# A bespoke data linkage of an IVF clinical quality registry to population health datasets; methods and performance

Georgina M Chambers[1,2,*], Stephanie K.Y. Choi[1,2], Katie Irvine[3], Christos Venetis[1,2], Katie Harris[4], Alys Havard[1,5], Robert J Norman[6], Kei Lui[2], William Ledger[2], and Louisa R Jorm[1]

[1]Centre for Big Data Research in Health, Faculty of Medicine, University of New South Wales, Sydney, Australia
[2]School of Women's and Children's Health, Faculty of Medicine, University of New South Wales, Sydney, Australia
[3]Centre for Health Record Linkage, Ministry of Health, New South Wales, Australia
[4]The George Institute for Global Health, Faculty of Medicine, University of New South Wales, Sydney, Australia
[5]National Drug and Alcohol Research Centre, Faculty of Medicine, University of New South Wales, Sydney, Australia
[6]The Robinson Research Institute, The University of Adelaide, Adelaide, South Australia, Australia

## Abstract

### Introduction

Assisted reproductive technologies (ART), such as in-vitro fertilisation (IVF), have revolutionised the treatment of infertility, with an estimated 8 million babies born worldwide. However, the long-term health outcomes for women and their offspring remain an area of concern. Linking IVF treatment data to long-term health data is the most efficient method for assessing such outcomes.

### Objectives

To describe the creation and performance of a bespoke population-based data linkage of an ART clinical quality registry to state-based and national administrative datasets.

### Methods

The linked dataset was created by deterministically and probabilistically linking the Australia and New Zealand Assisted Reproduction Database (ANZARD) to New South Wales (NSW) and Australian Capital Territory (ACT) administrative datasets (performed by NSW Centre for Health Record Linkage (CHeReL)) and to national claims datasets (performed by Australian Institute of Health and Welfare (AIHW)). The CHeReL's Master Linkage Key (MLK) was used as a bridge between ANZARD's partially identifiable patient data (statistical linkage key) and NSW and ACT administrative datasets. CHeReL then provided personal identifiers to the AIHW to obtain national content data. The results of the linkage were reported, and concordance between births recorded in ANZARD and perinatal data collections (PDCs) was evaluated.

### Results

Of the 62,833 women who had ART treatment in NSW or ACT, 60,419 could be linked to the CHeReL MLK (linkage rate: 96.2%). A reconciliation of ANZARD-recorded births among NSW residents found that 94.2% (95% CI: 93.9–94.4%) of births were also recorded in state/territory-based PDCs. A high concordance was found in plurality status and birth outcome ($\geq$99% agreement rate, Cohen's kappa ranged: 0.78–0.98) between ANZARD and PDCs.

### Conclusion

The data linkage resource demonstrates that high linkage rates can be achieved with partially identifiable data and that a population spine, such as the CHeReL's MLK, can be successfully used as a bridge between clinical registries and administrative datasets.

### Keywords

assisted reproductive techniques; infertility; data linkage; pregnancy outcome; australia

*Corresponding Author:
*Email Address:* g.chambers@unsw.edu.au (Georgina M. Chambers)

# Introduction

Infertility affects one in six couples [1], resulting in significant personal suffering, and representing an important and increasingly prevalent public health problem [2, 3]. Fortunately, there are a number of Medically Assisted Reproduction (MAR) treatments that allow many infertile individuals to achieve parenthood. The most advanced of these are assisted reproductive technologies (ART), such as in vitro fertilisation (IVF), which involve the fertilisation of human eggs outside of the body before transferring the resulting embryos into the uterus in the hope of achieving a pregnancy. ART represents one of the most significant medical and social achievements of the past century, leading to the birth of an estimated 8 million babies over the last four decades [4]. Non-ART treatment using ovulation induction (OI) with or without intrauterine insemination (IUI) is a more traditional form of MAR treatment in which fertilisation occurs within the woman's reproductive tract, but is still widely used as part of evidence-based management [5]. The increasing demand for MAR treatment, both ART and non-ART, reflects the social trend to delayed childbearing, changes in family structures, rising levels of sexually transmitted disease, obesity, and declining sperm quality [6–10].

Australia has one of the highest rates of ART utilisation per capita in the world [11]. Over the last two decades, Australia has experienced a 192% increase in ART utilisation, and in 2018, around 4.9% of Australian children were conceived using ARTs [2, 11–13]. However, little is known about the number of children conceived through OI/IUI (non-ART) treatment or by spontaneous conception for women with a history of subfertility.

