	254 32%	112 4%	44 7%	396 35%	552
no indwelling device	904 96%	2,459 93%	477 57%	2,999 98%	382 69%	459 57%	2,626 100%	593 91%	621 55%	3,840
91days<1yr last insertion	18 2%	75 3%	57 7%	39 1%	57 10%	54 7%	4 0%	2 0%	144 13%	150
31-90days last insertion	4 0%	80 3%	108 13%	4 0%	87 16%	101 13%	2 0%	3 0%	187 17%	192
7-30days last insertion	1 0%	21 1%	138 16%	1 0%	27 5%	132 16%	0 0%	34 5%	126 11%	160
1-6days last insertion	14 1%	0 0%	59 7%	12 0%	4 1%	57 7%	4 0%	17 3%	52 5%	73
Total	941	2,635	839	3,055	557	803	2,636	649	1,130	4,415

6.3.2.2 Children aged 1-5 years

In all three approaches, children with CA BSI (Table 6.4) had the lowest proportion of prematurity at birth or low birth-weight, chronic conditions, congenital conditions, previous procedure, previous discharge or previous invasive device compared to children with HCA or HA BSI. This was most pronounced for patients with indwelling devices, where ≤1% of patients had any mention of an indwelling device in the past 1-90 days.

The HA BSI category contained the largest proportion of children with more than two chronic conditions for approaches A, B and C (50%, 49% and 43% respectively), with the lowest proportion found in children with CA BSI (7%, 6% and 2% respectively).

The HCA BSI category in approaches B and C contained a larger proportion of children with congenital conditions (B = 42%, C = 39%) than the HA BSI category (B = 30%, C = 19%), and the largest proportion of children with a previous discharge in the past year (particularly in the past 1-30 days) compared to the CA or HA BSI categories. The HCA category in approach A contained a smaller proportion of children with congenital or previous discharge compared to the HA BSI category.

In all three approaches, children with HA BSI had the highest proportion of indwelling devices or invasive procedures inserted in the past 1-6 days (8-13% and 31-49% respectively) compared to children with HCA BSI (1-7% and 1-21% respectively) and CA BSI (0% and 1-3% respectively). Little difference was observed by sex or ethnicity between the CA, HCA or HA BSI categories.

Table 6.4. Distribution of demographic and clinical characteristics by approach A, B and C for children aged 1 - 5 years (HA2 threshold)

Clinical and demographic characteristics	Approach A			Approach B			Approach C			Total
	CA %	HCA %	HA2 %	CA %	HCA/ HA %	HA %	Group 1 (most like CA) %	Group 2 (most like HCA) %	Group 3 (most like HA) %	
boy	1,054 57%	2,022 56%	407 57%	2,699 57%	391 55%	393 57%	2,331 55%	679 62%	473 53%	3,483
girl	785 43%	1,599 44%	305 43%	2,074 43%	320 45%	295 43%	1,874 45%	422 38%	393 44%	2,689
White	1,268 69%	2,449 68%	446 63%	3,251 68%	481 68%	431 63%	2,844 68%	717 65%	602 68%	4,163
Mixed	72 4%	112 3%	14 2%	170 4%	14 2%	14 2%	136 3%	40 4%	22 2%	198
Indian subcontinent	133 7%	343 9%	76 11%	412 9%	69 10%	71 10%	360 9%	140 13%	52 6%	552
Any other Asian	33 2%	119 3%	22 3%	115 2%	38 5%	21 3%	113 3%	39 4%	22 2%	174
Black	98 5%	258 7%	41 6%	324 7%	32 5%	41 6%	277 7%	86 8%	34 4%	397
other ethnic group	44 2%	124 3%	34 5%	143 3%	26 4%	33 5%	132 3%	24 2%	46 5%	202
NK	191 10%	216 6%	79 11%	358 8%	51 7%	77 11%	343 8%	55 5%	88 10%	486
normal	1,702 93%	3,306 91%	610 86%	4,429 93%	596 84%	593 86%	3,955 94%	875 79%	788 89%	5,618
premature/low birthweight	137 7%	315 9%	102 14%	344 7%	115 16%	95 14%	250 6%	226 21%	78 9%	554
no chronic condition	881 48%	1,833 51%	65 9%	2,676 56%	38 5%	65 9%	2,626 62%	115 10%	38 4%	2,779
1chronic	660 36%	880 24%	158 22%	1,394 29%	149 21%	155 23%	1,259 30%	236 21%	203 23%	1,698
2chronics	177 10%	466 13%	133 19%	435 9%	210 30%	131 19%	250 6%	284 26%	242 27%	776
>2chronic	121 7%	442 12%	356 50%	268 6%	314 44%	337 49%	70 2%	466 42%	383 43%	919
no congenital condition	1,764 96%	3,241 90%	494 69%	4,609 97%	409 58%	481 70%	4,131 98%	674 61%	694 78%	5,499
congenital condition	75 4%	380 10%	218 31%	164 3%	302 42%	207 30%	74 2%	427 39%	172 19%	673
no invasive procedure	1,630 89%	2,589 71%	146 21%	4,122 86%	97 14%	146 21%	4,118 98%	215 20%	32 4%	4,365
91days<1yr	80 4%	284 8%	39 5%	257 5%	107 15%	39 6%	0 0%	402 37%	1 0%	403
31-90days	40 2%	252 7%	72 10%	146 3%	146 21%	72 10%	9 0%	306 28%	49 6%	364
7-30days	31 2%	281 8%	229 32%	111 2%	210 30%	220 32%	26 1%	166 15%	349 39%	541
1-6days	58 3%	215 6%	226 32%	137 3%	151 21%	211 31%	52 1%	12 1%	435 49%	499
no previous discharge	1,405 76%	1,892 52%	235 33%	3,289 69%	13 2%	230 33%	3,360 80%	33 3%	139 16%	3,532
91days<1yr	182 10%	476 13%	75 11%	625 13%	34 5%	74 11%	387 9%	314 29%	32 4%	733
31-90days	84 5%	309 9%	67 9%	292 6%	104 15%	64 9%	128 3%	300 27%	32 4%	460
7-30days	64 3%	432 12%	119 17%	258 5%	242 34%	115 17%	129 3%	361 33%	125 14%	615
1-6days	104 6%	512 14%	216 30%	309 6%	318 45%	205 30%	201 5%	93 8%	538 61%	832
no indwelling device	1,783 97%	3,079 85%	345 48%	4,681 98%	190 27%	336 49%	4,203 100%	724 66%	280 32%	5,207
91days<1yr last insertion	30 2%	235 6%	80 11%	66 1%	201 28%	78 11%	2 0%	270 25%	73 8%	345
31-90days last insertion	11 1%	158 4%	93 13%	11 0%	162 23%	89 13%	0 0%	100 9%	162 18%	262
7-30days last insertion	6 0%	104 3%	136 19%	6 0%	110 15%	130 19%	0 0%	6 1%	240 27%	246
1-6days last insertion	9 0%	45 1%	58 8%	9 0%	48 7%	55 8%	0 0%	1 0%	111 13%	112
Total	1,839	3,621	712	4,773	711	688	4,205	1,101	886	6,172

6.3.3 Section summary

- As expected, children with CA BSI had the smallest proportion of risk factors indicating child susceptibility or direct exposure to infection compared to the HCA or HA BSI categories. This is because healthy children are much less likely to develop a BSI from opportunistic pathogens that rely on immune-suppression to multiply, as found in children with HCA or HA BSI, and therefore the CA BSI group should contain the highest number of previously healthy children prior to a BSI, even though some children with CA BSI will also have underlying medical conditions.
- The younger children (aged 1 month < 1 year) categorised as having HA BSI had more clinical characteristics than children with HCA BSI, whereas with an increase in age, children aged 1-5 years with HCA BSI had proportionally the same or more clinical characteristics than children with HA BSI.
- Not all the clinical characteristics were clear predictors of CA, HCA or HA BSI.
- A few risk factors had a low prevalence but a high predictive value to distinguish between CA and HCA BSI in all approaches and age-groups: more than two chronic and congenital conditions, indwelling devices in the past 1-30 days, and invasive procedures in the past 1-30 days.

6.4 Distribution of bacterial species by time and BSI category

This section presents the distribution of pathogens isolated from blood specimens in relation to the timing of hospital admission and investigates how well the predicted CA, HCA and HA BSI groups in approaches A, B and C can predict the causative pathogen to inform treatment strategies (Figure 6.6).

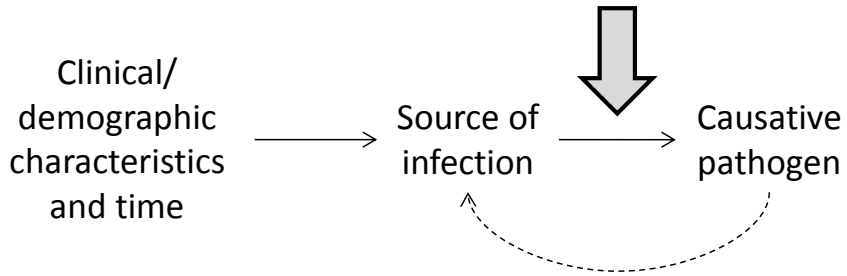


Figure 6.6. The predicted source of infection used to predict the causative pathogen

6.4.1 Methods

The distribution and frequency of pathogen species in relation to hospital admission were tabulated and shown graphically by age-group for the 16 most frequently isolated pathogens from children with a BSI: CoNS, *S. aureus* (including the breakdown into MRSA and MSSA), GAS, GBS, *S. pneumoniae,* non-pyogenic streptococcus, *Enterococcus* spp., *Micrococcus* spp., *E. coli, Klebsiella* spp., *Enterobacter* spp., *P. aeruginosa, Acinetobacter* spp., *Salmonella* spp., *N. meningitidis* and *H. influenzae.* The remaining less frequently reported pathogens were grouped into two "other" categories for the remaining Gram-positive or Gram-negative pathogen species.

6.4.2 Results

A total of 4,650 pathogens from 52 genera were isolated from the 4,415 children with BSI aged 1 month < 1 year, of whom 5% (n = 219 children) had polymicrobial BSI. In children aged 1-5 years, 6,629 pathogens from 63 genera were isolated from blood, of which 7% (n = 409 children) had polymicrobial BSI. The proportions of BSIs by pathogen and specimen date in relation to hospital admission for all children are summarised in Table 6.5 and Table 6.6. The frequency and distribution of positive blood samples by day in relation to before and the first 30 days of hospital admission are illustrated in Figure 6.7 (children aged 1 month < 1 year) and Figure 6.8 (children aged 1-5 years).

Table 6.5. Proportional distribution of BSI by pathogen and timing of the positive specimen in relation to hospital admission for children aged 1 month < 1 year

		Days either side of admission to hospital (day=0)							
Pathogen	**Total**	**-5d to +1d**	**(95% CI)**	**+2d to +9d**	**(95% CI)**	**+10d to +30d**	**(95% CI)**	**>30 d**	**(95% CI)**
Gram-positive									
CoNS	1,231	76%	(74-79)	9%	(8-11)	8%	(6-9)	7%	(6-8)
S. aureus	392	68%	(64-73)	12%	(9-15)	10%	(8-13)	10%	(7-13)
MSSA	*337*	*69%*	*(64-73)*	*12%*	*(9-16)*	*10%*	*(7-13)*	*9%*	*(7-13)*
MRSA	*25*	*60%*	*(41-76)*	*8%*	*(2-25)*	*20%*	*(9-39)*	*12%*	*(4-30)*
Non-pyogenic streptococcus	435	89%	(85-91)	8%	(6-11)	2%	(1-4)	2%	(1-3)
Enterococcus spp.	193	38%	(31-45)	18%	(13-24)	17%	(12-23)	27%	(21-34)
Micrococcus spp.	102	93%	(86-97)	5%	(2-11)	0%	(0-4)	2%	(1-7)
Group B streptococcus	213	93%	(89-96)	3%	(1-6)	2%	(1-5)	1%	(0-4)
S. pneumoniae	340	94%	(91-96)	6%	(4-8)	1%	(0-2)	0%	(0-1)
Group A streptococcus	117	98%	(93-99)	2%	(1-7)	0%	(0-3)	0%	(0-3)
Other	143	90%	(83-93)	4%	(2-9)	3%	(1-7)	3%	(2-8)
Gram-negative									
E. coli	582	91%	(88-93)	4%	(3-6)	2%	(1-4)	3%	(2-5)
Klebsiella spp.	121	52%	(43-61)	12%	(8-19)	12%	(8-19)	23%	(17-31)
Enterobacter spp.	62	21%	(13-33)	18%	(10-29)	19%	(11-31)	42%	(30-54)
P. aeruginosa	48	52%	(38-65)	10%	(5-22)	19%	(10-32)	19%	(10-32)
Acinetobacter spp.	54	85%	(74-92)	7%	(3-17)	2%	(0-10)	5%	(2-15)
Salmonella spp.	25	84%	(65-93)	16%	(6-35)	0%	(0-13)	0%	(0-13)
N. meningitidis	367	95%	(92-97)	5%	(3-8)	0%	(0-1)	0%	(0-1)
H. influenzae	52	88%	(77-94)	6%	(2-16)	4%	(1-13)	2%	(0-10)
Other	173	68%	(61-75)	7%	(4-12)	12%	(8-18)	12%	(8-18)
Total	**4,650**	**80%**	**(79-81)**	**8%**	**(7-9)**	**6%**	**(5-6)**	**6%**	**(6-7)**

Table 6.6. Proportional distribution of BSI by pathogen and timing of the positive specimen in relation to hospital admission for children aged 1-5 years

		Days either side of admission to hospital (day=0)							
Pathogen	**Total**	**-5d to +1d**	**(95% CI)**	**+2d to +9d**	**(95% CI)**	**+10d to +30d**	**(95% CI)**	**>30 d**	**(95% CI)**
Gram-positive									
CoNS	1,692	87%	(85-88)	7%	(5-8)	4%	(3-5)	3%	(2-4)
S. aureus	571	84%	(81-87)	8%	(6-10)	6%	(4-8)	3%	(2-4)
MSSA	*477*	*86%*	*(82-89)*	*7%*	*(5-10)*	*5%*	*(3-7)*	*2%*	*(1-4)*
MRSA	*18*	*56%*	*(34-75)*	*17%*	*(6-39)*	*17%*	*(6-39)*	*11%*	*(3-33)*
Non-pyogenic streptococcus	871	91%	(89-93)	5%	(3-6)	3%	(2-4)	2%	(1-3)
Enterococcus spp.	176	56%	(49-63)	13%	(9-19)	18%	(13-25)	13%	(8-18)
Micrococcus spp.	216	93%	(89-96)	5%	(3-8)	1%	(0-3)	1%	(0-4)
Group B streptococcus	10	80%	(52-94)	10%	(2-38)	0%	(0-26)	10%	(2-38)
S. pneumoniae	707	96%	(94-97)	3%	(2-5)	<1%	(0-1)	0%	(0-1)
Group A streptococcus	394	96%	(93-97)	4%	(3-6)	<1%	(0-1)	0%	(0-1)
Other	305	89%	(84-92)	7%	(4-10)	3%	(2-6)	2%	(1-4)
Gram-negative									
E. coli	170	66%	(59-73)	12%	(11-12)	16%	(11-22)	6%	(4-11)
Klebsiella spp.	96	61%	(51-71)	13%	(8-22)	14%	(8-22)	13%	(7-21)
Enterobacter spp.	102	65%	(55-73)	12%	(9-23)	15%	(0-23)	9%	(5-16)
P. aeruginosa	86	61%	(50-70)	16%	(10-25)	16%	(10-25)	7%	(3-14)
Acinetobacter spp.	123	90%	(84-94)	7%	(1-7)	2%	(1-7)	1%	(0-4)
Salmonella spp.	103	90%	(83-95)	10%	(0-4)	0%	(0-4)	0%	(0-4)
N. meningitidis	508	94%	(92-96)	6%	(0-1)	0%	(0-1)	0%	(0-1)
H. influenzae	90	88%	(80-93)	6%	(2-12)	6%	(2-12)	1%	(0-6)
Other	409	86%	(82-89)	8%	(2-6)	4%	(2-6)	2%	(1-4)
Total	**6,629**	**87%**	**(86-88)**	**7%**	**(6-7)**	**4%**	**(3-4)**	**2%**	**(2-3)**

The 16 most frequently isolated pathogens accounted for 93% of BSI in children aged 1 month < 1 year and 89% of BSI in children aged 1-5 years. Children aged 1 month < 1 year had double the proportion of BSI occuring more than 10 days after hospital admission (12%) compared to the children aged 1-5 years (6%), but had similar proportions between day +2 to +9 (1 month < 1 year: 8%; 1-5 years: 7%).

6.4.2.1 Children aged 1 month < 1 year

A total of 46/52 (88%) genera were isolated between day -5 and +1 of admission, accounting for the majority of all pathogens (80%; 95% CI: 79-81%). *Enterococcus* spp. (62%, n=120/193) and *Enterobacter* spp. (79%, n=49/62) were the only pathogens more frequently isolated in hospital (two or more days after hospital admission) compared to on or before admission (day one or less after hospital admission), and the proportion of *Klebsiella* spp. and *P. aeruginosa* was split evenly before and after the HA2 threshold. Only *N. meningitidis, S. pneumoniae*, GAS, *Salmonella* spp. were not isolated beyond 11 days into hospital admission. GBS was isolated over a longer period of time during IP hospital stay, most likely due to late-onset infection.

Figure 6.7 demonstrates that all pathogens were isolated on or before admission to hospital. With the exception of the *a priori* pathogens, all other frequently reported pathogens were isolated consistently throughout hospital admission.

6.4.2.2 Children aged 1-5 years

A total of 56/63 (89%) genera were isolated between day -5 and +1 of admission, accounting for 87% (95% CI: 86-88%) of pathogens. All the pathogen species were more frequently isolated on day 1 or before of hospital admission; *Enterococcus* spp. was the only pathogen species that was most evenly distributed before and after the HA2 threshold (day 1 or less of hospital admission: 56%, n=99/176; 2 days or more: 44%, n=77/176). The range of isolation of the *a priori* CA pathogens after hospital admission was variable (*N. meningitidis:* 0-8 days; *Salmonella* spp.: 0-9 days; GAS: 0-11 days; *S. pneumoniae:* 0-17 days). With the exception of 1 isolate, GBS was not isolated any later than day 7 after hospital admission in the older age-group.

Figure 6.8 demonstrates that, with the exception of the *a priori* pathogens (*S. pneumoniae*, Group A streptococcus, *Salmonella* spp. and *N. meningitidis*) and GBS, all other frequently reported pathogens were isolated consistently throughout hospital admission.

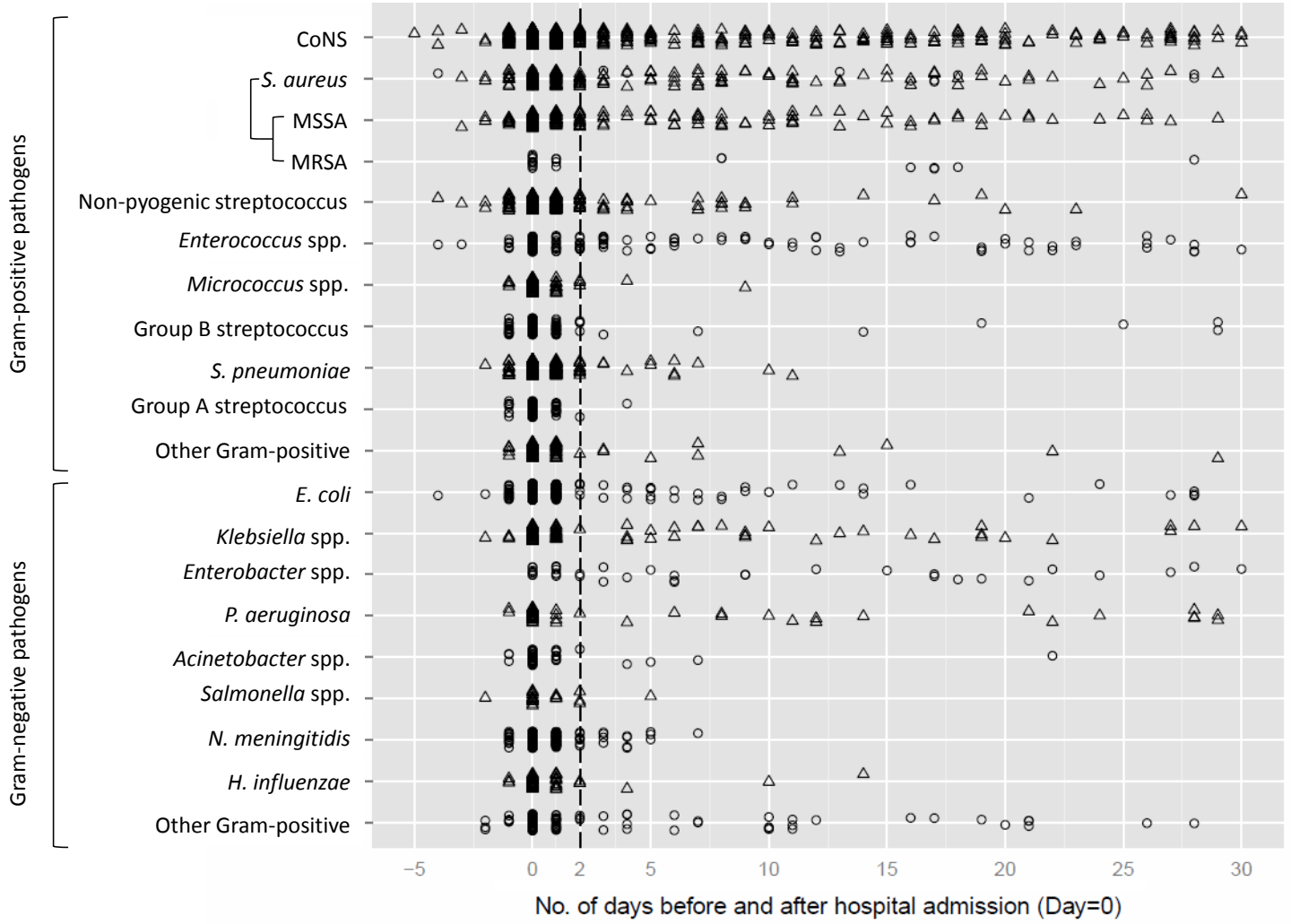


Figure 6.7. Distribution of BSI by pathogen and timing of the positive specimen in relation to hospital admission for children aged 1 month < 1 year; triangles and circles distinguish adjacent pathogens from each other; the dashed line indicates the HA2 threshold

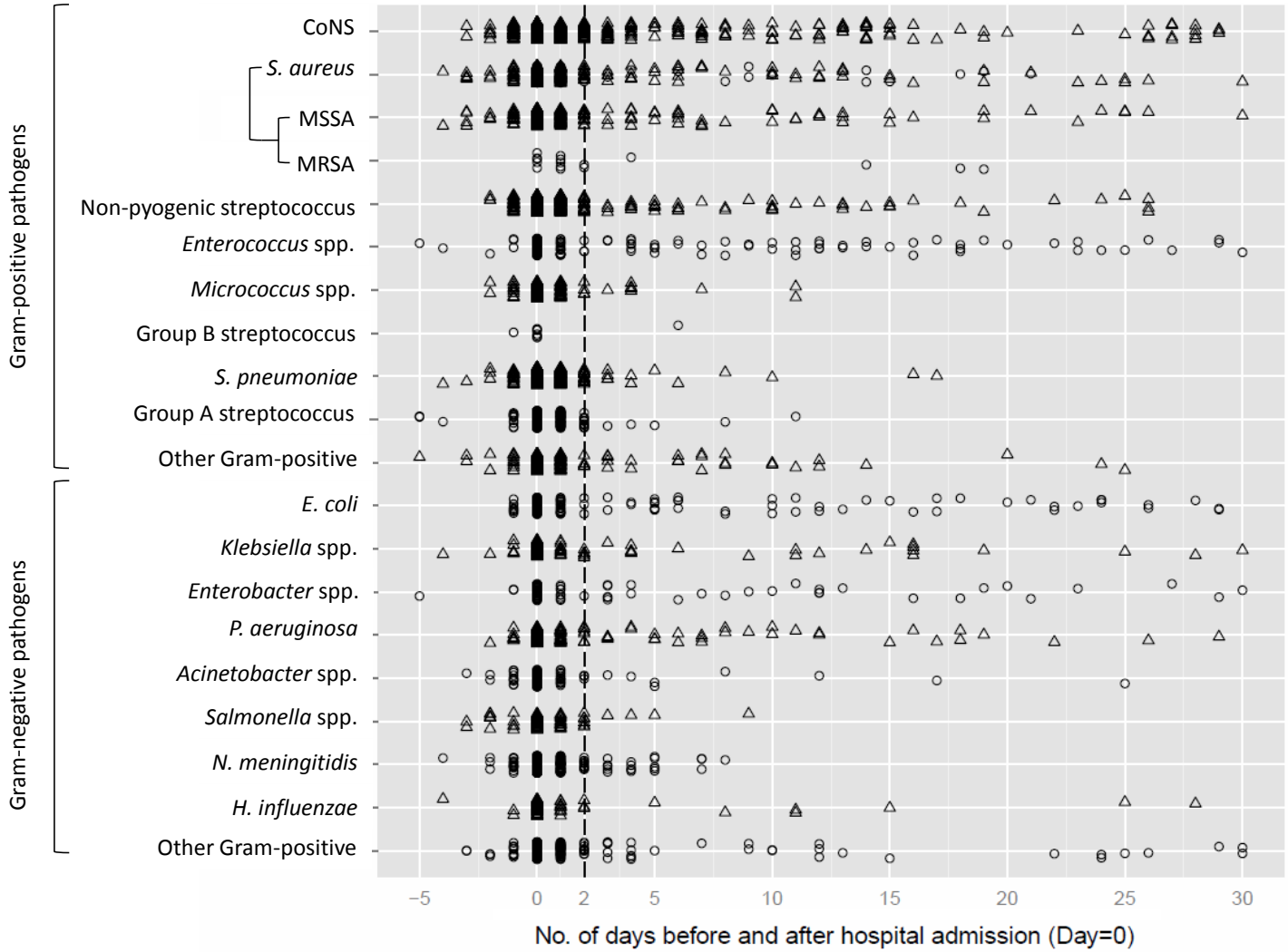


Figure 6.8. Distribution of BSI by pathogen and timing of the positive specimen in relation to hospital admission for children aged 1 – 5 years; triangles and circles distinguish adjacent pathogens from each other; the dashed line indicates the HA2 threshold

6.4.3 Section summary

- These results demonstrate the mixture of pathogens causing BSI in children on admission to hospital, where the timing of pathogens historically associated with hospital infection (e.g. *Enterococcus* spp., *Klebsiella* spp.) are also being isolated earlier before and within the first two days in hospital.
- A copy of the peer-reviewed manuscript on the timing of bacterial BSI in children in relation to hospital admission for one year of linked data (April 2009 - March 2010), published in the Epidemiology and Infection journal, can be found in Appendix 6.(187)

6.5 Pathogen distribution by predicted CA, HCA and HA BSI category

So far, I have looked at the clinical and demographic characteristic distribution by approach A, B and C for children with CA, HCA and HA BSI to gain an understanding of how the children are distributed between the three groups and whether the classifications are feasible given their characteristics. The purpose of this section is to reveal the distribution of pathogens according to each predicted category (CA, HCA or HA) and determine whether there is a clear divide (or not) of the causative organism by predicted category.

6.5.1 Methods

The distribution of pathogen species according to the predicted source of infection (CA, HCA or HA) were cross-tabulated and compared. As a sensitivity analysis for assessing the impact of bacteria commonly associated with contaminating blood cultures, although they do also cause real invasive infection, all monomicrobial BSIs that were caused by either CoNS, *Micrococcus* spp. or diphtheroids were removed.(188-190)

6.5.2 Results

Table 6.7 and Table 6.9 show the distribution of pathogens between CA, HCA and HA BSI for approaches A, B and C; the results for approaches A and B use the HA2 threshold.

6.5.2.1 Children aged 1 month < 1 year

The results in Table 6.7 and Table 6.8 depict the same numbers of pathogens, but the percentages are calculated differently. Table 6.7 displays the percentages of the predicted

proportion (calculated horizontally) that each pathogen is distributed to between CA, HCA and HA BSI for approach A, B and C. Table 6.8 displays the percentages of the proportional frequency of the pathogen species found within each CA, HCA and HA group (calculated vertically), which is more informative for clinical decision making.

CoNS were the most frequently reported pathogens (25-33%) in the CA, HCA and HA BSI categories irrespective of model approach, with the exception of *N. meningitidis* (38%) for CA BSI in approach A (Table 6.7). The majority (68-77%) of children with BSI caused by *S. pneumoniae*, GAS, *N. meningitidis* or *Salmonella* spp. in approach C were classified into group 1 (most resembling CA BSI) and a small proportion (6-12%) were classified into group 3 (most resembling HA BSI). GBS was spread between the CA, HCA and HA BSI categories.

