

A comprehensive methodology report for maternity and perinatal mortality research in the New Zealand Integrated Data Infrastructure

Esti C de Graaff, Lynn Sadler, Lesley McCowan, Robin Cronin, Ngaire Anderson

ABSTRACT

AIM: The highest quality perinatal data in New Zealand is collected and collated by the Perinatal and Maternal Mortality Review Committee (PMMRC) and is made available to a limited number of researchers. Therefore, maternity, and perinatal mortality studies are generally performed on Government-held data. This report offers an alternative approach with in-depth justification for the methodology, while simultaneously improving the understanding of the data sources.

METHOD: A standardised method for creating a comprehensive maternity dataset within the Statistics New Zealand Integrated Data Infrastructure (IDI) was developed and a validation dataset was created to include all births between 2008 and 2017.

RESULTS: A close approximation to the PMMRC annual report data was found, with 4.0% over-reporting of perinatal deaths and 0.05% over-reporting of live births in the IDI dataset. Several variables, including important pregnancy risk factors, were validated for use. Limitations to the datasets were explored and additional tables in the IDI were proposed, to include variables on pregnancy complications, ethnicity and country of birth, and socio-economic data.

CONCLUSION: This methodological report describes an opportunity for standardised, high-quality maternity research in New Zealand using the IDI, including a variety of national data sources. Recommendations for further enhancement of these resources have been offered.

Observational studies are a useful tool in epidemiology.¹ One study revealed that approximately 68% of published articles in the four leading United States obstetrics and gynaecology journals were of observational nature.² In obstetric research, they provide the opportunity to study relatively rare adverse events like stillbirth or neonatal death (NND). Despite their value, observational studies come with biases and investigators have an obligation to identify and mitigate these,³ which includes adequate reporting of the study design and methodology.⁴

Epidemiological maternity research in New Zealand is usually performed on national Government-held data. The highest quality perinatal data in the country is collected and held by the Perinatal and Maternal Mortality Review Committee (PMMRC), which is an independent committee under the Health Quality & Safety Commission New Zealand. PMMRC data is only made available to a limited number of researchers, due to data sovereignty issues. An alternative data source for maternity research in New Zealand is provided by the Statistics New Zealand (NZ) Integrated

Data Infrastructure (IDI). The IDI is a collection of New Zealand Government and non-government administrative and survey data, held by Statistics NZ. Researchers are granted access to merge and interrogate data sources only available within the strict privacy rules of the IDI environment; as projects are required to meet all “five safes” (safe people, safe projects, safe settings, safe data, and safe output).⁵ Data in the IDI is linked at the individual level, which allows personal information to be connected across different sectors (such as income, migration, and health). Each person is provided with a unique identification (ID) number in a central “spine”, by which the various datasets can be joined. All people ever resident in New Zealand (i.e., citizens or those with visas that allow residency, work or study) and captured in one of the data sources, are included. The IDI data are “refreshed” (i.e., updated, to include newer data, and additional data sources) up to four times a year. By connecting data from multiple sources not otherwise linked on an individual level, questions around complex issues can be researched with high quality across the population.⁶

As both maternal and infant data are generally necessary in perinatal analyses, creating datasets for maternity research can be complicated. This methodological report was prepared to assist New Zealand researchers in developing comprehensive datasets for national pregnancy studies, with a focus on perinatal death. We describe a standardised method for creating a “core” dataset within the IDI, allowing for consistent national reporting, and include suggestions for additional tables. We additionally aim to improve the understanding of the used datasets and variables.

Method

An application was made with Statistics NZ to use the data within the IDI. Once approved, a dataset for perinatal research was built in Microsoft SQL Management Server Studio (using IDI refresh IDI_Clean_20211020). National data sources used, were the Maternity Collection (MAT),⁷ Births, Deaths and Marriages (BDM),^{8,9} the National Minimum Dataset (NMDS),¹⁰ the Mortality Collection (MORT),¹¹ the Chronic Conditions dataset (CC),¹² Census,¹³ and Immigration data.¹⁴ “General” or “central” IDI tables used included the full birth date, full death date, and address notification tables.¹⁵ A succinct and essential summary of these data sources has been provided in Appendix 1. See Table 1 for an overview of abbreviations used in this methodological report. Finally, PMMRC publicly available annual report data was used as the gold standard to validate the tables created.¹⁶

A numerator table (including all perinatal deaths) and a denominator table (including all births) were created separately, to include all births between 2008 and 2017. The main dataset used was the MAT collection. Data quality of the MAT collection varies by item but has improved significantly since 2008.¹⁷ Thus, data from 2008 onward are most useful for perinatal research. The proposed method for creating the core dataset, as well as two corresponding full SQL codes (for a table in and excluding multiples) are provided in Appendices 2–4.

Pregnancy risk factors

Even though data quality has improved since 2008, some variables still have a high degree of missing data. This is particularly true for women cared for exclusively by district health boards (DHB), due to a funding change introduced in

July 2007. DHB employed midwifery teams are no longer required to claim for primary maternity services, while self-employed community-based lead maternity carers (LMCs; midwives, general practitioners, or obstetric specialists) must submit pregnancy data prior to payment from the Ministry of Health (MoH).¹⁸ As a result, completeness of some data for DHB-registered pregnancies varies widely, while the same data for non-DHB pregnancies is mostly complete. It should also be noted that since 2008 the variable “LMC type” is inaccurate for women under DHB care.¹⁷ Finally, when considering data quality issues, some variables can be taken from either MAT or BDM data sources. We validated the following variables for use: maternal age, fetal sex, gestational age, and birthweight.