While evidence regarding the health outcomes of ART-conceived children is generally reassuring, several studies have suggested a higher risk of poorer perinatal outcomes, and longer term metabolic risks [14–18]. This is primarily because they are at a greater risk of being born as part of multiple gestation pregnancies (e.g. twins and triplets), but even singletons are at a marginally higher risk of low birth weight, small for gestational age, congenital anomalies, perinatal death (stillbirth and neonatal death), as well as maternal morbidity compared to spontaneously conceived children [14–16, 18]. The reasons for these increased risks to mothers and babies are not well understood. Interestingly, it appears that couples who experience subfertility but achieve a spontaneous conception also have similar adverse risk profiles [14, 19, 20]. Furthermore, there is a lack of evidence on the health outcomes of children conceived using OI/IUI (non-ART) treatment [21].

The lack of clarity on the potential risks of MAR treatments (ART and non-ART) and the possible confounding role of subfertility is an enduring evidence gap when advising patients, clinicians and policymakers on the use of MAR treatments.

To address this gap, the National Perinatal Epidemiology and Statistics Unit (NPESU) of the University of New South Wales created a MAR data linkage by linking a regional ART treatment registry (Australian and New Zealand Assisted Reproduction Database, ANZARD) to a number of other jurisdiction-based and national administrative databases. The resulting dataset contains longitudinal health records for women who have either undergone MAR (ART and non-ART), or who have conceived spontaneously, and their resulting children. The overarching objective of establishing the MAR data linkage resource was to quantify the risk of adverse health outcomes in children conceived from ART and non-ART treatments after accounting for confounders, in particular underlying subfertility, and to assess if specific forms of ART contribute differently to these outcomes.

Central to the MAR data linkage resource is the ANZARD, which is the oldest national ART registry in the world incorporating all accredited fertility clinics operating in Australia and New Zealand (currently over 90 clinics) and providing demographic, treatment, laboratory and outcome data on all ART cycles and donor insemination (DI) cycles (currently over 80,000 cycles per year) [2]. The submission of data to ANZARD is a requirement of a clinic's accreditation to practice and thus complete ascertainment of ART cycles is assumed [22]. ANZARD does not currently collect data from non-ART treatments such as OI and IUI.

This paper describes the data linkage methodology and results between the ANZARD and the state/territory and national data sources, and describes the concordance between the births recorded in ANZARD and those in the state perinatal data collections (PDCs).

# Methods

## The MAR data linkage

New South Wales (NSW) and Australian Capital Territory (ACT) are two of the eight Australian states and territories, with their combined population of approximately 8 million residents accounting for one-third of the total Australian population [23]. ANZARD was linked to NSW and ACT perinatal, births, deaths, hospital admissions, and congenital anomaly routinely collected databases, as well as national medical and pharmaceutical claims databases.

The linkage was possible because since 2009 ANZARD has collected the first two letters of female patients' first and last names. These personal identifiers were combined with the female patients' date of birth (DOB), residential postcode, and their partners' DOB to form a Statistical Linkage Key (SLK). Combinations of the components of the SLK were the foundation for linkage with the administrative datasets [2].

The NSW Ministry of Health's Centre for Health Record Linkage (CHeReL) and the Australian Institute of Health and Welfare (AIHW) undertook the required linkages before transferring the data into a secure research environment for cleaning and analysis by the researchers. Figure 1 summarise the data linkage process of ANZARD to 5 NSW and ACT administrative and 2 Commonwealth datasets.

A key strategy to enable the linkage of ANZARD's partially identifiable data (components of the SLK) to the administrative datasets was the ability of the CHeReL to use their Master Linkage Key (MLK) as a bridge between ANZARD and the NSW and ACT Perinatal Data Collections (PDC) to identify births to women who had conceived using ART and those who had conceived spontaneously (without ART). The MLK is constructed by the CHeReL using probabilistic record linkage methods and ChoiceMaker software using a best practice approach to privacy-preserving

