

1 **Sepsis Recording in Primary Care Electronic Health Records, Linked Hospital**
2 **Episodes and Mortality Records: Population-based Cohort Study in England**

3

4 **Emma Rezel-Potts PhD^{1,2*} and Martin C. Gulliford FFPH^{1,2} and the SafeAB Study Group**

5 **¹ King's College London, School of Population Health and Environmental Sciences,**
6 **London, UK;**

7 **² National Institute for Health Research Biomedical Research Centre (BRC), Guy's and**
8 **St Thomas' NHS Foundation Trust and King's College London, London, UK**

9

10 **Correspondence: Dr Emma Rezel-Potts**
11 **Addison House, Guy's Campus,**
12 **King's College London,**
13 **London SE1 1UL**
14 **Email: emma.rezel-potts@kcl.ac.uk**
15 **Tel: 0207 848 6631**
16 **Fax: 0207 848 6620**

17

18

19

20 **Word count: Text: 2,926 words**
21 **Abstract: 290 words**
22 **Tables: 2**
23 **Figures: 2**

24

25

26 **Abstract**

27

28 **Objectives:** Sepsis is a growing concern for health systems, but the epidemiology of sepsis
29 is poorly characterised. We evaluated sepsis recording across primary care electronic
30 records, hospital episodes and mortality registrations.

31 **Methods and Findings:** Cohort study including 378 general practices in England from
32 Clinical Practice Research Datalink (CPRD) GOLD database from 2002 to 2017 with
33 36,209,676 patient-years of follow-up with linked Hospital Episode Statistics (HES) and
34 Office for National Statistics (ONS) mortality registrations. Incident sepsis episodes were
35 identified for each source. Concurrent records from different sources were identified and
36 age-standardised and age-specific incidence rates compared. Logistic regression analysis
37 evaluated associations of gender, age-group, fifth of deprivation and period of diagnosis with
38 concurrent sepsis recording.

39 There were 20,206 first episodes of sepsis from primary care, 20,278 from HES and 13,972
40 from ONS. There were 4,117 (20%) first HES sepsis events and 2,438 (17%) mortality
41 records concurrent with incident primary care sepsis records within 30 days. Concurrent
42 HES and primary care records of sepsis within 30 days before or after first diagnosis were
43 higher at younger or older ages and for patients with the most recent period of diagnosis with
44 those diagnosed during 2007:2011 less likely to have a concurrent HES record given CPRD
45 compared to those diagnosed during 2012 to 2017 (odd ratio 0.65, 95% confidence interval
46 0.60 to 0.70). At age 85 and older, primary care incidence was 5.22 per 1,000 patient years
47 (95% CI 1.75 to 11.97) in men and 3.55 (0.87 to 9.58) in women which increased to 10.09
48 (4.86 to 18.51) for men and 7.22 (2.96 to 14.72) for women after inclusion of all three
49 sources.

50 **Conclusion:** Explicit recording of sepsis is inconsistent across healthcare sectors with a
51 high proportion of non-concurrent records. Incidence estimates are higher when linked data
52 are analysed.

53

54

55 **Keywords:** electronic health records; data linkage; sepsis; septicaemia; primary care

56 **Introduction**

57 Sepsis is a growing concern for health systems. In the UK, sepsis is estimated to account for
58 36,900 deaths and 123,000 hospital admissions annually [1]. The Global Burden of Disease
59 Study estimated that there were nearly 50 million incident cases of sepsis worldwide in 2017,
60 with 11 million deaths representing 19.7% of global deaths [2]. The term sepsis was
61 introduced by ancient Greek physicians, but only in recent years has sepsis come to be
62 defined as a syndrome resulting from the interaction between an acute infection and host
63 response leading to new organ dysfunction [3]. Sepsis is an intermediate state that links an
64 infection, or an infection-causing condition, to adverse health outcomes. The term sepsis is
65 now more commonly used than the term 'septicaemia', which refers to blood-stream
66 infection. In the health care systems of high-income countries, records of 'sepsis' have been
67 increasing in both hospital and primary care settings [4-6]. A study from the U.S.
68 Massachusetts General Hospital [7] found that recording of severe sepsis or septic shock
69 increased by 706% in the decade between 2003 and 2012, while objective markers of
70 severe infection, including positive blood cultures, remained stable or decreased. In a large
71 study from UK primary care the incidence of sepsis diagnoses increased throughout the
72 period from 2002 to 2017, with an especially rapid increase since 2011 [6]. Alongside
73 increasing use of the term sepsis, case definitions have expanded to include patients with
74 evidence of both acute infection and acute organ dysfunction as having 'implicit sepsis' even
75 when sepsis was not explicitly diagnosed [8, 2].

76
77 Accurate surveillance of sepsis and other bacterial infections is of importance in the context
78 of antimicrobial resistance which can limit options for effective treatment [9]. Antimicrobial
79 stewardship is a strategy to optimise antibiotic prescribing practices to prevent increasing
80 antimicrobial resistance, aiming to strike a balance between effective management of
81 suspected infection and avoidance of inappropriate or unnecessary antibiotic use [10]. The
82 recent increase in sepsis has been accompanied by heightened awareness of antimicrobial

83 resistance and antimicrobial stewardship, raising safety concerns about the potential for
84 increased rates of serious bacterial infections if antibiotics are not used when needed [11].

85

86 Electronic health records provide an important data resource for epidemiological research
87 and health surveillance, especially in health care settings. Data linkage provides
88 opportunities to enhance the completeness of ascertainment of health events across health
89 service sectors and population health registries. The advantages of linked records for case
90 ascertainment have been demonstrated for long-term conditions [12-15], but research into
91 the use of linked records for the evaluation of infectious diseases has been limited. However,
92 studies investigating the incidence of community-acquired pneumonia indicate that primary
93 care data alone may lead to under-estimation of the burden of infections [16, 17].

94

95 This study aimed to exploit data linkage to evaluate the recording of sepsis across primary
96 care electronic health records, hospital episodes and mortality registrations for individual
97 patients registered at general practices in England. We conducted a population-based cohort
98 study to compare simultaneous recording of sepsis in primary care, hospital episodes and
99 mortality data and to estimate the incidence of sepsis from different data sources.

