From data linkage to policy and back again: A preliminary report on the linkage of neonatal data in Victoria

CA Stone, CA Ramsden, JA Howard, M Roberts and JH Halliday

The Data and Information Sub-committee, Neonatal Services Advisory Committee, Department of Human Services Victoria

Little Gidding

What we call the beginning is often the end.
And to make an end is to make a beginning.
The end is where we start from …

… We shall not cease from exploration
And the end of all our exploring
Will be to arrive where we started
And know the place for the first time.

TS Eliot

Abstract

The Neonatal Services Advisory Committee (NSAC) was established in 2000 to 'advise and assist the Department in the provision of the best possible standard of neonatal care to Victorian babies'. The NSAC established a Data and Information Sub-Committee (DISC) to ensure that data collection and analysis contribute to the effective monitoring and planning of neonatal services. In order to support DISC we set up the Neonatal Data Record Linkage Project which was to explore the feasibility and usefulness of linking maternal and neonatal data from routinely collected datasets, the Perinatal Data Collection Unit (PDCU), the Victorian Admitted Episode Dataset (VAED) and the Australian and New Zealand Neonatal Network (ANZNN). The aim of this paper is to describe the work, and the technological and political issues that arose during the process of this project.

We linked the records of Victorian-born neonates, their mothers and their subsequent transfers using probabilistic record linkage software, AutoMatch and a complex multistage approach. The linkage process was highly successful in that a high percentage of those that could be matched were. This was in spite of the complexity of the process and the political process of working across organisations with different datasets requirements.

In summary, record linkage of routinely collected data provides potential to inform policy however there is still work to be done to improve policy around record linkage. In a field that is expanding so rapidly, we believe that this is an iterative process.

Introduction

Comprehensive and valid neonatal data are required for effective monitoring and planning of neonatal services in Victoria. The most recent Department of Human Services (DHS) review of neonatal services in Victoria, the DHS Wilkinson Review(1), was undertaken in 1998, to 'assess the ability of existing structures and resources to meet current and future demand for neonatal care of an appropriate standard and make recommendations for systems improvement'.

As a result of the recommendations of the Wilkinson Review, the Minister for Health launched the Neonatal Services Advisory Committee (NSAC) in April 2000, which was to 'advise and assist the Department in the provision of the best possible standard of neonatal care to Victorian babies'. The NSAC established a data and information sub-committee (DISC) to ensure that data collection and analysis contribute to the effective monitoring and planning of neonatal services.

Currently, a large amount of data on neonates and their care is available through the Victorian Admitted Episodes Dataset (VAED), the Perinatal Data Collection Unit (PDCU) and the Australian and New Zealand Neonatal Network (ANZNN) datasets. Combining the variables from the three sets will increase the 'power' of the datasets for the purpose of assisting the DISC in its service planning and monitoring roles. However, in the absence of a unique identifier, linking a neonate's record across the three datasets is a complex process.

The Neonatal Data Record Linkage Project was established to explore the feasibility of linking maternal and neonatal data from the datasets. The project was approved by the DHS Ethics Committee. The resulting composite dataset would provide the most comprehensive data on Victorian babies for a) developing performance indicators for service planning and monitoring, and b) undertaking neonatal research. The aim of this paper is to describe the work undertaken, and technological and political issues that arose during this data linkage project.

Methods

The data sources were the Perinatal Data Collection Unit (PDCU), which is the most complete collection with respect to numbers of women giving birth and their babies in Victoria(2). The PDCU has information on almost all births in both public and private hospitals and also contains gestational age and weight, but has little detail about neonatal morbidity. The Victorian Admitted Episodes Dataset (VAED), which has almost all mothers and all babies born in public hospitals but only those babies admitted to special care nurseries (SCNs) in private hospitals. The Australia and New Zealand Neonatal Network (ANZNN), which has high-risk newborns (gestation <32W or weight <1500 gm or those ventilated or undergoing major surgery).

The records were linked using probabilistic record linkage software, Automatch(3), in a sequential multistage approach. The first stage was the linking of the baby-mother record in the PDCU (index cases) to the mothers records in the VAED. In the second stage: the linked dataset (PDCU-mVAED) was then linked to the neonates' record in the VAED. The third stage consisted of linking the file (PDCU-mVAED-nVAED) with the neonates' transfer records in the VAED. The fourth stage will be linking the file (PDCU-mVAED-nVAED-tVAED) to the records of the Victorian neonates in the ANZNN. The final product will be a file that contains a composite record for each baby born in Victoria including its first transfer after birth.