Of the pathogens commonly associated with causing HA BSI (in the literature), approach A classified a very low proportion as being CA BSI: *Enterococcus* spp. (CA = 1%), *Klebsiella* spp. (CA = 2%), *Enterobacter* spp. (CA = 0%), *Acinetobacter* spp. (CA = 4%) and *P. aeruginosa* (CA = 4%). However, approaches B and C categorised a higher proportion of these pathogens as CA BSI: *Enterococcus* spp. (CA = 21-23%), *Klebsiella* spp. (CA = 23-25%), *Enterobacter* spp. (CA = 10-11%), *Acinetobacter* spp. (CA = 63-69%) and *P. aeruginosa* (CA = 21-27%). These pathogens, for which the majority are Gram-negative, are harder to treat because of innate or acquired antibiotic resistance within the hospital environment, particularly when the choice of BNF-C first-line antibiotics for CA BSI alone are directed more towards treating Gram-positive than Gram-negative infections.(56;191) This said, more than 50% of these pathogens were categorised as causing HA BSI for all three approaches.

Table 6.7. BSI causing pathogen breakdown by approach A, B and C for children aged 1 month < 1 year (percentage calculated by pathogen between CA, HCA and HA groups)

Pathogen	Approach A			Approach B			Approach C			Total
	CA %	HCA %	HA %	CA %	HCA/HA %	HA %	Group 1 (most like CA) %	Group 2 (most like HCA) %	Group 3 (most like HA) %	
Gram-positive										
CoNS	42 3%	888 72%	301 24%	787 64%	154 13%	290 24%	687 56%	208 17%	336 27%	1,231
Non-pyogenic streptococcus	27 6%	355 82%	53 12%	335 77%	51 12%	49 11%	299 69%	59 14%	77 18%	435
S. aureus	15 4%	248 63%	129 33%	184 47%	84 21%	124 32%	174 44%	65 17%	153 39%	392
MSSA	13 4%	215 64%	109 32%	159 47%	73 22%	105 31%	148 44%	58 17%	131 39%	337
MRSA	2 8%	13 52%	10 40%	11 44%	4 16%	10 40%	10 40%	6 24%	9 36%	25
*S. pneumoniae**	340 100%	0 0%	0 0%	340 100%	0 0%	0 0%	250 74%	50 15%	40 12%	340
Group B streptococcus	2 1%	197 92%	14 7%	146 69%	53 25%	14 7%	131 62%	16 8%	66 31%	213
Enterococcus spp.	1 1%	70 36%	122 63%	45 23%	28 15%	120 62%	40 21%	34 18%	119 62%	193
Group A streptococcus*	117 100%	0 0%	0 0%	117 100%	0 0%	0 0%	87 74%	21 18%	9 8%	117
Micrococcus spp.	2 2%	92 90%	8 8%	86 84%	9 9%	7 7%	70 69%	16 16%	16 16%	102
Other Gram-positive	2 2%	74 73%	26 25%	63 62%	16 16%	23 23%	52 51%	23 23%	27 26%	102
Gram-negative										
E. coli	9 2%	517 89%	56 10%	441 76%	87 15%	54 9%	410 70%	57 10%	115 20%	582
*N. meningitidis**	367 100%	0 0%	0 0%	367 100%	0 0%	0 0%	282 77%	64 17%	21 6%	367
Klebsiella spp.	2 2%	55 45%	64 53%	23 19%	40 33%	58 48%	25 21%	17 14%	79 65%	121
Enterobacter spp.	0%	11 18%	51 82%	6 10%	7 11%	49 79%	7 11%	10 16%	45 73%	62
Acinetobacter spp.	2 4%	42 78%	10 19%	37 69%	9 17%	8 15%	34 63%	5 9%	15 28%	54
H. influenzae	9 17%	36 69%	7 13%	39 75%	7 13%	6 12%	38 73%	6 12%	8 15%	52
P. aeruginosa	2 4%	23 48%	23 48%	10 21%	15 31%	23 48%	13 27%	6 13%	29 60%	48
Salmonella spp.*	25 100%	0 0%	0 0%	25 100%	0 0%	0 0%	17 68%	6 24%	2 8%	25
Other Gram-negative	16 7%	150 70%	48 22%	133 62%	34 16%	47 22%	131 61%	28 13%	55 26%	214
Total	980 21%	2,758 59%	912 20%	3,184 68%	594 13%	872 19%	2,747 59%	691 15%	1,212 26%	4,650

* *a priori* CA for Approaches A and B

Table 6.8 BSI causing pathogen breakdown by approach A, B and C for children aged 1 month < 1 year (percentage calculated for all pathogens within each individual CA, HCA and HA group)

Pathogen	Approach A			Approach B			Approach C			Total
	CA %	HCA %	HA %	CA %	HCA/HA %	HA %	Group 1 (most like CA) %	Group 2 (most like HCA) %	Group 3 (most like HA) %	
Gram-positive	**548 56%**	**1,924 70%**	**653 72%**	**2,103 66%**	**395 66%**	**627 72%**	**1,790 65%**	**492 71%**	**843 70%**	**3,125**
CoNS	42 4%	888 32%	301 33%	787 25%	154 26%	290 33%	687 25%	208 30%	336 28%	1,231
Non-pyogenic streptococcus	27 3%	355 13%	53 6%	335 11%	51 9%	49 6%	299 11%	59 9%	77 6%	435
S. aureus	15 2%	248 9%	129 14%	184 6%	84 14%	124 14%	174 6%	65 9%	153 13%	392
MSSA	13 1%	215 8%	109 12%	159 5%	73 12%	105 12%	148 5%	58 8%	131 11%	337
MRSA	2 0%	13 0%	10 1%	11 0%	4 1%	10 1%	10 0%	6 1%	9 1%	25
*S. pneumoniae**	340 35%	0 0%	0 0%	340 11%	0 0%	0 0%	250 9%	50 7%	40 3%	340
Group B streptococcus	2 0%	197 7%	14 2%	146 5%	53 9%	14 2%	131 5%	16 2%	66 5%	213
Enterococcus spp.	1 0%	70 3%	122 13%	45 1%	28 5%	120 14%	40 1%	34 5%	119 10%	193
Group A streptococcus*	117 12%	0 0%	0 0%	117 4%	0 0%	0 0%	87 3%	21 3%	9 1%	117
Micrococcus spp.	2 0%	92 3%	8 1%	86 3%	9 2%	7 1%	70 3%	16 2%	16 1%	102
Other Gram-positive	2 0%	74 3%	26 3%	63 2%	16 3%	23 3%	52 2%	23 3%	27 2%	102
Gram-negative	**432 44%**	**834 30%**	**259 28%**	**1,081 34%**	**199 34%**	**245 28%**	**957 35%**	**199 29%**	**369 30%**	**1,525**
E. coli	9 1%	517 19%	56 6%	441 14%	87 15%	54 6%	410 15%	57 8%	115 9%	582
*N. meningitidis**	367 37%	0 0%	0 0%	367 12%	0 0%	0 0%	282 10%	64 9%	21 2%	367
Klebsiella spp.	2 0%	55 2%	64 7%	23 1%	40 7%	58 7%	25 1%	17 2%	79 7%	121
Enterobacter spp.	0%	11 0%	51 6%	6 0%	7 1%	49 6%	7 0%	10 1%	45 4%	62
Acinetobacter spp.	2 0%	42 2%	10 1%	37 1%	9 2%	8 1%	34 1%	5 1%	15 1%	54
H. influenzae	9 1%	36 1%	7 1%	39 1%	7 1%	6 1%	38 1%	6 1%	8 1%	52
P. aeruginosa	2 0%	23 1%	23 3%	10 0%	15 3%	23 3%	13 0%	6 1%	29 2%	48
Salmonella spp.*	25 3%	0 0%	0 0%	25 1%	0 0%	0 0%	17 1%	6 1%	2 0%	25
Other Gram-negative	16 2%	150 5%	48 5%	133 4%	34 6%	47 5%	131 5%	28 4%	55 5%	214
Total	**980 100%**	**2,758 100%**	**912 100%**	**3,184 100%**	**594 100%**	**872 100%**	**2,747 100%**	**691 100%**	**1,212 100%**	**4,650**

No pathogen was categorised as 100% CA or HCA by any of the three approaches. *Enterobacter* spp. was the only pathogen that was classified as causing more HCA than CA BSI by all three approaches, and the four *a priori* pathogens were the only pathogens classified as causing more CA than HCA BSI.

6.5.2.2 Children aged 1-5 years

The results in Table 6.9 and Table 6.10 depict the same numbers of pathogens, but the percentages are calculated differently. Table 6.9 displays the percentages of the predicted proportion (calculated horizontally) that each pathogen is distributed to between CA, HCA and HA BSI for approach A, B and C. Table 6.10 displays the percentages of the proportional frequency of the pathogen species found within each CA, HCA and HA group (calculated vertically), which is more informative for clinical decision making.

CoNS were the most frequent pathogen (25-35%) in each CA, HCA and HA category irrespective of approach for children aged 1-5 years, with the exception of CA BSI in approach A, where *S. pneumoniae* (38%) was the most frequently reported pathogen (Table 6.9).

A high proportion (81-93%) of the pathogens commonly associated with CA BSI (*S. pneumoniae*, GAS, *N. meningitidis* and *Salmonella* spp.) in approach C were clustered in the group considered to most resemble CA (group 1), where only 2-5% were clustered in the HA BSI group and 5-15% in the HCA group (Table 6.10).

Of the pathogens commonly associated with HA BSI, approach A classified a very low proportion as being CA BSI: *Enterococcus* spp. (CA = 1%), *Klebsiella* spp. (CA = 0%), *Enterobacter* spp. (CA = 1%), *Acinetobacter* spp. (CA = 4%) and *P. aeruginosa* (CA = 0%). As with the younger age-group, approach B categorised a higher proportion of these pathogens as being CA than what was expected: *Enterococcus* spp. (CA = 35%), *Klebsiella* spp. (CA = 18%), *Enterobacter* spp. (CA = 27%), *P. aeruginosa* (CA = 26%).

As with the younger age-group, no pathogen was categorised as uniquely CA or HCA. Compared to the younger age-group, *P. aeruginosa* and *Klebsiella* spp., in addition to *Enterobacter* spp., were predicted to cause more HCA than CA BSI (>50%). The four *a priori* pathogens were predicted to predominantly cause more CA BSI than HCA BSI in all three approaches.

Table 6.9. BSI causing pathogen breakdown by approach A, B and C for children aged 1-5years (percentage calculated by pathogen between CA, HCA and HA groups)

Pathogen	Approach A			Approach B			Approach C			Total
	CA %	HCA %	HA %	CA %	HCA/HA %	HA %	Group 1 (most like CA) %	Group 2 (most like HCA) %	Group 3 (most like HA) %	
Gram-positive										
CoNS	76 4%	1,388 82%	228 13%	1,263 75%	208 12%	221 13%	1,110 66%	313 18%	269 16%	1,692
Non-pyogenic streptococcus	39 4%	749 86%	83 10%	700 80%	91 10%	80 9%	654 75%	131 15%	86 10%	871
*S. pneumoniae**	707 100%	0 0%	0 0%	707 100%	0 0%	0 0%	578 82%	104 15%	25 4%	707
S. aureus	25 4%	452 79%	94 16%	374 65%	105 18%	92 16%	338 59%	118 21%	115 20%	571
MSSA	19 4%	392 81%	71 15%	325 67%	88 18%	69 14%	292 61%	101 21%	89 18%	482
MRSA	0 0%	11 55%	9 45%	8 40%	3 15%	9 45%	8 40%	6 30%	6 30%	20
Group A streptococcus*	394 100%	0 0%	0 0%	394 100%	0 0%	0 0%	341 87%	35 9%	18 5%	394
Micrococcus spp.	7 3%	193 89%	16 7%	176 81%	25 12%	15 7%	167 77%	33 15%	16 7%	216
Enterococcus spp.	1 1%	98 56%	77 44%	64 36%	35 20%	77 44%	46 26%	51 29%	79 45%	176
Group B streptococcus	1 10%	7 70%	2 20%	8 80%	0 0%	2 20%	8 80%	0%	2 20%	10
Other Gram-positive	6 3%	179 83%	31 14%	138 64%	48 22%	30 14%	119 55%	45 21%	52 24%	216
Gram-negative										
*N. meningitidis**	508 100%	0 0%	0 0%	508 100%	0 0%	0 0%	473 93%	22 4%	13 3%	508
E. coli	2 1%	107 63%	61 36%	66 39%	47 28%	57 34%	59 35%	57 34%	54 32%	170
Acinetobacter spp.	7 6%	103 84%	13 11%	100 81%	11 9%	12 10%	86 70%	25 20%	12 10%	123
Salmonella spp.*	103 100%	0 0%	0 0%	103 100%	0 0%	0 0%	87 84%	15 15%	1 1%	103
Enterobacter spp.	2 2%	63 62%	37 36%	29 28%	37 36%	36 35%	17 17%	40 39%	45 44%	102
Klebsiella spp.	0 0%	58 60%	38 40%	17 18%	42 44%	37 39%	8 8%	44 46%	44 46%	96
H. influenzae	6 7%	73 81%	11 12%	70 78%	9 10%	11 12%	61 68%	19 21%	10 11%	90
P. aeruginosa	0 0%	50 58%	36 42%	22 26%	31 36%	33 38%	13 15%	29 34%	44 51%	86
Other Gram-negative	17 3%	415 83%	66 13%	335 67%	99 20%	64 13%	299 60%	121 24%	81 16%	498
Total	1,901 29%	3,935 59%	793 12%	5,074 77%	788 12%	767 12%	4,464 67%	1,202 18%	966 15%	6,629

* *a priori* CA for Approaches A and B

Table 6.10 BSI causing pathogen breakdown by approach A, B and C for children aged 1-5 years (percentage calculated for all pathogens within each individual CA, HCA and HA group)

Pathogen	Approach A			Approach B			Approach C			Total
	CA %	HCA %	HA %	CA %	HCA/HA %	HA %	Group 1 (most like CA) %	Group 2 (most like HCA) %	Group 3 (most like HA) %	
Gram-positive	**1,256 66%**	**3,066 78%**	**531 67%**	**3,824 75%**	**512 65%**	**517 67%**	**3,361 75%**	**830 69%**	**662 69%**	**4,853**
CoNS	76 4%	1,388 35%	228 29%	1,263 25%	208 26%	221 29%	1,110 25%	313 26%	269 28%	1,692
Non-pyogenic streptococcus	39 2%	749 19%	83 10%	700 14%	91 12%	80 10%	654 15%	131 11%	86 9%	871
*S. pneumoniae**	707 37%	0 0%	0 0%	707 14%	0 0%	0 0%	578 13%	104 9%	25 3%	707
S. aureus	25 1%	452 11%	94 12%	374 7%	105 13%	92 12%	338 8%	118 10%	115 12%	571
MSSA	19 1%	392 10%	71 9%	325 6%	88 11%	69 9%	292 7%	101 8%	89 9%	482
MRSA	0 0%	11 0%	9 1%	8 0%	3 0%	9 1%	8 0%	6 0%	6 1%	20
Group A streptococcus*	394 21%	0 0%	0 0%	394 8%	0 0%	0 0%	341 8%	35 3%	18 2%	394
Micrococcus spp.	7 0%	193 5%	16 2%	176 3%	25 3%	15 2%	167 4%	33 3%	16 2%	216
Enterococcus spp.	1 0%	98 2%	77 10%	64 1%	35 4%	77 10%	46 1%	51 4%	79 8%	176
Group B streptococcus	1 0%	7 0%	2 0%	8 0%	0 0%	2 0%	8 0%	0%	2 0%	10
Other Gram-positive	6 0%	179 5%	31 4%	138 3%	48 6%	30 4%	119 3%	45 4%	52 5%	216
Gram-negative	**645 34%**	**869 22%**	**262 33%**	**1,250 25%**	**276 35%**	**250 33%**	**1,103 25%**	**372 31%**	**304 31%**	**1,776**
*N. meningitidis**	508 27%	0 0%	0 0%	508 10%	0 0%	0 0%	473 11%	22 2%	13 1%	508
E. coli	2 0%	107 3%	61 8%	66 1%	47 6%	57 7%	59 1%	57 5%	54 6%	170
Acinetobacter spp.	7 0%	103 3%	13 2%	100 2%	11 1%	12 2%	86 2%	25 2%	12 1%	123
Salmonella spp.*	103 5%	0 0%	0 0%	103 2%	0 0%	0 0%	87 2%	15 1%	1 0%	103
Enterobacter spp.	2 0%	63 2%	37 5%	29 1%	37 5%	36 5%	17 0%	40 3%	45 5%	102
Klebsiella spp.	0 0%	58 1%	38 5%	17 0%	42 5%	37 5%	8 0%	44 4%	44 5%	96
H. influenzae	6 0%	73 2%	11 1%	70 1%	9 1%	11 1%	61 1%	19 2%	10 1%	90
P. aeruginosa	0 0%	50 1%	36 5%	22 0%	31 4%	33 4%	13 0%	29 2%	44 5%	86
Other Gram-negative	17 1%	415 11%	66 8%	335 7%	99 13%	64 8%	299 7%	121 10%	81 8%	498
Total	**1,901 100%**	**3,935 100%**	**793 100%**	**5,074 100%**	**788 100%**	**767 100%**	**4,464 100%**	**1,202 100%**	**966 100%**	**6,629**

Although there is increasing recognition of the evolving epidemiology of bacteria causing BSIs and some pathogens, previously thought of as causing HA BSI, may be isolated on admission to hospital from children with genuine CA BSI, it is difficult to accept such high proportions of CA BSI that approaches B and C report. These results have been challenged and discussed with microbiology and clinical colleagues around the idea that the true proportions lie somewhere between the results of approach A and approach B/C. More information about other forms of healthcare contact for these young patients, in addition to the IP data presented in this thesis, may help to further differentiate between children with different pathogens causing BSI.

6.5.2.3 Sensitivity analysis: removal of CoNS, *Micrococcus* spp. and diphtheroids

The total proportion of mono-microbial BSI removed from the dataset was 21% (1,294/6,172) in children aged 1 month < 1 year (CoNS=1,126; *Micrococcus* spp.=99; diphtheroids=69) and 42% (1,839/4,422) in children aged 1-5 years (CoNS=1,515; *Micrococcus* spp.=201; diphtheroids=123).

The results of the proportion of BSI categorised as CA, HCA or HA BSI after removing the mono-microbial BSIs from the models are shown in Table 6.11 and Table 6.12. There was very little change in the proportional ranking of CA, HCA and HA BSI in the younger age-group; however, the gap between the lower and upper threshold of CA (range: 40%) and HCA BSI (range: 38%) was narrower compared to the 48% range for both in the original approaches. The most change was seen in approach A, where the proportion of CA BSI increased from 21% to 34% (and there was an increase in the numbers of children with CA BSI from 941 to 1,066), and there was a 12% reduction in the proportion of HCA BSI; the proportion of HA BSI varied by 1%.

In the older age-group, the removal of the pathogens had the largest effect in approach A for children aged 1-5 years. The proportion of CA BSI increased from 30% to 56% of all BSI in that age-group, with more children being categorised into this group compared to the other two categories, where the proportion of HCA BSI decreased and the proportion of HA BSI stayed the same. In this age-group, all three approaches showed the majority of BSIs were attributed to the CA BSI group.

Table 6.11 Results of proportion of CA, HCA and HA BSI using the day 2 *defined* HA threshold when CoNS, *Micrococcus* spp. and diphtheroids were removed for children aged 1 month < 1 year

	Total	CA BSI	%	HCA BSI	%	HA BSI	%
approach A	3,121	1,066	34%	1,508	48%	547	18%
approach B	3,121	2,300	74%	297	10%	524	17%
approach C	3,121	1,965	63%	628	20%	528	17%

Table 6.12 Results of proportion of CA, HCA and HA BSI using the day 2 *defined* HA threshold when CoNS, *Micrococcus* spp. and diphtheroids were removed for children aged 1-5 years

	Total	CA BSI	%	HCA BSI	%	HA BSI	%
approach A	4,333	2,412	56%	1,453	34%	468	11%
approach B	4,333	3,499	81%	381	9%	453	10%
approach C	4,333	2,965	68%	771	18%	597	14%

6.5.3 Section summary

- The distribution of pathogen species for CA BSI approach A contained very few infections caused by the Gram-negative pathogens and *Enterococcus* spp. traditionally associated with HA BSI, which was the opposite in approaches B and C
- There was little clarity about pathogens causing CA or HCA BSI for the majority of the most frequently isolated pathogens in both age-groups, except for *N. meningitidis*, *S. pneumoniae*, GAS and *Salmonella* spp. which were considered to predominantly cause CA BSI
- *Enterobacter* spp. (in both age-groups), *Klebsiella* spp. and *P. aeruginosa* (in children aged 1-5 years) were the only pathogens to be classified as causing predominantly HCA BSI amongst the subgroup who were not classified as HA.
- The removal of CoNS, *Micrococcus* spp., and diphtheroids, which are often considered as contaminants of blood cultures, did not appreciably alter the predicted proportion of BSI in children.

6.6 Discussion

This chapter reports the first national estimates of the proportion of CA, HCA and HA BSI in children in England. There is a consistent predicted proportion of HA BSI in both children aged 1 month < 1 year (≤25%) and children aged 1-5 years (≤13%). However, there was a large overlap in the predicted proportion of CA and HCA BSI, illustrating that the clinical risk factors in this analysis alone are not sufficient to fully predict the proportion of CA and HCA BSI. There was some evidence that a few clinical predictors were useful in distinguishing between CA and

HCA BSI (e.g. indwelling devices in the last 1-30 day, more than two chronic conditions, and invasive procedures in the last 1-30 days before the BSI), but their observed prevalence is low.

This is also the first time that the national distribution of pathogens in relation to hospital admission has been investigated. The most striking observation is that there was a complete mix of pathogens isolated from children before or on admission to hospital, and particularly a proportion of HCA BSI that might be wrongly categorised as CA if using the historic 48-hour distinction between pathogens causing CA BSI or HA BSI observed in the past. The distribution of pathogens between the CA, HCA and HA BSI categories for the three approaches also show that the models are not accurate enough to differentiate fully between pathogens causing CA as opposed to HCA BSI. From a clinical decision making point of view, the distribution of pathogens within each CA, HCA and HA group highlights the proportion of different species causing BSI, where proportionally more Gram-negative pathogens causing BSIs in children with HCA or HA BSI than for CA BSI across all the approaches.

The results presented here emphasise the movement of bacterial species commonly associated with the hospital environment into the community. The modelling of a child's movement between hospitals and the community will become an increasingly important part of monitoring and modelling the evolving epidemiology of BSI, particularly in terms of not just the pathogen species but also the antimicrobial susceptibility and administered antibiotics.

Chapter 7. Discussion

7.1 Aim of the thesis

The aim of this thesis was to use demographic and clinical characteristics from national, administrative data to predict the relative proportion of CA, HCA and HA BSI in children in order to inform population based BSI prevention strategies.

7.1.1 Objectives

The three objectives of this thesis were:

1. To link microbiology surveillance and clinical data and evaluate the effect of match-weight choice on the final linked dataset
2. To develop statistical models to identify patient and clinical characteristics associated with CA, HCA and HA BSI in children
3. To predict the proportion of CA, HCA and HA BSI occurring in children in England and the distribution of causative pathogens in each group

The six previous chapters included in-depth discussions of the three objectives. This chapter aims to discuss the strengths and limitations of the study as a whole, and its clinical, public health and research implications. Conference presentations arising from the work in this thesis are described in Appendix 7.

7.2 Key findings

The key findings of this thesis were:

1. Using probabilistic linkage, I successfully linked 80% of all positive blood specimen records from children in England between 1st April 2007 and 31st March 2011 to clinical HES IP data. Of the unlinked specimen records, 49% were not eligible for inclusion based on country (e.g. Wales) or age (older children and adults potentially included due to lack of date of birth data). Of the other 51% unlinked specimen records, 43% had no NHS number, which is the strongest and most reliable patient identifier when linking. Alternative strategies for calculating match-weights did not appreciably improve the linkage rate.

2. Three statistical models were developed using the clinical and demographic characteristics identified in the linked dataset. However, few characteristics (more than two chronic conditions, indwelling devices, and hospital discharge in the month prior to the BSI) had high positive predictive value for pathogens causing HCA-BSI compared to CA-BSI. Moreover, there were few children with these characteristics.
3. Using three different statistical models, the prevalence of CA BSI ranged from 21-77% and the HCA BSI ranged from 12-60%. These large ranges likely reflect the relatively small number of children with predictive characteristics that differentiate CA from HCA BSI. The three models consistently attributed a low proportion of BSI to HA BSI, which accounted for a quarter or less of all children aged 1 month < 1 year, and a sixth or less of all children aged 1-5 years.
4. A total of 80% (Table 6.5) and 87% (Table 6.6) of all isolated pathogen species causing BSI in children aged 1 month < 1 year and 1-5 years respectively (including the 16 most frequently isolated bacterial species) were isolated prior to, or on day 0 or 1 of hospital admission, (i.e. before the 2-day threshold for distinguishing between CA and HA BSI). Bacterial species historically associated with causing HA BSI were more frequently isolated on or before hospital admission than during hospital admission in this study (except for *Enterococcus* spp. and *Enterobacter* spp. in children aged 1 month < 1 year). This suggests an important role of HCA BSI, or that HA type pathogens are circulating in the community, reflecting the shortening of average length of hospital stay and increasing healthcare contact.(192;193)

7.3 Strengths and limitations

7.3.1 Objective 1: Link microbiology surveillance and clinical data and evaluate the effect of match-weight choice on the final linked dataset

I created a linked dataset that is nationally representative of children with microbiologically confirmed BSI who were admitted to hospital in England. It is representative because clinicians who suspect a BSI would take a blood sample to determine the invasive organism and its antibiotic susceptibility to treat accordingly, and the child would be admitted to hospital due to the severity of the infection. I have also estimated the rate per 100,000 population by age-group to inform the overall impact that BSIs have on the paediatric population in England, demonstrating the higher burden of BSI in the younger age-group (1 month < 1 year) compared to the older children (1-5 years) in terms of the general child population. One

limitation that couldn't be measured was the proportion of children with a clinical diagnosis of BSI but a negative blood culture, potentially resulting from the administration of antibiotics before the blood sample was taken.

Between April 2007 and March 2011, the proportion of linked records increased by 8% (from 76.7% to 84.4%) as the patient identifier completion improved over time. Despite only 51-70% completion of the strongest identifier, NHS number, in LabBase2 compared to the 91-97% completion in HES (between 2007 and 2011), the probabilistic linkage method linked 25% more records (MW-Child total linked: 81%, n=17,141/21,074 OPIEs) than linkage using NHS number alone (total linked: 58%, n=12,212/21,074 OPIEs; NHS number deterministic linkage). A limitation of the linkage was the variable detail about the unlinked child records, where 92% of these were either reported from a UK laboratory external to England, or had no identifiable age, or had incomplete NHS number record. Where data was reported, there were no statistically significant differences by age or sex between the linked and unlinked LabBase2 records.

After the final dataset was chosen following the probabilistic linkage, I was able to determine linkage error within HES when comparing the results to the linked OPIEs (Figure 3.17). I identified a small proportion of multiple HESids that were linked to the same OPIE (i.e. indicating multiple children linking to one infection episode), but on closer inspection the multiple HESids corresponded to the same child. This has been found in a similar study looking at anonymised HES data.(194) However, I did not investigate internal linkage error within the HES data when the same HESid was assigned; I assumed that records with the same HESid were for the same child.

The calculation of match-weights and the linkage was possible because of my privileged access to both LabBase2 and HES data within the HPA (now PHE). The linkage methods discussed here are now being used for other applications within PHE, alongside other departments who are investigating the use of probabilistic linkage for other datasets.(195) Although the resulting differences in match-weights were minimal, they could be very different in other scenarios and therefore highlight the important consideration of how match-weights are calculated when linking new datasets.

7.3.2 Objective 2: Develop statistical methods to identify patient and clinical characteristics associated with CA, HCA and HA BSI in children

Three statistical approaches were used to estimate the proportion of BSI attributable to CA, HCA and HA based on clinical characteristics, as there was no established gold standard method. This is the first time that the contribution of CA, HCA and HA BSI in children in England has been examined at a national scale. The HES data allows for an accurate, retrospective search of previous IP contact, providing a rich profile of IP hospital records. I divided children into two age-groups because of the large variation in the number of BSIs in children aged 1 month < 1 year and 1-5 years and the varying clinical lookback period prior to each child's index case of BSI, due to practical age restrictions.

The linked dataset enabled the timing of the BSI detection in relation to day of hospital admission to be incorporated into the models, in addition to the added value of clinical characteristics to a microbiology dataset that is entirely lacking the clinical context. Clinical indicators, such as invasive surgical procedures and presence of chronic conditions, where reported in HES, were strongly associated with the risk of HA BSI. Additional information on indwelling devices inserted at the bedside and previous antibiotic treatment, which were not available in the dataset, may have strengthened the models.

In the models used in this thesis, the HA definition was adjusted along a time-threshold to overlap the commonly used 2-day (after hospital admission) threshold. A limitation in interpreting whether the 2-day threshold was still an appropriate time-measurement was biased due to blood sampling practice, where the majority of samples were taken on day 0 and day 1 of hospital admission. In considering the relationship of the timing of specimen date in relation to hospital admission, a recent Danish study in adults found no precise time threshold between CA, HCA and HA BSI.(21) The identification of BSIs later during a hospital admission (after day 1) may go undetected due to lower levels of sampling, or where sampling is performed after a course of antibiotics has begun that reduces the likelihood of a positive culture. Therefore, later detection of BSI in a hospital admission may be due to either the start of a true infection, or the re-emergence of an ongoing infection following unsuccessful antibiotic treatment.

