The real divide: the use of algorithm-derived Indigenous status to measure disparities in sudden unexpected deaths in infancy in Queensland

Rebecca A. Shipstone,[1] Jeanine Young,[1] John M.D. Thompson[1,2]

Abstract

Objective: To investigate the under-identification of Indigenous[*] infants in death records and examine the impact of a multi-stage algorithm on disparities in sudden unexpected deaths in infancy (SUDI).

Methods: Data on SUDI in Queensland between 2010 and 2014 were linked to birth and death registrations, health data, and child protection and coronial records. An algorithm was applied to cases of SUDI and population data to derive Indigenous status. Numbers, proportions and rates of SUDI were compared.

Results: Using multiple sources of Indigenous status resulted in a 64.9% increase in the number of infants identified as Indigenous. The Indigenous SUDI rate increased by 54.3%, from 1.38 to 2.13 per 1,000 live births after applying the algorithm to SUDI and live births data.

Conclusions: Applying an algorithm to both numerator and denominator data reduced numerator-denominator incompatibility, to more accurately report rates of Indigenous SUDI and measure the gap in Indigenous infant mortality.

Implications for public health: Estimation of the true magnitude of the disparity is restricted by under-identification of Indigenous status in death records. Data linkage improved the reporting of Indigenous infant mortality. Accuracy in reporting of measures is integral to determining genuine progress towards Closing the Gap.

Key words: Aboriginal, data linkage, Indigenous, mortality, sudden infant death syndrome (SIDS)

[*] Due to repeated use throughout this paper the term Indigenous people will be respectfully used when referring to individuals who identify as Aboriginal and/or Torres Strait Islander.

Aboriginal and Torres Strait Islander people experience considerably poorer health than non-Indigenous Australians, with disparities persisting across all health outcomes routinely reported by Australian governments.[1-3] This has been recognised by the Council of Australian Governments' (COAG) commitment to halving the gap in infant and young child mortality as one of seven national targets to achieve health equity.[4,5] Infant mortality (deaths of infants under one year of age) contributes more than 80% of mortality for children aged less than five years of age and is a well-established measure of child and population health and social prosperity.[6] Infant mortality rates reflect infant deaths from all causes, both preventable and unpreventable. While deaths in the neonatal period are predominantly due to less easily avoidable causes such as extreme prematurity and congenital anomalies, deaths in the postneonatal period are, in the main, considered preventable and largely associated with social determinants of health.[7] Sudden unexpected deaths in infancy (SUDI) is a broad category of infant death that includes sudden infant death syndrome (SIDS), and is defined as the sudden, unexpected death of an infant in which the cause of death is not immediately obvious.[8,9] SUDI contributes notably to infant mortality. It is in the top three categories of death overall and is the leading category of death in the postneonatal period, accounting for more than 90% of postneonatal deaths. SUDI has also repeatedly been shown to be higher among Indigenous infants.[9]

SUDI are particularly sensitive to public health initiatives, with previous rate reductions of up to 86% in Australia and internationally attributed to successful safe sleeping campaigns.[10,11] More recently, the introduction of culturally valued safe sleep interventions has resulted in a marked reduction in Māori SUDI in New Zealand, after a decade of little change.[7] In this context, SUDI mortality serves as a litmus test for the effectiveness of initiatives to reduce disparities in Indigenous infant mortality. Measuring and quantifying the gap in SUDI mortality is reliant on administrative data. However, the recording of Indigenous status is known to be poor in almost all datasets and jurisdictions.[3,12-14] This is due to Indigenous people's degree of willingness to self-identify, including context dependent changes in an individual's Indigenous status, and the

1. School of Nursing, Midwifery, and Paramedicine, University of the Sunshine Coast, Queensland
2. Departments of Paediatrics, Child and Youth Health, and Obstetrics and Gynaecology, Faculty of Medical and Health Science, University of Auckland, New Zealand
Correspondence to: Ms Rebecca Shipstone, School of Nursing, Midwifery, and Paramedicine, University of the Sunshine Coast, ML 40 USC Locked Bag 4 Maroochydore DC, Queensland 4558; e-mail: rebecca.shipstone@research.usc.edu.au

Submitted: April 2019; Revision requested: July 2019; Accepted: September 2019
The authors have stated they have no conflict of interest.