Pregnancy complications

Pregnancy complications such as gestational diabetes (GDM) or hypertensive disorders of pregnancy are important outcomes in routine maternity research. The MAT delivery table in the IDI identifies births that were complicated by pre-eclampsia or eclampsia, identified by NMDS. MAT, however, does not identify pregnancies of mothers with pre-existing hypertensive disorders or gestational hypertension. Secondly, the indicator for GDM is not offered in the IDI as this field is incorrectly labelled. It indicates both pre-existing diabetes and GDM, and according to the MoH no validation process is undertaken for this field. For that reason, we propose to add NMDS and CC datasets to the core dataset. Unfortunately, data on primary care diagnoses are not available in the IDI. Hospital admissions can be joined to the correct pregnancy by maternal ID and admission dates. Following a similar method, mothers with pre-existing diabetes (as an important risk factor) can be identified by joining the CC dataset.

Ethnicity and country of birth

Ethnicity is an important demographic risk factor widely used in pregnancy research in New Zealand, and key to examining health equity. Characteristics of ethnicity recorded in New Zealand include that it is self-defined, it can change over time and an individual may identify with more than one ethnic group. The use of ethnicity data in health research is addressed by the *Ethnicity Data Protocols for the Health and Disability Sector* by the MoH.¹⁹ According to this protocol, ethnicity data can be categorized at four different “levels” following a hierarchical structure;

with level four being the most detailed level of reporting (containing 231 ethnicity codes). These codes are then aggregated into ethnicity levels three to one. As an illustration, code “43112” stands for Fijian Indian and aggregates into “431” Indian (level three), “43” Indian (level two) and “4” Asian (level one). Generally, level two ethnicity data are used in health research for reporting, which includes 22 groups. In this aggregation a high level of detail is maintained for some ethnicities (such as Māori, Pacific Peoples, Chinese or

Indian), while other minority groups are merged despite large heterogeneity (such as other Asian ethnicities, African or Latin American).

Unfortunately, the MAT dataset only holds level two ethnicity data. Due to the data collection methods of MAT, ethnicity data may also have been completed by a healthcare provider, leading to potential inaccuracy or lack of detail. Moreover, someone’s ethnicity can reflect a contextual response. This might occur, when a mother believes she will receive better care

Table 1: Abbreviations used in this methodological report (in alphabetical order).

Abbreviation	Definition
BDM	Births, Deaths and Marriages
BMI	Body mass index
CC	Chronic Conditions
COB	Country of birth
DHB	District Health Board
GDM	Gestational diabetes
ID	Identification
IDI	Integrated Data Infrastructure
LMC	Lead maternity carer
MAT	Maternity Collection
MOH	Ministry of Health
MORT	Mortality Collection
NMDS	National Minimum Dataset
NND	Neonatal death
NZ	New Zealand
NZDep	New Zealand Socioeconomic Deprivation Indices
PMMRC	Perinatal and Maternal Mortality Review Committee

when reporting a different ethnicity. In contrast, BDM birth registration provides high-quality level four ethnicity data, including information reported by the parents directly outside of the healthcare setting. The requirement for parents to complete birth registration separately from the LMC leads to ethnicity data akin to ethnicity reported in the national Census and is generally considered second choice to Census ethnicity data.²⁰ However, since ethnicity can change and BDM is recorded closer to the birth event than Census, we argue that BDM should be used as the main source in maternity studies. Note that in cases where ethnicity data is missing, the source-ranked ethnicity table (“central” table in the IDI) may be consulted, although only level one ethnicity is provided.

Some research questions will require both ethnicity (i.e., a subjective belief, related to cultural behaviours and practices) and country of birth (COB; i.e., an objective measure, more closely linked to ethnic origin), considering increasing migration and ethnic diversity globally.²¹ Combining these variables in analyses might provide a better understanding of pregnancy risk factors, since common issues associated with migration in first generation migrants (such as socio-economic deprivation or diverse health literacy) may result in differential health outcomes compared to second and third generation women.²²⁻²⁴ COB data are not available from the standard birth tables and should therefore be obtained from alternative datasets. Census or immigration data present as the highest quality sources for this variable. Since COB is fixed, the datasets can simply be linked by maternal ID, regardless of the correct pregnancy event. An alternative method for consideration, when solely interested whether a mother was New Zealand born or not, is to join parent ID on the infants BDM birth record, with the BDM births table. If the mother’s birth was registered, she was New Zealand born. In contrast, if the mother’s birth cannot be identified in BDM births, she was most likely born overseas.