A limitation for two of the statistical approaches (A and B) was assuming that four pathogens could only cause *a priori* CA BSI rather than HCA or HA BSI. Although based on evidence in the literature, in practice the time it would take for the bacteria to multiply between the onset of infection to being detectable in the blood, would vary depending on the bacteria and the child.

There was also no method to determine if all these BSI were truly CA BSI, which is why approach C was applied to test any other associations. A small number of children with *a priori* CA BSI had a catheter inserted shortly before their positive sample was taken, which according to European Centre for Disease Prevention and Control guidance may suggest that they had a HCA BSI, however, these four pathogens are not commonly associated with causing infection via indwelling devices and therefore the insertion was considered to be unrelated.(8;72;75;78;165) The timing in hours of hospital admission and when the blood specimen was taken was lacking from both data sources, which may have provided greater precision in the models in indicating, within a 24-hour period, how closely these two variables occurred in relation to one another. I have previously discussed the strengths and limitations of the probability thresholds and best subset framework selection in Chapter 5.

7.3.3 Objective 3: Predict the proportion of CA, HCA and HA BSI occurring in children and the distribution of causative pathogens in each group

My results show that at most, a quarter of all microbiology confirmed BSIs were estimated to be HA, although the timing is partly collinear with blood sampling predominantly occurring at the time of hospital admission and length of hospital stay. The remaining proportion of BSI ranged from 21-77% for CA and 60-12% for HCA BSI. Although the estimates were not consistent, the results showed that HCA BSI was estimated to be an important component in the epidemiology of BSI in children, as reflected in the limited published literature.(8;72)

The daily recording of pathogens causing BSI by day of admission shows unequivocally that i) 95% of pathogens classified as causing *a priori* CA BSI (*S. pneumoniae*, GAS, *N. meningitidis* and *Salmonella* spp.) were isolated close to the day of hospital admission, and ii) pathogens commonly associated with HA-BSI were also frequently isolated early on in children's hospital admissions. Overall, 80% (children aged 1 month < 1 year) and 87% (children aged 1-5 years) of pathogen species were isolated less than 2 days after hospital admission. This reflects the dilemma over the large range of pathogens in the first two days of hospital admission when considering empiric antibiotic therapy.

The putative source of where a BSI is acquired is important to target preventative strategies in order to reduce infection and to indicate antibiotic treatment to which the pathogens are likely to be sensitive. There were two potential reasons why the models failed to consistently discriminate between sources of infection:

i. Information on clinical risk factors was not sufficiently detailed to capture likely source of exposure, particularly with respect to the severity of underlying medical conditions or invasive procedures, and therefore insufficient in their ability to fully differentiate between CA, HCA and HA BSI. This may introduce unmeasured confounding and therefore more accurate recording of these risk factors is needed. Data on prior antibiotic treatment was also unavailable, which would be useful in indicating treatment for a bacterial infection prior to the BSI and additional information of previous healthcare contact.(9;23)

ii. Historically, pathogens that were circulating in separate environments (e.g. community, hospital) have now disseminated into the different populations, seen with the emergence of HCA BSI, where healthcare is expanding into the community and is becoming an extension of the hospital environment, leading to less separation between the different environments. Therefore the assumption that the type of patient environment (e.g. hospital, community) reflects the type of pathogen and hence the type of antibiotic susceptibility, no longer applies.

This thesis included children with BSI, denoted by a positive blood culture. Children with clinical diagnosis of a BSI and a negative culture result who would also have been treated empirically with antibiotics were not included, and were unquantifiable, because LabBase2 did not collect negative culture data, and so no information on the invasive pathogen would have been available.

I did not examine patterns of antibiotic resistance in recorded BSIs in the cohort because the antibiotic susceptibility was not consistently tested and/or reported by the laboratories. Therefore, I did not investigate antibiotic susceptibility within the cohort by CA, HCA and HA BSI category. I was also not able to accurately determine rates of BSI in the population, as there was no clear denominator population of children at risk within the hospital setting versus the community setting who did not acquire a BSI. In addition, the effects of different health interventions, such as the introduction of care bundles and antibiotic policies, were not adjusted for when analysing the data, however, this was not what this thesis was concerned with investigating and therefore was outside the remit of the work.

My study is based on the assumption that clinical characteristics and the timing of the BSI can help predict the source of the invasive pathogen, and therefore its antibiotic susceptibility. Although estimated predictions in my study are not precise or sufficiently detailed, they do suggest that the growing mix of children acquiring CA and HCA BSI from the community setting (as well as HA BSI in hospital), supported by the findings in the literature review.(8;61;62;65-

68;71) This is likely to increase as the mix of children becomes more heterogeneous in the future as healthcare provision outside of the hospital continues to diversify.

7.4 What this thesis adds

7.4.1 Implications for clinical practice

The results of this thesis demonstrate that the majority of BSIs in children in England are CA or HCA and are caused by a wide range of pathogens, including those usually associated with causing HA BSI. The treatment for the wide distribution of pathogens isolated before or on admission to hospital is not reflected in the current BNF-C guidance for pathogens, which currently considers only one group: CA BSI.(56) An implication of this work would be to utilise the results found in this thesis, enhanced with future work to incorporate antibiotic susceptibility data, to determine the distribution of pathogens and their antimicrobial susceptibility patterns in relation to timing of hospital admission. For clinicians, this would give a probability of confirmed BSI to inform future empiric antibiotic therapy guidelines and contribute towards work that has been taking place at the European level to investigate the appropriateness of empiric antibiotic regimens using pooled antimicrobial resistance laboratory data from select European hospitals.(196)

The implication that the poor performance of the clinical and time predictors used in the models to determine CA, HCA and HA BSI at a national level has highlighted the level of uncertainty in knowing what antibiotic treatment is most likely to be effective. Bedside tests would offer a more rapid method to identify the antimicrobial susceptibility of the pathogen, overcoming the 48-hour wait for a blood culture result, providing a quick decision regarding first line antibiotic treatment.(197)

7.4.2 Implications for Public Health

This thesis shows that children are presenting to hospital with BSIs caused by a wide range of pathogens. I have shown how to perform the data linkage and the results from my work can be used, going forward, to monitor the patterns of pathogens causing BSI to inform preventive BSI strategies. The limited clinical details in HES and lack of information on antibiotic treatment in HES could be potentially addressed through primary care data linkage, which captures a more comprehensive, although not perfect, longitudinal picture of patient care over time as the

majority of medical contact takes place in primary care, including antibiotic prescriptions.(198;199)

Due to a recent change in data collection and replacement of the LabBase2 system with an upgraded version (Second Generation Surveillance System; SGSS), pathogen and antimicrobial sensitivity for all positive tests performed in participating laboratories can now be reported and monitored centrally. In future work, consistently tested and reported antibiotic susceptibility results would provide nationally representative scientific evidence to further inform empiric therapy of children with BSI. The SGSS database also intends to collect negative blood culture information in the future, which will increase our understanding of children with clinical sepsis (identified in HES) but with a negative blood culture. To put the size of this into context, a regional, south-west London study investigating invasive paediatric bacterial infections found that only 2.6% (n=482/18,366) of all blood cultures taken during a 3-year study period (2009-2011) were positive and determined to be clinically significant.(163) Although not possible at a national level yet, additional linkage to hospital prescribing data would also affirm whether an antibiotic was prescribed during the hospital spell.

The results from this thesis and the linkage methodology, as well as the future work to incorporate antibiotic susceptibility data, is part of the work that contributes towards the UK Government's five-year (2013-2018) antimicrobial resistance strategy.(10) This work is also timely because the National Institute for Health and Care Excellence are currently developing new sepsis guidelines on behalf of NHS England, who have made sepsis a clinical priority to improve the outcome of patients in 2015 and 2016.(200) Most recently in May 2016, David Cameron announced an aim of halving the number of HCA BSI by 2020, with a focus on Gram-negative species.(201) This public health call highlights the balance between setting a simple, understandable ambition and the complexity of the attribution of BSI highlighted by this thesis.

7.4.3 Implications for Research

This thesis has demonstrated the value of enhancing individual datasets by linking them together to produce information that would otherwise be unobtainable, both at an event specific time-point (the BSI) and from a longitudinal perspective (of medical records for each child prior to the BSI). One next step would be to focus on condition-specific, prospective cohorts, such as children with cancer or children with indwelling catheters, to determine the

risks of BSI and antibiotic resistance. This approach could provide more specific indicators for prevention strategies in these high risk groups.

Another step would be to link the new SGSS antimicrobial resistance data to the clinical HES data. This would enhance the understanding of the bacterial species distribution in relation to timing of hospital admission by adding information on resistance patterns of the bacteria isolated from children already in hospital compared to bacteria isolated from children presenting to hospital from the community.

Thirdly, the extension of linkage to other datasets, such as GP, A&E, OP and/or prescribing data would contribute towards further evaluation of the data quality, and would help to better understand previous healthcare contact and prior administration of antibiotics, indicating previous infection.

The results from this thesis inform the epidemiology of BSI in children but do not directly impact on antibiotic treatment choices. Given the limitations of the data's capabilities in distinguishing between CA, HCA and HA BSI and therefore guiding empiric therapy, more direct and practical solutions are needed to inform the antibiotic treatment, such as point-of-care (bedside) tests. Bedside tests for BSI detection are currently being developed and would provide an efficient testing method to tailor antibiotic treatment to each patient, avoiding time wastage, improving patient outcome and reducing the exposure to unnecessary antibiotics.(197) Results of such bedside tests could eventually be monitored nationally.

This PhD demonstrates the proof of principal for the internal linkage of HES and LabBase2. This work highlights the value and importance of combining these datasets. The steps that were taken to access the data took over a year and highlights the need for streamlining the data linkage processes without compromising data security and information governance; this will enhance research capabilities in academia and public health. This thesis has already been used as the foundation for developments of further work in the field of healthcare-associated infection epidemiology. The linkage methods used have contributed towards the literature that has previously been described as lacking guidance on methods of data linkage.(202) This thesis was also one of the key drivers in gaining access to identifiable HES data for the HCAI & AMR department for a selection of departmental public health and surveillance projects. Subsequently, there has been a central effort from multiple PHE departments to develop the permissions and deliver a PHE-wide agreement with the HSCIC for access to HES data for a much wider range of infectious disease surveillance strategies. This has already made a

considerable contribution to outbreak responses, case finding and long-term public health epidemiology projects.

7.5 Closing remarks

This thesis demonstrates the value and power of using established administrative datasets to enhance the epidemiology of BSI in children. I described and summarised the complex and dynamic process of categorising and interpreting infection with the aim of reducing and appropriately treating BSI. Although this thesis highlights the limitations of the clinical characteristics captured in HES to differentiate between CA or HCA BSI in the models, the strength is that there is now national evidence to show that the majority of BSIs occur in children presenting to hospital from the community caused by a wide range of pathogens, of which HCA BSI plays an important role. The focus of further surveillance should be on better understanding exposure to infection in the community setting and investigating the antimicrobial susceptibility of these pathogens.

References

(1) Ladhani S, Pebody RG, Ramsay ME, Lamagni TL, Johnson AP, Sharland M. Continuing impact of infectious diseases on childhood deaths in England and Wales, 2003-2005. *Pediatr Infect Dis J* 2010;29(4):310-3.

(2) Livermore DM, Wain J. Revolutionising bacteriology to improve treatment outcomes and antibiotic stewardship. *Infect Chemother* 2013;45(1):1-10.

(3) MacFadden DR, Leis JA, Mubareka S, Daneman N. The opening and closing of empiric windows: the impact of rapid microbiologic diagnostics. *Clin Infect Dis* 2014;59(8):1199-200.

(4) Weinstein MP, Reller LB, Murphy JR, Lichtenstein KA. The clinical significance of positive blood cultures: a comprehensive analysis of 500 episodes of bacteremia and fungemia in adults. I. Laboratory and epidemiologic observations. *Rev Infect Dis* 1983;5(1):35-53.

(5) Garner JS, Jarvis WR, Emori TG, Horan TC, Hughes JM. CDC definitions for nosocomial infections, 1988. *Am J Infect Control* 1988;16(3):128-40.

(6) Garner JS, Jarvis WR, Emori TG, Horan TC, Hughes JM. CDC definitions for nosocomial infections. In: Olmsted RN, editor. *APIC Infection control and applied epidemiology: Principles and practice*. St. Louis: Mosby; 1996. p. A1-A20.

(7) Gray JW. A 7-year study of bloodstream infections in an English children's hospital. *Eur J Pediatr* 2004;163(9):530-5.

(8) Laupland KB, Gregson DB, Vanderkooi OG, Ross T, Kellner JD. The changing burden of pediatric bloodstream infections in Calgary, Canada, 2000-2006. *Pediatr Infect Dis J* 2009;28(2):114-7.

(9) Friedman ND, Kaye KS, Stout JE, McGarry SA, Trivette SL, Briggs JP, et al. Health care--associated bloodstream infections in adults: a reason to change the accepted definition of community-acquired infections. *Ann Intern Med* 2002;137(10):791-7.

(10) Department of Health, Department for Environment Food & Rural Affairs. UK Five Year Antimicrobial Resistance Strategy 2013 to 2018 [online]. DH: London; 2013 [accessed 16 June 2016]. Available from:

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/244058/20130902_UK_5_year_AMR_strategy.pdf

(11) Sharland M. The use of antibacterials in children: a report of the Specialist Advisory Committee on Antimicrobial Resistance (SACAR) Paediatric Subgroup. *J Antimicrob Chemother* 2007;60 Suppl 1:i15-i26.

(12) Roberts GJ. Dentists are innocent! "Everyday" bacteremia is the real culprit: a review and assessment of the evidence that dental surgical procedures are a principal cause of bacterial endocarditis in children. *Pediatr Cardiol* 1999;20(5):317-25.

(13) Dellinger RP, Levy MM, Rhodes A, Annane D, Gerlach H, Opal SM, et al. Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock: 2012. *Crit Care Med* 2013;41(2):580-637.

(14) Lesch JE. Prontosil. *The first miracle drugs: how the sulfa drugs transformed medicine*. Oxford: Oxford University Press; 2007. p. 51.

(15) Bates DW, Lee TH. Rapid classification of positive blood cultures. Prospective validation of a multivariate algorithm. *JAMA* 1992;267(14):1962-6.

(16) Weinstein MP. Blood culture contamination: persisting problems and partial progress. *J Clin Microbiol* 2003;41(6):2275-8.

(17) Weinstein MP, Towns ML, Quartey SM, Mirrett S, Reimer LG, Parmigiani G, et al. The clinical significance of positive blood cultures in the 1990s: a prospective comprehensive evaluation of the microbiology, epidemiology, and outcome of bacteremia and fungemia in adults. *Clin Infect Dis* 1997;24(4):584-602.

(18) Stefani S. Diagnostic techniques in bloodstream infections: where are we going? *Int J Antimicrob Agents* 2009;34 Suppl 4:S9-12.

(19) Hall KK, Lyman JA. Updated review of blood culture contamination. *Clin Microbiol Rev* 2006;19(4):788-802.

(20) Public Health England. Annual Epidemiological Commentary: Mandatory MRSA, MSSA and E. coli bacteraemia and C. difficile infection data, 2013/14 [online]. PHE: London; 2014 Jul [accessed 16 June 2016]. Available from:

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/330529/HCAI_mandatory_surveillance_annual_epidemiological_commentary_2013_14.pdf.

(21) Gradel KO, Nielsen SL, Pedersen C, Knudsen JD, Ostergaard C, Arpi M, et al. No specific time window distinguishes between community-, healthcare-, and hospital-acquired bacteremia, but they are prognostically robust. *Infect Control Hosp Epidemiol* 2014;35(12):1474-82.

(22) Leibovici L, Schonheyder H, Pitlik SD, Samra Z, Moller JK. Bacteraemia caused by hospital-type micro-organisms during hospital stay. *J Hosp Infect* 2000;44(1):31-6.

(23) Siegman-Igra Y, Fourer B, Orni-Wasserlauf R, Golan Y, Noy A, Schwartz D, et al. Reappraisal of community-acquired bacteremia: a proposal of a new classification for the spectrum of acquisition of bacteremia. *Clin Infect Dis* 2002;34(11):1431-9.

(24) Stephens DS. Neisseria meningitidis. *Infect Control* 1985;6(1):37-40.

(25) Sood S, Malhotra M, Das BK, Kapil A. Enterococcal infections & antimicrobial resistance. *Indian J Med Res* 2008;128(2):111-21.

(26) Flaherty JP, Weinstein RA. Nosocomial infection caused by antibiotic-resistant organisms in the intensive-care unit. *Infect Control Hosp Epidemiol* 1996;17(4):236-48.

(27) Struelens MJ. The epidemiology of antimicrobial resistance in hospital acquired infections: problems and possible solutions. *BMJ* 1998;317(7159):652-4.

(28) Billett J, Majeed A, Gatzoulis M, Cowie M. Trends in hospital admissions, in-hospital case fatality and population mortality from congenital heart disease in England, 1994 to 2004. *Heart* 2008;94(3):342-8.

(29) Barr DA, Semple L, Seaton RA. Outpatient parenteral antimicrobial therapy (OPAT) in a teaching hospital-based practice: a retrospective cohort study describing experience and evolution over 10 years. *Int J Antimicrob Agents* 2012;39(5):407-13.

(30) Department of Health. NHS at Home: Community Children's Nursing services [online]. DH: London; 2011 [accessed 15 October 2016]. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/215708/dh_124900.pdf

(31) Action for Sick Children. History: Action for Sick Children [online]. Stockport; 2016 [accessed 15 October 2016]. Available from: http://www.actionforsickchildren.org.uk/our-history

(32) Great Ormond Street Hospital Charity. Great Ormond Street Hospital (GOSH) takes palliative care to the family home [online]. London; 2016 [accessed 15 October 2016]. Available from: http://www.gosh.org/great-ormond-street-hospital-gosh-takes-palliative-care-family-home

(33) Department of Health. A simple guide to payment by results [online]. DH: London; 2012 [accessed 15 October 2016]. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/213150/PbR-Simple-Guide-FINAL.pdf

(34) Tatman MA, Woodroffe C. Paediatric home care in the UK. *Arch Dis Child* 1993;69(6):677-80.

(35) While AE, Dyson L. Characteristics of paediatric home care provision: the two dominant models in England. *Child Care Health Dev* 2000;26(4):263-76.

(36) Whiting M. Royal College of Nursing Directory of Community Children's Nursing Services. London; 2007.

(37) Gastmeier P. Healthcare-associated versus community-acquired infections: a new challenge for science and society. *Int J Med Microbiol* 2010;300(6):342-5.

(38) Cooke FJ, Brown NM. Community-associated methicillin-resistant Staphylococcus aureus infections. *Br Med Bull* 2010;94:215-27.

(39) Gonzalez BE, Martinez-Aguilar G, Hulten KG, Hammerman WA, Coss-Bu J, Avalos-Mishaan A, et al. Severe Staphylococcal sepsis in adolescents in the era of community-acquired methicillin-resistant Staphylococcus aureus. *Pediatrics* 2005;115(3):642-8.

(40) Lesens O, Hansmann Y, Brannigan E, Hopkins S, Meyer P, O'Connel B, et al. Healthcare-associated Staphylococcus aureus bacteremia and the risk for methicillin resistance: is the Centers for Disease Control and Prevention definition for community-acquired bacteremia still appropriate? *Infect Control Hosp Epidemiol* 2005;26(2):204-9.

(41) Shorr AF, Tabak YP, Killian AD, Gupta V, Liu LZ, Kollef MH. Healthcare-associated bloodstream infection: A distinct entity? Insights from a large U.S. database. *Crit Care Med* 2006;34(10):2588-95.

(42) Valles J, Calbo E, Anoro E, Fontanals D, Xercavins M, Espejo E, et al. Bloodstream infections in adults: importance of healthcare-associated infections. *J Infect* 2008;56(1):27-34.

(43) Wunderink RG. Healthcare-associated bacteremia: Stirring the mud. *Crit Care Med* 2006;34(10):2685-6.

(44) Henderson KL, Johnson AP, Muller-Pebody B, Charlett A, Gilbert R, Sharland M. The changing aetiology of paediatric bacteraemia in England and Wales, 1998-2007. *J Med Microbiol* 2010;59(Pt 2):213-9.

(45) Henderson KL, Muller-Pebody B, Blackburn RM, Johnson AP. Reduction in erythromycin resistance in invasive pneumococci from young children in England and Wales. *J Antimicrob Chemother* 2010;65(2):369-70.

(46) Watson RS, Carcillo JA, Linde-Zwirble WT, Clermont G, Lidicker J, Angus DC. The epidemiology of severe sepsis in children in the United States. *Am J Respir Crit Care Med* 2003;167(5):695-701.

(47) Skogberg K, Lyytikainen O, Ruutu P, Ollgren J, Nuorti JP. Increase in bloodstream infections in Finland, 1995-2002. *Epidemiol Infect* 2008;136(1):108-14.

(48) Luthander J, Bennet R, Giske CG, Nilsson A, Eriksson M. Age and risk factors influence the microbial aetiology of bloodstream infection in children. *Acta Paediatr* 2013;102(2):182-6.

(49) Balmer P, Borrow R, Miller E. Impact of meningococcal C conjugate vaccine in the UK. *J Med Microbiol* 2002;51(9):717-22.

(50) Ladhani S, Slack MP, Heath PT, von GA, Chandra M, Ramsay ME. Invasive Haemophilus influenzae Disease, Europe, 1996-2006. *Emerg Infect Dis* 2010;16(3):455-63.

(51) National Health Service. The NHS vaccination schedule [online]. NHS; London 2014 [accessed 16 June 1016]. Available from:

http://www.nhs.uk/conditions/vaccinations/pages/vaccination-schedule-age-checklist.aspx

(52) National Health Service. Haemophilus influenzae type b (Hib) [online]. NHS; London 2014 [accessed 16 June 1016]. Available from: http://www.nhs.uk/conditions/hib/Pages/Introduction.aspx

(53) Department of Health. Meningococcal: meningitis and septicaemia notifiable (Chapter 22) [online]. In: *The Green Book*. London: DH; 2015 [accessed 16 June 2016]. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/462629/2904512_Green_Book_Chapter_22_v6_0W.pdf

(54) Department of Health. Pneumococcal: pneumococcal meningitis notifiable (Chapter 25) [online]. In: *The Green Book*. London: DH; 2013. [accessed 16 June 2016]. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/263318/Green-Book-Chapter-25-v5_2.pdf

(55) Miller E, Andrews NJ, Waight PA, Slack MP, George RC. Herd immunity and serotype replacement 4 years after seven-valent pneumococcal conjugate vaccination in England and Wales: an observational cohort study. *Lancet Infect Dis* 2011;11(10):760-8.

(56) Paediatric Formulary Committee. *BNF for Children 2014-2015*. London, BMJ Group: Pharmaceutical Press and RCPCH Publications; 2015.

(57) Blackburn RM, Muller-Pebody B, Planche T, Johnson A, Hopkins S, Sharland M, et al. Neonatal sepsis--many blood samples, few positive cultures: implications for improving antibiotic prescribing. *Arch Dis Child Fetal Neonatal Ed* 2012;97(6):F487-F488.

(58) Health Protection Agency. English National Point Prevalence Survey on Healthcare-associated Infections and Antimicrobial Use, 2011: Preliminary data [online]. HPA: London; 2011. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/331871/English_National_Point_Prevalence_Survey_on_Healthcare_associated_Infections_and_Antimicrobial_Use_2011.pdf

(59) Okike IO, Johnson AP, Henderson KL, Blackburn RM, Muller-Pebody B, Ladhani SN, et al. Incidence, Aetiology and Outcome of Bacterial Meningitis in Infants Aged <90 days in the UK and Republic of Ireland: prospective, enhanced, national population-based surveillance. *Clin Infect Dis* 2014:59(10):e150-7.

(60) The World Bank Group. Country classifications: country and lending groups [online]. The World Bank Group: Washington, DC; 2016 [accessed 16 June 2016]. Available from: http://data.worldbank.org/about/country-classifications/country-and-lending-groups

(61) Hounsom L, Grayson K, Melzer M. Mortality and associated risk factors in consecutive patients admitted to a UK NHS trust with community acquired bacteraemia. *Postgrad Med J* 2011;87(1033):757-62.

(62) Bloomfield M, Briant R. Bacteraemia at Gisborne hospital. *N Z Med J* 2008;121(1271):125-9.

(63) Ang L, Laskar R, Gray JW. A point prevalence study of infection and antimicrobial use at a UK children's hospital. *J Hosp Infect* 2008;68(4):372-4.

(64) Nicoletti G, Schito G, Fadda G, Boros S, Nicolosi D, Marchese A, et al. Bacterial isolates from severe infections and their antibiotic susceptibility patterns in Italy: a nationwide study in the hospital setting. *J Chemother* 2006;18(6):589-602.

(65) Raymond NJ, Blackmore TK, Humble MW, Jones MR. Bloodstream infections in a secondary and tertiary care hospital setting. *Intern Med J* 2006;36(12):765-72.

(66) Douglas MW, Lum G, Roy J, Fisher DA, Anstey NM, Currie BJ. Epidemiology of community-acquired and nosocomial bloodstream infections in tropical Australia: a 12-month prospective study. *Trop Med Int Health* 2004;9(7):795-804.

(67) Wyllie DH, Walker AS, Peto TE, Crook DW. Hospital exposure in a UK population, and its association with bacteraemia. *J Hosp Infect* 2007;67(4):301-7.

(68) Marchaim D, Zaidenstein R, Lazarovitch T, Karpuch Y, Ziv T, Weinberger M. Epidemiology of bacteremia episodes in a single center: increase in Gram-negative isolates, antibiotics resistance, and patient age. *Eur J Clin Microbiol Infect Dis* 2008;27(11):1045-51.

(69) Luzzaro F, Vigano EF, Fossati D, Grossi A, Sala A, Sturla C, et al. Prevalence and drug susceptibility of pathogens causing bloodstream infections in northern Italy: a two-year study in 16 hospitals. *Eur J Clin Microbiol Infect Dis* 2002;21(12):849-55.

(70) Raveh D, Levy Y, Schlesinger Y, Greenberg A, Rudensky B, Yinnon AM. Longitudinal surveillance of antibiotic use in the hospital. *QJM* 2001;94(3):141-52.

(71) Khan FY, Elshafie SS, Almaslamani M, Abu-Khattab M, El Hiday AH, Errayes M, et al. Epidemiology of bacteraemia in Hamad general hospital, Qatar: a one year hospital-based study. *Travel Med Infect Dis* 2010;8(6):377-87.

(72) Gray J, Gossain S, Morris K. Three-year survey of bacteremia and fungemia in a pediatric intensive care unit. *Pediatr Infect Dis J* 2001;20(4):416-21.

(73) Jensen AG, Kirstein A, Jensen I, Scheibel J, Espersen F. A 6-month prospective study of hospital-acquired bacteremia in Copenhagen county. *Scand J Infect Dis* 1996;28(6):601-8.

(74) McGregor AR, Collignon PJ. Bacteraemia and fungaemia in an Australian general hospital--associations and outcomes. *Med J Aust* 1993;158(10):671-4.

(75) Ashkenazi S, Leibovici L, Churi C, Drucker M, Konisberger H, Samra Z. Childhood bacteremia in Israel: causes, age relation, predisposing factors and source. *Isr J Med* Sci 1994;30(8):610-6.

(76) Gomez J, Banos V, Ruiz J, Herrero F, Perez M, Pretel L, et al. Clinical significance of anaerobic bacteremias in a general hospital. A prospective study from 1988 to 1992. *Clin Investig* 1993;71(8):595-9.

(77) Iversen G, Scheel O. Epidemiological aspects of community-acquired and hospital-acquired bacteremia in northern Norway. *Scand J Infect Dis* 1993;25(4):465-70.

(78) Cheng AF, Fok TF, Duthie R, French GL. A five-year prospective study of septicaemia in hospitalized children in Hong Kong. *J Trop Med Hyg* 1991;94(5):295-303.

(79) McLaws ML, Gold J, King K, Irwig LM, Berry G. The prevalence of nosocomial and community-acquired infections in Australian hospitals. *Med J Aust* 1988;149(11-12):582-90.

(80) Eykyn SJ, Gransden WR, Phillips I. The causative organisms of septicaemia and their epidemiology. *J Antimicrob Chemother* 1990;25 Suppl C:41-58.

(81) Kempker RR, Farley MM, Ladson JL, Satola S, Ray SM. Association of methicillin-resistant Staphylococcus aureus (MRSA) USA300 genotype with mortality in MRSA bacteremia. *J Infect* 2010;61(5):372-81.

(82) Kallen AJ, Mu Y, Bulens S, Reingold A, Petit S, Gershman K, et al. Health care-associated invasive MRSA infections, 2005-2008. *JAMA* 2010;304(6):641-8.

(83) Centers for Disease Control and Prevention. Bloodstream Infection. In: *National nosocomial infections study site definitions manual*. Atlanta, GA: CDC; 1972. p. 82-4.

(84) Glynn A, Ward V, Wilson J, Charlett A, Cookson BD, Taylor L, *et al*. Hospital-acquired infection: surveillance, policies and practice. London: PHLS; 1997.

(85) Gransden WR, Eykyn SJ, Phillips I, Rowe B. Bacteremia due to Escherichia coli: a study of 861 episodes. *Rev Infect Dis* 1990;12(6):1008-18.