Aust NZ J Public Health. 2019; 43:570-6; doi: 10.1111/1753-6405.12951

opportunities afforded for them to do so.[3] Indigenous status may not be asked of individuals who do not appear Indigenous, particularly in crisis situations such as the completion of a death certificate following an unexpected infant death. Studies in Western Australia and New South Wales suggest that Indigenous people are under-enumerated in official death records by 23% and 27%, respectively.[15,16] Misclassification of Indigenous status affects estimates of Indigenous infant mortality and other health outcomes,[12-15] and undermines the credibility of official measures of social progress.[17]

The problem of under-recording of Indigenous status can be substantially diminished through data linkage, which has been used to improve reporting of mortality data in Australia,[14-16,18] New Zealand[19] and Canada.[20] A number of estimation algorithms for determining Indigenous status from linked data sources have been previously reported.[1,12-15,21-25] The aim of this study was to apply the algorithm proposed by the 'Getting our Story Right' cross agency data linkage project (GOSR algorithm)[12] to improve Indigenous identification in both cases of SUDI (numerator) and infant population data (denominator) to compare the incidence of SUDI and measure disparities between Indigenous and non-Indigenous infants.

Methods

Study sample and data collections

All cases of SUDI that occurred in Queensland between 1 January 2010 and 31 December 2014 were extracted from the Queensland Child Death Register for analysis. Cases of SUDI were identified following review of all coronial infant deaths. Deaths that met all the following criteria were classified as SUDI:

- an infant aged less than one year
- death was sudden, with no previously known condition likely to cause death
- death was unexpected (death was not anticipated as a significant possibility in the preceding 24 hours), and
- the cause of death was not immediately obvious.

This included infants who were temporarily but unsuccessfully resuscitated. The data selected for this study included five administrative collections: Death Registrations, Birth Registrations, Perinatal Data Collection (hereafter perinatal), Queensland Hospital Admitted Patient Data Collection (hereafter inpatient hospitalisations), and the Emergency Department Information System (hereafter emergency presentations). Details of these administrative data collections are provided in Supplementary Table 1. Additionally, an infant's Indigenous status as captured in coronial and child protection records was extracted manually and entered into the study database following a review of coronial investigation documents and Queensland Family and Child Commission case files (see Child Death Register entry in Supplementary Table 1).

Data linkage

Linked Birth and Death Registration data were extracted from the Queensland Child Death Register and provided by the Queensland Family and Child Commission. Linked perinatal, inpatient hospitalisations, and emergency presentations data for infants and their mothers (during the period from conception to an infant's death) were obtained from Queensland Health's Master Linkage File. The Master Linkage File contains linked references to multiple health-related data collections and registries in Queensland. The Statistical Analysis and Linkage Unit were provided with the following identifying information from the Child Death Register extract to facilitate data linkage: death registration number, infant's full name, infant's date of birth, infant's date of death, infant's sex, infant's address, mother's full name, mother's maiden name, mother's date of birth, and mother's address.

The Statistical Analysis and Linkage Unit (Queensland Health) performed the data linkage using deterministic and probabilistic matching based on multiple data fields, name compression algorithms, and multiple matching passes. Extensive clerical review of all potential linkages that were not identified as definite matches was also undertaken to improve the linkage rate. The researchers were provided anonymised linkage keys to merge the information from the linked perinatal, inpatient hospitalisations and emergency presentations data with the Child Death Register extract.

Multi-stage algorithm to derive Indigenous status

To address the false negative reporting of Indigenous status in SUDI mortality data, the GOSR multistage algorithm was used in this study.[12] Table 1 shows the methodology used to combine an infant's Indigenous statuses within and between collections to determine their overall Indigenous status. Within hospital data collections (perinatal, inpatient hospitalisations, emergency presentations), both maternal and infant data were utilised in a stepwise manner to derive an infant's Indigenous status for each collection. In deriving an infant's Indigenous status from maternal data, because the father's Indigenous status was not collected, the father's Indigenous status was assumed to be missing. The decision process used to derive an infant's Indigenous status from parental data is described in the original GOSR paper and outlined in Supplementary Table 2.

An overall Indigenous status for each infant was derived from the collection derived Indigenous statuses in the seven datasets as follows:

1. Death registration – infant status as reported
2. Birth registration – mother and father used to derive infant status
3. Perinatal
    a. Perinatal mother – mother status with father status as missing used to derive infant status
    b. Perinatal infant – infant status as reported, from 1 July 2010
4. Inpatient hospitalisations
    a. Inpatient hospitalisations mother – mother status with father status as missing used to derive infant status
    b. Inpatient hospitalisations infant – infant status as reported
5. Emergency presentations
    a. Emergency presentations mother – mother status with father status as missing used to derive infant status
    b. Emergency presentations infant – infant status as reported
6. Coronial infant – infant status derived from manual review of coronial investigation documents
7. Child protection infant – infant status derived from manual review of child protection documents.

All seven collections reported people as belonging to one of the following groups: Aboriginal; Torres Strait Islander; both Aboriginal and Torres Strait Islander; neither Aboriginal nor Torres Strait Islander; or not stated/unknown. In this study, Aboriginal and/or Torres Strait Islander people were considered as one group and referred to as Indigenous.