100

101

102 **Methods**

103

104 **Study Population & Data Sources**

105 The study employed the UK Clinical Practice Research Datalink (CPRD) GOLD database.
106 The CPRD GOLD is a primary care database of anonymised electronic health records for
107 general practices in the UK. The high quality of CPRD GOLD data is well-established [18].
108 CPRD GOLD has a coverage of some 11.3 million patients, including approximately 7% of
109 UK population, of which it is broadly representative in terms of age and sex [19]. Consenting
110 practices in England participate in a data linkage scheme [20]. Approximately 74% of all
111 CPRD GOLD practices in England are eligible for linkage. Linkages are available for the
112 Hospital Episode Statistics (HES) and mortality registration data from the Office for National
113 Statistics (ONS). HES admitted patient care data include admission and discharge dates and
114 diagnostic data coded using the International Classification of Diseases 10th revision
115 (ICD10). Mortality registration data include information for the date and causes of death
116 coded using ICD10. ONS identifies one underlying cause of death and secondary causes of
117 death including up to 15 additional causes of death. Linked area-based measures of
118 deprivation include the Indices of Multiple Deprivation (IMD) based on a weighted profile of
119 indicators [21]. We employed deprivation for the general practice postcode for this study
120 because of the low proportion of missing values. The protocol was approved by the CPRD
121 Independent Scientific Advisory Committee (ISAC protocol 18-041R).

122

123 **Main measures**

124 We included patient records between 1st January 2002 to 31st December 2017. The start of
125 the patient record was the later of the patient registration date or the date the general
126 practice joined CPRD. The end of the patient record was the earliest of the last data
127 collection date, the end of registration and the date of death. We evaluated first records of
128 sepsis more than 12 months after start of registration in primary care electronic health

129 records, or as a primary diagnosis in HES, or sepsis as any mentioned cause of death in
130 mortality records. In UK primary care records, diagnoses recorded at consultations or
131 referrals to or from hospitals were coded, at the time of this study, using Read codes. We
132 identified sepsis records using a list of 77 eligible Read codes. Incident episodes of sepsis in
133 CPRD were recorded using 55 Read codes with four codes accounting for 92% of events,
134 including 'Sepsis' (two codes) (64%), 'Septicaemia' (18%), and 'Urosepsis' (10%). In HES
135 and death registry records, sepsis diagnoses and sepsis deaths were defined using 23
136 ICD10 codes for sepsis. In HES records we evaluated the primary diagnosis, which accounts
137 for the majority of the length of stay of the episode, with other diagnoses being referred to as
138 comorbidities [22]. Incident diagnoses of sepsis in HES were coded with 20 ICD10 codes
139 with three codes accounting for 89% of events, including 'Sepsis, unspecified' (72%),
140 'Sepsis due to other Gram-negative organisms' (13%) and
141 'Sepsis due to Staphylococcus aureus' (5%). In mortality data, we included all mentioned
142 causes of death because sepsis may be part of a sequence of morbid events and not always
143 an underlying cause of death [23]. 'Sepsis, unspecified' accounted for 93% of causes of
144 death among those in the ONS death registry with sepsis as any mentioned cause of death.

145

146 **Analysis**

147 Incident sepsis events were identified for each data source. We calculated person-time at
148 risk from the start to the end of the patient record. Person-time was grouped by gender and
149 age-group from zero to four, five to nine and 10 to 14 and then 10 years age groups up to 85
150 years and over. Incidence and mortality rates were age-standardised using the European
151 standard population for reference. We searched for concurrent events across data sources
152 using a 30-day time-window. We calculated age-specific incidence rates using primary care
153 EHRs and then adding HES records, mortality records or both. Finally, we fitted a logistic
154 regression model to evaluate associations of gender, age-group, fifth of deprivation and

155 period of diagnosis with concurrent sepsis recording. All data were analysed in R, version
156 3.6.3.

157

158 **Sensitivity analyses**

159 In order to consider recurrent sepsis events, we conducted a sensitivity analysis using
160 CPRD and HES where incident events were first sepsis records during each calendar year
161 during the study period. We also evaluated the effect of extending the time-window for
162 concurrent events from 30 to 90 days.

163

164

165 **Results**

166

167 **Incidence of sepsis in each data source**

168 There were 4,081,214 registered patients from 378 general practices, who were eligible for
169 linkage and contributed 36,209,676 patient-years (PY) of follow-up. There were 20,206
170 patients with a first episode of sepsis recorded in primary care EHRs, 20,278 in HES and
171 13,972 in ONS. The characteristics of patients with sepsis from each data source are
172 presented in Table 1. Each data source showed a slightly higher proportion of females than
173 males (CPRD: 51%; HES: 52%; ONS: 55%). The frequency of sepsis increased with age,
174 with a maximum over 75 years of age. The most deprived IMD quintile had the highest
175 proportion of sepsis cases for each data source for men and women (25% to 26%) and the
176 least deprived quintile consistently had the lowest (15% to 17%). The number of first sepsis
177 events increased over time for primary care and hospital records with more first episode
178 sepsis cases in the period 2012:2017 for both males and females. In CPRD, the period
179 2012:2017 accounted for 44% of sepsis episodes for men and 42% for women; in HES the
180 latest period accounted for 50% of cases for both genders. This increasing trend was not
181 apparent in mortality records. The period with the highest number of sepsis deaths in the
182 ONS file was 2007:2011.