Socio-economic data

The MAT or BDM datasets do not contain any information on individual level socio-economic status. The current classification system used to monitor deprivation, which is widely used in health and social research, is the New Zealand Socioeconomic Deprivation Indices (NZDep).²⁵ NZDep is a decile score based on area of domicile, divided into meshblocks or larger Census

Area Units where a higher level of confidentiality is required. The NZDep is based on census data; with decile one representing 10% of the population who live in the least deprived areas and decile ten representing 10% of the population who live in the most deprived areas in New Zealand. The NZDep2013 is the fifth updated version since 1991 and combines nine variables from the 2013 Census. A limitation of this method is that the NZDep represents area-level deprivation and does not necessarily reflect the socio-economic status of an individual, although it is a close approximation and may be used as a proxy in large datasets.²⁶ Alternative proxy variables for consideration, include the New Zealand Indices of Multiple Deprivation,²⁷ region of domicile (sourced from the address notifications table), DHB of domicile (sourced from MAT), or personal income level (sourced from Census), although the researchers of this paper have no experience using these alternative sources.

To provide an estimation of socio-economic status in our birth cohort, the registered address closest to the date of birth, and prior to delivery, was chosen for each delivery event, to allow for geographic movement over time, and to best capture mothers socio-economic status during pregnancy. Note that where a full date of birth is missing (primarily among perinatal deaths), birth year and month sourced from the BDM births or MAT table may be used. The meshblock associated with this address was then extracted and linked to the corresponding NZDep2006 and 2013 decile scores (births after 2008 and before 2013 linked to NZDep2006 and from 2013 linked to NZDep2013).

Results

Birth outcome

Six thousand, seven hundred and ninety perinatal deaths (4,768 stillbirths and 2,022 NNDs) and 617,375 live births were identified in our dataset. In comparison, PMMRC annual report data comprised 6,518 perinatal deaths (4,779 stillbirths, including 1,456 terminations, and 1,739 NNDs) and 617,321 live births.¹⁶ As a result, our numerator dataset includes 272 more perinatal mortalities compared to the gold standard (an approximate 4.0% overreporting in deaths). Our denominator dataset includes 54 more live births compared to PMMRC report data (a 0.05% overreporting in live births). Thus overall, our datasets represent a close approximation to the gold standard. We found that 99.2% of all births (99.5% of

live births and 77.4% of perinatal deaths) in our final dataset had a record from both MAT delivery and BDM births; providing complete data for the majority of cases. Cases that could not be joined to both sources, were mostly missing BDM data (example in Table 2).

Pregnancy risk factors

Smoking status at registration was unknown for 44.0% of women with a DHB-registered pregnancy, while this was only 0.04–1.5% for women under the care of an LMC provider. This was even higher for maternal BMI, with 58.0% and 0.1–2.2% of missing data among these groups respectively. The variable “booking trimester” was missing for 48.5% of women with a DHB-registered pregnancy, compared to 0.01–0.2% of women booked with another LMC type. 37.1% of DHB cases were noted as LMC type “unknown” in our dataset. The following variables were validated.

Maternal age

While maternal age is provided by MAT, it was calculated from maternal birth year, month and delivery date in BDM. In our cohort, there was a 95.6% overlap where this variable was available from both datasets. 95.4% of the 4.4% non-matches only differed by one year. Consequently, the MAT dataset may be used for maternal age (accounting for 99.9% of all cases).

Fetal sex

The overlap in fetal sex was 99.96% where both MAT and BDM variables were available. Hence, either variable can be used in analysis (accounting for 98.2% of all cases).

Gestational age

The MAT dataset notes gestational age in weeks, while BDM also registers days. There was an 89.1% overlap in gestational age week where both variables were available. Of the remaining 10.9%, it appeared that 0.7% only differed by one day (e.g., 37 and 36+6), while 7.0% differed by one day to one week (e.g., 39 and 40+0). Among cases in which a larger difference existed (3.2% of total), birthweight was more likely to correlate with MAT gestational age and therefore MAT may be prioritised for use. However, BDM should be used in analyses including customised birthweight centiles,²⁸ where the absence of gestation in days leads to systematic over-estimation of birthweight centiles. Using both tables, 98.0% of cases are accounted for.

Birthweight

The overlap in birthweight was 96.6% where both MAT and BDM variables were available. Of the remaining 3.4%, 1.7% only differed by 100 grams and 0.6% differed by more than 500 grams. Either variable can be used in analysis, accounting for 94.0% of all cases.

Socio-economic data

Close to 99.0% of deliveries could be linked to a NZDep score.

Ethnicity and country of birth

Level four ethnicity from the 2013 Census had an 89.4% overlap with BDM ethnicity, in cases where only one ethnic group was recorded in both sources (N=392,004). The corresponding overlap for level three ethnicity was 90.6% and level two ethnicity 95.7%. Therefore, if BDM is missing, Census provides a good alternative. If both are missing, MAT ethnicity can be used as a surrogate. This method may also mitigate some data quality differences between ethnicities, as the availability of BDM ethnicity data for perinatal deaths differs per group (Table 2).

Between 2008 and 2017 88.3% of mothers had a known COB from the 2018 Census. Where Census 2018 data was missing, Census 2013 data was used, with a 99.2% agreement between the two surveys among women where both were available. If both were missing, then immigration data was used, with an 89.7% and 88.8% agreement with Census 2018 and 2013 respectively. Immigration metadata suggests using nationality over COB, however in our dataset this resulted in less agreement with Census (82.1% and 80.7%). Finally, in this report nationality was used as a surrogate for COB if all other COB data was missing. This is justified by an 85.4% agreement between COB and nationality in the immigration dataset. By combining all four variables, COB was available for 98.9% of all mothers.