An infant's Indigenous status as recorded in the Death Registration data was used as the Indigenous status variable to provide a baseline for comparison for algorithm-derived data. Death Registration data forms the basis of ABS mortality data and is used to report Indigenous status in the Queensland Family and Child Commission's annual reports of child deaths in Queensland.[9,26]

Table 1: Getting Our Story Right (GOSR) multi-stage algorithm.

Collection Derived Indigenous Status (within collections)	Step 1	Missing records within a collection cannot be used to determine a derived Indigenous status. Where all records are missing Indigenous status within a collection then collection-derived status = missing (for that collection)
	Step 2	Where there is only one non-missing record within a collection, this becomes the infant's derived Indigenous status for that collection.
	Step 3	Where there are two non-missing records within a collection and at least one identifies a person as Indigenous then collection-derived Indigenous status = Indigenous. If both records = non-Indigenous then collection-derived Indigenous status = non-Indigenous.
	Step 4	Where there are three or more non-missing records within a collection and two or more = Indigenous then collection-derived Indigenous status = Indigenous. If there are three or more non-missing records and two or more = non-Indigenous then collection- derived Indigenous status = non-Indigenous.
Steps 1-4 give a person a derived Indigenous status for each collection in which they have records. This process is then then repeated with the derived status for each collection in steps 5-8 below.		
Overall Indigenous Status (between collections)	Step 5	Collections with collection-derived statuses of missing cannot be used to determine an overall Indigenous status. If an infant only has records in collections where collection-derived status = missing, then overall Indigenous status = missing.
	Step 6	Where there is only one non-missing collection-derived status, this becomes the infant's overall Indigenous status.
	Step 7	Where there is a total of two non-missing collections and at least one collection derived Indigenous status = Indigenous then overall Indigenous status = Indigenous. If both collection-derived statuses = non-Indigenous, then overall Indigenous status = non-Indigenous.
	Step 8	Where there are three or more non-missing collections and two or more = Indigenous then overall Indigenous status = Indigenous. Otherwise overall Indigenous status = non-Indigenous.

Note: Based on Appendix One: the multistage median algorithm methodology in Christensen D, Davis G, Draper G, Mitrou F, McKeown S, Lawrence D, et al. Evidence for the use of an algorithm in resolving inconsistent and missing Indigenous status in administrative data collections. *Australian Journal of Social Issues.* 2014;49(4):423-43.

Calculations of infant death rates

The above described GOSR multi-stage algorithm was applied to linked birth registration, perinatal, inpatient hospitalisations, and emergency presentations data to provide enhanced Indigenous status in Queensland live births data for the calculation of infant death rates. The Perinatal Data Collection was defined as the primary data collection for the extract of all live births that occurred in Queensland between 2010 and 2014. Data linkage and extraction was performed by the Statistical Analysis and Linkage Unit (Queensland Health) using the Master Linkage File. The researchers were provided with confidentialised, aggregate data, broken down by Indigenous status, remoteness areas (Accessibility/Remoteness Index of Australia [ARIA+]) and the Socio-Economic Indexes for Areas (SEIFA).

Comparative, unenhanced infant mortality rates were calculated using the baseline Indigenous status variable and ABS live births data for the five-year period to 31 December 2014.[27] This is based on calendar year of live births registrations in Queensland.

Analysis

Data linkage rates, the contribution of individual data sources, and the contribution of maternal data to the overall enhancement were assessed. Using frequencies and chi-square tests, the impact of selected sociodemographic factors (geographical location, area level disadvantage, family type, parental age, parental employment status) on the recording of Indigenous status in baseline and algorithm-derived data were assessed. Numbers, proportions, and rates of Indigenous SUDI between 2010 and 2014 resulting from the application of the GOSR multi-stage algorithm to SUDI cases and live births data were compared with baseline data. Percentage enhancement was calculated by expressing the difference between the number and rate of algorithm-derived deaths and the baseline as a fraction of the baseline number of deaths.

Ethics and related approvals

Ethical approval for this study was provided by the University of the Sunshine Coast Human Research Ethics Committee (HREC: S/15/805). Ethical review specifically considered the extent to which this study accorded with the National Health and Medical Research Council's principles and guidelines for research concerning Aboriginal and Torres Strait Islander peoples, and the extent to which the research was respectful and of benefit to Indigenous people and communities. The original cross agency data linkage project through which the GOSR algorithm was developed included Aboriginal representation on the research team as well as consultation with community stakeholders, data custodians and Aboriginal researchers regarding self-identification, the development of a derived Indigenous status indicator and potentially sensitive issues.[12] Permission to access Child Death Register data was granted under the *Family and Child Commission Act 2014*. Permission to access coronial documents was granted under the *Coroners Act 2003*. The *Public Health Act 2005* approval provided permission to access and link perinatal, in-patient hospitalisations and emergency presentations data.

