

## Evaluating record linkage of birth registration and notification records to Hospital Episode Statistics: Singleton births in 2005 and 2006 across England

Coathup, V<sup>1\*</sup>, Macfarlane, A<sup>2</sup>, and Quigley, M<sup>1</sup>

<sup>1</sup>University of Oxford

<sup>2</sup>City University

### Background with rationale

Linked administrative datasets are particularly useful within the field of perinatal epidemiology. By linking multiple datasets, researchers can create longitudinal datasets, which allow them to explore research questions relating to early exposures and outcomes later in life.

### Main aim

The aims of this study were to describe the methods used to deal with duplicate hospital admission records, assess the quality of linkage between babies birth registration records and subsequent hospital admissions, and to evaluate the potential bias that may be introduced as a result of these methods.

### Methods

Three routinely collected datasets were linked for use within this study and included data from birth registration, NHS Numbers for Babies (NN4B) and Hospital Episode Statistics (HES) for babies born in England between 1<sup>st</sup> January 2005 and 31<sup>st</sup> December 2006. A number of stages to cleaning were undertaken, including dealing with duplicate HES records and assessing the quality of the linkage using a deterministic algorithm. Internal and external validity was also assessed.

### Results

There were a total of 1,170,970 live, singleton births, occurring in NHS hospitals, to mothers who normally reside in England in 2005 and 2006 combined. Of these, approximately 92% were successfully linked with a HES birth record. Data quality was somewhat poorer in HES birth records compared to birth registration and NN4B. The quality assurance algorithms identified 1,456 incorrect linkages (<1%) and examination of external validity identified children that were not linked were slightly more likely to be born to mothers who were older and of higher socio-economic status.

\*Corresponding Author:

Email Address: [victoria.coathup@npeu.ox.ac.uk](mailto:victoria.coathup@npeu.ox.ac.uk) (V Coathup)

### Conclusion

It is possible to create valuable longitudinal datasets allowing researchers to explore important questions about exposures and childhood outcomes using administrative datasets, however, missing data and coding errors and inconsistencies mean it is important that the quality of linkage is assessed prior to analysis.