Discussion

This methodological paper describes an opportunity for standardised, high-quality maternity research in New Zealand using the IDI. A strength of this proposed approach is the ability to create a comprehensive dataset including perinatal deaths and live births from a variety of national sources, using our combined knowledge, and defining PMMRC data as the gold standard; thereby utilising the best quality data from

Table 2: Availability of ethnicity data per data source, for all perinatal deaths between 2008–2017.

	European		Māori		Pacific Peoples		Asian		MELAA	
	PND (N=3,036)	LB (N=316,284)	PND (N=1,710)	LB (N=140,553)	PND (N=855)	LB (N=62,982)	PND (N=963)	LB (N=86,781)	PND (N=126)	LB (N=10,221)
BDM births	2,646 (87.2%)	313,902 (99.2%)	1,248 (73.0%)	138,663 (98.7%)	621 (72.6%)	62,424 (99.1%)	783 (81.3%)	85,392 (98.4%)	93 (73.8%)	9,390 (91.9%)
Census 2013	2,592 (85.4%)	277,251 (87.7%)	1,215 (71.1%)	102,936 (73.2%)	579 (67.7%)	44,028 (69.9%)	666 (69.2%)	62,178 (71.6%)	75 (59.5%)	6,873 (67.2%)
MAT delivery	2,943 (96.9%)	315,996 (99.9%)	1,668 (97.5%)	140,445 (99.9%)	834 (97.5%)	62,835 (99.8%)	936 (97.2%)	86,670 (99.9%)	120 (95.2%)	10,194 (99.7%)

Abbreviations: PND = perinatal deaths; LB = live births; MELAA = Middle Eastern/Latin American/African; BDM = births, deaths, and marriages; MAT = maternity collection.
 Prioritisation: BDM births, Census 2013, MAT delivery.

each dataset available. All steps in creating this dataset have been justified and validated. Complete understanding of the data sources, including the quality of the variables used and general inconsistencies in metadata, will also improve the accuracy of research output. Since these data sources are available to all researchers who are granted permission to use the Statistics NZ IDI, this will increase accessibility.

In developing this methodology, some limitations to the IDI were discovered, such as restricted use of the MORT dataset. Even though MORT is considered the best source for stillbirths, this methodology uses the MAT and BDM datasets to create the numerator table. This is justified as while MORT identified 6,270 perinatal deaths between 2008 and 2017, only 1,955 (31.2%) of these could be matched to MAT or BDM tables. Thus, clinical data including important variables would be unavailable for almost 70% of all mortalities. We suspect this low matching rate is due to a linking error within the IDI, which should be addressed by Statistics NZ. In addition, only 59 cases had a different birth status according to MORT, validating this approach.

We also offer recommendations to improve the quality of perinatal data, to further enhance these resources. Firstly, the transfer of BMI and smoking data from DHB primary care facilities to the MAT datasets should be facilitated to eliminate systematic bias in analyses that control for these variables, as the highest degree of missing data is among high-risk mothers under DHB care, and who are also more likely to suffer perinatal mortalities. Consideration should be given to excluding DHB-registered pregnancies from analyses that require adjustment for these variables. For many years the PMMRC has recommended the MoH to “urgently require DHBs to provide complete and accurate registration data to the MAT dataset”, without success.²⁹ Additionally, the variable “booking trimester” was missing for almost half of women with a DHB-registered pregnancy, despite late booking being associated with poorer perinatal outcomes and socio-economic deprivation.^{30,31}

Collection of important obstetric risk factors such as maternal pre-existing chronic conditions, should be included in the MAT dataset. Others, such as GDM or hypertensive disorders of pregnancy, may need to be validated as the quality

of these variables is not clear. For instance, Lawrence et al. investigated the prevalence of GDM according to some commonly used data sources in New Zealand and found an underreporting in NMDS (3.8%, compared to 5.9% reported by DHBs or 6.9% reported by laboratories).³² There was 70% agreement on the presence of GDM between the data sources. We also suggest that validation studies of routine maternity datasets are performed. This will assist researchers in the interpretation of results of a widely used data source. Furthermore, pregnancy research requires both mother and infant data in most analyses. Mothers may appear in a table more than once with consecutive pregnancies or a pregnancy may result in multiple infants, which complicates the building of a perinatal dataset. Including stillbirths into the MAT infant dataset will provide more detailed information about the birth and simplify the process of creating a dataset. However importantly, although this methodology report offers quality improvement for maternity research, making the PMMRC national dataset available within the IDI should be considered. Use of this dataset would eliminate many data quality issues described in this paper associated with perinatal mortality studies.

Even though the IDI provides a promising avenue for perinatal studies, there are barriers to accessing the data. Each new project requires a comprehensive application process. New research projects are assessed seven times a year, with a turnaround time of approximately six weeks. Successful applications are required to pay a one-off fee (\$500). Once approved, specific users of the IDI will need to be authorised by Statistics NZ, undergo confidentiality training, and any changes to the project are subject to evaluation. Researchers are granted access to merge and interrogate data sources only available within the strict privacy rules of the DataLab; available in a few cities across New Zealand. Finally, researchers are recommended to have intermediate SQL coding skills.