(86) Klevens RM, Morrison MA, Fridkin SK, Reingold A, Petit S, Gershman K, et al. Community-associated methicillin-resistant Staphylococcus aureus and healthcare risk factors. *Emerg Infect Dis* 2006;12(12):1991-3.

(87) Cohen AL, Calfee D, Fridkin SK, Huang SS, Jernigan JA, Lautenbach E, et al. Recommendations for metrics for multidrug-resistant organisms in healthcare settings: SHEA/HICPAC Position paper. *Infect Control Hosp Epidemiol* 2008;29(10):901-13.

(88) Horan TC, Gaynes RP, Martone WJ, Jarvis WR, Emori TG. CDC definitions of nosocomial surgical site infections, 1992: a modification of CDC definitions of surgical wound infections. *Infect Control Hosp Epidemiol* 1992;13(10):606-8.

(89) Gaynes RP, Horan TC. National Nosocomial Infections Surveillance Systems: Surveillance of nosocomial infection. In: Mayhall CG, editor. *Hospital Epidemiology and Infection Control*. Baltimore: Williams and Wilkins; 1996. p. 1017-31.

(90) Department of Health. Technical Guidance for the 2012/13 Operating Framework [online]. DH: London; 2011 [accessed 16 June 2016]. Available from: http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_132045.pdf

(91) Horan TC, Andrus M, Dudeck MA. CDC/NHSN surveillance definition of health care-associated infection and criteria for specific types of infections in the acute care setting. *Am J Infect Control* 2008;36(5):309-32.

(92) Salgado CD, Farr BM, Calfee DP. Community-acquired methicillin-resistant Staphylococcus aureus: a meta-analysis of prevalence and risk factors. *Clin Infect Dis* 2003;36(2):131-9.

(93) Diekema DJ, Beekmann SE, Chapin KC, Morel KA, Munson E, Doern GV. Epidemiology and outcome of nosocomial and community-onset bloodstream infection. *J Clin Microbiol* 2003;41(8):3655-60.

(94) Tacconelli E, Venkataraman L, De Girolami PC, DAgata EM. Methicillin-resistant Staphylococcus aureus bacteraemia diagnosed at hospital admission: distinguishing between community-acquired versus healthcare-associated strains. *J Antimicrob Chemother* 2004;53(3):474-9.

(95) Hota B. Contamination, disinfection, and cross-colonization: are hospital surfaces reservoirs for nosocomial infection? *Clin Infect Dis* 2004;39(8):1182-9.

(96) Miller R, Esmail H, Peto T, Walker S, Crook D, Wyllie D. Is MRSA admission bacteraemia community-acquired? A case control study. *J Infect* 2008;56(3):163-70.

(97) Vanderkooi OG, Gregson DB, Kellner JD, Laupland KB. Staphylococcus aureus bloodstream infections in children: A population-based assessment. *Paediatr Child Health* 2011;16(5):276-80.

(98) Pearson A, Chronias A, Murray M. Voluntary and mandatory surveillance for methicillin-resistant Staphylococcus aureus (MRSA) and methicillin-susceptible S. aureus (MSSA) bacteraemia in England. *J Antimicrob Chemother* 2009;64 Suppl 1:i11-i17.

(99) Health Protection Scotland. ECOSS (The Electronic Communication of Surveillance in Scotland). HPS: Glasgow;2016 [accessed 16 June 2016]. Available from: http://www.hps.scot.nhs.uk/surveillance/SystemsDetail.aspx?id=248

(100) Freeman R, Charlett A, Hopkins S, O'Connell AM, Andrews N, Freed J, et al. Evaluation of a national microbiological surveillance system to inform automated outbreak detection. *J Infect* 2013;67(5):378-84.

(101) Health Protection Agency. New LabBase statistical deduplication specification [internal document only]. HPA: London; 2009

(102) Health Protection Agency. Laboratory reporting to the Health Protection Agency: Guide for diagnostic laboratories [online]. HPA: London; 2012 [accessed 16 June 2016]. Available from: http://phenet.phe.gov.uk/Resources/laboratory-reporting-surveillance/LabBase/Documents/Guidelines%20for%20laboratories%20reporting%20to%20PHE.pdf

(103) Public Health England. Escherichia coli (E. coli): guidance, data and analysis [online]. PHE: London; 2015 [accessed 16 June 2016]. Available from: https://www.gov.uk/government/collections/escherichia-coli-e-coli-guidance-data-and-analysis

(104) Health Protection Agency. Voluntary reporting of Staphylococcus aureus bacteraemia in England, Wales and Northern Ireland, 2011 [online]. HPA: London; 2012 [accessed 16 June 2016]. Available from: http://webarchive.nationalarchives.gov.uk/20140629102627/http://www.hpa.org.uk/webc/HPAwebFile/HPAweb_C/1317135574852

(105) Public Health England. Voluntary surveillance of Escherichia coli bacteraemia in England, Wales and Northern Ireland: 2009-2013 [online]. PHE: London; 2014 [accessed 16 June 2016]. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/339992/Ecoli_bacteraemia_in_Eng_Wales_and_NI_2009_to_2013.pdf

(106) Harron K, Goldstein H, Wade A, Muller-Pebody B, Parslow R, Gilbert R. Linkage, evaluation and analysis of national electronic healthcare data: application to providing enhanced blood-stream infection surveillance in paediatric intensive care. *PLoS One* 2013;8(12):e85278.

(107) Harron K. Evaluating data linkage techniques for the analysis of blood-stream infection in paediatric intensive care. PhD thesis, Institute of Child Health, University College London; 2014.

(108) NHS Connecting for Health. Demographics Batch Service (DBS) [online]. NHS 2014 [accessed 16 June 2016]. Available from: http://developer.nhs.uk/library/systems/demographic-batch-service-dbs/

(109) Health & Social Care Information Centre. NHS number [online]. HSCIC: Leeds; 2016 [accessed 16 June 2016]. Available from: http://systems.hscic.gov.uk/nhsnumber

(110) Health & Social Care Information Centre. Hospital Episode Statistics [online]. HSCIC: Leeds; 2016 [accessed 16 June 2016]. Available from: http://www.hscic.gov.uk/hes

(111) Health & Social Care Information Centre. Secondary Uses Service (SUS) [online]. HSCIC: Leeds; 2016 [accessed 16 June 2016]. Available from: http://www.hscic.gov.uk/sus

(112) Health & Social Care Information Centre. The HES processing cycle and data quality [online]. HSCIC: Leeds; 2014 [accessed 16 June 2016]. Available from: http://www.hscic.gov.uk/media/1366/The-HES-processing-cycle-and-HES-data-quality/pdf/The_HES_Processing_Cycle_and_HES_Data_Quality.pdf

(113) World Health Organisation. ICD-10 Version: 2010 [online]. WHO: Geneva; 2010 [accessed 16 June 2016]. Available from: http://apps.who.int/classifications/icd10/browse/2010/en

(114) NHS Classifications Service. OPCS-4 Classification. HSCIC: Leeds; 2012 [accessed 16 June 2016]. Available from: http://systems.hscic.gov.uk/data/clinicalcoding/codingstandards/opcs4

(115) Campbell SE, Campbell MK, Grimshaw JM, Walker AE. A systematic review of discharge coding accuracy. *J Public Health Med* 2001;23(3):205-11.

(116) Williams JG, Mann RY. Hospital episode statistics: time for clinicians to get involved? *Clin Med* 2002;2(1):34-7.

(117) Brennan L, Watson M, Klaber R, Charles T. The importance of knowing context of hospital episode statistics when reconfiguring the NHS. *BMJ* 2012;344:e2432.

(118) Audit Commission. PrB data assurance framework 2007/08 [online]. Audit Commission: London; 2008 [accessed 16 June 2016]. Available from: http://collections.europarchive.org/tna/20121205001956/http://audit-commission.gov.uk/SiteCollectionDocuments/AuditCommissionReports/NationalStudies/PbRreport.pdf.

(119) Bottle A, Hansell A, Aylin P. Hospital episode statistics: time for clinicians to get involved? *Clin Med* 2002;2(5):483-4.

(120) Health & Social Care Information Centre. Methodology for creation of the HES Patient ID (HESID) [online]. HSCIC: Leeds; 2014 [accessed 16 June 2016]. Available from: http://www.hscic.gov.uk/media/1370/HES-Hospital-Episode-Statistics-Replacement-of-the-HES-patient-ID/pdf/HESID_Methodology.pdf

(121) Ministry of Justice UK Government. Women Prisoners [online]. Ministry of Justice: London; 2012 [accessed 16 June 2016]. Available from: http://www.justice.gov.uk/offenders/types-of-offender/women

(122) Melin P. Neonatal group B streptococcal disease: from pathogenesis to preventive strategies. *Clin Microbiol Infect* 2011;17(9):1294-303.

(123) Escobar GJ. What have we learned from observational studies on neonatal sepsis? *Pediatr Crit Care Med* 2005;6(3 Suppl):S138-S145.

(124) Wijlaars LP, Hardelid P, Woodman J, Allister J, Cheung R, Gilbert R. Contribution of recurrent admissions in children and young people to emergency hospital admissions: retrospective cohort analysis of hospital episode statistics. *Arch Dis Child* 2015;100(9):845-9.

(125) Farr W. Letter to the Registrar General, in: 'Supplement to 35th Annual Report of the Registrar General'. London: HMSO; 1875.

(126) Dunn HL. Record Linkage. *Am J Public Health Nations Health* 1946;36(12):1412-6.

(127) Weed LL. *Medical records, medical education, and patient care*. Cleveland: Case Western Reserve University Press; 1969.

(128) Acheson ED. Medical record linkage--The method and its applications. *R Soc Health J* 1966;86(4):216-20.

(129) Gertzog J. The changing concept of medical records. *Public Health Rep* 1970;85(8):673-9.

(130) Acheson ED. Linkage of medical records. *Br Med Bull* 1968;24(3):206-9.

(131) Holman CD, Bass AJ, Rouse IL, Hobbs MS. Population-based linkage of health records in Western Australia: development of a health services research linked database. *Aust N Z J Public Health* 1999;23(5):453-9.

(132) Lyons RA, Ford DV, Moore L, Rodgers SE. Use of data linkage to measure the population health effect of non-health-care interventions. *Lancet* 2014;383(9927):1517-9.

(133) Holman CD, Bass AJ, Rosman DL, Smith MB, Semmens JB, Glasson EJ, et al. A decade of data linkage in Western Australia: strategic design, applications and benefits of the WA data linkage system. *Aust Health Rev* 2008;32(4):766-77.

(134) Newcombe HB. *Handbook of Record Linkage*. New York: Oxford University Press; 1988.

(135) Fellegi IP, Sunter AB. A theory for record linkage. *J Am Stat Assoc* 1969;64:1183-210.

(136) Jaro MA. Probabilistic linkage of large public health data files. *Stat Med* 1995;14(5-7):491-8.

(137) Scottish Government eHealth Division. The use of the CHI (Community Health Index) across the NHS in Scotland [online]. The Knowledge Network; Glasgow: 2012 [accessed 16 June 2016]. Available from: http://www.knowledge.scot.nhs.uk/media/CLT/ResourceUploads/4014634/CHI%20Guidance%20-%20FINAL.pdf

(138) Administrative Data Liaison Service. Secure Anonymised Information Linkage Databank [online]. ADLS: St. Andrew's, Scotland; 2013 [accessed 16 June 2016]. Available from: http://www.adls.ac.uk/secure-anonymised-information-linkage-databank/

(139) Gill L, Goldacre M. English National Record Linkage of Hospital Episode Statistics and Death Registration Records [online]. Oxford University: Oxford; 2003 [accessed 16 June 2016]. Available from: http://nchod.uhce.ox.ac.uk/NCHOD%20Oxford%20E5%20Report%201st%20Feb_VerAM2.pdf

(140) Newcombe HB, Kennedy JM, Axford SJ, James AP. Automatic linkage of vital records. *Science* 1959;130(3381):954-9.

(141) Grant A. A Proposal to replace the load, de-duplication and merge functions for the national laboratory reporting system (LabBase) [internal document only]. HPA: London; 2009.

(142) Potz N, Powell D, Pebody R, Lamagni TL, Bridger D, Duckworth G. Development of a method to link infection and mortality data [online]. HPA: London; 2006 [accessed 16 June 2016]. Available from: http://webarchive.nationalarchives.gov.uk/20140714084352/http:/www.hpa.org.uk/webc/HPAwebFile/HPAweb_C/1200660056163.

(143) Office for National Statistics. Postal Geography [online]. ONS: London; 2014 [accessed 16 June 2016]. Available from: http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/postal/index.html

(144) Blackburn RM, Henderson KL, Minaji M, Muller-Pebody B, Johnson AP, Sharland M. Exploring the epidemiology of hospital-acquired bloodstream infections in children in England (January 2009-March 2010) by linkage of national hospital admissions and microbiological databases. *J Pediatr Infect Dis Soc* 2012;1(4):284-92.

(145) StataCorp LP. STATA: data analysis and statistical software [online]. StataCorp LP: Texas; 2015 [accessed 16 June 2016]. Available from: http://www.stata.com/stata13/

(146) Bohensky MA, Jolley D, Sundararajan V, Evans S, Pilcher DV, Scott I, et al. Data linkage: a powerful research tool with potential problems. *BMC Health Serv Res* 2010;10:346.

(147) Goldstein H, Harron K, Wade A. The analysis of record-linked data using multiple imputation with data value priors. *Stat Med* 2012;31(28):3481-93.

(148) Kim G, Chambers R. Regression Analysis under probabilistic multi-linkage. *Statistica Neerlandica* 2012;66(1):64-79.

(149) Zhu VJ, Overhage MJ, Egg J, Downs SM, Grannis SJ. An empiric modification to the probabilistic record linkage algorithm using frequency-based weight scaling. *J Am Med Inform Assoc* 2009;16(5):738-45.

(150) Andreas NJ, Kampmann B, Mehring Le-Doare K. Human breast milk: A review on its composition and bioactivity. *Early Hum Dev* 2015;91(11):629-35.

(151) Simon AK, Hollander GA, McMichael A. Evolution of the immune system in humans from infancy to old age. *Proc Biol Sci* 2015;282(1821):20143085.

(152) Clark RA. Activation of the neutrophil respiratory burst oxidase. *J Infect Dis* 1999;179 Suppl 2:S309-S317.

(153) Otter JA, French GL. Survival of nosocomial bacteria and spores on surfaces and inactivation by hydrogen peroxide vapor. *J Clin Microbiol* 2009;47(1):205-7.

(154) French GL, Otter JA, Shannon KP, Adams NM, Watling D, Parks MJ. Tackling contamination of the hospital environment by methicillin-resistant Staphylococcus aureus (MRSA): a comparison between conventional terminal cleaning and hydrogen peroxide vapour decontamination. *J Hosp Infect* 2004;57(1):31-7.

(155) Walther BA, Ewald PW. Pathogen survival in the external environment and the evolution of virulence. *Biol Rev Camb Philos Soc* 2004;79(4):849-69.

(156) Office for National Statistics. Population estimates [online]. ONS: London; 2016. [accessed 15 October 2016]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates

(157) Kagen J, Zaoutis TE, McGowan KL, Luan X, Shah SS. Bloodstream infection caused by Stenotrophomonas maltophilia in children. *Pediatr Infect Dis J* 2007;26(6):508-12.

(158) Olowokure B, Spencer NJ, Hawker JI, Blair I. Ethnic differences in invasive Haemophilus influenzae disease in the West Midlands region of England, UK. *Eur J Clin Microbiol Infect Dis* 2002;21(1):65-6.

(159) Murray J, Saxena S, Modi N, Majeed A, Aylin P, Bottle A. Quality of routine hospital birth records and the feasibility of their use for creating birth cohorts. *J Public Health* (Oxf) 2013;35(2):298-307.

(160) Hardelid P, Dattani N, Davey J, Pribramska I, Gilbert R. Child Health Reviews - UK: Overview of child deaths in the four UK countries [online]. Royal College of Paediatrics and Child Health: London; 2013 [accessed 16 June 2016]. Available from: http://www.rcpch.ac.uk/system/files/protected/page/CHRUK_Module%20A%20low%20res%20(2).pdf

(161) Charlson M, Szatrowski TP, Peterson J, Gold J. Validation of a combined comorbidity index. *J Clin Epidemiol* 1994;47(11):1245-51.

(162) Public Health England. Protocol for the surveillance of surgical site infection [online]. PHE: London; 2013 [accessed 16 June 2016]. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/364412/Protocol_for_surveillance_of_surgical_site_infection_June_2013.pdf

(163) Le DK, Nichols AL, Payne H, Wells R, Navidnia S, Appleby G, et al. Very low rates of culture-confirmed invasive bacterial infections in a prospective 3-year population-based surveillance in Southwest London. *Arch Dis Child* 2014;99(6):526-31.

(164) Office for National Statistics. Linking maternity data for England 2007: methods and data quality. Report No: 53 [online]. ONS: London; 2012. Available from: http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/rel/hsq/health-statistics-quarterly/no--53--spring-2012/linkage-of-maternity-hospital-episode-statistics-data.html.

(165) Goossens H, Sprenger MJ. Community acquired infections and bacterial resistance. *BMJ* 1998;317(7159):654-7.

(166) Scott E, Bloomfield SF. The survival and transfer of microbial contamination via cloths, hands and utensils. *J Appl Bacteriol* 1990;68(3):271-8.

(167) Kramer A, Schwebke I, Kampf G. How long do nosocomial pathogens persist on inanimate surfaces? A systematic review. *BMC Infect Dis* 2006;6:130.

(168) Kusumaningrum HD, Riboldi G, Hazeleger WC, Beumer RR. Survival of foodborne pathogens on stainless steel surfaces and cross-contamination to foods. *Int J Food* Microbiol 2003;85(3):227-36.

(169) Heath PT, Balfour GF, Tighe H, Verlander NQ, Lamagni TL, Efstratiou A. Group B streptococcal disease in infants: a case control study. *Arch Dis Child* 2009;94(9):674-80.

(170) Easmon CS, Hastings MJ, Clare AJ, Bloxham B, Marwood R, Rivers RP, et al. Nosocomial transmission of group B streptococci. *Br Med J (Clin Res Ed)* 1981;283(6289):459-61.

(171) Lenz R, Leal JR, Church DL, Gregson DB, Ross T, Laupland KB. The distinct category of healthcare associated bloodstream infections. *BMC Infect Dis* 2012;12:85.

(172) Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *J of the Royal Statistical Society: Applied Statistics* 1994;43(3):429-67.

(173) Royston P, Ambler G, Sauerbrei W. The use of fractional polynomials to model continuous risk variables in epidemiology. *Int J Epidemiol* 1999;28(5):964-74.

(174) Akaike H. Information theory and an extension of the maximum likelihood principle. In: Petrov,B.N., Caski,F editors. Proceedings of the Second International Symposium on Information Theory, 1973, Budapest, Hungary. Springer; 1973. p. 267-81.

(175) Bozdogan H. Akaike's Information Criterion and Recent Developments in Information Complexity. *J Math Psychol* 2000;44(1):62-91.

(176) Kass RE, Raftery AE. Bayes factors. *J Am Stat Assoc* 1995;90(430):773-95.

(177) Wagenmakers E, Farrell S. AIC model selection using Akaike weights. *Psychon Bull Rev* 2004;11(1):192-6.

(178) Goldacre M. *Correspondence analysis in practice*. 2 ed. Boca Raton Florida: Taylor & Francis Group; 2007.

(179) StataCorp LP. Multiple and joint correspondence analysis [online]. StataCorp LP: Texas; 2015 [accessed 16 June 2016]. Available from: http://www.stata.com/manuals13/mvmca.pdf

(180) Bradley PS, Mangasarian OL, Street WN. *Clustering via concave minimization*. 1997; Cambridge, MA: MIT Press; 1997 p. 368-74.

(181) von Eiff C, Peters G, Heilmann C. Pathogenesis of infections due to coagulase-negative staphylococci. *Lancet Infect Dis* 2002;2(11):677-85.

(182) Berk KN. Comparing Subset Regression Procedures. *Technometrics* 1978;20(1):1-6.

(183) Sawa T. Criteria for discriminating among alternative regression models. *Econometrica* 1978;46(6):1273-91.

(184) Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med* 2000;19(4):453-73.

(185) Schumacher M, Hollander N, Sauerbrei W. Resampling and cross-validation techniques: a tool to reduce bias caused by model building? *Stat Med* 1997;16(24):2813-27.

(186) Steyerberg EW, Harrell FE, Jr., Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol* 2001;54(8):774-81.

(187) Henderson KL, Muller-Pebody B, Wade A, Sharland M, Minaji M, Johnson AP, et al. Timing of positive blood samples does not differentiate pathogens causing healthcare-associated from community-acquired bloodstream infections in children in England: a linked retrospective cohort study. *Epidemiol Infect* 2015;143(11):2440-5.

(188) Wisplinghoff H, Seifert H, Tallent SM, Bischoff T, Wenzel RP, Edmond MB. Nosocomial bloodstream infections in pediatric patients in United States hospitals: epidemiology, clinical features and susceptibilities. *Pediatr Infect Dis J* 2003;22(8):686-91.

(189) Wisplinghoff H, Bischoff T, Tallent SM, Seifert H, Wenzel RP, Edmond MB. Nosocomial bloodstream infections in US hospitals: analysis of 24,179 cases from a prospective nationwide surveillance study. *Clin Infect Dis* 2004;39(3):309-17.

(190) Huebner J, Goldmann DA. Coagulase-negative staphylococci: role as pathogens. *Annu Rev Med* 1999;50:223-36.

(191) Le DK, Nichols AL, Payne H, Wells R, Navidnia S, Appleby G, et al. Very low rates of culture-confirmed invasive bacterial infections in a prospective 3-year population-based surveillance in Southwest London. *Arch Dis Child* 2014;99(6):526-31.

(192) NHS. Length of stay - reducing length of stay [online]. NHS: London; 2008 [accessed 16 June 2016]. Available from: http://www.institute.nhs.uk/quality_and_service_improvement_tools/quality_and_service_improvement_tools/length_of_stay.html

(193) Goncalves-Bradley DC, Lannin NA, Clemson LM, Cameron ID, Shepperd S. Discharge planning from hospital. *Cochrane Database Syst Rev* 2016;1:CD000313.

(194) Hagger-Johnson G, Harron K, Gonzalez-Izquierdo A, Cortina-Borja M, Dattani N, Muller-Pebody B, et al. Identifying Possible False Matches in Anonymized Hospital Administrative Data without Patient Identifiers. *Health Serv Res* 2015;50(4):1162-78.

(195) Aldridge RW, Shaji K, Hayward AC, Abubakar I. Accuracy of Probabilistic Linkage Using the Enhanced Matching System for Public Health and Epidemiological Studies. *PLoS One* 2015;10(8):e0136179.

(196) Bielicki JA, Sharland M, Johnson AP, Henderson KL, Cromwell DA. Selecting appropriate empirical antibiotic regimens for paediatric bloodstream infections: application of a Bayesian decision model to local and pooled antimicrobial resistance surveillance data. *J Antimicrob Chemother* 2016;71(3):794-802.

(197) Dinarelli S, Girasole M, Kasas S, Longo G. Nanotools and molecular techniques to rapidly identify and fight bacterial infections. *J Microbiol Methods* 2016 (*epub ahead of print*).

(198) Gnani S, Majeed A. A user's guide to data collected in primary care in England. Eastern Region Public Health Observatory (erpho) on behalf of the Association of Public Health Observatories: London ; 2006 Jan. Available from: http://www.institute.nhs.uk/quality_and_service_improvement_tools/quality_and_service_improvement_tools/length_of_stay.html

(199) Williams T, van ST, Puri S, Eaton S. Recent advances in the utility and use of the General Practice Research Database as an example of a UK Primary Care Data resource. *Ther Adv Drug Saf* 2012;3(2):89-99.

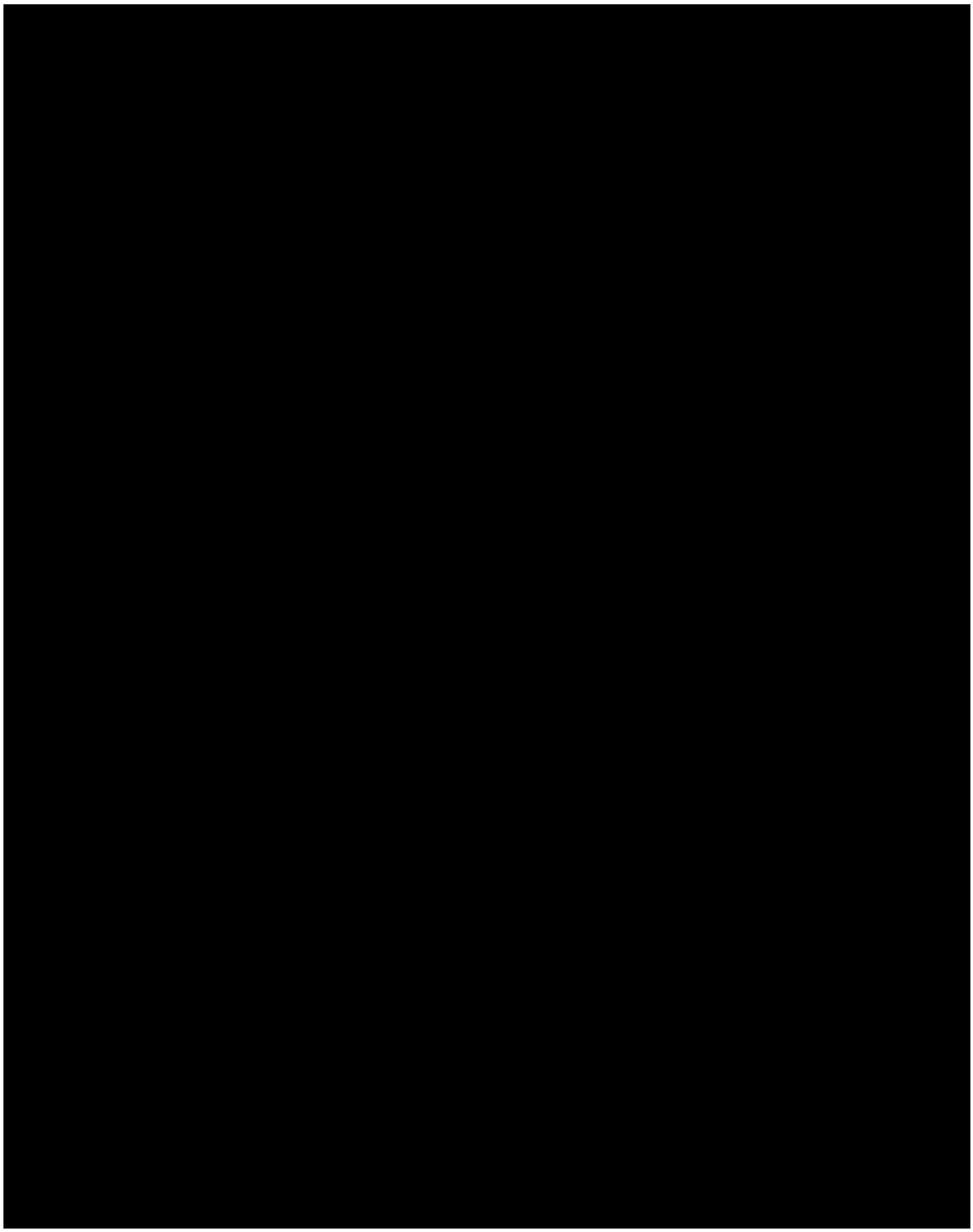
(200) National Institute for Health and Care Excellence. NICE sepsis guidance: under consultation [online]. NICE: London; 2014 [accessed 16 June 2016]. Available from: https://www.nice.org.uk/guidance/indevelopment/gid-cgwave0686

(201) Department of Health Media Centre. UK leading the global fight against drug resistant bugs. 27 May [online] 2016. [accessed 16 June 2016]. Available from: https://healthmedia.blog.gov.uk/2016/05/27/amr/

(202) Harron K, Wade A, Muller-Pebody B, Goldstein H, Gilbert R. Opening the black box of record linkage. *J Epidemiol Community Health* 2012;66(12):1198.