183

184 Annual age- and gender-standardised incidence of sepsis is shown for each data source in
185 Figure 1. Primary care electronic records showed a steady increase in sepsis cases over the
186 study period, with an acceleration from 2012 to 2017. In the CPRD in 2002, the age-
187 standardised incidence was 0.35 (95% confidence interval 0.32 to 0.39) per 1,000 PYs in
188 males and 0.34 (0.30 to 0.37) per 1,000 PYs in females. By 2017, this increased to 1.15
189 (1.04 to 1.26) per 1,000 PYs among males and 1.10 per 1,000 PYs among females (95% CI
190 1.00 to 1.19). Consistent with primary care data, HES data showed a steep increase in
191 sepsis cases over the study period, particularly from 2012 to 2017. In HES records for 2002,
192 the age-standardised incidence was 0.29 (0.25 to 0.32) per 1,000 PYs among males and
193 0.25 (0.22 to 0.28) per 1,000 PYs among females. By 2017, this had increased to 2.52 (2.35
194 to 2.68) per 1,000 PYs among males and 2.10 (1.96 to 2.23) per 1,000 PYs among females.
195 In contrast to primary care and hospital records, there was no consistent trend in sepsis
196 recording as a cause of death during the study period. In 2002 sepsis mortality was 0.34
197 (0.31 to 0.38) per 1,000 PYs among males and 0.28 (0.26 to 0.31) per 1,000 PYs among
198 females. By 2017, this had risen to 0.42 (0.35 to 0.49) per 1,000 PYs among males and 0.30
199 (0.25 to 0.35) per 1,000 PYs among females although the highest peak in mortality rate was
200 in 2006 for females and 2007 for males. Supplementary Figure 1 shows equivalent age-
201 standardised rates when first sepsis records in each calendar year are included, rather than
202 only the first in the study period.

203

204

205 **Concurrent recording of sepsis events across data sources**

206

207 There were 4,117 (20%) first HES sepsis events and 2,438 (17%) mortality records that
208 were concurrent with incident primary care sepsis records, based on a 30-day time-window
209 (Figure 2 and Supplementary Table 1). Among the 13,972 deceased patients with sepsis
210 listed as any cause of death, 2,438 (17%) had an incident sepsis event recorded in primary
211 care EHRs and 3,397 (24%) had incident sepsis events recorded in HES in the same period.
212 Only 614 patients had index sepsis events recorded across all three data sources within 30

213 days of the date of event or date of death. We evaluated whether extending the time-window
214 influenced conclusions. Including both first and subsequent sepsis records and a 90-day
215 time window gave 4,770 (24%) HES records and 2,635 (19%) of mortality records
216 concurrent with an incident primary care sepsis record. (Supplementary Table 1 and
217 Supplementary Figures 3 and 4).

218 Logistic regression analysis of variables associated with concurrent recording in more than
219 one data source for patients with sepsis found that among patients with sepsis events
220 recorded in primary care, concurrent HES records of sepsis within 30 days before or after
221 first diagnosis were higher at younger or older ages and for patients with the most recent
222 period of diagnosis (Figure 3). For index sepsis events recorded in HES, concurrent
223 recording in primary care EHRs was not consistently associated with age, deprivation nor
224 period. Among patients with sepsis events recorded in ONS, there were lower odds of a
225 concurrent CPRD record of sepsis for the age range 35 to 54 years, being registered with a
226 practice in the least deprived quintiles and the most recent period of diagnosis (Figure 4).

227

228 **3.4 Incidence of sepsis from linked data sources**

229 Table 3 shows age-specific sepsis incidence rate per 1,000 patient-years from primary care
230 EHRs and for primary care EHRs combined with HES and ONS, stratified. With primary care
231 EHRs alone, sepsis incidence rates increased from 0.19 in males and 0.20 in females per
232 1,000 PY under five years of age, to 5.22 in males and 3.55 in females at 85 years and
233 older. Estimated incidence rates were substantially higher when either HES or mortality
234 records were included. When all three data sources were combined the incidence of sepsis
235 was 0.32 per 1,000 PY in males and 0.36 per 1,000 PY in females under five years
236 increasing to 10.09 per 1,000 PY in males and 7.22 per 1,000 PY in females at age 85 years
237 and older. Supplementary Figure 2 shows equivalent age-specific incidence when first
238 sepsis records in each calendar year, rather than the first in the study period, were included.

239

240 **Discussion**

241

242 **Summary of results**

243 This study evaluated sepsis recording in primary care electronic records, hospital episodes
244 and mortality records for a large population registered with general practices in England. We
245 found that, over a 16-year period, a similar number of incident sepsis events were recorded
246 in primary care and hospital records, However, a high proportion of these records were not
247 concurrent across data sources: the majority of sepsis events recorded in primary care were
248 not recorded in hospital episodes, and the majority of hospital episodes were not recorded in
249 primary care. This conclusion held even when events were evaluated over a longer time-
250 window or if recurrent as well as incident events were included. There were a smaller
251 number of records of sepsis from mortality registration but a majority of these were not
252 associated with concurrent primary care or hospital records for sepsis. Analyses of
253 associations between concurrent recording between sources and available patient
254 covariates were not consistent across linkages, suggesting that coding variations were
255 largely unexplained. Estimates for age-specific incidence rates may be up to twice as high if
256 linked data sources are employed.

257

258 **Interpretation**

259

260 Current estimates suggest that about 50-70% of all sepsis events are community-acquired
261 [24, 25]. Clinical guidelines recommend that patients with suspected sepsis should be
262 referred for management in hospital, which suggests that most patients seen in primary care
263 may be admitted to hospital [1]. However, some patients with sepsis may access hospital
264 services directly without first presenting in primary care. Investigations in hospital may
265 identify underlying causes for sepsis, which might be coded as the reason for admission.
266 Just under half of severe sepsis cases admitted to intensive care in England and Wales are

267 associated with a fatal outcome [26]. It may be expected that a sepsis diagnosis will be
268 communicated to the patient's general practice or, in the event of a fatal outcome, recorded
269 into mortality records. Our results indicate that the process of recording sepsis episodes
270 across different health information systems is highly inconsistent. Health professionals may
271 make varying use of the concept of sepsis in the clinical recording of patients' conditions.
272 'Sepsis' may sometimes form an element of the clinical narrative but, on other occasions, an
273 underlying cause may be given greater prominence, as in the COVID-19 pandemic. The
274 NHS in England produced a cross-sectoral plan to improve outcomes for patients with sepsis
275 [27]. This aimed to improve training between different healthcare professional groups and
276 reduce differences in coding practices between organisations. A National Early Warning
277 Score (NEWS) has been rolled out to improve early recognition of signs of sepsis but this
278 has been accompanied by concerns for false positive alerts and possible over-diagnosis of
279 sepsis [28].