In conclusion, this methodological report aims to improve the quality of routine maternity studies in New Zealand by offering an alternative approach to conventional data sources, while simultaneously increasing knowledge and accessibility.

COMPETING INTERESTS

Nil

ACKNOWLEDGEMENT

The authors would like to acknowledge the University of Auckland and the NHMRC Centre of Research Excellence in Stillbirth for scholarship support.

DISCLAIMER

These results are not official statistics. They have been created for research purposes from the Integrated Data Infrastructure (IDI) which is carefully managed by Stats NZ. For more information about the IDI please visit <https://www.stats.govt.nz/integrated-data/>.

AUTHOR INFORMATION

Esti C de Graaff: PhD Candidate, The University of Auckland Faculty of Medical and Health Sciences, Obstetrics & Gynaecology, Auckland, New Zealand.
Lynn Sadler: Honorary Research Fellow Obstetrics and Gynaecology, PhD co-supervisor, The University of Auckland Faculty of Medical and Health Sciences, Obstetrics & Gynaecology, Auckland, New Zealand.
Lesley McCowan: Emeritus Professor Obstetrics and Gynaecology, PhD co-supervisor, The University of Auckland Faculty of Medical and Health Sciences, Obstetrics & Gynaecology, Auckland, New Zealand.
Robin Cronin: Honorary Academic, The University of Auckland Faculty of Medical and Health Sciences, Obstetrics & Gynaecology, Auckland, New Zealand.
Ngairé Anderson: Senior Lecturer Obstetrics and Gynaecology, PhD main supervisor, The University of Auckland Faculty of Medical and Health Sciences, Obstetrics & Gynaecology, Auckland, New Zealand.

CORRESPONDING AUTHOR

Esti de Graaff: Faculty of Medical and Health Sciences, Department of Obstetrics and Gynaecology, The University of Auckland. Building 507, level 1, 22–30 Park Avenue, Grafton, Auckland 1023, New Zealand. Ph: 0064 21 023 78166. E: e.degraaff@auckland.ac.nz.

REFERENCES

- Hoppe DJ, Schemitsch EH, Morshed S, et al. Hierarchy of Evidence: Where Observational Studies Fit in and Why We Need Them. *J Bone Joint Surg Am*. 2009;91(Supplement_3).
- Funai EF, Rosenbush EJ, Lee MJ, Del Priore G. Distribution of Study Designs in Four Major US Journals of Obstetrics and Gynecology. *Gynecol Obstet Invest*. 2001;51(1):8-11.
- Grimes DA, Schulz KF. Bias and causal associations in observational research. *Lancet*. 2002;359(9302):248-52.
- Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Int J Surg*. 2014;12(12):1500-24.
- Milne B, Atkinson J, Blakely T, et al. Data Resource Profile: The New Zealand Integrated Data Infrastructure (IDI). *Int J Epidemiol*. 2019;48.
- Atkinson J, Blakely T. New Zealand's Integrated Data Infrastructure (IDI): Value to date and future opportunities. *Int J Popul Data Sci*. 2017;1.
- National Health Board Business Unit. National Maternity Collection Data Mart Data Dictionary. 2011.
- Statistics New Zealand. IDI DIA Life Event data. 2021.
- Births, Deaths, Marriages, and Relationships Registration (Prescribed Information) Regulations 1995 (SR 1995/183). 2021.
- National Health Board. National Minimum Dataset (Hospital Events) Data Dictionary. 2014.
- National Health Board. Mortality Collection Data Dictionary v1.8. 2021.
- Statistics New Zealand. IDI MOH Chronic Condition/ Significant Health Event Cohort data. 2021.
- Statistics New Zealand. the Statistics Act 1975, version as at 28 October 2021.
- Statistics New Zealand. IDI Data Dictionary: Immigration data (July 2015 edition). 2021.
- Statistics New Zealand. Statistical standard for meshblock. Available from www.stats.govt.nz; 2016.
- Perinatal and Maternal Mortality Review Committee. Thirteenth Annual Report of the Perinatal and Maternal Mortality Review Committee | Te Pūrongo ā-Tau Tekau mā Whā o te Komiti Arotake Mate Pēpi, Mate Whaea Hoki: Reporting mortality and morbidity 2017 | Te tuku pūrongo mō te mate me te whakamate 2017. Wellington: Health Quality & Safety Commission; 2019.
- Perinatal and Maternal Mortality Review Committee. Methodology and definitions for Perinatal and Maternal Mortality Review Committee (PMMRC) reporting. 2018.
- Ministry of Health. Primary Maternity Services Notice Pursuant to Section 88 of the New Zealand Public Health And Disability Act 2000. 2007.
- Ministry of Health. HISO 10001:2017 Ethnicity Data Protocols. 2017.
- Reid G, Bycroft C, Gleisner F. Comparison of ethnicity information in administrative data and the census. Available from www.stats.govt.nz; 2016.
- Stronks K, Kulu-Glasgow I, Agyemang C. The utility of 'country of birth' for the classification of ethnic groups in health research: the Dutch experience. *Ethn Health*. 2009;14(3):255-69.
- Premkumar A, Debbink MP, Silver RM, et al. Association of Acculturation With Adverse