Figure 1: An overview of the data linkage process for the medically assisted reproduction (MAR) data linkage
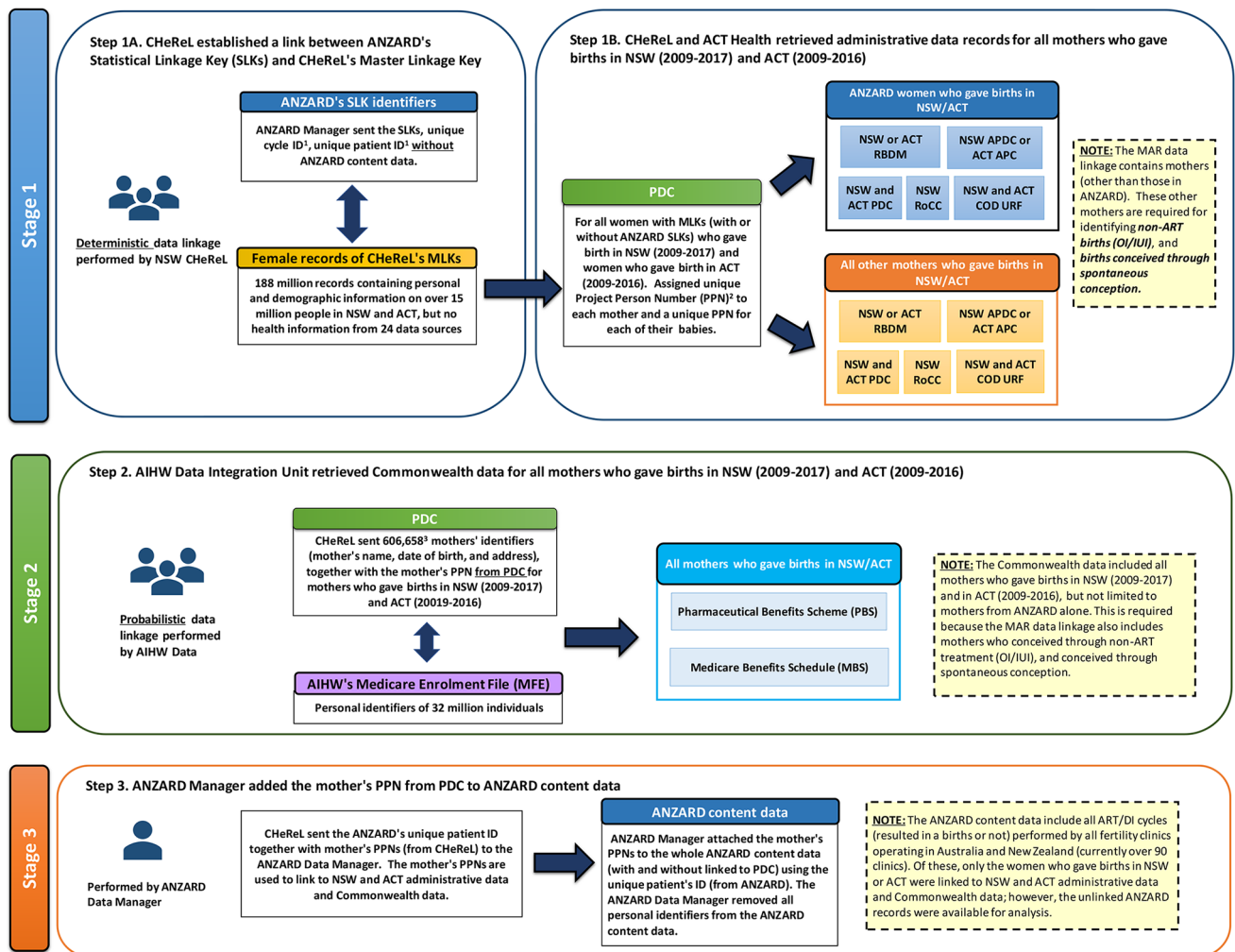


PDC = Perinatal Data Collection; NSW APDC = NSW Admitted Patient Data Collection; RBDM = Registry of Births, Deaths and Marriages; RoCC = Register of Congenital Conditions; ACT APC = ACT Admitted Patient Care; PBS = Pharmaceutical Benefits Scheme; MBS = Medicare Benefits Schedule; ANZARD = Australia and New Zealand Assisted Reproduction Database; COD URF = Cause of Death Unit Record File Registry

[1] ANZARD's Patient IDs were used to link back to the ANZARD content data in stage 3 by ANZARD manager. Please note that ANZARD includes all ART and DI cycles performed by all fertility clinics operating in Australia and New Zealand (currently over 90 clinics)

[2] Project Person Number (PPN) is a unique person ID for each individual in the linked data. It varies from project to project to prevent linking individual-level records across different projects. This is required to ensure privacy and confidentiality in Australia.

