

Meeting the challenge of data linkage for special populations

Hills, B¹

¹Population Data BC, University of British Columbia

Introduction

A file's overall linkage rate may hide significant bias in unlinked records. Newborns, for example, may represent a small percentage of a file but a large proportion of unlinked records. This creates challenges for researchers focused on the affected populations. Decreasing bias is a clear goal for data linkage science.

different data sets, but the linkage of babies with adverse outcomes required considerably more change, including moving beyond our usual spine linkage approach.

Objectives and Approach

A standard spine-based linkage process with a typical set of identifiers was used to link perinatal records in British Columbia. This included multiple iterations and some manual linkage, stopping when there were only minor improvements. Despite high overall linkage rates, poor rates existed for babies who were stillborn or died near birth. Given research interests in stillbirth, there was a desire to use a creative approach to improve linkage for that sub-population. The solution was to expand the scope of the spine linkage to include birth and stillbirth records from other files, e.g. hospital discharge and Vital Statistics Registrations.

Results

Original linkage methods produce overall baby linkage rates of 98.1% but only 3.9% of records discharged to death/stillbirth were linked. Revised methods provide a small change in the overall rate to 99.4% but increase the discharge to death/stillbirth linkage rate to 90.3%. All numbers exclude terminations. Additions to our base representation of a linked entity enabled the association of records for babies either stillborn or never registered for health insurance coverage with the Ministry of Health. Additional variables addressed high rates of missing identifiers, while new linkage processing to Vital Statistics data addressed gaps within multiple newborn related datasets.

Conclusion/Implications

The complexity of adding BC Perinatal Data as a linked dataset to Population Data BC holdings was underappreciated. There are always adjustments in linkage approach across