- Pregnancy Outcomes. *Obstet Gynecol.* 2020;135(2):301-9.
23. Bakken KS, Skjeldal OH, Stray-Pedersen B. Obstetric Outcomes of First- and Second-Generation Pakistani Immigrants: A Comparison Study at a Low-Risk Maternity Ward in Norway. *J Immigr Minor Health.* 2017;19(1):33-40.
 24. Horner J, Ameratunga SN. Monitoring immigrant health and wellbeing in New Zealand: addressing the tyranny of misleading averages. *Aust Health Rev.* 2012;36(4):390-3.
 25. Atkinson J, Salmond C, Crampton P. NZDEP2013 Index of Deprivation. Department of Public Health, University of Otago, Wellington; 2014.
 26. Crampton P, Salmond C, Atkinson J. A comparison of the NZDep and New Zealand IMD indexes of socioeconomic deprivation. *Kōtuitui.* 2020;15(1):154-69.
 27. Exeter DJ, Zhao J, Crengle S, et al. The New Zealand Indices of Multiple Deprivation (IMD): A new suite of indicators for social and health research in Aotearoa, New Zealand. *PLOS ONE.* 2017;12(8):e0181260.
 28. Anderson N, Sadler L, Stewart A, McCowan L. Maternal and pathological pregnancy characteristics in customised birthweight centiles and identification of at-risk small-for-gestational-age infants: a retrospective cohort study. *BJOG.* 2012;119(7):848-56.
 29. Perinatal and Maternal Mortality Review Committee. Fourteenth Annual Report of the Perinatal and Maternal Mortality Review Committee | Te Pūrongo ā-Tau Tekau mā Whā o te Komiti Arotake Mate Pēpi, Mate Whaea Hoki: Reporting mortality and morbidity 2018 | Te tuku pūrongo mō te mate me te whakamate 2018. Wellington: Health Quality & Safety Commission; 2021.
 30. Flenady V, Koopmans L, Middleton P, et al. Major risk factors for stillbirth in high-income countries: a systematic review and meta-analysis. *Lancet.* 2011;377(9774):1331-40.
 31. Bartholomew K, Morton SMB, Atatoa Carr PE, et al. Early engagement with a Lead Maternity Carer: Results from Growing Up in New Zealand. *Aust N Z J Obstet Gynaecol.* 2015;55(3):227-32.
 32. Lawrence RL, Wall CR, Bloomfield FH. Prevalence of gestational diabetes according to commonly used data sources: an observational study. *BMC Pregnancy Childbirth.* 2019;19(1):34

Appendices

Appendix 1: Background to Integrated Data Infrastructure datasets.

The following tables were used for this methodological approach:

- The Maternity Collection (MAT):** This dataset, compiled from various sources, is held by the Ministry of Health (MoH) and includes all births in New Zealand from 20 weeks' gestation. MAT consists of two separate tables, one including data on all birth events ("MAT delivery") and the other including data on the individual infants ("MAT infant"). Despite inconsistencies in metadata, we understand MAT integrates information from three sources: 1) data collection by maternity care providers, including Lead Maternity Carers (LMCs; midwives, general practitioners (GPs) or obstetric specialists), reported to the MoH for funding purposes, and hospital providers of primary maternity care; 2) public and private hospital discharge coded data with health event information, sourced from the National Minimum Dataset (NMDS); and 3) demographic information from the National Health Index (NHI) dataset. Each live born infant is assigned a unique NHI number and is included in the NMDS and both MAT datasets. Since 1 July 2008, registration of maternal NHI has been mandatory for all births in NMDS, enabling mother and infant records to be linked. Stillborn infants, however, are not officially admitted to hospital and thus no individual discharge event is reported to the NMDS. Consequently, stillbirths are excluded from the MAT infants table, but included in the MAT delivery table through maternal discharge record data. It should be noted that stillbirths were not always included in the MAT dataset, although consistent assignment of NHI numbers for stillbirths is evident by 2009. MAT provides demographic and clinical information on women and infants for the duration of pregnancy until three months post-partum. It is used for annual national reporting of clinical indicators, evaluation of provided maternity services and can be requested for research purposes. Within the MAT delivery table in the Integrated Data Infrastructure (IDI),
- each birth is registered as a single entry and provided with a specific pregnancy-key (i.e., with a shared key for twins or higher order multiple pregnancies). Each individual mother also has a unique ID, allowing for identification of sequential pregnancies to the same mother.
- Births, Deaths and Marriages (BDM):** This dataset is held by the Department of Internal Affairs (DIA). BDM consists of three separate tables; "births", "deaths" and "marriages" of which the first two were used in this report. Infant and parent data on all deliveries in New Zealand are registered in the births table. Parents and LMCs are mutually responsible for providing information on the birth event to BDM. The LMC (or the public hospital in case of a hospital birth) is expected to provide an online birth notice within five working days, including information on infant birthweight and gestational age. Parents supply data on infant sex, birth status (i.e., live or stillborn), plurality (i.e., a singleton or multiple birth), location of birth, parental demographics, occupation, ethnicity and relationship status through a birth registration form. This is generally done online within two months after birth. Birth registration is only completed when both parties have provided this notification, with a few exceptions (i.e., where one parent is unknown, deceased, or missing; unable to complete the form because of a mental or medical condition; overseas with no delivery address or contact details; or a danger to the mother or child). In the case of a stillbirth ≥ 20 weeks' gestation or ≥ 400 grams in birthweight, the birth is registered with an indication that the infant was stillborn, in accordance with the Births, Deaths, Marriages, and Relationships Registration Act 1995. The deaths table registers all deaths in New Zealand, excluding stillbirths. Thus, only neonatal deaths (i.e., death of a live born infant under the 28th day of life) are noted here. The person handling the disposal of the body is responsible for notifying BDM of the death, by a death registration form.
- National Minimum Dataset (NMDS):** This dataset is held by the MoH and includes public and private hospital discharge data (as two separate tables) for inpatients and day-patients admitted for three hours or