[3] These 606,658 mother's identifiers included duplicates of mothers who gave birth in both NSW and ACT. We removed the duplicates from the NSW and ACT PDC data, resulting in 606,549 mothers in Figure 2.

record linkage. The MLK comprises over 188 million records containing personal and demographic information, but no health information, on over 15 million people in NSW and ACT from a range of population-based health and health-related data collections [24]. The CHeReL uses the following personal information to link records for the same person to create the MLK: full name, address, sex, DOB, country of birth, and uses relevant event information such as hospital code, medical record number, event dates (e.g., hospital dates of admission and discharge), hospital transferred to, hospital transferred from, and date of death. The entire linked NSW and ACT administrative data has less than 5/1000 missed links and 3/1000 false positive links [25]. In addition to person links, the MLK contains a family structure, by virtue

of data sources that contain details of a child and up to two parents.

The data linkage between ANZARD and NSW/ACT administrative data and the Commonwealth data involved three stages that enabled the construction of the MAR data linkage while abiding by the principles of data separation to protect patient privacy.

**Stage 1: Data linkage between ANZARD and NSW and ACT administrative data**

The ANZARD Data Manager (who is independent from the research team) transferred to CHeReL a cycle ID (an anonymous unique cycle identifier) and a patient ID (an

anonymous unique patient identifier) together with the SLK components associated with the 638,036 ART and DI cycles (with 195,490 SLKs) performed in Australia between 1st January 2009 and 31st December 2016.

The CHeReL then deterministically linked ANZARD personal SLK identifiers (SLK person ID) to the MLK identifiers (MLK person ID) for females. Several strategies were adopted to improve the data linkage rate between the ANZARD identifiers and the MLK. Because linkage using the SLK may have low sensitivity depending on data quality, the deterministic linkage on person characteristics was combined with event-based linkage using ANZARD and hospital procedures dates for selected procedure codes relating to oocyte retrieval. An ART cycle involves mostly outpatient services; however, almost all egg retrieval procedures are undertaken under sedation as part of an inpatient admission which is recorded in the hospital admission data collections. Thus, procedure codes related to egg retrieval procedures were used for the event-based linkage to supplement the linkage based on personal identifiers. Matches were initially performed on all personal identifiers and event information; then, restrictions were progressively relaxed to allow a higher rate of matching to the SLK. Where multiple MLK person IDs matched to a single ANZARD SLK person ID clerical review was performed. Details of the results of each pass in the linkage process between the ANZARD SLK and MLK person ID and the corresponding linkage rate are shown in the results section below.

The MAR data linkage is a birth cohort for all births in NSW/ACT, comprising births that were conceived through ART treatment, non-ART treatment, or spontaneous conception. All MLK person IDs that linked to PDC records indicated that a woman had given birth between 1st January, 2009 and 31st December, 2017 in NSW and women who gave birth between 1st January, 2009 and 31st December, 2016 in the ACT. These PDC records (including those with a link to an ANZARD SLKs and those that did not) were then linked to individual-level content data from the various NSW and ACT administrative databases by the CHeReL and ACT Health (see more details of the administrative databases included in Table S1, Supplementary Appendix).

Once the linkage rate and accuracy were considered to be maximised based on available identifiers, the CHeReL created a Project Person Number (PPN) for each woman. The PPNs were later used by the research team to merge all datasets. The CHeReL loaded the PPNs and the content data from the NSW and ACT administrative databases into the Sax Institute's Secure Unified Research Environment (SURE) [26]. The SURE is a central, secure, online remote-access computing environment for analysing sensitive human research data.

**Stage 2: Data linkage between the NSW and ACT MLKs for mothers who gave birth and the Commonwealth Medicare Benefits Schedule and Pharmaceutical Benefits Scheme data**

The CHeReL sent identifying information only (mother's name, her DOB and residential address) representing the 606,658 (include duplicates) women who gave birth between 1st January, 2009 and 31st December, 2017 in NSW and

1st January, 2009 and 31st December, 2016 in ACT and the corresponding PPNs to the AIHW Data Integration Unit.

The AIHW Data Integration Unit undertook a probabilistic linkage between the MLKs and the personal identifiers from the Medicare Enrolment File (MEF) of 32,378,696 individuals registered to Australia's national health care scheme, Medicare. The MEF linkage procedure involved creating record pairs between MLKs and MEF's personal identifiers based on a combination of seven personal identifiers: surname; given name; sex; day, month, and year of birth; day, month, and year of death when applicable; residential postcode; upper case of the first six characters of the address after removing the punctuations and words such as unit, flat, PO box, etc. A total of 18 passes were undertaken to create the final linked dataset.

