

## 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>1</sup>

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

<sup>2</sup>University College London

<sup>3</sup>Public Health England

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

First, 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). Second, we will use multiple imputation when link status cannot be determined due to missing identifiers.

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 infection episodes (2% of 497,936 admissions and 41% of 37,660 infections). The proportion of links that were deterministic changed from 83% (1,089/1,307) in 2010 to 96% (968/1,008) in 2017. Link status could not be determined for 12,094 infections due to missing identifiers; multiple imputation will be used to determine if any are links.

Spurious infection incidence rates can arise from changes in data quality, impacting the quality of linkage to clinical data. Linkage and imputation of missing data minimises spurious findings due to data quality.

\*Corresponding Author:

Email Address: [caroline.fraser@ucl.ac.uk](mailto:caroline.fraser@ucl.ac.uk) (C Fraser)