more, all birth events and all admission where a procedure was performed. Information is provided by hospitals in an agreed (electronic) file format and is used for both statistical and funding purposes. Clinical coded data follows the Australian modification of the International Classification of Diseases (ICD-AM), which is updated over time. Mapping between subsequent ICD-AM editions can be done within the IDI. All hospital births are regarded as admission events and will therefore have been recorded. All complications of that pregnancy should be included in this coding.

- **The Mortality Collection (MORT):** This dataset is held by the MoH. It is a collection of data, recording cause of death by ICD-AM on all mortalities in New Zealand and used for annual reporting of vital statistics. MORT integrates information from (among other sources) BDM, NMDS, certificates by funeral directors and coroners' reports. A subset of MORT includes all fetal and infant deaths sourced from BDM and medical certificates or post-mortem reports.
- **The Chronic Conditions dataset (CC):** This dataset is held by the MoH, as a subset of data from several sources, for all residents in contact with the healthcare system during at least two quarters of that year. Individuals with one or more out of eight chronic conditions or significant health events are recorded here: acute myocardial infarction, cancer, coronary heart disease, chronic obstructive pulmonary disease, diabetes, gout, stroke and traumatic brain injury. The table also tracks the first and last incidence dates for the condition registered. In maternity research, this dataset can be used to identify mothers with pre-existing diabetes. Note that gestational diabetes is not recorded in this table. This dataset was supplied as a one-off by the MoH. A limitation to this dataset is that currently no updates are planned by Statistics NZ, with the latest incidence dates reported 31 December 2017. Studies including births after 2018 should be cautious of using this data source.
- **Census:** The Census of Population and Dwellings is a survey on all people living in New Zealand at a defined point in time. The last Census was completed in 2018

and is usually carried out once every five years, as set out in The Statistics Act 1975. Each individual is required to answer questions on personal affairs such as country of birth and religion, and socio-demographic variables such as housing and income. The Census is used for population counts, research purposes and planning by government and city councils. The IDI currently holds data from the 2013 and 2018 Census.

- **Immigration data:** This dataset is held by the Department of Labour, within the Ministry of Business, Innovation and Employment. It provides information on an individual's border movements, country of birth and visa type, sourced from immigration visa application forms and electronic passenger movement records. With increasing global migration this data may be useful to identify mothers born outside New Zealand.

Besides administrative datasets and surveys, the IDI provides some "general" or "central" tables. The following general tables were used for this methodological approach:

- **Full birth date, and full death date tables:** This data is available for all individuals in the IDI, sourced from the MoH. Generally, only a person's birth/death "month" and birth/death "year" are provided to researchers for privacy reasons. For a full date of birth/death (in the format dd/mm/yyyy), this table needs to be requested from Statistics NZ.
- **Address notification table:** This table holds data on address notifications of all individuals in the IDI. In addition, it identifies when a change in postal address occurred. Naturally, this depends on an individual updating address information. Geographical unit codes such as meshblocks (the smallest reported geographical units, representing on average 30 to 60 dwellings, with an average population of 100 people per unit) can be identified and used for other purposes, for example, area-based measures of socio-economic status.

The "gold standard"

In New Zealand, details of the deaths of still-born fetuses and infants (all perinatal related

mortalities from 20 weeks gestation up to the 28th day after birth) are collected by the Perinatal and Maternal Mortality Review Committee (PMMRC) using a standardised reporting tool. The PMMRC publishes annual reports on mortality. In addition, the Otago Data Group, contracted to the PMMRC, creates a total perinatal dataset by

Appendix 2: Creating the core dataset.

The following steps describe the method used, to create a “core” dataset for maternity research in the Integrated Data Infrastructure (IDI). A numerator table (including all perinatal deaths) and a denominator table (including all births) were created separately. Note that the final tables are not available for other users in the IDI. Users must run the code with the appropriate IDI refresh and adopt the code to each specific research project.

Step one: stillbirths

Statistics New Zealand (NZ) metadata recommends that stillbirths should not be identified from the Maternity Collection (MAT) or Births, Deaths and Marriages (BDM) datasets, as the available variables have not been validated. This was confirmed by data custodians at the Ministry of Health (MoH) and the Department of Internal Affairs. Instead, the Mortality collection (MORT) is considered the best source for stillbirths. Besides birth outcome, however, MORT does not provide any clinical information on the delivery event, maternal demographics, or other pregnancy risk factors.