Following the completion of the probabilistic linkage, a sample-based clerical review including 32 batches each containing between 78,437 and 1,250,122 records was performed to determine the linkage status for record pairs with similar linkage weights.

Once all linkages were maximised, the AIHW retrieved the requested individual-level content data from the Medicare Benefits Schedule (fertility-related services) and Pharmaceutical Benefits Scheme (fertility medicines and other medicines) and uploaded the PPNs and the content data from the Medicare Benefits Schedule and Pharmaceutical Benefits Scheme into SURE.

**Stage 3: Retrieved ANZARD treatment and outcome data**

The CHeReL sent the PPNs and both linked and unlinked ANZARD's unique patient IDs back to the study-independent ANZARD Data Manager. The ANZARD Data Manager removed all personal identifiers from the ANZARD content data and attached the PPNs and ANZARD's unique patient ID to the ANZARD content data. The ANZARD Data Manager loaded the ANZARD content data (with PPNs and ANZARD's unique patient ID) for all ANZARD treatment performed during the study period to SURE ready for the researchers to merge all data collections using the common PPN.

## Final MAR data linkage

Table S1 (Supplementary Appendix) describes all data sources included in the final MAR data linkage. Briefly, the MAR data linkage constitutes 581,241 mothers who had conceived 874,922 babies born between 2009 and 2017 plus 261,711 siblings of these babies born before 2009 in NSW, and 27,631 mothers who had conceived 36,964 babies born between 2009 and 2016 plus 8,088 siblings of these babies born before 2009 in the ACT. The resulting longitudinal health record provided up to 10.25 years of follow-up. Of these NSW/ACT mothers, 37,443 (6.2%) had at least one ANZARD ART treatment cycle record. ANZARD was only linked to state and Commonwealth administrative datasets where a woman was identified as giving birth in NSW or ACT. Unlinked ANZARD records for treatment information for women who had undergone ART and DI treatment and who had not given birth were also uploaded to SURE.

Table 1: Criteria to construct birth windows for examining concordance between ANZARD's birth records and PDC's birth records

| Criteria | embryo stage | Birth window |
|---|---|---|
| 1 | | 1. Babies' month and year of birth recorded in ANZARD matched exactly with that of the PDC |
| | | 2. For the remaining unmatched ANZARD births, we progressively relaxed by a grace period of 16 days |
| | | Lower limit: Baby's DOB from ANZARD data - grace period[1] |
| | | Upper limit: Baby's DOB from ANZARD data + grace period[1] |
| 2 | Cleavage (3 day old embryo) | Lower limit: Embryo transfer date - 3 days - 14 days + 7×gestational weeks - grace period[2] |
| | | Upper limit: Embryo transfer date - 3 days - 14 days + 7×gestational weeks + grace period[2] |
| | Blastocysts (5 day old embryo) | Lower limit: Embryo transfer date - 5 days - 14 days + 7×gestational weeks - grace period[2] |
| | | Upper limit: Embryo transfer date - 5 days - 14 days + 7×gestational weeks + grace period[2] |
| | Both cleavage and blastocysts | Lower limit: Embryo transfer date - 5 days - 14 days + 7×gestational weeks - grace period[2] |
| | | Upper limit: Embryo transfer date - 3 days - 14 days + 7×gestational weeks + grace period[2] |
| | Donor Insemination (DI) | Lower limit: DI date - 14 days + 7×gestational weeks - grace period[2] |
| | | Upper limit: DI date - 14 days + 7×gestational weeks + grace period[2] |

PDC = Perinatal Data Collection; ANZARD = Australia and New Zealand Assisted Reproduction Database; DI = Donor insemination; DOB = date of birth.
[1] Grace period was assumed as a 16-day period.
[2] Grace period was assumed as a 10-day period.

The MAR data linkage resource includes a wide range of data sources. For mothers, information is available on their use of fertility medicines and other medicines, their use of fertility-related services, hospital admissions, history of health conditions, socio-demographics, Aboriginal and/or Torres Strait Islander status, and pregnancy-related risk factors such as hypertension and gestational diabetes (Table 1, Supplementary Appendix). For their offspring, information is available on birth outcomes, birth defects, long-term and short-term adverse health conditions, hospital admissions, death, and cause of death (Table 1, Supplementary Appendix).