Appendices

Appendix Chapter 1a: Peer-reviewed literature review published in Journal of Hospital Infection



Appendix Chapter 1b: PubMed search strategy

Last extraction: March 6th 2012

Limits used: Humans

Search	CA	No. of papers
#1	community-acquired Infections	13,121
	HA *(including HCA)*	
#2	hospital-associated infections	262
#3	healthcare-related infections	85
#4	cross infection	60,476
#5	catheter-related infections	3,843
#6	nosocomial infections	64,091
#7	hospital infections	178,362
#8	hospital infection	174,468
#9	hospital acquired	40,473
#10	hospital-acquired	3,852
#11	hospital-acquired infections	3,420
#12	#2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10	263,184
	BSI	
#13	bloodstream infections	9,108
#14	bloodstream infection	8,078
#15	bacteremia	23,339
#16	bacteraemia	23,339
#17	septicemia	95,414
#18	septicaemia	95,414
#19	sepsis	90,537
#20	septic shock	16,510
#21	#12 OR #13 OR #14 OR #15 OR #17 OR #19	107,634
	Paediatric cases	
#22	child	1,463,568
	Systematic review search for adults and children	
#23	systematic[sb] AND (#1 AND #12 AND #21)	28
	Literature search for adults and children, and children only	
#24	#1 AND #12 AND #21	1,424
#25	#1 AND #12 AND #21 AND #22	303

Appendix Chapter 4a: ICD-10 codes identifying chronic conditions in children

Chronic condition	ICD-10 Codes
Mental health/behavioural	E24.4, F00-F01, F02.8, F03-F48, F50, F53, F54, F55, F59, F60-79, F80.0-F80.2, F80.8, F80.9, F81-F84, F88, F89, F90-F99, G24.0, G31.2, G40.5, G62.1, G72.0, G72.1, I42.6, K29.2, K70, K85.2, K85.3, K86.0, O35.4, R78.1-R78.5, X60-X84, Y10-Y34, Y47, Y49, , Y87.0, Y87.2, Z09.3, Z50.2, Z50.3, Z50.4, Z71.4, Z71.5, Z72.2, Z86.4, Z86.5, Z91.4, Z91.5
Cancer/blood disorders	C00-C97, D00-D02, D05-D09, D12, D13, D14.1-D14.4, D15, D20, D32-D35, D37-D48, D50, D56.0-D56.2, D56.4, D56.8, D56.9, D57.0-D57.2, D57.8, D58, D61.0, D61.9, D63.0, D64, D66, D67, D68.0-D68.2, D68.4-D68.9, D69, D70-D76, D80-D84, E34.0, E88.3, G13.0, G13.1, G53.2, G53.3, G55.0, G63.1, G73.1, G73.2, G94.1, M36.0, M36.2-M36.4, M36.1, M49.5, M82.0, M90.4, M90.6, M90.7, N08.1, N08.2, N16.1, Y43.1-Y43.3, Y84.2, Z08, Z51.0-Z51.2, Z54.1, Z54.2, Z85, Z86.0, Z86.2, Z92.3
Chronic infections	A15-A19, A50, A81, B18, B20-B24, B37.1, B37.5, B37.6, B37.7, B38.1, B39.1, B40.1, B44.0, B44.7, B45, B46, B48.7, B50.0, B50.8*, B51.0, B51.8*, B52.8*, B52.0, B55, B57.2-B57.5, B58.0, B59, B67, B69, B73, B74, B78.7, B90-B94, E35.0, F02.1, F02.4, K23.0, K23.1, K67.3, K93.0, K93.1, M00, M01.1, M49.0, N33.0, P35.0-P35.2, P35.8, P35.9, P37.0, P37.1, R75, Z21
Respiratory	E84, G47.3, J41-J47, J60-J70, J80-J86, J96.1, J98, P27, P75, S17, S27, S28, T27, T91.4, Y55.6, Z43.0, Z93.0, Z94.2
Metabolic/endocrine/ digestive/ renal/genitourinary	D55, D63.8, E00, E03.0, E03.1, E07.1, E10-E14, E22.0, E23.0, E25, E26.8, E29.1, E31, E34.1, E34.2, E34.5, E34.8, E66, E70-E72, E74-E78, E79.1-E79.9,E80.0-E80.3, E80.5, E80.7, E83, E85, E88.0, E88.1, E88.2, E88.8, E88.9, G13.2, G59.0, G63.2, G63.8, G63.3, G73.5, G73.6, G99.0, G99.8, I68.8, I79.2, K20, K21.0, K22, K23.8, K25-K28, K29.0, K29.1, K29.3-K29.9, K31, K50-K52, K55, K57, K59.2, K63.0-K63.3, K66, K72-K76, K80-K83, K85.0, K85.1, K85.8, K85.9, K86.1-K86.9, K87.0, K90, L99.0, M07.4, M07.5, M09.1, M09.2, M14.2-M14.5, M90.8, N00-N05, N07, N08.3, N08.4, N11-N15, N16.0, N16.2, N16.3, N16.4, N16.5, N16.8, N18, N19, N20-N23, N25, N26, N28, N29, N31, N32, N33.8, N35, N36, N39.1, N39.3, N39.4, N40-N42, N70-N74, N80-N82, N85, N86, N87,N88, N92, O24, S36, S37, S38, S39.6, S39.7, T06.5, T28, P96.0, T82.4, T83.1, T83.2, T83.4-T83.9, T85.5, T86.1, T86.4, Y42.1, T91.5, Y42.3, Y60.2, Y61.2, Y62.2, Y84.1, Z43.2-Z43.4, Z46.5, Z49, Z86.3, Z90.3, Z90.4, Z93.2-Z93.5, Z93.6, Z93.8, Z94.0, Z99.2
Musculoskeletal/skin	G55.1-G55.3, G63.5, G63.6, G73.7, J99.0, J99.1, L10, L11.0, L11.8, L11.9, L12-L14, L28, L40-L45, L57, L58.1, L59,L62.0, L87, L88, L90, L92, L95, L93, L98.5, M05, M06, M07.0-M07.3, M07.6, M08, M09.0, M09.8, M10-M13, M14.0, M14.6, M14.8, M30-M35, M40-M43, M45- M48,M50-M54, M60-M62, M63.8, M80.1-M80.9, M81.1-M81.9, M82.1, M82.8, M84.0-M84.2, M84.8, M84.9, M85, M86.3-M86.6, M89, M90.0, M91-M94, N08.5, S13, S22.0-S22.2, S22.5, S23, S32, S33, S68.3, S68.4, S68.8, S77, S78, S87, S88, S97, S98.0, S98.2-S98.4, T02, T04, T05, T20.3, T20.7, T21.3, T21.7,

Chronic condition	ICD-10 Codes
	T22.3, T22.7, T23.2, T23.3, T23.6, T23.7, T24.3, T24.7, T25.2, T25.3, T25.6, T25.7, T29.3, T29.7, T30.3, T30.7, T31.2-T31.9, T32.2-T32.9, T87.3-T87.6, T91.2 T91.8, T92.6, T93.1, T93.4, T93.6, T94.0, T94.1, T95.0, T95.1, T95.4, T95.8, T95.9, Y45.4, Y83.5, Z89.1, Z89.2, Z89.5-Z89.8, Z97.1
Neurological	F80.3, G40.0-G40.4, G40.6-G40.9, G41, G80-G83, H05.1-H05.9, H13.3, H17, H18, H19.3, H19.8, H21, H26, H27, H28.0-H28.2, H31, H32.8, H33, H34, H35, H40, H42.0, H43, H44, H47, H54.0- H54.2, H54.4, H60.2, H65.2-H65.4, H66.1-H66.3, H69.0, H70.1, H73.1, H74.0-H74.3, H75.0, H80, H81.0, H81.4, H83.0, H83.2, H90.0, H90.3, H90.5, H90.6, H91, I00-I28, I31-I39, I41, I42.0-I42.5, I42.7-I42.9, I43.0, I43.1, I43.2-I43.8, I44.1-I44.7, I45.1-I45.9, I46-I51, I52.8, I70-I71, I72.1-I72.4, I72.8, I72.9, I73-I77, I79.0, I79.1, I79.8, I81-I82, I98-I99, M03.6, N08.8, P10, P21.0, P52, P57, P90, P91.1, P91.2, P91.6, R56.8, S05-S08, S12, S14, S24, S26, S34, S44, S54, S64, S74, S84, S94, T06.0-T06.2, T26, T82.0-T82.3, T82.5-T82.9, T85.2, T85.3, T86.2, T90.4, T90.5, T91.1, T91.3, T92.4, Y46.0-Y46.6, Y60.5, Y61.5, Y62.5, Y84.0, Z44.2, Z45.0, Z45.3, Z50.0, Z94.1, Z95
Codes indicating non-specific chronic condition	R62, R63.3, Z43.1, Z51.5, Z75.5, Z93.1, Z99.3

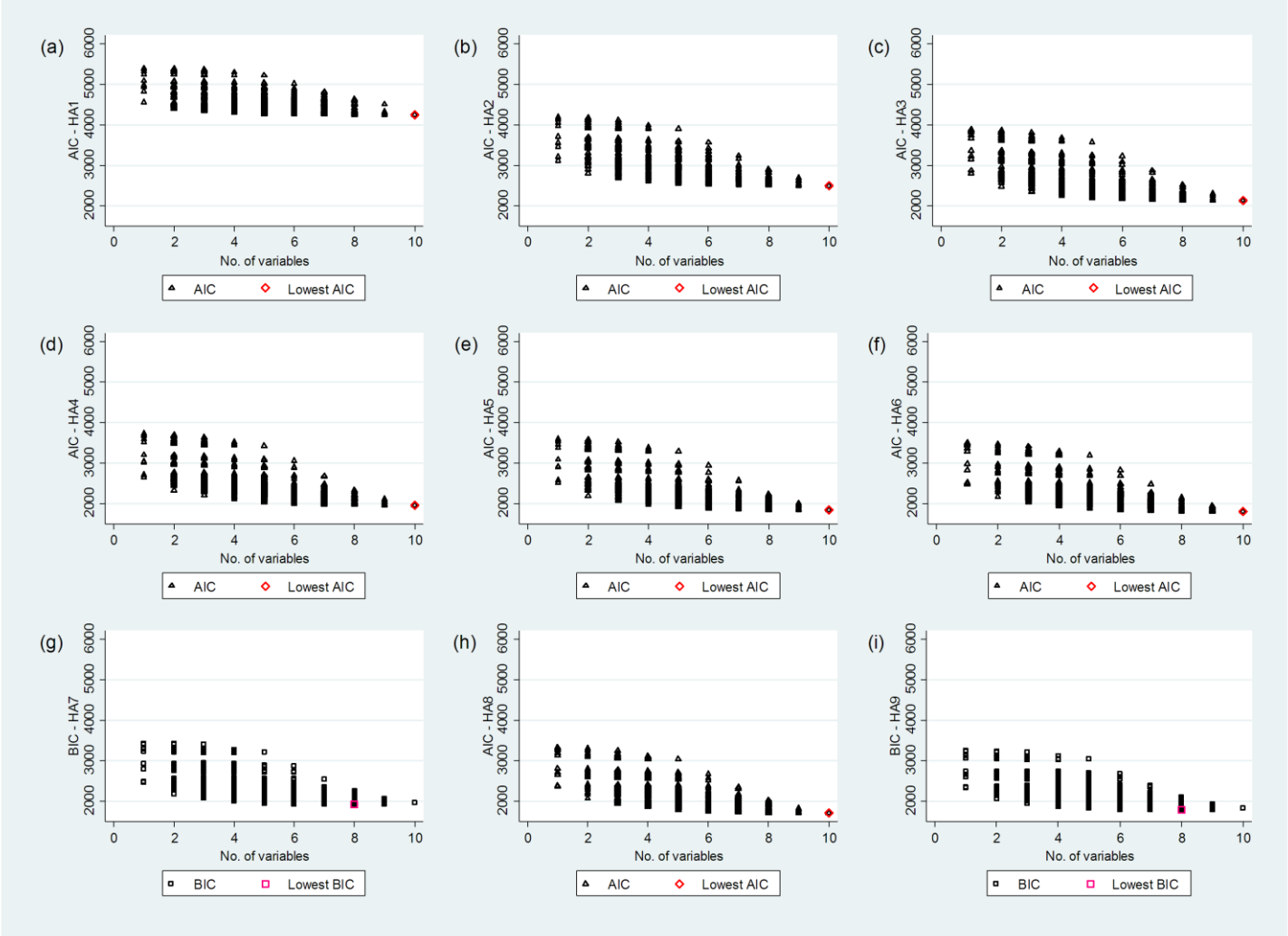
Appendix Chapter 4b: OPCS codes identifying indwelling devices

OPCS Code	Description
A123 – A129	Insertion of stent or creation of connection from ventricle of brain
A131 – A145	Procedure associated with cerebroventricular shunt
G341 – G349	Procedure associated with artificial opening into stomach (e.g. gastrostomy)
G445	Fibre-optic endoscopic percutaneous insertion of gastrostomy (PEG)
J661	Percutaneous drainage of lesion of pancreas and insertion of cystogastrostomy tube NEC
K634 – K635	Coronary arteriography using one or two catheters
K651 – K659	Catheterisation of heart
L704	Open cannulation of artery
L911 – L919	Vein related operations associated with insertion or removal of central venous catheter
L921 – L929	Unblocking on access catheter
L934	Open cannulation of vein
L942 – L949	Percutaneous transluminal cannulation of vein, insertion/replacement of subcutaneous port or stent, venoplasty or angioplasty
L997	Percutaneous transluminal insertion of peripherally inserted central catheter
X411 – Z412	Insertion/removal of ambulatory peritoneal dialysis catheter
X421	Insertion of temporary peritoneal dialysis catheter
Y512	Approach to organ through gastrostomy

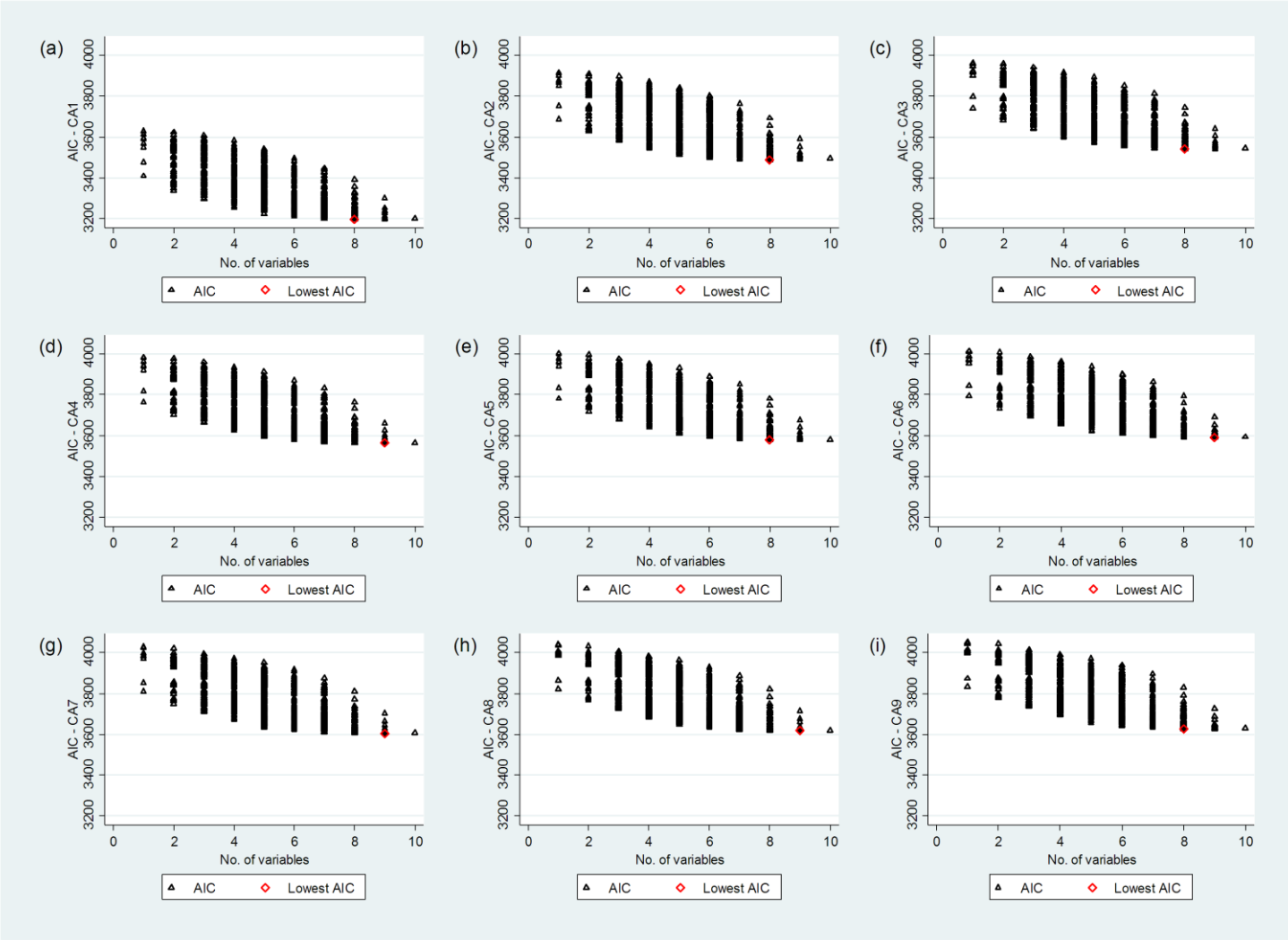
OPCS codes: identifying non-invasive/low risk for adverse event procedures

Code	Description
E852	Non-invasive ventilation
R36-R43	Obstetric scans
U01-U21	Diagnostic imaging
U22-U40	Diagnostic tests
U50-U54	Rehabilitation
X44	Administration of vaccine
X48-X49	Plaster case/support of limb
X52	Oxygen therapy
X60	Rehabilitation assessment
X61	Complimentary therapy
X62	Assessments
X63-X64	Retired codes
X65	Radiotherapy
X66	Cognitive behavioural therapy
X67-X68	Preparation for radio/brachytherapy
X70-X97	Specified drug therapy
Y95-Y96	Non-operations relating to fetus and embryo
Y90	Non-operations
Y99	Donor status

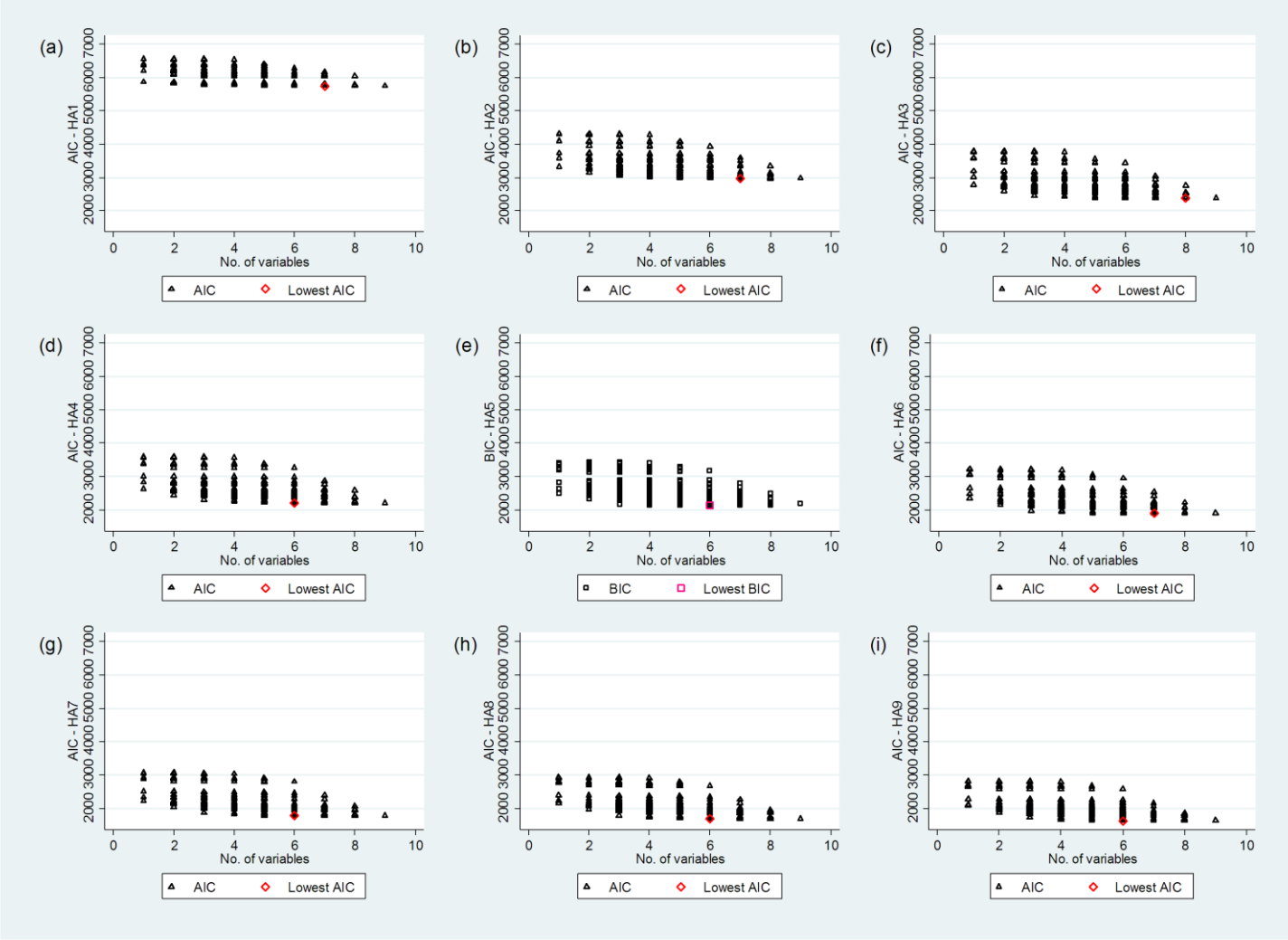
Appendix Chapter 5: Summary figures depicting all the AIC and BIC values from the approach A and approach B model



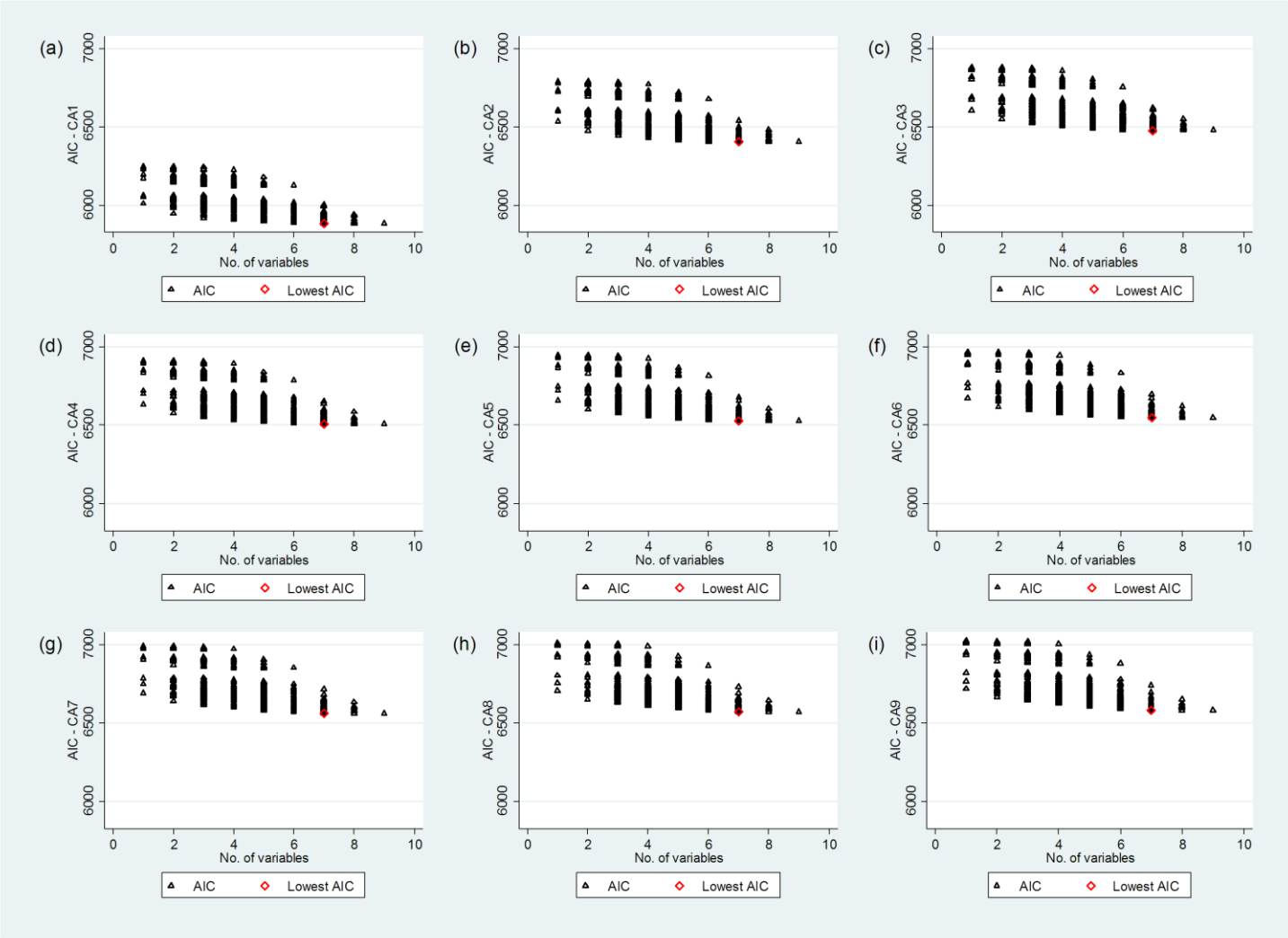
Appendix 5Ai. Approach A: best fitting AIC or BIC model (in red) in comparison to all other models (total 1,023 models) to predict HA BSI in children aged 1 month < 1 year



Appendix 5Aii. Approach A: best fitting AIC or BIC model (in red) in comparison to all other models (total 1,023 models) to predict CA BSI in children aged 1 month < 1 year



Appendix 5Aiii. Approach A: best fitting AIC or BIC model (in red) in comparison to all other models (total 551models) to predict HA BSI in children aged 1-5 years



Appendix 5A iv. Approach A: best fitting AIC or BIC model (in red) in comparison to all other models (total 551 models) to predict CA BSI in children aged 1-5 years

