

# International Journal of Population Data Science

Journal Website: [www.ijpds.org](http://www.ijpds.org)



Swansea University  
Prifysgol Abertawe

## Challenges in linking administrative data for monitoring bloodstream infection in neonatal units in England and Wales

Fraser, C<sup>1</sup>, Gilbert, R<sup>1</sup>, Blackburn, R<sup>2</sup>, Muller-Pebody, B<sup>3</sup>, and Harron, K<sup>2</sup>

<sup>1</sup>UCL Institute of Child Health

<sup>2</sup>University College London, UK

<sup>3</sup>Public Health England

### Introduction

Monitoring risk-adjusted trends of neonatal bloodstream infection (BSI) is vital and linkage of neonatal electronic health records to national infection surveillance enables this. We demonstrate why changes in data quality and collection methods over time must be accounted for to minimise spurious findings.

### Objectives and Approach

First, we determined the effect of a system change in 2014 (changed from only clinically relevant BSI to automated reporting of all BSI), by investigating changes in number of all BSI and BSI excluding the contaminants coagulase-negative staphylococci for infants aged <1 year reported to infection surveillance, using interrupted-time-series Poisson regression. Second, we evaluated the impact of changes in identifier completeness over time in each database, and determined variation in infection rates according to linkage method (deterministic linkage on NHS number or probabilistic linkage). Third, we will use multiple imputation when link status cannot be determined due to missing identifiers.

### Results

The number of BSI reported to infection surveillance system following the change in data collection increased by 34% (incidence rate ratio (IRR) of 1.34, 95% confidence interval 1.28-1.40) for all BSI compared to 19% (IRR 1.19, 1.12-1.27) excluding coagulase-negative staphylococci. Completeness of NHS number in infection surveillance increased from 69% (3,296/4,792) in 2010 to 92% (3,037/3,307) in 2017. We linked 12,003 neonatal admissions to 15,571 BSI episodes (2% of 497,936 admissions and 41% of 37,660 BSI). The proportion of links that were deterministic changed from 83% (1,089/1,307) in 2010 to 96% (968/1,008) in 2017. There were 12,094 BSI for which the link status could not be determined due to missing identifiers; multiple imputation will be used to determine if any are links.

### Conclusion/Implications

Spurious trends in infection incidence can arise from changes in data collection and quality, impacting the quality of linkage to clinical data. Data quality and system changes must be explored in each source dataset before analysis. Probabilistic linkage and imputation of missing data minimises spurious findings due to data quality.