To generate a table with all stillbirths, a full outer join was performed between the MAT delivery and BDM births datasets on maternal ID (“parent1”). Two additional joins on year and month of delivery ensure the link to the correct birth event. A filter will need to be applied to the variables “delivery outcome” (MAT) and “stillbirth indication” (BDM). A second filter for delivery year(s) will specify the targeted birth cohort. By performing a full outer join we ensure that all possible perinatal deaths are included, and no cases are missed. A disadvantage of this method is the risk of including some infants twice, where a wrong entry of birth month in either dataset will result in a non-match. Cases where linking seems appropriate (for instance, where only a one-month difference exists between delivery

merging PMMRC Data with MAT. This data collection is only available for specific perinatal-related research after formal application to the PMMRC and is not offered within the IDI. Due to the superior quality of these data, PMMRC annual report data were used as the gold standard to validate the tables created in this report.

dates, plus overlapping maternal demographics), can be merged in a separate step. In contrast, performing an inner join between the two datasets will provide an underrepresentation of the number of stillbirths in a cohort. This can be explained by the following rhetoric: if a birth registration has not been completed (e.g., when parents do not register a stillbirth), then DIA do not provide a record of this birth to Statistics NZ. However, for this same birth, the MoH may have provided Statistics NZ with the birth notice information (MAT). Finally, a small number of the identified stillbirths will be able to be joined to the MAT infant table (through their pregnancy-key), or BDM deaths datasets (by infant ID). These most likely include cases where it was unclear if the infant died before or soon after birth (as a day 0 neonatal death [NND], i.e., death of a live born infant up to the 28th day of life).

Step two: neonatal deaths

As no datasets in the IDI classify NNDs (and therefore do not distinguish them from infant mortalities ≥ 28 days of life), it is not possible to perform outer joins. A temporary table created from the full birth and death date tables will firstly identify all perinatal mortalities. To make the distinction between stillbirths and day 0 NNDs, cases with similar dates of birth and death can be matched to the MAT infant table by an inner join on infant ID. As stillbirths are not registered in the MAT infant dataset, this step should only return NNDs. A second link can then be made with the MAT delivery table, by performing a left join on pregnancy-key. Here, an inner join would result in a slight underestimate of NNDs as the tables don’t match perfectly, while a full outer join would not distinguish NNDs from other live births. A subsequent inner join by infant ID will match to the correct case in the BDM births dataset. As a final step, overlapping cases between the temporary stillbirth and NNDs tables should be investigated and adjusted where necessary.

Step three: perinatal deaths

To create a full numerator table including all perinatal deaths, the temporary stillbirth and NND tables should be combined. A conservative linking method applied by Statistics NZ estimates a <2% false positive linking rate of individuals in the IDI (i.e., where records of two people are incorrectly matched), thus allowing for a relatively high false negative linking rate (i.e., where two records of one person have not been matched). In addition, the use of full outer and left joins rather than inner joins as described in this report, will result in a further increase in duplicates of cases. The newly created perinatal deaths table is therefore expected to hold a larger number of cases compared to gold standard Perinatal and Maternal Mortality Review Committee (PMMRC) numerator data at this point. To correct for this difference as much as possible, we propose further probabilistic linking by using overlapping variables between the various datasets (e.g., delivery year and month, maternal birth year and month, gestational age, birthweight, area of domicile, District Health Board of birth, or facility code). For example, BDM cases without a MAT record can be joined to MAT cases without a BDM record, and vice versa. This approach should result in a dataset similar to our gold standard, while still guaranteeing conservative linking.

Where possible, a join to the BDM deaths dataset can be performed by infant ID. A new variable for delivery outcome will most accurately represent the true birth outcome, by applying the following prioritisation: 1) If a death can be joined to the MORT dataset, this delivery outcome should

be used as the best available source; else 2) if a death is registered in the BDM deaths dataset, it should be considered a NND; else 3) if a death is registered as a stillbirth in BDM births, it should be considered a stillbirth; else 4) if a death is registered in the MAT infant table, it should be considered a NND; and else 5) if MORT, BDM and MAT infant data are missing, the MAT delivery table birth outcome should be used.

Step four: live births

To create a denominator table including all births, all live births will need to be identified first. A full outer join between the MAT delivery and infant datasets on pregnancy-key, plus a full outer join on BDM births by infant or maternal ID, will return all matches and non-matches of any birth outcome. A filter by delivery year(s), “delivery outcome” (MAT) and “stillbirth indication” (BDM) will allow selection of all live births. Duplicates with the numerator table, should be deleted. Further probabilistic linking can be performed by matching between overlapping variables from the various datasets as previously described. Finally, a full denominator table is created by unifying the numerator set with all live births.

The PMMRC and BDM use marginally different definitions for stillbirth. For consistency in this report, we followed PMMRC definitions. Infants born <20 weeks’ gestation (or <400g of birthweight if gestational age was unknown) should be removed from the dataset, since these infants do not meet the PMMRC criteria for perinatal mortality. Further data cleaning can be based on rules outlined in the PMMRC methodology report.

Appendices 3 & 4:

https://uploads-ssl.webflow.com/5e332a62c703f6340a2faf44/63c9c0dd41fd76cd48e8b15b_5912%20-%20Appendices%203%20%26%204_final.pdf

To see copies of the original SQL files, please contact corresponding author Esti C de Graaff at e.degraaff@auckland.ac.nz