280

281 **Limitations**

282 The present results derive from a large, representative population. However, a disadvantage
283 of using linked, rather than stand-alone CPRD records is that linkage eligibility restrictions
284 reduces the sample size and possibly representativeness. GPs must be eligible for linkage
285 which requires meeting standards of data completeness, thus biasing the practices within
286 the linked sample toward those that record disease events with greater accuracy. It is also
287 possible that there are some inaccuracies with linkage across sources. Discrepancies
288 between CPRD GOLD and the ONS death data have been highlighted, particularly in the
289 years prior to 2013 [29]. This study does not differentiate between community-acquired and
290 hospital-acquired sepsis. Further research is required to understand how sepsis incidence in
291 the combined sources can reveal more about the impact of antimicrobial stewardship
292 strategies.

293

294 **Findings in relation to other studies**

295 This study is broadly consistent with the growing body of literature that advocates the use of
296 linked data sources [13-16]. However, it also indicates that stand-alone CPRD data may
297 provide accurate estimates of changes in the burden of sepsis. Millet et al. (2016) found that
298 population-averaged community-acquired pneumonia incidence was 39% higher using the
299 linked data than stand-alone data [16]. It may be that increased awareness and
300 standardisation of detection and recording have prevented such discrepancies being
301 observed for sepsis.

302

303 **Conclusion**

304 Analysing linked data enhances the completeness of ascertainment of health events across
305 health service sectors and population health registries particularly for age-groups at highest
306 risk. Further standardisation of coding practices across linked sources in addition to more
307 timely and accurate recording of secondary care and mortality events in GP records would
308 help to improve comparability. Further research is required to investigate the reasons for any
309 divergent trends across the data sources and to differentiate trends in community-acquired
310 versus hospital-acquired sepsis.

311

312 **Declarations**

313

314 **Funding**

315 The study is funded by the National Institute for Health Research (NIHR) Health Services
316 and Delivery Programme (16/116/46). ERP and MG were supported by the NIHR Biomedical
317 Research Centre at Guy's and St Thomas' Hospitals. The views expressed are those of the
318 authors and not necessarily those of the NHS, the NIHR, or the Department of Health. The
319 funder of the study had no role in study design, data collection, data analysis, data
320 interpretation, or writing of the report. The authors had full access to all the data in the study
321 and both authors shared final responsibility for the decision to submit for publication.

322

323 **Conflict of Interest**

324 The authors have no conflicts of interest.

325

326 **Data sources**

327 The study is based in part on data from the Clinical Practice Research Datalink obtained
328 under license from the UK Medicines and Healthcare products Regulatory Agency. However,
329 the interpretation and conclusions contained in this report are those of the authors alone.

330

331 **Availability of data**

332 Requests for access to data from the study should be addressed to
333 martin.gulliford@kcl.ac.uk. All proposals requesting data access will need to specify planned
334 uses with approval of the study team and CPRD before data release.

335

336 **CRedit authorship contribution statement**

337

338 **Emma Rezel-Potts:** Data curation, Formal analysis, Methodology, Software, Writing –
339 Original Draft, Visualization. **Martin Gulliford:** Conceptualization, Validation, Writing –
340 Review & Editing, Supervision, Funding acquisition. **SafeAB Study Group:**
341 Conceptualization, Validation.

342 The SafeAB Study Group also includes Caroline Burgess, Vasa Curcin and James Shearer,
343 Judith Charlton, Joanne R. Winter, Xiaohui Sun, Catey Bunce, Robin Fox, Paul Little,
344 Alastair D Hay, Michael V. Moore, Mark Ashworth.

345

346

347

References

1. The National Institute of Health and Care Excellence. Sepsis: recognition, diagnosis and early management 2016. <https://www.nice.org.uk/guidance/ng51> (April 2020, date last accessed).
2. Rudd KE, Johnson SC, Agesa KM, et al. Global, regional, and national sepsis incidence and mortality, 1990-2017: analysis for the Global Burden of Disease Study. *The Lancet* 2020; 395: 200-11.
3. Funk DJ, Parrillo JE, Kumar A. Sepsis and Septic Shock: A History. *Crit Care Clin* 2009; 25: 83-101.
4. Martin GS, Mannino DM, Eaton S, Moss M. The Epidemiology of Sepsis in the United States from 1979 through 2000. *New England Journal of Medicine* 2003; 348: 1546-54.
5. Burki TK. Sharp rise in sepsis deaths in the UK. *Lancet Respir Med* 2018; 6: 826.
6. Gulliford MC, Sun X, Charlton J, et al. Serious bacterial infections and antibiotic prescribing in primary care: cohort study using electronic health records in the UK. *BMJ Open* 2020;**10**:e036975.
7. Rhee C, Murphy MV, Li L, Platt R, Klompas M. Comparison of trends in sepsis incidence and coding using administrative claims versus objective clinical data. *Clin Infect Dis* 2015; 60: 88-95.
8. Angus DC, Linde-Zwirble WT, Lidicker J, Clermont G, Carcillo J, Pinsky MR. Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care. *Crit Care Med* 2001; 29: 1303-10.
9. HM Government. Tackling antimicrobial resistance 2019-2024: The UK's five year national action plan 2019. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/784894/UK_AMR_5_year_national_action_plan.pdf (April 2020, date last accessed).