Results

A total of 228 cases of SUDI occurred in Queensland between 1 January 2010 and 31 December 2014. Death Registration data and coronial records were available for all cases. Linkage of death registration to births registration data was poor, with only 177 birth registration records able to be linked (77.6%). Linkage rates were high for both the perinatal and inpatient hospitalisations data (n=220, 96.5% each). Seven infants were born outside Queensland (3.1%); this reduced the linkage rates of birth registration, perinatal and inpatient hospitalisations data. The data from the Emergency Department Information System were limited (n=124, 54.4%), as an emergency department presentation was required during a mother's pregnancy or an infant's life for an individual to be identified in this dataset. Likewise, child protection data were also limited, as this is only captured in QFCC records if an infant's family had contact with child protection services in the three years prior to the infant's death.

Table 2 shows the data linkage rates and the contribution of individual data sources to the algorithm-derived Indigenous status. Overall the use of multiple sources of Indigenous status resulted in a 64.9% increase in the number of infants identified as Indigenous, with 37 Indigenous infants identified in death registration data (baseline), compared to 61 in algorithm-derived data. The use of multiple infant data sources accounted for 91.7% (22/24) of the increase in Indigenous infants. However, the addition of no single data collection alone (to the death registration data) was sufficient to account for more than 50% of the enhancement (11/22). Coronial data provided the best individual improvement in the reporting of Indigenous status, accounting for 50% of the enhancement. However, in this study, this data source required manual extraction of Indigenous status from narrative records and is not commonly used in routine data linkage activities. (The National Coronial Information System – an online repository of coronial data from Australia and New Zealand – may be used an alternative to manual extraction to provide Indigenous status information for reportable deaths.) The combination of infant, perinatal, inpatient hospitalisations, and birth registration data accounted for a large proportion of the enhancement (18/22, 81.8%). The addition of maternal health data sources only led to the identification of two further infants.

Under-identification was significantly higher in the highest socioeconomic areas (7/10, 70.0%) compared to moderate and low socioeconomic areas (18/51, 35.3%), *p*=0.04. Under-identification of Indigenous status was not associated with any other demographic factor examined (see Supplementary Table 3).

Applying the GOSR algorithm at population level, using both infant and maternal data, increased the proportion of Indigenous live births from 8.4% (ABS live births data) to 9.1% (algorithm-derived live births data based on perinatal data). To test whether infant data also accounted for the vast majority of the increase in Indigenous identification at population level, maternal health data was excluded from calculation of algorithm-derived live births data. Using multiple infant data sources only, the proportion of Indigenous live births increased to 8.8%.

Table 3 compares the number and rates of SUDI per 1000 live births, using the baseline (unenhanced) and the algorithm-derived Indigenous status variable. The Indigenous SUDI rate increased from 1.38 per 1,000 live births to 2.13 per 1,000 live births following the application of the of GOSR algorithm to numerator (cases of SUDI) and denominator (live births) data derived from both infant and maternal data sources. This represents a rate increase of 0.75 per 1,000 live births, an increase of 54.3%.

Discussion

Data linkage can be used to improve Indigenous identification in death records. A wide range of algorithms to address quality concerns with Indigenous status data were initially considered in this project.[12,13,15,22-24] A number have been evaluated previously, showing varying levels of enhancement over and above the baseline Indigenous status variable, as well as variations in estimated disparities between Indigenous and non-Indigenous people in morbidity and mortality outcomes.[12,13,17,28] Multi-stage algorithms – which combine information from both within and across datasets such that an individual is ascribed an Indigenous status for each collection, with the collection-derived statuses combining into an overall Indigenous status for that individual – provide a more nuanced version of simple frequency-based algorithms, and reach a compromise between the over- and under-inclusive approaches.[12,13] The GOSR multi-stage algorithm used in this study has previously been used to inform the National Best Practice Guidelines for Data Linkage activities Relating to Aboriginal and Torres Strait Islander people.[21] It has been shown to produce results robust against error in individual collections by combining all available information about the Indigenous status of any given individual within a linked dataset to find a credible derived status.[12]

In this study, the GOSR algorithm was highly effective at increasing Indigenous identification among infants who died suddenly and unexpectedly, even with a median age of just 10.8 weeks (IQR 6.7-18.6). The use of multiple sources of Indigenous status resulted in an additional 24 cases of SUDI (10.5% of all cases) being identified as Indigenous infants. Indigenous status in mortality data collections is considered acceptable for data from the Northern Territory, South Australia, Western Australia, and Queensland.[26] In Queensland, this was based solely on death registration data during the study period. Indigenous status is also derived from the Medical Certificate of Cause of Death (MCCD) for South Australia, Western Australia, Tasmania, the Northern Territory and the Australian Capital Territory from 2007. Since 2015, the Queensland Registry of Births, Deaths and Marriages also used MCCD information to derive Indigenous status. In this study, only 16.2% (37/228) of

Table 2: Number of infants identified as Indigenous, linkage rates and contribution of infant data sources to enhancement of infant Indigenous status for the Queensland SUDI study 2010-2014.