Approach A	HA1				HA2				HA3				HA4				HA5				HA6				HA7				HA8				HA9			
1m<1y HA outcome	AOR	P-value	95% CI		AOR	P-value	95% CI		AOR	P-value	95% CI		AOR	P-value	95% CI		AOR	P-value	95% CI		AOR	P-value	95% CI		AOR	P-value	95% CI		AOR	P-value	95% CI		AOR	P-value	95% CI	
Age in days	*(FP=1; linear)*				*(FP=1)*				*(FP=1)*				*(FP=1)*				*(FP=1)*				*(FP=1)*				*(FP=1)*				*(FP=1)*				*(FP=1)*			
Fractional polynomial (FP) 1	0.99	<0.001	0.99	0.99	1.00	<0.001	0.99	0.99	1.00	<0.001	0.99	0.99	1.00	<0.001	0.99	0.99	1.00	<0.001	0.99	0.99	1.00	<0.001	0.99	0.99	1.00	<0.001	0.99	0.99	1.00	<0.001	0.99	0.99	1.00	<0.001	0.99	0.99
Postnatal stay	*(FP = 0)*				*(FP = -2, -0.5)*				*(FP = -2,0)*				*(FP = -0.5,-0.5)*				*(FP = -1,0)*				*(FP = -2,0)*				*(FP = -2,0)*				*(FP = -2,0)*				*(FP = -2,0)*			
Fractional polynomial (FP) 1	0.88	0.001	0.82	0.95	0.99	<0.001	0.99	0.99	0.99	<0.001	0.99	0.99	2.58	<0.001	1.84	3.62	0.99	<0.001	0.99	0.99	0.99	<0.001	0.99	0.99	0.99	<0.001	0.99	0.99	0.99	<0.001	0.99	0.99	0.99	<0.001	0.99	0.99
Fractional polynomial (FP) 2					1.19	<0.001	1.12	1.28	0.71	<0.001	0.63	0.80	1.17	<0.001	1.10	1.24	0.65	<0.001	0.55	0.76	0.69	<0.001	0.61	0.78	0.68	<0.001	0.60	0.78	0.66	<0.001	0.58	0.75	0.68	<0.001	0.60	0.78
Ethnicity																																				
White	1.00				1.00				1.00				1.00				1.00				1.00								1.00							
Mixed	1.21	0.386	0.79	1.85	1.32	0.357	0.73	2.37	1.05	0.890	0.54	2.03	0.85	0.650	0.41	1.73	0.87	0.713	0.41	1.83	0.83	0.631	0.38	1.79					0.95	0.889	0.43	2.06				
Indian subcontinent	1.45	0.004	1.13	1.87	1.68	0.003	1.19	2.37	1.48	0.043	1.01	2.16	1.29	0.211	0.87	1.93	1.38	0.121	0.92	2.08	1.33	0.179	0.88	2.01					1.30	0.227	0.85	1.99				
Any other Asian	1.30	0.236	0.84	2.02	1.50	0.159	0.85	2.64	1.38	0.297	0.75	2.52	1.44	0.247	0.78	2.68	1.26	0.488	0.66	2.39	1.25	0.495	0.66	2.39					1.08	0.816	0.55	2.11				
Black	1.15	0.417	0.82	1.61	1.30	0.261	0.82	2.07	1.18	0.529	0.71	1.96	1.26	0.382	0.75	2.13	1.28	0.373	0.75	2.19	1.35	0.279	0.78	2.32					1.09	0.762	0.62	1.93				
other ethnic group	1.03	0.872	0.69	1.56	1.79	0.023	1.09	2.96	1.69	0.056	0.99	2.89	1.43	0.209	0.82	2.50	1.50	0.161	0.85	2.65	1.54	0.138	0.87	2.72					1.44	0.222	0.80	2.57				
NK	1.30	0.069	0.98	1.72	1.89	<0.001	1.32	2.70	1.88	0.001	1.28	2.76	2.02	<0.001	1.36	2.99	2.17	<0.001	1.45	3.23	2.21	<0.001	1.48	3.29					2.16	<0.001	1.44	3.24				
boy	1.00				1.00				1.00				1.00				1.00				1.00															
girl	1.14	0.089	0.98	1.33	1.26	0.035	1.02	1.55	1.24	0.066	0.99	1.56	1.27	0.053	1.00	1.61	1.20	0.144	0.94	1.54	1.31	0.035	1.02	1.69												
normal birth	1.00				1.00				1.00				1.00				1.00				1.00															
Premature/low birthweight	1.34	0.013	1.07	1.69	1.83	<0.001	1.38	2.44	1.97	<0.001	1.45	2.68	2.00	<0.001	1.45	2.74	2.17	<0.001	1.57	2.99	2.16	<0.001	1.56	2.99	2.14	<0.001	1.55	2.95	2.10	<0.001	1.51	2.92	1.92	<0.001	1.38	2.67
Chronic conditions																																				
no chronic	1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00			
1 chronic	0.82	0.046	0.67	1.00	1.59	0.002	1.19	2.12	1.79	0.001	1.28	2.50	2.01	<0.001	1.40	2.87	2.12	<0.001	1.44	3.11	2.11	<0.001	1.43	3.14	2.06	<0.001	1.38	3.07	2.04	0.001	1.34	3.10	2.18	<0.001	1.42	3.37
2 chronic	1.34	0.029	1.03	1.75	3.43	<0.001	2.46	4.77	4.28	<0.001	2.97	6.16	4.51	<0.001	3.06	6.64	4.89	<0.001	3.25	7.35	4.47	<0.001	2.94	6.80	4.77	<0.001	3.13	7.27	4.70	<0.001	3.03	7.30	5.12	<0.001	3.27	8.03
>2 chronic	2.55	<0.001	1.93	3.37	5.87	<0.001	4.19	8.22	7.30	<0.001	5.05	10.54	8.40	<0.001	5.70	12.39	9.43	<0.001	6.29	14.16	8.79	<0.001	5.81	13.30	9.54	<0.001	6.29	14.45	9.60	<0.001	6.23	14.80	9.84	<0.001	6.32	15.31
not recorded	1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00			
Congenital condition(s)	2.57	<0.001	2.09	3.16	4.46	<0.001	3.48	5.72	4.98	<0.001	3.82	6.48	4.99	<0.001	3.79	6.56	4.68	<0.001	3.54	6.19	4.81	<0.001	3.62	6.40	4.50	<0.001	3.39	5.97	4.49	<0.001	3.35	6.02	4.11	<0.001	3.07	5.52
No. of days since last indwelling device insertion																																				
no device	1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00			
91 days<1 yr	1.15	0.491	0.77	1.73	1.08	0.749	0.69	1.69	1.45	0.118	0.91	2.31	1.30	0.279	0.81	2.11	1.20	0.455	0.74	1.96	1.22	0.428	0.74	2.00	1.27	0.343	0.78	2.07	1.45	0.146	0.88	2.40	1.39	0.203	0.84	2.29
31-90 days	1.48	0.037	1.02	2.15	1.75	0.005	1.18	2.61	1.72	0.009	1.14	2.58	1.96	0.001	1.30	2.95	1.96	0.001	1.30	2.96	1.90	0.002	1.26	2.88	2.01	0.001	1.33	3.04	2.23	<0.001	1.47	3.38	2.18	<0.001	1.44	3.29
7-30 days	3.68	<0.001	2.20	6.18	4.87	<0.001	2.94	8.06	5.79	<0.001	3.48	9.63	6.83	<0.001	4.07	11.46	6.78	<0.001	4.05	11.34	6.56	<0.001	3.95	10.88	6.62	<0.001	4.02	10.92	7.17	<0.001	4.33	11.88	8.15	<0.001	4.94	13.44
1-6 days	1.90	0.056	0.98	3.67	4.36	<0.001	2.27	8.38	4.91	<0.001	2.59	9.29	6.07	<0.001	3.20	11.54	5.58	<0.001	2.98	10.44	6.68	<0.001	3.58	12.47	4.94	<0.001	2.70	9.02	5.93	<0.001	3.21	10.96	4.74	<0.001	2.58	8.70
No. of days since last invasive procedure																																				
no procedure	1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00			
91 days<1 yr	1.37	0.058	0.99	1.89	1.44	0.110	0.92	2.26	1.49	0.124	0.90	2.47	1.58	0.099	0.92	2.71	1.72	0.058	0.98	3.02	1.98	0.019	1.12	3.50	2.02	0.015	1.15	3.57	2.01	0.021	1.11	3.63	2.52	0.003	1.38	4.59
31-90 days	1.50	0.006	1.12	2.00	2.24	<0.001	1.55	3.25	2.63	<0.001	1.77	3.92	2.90	<0.001	1.92	4.40	3.31	<0.001	2.15	5.09	3.62	<0.001	2.33	5.62	3.66	<0.001	2.36	5.68	3.75	<0.001	2.38	5.89	4.36	<0.001	2.77	6.88
7-30 days	3.25	<0.001	2.39	4.41	5.17	<0.001	3.61	7.42	6.43	<0.001	4.39	9.43	7.56	<0.001	5.08	11.25	9.23	<0.001	6.12	13.94	9.68	<0.001	6.38	14.71	9.46	<0.001	6.24	14.36	9.58	<0.001	6.25	14.71	9.40	<0.001	6.09	14.51
1-6 days	8.44	<0.001	6.27	11.37	7.57	<0.001	5.42	10.58	7.39	<0.001	5.13	10.65	6.50	<0.001	4.42	9.56	6.47	<0.001	4.32	9.68	4.82	<0.001	3.17	7.33	3.88	<0.001	2.54	5.92	3.11	<0.001	1.99	4.85	3.17	<0.001	2.02	4.97
No. of days between last discharge and current admission																																				
no previous discharge	1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00			
91 days<1 yr	0.63	0.001	0.48	0.83	0.20	<0.001	0.11	0.35	0.13	<0.001	0.07	0.24	0.10	<0.001	0.05	0.19	0.10	<0.001	0.05	0.20	0.09	<0.001	0.05	0.18	0.07	<0.001	0.03	0.14	0.07	<0.001	0.04	0.15	0.05	<0.001	0.03	0.11
31-90 days	0.61	<0.001	0.46	0.81	0.16	<0.001	0.09	0.27	0.10	<0.001	0.05	0.17	0.08	<0.001	0.05	0.15	0.09	<0.001	0.05	0.16	0.07	<0.001	0.04	0.13	0.06	<0.001	0.03	0.11	0.06	<0.001	0.03	0.11	0.05	<0.001	0.03	0.10
7-30 days	0.59	0.001	0.43	0.80	0.18	<0.001	0.11	0.30	0.12	<0.001	0.07	0.22	0.10	<0.001	0.05	0.17	0.10	<0.001	0.05	0.17	0.09	<0.001	0.05	0.17	0.07	<0.001	0.04	0.14	0.08	<0.001	0.04	0.14	0.07	<0.001	0.04	0.13
1-6 days	0.77	0.124	0.56	1.07	0.28	<0.001	0.16	0.47	0.18	<0.001	0.10	0.32	0.15	<0.001	0.09	0.27	0.16	<0.001	0.09	0.28	0.16	<0.001	0.09	0.29	0.14	<0.001	0.07	0.25	0.15	<0.001	0.08	0.27	0.15	<0.001	0.08	0.27

Appendix 5Bi. Approach A: Adjusted OR results for explanatory variables in the final models (HA1-HA9) with HA BSI as the outcome in children aged 1 month < 1 year

	No. of obs = 3090				No. of obs = 3576				No. of obs = 3661				No. of obs = 3704				No. of obs = 3736				No. of obs = 3759				No. of obs = 3781				No. of obs = 3808				No. of obs = 3827			
Approach A	HA1				HA2				HA3				HA4				HA5				HA6				HA7				HA8				HA9			
1m<1y CA outcome	AOR	P-value	95% CI		AOR	P-value	95% CI		AOR	P-value	95% CI		AOR	P-value	95% CI		AOR	P-value	95% CI		AOR	P-value	95% CI		AOR	P-value	95% CI		AOR	P-value	95% CI		AOR	P-value	95% CI	
Age in days	*(FP = 0.5,2)*				*(FP = 0.5,2)*				*(FP = 0.5,2)*				*(FP = 0.5,2)*				*(FP = 0.5,2)*				*(FP = 1,2)*				*(FP = 1,2)*				*(FP = 1,2)*				*(FP = 0.5,2)*			
Fractional polynomial (FP) 1	23.93	<0.001	12.55	45.60	22.31	<0.001	11.93	41.75	22.83	<0.001	12.22	42.67	22.57	<0.001	12.10	42.12	22.53	<0.001	12.08	42.03	5.74	<0.001	3.66	9.02	5.69	<0.001	3.63	8.92	5.60	<0.001	3.58	8.78	21.82	<0.001	11.72	40.62
Fractional polynomial (FP) 2	0.83	<0.001	0.78	0.88	0.84	<0.001	0.79	0.89	0.83	<0.001	0.78	0.88	0.83	<0.001	0.78	0.88	0.83	<0.001	0.78	0.88	0.72	<0.001	0.64	0.80	0.72	<0.001	0.64	0.80	0.72	<0.001	0.65	0.80	0.83	<0.001	0.79	0.89
Postnatal stay	*(FP = -2,0)*				*(FP = -2,0)*				*(FP = -0.5,-0.5)*				*(FP = -0.5,-0.5)*				*(FP = -0.5,-0.5)*				*(FP = -2,0)*				*(FP = -2,0)*				*(FP = -2,0)*				*(FP = -2,0)*			
Fractional polynomial (FP) 1	1.00	<0.001	1.00	1.00	1.00	<0.001	1.00	1.00	2.02	<0.001	1.50	2.73	2.01	<0.001	1.49	2.70	1.98	<0.001	1.47	2.67	1.00	0.001	1.00	1.00	1.00	0.001	1.00	1.00	1.00	0.001	1.00	1.00	1.00	<0.001	1.00	1.00
Fractional polynomial (FP) 2	0.73	<0.001	0.63	0.85	0.72	<0.001	0.63	0.83	1.12	<0.001	1.07	1.17	1.12	<0.001	1.07	1.17	1.11	<0.001	1.06	1.17	0.74	<0.001	0.64	0.85	0.73	<0.001	0.63	0.85	0.73	<0.001	0.63	0.85	0.74	<0.001	0.64	0.85
Ethnicity																																				
White	1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00			
Mixed	0.60	0.048	0.36	1.00	0.64	0.072	0.39	1.04	0.60	0.042	0.37	0.98	0.59	0.033	0.36	0.96	0.59	0.035	0.36	0.96	0.59	0.037	0.36	0.97	0.60	0.042	0.37	0.98	0.61	0.046	0.37	0.99	0.60	0.041	0.37	0.98
Indian subcontinent	0.61	0.004	0.43	0.85	0.60	0.002	0.44	0.84	0.59	0.001	0.42	0.81	0.58	0.001	0.42	0.80	0.58	0.001	0.42	0.80	0.58	0.001	0.42	0.80	0.58	0.001	0.42	0.80	0.58	0.001	0.42	0.80	0.58	0.001	0.42	0.79
Any other Asian	0.23	<0.001	0.10	0.51	0.23	<0.001	0.10	0.52	0.23	<0.001	0.10	0.51	0.23	<0.001	0.10	0.51	0.22	<0.001	0.10	0.49	0.22	<0.001	0.10	0.50	0.21	<0.001	0.10	0.48	0.21	<0.001	0.10	0.47	0.22	<0.001	0.10	0.48
Black	0.36	<0.001	0.22	0.57	0.36	<0.001	0.23	0.58	0.36	<0.001	0.23	0.57	0.36	<0.001	0.23	0.58	0.36	<0.001	0.23	0.57	0.36	<0.001	0.23	0.57	0.36	<0.001	0.23	0.57	0.34	<0.001	0.22	0.55	0.34	<0.001	0.21	0.53
other ethnic group	0.60	0.046	0.36	0.99	0.68	0.135	0.41	1.13	0.67	0.116	0.41	1.10	0.67	0.115	0.41	1.10	0.68	0.126	0.41	1.11	0.68	0.134	0.42	1.12	0.67	0.113	0.41	1.10	0.68	0.121	0.41	1.11	0.66	0.097	0.40	1.08
NK	0.80	0.204	0.57	1.13	0.86	0.367	0.61	1.20	0.84	0.307	0.60	1.17	0.85	0.330	0.61	1.18	0.85	0.345	0.61	1.19	0.83	0.268	0.60	1.15	0.82	0.232	0.59	1.14	0.83	0.253	0.59	1.15	0.82	0.228	0.59	1.13
normal birth	1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00			
Premature/low birthweight	1.56	0.009	1.12	2.18	1.62	0.004	1.17	2.23	1.64	0.003	1.19	2.28	1.63	0.003	1.18	2.25	1.62	0.003	1.17	2.24	1.64	0.003	1.19	2.26	1.63	0.003	1.18	2.25	1.63	0.003	1.18	2.24	1.57	0.006	1.14	2.15
Chronic conditions																																				
no chronic	1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00			
1 chronic	1.96	<0.001	1.62	2.38	2.05	<0.001	1.70	2.47	2.02	<0.001	1.68	2.43	2.03	<0.001	1.69	2.44	2.02	<0.001	1.68	2.42	2.05	<0.001	1.71	2.47	2.05	<0.001	1.70	2.46	2.05	<0.001	1.70	2.47	2.01	<0.001	1.68	2.42
2 chronic	2.37	<0.001	1.67	3.36	2.57	<0.001	1.84	3.61	2.61	<0.001	1.87	3.65	2.54	<0.001	1.82	3.55	2.51	<0.001	1.80	3.49	2.49	<0.001	1.79	3.46	2.45	<0.001	1.76	3.40	2.42	<0.001	1.74	3.36	2.34	<0.001	1.69	3.25
>2 chronic	1.62	0.043	1.02	2.57	1.74	0.014	1.12	2.72	1.69	0.018	1.10	2.62	1.68	0.019	1.09	2.60	1.70	0.017	1.10	2.62	1.69	0.017	1.10	2.61	1.64	0.026	1.06	2.52	1.62	0.028	1.05	2.49	1.48	0.073	0.96	2.26
not recorded	1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00			
Congenital condition(s)	0.62	0.006	0.45	0.87	0.66	0.013	0.48	0.92	0.65	0.008	0.47	0.89	0.63	0.004	0.46	0.86	0.59	0.001	0.43	0.81	0.61	0.002	0.44	0.84	0.58	0.001	0.42	0.80	0.56	<0.001	0.41	0.77	0.53	<0.001	0.39	0.72
No. of days since last indwelling device insertion																																				
no device	1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00			
91 days<1 yr	0.64	0.180	0.34	1.23	0.75	0.370	0.41	1.40	0.79	0.453	0.43	1.46	0.75	0.362	0.41	1.38	0.74	0.331	0.40	1.36	0.76	0.382	0.42	1.40	0.79	0.432	0.43	1.44	0.81	0.502	0.44	1.49	0.78	0.416	0.43	1.42
31-90 days	0.17	0.002	0.06	0.53	0.24	0.009	0.08	0.70	0.22	0.005	0.08	0.64	0.23	0.007	0.08	0.67	0.23	0.007	0.08	0.67	0.26	0.013	0.09	0.75	0.27	0.015	0.09	0.77	0.28	0.017	0.10	0.79	0.25	0.011	0.09	0.73
7-30 days	0.11	0.042	0.01	0.92	0.19	0.121	0.02	1.55	0.21	0.146	0.03	1.71	0.23	0.163	0.03	1.82	0.23	0.160	0.03	1.80	0.27	0.210	0.03	2.11	0.26	0.203	0.03	2.06	0.27	0.209	0.03	2.10	0.26	0.199	0.03	2.03
1-6 days	3.70	0.121	0.71	19.33	8.16	0.009	1.70	39.21	6.17	0.008	1.62	23.52	6.92	0.004	1.83	26.23	6.31	0.003	1.84	21.59	7.15	0.002	2.08	24.61	4.09	0.006	1.51	11.09	3.24	0.012	1.29	8.13	2.49	0.040	1.04	5.94
No. of days since last invasive procedure																																				
no procedure	1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00			
91 days<1 yr	0.83	0.337	0.58	1.21	0.73	0.087	0.52	1.05	0.71	0.059	0.50	1.01	0.71	0.057	0.50	1.01	0.72	0.064	0.51	1.02	0.74	0.095	0.52	1.05	0.75	0.112	0.53	1.07	0.76	0.126	0.54	1.08	0.76	0.123	0.54	1.08
31-90 days	0.65	0.057	0.41	1.01	0.58	0.017	0.37	0.91	0.57	0.014	0.37	0.89	0.57	0.012	0.36	0.88	0.57	0.012	0.37	0.88	0.65	0.060	0.41	1.02	0.66	0.074	0.42	1.04	0.67	0.081	0.43	1.05	0.60	0.023	0.39	0.93
7-30 days	0.35	0.002	0.18	0.69	0.34	0.001	0.18	0.66	0.33	0.001	0.17	0.65	0.34	0.001	0.18	0.66	0.35	0.002	0.18	0.68	0.44	0.021	0.22	0.88	0.47	0.030	0.23	0.93	0.45	0.022	0.23	0.89	0.35	0.001	0.18	0.67
1-6 days	3.41	<0.001	2.01	5.81	1.44	0.088	0.95	2.20	1.20	0.379	0.80	1.81	1.06	0.775	0.71	1.59	1.01	0.950	0.68	1.51	0.94	0.749	0.63	1.40	0.84	0.405	0.57	1.26	0.81	0.305	0.55	1.21	0.76	0.173	0.52	1.13
No. of days between last discharge and current admission																																				
no previous discharge																					1.00				1.00				1.00							
91 days<1 yr																					0.83	0.442	0.52	1.33	0.81	0.370	0.51	1.29	0.81	0.384	0.51	1.30				
31-90 days																					0.69	0.136	0.43	1.12	0.67	0.100	0.41	1.08	0.66	0.092	0.41	1.07				
7-30 days																					0.55	0.024	0.32	0.93	0.53	0.018	0.32	0.90	0.53	0.018	0.32	0.90				
1-6 days																					0.71	0.195	0.43	1.19	0.70	0.165	0.42	1.16	0.71	0.187	0.43	1.18				

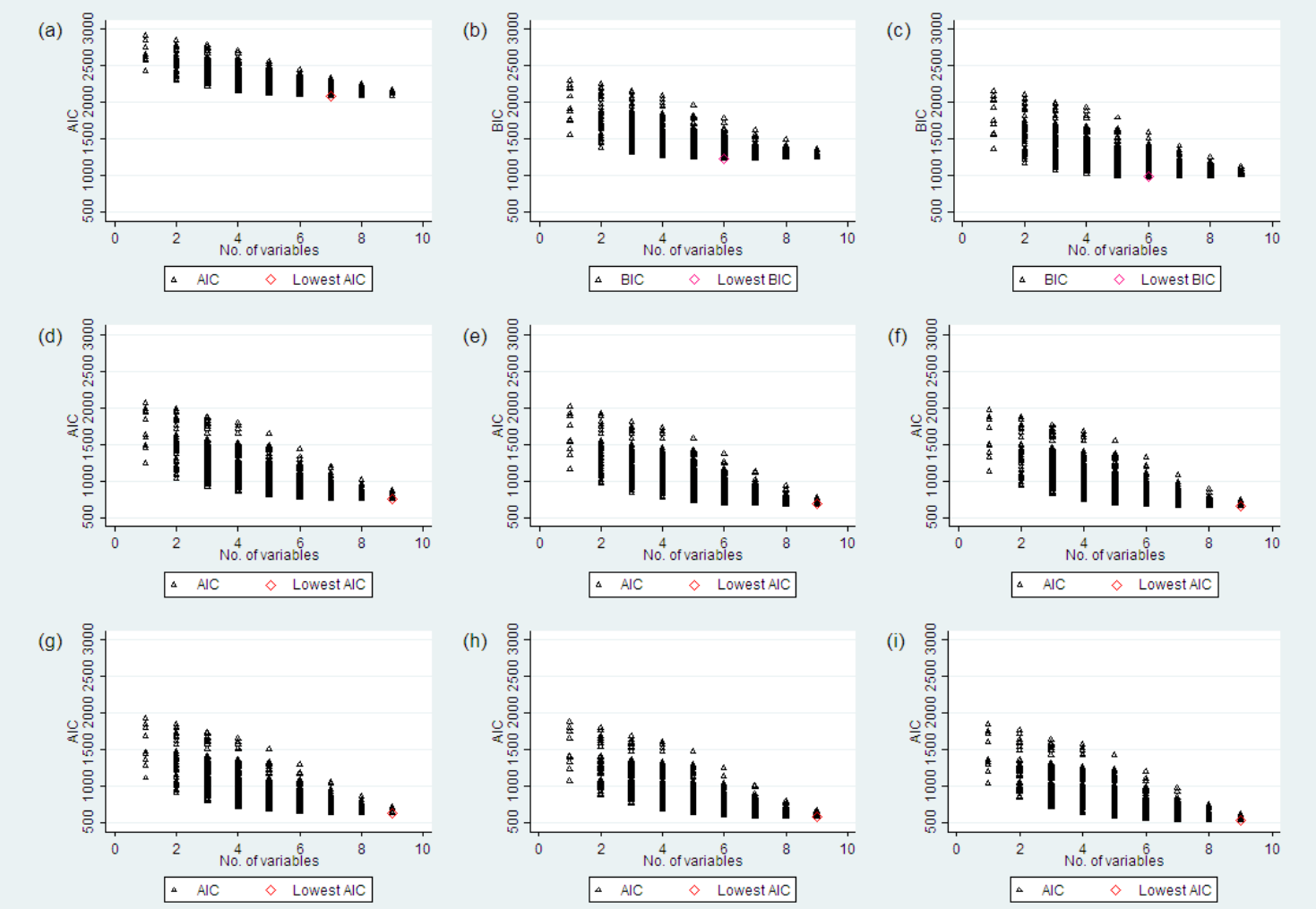
Appendix 5Bii. Approach A: Adjusted OR results for explanatory variables in the final models (HA1-HA9) with CA BSI as the outcome in children aged 1 month < 1 year

Approach A	HA1				HA2				HA3				HA4				HA5				HA6				HA7				HA8				HA9			
1-5yrs HA outcome	AOR	P-value	95% CI		AOR	P-value	95% CI		AOR	P-value	95% CI		AOR	P-value	95% CI		AOR	P-value	95% CI		AOR	P-value	95% CI		AOR	P-value	95% CI		AOR	P-value	95% CI		AOR	P-value	95% CI	
Age in days	*(FP = 1)*				*(FP = 1)*				*(FP = 1)*				*(FP = 1)*				*(FP = 1)*				*(FP = 1)*				*(FP = 1)*				*(FP = 1)*				*(FP = 1)*			
Fractional polynomial (FP) 1	0.99	0.002	0.99	0.99	0.99	0.007	0.99	0.99	0.99	0.010	0.99	0.99	0.99	0.003	0.99	0.99	0.99	<0.001	0.99	0.99	0.99	<0.001	0.99	0.99	0.99	<0.001	0.99	0.99	0.99	<0.001	0.99	0.99	0.99	<0.001	0.99	0.99
Ethnicity																																				
White	1.00				1.00				1.00												1.00															
Mixed	0.65	0.046	0.42	0.99	0.65	0.194	0.34	1.24	0.67	0.270	0.33	1.36									0.68	0.350	0.30	1.53												
Indian subcontinent	1.25	0.056	0.99	1.56	1.53	0.012	1.10	2.13	1.46	0.052	1.00	2.13									1.49	0.066	0.97	2.29												
Any other Asian	1.22	0.310	0.83	1.80	1.20	0.525	0.68	2.13	0.85	0.634	0.43	1.67									0.72	0.410	0.33	1.57												
Black	1.00	0.975	0.75	1.32	1.36	0.142	0.90	2.05	1.21	0.435	0.75	1.95									1.58	0.081	0.95	2.64												
other ethnic group	0.93	0.693	0.64	1.35	1.51	0.094	0.93	2.45	1.50	0.144	0.87	2.57									1.75	0.066	0.96	3.16												
NK	1.25	0.064	0.99	1.59	1.60	0.005	1.15	2.22	1.79	0.001	1.25	2.57									1.51	0.049	1.00	2.27												
normal birth									1.00																											
Premature/low birthweight									1.28	0.134	0.93	1.78																								
Chronic conditions																																				
no chronic	1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00			
1 chronic	1.17	0.079	0.98	1.39	2.76	<0.001	2.00	3.82	4.29	<0.001	2.77	6.64	4.32	<0.001	2.69	6.95	5.88	<0.001	3.34	10.37	5.32	<0.001	2.89	9.80	5.89	<0.001	3.00	11.58	6.98	<0.001	3.35	14.55	7.94	<0.001	3.50	18.02
2 chronic	1.20	0.138	0.94	1.51	3.74	<0.001	2.58	5.43	6.63	<0.001	4.11	10.68	7.55	<0.001	4.53	12.58	11.29	<0.001	6.21	20.51	11.84	<0.001	6.28	22.33	15.24	<0.001	7.62	30.49	16.28	<0.001	7.65	34.65	19.55	<0.001	8.48	45.06
>2 chronic	1.99	<0.001	1.57	2.53	8.98	<0.001	6.25	12.89	16.89	<0.001	10.60	26.91	19.71	<0.001	11.98	32.42	31.56	<0.001	17.58	56.65	32.05	<0.001	17.19	59.75	41.34	<0.001	20.88	81.84	50.16	<0.001	23.89	105.32	60.07	<0.001	26.40	136.67
not recorded	1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00			
Congenital condition(s)	1.80	<0.001	1.47	2.20	2.24	<0.001	1.76	2.86	2.35	<0.001	1.80	3.06	2.48	<0.001	1.90	3.24	2.29	<0.001	1.73	3.03	2.19	<0.001	1.64	2.93	2.20	<0.001	1.64	2.95	2.36	<0.001	1.74	3.19	2.36	<0.001	1.74	3.20
No. of days since last indwelling device insertion																																				
no device	1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00			
91 days<1 yr	0.86	0.301	0.64	1.15	0.96	0.838	0.68	1.37	0.98	0.930	0.67	1.44	0.95	0.810	0.64	1.41	0.93	0.728	0.62	1.40	0.88	0.564	0.58	1.35	0.92	0.715	0.59	1.43	1.03	0.908	0.66	1.61	1.13	0.608	0.71	1.78
31-90 days	1.08	0.641	0.78	1.49	1.46	0.040	1.02	2.10	1.68	0.008	1.14	2.47	1.77	0.005	1.19	2.62	1.67	0.015	1.11	2.52	1.77	0.008	1.16	2.70	1.91	0.003	1.24	2.93	2.09	0.001	1.34	3.27	2.38	<0.001	1.51	3.73
7-30 days	2.15	<0.001	1.55	2.99	2.55	<0.001	1.79	3.64	3.03	<0.001	2.08	4.42	3.15	<0.001	2.15	4.62	3.59	<0.001	2.42	5.33	3.89	<0.001	2.59	5.83	4.31	<0.001	2.85	6.52	4.96	<0.001	3.24	7.60	5.37	<0.001	3.49	8.28
1-6 days	1.57	0.048	1.00	2.45	2.06	0.003	1.28	3.32	1.97	0.009	1.19	3.27	2.12	0.004	1.27	3.54	2.22	0.003	1.30	3.76	2.13	0.007	1.23	3.67	2.03	0.014	1.15	3.59	2.04	0.019	1.13	3.70	2.00	0.026	1.09	3.69
No. of days since last invasive procedure																																				
no procedure	1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00			
91 days<1 yr	1.11	0.512	0.82	1.51	1.89	0.008	1.18	3.02	2.54	0.001	1.48	4.36	2.87	<0.001	1.64	5.02	2.78	0.001	1.53	5.04	3.29	<0.001	1.78	6.08	3.19	<0.001	1.69	6.03	3.24	<0.001	1.69	6.22	2.94	0.001	1.52	5.71
31-90 days	1.96	<0.001	1.42	2.71	5.57	<0.001	3.60	8.62	8.06	<0.001	4.91	13.24	8.28	<0.001	4.94	13.90	9.46	<0.001	5.49	16.31	11.83	<0.001	6.73	20.80	10.97	<0.001	6.12	19.66	10.75	<0.001	5.91	19.58	9.66	<0.001	5.26	17.74
7-30 days	4.54	<0.001	3.43	6.00	15.57	<0.001	10.77	22.51	26.68	<0.001	17.55	40.54	30.75	<0.001	19.89	47.55	34.24	<0.001	21.62	54.21	40.01	<0.001	24.72	64.77	40.15	<0.001	24.42	66.02	37.95	<0.001	22.77	63.23	30.87	<0.001	18.39	51.84
1-6 days	7.30	<0.001	5.61	9.51	15.63	<0.001	11.15	21.90	21.21	<0.001	14.45	31.14	20.47	<0.001	13.74	30.51	19.72	<0.001	12.93	30.10	20.05	<0.001	12.82	31.35	16.60	<0.001	10.44	26.40	13.61	<0.001	8.43	21.98	11.21	<0.001	6.84	18.36
No. of days between last discharge and current admission																																				
no previous discharge	1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00			
91 days<1 yr	1.13	0.311	0.89	1.42	0.56	0.004	0.38	0.83	0.45	0.001	0.28	0.71	0.43	<0.001	0.26	0.69	0.44	0.001	0.26	0.72	0.45	0.003	0.27	0.76	0.42	0.002	0.25	0.72	0.36	<0.001	0.21	0.63	0.39	0.001	0.22	0.68
31-90 days	0.84	0.230	0.63	1.12	0.30	<0.001	0.20	0.47	0.22	<0.001	0.13	0.35	0.21	<0.001	0.13	0.34	0.18	<0.001	0.10	0.30	0.17	<0.001	0.10	0.29	0.16	<0.001	0.09	0.27	0.14	<0.001	0.08	0.24	0.13	<0.001	0.07	0.24
7-30 days	0.69	0.006	0.53	0.90	0.19	<0.001	0.13	0.28	0.10	<0.001	0.07	0.16	0.09	<0.001	0.05	0.13	0.07	<0.001	0.04	0.11	0.06	<0.001	0.04	0.11	0.06	<0.001	0.03	0.09	0.05	<0.001	0.03	0.08	0.06	<0.001	0.03	0.10
1-6 days	0.63	<0.001	0.49	0.80	0.23	<0.001	0.16	0.32	0.14	<0.001	0.10	0.21	0.12	<0.001	0.08	0.18	0.10	<0.001	0.07	0.16	0.10	<0.001	0.07	0.16	0.09	<0.001	0.06	0.14	0.08	<0.001	0.05	0.12	0.08	<0.001	0.05	0.13