The programming for each of the linkage stages was piloted with small datasets to ensure the linkage program provided the optimum number of correct linkages, and the minimum number of incorrect linkages and clerical reviews. Some mothers had two birth events in one year. When linking mothers it was important to ensure that the same event was linked rather than the other event. Similarly, when linking babies from a multiple birth programming had to ensure there was linkage of the same baby and reduce the risk of linking to its very similar sibling. This part of the process involved a large amount of quality control and provided a checking process for the linkage overall.

Results

Study population

The study population consisted of the 62,091 births that reportedly occurred in Victoria in 1998, according to the PDCU. The maximum numbers of babies with the same birthday (one of the linking variables) was 229. There were 37 sets of triplets and 983 sets of twins. In addition to the risk of linking babies with similar birth data, some mothers have two births in the same year (approximately 300). There were 99 homebirths and 406 stillbirths in the PDCU but these are not recorded in VAED although we have included them in our denominator.

Success of linkages

In stage 1, 61,207 maternal records in the VAED were linked to the mother-baby records (the index cases) from the PDCU. The linkage success rate equals 98.6%. In stage 2, 47,801 of the neonates records in the VAED were linked to the 62,091 index cases (now a linked file called PDCU-mVAED). This equates to a linkage success rate of 77.0%. If we look at public and private hospitals separately, 41,475 out of 41,702 (99.4%) of public hospital babies could be linked, whereas only 5,860 of 18,919 (31.0%) of private hospital babies were.

Stage 3. Out of 62,091 index cases (PDCU-mVAED-nVAED) we were able to link 1,100 or 1.8% to a direct transfer in the VAED. The true denominator was difficult to determine for a number of reasons: it was not known how many babies were actually transferred after a birth admission; not all private hospital babies were admitted; and not all babies were linked. Assuming that the VAED variable septype='T' is correctly coded, not underreported and all private hospital babies that were sick enough to be transferred were officially admitted, there were at least 2,157 babies transferred after their birth admission. The success rate was therefore 1,100/2,157 or 51%. In stage 4, the final linkage will be to the babies' records in the ANZNN.

Discussion

The major issue in relation to record linkage is the lack of a unique identifier for health data in Victoria. This means that for population level information, information collected in routinely collected datasets needs to be linked using automated methods rather than a simple merge or manual method. While record linkage is a valuable tool to combine and validate the information, it can be very resource intensive.

A major part of the work is the exploration of the dataset, data cleaning, data recoding and formatting, construction of new variables and programming and piloting of the linkage strategy.

Record linkage of birth data at a population level is complicated because there are many similar records, which may result in incorrect linkages. Our previous linkage work has involved linking mainly maternal data within a single dataset such as constructing reproductive histories (4), or across datasets, one of which has relatively small numbers, such as in the Koori Births project. (5-7) Linking the neonates is another level of difficulty. Date of birth is often used as one of the common variables for linkage however because this is the full birth cohort, it has the maximum number of individuals with the same birth date. In Victoria in 1998 the maximum number of babies born on the same day was 229. There were 37 sets of triplets and 983 sets of twins. In addition to linking babies with similar birth data, some mothers have two births in the same year (approximately 300).

The differences in the datasets, the quality control and the incentives provided for complete collection means that a large amount of resources are committed to data cleaning and data recoding. Record linkage of birth data at a population level is complicated because variables may have different data definitions. In this case, episodes do not correspond between datasets. A single birth episode in the PDCU record may equate to multiple admissions for the neonate in the VAED (eg Women's Hospital, Children's Hospital and back to the Women's Hospital) and a single admission under first hospital in the ANZNN.