## Agreement between births in ANZARD and PDC

A primary research question to be addressed by the MAR data linkage is whether health outcomes are different between ART conceived versus non-ART conceived children (from other fertility treatment or spontaneously conceived). Therefore, an assessment was undertaken of the concordance between ANZARD recorded *births* and those recorded in the PDCs to identify births to women who had ART treatment, who may have conceived using non-ART treatment or spontaneously. This agreement analysis was only conducted for births resulting from ART or DI treatment by NSW residents and birthing in NSW or ACT because the ACT PDC data only covers births delivered in ACT public hospitals. Therefore, births to women residing in the ACT and who gave birth in a private hospital are missing from the ACT PDC, estimated to be about 20–25% of ACT births [12]. The PDCs encompass all live births and stillbirths of at least 20 weeks gestation or ≥400 grams birth weight. The record of birth and pregnancy information by ANZARD relies on the ART clinic staff following-up with women after their ART treatment, while the PDC relies on the attending midwife or medical practitioner completing a record of birth and pregnancy information.

We relied on baby's DOB, gestational age, embryo transfer or DI date, and the age of embryo at transfer to match each ANZARD birth to a PDC birth of the corresponding mother. Where the babies' DOB in ANZARD did not exactly match with a PDC-recorded month and year of birth, we progressively relaxed the matching requirements, allowing a grace period of 16 days (see criterion 1 in Table 1). For the remaining unmatched ANZARD births, we then used the embryo transfer or DI date (from ANZARD data) to extrapolate a birth window of an expected DOB to match to the PDC's DOB (see criterion 2 in Table 1). We progressively relaxed the birth window to account for uncertainty in the embryo transfer or DI date and ANZARD recorded DOB by a grace period of 10 days (Table 1). This was necessary because only the babies' month and year of birth were provided by the PDC.

The data concordance rate was calculated as a total number of births in an agreement between ANZARD and PDC data divided by a total number of ANZARD treatment records from NSW residents with an ANZARD birth recorded or without an ANZARD birth recorded due to loss of follow-up. Fact of birth plus key birth outcomes (live or stillbirth) and plurality (singleton or multiple births) were chosen to evaluate concordance because these are recorded in ANZARD and the PDCs.

To examine the impact of these grace periods (criterion 1: ANZARD recorded DOB and criterion 2: embryo transfer or DI date) on the concordance rate, we performed several sensitivity analyses by varying the grace period of criterion 1 from 16 days to 5 or 31 days; and by varying the grace period of criterion 2 from 10 days to 5 or 15 days.

### Concordance of plurality status and birth outcomes between ANZARD and the PDCs

For births in the agreement between the PDC and ANZARD birth records for NSW residents, we also examined agreement

Table 2: Linkage results for the data linkage between ANZARD statistical linkage key (SLKs) and New South Wales (NSW) and Australian capital territory (ACT) master linkage key (MLKs), number and percentage of total linked records

| Pass | Description | ANZARD SLK IDs | | Multi-links to MLK |
| --- | --- | --- | --- | --- |
| | | n | % | |
| 1 | First name code, surname code, DOB, partner DOB, postcode, ANZARD procedure date matches APDC subset dates | 22,983 | 35.8% | 11 |
| 2 | First name code, surname code, DOB, partner DOB, postcode, ANZARD procedure date not NULL, and MLK in procedure subset | 156 | 0.2% | 0 |
| 3 | First name code, surname code, DOB, partner DOB, postcode | 9,385 | 14.6% | 6 |
| 4 | First name code, surname code, DOB, postcode, ANZARD procedure date matches APDC subset dates | 17,339 | 27.0% | 19 |
| 5 | First name code, surname code, DOB, postcode, where either ANZARD or MLK partner DOB is NULL | 12,296 | 19.2% | 118 |
| 6 | First name code, surname code, DOB, edit distance between partner DOB and MLK partner DOB $\leq$ 2, postcode | 105 | 0.2% | 0 |
| 7 | First name code, surname code, DOB, partner DOB, edit distance between postcode and MLK postcode $\leq$ 1 | 241 | 0.4% | 2 |
| 8 | First name code, surname code, DOB, postcode in NSW or ACT, MLK in procedure subset | 688 | 1.1% | 2 |
| 9 | First name code, surname code, DOB, partner DOB | 1,013 | 1.6% | 2 |
| **Total Linked ANZARDSLKs to NSW and ACT MLKs** | | **64,206** | | |
| **Calculating linkage rate between ANZARD SLKs and MLKs** | | | | |
| Total number of ANZARD SLK (with a NSW/ACT postcode) (A) | | 62,833 | | |
| Total number of ANZARD SLKs (with a NSW/ACT postcode) linked to MLKs (B) | | 60,419 | | |
| *Linkage rate between ANZARD SLKs and MLKs (B/A)* | | *96.2%* | | |