Data source	Number identified using each data source alone (infant data only)	Number identified using each data source alone (mother & infant data)	Number identified adding each data source to death registration data (infant only)	Number identified adding each data source to death registration data (mother & infant)	Number identified when data source excluded (infant only)	Number identified when data source excluded (mother & infant)	Number (%) of SUDI cases linked to each data source
Death registration	37[a]	37[a]	37[a]	37[a]	59[b]	61[b]	
Birth registration	43		55		54		177 (77.6%)
Perinatal	41	50	53	58	57	59	220 (96.5%)
Inpatient hospitalisations	43	50	54	57	58	60	220 (96.5%)
Emergency presentations	17	27	43	48	59	61	124 (54.4%)
Coronial	48		54		55		228 (100%)
Child protection	21		43		57		91 (39.9)

Notes:

a: Number of Indigenous infants identified using death registration as the sole data source from which official infant mortality rates are determined (9, 26).

b: Number of Indigenous infants identified with enhancement from all data sources

SUDI were identified as Indigenous using death registration (baseline) data, compared with 26.8% (61/228) of cases using algorithm-derived data. Overall, this constitutes a 64.9% increase in the number of sudden infant deaths identified as Indigenous. The high rate of under-identification of Indigenous status in death records has previously been noted.[15,16] Previous studies have found that under-enumeration of Indigenous people occurs more frequently in urban areas where the proportion of the population that is Indigenous is lowest, as well as among those that are least disadvantaged.[13,16] However, in this study under-identification of Indigenous infants was only found to be related to area-level socioeconomic advantage.

The majority of studies of health outcomes using data linkage to increase Indigenous identification have involved adult populations, with relatively few focusing on infants or very young children.[17,29-32] A specific challenge arises when using administrative data to measure infant mortality. Due to their young age at death, most infants will have had limited contact with health and other administrative systems and lack the opportunity to have their Indigenous status captured in multiple presentations across a range of datasets. Greater increases in Indigenous identification occur among older age groups, due to their increased opportunity to utilise health services.[13] As this study involves infants under the age of one year, it was hypothesised that relying solely on infant records was likely to under-identify Indigenous infants. To ameliorate this potential limitation, maternal hospital data were also used to derive an infant's Indigenous status. However, the additional use of multiple maternal data sources was found to account for only 8.4% (2/61) of the enhancement. To test whether the infant data had the same impact at population level, a re-calculation of live births data, excluding maternal health data, was obtained. In contrast to SUDI cases, maternal data had a substantial impact on population data, accounting for nearly half of the enhancement. At population level, the use of infant data sources only significantly under-estimated the number of Indigenous live births. This suggests that the use of both maternal and infant data from multiple sources to derive Indigenous status in both numerator and denominator data is optimal in research involving infant and young child populations.[32]

Table 3: Numbers of deaths, populations, and SUDI rates and changes by infant Indigenous status.

	Indigenous Deaths n (%)	Non-Indigenous Deaths n (%)	Indigenous Births n (%)	Non-Indigenous Births n (%)	Indigenous SUDI rate per 1,000 live births	Non-Indigenous SUDI rate per 1,000 live births
Indigenous status as per death registration[a]	37	191	26,726	291,307	1.38	0.66
Indigenous status as enhanced by infant data sources only	59	169	28,696	286,239	2.06	0.59
Change from death registration data (percentage change)	+22 (+59.5%)	-22 (-11.5%)	1970 (+7.4%)	-1970 (-0.7%)	+0.68 (+49.3%)	-0.07 (-11.9%)
Indigenous status as enhanced by infant and maternal data sources	61	167	28,696	286,239	2.13	0.58
Change from death registration data (percentage change)	+24 (+64.5%)	-24 (-14.2%)	1970 (+7.4%)	-1970 (-0.7%)	0.75 (+54.3%)	-0.08 (-12.1%)

Note:

a: official published Indigenous rate is based on an Indigenous population of 26,726 according to ABS Live Births Cat. No. 3301.0 2015, calculations in this manuscript have used data prepared by the Statistical Analysis and Linkage Unit based on live births identified in the Perinatal Data Collection. Total births estimate was 318,033 for official estimate and is 314,935 for GOSR algorithm-derived rates. For the purpose of the calculation of rates using the GOSR algorithm-derived live births data, non-Indigenous is an aggregate category including both non-Indigenous and missing/unknown status. The overall SUDI rate was 0.72 per 1000 live births using either denominator.