10. Fitzpatrick F, Tarrant C, Hamilton V, Kiernan FM, Jenkins D, Krockow EM. Sepsis and antimicrobial stewardship: two sides of the same coin. *BMJ QUAL SAF* 2019; 28: 758.
11. Gulliford MC, Moore MV, Little P, et al. Safety of reduced antibiotic prescribing for self-limiting respiratory tract infections in primary care: cohort study using electronic health records. *BMJ* 2016; 354: i3410.
12. Herrett E, Shah AD, Boggon R, et al. Completeness and diagnostic validity of recording acute myocardial infarction events in primary care, hospital care, disease registry, and national mortality records: cohort study. *BMJ* 2013; 346: f2350.
13. Crooks CJ, Card TR, West J. Defining upper gastrointestinal bleeding from linked primary and secondary care data and the effect on occurrence and 28 day mortality. *BMC Health Serv Res* 2012; 12: 392.
14. Vollmer WM, O'Connor EA, Heumann M, et al. Searching multiple clinical information systems for longer time periods found more prevalent cases of asthma. *J Clin Epidemiol* 2004; 57: 392-7.
15. Morgan CL, Currie CJ, Stott NCH, Smithers M, Butler CC, Peters JR. Estimating the prevalence of diagnosed diabetes in a health district of Wales: the importance of using primary and secondary care sources of ascertainment with adjustment for death and migration. *Diabetic Med* 2000; 17: 141-5.
16. Millett ERC, Quint JK, De Stavola BL, Smeeth L, Thomas SL. Improved incidence estimates from linked vs. stand-alone electronic health records. *J Clin Epidemiol* 2016; 75: 66-9.
17. Williams NP, Coombs NA, Johnson MJ, et al. Seasonality, risk factors and burden of community-acquired pneumonia in COPD patients: a population database study using linked health care records. *Int J Chron Obstruct Pulmon Dis* 2017; 12: 313-22.

18. Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. *Br J Clin Pharmacol* 2010; 69: 4-14.
19. Herrett E, Gallagher AM, Bhaskaran K, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol* 2015; 44: 827-36.
20. Medicines & Healthcare Products Regulatory Agency, National Institute for Health Research. CPRD linked data 2020. <https://cprd.com/linked-data> (April 2020, date last accessed).
21. Department for Communities and Local Government. The English Index of Multiple Deprivation (IMD) 2015 - Guidance. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/464430/English_Index_of_Multiple_Deprivation_2015_-_Guidance.pdf (April 2020, date last accessed).
22. Herbert A, Wijlaars L, Zylbersztejn A, Cromwell D, Hardelid P. Data Resource Profile: Hospital Episode Statistics Admitted Patient Care (HES APC). *Int J Epidemiol* 2017; 46: 1093-i.
23. McPherson D, Griffiths C, Williams M, Baker A, Klodawski E, Jacobson B et al. Sepsis-associated mortality in England: an analysis of multiple cause of death data from 2001 to 2010. *BMJ Open*. 2013;3(8):e002586.
24. Page DB, Donnelly JP, Wang HE. Community-, Healthcare-, and Hospital-Acquired Severe Sepsis Hospitalizations in the University Health System Consortium. *Crit Care Med* 2015; 43: 1945-51.
25. Adrie C, Alberti C, Chaix-Couturier C, et al. Epidemiology and economic evaluation of severe sepsis in France: age, severity, infection site, and place of acquisition (community, hospital, or intensive care unit) as determinants of workload and cost. *J Crit Care* 2005; 20: 46-58.

26. Padkin A, Goldfrad C, Brady AR, Young D, Black N, Rowan K. Epidemiology of severe sepsis occurring in the first 24 hrs in intensive care units in England, Wales, and Northern Ireland. *Crit Care Med* 2003; 31: 2332-8.
27. NHS England. Improving outcomes for patients with sepsis: a cross system action plan 2015. <https://www.england.nhs.uk/wp-content/uploads/2015/08/Sepsis-Action-Plan-23.12.15-v1.pdf> (April 2020, date last accessed).
28. The Lancet Respiratory Medicine. Crying wolf: the growing fatigue around sepsis alerts. *Lancet Respir Med* 2018; 6: 161.
29. Gallagher AM, Dedman D, Padmanabhan S, Leufkens HG, de Vries F. The accuracy of date of death recording in the Clinical Practice Research Datalink GOLD database in England compared with the Office for National Statistics death registrations. *Pharmacoepidemiol Drug Saf* 2019; 28: 563-9.

Table 1: Characteristics of patients with incident sepsis events in three data sources. Figures are frequencies (% of column total).

		Primary care (CPRD)		Hospital episodes (HES)		Mortality data (ONS)	
		Male	Female	Male	Female	Male	Female
Total		9 893 (49)	10 313 (51)	9 796 (48)	10 482 (52)	6 245 (45)	7 727 (55)
Age-group (years)	0-4	137 (1)	138 (1)	122 (1)	142 (1)	5 (0)	5 (0)
	5-14	194 (2)	181 (2)	131 (1)	82 (1)	13 (0)	6 (0)
	15-24	230 (2)	293 (3)	114 (1)	121 (2)	19 (0)	17 (0)
	25-34	257 (3)	464 (4)	191 (2)	238 (2)	33 (0)	19 (0)
	35-44	461 (5)	667 (6)	353 (4)	430 (4)	91 (1)	83 (1)
	45-54	786 (8)	911 (9)	651 (7)	840 (8)	219 (4)	219 (3)
	55-64	1 410 (14)	1 351 (13)	1 298 (13)	1 180 (11)	520 (8)	468 (6)
	65-74	2 186 (22)	1 756 (17)	2 178 (22)	1 809 (17)	1 172 (19)	962 (12)
	75-84	2 639 (27)	2 281 (22)	2 855 (29)	2 661 (25)	2 169 (35)	2 423 (31)
	85+	1 593 (16)	2 271 (22)	1 903 (19)	2 979 (28)	2 004 (32)	3 525 (46)
Deprivation fifth	Least deprived	1 570 (16)	1 502 (15)	1 630 (17)	1 662 (16)	923 (15)	1 123 (15)
	2 nd	1 570 (16)	1 502 (15)	1 630 (17)	1 662 (16)	923 (15)	1 123 (15)
	3 rd	2 126 (21)	2 403 (23)	1 892 (19)	2 004 (19)	1 258 (20)	1 525 (20)
	4 th	1 852 (19)	1 817 (18)	1 892 (19)	1 960 (19)	1 150 (18)	1 408 (18)
	Most deprived	1 889 (19)	1 914 (19)	1 944 (20)	2 178 (21)	1 308 (21)	1 657 (21)
	Data missing	1 (0)	2 (0)	1 (0)	4 (0)	6 (0)	4 (0)
Period	2002:2006	2 271 (23)	2 604 (25)	1 911 (19)	2 133 (20)	2 077 (33)	2 550 (33)
	2007:2011	3 234 (33)	3 422 (33)	2 964 (30)	3 155 (30)	2 398 (38)	3 086 (40)
	2012:2017	4 388 (44)	4 287 (42)	4 921 (50)	5 194 (50)	1 770 (28)	2 091 (27)