DOB = date of birth.
ANZARD = Australian and New Zealand Assisted Reproduction Database.
APDC = New South Wales and Australian Capital Territory Admitted Patient Data Collection.

in the plurality (i.e., singleton, twins, and triplets) and birth outcome status (live birth or perinatal death) fields. We relied on the birth registry's plurality status, and death registry's perinatal death information if information in the PDC differed from that of the birth or death registries. We used both PDC and the Registry of Births, Deaths and Marriages (RBDM) (birth and death registries) as a gold standard reference when estimating sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). The RBDM is a statutory registry in NSW and ACT. Agreement rate, sensitivity, specificity, PPV, NPV, area under the receiver operating characteristic curve (AUC), and Cohen's Kappa statistics [27, 28] were reported.

# Results

## Results of the deterministic data linkage for women recorded by ANZARD SLKs and NSW/ACT MLKs

Between 1st January, 2009 and 31st December, 2016, a total of 195,490 ANZARD SLKs (each representing a woman receiving ART treatment in Australia and New Zealand) were deterministically linked to MLKs of females recorded by CHeReL. Table 2 presents the linkage results of the nine passes of this data linkage process. Of the 195,490 ANZARD SLKs, 64,206 were linked to a single MLK person ID. There were 160

ANZARD records that linked to multiple MLK records. For these, a clerical review was performed to identify duplicates and use other data items to identify likely links. The linkage completed by the CHeReL was also checked for false-positive links by selecting a random sample of 1000 Person IDs from the linked data for review. The results indicated a false positive rate of 5/1000 Person IDs. Table 2 presents the linkage rate between ANZARD SLKs and the MLK for female patients with a NSW or ACT residential postcode (aligning with the state administrative datasets including the PDCs), or an unknown postcode. Of the 195,490 ANZARD SLKs, there were 62,833 ANZARD SLKs with a NSW or ACT residential postcode. Of these, 60,419 (including NSW or ACT) were linked to an MLK person ID, resulting in a 96.2% linkage rate between women residing in NSW or ACT who underwent ART treatment and being able to be identified in the MLK (i.e., 60,419/62,833).

## Results of the probabilistic data linkage between the NSW/ACT MLKs for *Mothers who gave birth* and the Commonwealth Medicare Benefits Schedule and Pharmaceutical Benefits Scheme data

Overall, there were 606,658 PPNs from mothers who gave birth in NSW between 2009 and 2017, and from mothers who gave births in the ACT between 2009 and 2016, respectively. Of these, 597,549 (98.49%) were linked to the MEF's personal

identifiers representing a linkage accuracy of 99% between the PDCs and national Medicare enrolments.

## Overall concordance rate for *births* recorded in ANZARD and the PDC data

To assess the concordance rate between ANZARD and the PDC, the cohort was restricted to ANZARD's cycles performed between 2009 and 2015 (rather than 2016) because ACT PDC data was only available up to 31st December, 2016, and one year of follow-up after an ART or DI cycle was needed to capture births in the PDCs. Between 1st January, 2009 and 31st December, 2015, there were 400,853 cycles of ART or DI treatment records recorded in ANZARD (Figure 2). Of these, 28,937 cycles were from women residing in NSW with a birth outcome or unknown outcome due to loss follow-up. Of these, 27,248 were matched to an NSW or ACT PDC birth record via baby's DOB, gestational age, embryo transfer or DI date, and the age of embryo at transfer, resulting in a concordance rate of 94.2% (27,248/28,937) (95% CI: 93.9–94.4%). Of these 27,248 births from women with ART or DI cycles in NSW, 158 (0.58%) had a cross-state delivery in an ACT public hospital. The concordance rate remained stable by year (Figure S1, Supplementary Appendix).

Our sensitivity analysis, in which we used different combinations of grace periods, showed consistent results of the concordance rate between ART/DI treatment cycles and NSW and ACT PDC records at 94.2% over our study period (Table S2, Supplementary Appendix).