Studies using data linkage to improve the completeness of Indigenous status usually calculate crude rates using Census-derived estimated resident population (ERP) data or live births data based on birth registrations.[13,14,22,23,28] Higher estimates of mortality and disease rates are thus to be expected, irrespective of the algorithm applied, as this increases the number of Indigenous events (and correspondingly decreases non-Indigenous events) but does not change the population denominators. The present study is the first study of SUDI to apply a multi-stage algorithm to enhance Indigenous status in both numerator (cases of SUDI) and denominator (live births) data. While the manual extraction of Indigenous status data from coronial and child protection records had the potential to result in incongruence between numerator and denominator, almost all of the increase in Indigenous identification came from routinely linked administrative data sources that used both numerator and denominator. This substantially reduces the risk that an increase in Indigenous SUDI mortality rates may be due to numerator–denominator incompatibility, a noted limitation of previous research in this area, which affects the ability to accurately measure the gap in mortality between Indigenous and non-Indigenous infants. Enhancement of both cases of SUDI and live births data in this study resulted in a significant increase in the Indigenous SUDI mortality rate, from 1.38 per 1,000 live births to 2.13 per 1,000 live births. While there have been some improvements in Indigenous infant mortality rates and a narrowing of the gap between Indigenous and non-Indigenous infants over the longer term,[5,9] official state and national child death statistics for Queensland continue to report an Indigenous SUDI rate only twice that of non-Indigenous infants.[9,33] The findings of the present study suggest that Indigenous infants in Queensland die suddenly and unexpectedly at a rate that is more than 3.5 times that of non-Indigenous infants; a much higher rate than previously estimated by official statistics

Limitations

Due to their young age at death, most infants have had limited contact with health and other administrative systems and lack the opportunity to have their Indigenous status captured in multiple presentations across a range of datasets. Consequently, studies of infant health outcomes rely heavily on birth registration and perinatal data collections.[29,31,34] A low rate of linkage of birth and death registration in the SUDI data was observed (77.6%; with only 3.1% explained by infants born outside Queensland), although the reasons for this are unclear. It may relate to operational processes since this data (obtained from the Child Death Register) was linked by the Registry of Births Deaths and Marriages, which does not have a dedicated data linkage and statistical analysis team. Compounding this issue, Indigenous births are registered at a significantly lower rate than non-Indigenous births.[24,31,34] A recent Queensland investigation found that 15–18% of births to Aboriginal mothers were not registered compared with an under-registration rate of 1.8% for births to non-Aboriginal mothers.[35] There is the potential for under-identification of Indigenous infants due to the low linkage rate of birth registrations in the SUDI data.

A study by Comino and colleagues[29] of the identification of Aboriginal infants within antenatal services showed that, in an urban area, identification based on the Indigenous status of an infant's mother only significantly under-identified Aboriginal infants, with nearly 40% of Aboriginal infants having a non-Indigenous mother. In the present study, the addition of maternal data yielded minimal enhancement over and above that obtained by using multiple infant records. It is possible that the inclusion of paternal inpatient hospitalisations and emergency presentations data may have resulted in greater enhancement; however, only maternal data was available.

An under-explored limitation of all enhancement algorithms concerns the aggregation of three distinct Indigenous statuses: Aboriginal, Torres Strait Islander, and both Aboriginal and Torres Strait Islander. This is due to the practical difficulty of developing an algorithm that accounts for all possible combinations. However, the potential impact of the inability to distinguish between these two distinct cultural groups should be acknowledged, particularly in research undertaken in Queensland, which has a significantly larger Torres Strait Islander population than any other state or territory.[36]

Strengths

This study shows that the use of unenhanced Indigenous status may lead to an under-enumeration of Indigenous persons and an under-reporting of key outcomes, such as deaths. It further demonstrates that the application of the GOSR algorithm to multiple linked infant and maternal data sources is highly effective at increasing Indigenous identification in studies of infant mortality. By applying the GOSR algorithm to both numerator and denominator data, this study ameliorated numerator–denominator incongruency, to more accurately report the rates of Indigenous infant mortality and measure the gap in SUDI mortality. As they become available, additional data collections can (and should) be added to this algorithmic approach to determining Indigenous status in both numerator and denominator data with relative ease, resulting in continuous improvement to data quality.

Implications for public health

This study illustrates that the linkage of existent data resources can substantially improve measures of Indigenous health that are based on mortality data. More specifically, with the rapid growth in the use of linked data, the GOSR algorithm has the potential to increase Indigenous status data quality in research and government reporting. Accurate and reliable epidemiological and statistical data provide the foundation for evidence based public policy and are critical to improving Indigenous health and reducing the health disparities between Indigenous and non-Indigenous people. It remains important to recognise that Indigenous identity and identification reflects a complex interplay of individual and contextual factors.[12] Ongoing work aimed at improving the quality of administrative data collections should continue.