CPRD, Clinical Practice Research Datalink; HES, Hospital Episode Statistics; ONS, Office for National Statistics

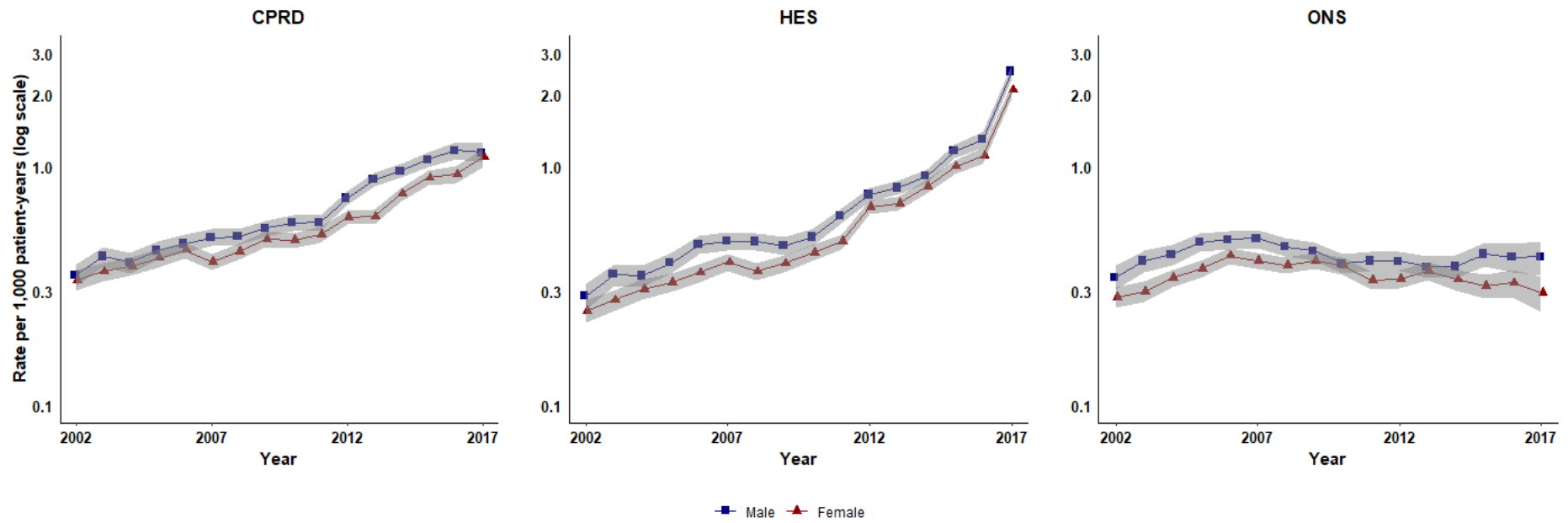


Figure 1 Age-standardised incidence of first cases of sepsis in CPRD and HES and age-standardised sepsis mortality in ONS reported from 2002-2017, stratified by gender.

CPRD, Clinical Practice Research Datalink; HES, Hospital Episode Statistics; ONS, Office for National Statistics