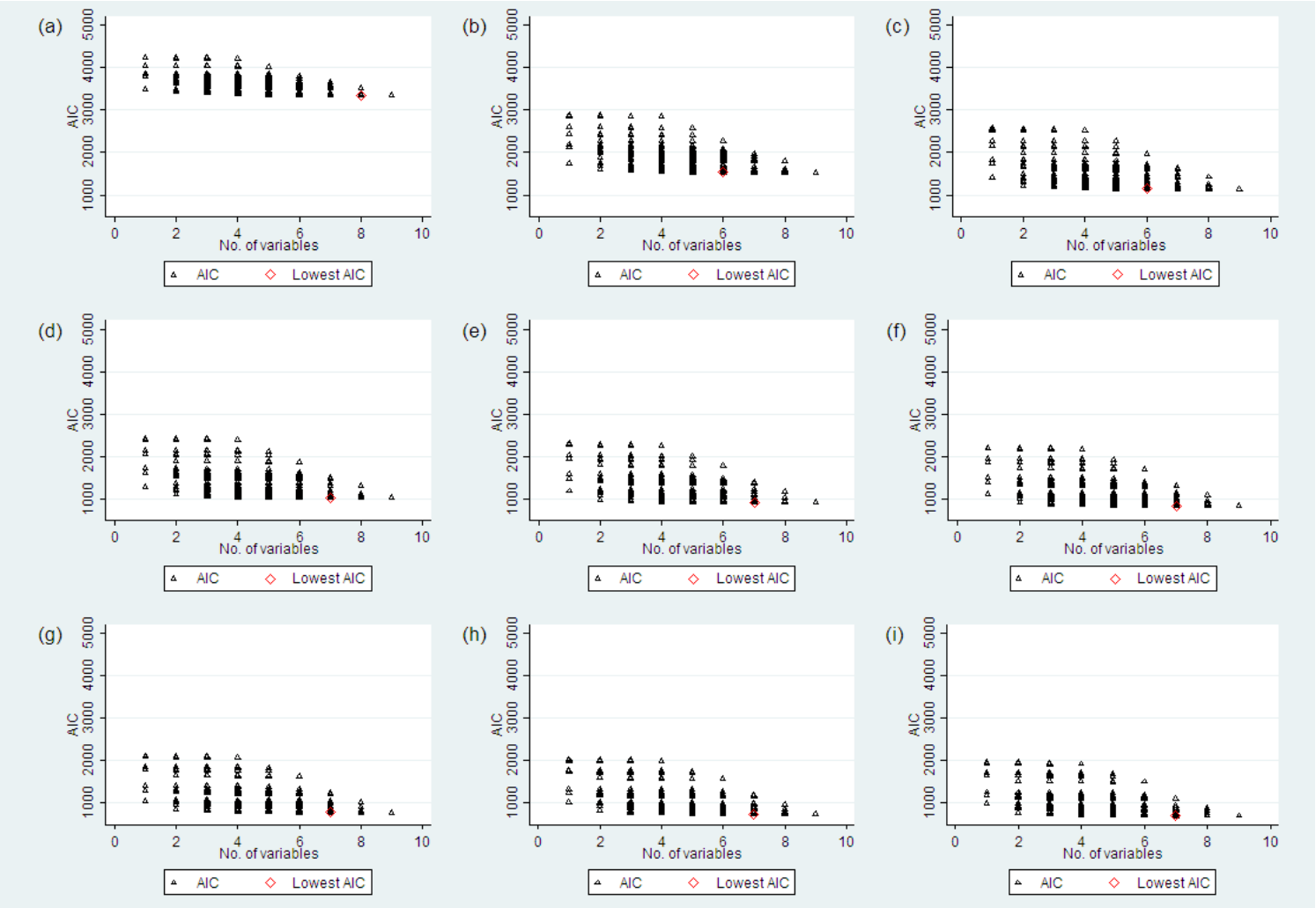
Appendix 5Biii. Approach A: Adjusted OR results for explanatory variables in the final models (HA1-HA9) with HA BSI as the outcome in children aged 1-5 years

	No. of obs = 4793				No. of obs = 5460				No. of obs = 5576				No. of obs = 5619				No. of obs = 5666				No. of obs = 5694				No. of obs = 5731				No. of obs = 5756				No. of obs = 5776			
Approach A	CA1				CA2				CA3				CA4				CA5				CA6				CA7				CA8				CA9			
1-5y CA outcome	AOR	P-value	95% CI		AOR	P-value	95% CI		AOR	P-value	95% CI		AOR	P-value	95% CI		AOR	P-value	95% CI		AOR	P-value	95% CI		AOR	P-value	95% CI		AOR	P-value	95% CI		AOR	P-value	95% CI	
Age in days	*(FP = 1)*				*(FP = -2,-1)*				*(FP = -2,-1)*				*(FP = -2,-1)*				*(FP = -2,-1)*				*(FP = -2,-1)*				*(FP = -2,-1)*				*(FP = -2,-1)*				*(FP = -2,-1)*			
Fractional polynomial (FP) 1	1.00	<0.001	1.00	1.00	1.36	<0.001	1.16	1.60	1.38	<0.001	1.17	1.62	1.37	<0.001	1.16	1.60	1.36	<0.001	1.16	1.60	1.36	<0.001	1.16	1.60	1.36	<0.001	1.16	1.60	1.37	<0.001	1.16	1.60	1.37	<0.001	1.16	1.60
Fractional polynomial (FP) 2					0.36	<0.001	0.22	0.58	0.35	<0.001	0.21	0.56	0.36	<0.001	0.22	0.58	0.36	<0.001	0.22	0.58	0.36	<0.001	0.22	0.58	0.36	<0.001	0.22	0.58	0.35	<0.001	0.22	0.57	0.35	<0.001	0.22	0.57
Ethnicity																																				
White	1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00			
Mixed	1.03	0.875	0.74	1.43	1.08	0.639	0.78	1.50	1.08	0.628	0.78	1.50	1.10	0.547	0.80	1.53	1.11	0.524	0.80	1.54	1.10	0.559	0.80	1.52	1.11	0.541	0.80	1.53	1.11	0.528	0.80	1.54	1.11	0.541	0.80	1.53
Indian subcontinent	0.80	0.050	0.64	1.00	0.79	0.035	0.64	0.98	0.78	0.028	0.63	0.97	0.78	0.023	0.63	0.97	0.77	0.021	0.62	0.96	0.78	0.022	0.63	0.96	0.77	0.020	0.62	0.96	0.77	0.020	0.62	0.96	0.78	0.021	0.62	0.96
Any other Asian	0.59	0.013	0.39	0.89	0.58	0.008	0.39	0.87	0.57	0.006	0.38	0.86	0.58	0.007	0.39	0.86	0.58	0.008	0.39	0.87	0.58	0.008	0.39	0.87	0.58	0.008	0.39	0.87	0.58	0.008	0.39	0.87	0.58	0.008	0.39	0.87
Black	0.71	0.008	0.55	0.91	0.74	0.020	0.58	0.95	0.74	0.016	0.57	0.94	0.74	0.018	0.58	0.95	0.74	0.017	0.58	0.95	0.75	0.020	0.58	0.96	0.73	0.014	0.57	0.94	0.73	0.013	0.57	0.94	0.73	0.011	0.57	0.93
other ethnic group	0.69	0.048	0.48	1.00	0.76	0.144	0.53	1.10	0.76	0.136	0.53	1.09	0.77	0.146	0.53	1.10	0.77	0.158	0.54	1.11	0.76	0.144	0.53	1.10	0.76	0.136	0.53	1.09	0.76	0.143	0.53	1.10	0.77	0.146	0.53	1.10
NK	1.25	0.060	0.99	1.57	1.23	0.070	0.98	1.53	1.23	0.060	0.99	1.54	1.21	0.091	0.97	1.50	1.20	0.102	0.96	1.49	1.19	0.118	0.96	1.48	1.18	0.129	0.95	1.47	1.18	0.131	0.95	1.47	1.17	0.162	0.94	1.45
Chronic conditions																																				
no chronic	1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00			
1 chronic	1.54	<0.001	1.33	1.78	1.53	<0.001	1.33	1.75	1.51	<0.001	1.32	1.74	1.50	<0.001	1.30	1.72	1.50	<0.001	1.30	1.72	1.48	<0.001	1.29	1.70	1.47	<0.001	1.28	1.69	1.48	<0.001	1.29	1.69	1.47	<0.001	1.28	1.69
2 chronic	1.45	0.002	1.15	1.84	1.50	<0.001	1.19	1.88	1.51	<0.001	1.20	1.89	1.51	<0.001	1.20	1.89	1.50	<0.001	1.20	1.88	1.49	<0.001	1.19	1.87	1.49	0.001	1.19	1.87	1.47	0.001	1.18	1.85	1.47	0.001	1.17	1.84
>2 chronic	1.70	<0.001	1.27	2.28	1.88	<0.001	1.42	2.50	1.84	<0.001	1.39	2.44	1.79	<0.001	1.35	2.36	1.78	<0.001	1.35	2.36	1.75	<0.001	1.33	2.32	1.75	<0.001	1.33	2.31	1.73	<0.001	1.31	2.28	1.68	<0.001	1.28	2.22
not recorded	1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00			
Congenital condition(s)	0.63	0.002	0.47	0.84	0.60	<0.001	0.46	0.80	0.60	<0.001	0.45	0.79	0.60	<0.001	0.45	0.79	0.57	<0.001	0.43	0.76	0.57	<0.001	0.43	0.75	0.56	<0.001	0.42	0.73	0.56	<0.001	0.42	0.73	0.55	<0.001	0.42	0.73
No. of days since last indwelling device insertion																																				
no device	1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00			
91 days<1 yr	0.43	<0.001	0.28	0.67	0.46	<0.001	0.29	0.71	0.46	0.001	0.30	0.71	0.46	0.001	0.30	0.72	0.47	0.001	0.30	0.72	0.47	0.001	0.30	0.72	0.47	0.001	0.30	0.73	0.48	0.001	0.31	0.74	0.49	0.001	0.32	0.76
31-90 days	0.20	<0.001	0.10	0.40	0.24	<0.001	0.12	0.47	0.24	<0.001	0.13	0.47	0.25	<0.001	0.13	0.48	0.25	<0.001	0.13	0.48	0.25	<0.001	0.13	0.49	0.26	<0.001	0.13	0.49	0.26	<0.001	0.13	0.50	0.27	<0.001	0.14	0.52
7-30 days	0.17	<0.001	0.07	0.42	0.20	<0.001	0.08	0.48	0.20	<0.001	0.09	0.48	0.21	<0.001	0.09	0.49	0.21	<0.001	0.09	0.51	0.21	<0.001	0.09	0.51	0.22	0.001	0.09	0.52	0.22	0.001	0.09	0.53	0.23	0.001	0.10	0.56
1-6 days	0.38	0.022	0.17	0.87	0.52	0.096	0.24	1.13	0.49	0.067	0.22	1.05	0.49	0.071	0.23	1.06	0.48	0.063	0.23	1.04	0.49	0.066	0.23	1.05	0.46	0.047	0.22	0.99	0.47	0.051	0.22	1.00	0.47	0.046	0.22	0.99
No. of days since last invasive procedure																																				
no procedure	1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00			
91 days<1 yr	0.85	0.315	0.61	1.17	0.85	0.305	0.62	1.16	0.85	0.304	0.62	1.16	0.86	0.335	0.63	1.17	0.86	0.339	0.63	1.17	0.86	0.353	0.63	1.18	0.86	0.324	0.63	1.17	0.86	0.331	0.63	1.17	0.85	0.316	0.63	1.16
31-90 days	0.64	0.037	0.41	0.97	0.67	0.055	0.44	1.01	0.66	0.047	0.43	0.99	0.65	0.043	0.43	0.99	0.65	0.044	0.43	0.99	0.65	0.043	0.43	0.99	0.64	0.036	0.43	0.97	0.64	0.031	0.42	0.96	0.63	0.026	0.42	0.95
7-30 days	0.61	0.038	0.38	0.97	0.59	0.026	0.37	0.94	0.58	0.020	0.36	0.92	0.58	0.019	0.36	0.91	0.56	0.013	0.35	0.88	0.55	0.010	0.35	0.87	0.53	0.007	0.34	0.84	0.52	0.005	0.33	0.82	0.49	0.002	0.31	0.77
1-6 days	1.69	0.017	1.10	2.60	1.01	0.954	0.69	1.47	0.95	0.775	0.66	1.37	0.87	0.445	0.60	1.25	0.80	0.229	0.56	1.15	0.75	0.118	0.53	1.07	0.68	0.031	0.48	0.97	0.64	0.010	0.45	0.90	0.60	0.003	0.42	0.84
No. of days between last discharge and current admission																																				
no previous discharge	1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00			
91 days<1 yr	0.64	<0.001	0.51	0.80	0.61	<0.001	0.49	0.76	0.62	<0.001	0.50	0.76	0.62	<0.001	0.50	0.76	0.62	<0.001	0.50	0.76	0.62	<0.001	0.50	0.77	0.63	<0.001	0.51	0.77	0.63	<0.001	0.51	0.77	0.63	<0.001	0.51	0.78
31-90 days	0.56	<0.001	0.41	0.76	0.53	<0.001	0.39	0.70	0.53	<0.001	0.40	0.71	0.53	<0.001	0.40	0.72	0.53	<0.001	0.40	0.71	0.53	<0.001	0.40	0.71	0.54	<0.001	0.40	0.72	0.54	<0.001	0.40	0.72	0.54	<0.001	0.41	0.72
7-30 days	0.38	<0.001	0.27	0.54	0.34	<0.001	0.25	0.48	0.34	<0.001	0.25	0.48	0.34	<0.001	0.24	0.47	0.34	<0.001	0.24	0.47	0.34	<0.001	0.25	0.48	0.35	<0.001	0.25	0.48	0.35	<0.001	0.25	0.49	0.36	<0.001	0.26	0.50
1-6 days	0.41	<0.001	0.31	0.55	0.44	<0.001	0.33	0.58	0.44	<0.001	0.33	0.57	0.44	<0.001	0.34	0.58	0.45	<0.001	0.34	0.59	0.46	<0.001	0.35	0.60	0.47	<0.001	0.36	0.61	0.47	<0.001	0.36	0.61	0.48	<0.001	0.37	0.62

Appendix 5Biv. Approach A: Adjusted OR results for explanatory variables in the final models (HA1-HA9) with CA BSI as the outcome in children aged 1-5 years



Appendix 5Ci. Approach B: best fitting AIC or BIC model (in red) in comparison to all other models (total 1,023 models) to predict HA BSI in children aged 1 month < 1 year



Appendix 5Cii. Approach B: best fitting AIC or BIC model (in red) in comparison to all other models (total 551 models) to predict HA BSI in children aged 1-5 years

	No. of obs = 2173				No. of obs = 1651				No. of obs = 1556				No. of obs = 1509				No. of obs = 1473				No. of obs = 1446				No. of obs = 1426				No. of obs = 1400				No. of obs = 1378			
Approach B	HA1				HA2				HA3				HA4				HA5				HA6				HA7				HA8				HA9			
1m<1y HA outcome	AOR	P>z	95% CI		AOR	P>z	95% CI		AOR	P>z	95% CI		AOR	P>z	95% CI		AOR	P>z	95% CI		AOR	P>z	95% CI		AOR	P>z	95% CI		AOR	P>z	95% CI		AOR	P>z	95% CI	
Age in days	*(FP = -0.5)*				*(FP = -0.5)*				*(FP = -0.5)*				*(FP = 1)*				*(FP = 1)*				*(FP = 1)*				*(FP = 1)*				*(FP = 1)*				*(FP = 1)*			
Fractional polynomial (FP) 1	6.19	<0.001	4.39	8.72	6.29	<0.001	3.94	10.04	9.36	<0.001	5.45	16.06	0.99	<0.001	0.99	0.99	0.99	<0.001	0.99	0.99	0.99	<0.001	0.99	0.99	0.99	<0.001	0.99	0.99	0.99	<0.001	0.99	0.99	0.99	<0.001	0.99	0.99
Postnatal stay	*(FP = -2)*				*(FP = -2)*				*(FP = -2)*				*(FP = 2,2)*				*(FP = 1)*				*(FP = 1)*				*(FP = 1)*				*(FP = 0)*				*(FP = 0)*			
Fractional polynomial (FP) 1	1.00	<0.001	1.00	1.00	1.00	<0.001	1.00	1.00	1.00	<0.001	1.00	1.00	0.22	0.007	0.07	0.66	0.99	0.090	0.98	1.00	0.99	0.083	0.98	1.00	0.99	0.095	0.98	1.00	0.77	0.014	0.63	0.95	0.76	0.011	0.62	0.94
Fractional polynomial (FP) 2													119.46	0.011	2.99	4775.93																				
Ethnicity																																				
White	1.00												1.00				1.00				1.00				1.00				1.00				1.00			
Mixed	1.26	0.457	0.68	2.35									0.35	0.100	0.10	1.22	0.29	0.047	0.08	0.99	0.25	0.033	0.07	0.89	0.25	0.037	0.07	0.92	0.24	0.033	0.06	0.89	0.23	0.034	0.06	0.90
Indian subcontinent	2.09	<0.001	1.42	3.08									2.23	0.026	1.10	4.53	2.67	0.008	1.29	5.51	2.55	0.014	1.20	5.39	2.30	0.032	1.07	4.91	2.70	0.013	1.24	5.88	2.65	0.023	1.14	6.15
Any other Asian	4.34	0.001	1.76	10.67									7.89	0.005	1.84	33.83	8.34	0.007	1.78	38.95	8.32	0.007	1.77	39.07	8.18	0.009	1.69	39.50	9.16	0.008	1.76	47.59	10.24	0.008	1.83	57.27
Black	2.62	0.001	1.50	4.57									3.90	0.005	1.50	10.12	2.85	0.049	1.00	8.08	2.98	0.041	1.04	8.49	2.82	0.060	0.96	8.31	1.46	0.543	0.43	4.90	1.10	0.887	0.31	3.84
other ethnic group	1.23	0.525	0.65	2.32									2.07	0.206	0.67	6.42	2.65	0.098	0.83	8.43	2.78	0.086	0.86	8.97	2.93	0.072	0.91	9.45	2.98	0.084	0.87	10.26	2.57	0.181	0.64	10.24
NK	1.21	0.360	0.80	1.82									2.01	0.039	1.04	3.88	2.10	0.035	1.05	4.19	2.03	0.050	1.00	4.10	1.95	0.071	0.94	4.04	2.19	0.043	1.02	4.66	2.34	0.034	1.07	5.15
normal birth													1.00				1.00				1.00				1.00				1.00				1.00			
Premature/low birthweight													2.20	0.004	1.28	3.79	2.09	0.006	1.23	3.54	2.16	0.005	1.26	3.71	2.23	0.004	1.29	3.85	2.83	0.001	1.50	5.33	3.11	0.001	1.59	6.11
Chronic conditions																																				
no chronic	1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00			
1 chronic	0.55	<0.001	0.43	0.71	1.08	0.667	0.75	1.56	1.27	0.279	0.82	1.97	1.28	0.334	0.78	2.11	1.35	0.274	0.79	2.33	1.33	0.311	0.76	2.33	1.33	0.319	0.76	2.35	1.33	0.355	0.73	2.45	1.44	0.275	0.75	2.75
2 chronic	0.73	0.094	0.50	1.06	2.31	<0.001	1.47	3.60	3.46	<0.001	2.09	5.72	3.39	<0.001	1.92	5.99	4.09	<0.001	2.24	7.48	3.80	<0.001	2.04	7.06	3.57	<0.001	1.90	6.72	4.16	<0.001	2.13	8.10	4.65	<0.001	2.30	9.38
>2 chronic	2.05	0.002	1.30	3.21	6.74	<0.001	4.01	11.30	10.52	<0.001	5.94	18.64	11.08	<0.001	5.81	21.13	13.91	<0.001	7.09	27.29	13.53	<0.001	6.79	26.94	13.49	<0.001	6.73	27.01	15.92	<0.001	7.64	33.15	18.11	<0.001	8.38	39.14
not recorded	1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00			
Congenital condition(s)	3.70	<0.001	2.68	5.11	6.97	<0.001	4.86	10.01	9.59	<0.001	6.43	14.30	11.03	<0.001	6.91	17.60	10.04	<0.001	6.18	16.30	9.91	<0.001	6.05	16.25	9.52	<0.001	5.79	15.66	10.23	<0.001	6.07	17.23	10.55	<0.001	6.12	18.19
No. of days since last indwelling device insertion																																				
no device	1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00			
91 days<1 yr	1.34	0.376	0.70	2.58	1.45	0.330	0.68	3.09	1.66	0.214	0.75	3.68	1.22	0.677	0.48	3.13	1.31	0.580	0.51	3.38	1.28	0.610	0.49	3.35	1.30	0.595	0.50	3.36	1.41	0.476	0.55	3.66	1.31	0.592	0.49	3.48
31-90 days	6.67	<0.001	2.29	19.39	8.68	<0.001	2.86	26.31	9.58	<0.001	3.01	30.47	11.27	<0.001	3.27	38.87	11.10	<0.001	3.21	38.33	10.75	<0.001	3.09	37.46	10.77	<0.001	3.10	37.36	13.96	<0.001	3.93	49.53	15.23	<0.001	4.18	55.40
7-30 days	15.57	0.008	2.08	116.74	24.27	0.002	3.22	182.96	29.71	0.001	3.91	225.96	30.42	0.001	3.83	241.42	31.91	0.001	3.98	255.98	37.80	0.001	4.60	310.40	41.13	0.001	4.93	343.28	47.24	<0.001	5.55	402.32	56.48	<0.001	6.47	492.95
1-6 days	0.97	0.936	0.46	2.03	1.63	0.235	0.73	3.64	1.92	0.136	0.81	4.55	2.19	0.096	0.87	5.50	2.15	0.119	0.82	5.63	2.69	0.051	1.00	7.27	2.63	0.063	0.95	7.31	2.81	0.056	0.97	8.09	2.26	0.137	0.77	6.63
No. of days since last invasive procedure																																				
no procedure	1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00			
91 days<1 yr	1.64	0.021	1.08	2.50	1.46	0.213	0.81	2.63	1.32	0.427	0.67	2.61	1.02	0.960	0.45	2.29	0.91	0.835	0.40	2.11	1.02	0.969	0.44	2.36	1.06	0.886	0.46	2.48	1.17	0.728	0.47	2.91	1.54	0.366	0.60	3.91
31-90 days	2.20	0.002	1.33	3.64	2.39	0.004	1.33	4.30	2.65	0.003	1.39	5.04	1.96	0.061	0.97	3.98	1.83	0.107	0.88	3.81	1.98	0.072	0.94	4.16	2.04	0.059	0.97	4.28	2.37	0.028	1.10	5.12	2.84	0.010	1.29	6.26
7-30 days	7.17	<0.001	3.74	13.76	11.45	<0.001	5.77	22.73	14.70	<0.001	7.16	30.21	15.56	<0.001	7.09	34.18	15.67	<0.001	7.09	34.65	16.92	<0.001	7.59	37.75	17.47	<0.001	7.83	38.94	19.45	<0.001	8.51	44.47	24.97	<0.001	10.54	59.12
1-6 days	4.37	<0.001	2.96	6.44	5.53	<0.001	3.52	8.71	5.82	<0.001	3.52	9.64	6.26	<0.001	3.54	11.09	6.44	<0.001	3.49	11.86	5.13	<0.001	2.70	9.77	4.58	<0.001	2.36	8.90	4.50	<0.001	2.23	9.10	5.49	<0.001	2.65	11.37
No. of days between last discharge and current admission																																				
no previous discharge													1.00				1.00				1.00				1.00				1.00				1.00			
91 days<1 yr													0.14	<0.001	0.08	0.27	0.12	<0.001	0.06	0.23	0.10	<0.001	0.05	0.21	0.10	<0.001	0.05	0.21	0.11	<0.001	0.05	0.25	0.07	<0.001	0.03	0.18
31-90 days													0.24	<0.001	0.13	0.45	0.21	<0.001	0.11	0.40	0.17	<0.001	0.09	0.34	0.18	<0.001	0.09	0.35	0.19	<0.001	0.09	0.39	0.16	<0.001	0.07	0.35
7-30 days													0.37	0.004	0.19	0.73	0.36	0.004	0.18	0.73	0.35	0.004	0.17	0.71	0.36	0.005	0.18	0.74	0.38	0.013	0.17	0.81	0.28	0.002	0.13	0.63
1-6 days													0.61	0.166	0.31	1.23	0.65	0.237	0.32	1.33	0.63	0.216	0.31	1.30	0.68	0.291	0.33	1.39	0.81	0.608	0.37	1.78	0.74	0.471	0.33	1.67

Appendix 5Di. Approach B: Adjusted OR results for explanatory variables in the final models (HA1-HA9) with HA BSI as the outcome in children aged 1 month < 1 year; the denominator contains either *a priori* CA or *defined* HA BSI cases