A key finding of this study – that Indigenous infants died suddenly at a rate more than 3.5 times that of non-Indigenous infants – has serious policy and practice implications. Specifically, it calls into question the continued use of unenhanced data in official Queensland child death statistics, including that of state-based child death review bodies,[9] the Australian Bureau of Statistics cause of death data,[26] and official reporting of against the COAG Closing the Gap initiative,[33,37] all of which presently rely on unenhanced Indigenous status as recorded in death certification and registration data. Much of this deficiency in official statistics could be overcome by the routine linking of readily available data sources, particularly birth registration and perinatal data.

Acknowledgements

We acknowledge and thank the data custodians – the Queensland Family and Child Commission, the Office of the State Coroner, and Queensland Health – and their research and data management staff for providing access to the data that made this research possible. We especially acknowledge the contribution of the Queensland Health Statistical Analysis and Linkage Unit (SALU) staff, in particular Dale Steinhardt, for data extraction and linkage.

Funding

Rebecca Shipstone was supported by a University of the Sunshine Coast Research Scholarship – PhD Candidate Funding.

Data accessibility statement

The datasets used and/or analysed during the current study are not available publicly as this was not specified in the original ethical approval request; however, they may be available in part from the corresponding author on reasonable request.

References

1. Thompson SC, Woods JA, Katzenellenbogen JM. The quality of Indigenous identification in administrative health data in Australia: Insights from studies using data linkage. *BMC Med Inform Decis Mak*. 2012;12:133.
2. Australian institute of Health and Welfare. *The Health and Welfare of Australia's Aboriginal and Torres Strait Islander Peoples 2015*. Catalogue IHW N0.: 147. Canberra (AUST): AIHW; 2015.
3. Campbell P, Biddle N, Paradies Y. Indigenous identification and transitions in Australia: Exploring new findings from a linked micro-dataset. *Population*. 2018;73(4):771-96.
4. Council of Australian Governments. *Glosing The Gap on Indigenous Disadvantage: The Challenge for Australia*. Canberra (AUST): COAG; 2009.
5. Department of the Prime Minister and Cabinet. *Closing The Gap Report 2019*. Canberra (AUST): Government of Australia; 2019.
6. Kim D, Saada A. The social determinants of infant mortality and birth outcomes in Western developed nations: A cross-country systematic review. *Int J Environ Res Public Health*. 2013;10(6):2296-335.
7. Mitchell EA, Cowan S, Tipene-Leach D. The recent fall in postperinatal mortality in New Zealand and the safe sleep programme. *Acta Paediatr*. 2016;105(11):1312-20.
8. Shipstone R, Young J, Kearney L. New frameworks for understanding sudden unexpected deaths in infancy (SUDI) in socially vulnerable families. *J Pediatr Nurs*. 2017;37:35-41.
9. Queensland Family and Child Commission. *Annual Report: Deaths of Children and Young People, Queensland, 2017–18*. Brisbane (AUST): The State of Government; 2018.
10. Australian Institute of Health and Welfare. *A Picture of Australia's Children 2012*. Catalogue No.: PHE 167. Canberra (AUST): AIHW; 2012.
11. Hauck FR, Tanabe KO. International trends in sudden infant death syndrome: Stabilization of rates requires further action. *Pediatrics*. 2008;122(3):660-6.
12. Christensen D, Davis G, Draper G, Mitrou F, McKeown S, Lawrence D, et al. Evidence for the use of an algorithm in resolving inconsistent and missing Indigenous status in administrative data collections. *Aust J Soc Issues*. 2014;49(4):423-43.
13. Randall DA, Lujic S, Leyland AH, Jorm LR. Statistical methods to enhance reporting of Aboriginal Australians in routine hospital records using data linkage affect estimates of health disparities. *Aust N Z J Public Health*. 2013;37(5):442-9.
14. Draper GK, Somerford PJ, Pilkington AAG, Thompson SC. What is the impact of missing Indigenous status on mortality estimates? An assessment using record linkage in Western Australia. *Aust N Z J Public Health*. 2009;33(4):325-31.