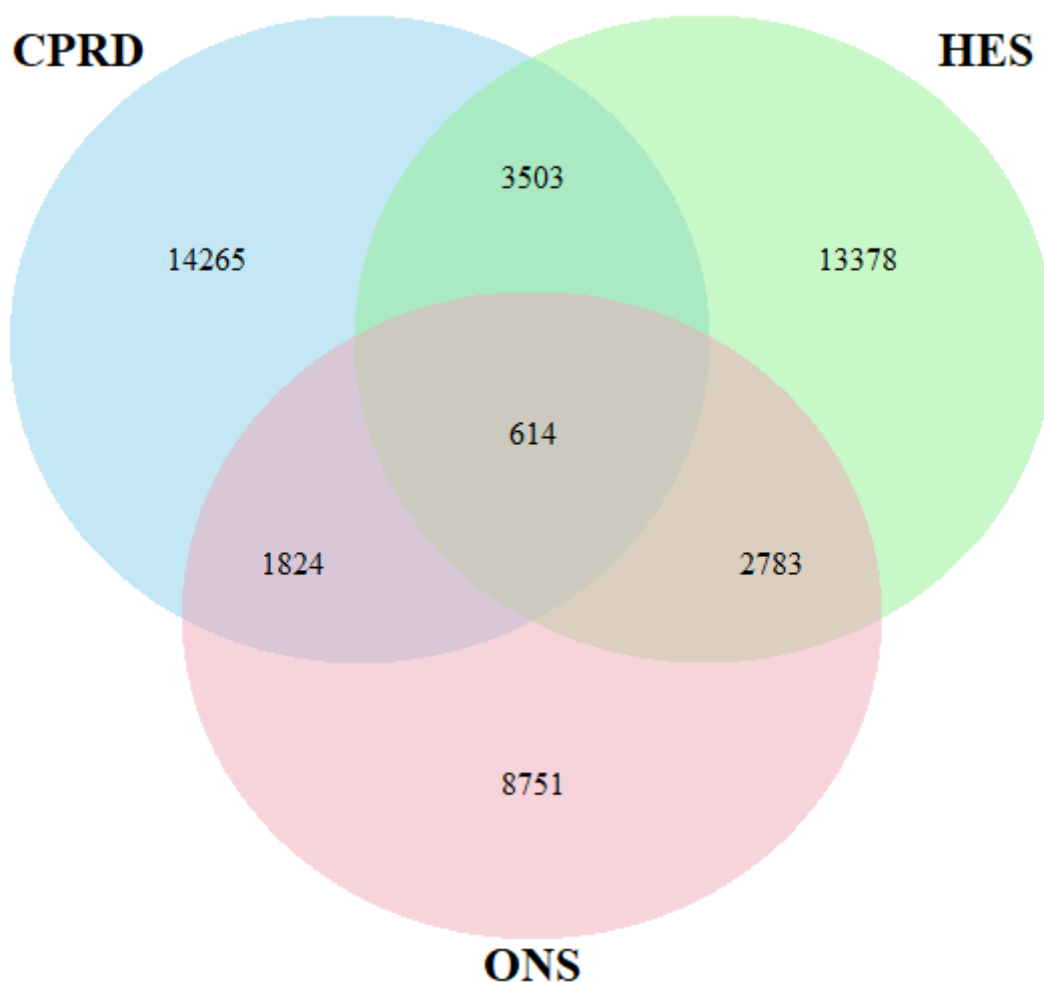


Figure 2 Incident first sepsis events in CPRD, HES and ONS using a 30-day time-window to evaluate concurrence.

CPRD, Clinical Practice Research Datalink; HES, Hospital Episode Statistics; ONS, Office for National Statistics

Table 2: Age-specific sepsis incidence rate (exact Poisson 95% confidence intervals) per 1,000 patient-years in each CPRD and CPRD combined with HES and ONS, stratified by gender.

Age group (years)	CPRD		CPRD + HES		CPRD + ONS		CPRD + ONS + HES	
	Male	Female	Male	Female	Male	Female	Male	Female
0 to 4	0.19 (0.00 to 4.08)	0.20 (0.00 to 4.10)	0.32 (0.00 to 4.33)	0.36 (0.00 to 4.40)	0.20 (0.00 to 4.09)	0.21 (0.00 to 4.11)	0.32 (0.00 to 4.33)	0.36 (0.00 to 4.40)
5 to 14	0.09 (0.00 to 3.87)	0.09 (0.00 to 3.87)	0.13 (0.00 to 3.96)	0.12 (0.00 to 3.93)	0.09 (0.00 to 3.88)	0.09 (0.00 to 3.88)	0.13 (0.00 to 3.96)	0.12 (0.00 to 3.93)
15 to 24	0.11 (0.00 to 3.92)	0.16 (0.00 to 4.02)	0.14 (0.00 to 3.99)	0.21 (0.00 to 4.11)	0.12 (0.00 to 3.94)	0.17 (0.00 to 4.03)	0.14 (0.00 to 3.99)	0.21 (0.00 to 4.11)
25 to 34	0.12 (0.00 to 3.93)	0.21 (0.00 to 4.12)	0.18 (0.00 to 4.05)	0.29 (0.00 to 4.28)	0.13 (0.00 to 3.95)	0.22 (0.00 to 4.13)	0.18 (0.00 to 4.05)	0.29 (0.00 to 4.28)
35 to 44	0.17 (0.00 to 4.03)	0.25 (0.00 to 4.19)	0.26 (0.00 to 4.21)	0.36 (0.00 to 4.41)	0.19 (0.00 to 4.08)	0.27 (0.00 to 4.24)	0.26 (0.00 to 4.21)	0.36 (0.00 to 4.41)
45 to 54	0.29 (0.00 to 4.27)	0.35 (0.00 to 4.38)	0.46 (0.00 to 4.60)	0.57 (0.00 to 4.81)	0.35 (0.00 to 4.39)	0.42 (0.00 to 4.52)	0.46 (0.00 to 4.60)	0.57 (0.00 to 4.81)
55 to 64	0.61 (0.00 to 4.87)	0.59 (0.00 to 4.84)	0.99 (0.02 to 5.55)	0.93 (0.02 to 5.45)	0.78 (0.01 to 5.19)	0.74 (0.01 to 5.11)	0.99 (0.02 to 5.55)	0.93 (0.02 to 5.45)
65 to 74	1.30 (0.07 to 6.09)	0.98 (0.02 to 5.53)	2.19 (0.31 to 7.53)	1.68 (0.15 to 6.72)	1.83 (0.19 to 6.95)	1.39 (0.08 to 6.24)	2.19 (0.31 to 7.53)	1.68 (0.15 to 6.72)
75 to 84	2.65 (0.47 to 8.23)	1.73 (0.16 to 6.79)	4.74 (1.86 to 11.29)	3.23 (0.72 to 9.12)	4.35 (1.27 to 10.75)	3.17 (0.69 to 9.02)	4.74 (1.48 to 11.29)	3.23 (0.72 to 9.12)
85+	5.22 (1.75 to 11.97)	3.55 (0.87 to 9.58)	10.09 (4.86 to 18.51)	7.22 (2.96 to 14.72)	10.42 (5.09 to 18.93)	7.93 (3.41 to 15.67)	10.09 (4.86 to 18.51)	7.22 (2.96 to 14.72)

CPRD, Clinical Practice Research Datalink; HES, Hospital Episode Statistics; ONS, Office for National Statistics