	No. of obs = 3,090				No. of obs = 2,399				No. of obs = 2,280				No. of obs = 2,234				No. of obs = 2,189				No. of obs = 2,160				No. of obs = 2,130				No. of obs = 2,105				No. of obs = 2,084			
Approach B	HA1				HA2				HA3				HA4				HA5				HA6				HA7				HA8				HA9			
1-5yrs HA outcome	AOR	P-value	95% CI		AOR	P>z	95% CI		AOR	P>z	95% CI		AOR	P>z	95% CI		AOR	P>z	95% CI		AOR	P>z	95% CI		AOR	P>z	95% CI		AOR	P>z	95% CI		AOR	P>z	95% CI	
Age in days	*(FP = -2,-1)*												*(FP = 1)*				*(FP = 1)*				*(FP = 1)*				*(FP = 1)*				*(FP = 1)*				*(FP = 1)*			
Fractional polynomial (FP) 1	0.64	<0.001	0.51	0.82									0.99	0.078	0.99	1.00	0.99	0.042	0.99	0.99	0.99	0.033	0.99	0.99	0.99	0.032	0.99	0.99	0.99	0.059	0.99	1.00	0.99	0.056	0.99	1.00
Fractional polynomial (FP) 2	4.14	<0.001	2.07	8.27																																
boy	1.00				1.00																															
girl	1.14	0.134	0.96	1.35	1.02	0.860	0.80	1.30																												
Ethnicity																																				
White	1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00			
Mixed	0.55	0.029	0.32	0.94	0.46	0.053	0.21	1.01	0.33	0.038	0.11	0.94	0.35	0.054	0.12	1.02	0.41	0.118	0.13	1.26	0.29	0.041	0.09	0.95	0.33	0.068	0.10	1.09	0.38	0.117	0.12	1.27	0.35	0.102	0.10	1.23
Indian subcontinent	1.39	0.028	1.04	1.87	1.52	0.044	1.01	2.27	1.89	0.023	1.09	3.28	1.83	0.045	1.01	3.31	1.95	0.042	1.02	3.71	1.96	0.049	1.00	3.82	1.97	0.057	0.98	3.96	2.17	0.032	1.07	4.41	2.08	0.047	1.01	4.30
Any other Asian	2.01	0.008	1.20	3.37	2.62	0.005	1.34	5.14	2.44	0.111	0.81	7.30	2.63	0.108	0.81	8.55	3.24	0.060	0.95	11.08	3.70	0.044	1.03	13.28	4.14	0.034	1.12	15.38	4.50	0.026	1.20	16.86	4.94	0.018	1.31	18.55
Black	1.12	0.521	0.79	1.61	1.32	0.253	0.82	2.13	1.21	0.563	0.63	2.33	1.37	0.359	0.70	2.69	1.57	0.215	0.77	3.21	1.62	0.205	0.77	3.42	1.39	0.422	0.62	3.09	1.25	0.604	0.54	2.91	1.06	0.900	0.43	2.59
other ethnic group	1.05	0.857	0.63	1.76	1.96	0.033	1.06	3.64	1.87	0.118	0.85	4.12	2.19	0.061	0.97	4.97	2.37	0.047	1.01	5.53	2.23	0.075	0.92	5.40	2.35	0.068	0.94	5.87	2.59	0.043	1.03	6.49	2.67	0.039	1.05	6.77
NK	1.10	0.541	0.82	1.48	1.60	0.017	1.09	2.35	2.10	0.005	1.25	3.52	1.84	0.033	1.05	3.24	1.94	0.031	1.06	3.53	1.80	0.074	0.94	3.41	2.03	0.033	1.06	3.89	1.99	0.046	1.01	3.90	1.80	0.100	0.89	3.64
normal birth					1.00				1.00				1.00																							
Premature/low birthweight					1.40	0.089	0.95	2.07	1.63	0.048	1.01	2.66	1.53	0.104	0.92	2.56																				
Chronic conditions																																				
no chronic	1.00								1.00				1.00				1.00				1.00				1.00				1.00				1.00			
1 chronic	0.91	0.332	0.74	1.11					3.05	<0.001	1.87	4.98	2.98	<0.001	1.75	5.10	4.29	<0.001	2.26	8.17	3.97	<0.001	1.97	7.98	4.51	<0.001	2.05	9.90	5.80	<0.001	2.44	13.79	6.93	<0.001	2.63	18.24
2 chronic	0.92	0.595	0.68	1.25					4.48	<0.001	2.58	7.79	4.92	<0.001	2.72	8.90	7.95	<0.001	3.95	16.00	9.18	<0.001	4.37	19.28	12.05	<0.001	5.32	27.30	13.95	<0.001	5.67	34.32	18.50	<0.001	6.87	49.86
>2 chronic	1.47	0.018	1.07	2.03					12.12	<0.001	7.14	20.57	13.27	<0.001	7.54	23.35	24.65	<0.001	12.61	48.20	28.25	<0.001	13.82	57.77	39.16	<0.001	17.71	86.59	49.55	<0.001	20.69	118.64	64.68	<0.001	24.64	169.77
not recorded	1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00			
Congenital condition(s)	2.44	<0.001	1.77	3.36	6.87	<0.001	4.83	9.77	3.11	<0.001	2.04	4.74	3.32	<0.001	2.15	5.12	3.68	<0.001	2.28	5.93	3.54	<0.001	2.16	5.80	3.59	<0.001	2.16	5.95	3.76	<0.001	2.24	6.30	3.79	<0.001	2.24	6.41
No. of days since last indwelling device insertion																																				
no device	1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00				1.00			
91 days<1 yr	1.51	0.102	0.92	2.49	4.80	<0.001	2.90	7.95	1.66	0.094	0.92	3.01	1.54	0.169	0.83	2.83	1.65	0.136	0.85	3.19	1.66	0.142	0.84	3.24	1.70	0.130	0.86	3.38	1.97	0.057	0.98	3.97	2.15	0.034	1.06	4.36
31-90 days	3.48	0.001	1.71	7.07	18.72	<0.001	9.50	36.87	4.72	<0.001	2.16	10.31	5.07	<0.001	2.29	11.23	6.01	<0.001	2.62	13.77	6.20	<0.001	2.68	14.32	6.61	<0.001	2.79	15.66	7.90	<0.001	3.27	19.11	8.58	<0.001	3.54	20.79
7-30 days	7.82	<0.001	3.28	18.66	75.06	<0.001	32.24	174.74	9.67	<0.001	3.91	23.89	10.03	<0.001	4.05	24.83	11.67	<0.001	4.68	29.10	12.28	<0.001	4.89	30.84	13.61	<0.001	5.38	34.40	15.49	<0.001	6.08	39.43	18.57	<0.001	7.21	47.81
1-6 days	2.82	0.008	1.31	6.09	28.40	<0.001	13.50	59.76	3.06	0.011	1.29	7.27	3.31	0.008	1.37	7.97	3.82	0.004	1.52	9.60	3.97	0.005	1.52	10.32	3.98	0.006	1.49	10.62	4.03	0.007	1.46	11.10	3.56	0.017	1.26	10.10
No. of days since last invasive procedure																																				
no procedure	1.00								1.00				1.00				1.00				1.00				1.00				1.00				1.00			
91 days<1 yr	1.22	0.335	0.82	1.81					1.65	0.100	0.91	2.98	1.86	0.048	1.00	3.46	2.74	0.006	1.34	5.61	3.29	0.001	1.58	6.85	3.21	0.003	1.49	6.90	3.12	0.005	1.42	6.87	3.10	0.006	1.39	6.94
31-90 days	1.87	0.012	1.15	3.04					4.28	<0.001	2.27	8.07	4.31	<0.001	2.21	8.42	5.59	<0.001	2.64	11.81	6.80	<0.001	3.15	14.66	6.72	<0.001	3.03	14.92	6.48	<0.001	2.86	14.71	6.52	<0.001	2.82	15.08
7-30 days	5.89	<0.001	3.74	9.27					22.74	<0.001	13.45	38.45	26.51	<0.001	15.44	45.51	36.24	<0.001	19.76	66.46	44.09	<0.001	23.46	82.86	49.19	<0.001	25.57	94.63	52.01	<0.001	26.68	101.39	44.37	<0.001	22.41	87.84
1-6 days	5.82	<0.001	4.12	8.23					19.41	<0.001	12.34	30.55	19.49	<0.001	12.10	31.39	20.93	<0.001	12.31	35.61	23.06	<0.001	13.16	40.43	21.65	<0.001	11.90	39.39	18.14	<0.001	9.67	34.03	16.35	<0.001	8.39	31.88
No. of days between last discharge and current admission																																				
no previous discharge	1.00				1.00												1.00				1.00				1.00				1.00				1.00			
91 days<1 yr	1.46	0.010	1.10	1.95	1.24	0.277	0.84	1.83									0.55	0.059	0.29	1.02	0.54	0.065	0.28	1.04	0.51	0.052	0.26	1.01	0.47	0.041	0.23	0.97	0.48	0.051	0.23	1.00
31-90 days	1.48	0.037	1.02	2.14	1.53	0.081	0.95	2.45									0.47	0.039	0.23	0.96	0.44	0.034	0.21	0.94	0.44	0.040	0.20	0.97	0.43	0.037	0.19	0.95	0.38	0.024	0.16	0.88
7-30 days	2.07	<0.001	1.43	3.00	2.78	<0.001	1.78	4.33									0.41	0.009	0.21	0.80	0.33	0.002	0.16	0.66	0.29	0.001	0.14	0.60	0.24	<0.001	0.11	0.51	0.23	<0.001	0.11	0.50
1-6 days	1.78	<0.001	1.30	2.43	4.19	<0.001	2.95	5.94									0.67	0.170	0.38	1.19	0.63	0.131	0.34	1.15	0.48	0.025	0.26	0.91	0.41	0.009	0.21	0.80	0.41	0.011	0.21	0.82

Appendix 5Dii. Approach B: Adjusted OR results for explanatory variables in the final models (HA1-HA9) with HA BSI as the outcome in children aged 1-5 years; the denominator contains either *a priori* CA or *defined* HA BSI cases

Appendix Chapter 6: Peer-reviewed article in Epidemiology and Infection (2015)

Epidemiol. Infect. (2015), **143**, 2440–2445. © Cambridge University Press 2014
doi:10.1017/S0950268814003306

Timing of positive blood samples does not differentiate pathogens causing healthcare-associated from community-acquired bloodstream infections in children in England: a linked retrospective cohort study

K. L. HENDERSON[1,2]*, B. MÜLLER-PEBODY[1], A. WADE[2], M. SHARLAND[3], M. MINAJI, A. P. JOHNSON[1] AND R GILBERT[2]

[1] *Department of Healthcare-Associated Infection and Antimicrobial Resistance, Centre for Infectious Disease Surveillance and Control, Public Health England, London, UK*
[2] *Paediatric Epidemiology and Biostatistics, UCL Institute of Child Health, London, UK*
[3] *Paediatric Infectious Diseases Unit, St George's Hospital, London, UK*

Received 13 August 2014; Final revision 13 November 2014; Accepted 13 November 2014; first published online 8 December 2014

SUMMARY

Paediatricians recognize that using the time-dependent community-acquired *vs.* hospital-acquired bloodstream infection (BSI) dichotomy to guide empirical treatment no longer distinguishes between causative pathogens due to the emergence of healthcare-associated BSIs. However, paediatric epidemiological evidence of the aetiology of BSIs in relation to hospital admission in England is lacking. For 12 common BSI-causing pathogens in England, timing of laboratory reports of positive paediatric (3 months to 5 years) bacterial blood isolates were linked to in-patient hospital data and plotted in relation to hospital admission. The majority (88·6%) of linked pathogens were isolated <2 days after hospital admission, including pathogens widely regarded as hospital acquired: *Enterococcus* spp. (67·2%) and *Klebsiella* spp. (88·9%). *Neisseria meningitidis, Streptococcus pneumoniae*, group A streptococcus and *Salmonella* spp. were unlikely to cause hospital-acquired BSI. Pathogens commonly associated with hospital-acquired BSI are being isolated <2 days after hospital admission alongside pathogens commonly associated with community-acquired BSI. We confirm that timing of blood samples alone does not differentiate between bacterial pathogens. Additional factors including clinical patient characteristics and healthcare contact should be considered to help predict the causative pathogen and guide empirical antibiotic therapy.

Key words: Bloodstream infections, hospital-acquired (nosocomial) infections, microbiology, paediatrics.

INTRODUCTION

There is increasing evidence that the traditional categorization of bloodstream infections (BSIs) in children as either community acquired (CA) or hospital acquired (HA) is no longer adequate to guide empirical treatment of sepsis, as patients presenting from the community may have infections that are actually healthcare-associated (HCA) and caused by pathogens acquired in a healthcare setting, rather than the community [1]. Since the 1980s, a 48-h post-admission threshold for CA- *vs.* HA-BSI has commonly been used to guide empirical antibiotic treatment based on the different pathogens circulating in the community *vs.* the hospital environment [2–9]. However,

* Author for correspondence: Miss K. L. Henderson, Department of HCAI & AMR, Public Health England, 61 Colindale Avenue, London, NW9 5EQ, UK.
(Email: Katherine.Henderson@phe.gov.uk)

This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/3.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

the epidemiology of paediatric BSIs has since evolved due to medical advances, shorter length of hospital stay and the ability to manage complex conditions, including those requiring the use of invasive devices, at home. As a result, children presenting to hospital with systemic infections may have either a CA- or HCA-BSI, with the latter more likely to be caused by pathogens commonly associated with HA-BSI, which require different empirical therapy to CA-BSI [1, 10–12]. Currently, the bacterial aetiology of paediatric BSIs in relation to hospital admission in England is unknown and evidence to guide the formulation of recommendations for treatment of paediatric BSIs is lacking.

A recent systematic review found only two regional studies, from Canada and England, that reported the proportions of paediatric HCA-BSI since 2000 [13]. To address this gap in our understanding of the aetiology and epidemiology of paediatric BSIs in England, we used national data to analyse the timing of positive blood samples from children in relation to their hospital admission for 12 common pathogens causing BSI.

METHODS

Linked national laboratory and clinical data

We used probabilistic methods to link retrospective national BSI microbiology surveillance data and hospital administrative data (Hospital Episode Statistics; HES) for children aged 3 months to 5 years admitted to a National Health Service (NHS) hospital or treated by the NHS in a private hospital in England between 1 April 2009 and 31 March 2010. Children aged 3 months to 5 years have the second highest incidence rate of paediatric BSI after neonates [14]. Neonates aged <3 months were excluded to avoid BSIs resulting from vertically transmitted pathogens and HA-BSI acquired by babies in intensive-care units since birth. The microbiology surveillance data were voluntarily submitted to Public Health England's (PHE) database (LabBase2) by laboratories across England. LabBase2 captures data from a substantial proportion of all positive blood specimens (for example, 80% of cases of BSI caused by methicillin-resistant *Staphylococcus aureus* (MRSA) reported via the national mandatory reporting scheme are also reported to LabBase2 [15].

The following patient identifiers were used to link the datasets: NHS number, hospital number, date of birth, sex and postcode. Match-weights for the probabilistic linkage were calculated using the following equation:

$$\text{Match-weight} = \log_2(m/u).$$

A standard reference dataset of linked LabBase2 and HES data from 2009 to 2010 considered to be true links was used to calculate the match-weights. The frequency of match-pairs that agreed, disagreed or were missing particular identifiers from this reference dataset were summed for each patient identifier (m value; matched). The true links match-pairs were then internally cross-matched to create all possible false-match combinations from the standard reference dataset to calculate the frequency counts of match-pairs that agreed, disagreed or were missing values by chance (u value; unmatched). The match-weight threshold for inclusion was selected following manual review of ≤100 randomly sampled match-pair observations per match-weight above the value of zero by K.L.H., and the highest match-weight for each laboratory blood specimen report was kept above this threshold. A χ^2 test was used to assess whether the unlinked and linked datasets were significantly different by gender, age and pathogen type to identify any linkage bias.

The objective of the analysis was to assess the association between the bacterial aetiology of BSI and the timing of positive blood specimens taken before and during admission to hospital. Following linkage, BSI reports were selected for inclusion in the analysis where the pathogen was isolated between 5 days before and up to 30 days after hospital admission. Only the first reported positive blood specimen per patient was included in the analysis and we focused on 12 frequently isolated pathogens within this age group: *S. aureus* [split into MRSA and methicillin-susceptible *S. aureus* (MSSA) where a methicillin susceptibility result was reported], non-pyogenic streptococci, *Enterococcus* spp., group B streptococcus (GBS; *Streptococcus agalactiae*), *Streptococcus pneumoniae*, group A streptococcus (GAS; *Streptococcus pyogenes*), *Escherichia coli, Klebsiella* spp., *Enterobacter* spp., *Pseudomonas aeruginosa, Salmonella* spp., and *Neisseria meningitidis*.

It should be noted that the exact time in hours of a blood specimen or the admission to hospital were not reported in either dataset; therefore 'day' is the closest approximation and may introduce a day-drift effect. For example two patients may be admitted on the same day, but one at 01:00 hours and other at 23:00 hours, leaving a 22-h difference. Patients may also

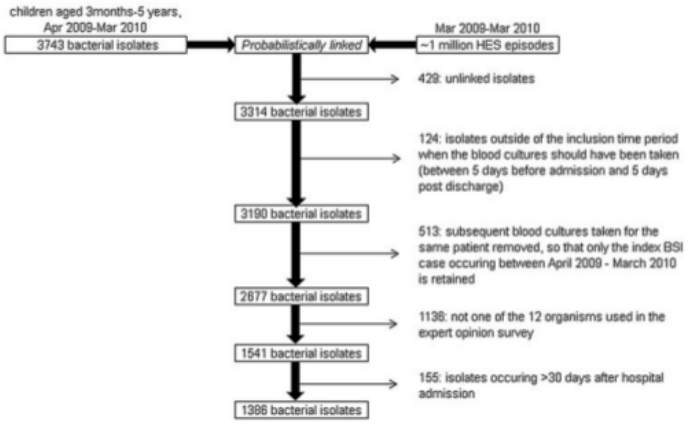


Fig. 1. Flow diagram illustrating the inclusion criteria for the linked national laboratory and clinical data. HES, Hospital Episode Statistics

have gone to an accident and emergency (A&E) department or NHS walk-in centre before being admitted to hospital as an in-patient. Records for A&E visits are held in a separate database that was not consulted for this study; however, we used data on the source of admission to explain positive blood specimens before admission to hospital. Isolates taken $\geqslant 2$ days after admission are consistent with the HA-BSI definition using a 48-h cut-off most widely used in the published literature [13].

Ethics approval

Public Health England has National Information Governance Board for Health and Social Care (NIGB) approval for the collation of surveillance data in accordance with section 251 of the NHS Act 2006.

RESULTS

Linkage of national laboratory and clinical data

There were 3743 reports of positive blood samples from children aged 3 months to 5 years during the study period, of which 89% ($n = 3314$) linked to HES records (Fig. 1); patient identifier completion was poor in the unlinked records (21% lacked NHS number, 31% lacked postcode). The type of pathogen was significantly different between the linked and unlinked datasets (χ^2 $P = 0{\cdot}001$). There was a higher proportion of GAS (8·2% *vs.* 3·9%), *N. meningitidis* (7·2% *vs.* 4·7%) and GBS (1·6% *vs.* 0·3%) in the unlinked compared to the linked datasets, and a lower proportion of *S. pneumoniae* (5·6% *vs.* 8·1%) and non-pyogenic streptococci (5·6% *vs.* 8·1%) in the unlinked dataset compared to the linked dataset; age and gender distributions were not significantly different. Of these, 124 isolates were removed if they did not occur between 5 days before and 5 days after hospital admission, and 513 subsequent blood specimens were removed so that only the index BSI case per child remained, leaving a total of 2677 isolates.

A total of 1386 index BSI records for the 12 common pathogens were reported for cases where the pathogen was isolated from blood between 5 days before and 30 days after admission to hospital. Of these, 57% were male and the median age was 1·73 years.

Characteristics of BSI caused by different pathogens

The number of positive blood samples together with their timing in relation to hospital admission is shown for each pathogen in Figure 2. Seventy (5%) patients had a positive blood specimen 2 days ($n = 68$) or >2 days ($n = 2$) before their admission to hospital. Of the 68 cases, 55 were admitted as emergency cases to hospital from the A&E department, seven were admitted as emergency cases from other sources including the A&E department of other

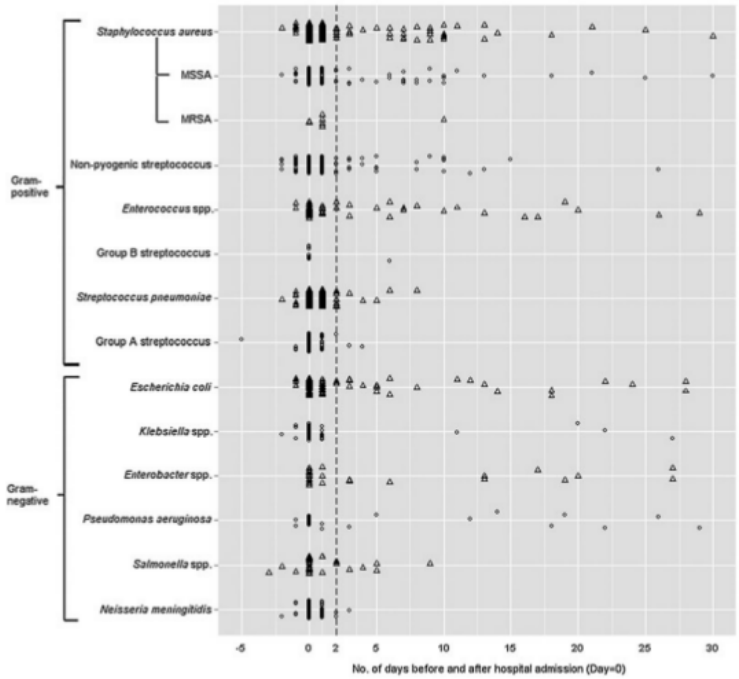


Fig. 2. Distribution of the timing of positive blood specimens in relation to hospital admission (day = 0) for the 12 most frequently reported pathogen groups. The dotted line illustrates 2 days threshold after hospital admission.

healthcare providers, three were emergency cases from general practitioners (GPs), one case was transferred from another hospital, one case had a planned elective admission and one case was admitted from an outpatient clinic. For the remaining two cases, the case whose blood specimen was taken on day 3 was admitted from the A&E department; the case whose blood specimen was taken 5 days before admission was an emergency GP admission.

Table 1 reports that 88·6% [95% confidence interval (CI) 88·5–88·7] of all 12 pathogen groups (range 60–98%), were isolated <2 days after hospital admission. Grouped by specific pathogens, 67·2% (95% CI 46·6–84·2) of *Enterococcus* spp., 88·9% (95% CI 49·0–100) of *Klebsiella* spp. and 64·0% (95% CI 0–100) of *P. aeruginosa*, and over 90% of all streptococci were isolated <2 days after hospital admission. Ninety-five per cent of *S. pneumoniae*, GAS, *N. meningitidis, Salmonella* spp. and GBS were isolated <2 days after hospital admission. Full details are shown for all pathogens causing BSIs between 5 days before and 30 days after hospital admission ($n = 2435$) in Supplementary Appendix A (online).

DISCUSSION

Almost 90% of BSIs for 12 common pathogens in children aged 3 months to 5 years were isolated from samples taken <2 days after hospital admission. As expected, this proportion was even higher for pathogens that are viewed as predominantly community acquired (*S. pneumoniae*, GAS, *Salmonella* spp., *N. meningitidis*). The finding that over two-thirds of pathogens (such as *Enterococcus* spp., *Klebsiella* spp., *P. aeruginosa*), which are usually considered to be hospital acquired, were isolated <2 days after hospital admission, indicates that children presenting with

Table 1. *The proportional distribution by time of 12 pathogen groups of 1386 positive bacterial isolates between 5 days before and 30 days after hospital admission in children aged 3 months to 5 years, England*

			Days either side of admission to hospital (day = 0)					
Gram	Pathogen	Total no. reported	−5 d to +1 d	(95% CI)	+2 to +5 d	(95% CI)	+6 to +30 d	(95% CI)
Positive	*Staphylococcus aureus*	227	83%	(81–85)	5%	(4–6)	12%	(11–15)
	MSSA	187	85%	(82–88)	5%	(4–7)	10%	(8–13)
	MRSA	7	86%	(0–100)	0%	(0–100)	14%	(0–100)
	Non-pyogenic streptococcus†	331	92%	(91–93)	5%	(4–6)	3%	(3–4)
	Enterococcus spp.	61	67%	(47–84)	8%	(0–24)	25%	(10–44)
	Group B streptococcus	8	88%	(0–100)	0%	(0–100)	13%	(0–100)
	Streptococcus pneumoniae	253	95%	(94–96)	4%	(3–5)	1%	(1–2)
	Group A streptococcus	123	98%	(94–99)	2%	(1–6)	0%	(0–3)
Negative	*Escherichia coli*	113	80%	(72–95)	8%	(5–14)	12%	(9–19)
	Klebsiella spp.	36	89%	(49–100)	0%	(0–27)	11%	(0–50)
	Enterobacter spp.	25	60%	(0–100)	8%	(0–76)	32%	(0–100)
	Pseudomonas aeruginosa	25	64%	(0–100)	8%	(0–76)	28%	(0–100)
	Salmonella spp.	33	79%	(28–100)	18%	(0–67)	3%	(0–40)
	Neisseria meningitidis	151	98%	(96–99)	2%	(1–4)	0%	(0–2)
	Total*	1386	89%	(89–89)	**5%**	(4–6)	**7%**	(7–7)

CI, Confidence interval.

* The total does not double count the *Staphylococcus aureus* isolates tested for methicillin susceptibility [methicillin-susceptible *S. aureus* (MSSA), methicillin-resistant *S. aureus* (MRSA)].

† Non-pyogenic streptococci includes: *Abiotrophia defectiva, Aerococcus* spp., *S. acidominimus, S. alactolyticus, S. anginosus, S. bovis, S. constellatus, S. dysgalactiae, S. gordonii*, group C streptococcus, group D streptococcus, *S. intermedius, S. milleri* group, *S. mitior, S. mitis, S. mutans, S. oralis, S. parasanguinis, S. salivarius, S. sanguinis, S. sobrinus, S. uberis, S. vestibularis, S. viridans*, other *Streptococcus* spp.

BSIs to hospital are likely to reflect a mixture of CA- and HCA-BSIs [13].

The overall high proportion of BSIs that were detected <2 days after admission is similar to the 81% proportion of CA- and HCA-BSI reported by Laupland *et al.* for children aged 30 days to <18 years in Canada [5]. Evidence is lacking on the timing of positive samples of BSI by pathogen in relation to hospital admission for children in England; an English, one-centre study reported BSIs in children by pathogen but only distinguished between children with and without serious underlying medical conditions rather than CA- or HA-BSI due to a recognition of the changing epidemiology of pathogens causing BSIs and the emergence of HCA infections [16].

The data linkage results, along with our previous review of the paediatric literature and other studies in adult patients argue that the timing of positive blood specimens in relation to hospital admission is insufficient on its own to differentiate CA- from HCA-BSI in order to inform empirical antibiotic therapy [10, 13]. Further analysis of these linked data is needed to develop statistical models to identify patient and clinical characteristics such as age, previous healthcare contact, invasive procedures and underlying chronic conditions that may help to differentiate CA-, HCA- and HA-BSI and to guide empirical treatment.

SUPPLEMENTARY MATERIAL

For supplementary material accompanying this paper visit http://dx.doi.org/10.1017/S0950268814003306.

ACKNOWLEDGEMENTS

No external funding was received for this study and it was completed within Public Health England's Healthcare Associated Infection & Antimicrobial Resistance departmental resources.

This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

HES data. Copyright © 2012, re-used with the permission of The Health and Social Care Information Centre. All rights reserved.

DECLARATION OF INTEREST

None.

REFERENCES

1. **Friedman ND, *et al.*** Health care-associated bloodstream infections in adults: a reason to change the accepted definition of community-acquired infections. *Annals of Internal Medicine* 2002; **137**: 791–797.
2. **Garner JS, *et al.*** CDC definitions for nosocomial infections, 1988. *American Journal of Infection Control* 1988; **16**: 128–140.
3. **Hounsom L, Grayson K, Melzer M.** Mortality and associated risk factors in consecutive patients admitted to a UK NHS trust with community acquired bacteraemia. *Postgraduate Medical Journal* 2011; **87**: 757–762.
4. **Raymond NJ, *et al.*** Bloodstream infections in a secondary and tertiary care hospital setting. *Internal Medicine Journal* 2006; **36**: 765–772.
5. **Laupland KB, *et al.*** The changing burden of pediatric bloodstream infections in Calgary, Canada, 2000–2006. *Pediatric Infectious Disease Journal* 2009; **28**: 114–117.
6. **Marchaim D, *et al.*** Epidemiology of bacteremia episodes in a single center: increase in Gram-negative isolates, antibiotics resistance, and patient age. *European Journal of Clinical Microbiology and Infectious Diseases* 2008; **27**: 1045–1051.
7. **Gray J, Gossain S, Morris K.** Three-year survey of bacteremia and fungemia in a pediatric intensive care unit. *Pediatric Infectious Disease Journal* 2001; **20**: 416–421.
8. **Jensen AG, *et al.*** A 6-month prospective study of hospital-acquired bacteremia in Copenhagen county. *Scandinavian Journal of Infectious Diseases* 1996; **28**: 601–608.
9. **Ashkenazi S, *et al.*** Childhood bacteremia in Israel: causes, age relation, predisposing factors and source. *Israel Journal of Medical Sciences* 1994; **30**: 610–616.
10. **Valles J, *et al.*** Bloodstream infections in adults: importance of healthcare-associated infections. *Journal of Infection* 2008; **56**: 27–34.
11. **Kollef MH, *et al.*** Epidemiology, microbiology and outcomes of healthcare-associated and community-acquired bacteremia: a multicenter cohort study. *Journal of Infection* 2011; **62**: 130–135.
12. **Diekema DJ, *et al.*** Epidemiology and outcome of nosocomial and community-onset bloodstream infection. *Journal of Clinical Microbiology* 2003; **41**: 3655–3660.
13. **Henderson KL, *et al.*** Community-acquired, healthcare-associated and hospital-acquired bloodstream infection definitions in children: a systematic review demonstrating inconsistent criteria. *Journal of Hospital Infection* 2013; **85**: 94–105.
14. **Henderson KL, *et al.*** The changing aetiology of paediatric bacteraemia in England and Wales, 1998–2007. *Journal of Medical Microbiology* 2010; **59**: 213–219.
15. **Pearson A, Chronias A, Murray M.** Voluntary and mandatory surveillance for methicillin-resistant Staphylococcus aureus (MRSA) and methicillin-susceptible *S. aureus* (MSSA) bacteraemia in England. *Journal of Antimicrobial Chemotherapy* 2009; **64** (Suppl. 1): i11–i17.
16. **Gray JW.** A 7-year study of bloodstream infections in an English children's hospital. *European Journal of Pediatrics* 2004; **163**: 530–535.

Appendix Chapter 7: Conference presentations

I presented work from my thesis at a variety of national and international conferences.

Oral presentations

Henderson KL, Minaji M, Muller-Pebody B, Wade A, Gilbert R. *How important are match-weights for probabilistic linkage?* ScottisH Informatics Programme, St. Andrew's, Scotland; August 2013

Henderson KL, Minaji M, Wade A, Gilbert R, Muller-Pebody B. *Linking HPA and Hospital Episode Statistics paediatric data: Methodology considerations*. Health Protection Agency Conference, Warwick, England; September 2012

Poster presentations

Henderson KL, Wade A, Gilbert R, Muller-Pebody B. *Estimating the relative proportion of community-acquired, healthcare-associated and hospital-acquired bloodstream infections in children in England from linked hospital and laboratory data*. European Scientific Conference on Applied Infectious Disease Epidemiology, Stockholm, Sweden; November 2015.

Henderson KL, Muller-Pebody B, Wade A, Sharland, M, Minaji M, Johnson AP, Gilbert R. *National data and expert opinions highlight the epidemiological change of invasive organisms causing bloodstream infections in children in England*. Public Health England Post-Graduate Poster Day, PHE National Institute for Biological Standards and Control, England; November 2013

Henderson KL, Minaji M, Muller-Pebody B, Johnson AP, Wade A, Sharland M, Gilbert R. *Complex timing of bloodstream infections in relation to inpatient hospital admission date in children in England, 2010-11.* European Society for Paediatric Infectious Diseases, Thessaloniki, Greece; May 2013

Seminar presentation

Henderson KL, Minaji M, Muller-Pebody B, Wade A, Gilbert R. *Paediatric bacteraemia: when, where and what pathogen? Calculating match-weights for the probabilistic linkage of laboratory & clinical data to investigate the current epidemiology*. London Farr Institute, England; May 2014