15. Neville SE, Taylor LK, Moore H, Madden R, Ring I, Pulver LJ, et al. Using linkage between hospital and ABS mortality data to enhance reporting of deaths among Aboriginal and Torres Strait Islander peoples. *Aust N Z J Public Health*. 2011;35(6):543-8.
16. Briffa TG, Sanfilippo FM, Hobbs MS, Ridout SC, Katzenellenbogen JM, Thompson PL, et al. Under-ascertainment of Aboriginality in records of cardiovascular disease in hospital morbidity and mortality data in Western Australia: A record linkage study. *BMC Med Res Methodol*. 2010;10:111.
17. Lawrence D, Christensen D, Mitrou F, Draper G, Davis G, McKeown S, et al. Adjusting for under-identification of Aboriginal and/or Torres Strait Islander births in time series produced from birth records: Using record linkage of survey data and administrative data sources. *BMC Med Res Methodol*. 2012;12(1):90.
18. Taylor LK, Bentley J, Hunt J, Madden R, McKeown S, Brandt P, et al. Enhanced reporting of deaths among Aboriginal and Torres Strait Islander peoples using linked administrative health datasets. *BMC Med Res Methodol*. 2012;12:91.
19. Blakely T, Tobias M, Robson B, Ajwani S, Bonne M, Woodward A. Widening ethnic mortality disparities in New Zealand 1981-99. *Soc Sci Med*. 2005;61(10):2233-51.
20. Anderson RN, Copeland G, Hayes JM. Linkages to improve mortality data for American Indians and Alaska Natives: A new model for death reporting? *Am J Public Health*. 2014;104 Suppl 3:258-62.
21. Australian Institute of Health and Welfare and Australian Bureau of Statistics. *National Best Practice Guidelines for Data Linkage Activities Relating to Aboriginal and Torres Strait Islander People*. Catalogue No.: IHW74. Canberra (AUST): AIHW; 2012.
22. Rowe SL, Cowie BC. Using data linkage to improve the completeness of Aboriginal and Torres Strait Islander status in communicable disease notifications in Victoria. *Aust N Z J Public Health*. 2016;40(2):148-53.
23. Watkins RE, Mak DB, Giele CM, Clews S. Aboriginal and non-Aboriginal sexually transmitted infections and blood borne virus notification rates in Western Australia: Using linked data to improve estimates. *BMC Public Health*. 2013;13:404.
24. Gialamas A, Pilkington R, Berry J, Scalzi D, Gibson O, Brown A, et al. Identification of Aboriginal children using linked administrative data: Consequences for measuring inequalities. *J Paediatr Child Health*. 2016;52(5):534-40.
25. Hunter B, Ayyar A. Undercounts in offender data and closing the gap between Indigenous and other Australians. *Aust J Soc Issues*. 2011;46(1):69-89.
26. Australian Bureau of Statistics. *3303.0 - Causes of Death, Australia 2017*. Canberra (AUST): ABS; 2018.
27. Australian Bureau of Statistics. *3301.0 - Births, Australia 2015*. Canberra (AUST): ABS; 2016.
28. Katzenellenbogen JM, Sanfilippo FM, Hobbs MS, Briffa TG, Ridout SC, Knuiman MW, et al. Incidence of and case fatality following acute myocardial infarction in Aboriginal and non-Aboriginal Western Australians (2000-2004): A linked data study. *Heart Lung Circ*. 2010;19(12):717-25.
29. Comino EJ, Titmuss A, Harris E, Craig P, Pulver LJ. Identification of Aboriginal infants at an urban hospital. *J Paediatr Child Health*. 2007;43(9):623-6.
30. Kennedy B, Cornes S, Martyn S, Wills R, Breckell C. *Measuring Indigenous Perinatal Outcomes - Should We use the Indigenous Status of the Mother, Father or Baby?* Technical Report No.: 2. Brisbane (AUST): Queensland Health; 2009.
31. Xu F, Sullivan EA, Madden RC, Black D, Jackson Pulver LR. Improvement of maternal Aboriginality in NSW birth data. *BMC Med Res Methodol*. 2012;12:8.
32. Falster K, Jorgensen M, Hanly M, Banks E, Brownell M, Eades S, et al. Data resource profile: Seeding success: A cross-sectoral data resource for early childhood health and development research in Australian Aboriginal and non-Aboriginal children. *Int J Epidemiol*. 2017;46(5):1365-6.
33. Australian Health Ministers' Advisory Council. *Aboriginal and Torres Strait Islander Health Performance Framework, 2017 Report*. Canberra (AUST): Department of Prime Minister and Cabinet; 2017.
34. Gibberd AJ, Simpson JM, Eades SJ. No official identity: A data linkage study of birth registration of Aboriginal children in Western Australia. *Aust N Z J Public Health*. 2016;40(4):388-94.
35. Queensland Ombudsman. *The Indigenous Birth Registration Report*. Brisbane (AUST): State Government of Queensland; 2018..
36. Australian Bureau of Statistics. *3238.0.55.001 - Estimates of Aboriginal and Torres Strait Islander Australians, June 2016*. Canberra (AUST): ABS; 2018.
37. Queensland Health. *Closing The Gap Performance Report 2017*. Brisbane (AUST): State Government of Queensland; 2018.

Supporting Information

Additional supporting information may be found in the online version of this article:

Supplementary Table 1: Data collections used to source information for the Queensland SUDI Study.

Supplementary Table 2: The decision process used to derive an infant's Indigenous status from parental data.

Supplementary Table 3: Associations of socio-demographic factors and Indigenous identification using the GOSR algorithm.