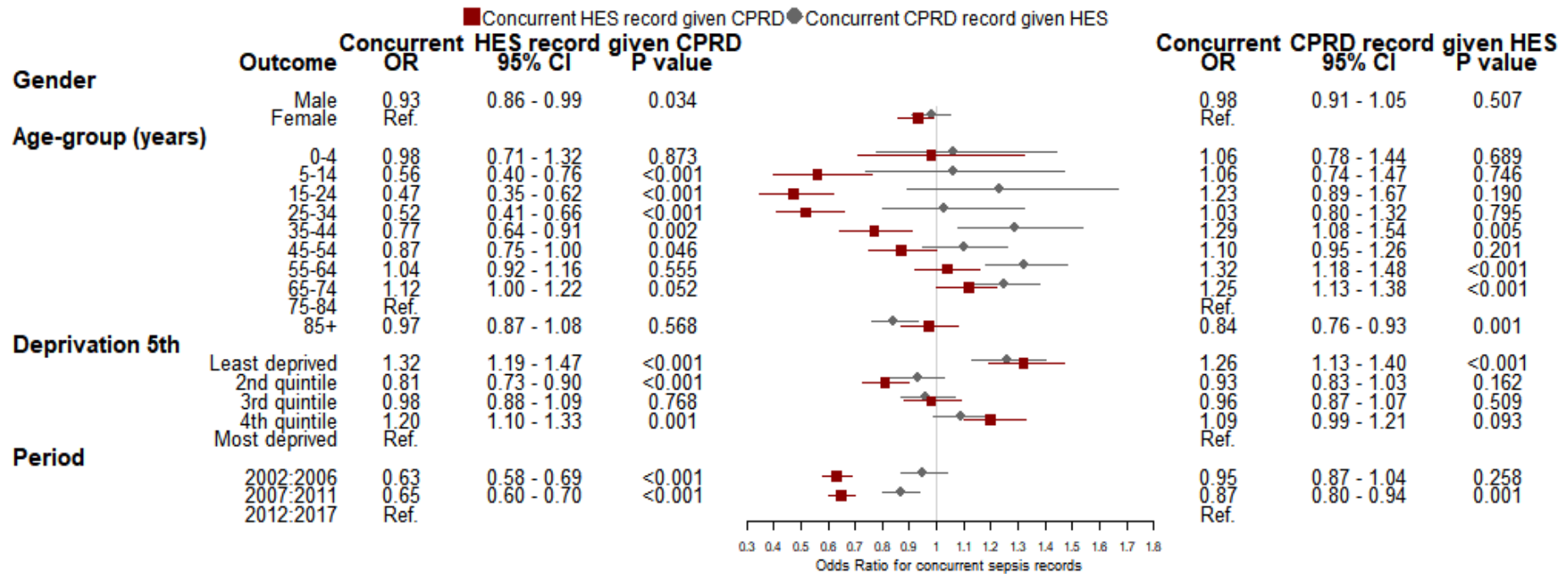


Figure 3 Logistic regression model of variables associated with the outcome of a) concurrent index sepsis recording HES given sepsis event in CPRD and b) concurrent index sepsis recording in CPRD given index sepsis event in HES.

CPRD, Clinical Practice Research Datalink; HES, Hospital Episode Statistics; ONS, Office for National Statistics

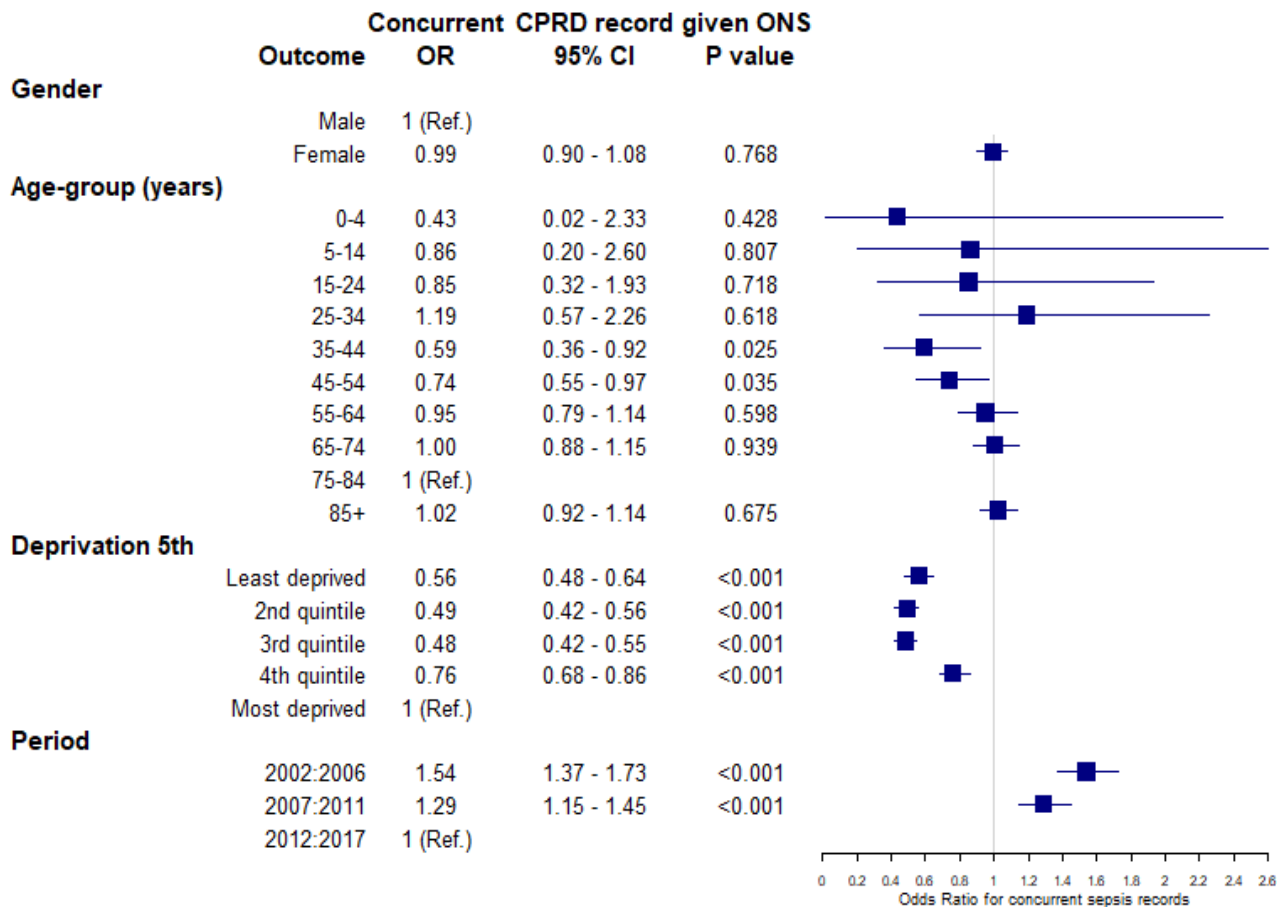


Figure 4 Logistic regression model of variables associated with the outcome of concurrent index sepsis recording in CPRD given recording of sepsis as any cause of death in ONS.

CPRD, Clinical Practice Research Datalink; ONS, Office for National Statistics