A measure of the quality control on the linkage process is the completeness of the linkage. When the recorded admissions are incomplete, as is the case with babies born in private hospitals, it is hard to determine whether those that were not linked were missed or were actually not recorded. Incomplete data also have an influence on the generalisability of the results. Of the 42,032 public hospital babies most should be in VAED whereas not all of the 20,059 private hospital babies are officially admitted so many are not in VAED, so in any analysis factors associated with private births may be underrepresented. The 99 homebirths are not in VAED so have a higher chance of not being in the linked dataset for analysis. The 406 stillbirths are not in VAED but they are not part of this study. Babies are often given mothers medicare number but up to 12.5% are missing medicare numbers in VAED. We recommend good quality control on data sets, consistent data dictionary, allowance for multiple admissions, admit and record all neonates and consistently identify babies in multiple births

Privacy/confidentiality issues are other major ethical issues, because the process by definition is dependent on being able to link records using identifiers within a dataset, although all results will be reported as aggregated data.. Many departmental datasets have sensitive data and the staff are well skilled in the issues of privacy and confidentiality. However this linkage is with an external dataset that has different protocols and there is substantial difficulty is devising a protocol that fulfils the varying requirements of each dataset. For example, the PDCU has an internal policy not to release data identifying individual hospitals. Many record linkers face similar problems and development of a standard protocol, addressing the ethical issues in particular, would benefit this type of epidemiological research.

The protocols for access to and management of databases often differ and at times may be in conflict. In this project one dataset has a restriction of only being on stand alone computers and is not able to be networked, whereas the other must be on a networked computer and not a stand-alone computer. These restrictions on who can do the work, what can be done, where and when it can be done, can make linkage rather difficult. This requires careful negotiation. On the other hand careful consideration of the data access issues is important to maintain the privacy and confidentiality of such sensitive information.

There have already been a number of beneficial outcomes of this project even before the linkage has been completed. We have been able to provide advice on changes to data items within the datasets such as the need to clearly identify the babies in multiple births. The DHS has been aware that not all babies in hospital are recorded in the VAED. These babies are referred to as ‘boarders’ and include babies that are transferred with their mothers, or their siblings of a multiple birth, or are transferred from a special nursery to a smaller hospital. We know from other sources how many occur as the birth admission but it is unknown how many are transferred afterwards. The data from the linkage project has been used to provide information on the number of babies who are admitted as ‘boarders’ after their birth admission. The issues about access and provision of composite datasets are being reviewed at the moment.

Conclusion

Record linkage is a valuable tool to combine, augment and validate the information in routinely collected datasets. Our composite dataset will provide a powerful tool to investigate the current status of neonatal care across Victoria and as such it can assist in the development of informed policy and planning. While birth data have been linked many times previously, this data linkage was more complex and resource intensive than the usual because of the complexity of birth data and differentiating multiple births, the multistage process that was necessary to gradually build up the linkage variables and the political process of working across organisations with different datasets requirements. It is important to have clear policies and protocols for all phases of the process from data collection through analysis, storage of composite dataset and final report.

With apologies to TS Eliot

What we call the beginning is often the end.
And to make our policy we do our linking.
To do our linking we make our policy.

… We shall not cease from policy making
and at the end of all our linkages
will be to arrive where we started
with the need to make policy again.

References

1. DHS V. Report on Victorian Neonatal Care Services. 1998.

2. Riley M, Griffin O. Validating a statewide data collection: differences in information technology resources between hospitals. Health Information Management 1997;27(2):67–8.

3. Jaro M. AutoMatch, Generalised record linakge system. In. Silver Spring, Maryland: Matchware Technologies; 1994.

4. Stone C, Halliday J, Lumley J, Brennecke S. Vaginal births after Caesarean (VBAC): a population study. Paediatr Perinat Epidemiol 2000;14(4):340–8.

5. Stone C, Owen T, Ross I, Halliday J. Koori Births Project: Linking mothers, linking organizations. In: Sansoni J, Tilley L, editors. Health Outcomes 2001; 2001; Canberra: The Australian Health Outcomes Collaboration, Centre for Health Service Development, University of Woolongong, NSW, Australia; 2001.

6. Koori Health Unit. Koori Health Counts: How many koori babies were born in Victoria in 1996. Melbourne: Koori Health Unit, Department of Human Services, Victoria, Australia; 1999.

7. Koori Health Unit. Koori Health Counts: How many koori babies were born in Victoria in 1998. Melbourne: Koori Health Unit, Department of Human Services, Victoria, Australia; 2000.