## Concordance of plurality status and birth outcomes between ANZARD and PDC data for the final MAR data linkage

Of the 27,248 ANZARD births (from NSW and ACT residents) linked to a NSW/ACT PDC birth, there was 99.7% (27,170/27,248) (95% CI: 99.6–99.8%) agreement with Cohen's kappa of 0.977 (95% CI: 0.971–0.982) in plurality recording between ANZARD and PDC data (Table S3, Supplementary Appendix). There was also a high degree of agreement (≥99%, Cohen's kappa ranged: 0.78–0.90) for live birth and perinatal death status between ANZARD and PDC records for the 25,758 singleton births, with a range of 87.0–99.9% for PPV and a range of 87.0–99.9% for NPV; for 1,412 plural births, with a range of 87.0–99.3% for PPV and a range of 88.9–99.2% for NPV (Table 3).

## Discussion

This paper describes the creation of a bespoke linked dataset (MAR data linkage) of a clinical quality registry (ANZARD) with state/territory and national administrative datasets. Despite limited personal identifiers being present in the registry, of the 62,833 women who had ART treatment in NSW or ACT, 60,419 could be linked to the CHeReL MLK population spine, representing a linkage rate: 96.2%. This means that only 3.8% of women who had ART treatment in NSW/ACT could not be linked to the NSW/ACT MLK created by CHeReL. This linkage rate is similar to other studies that have used limited identifiers as part of an SLK used to provide partial identifiers are part of clinical registries [29–31].

A reconciliation of the ART/DI cycles performed to women who resided in NSW and who were recorded as having a birth in ANZARD, found that 94.2% of the births were recorded in NSW and ACT PDCs. Possible reasons for the 5.8% of missing ANZARD births in the PDC, would include women who resided in NSW for ART/DI treatment but birthed in a private ACT hospital, in another Australian states/territories, or overseas, missing links between the SLK and MLK (3.8% linkage error), or births being erroneously recorded in ANZARD. An evaluation of the small percentage of NSW or ACT women recorded in ANZARD who could not be linked to the MLK population spine was not conducted as part of this study and could be a small source of linkage bias. However, because of Australia's universal health system it is unlikely that there would be a systematic bias in the linkage between ANZARD and the MLK based on demographics because the MLK contains 210 million records from 17 data collections with 15 average links per person [32].

A high concordance was found in plurality status (>99% agreement rate; Cohen's kappa: 0.977 (95% CI: 0.971–0.982)) and birth outcome (≥99% agreement rate; Cohen's kappa ranged: 0.78–0.90) between ANZARD and PDC birth records confirming the validity of the linkage. The high degree of concordance between births recorded in ANZARD and those recorded in jurisdictional perinatal data collections provides reassurance that fertility clinics in Australia are accurately recording the outcomes of ART treatment undertaken in their clinics, and that clinics are not artificially inflating their success rates in NSW/ACT. The accreditation of each fertility clinic in Australia is managed by the industry's Fertility Society of Australia and New Zealand under its Reproductive Technology Accreditation Committee's voluntary Code of Practice, under which clinics must submit their treatment and outcomes data to ANZARD [33]. The results of this concordance study and the high linkage rate reflects positively on the Fertility Society of Australia and New Zealand model of industry regulation being connected with clinical registry management.

Our linkage rate (96.2% at woman-level, 94.2% at birth-level) was higher than that achieved by the States Monitoring Assisted Reproductive Technology collaboration's linkage of the U.S. state-based and national ART data linkage to vital birth registrations (80–90.2%) [34–36] and the linkage to pregnancy data (89.7%) in the Massachusetts Outcomes Study of Assisted Reproductive Technology [37]. The Committee of Nordic Assisted Reproductive Technology and Safety was able to achieve a very high linkage between IVF registries and birth registries because of the existence of national personal identifiers [38].

Most of the earlier U.S. linkages conducted by the States Monitoring Assisted Reproductive Technology collaboration were cycle-based and adopted a deterministic or probabilistic linkage strategy to link treatment cycles to vital records based on less specific maternal or infant variables (e.g., DOB of mothers and infants, mother's postcode, or plurality, etc.) due to a lack of mothers' or infants' identifiers [34–36]. The results of the latest U.S. linkage by the Massachusetts Outcomes Study of Assisted Reproductive Technology that included maternal or parental identifiers (i.e. mother's first and last name, and father's last name) in the linkage strategy is the most comparable to that of the MAR data linkage (89.7%

Figure 2: Flowchart for concordance analysis between ANZARD births and PDC births, for New South Wales (NSW)[1] residents, with ART/DI treatment cycles between 2009 and 2015
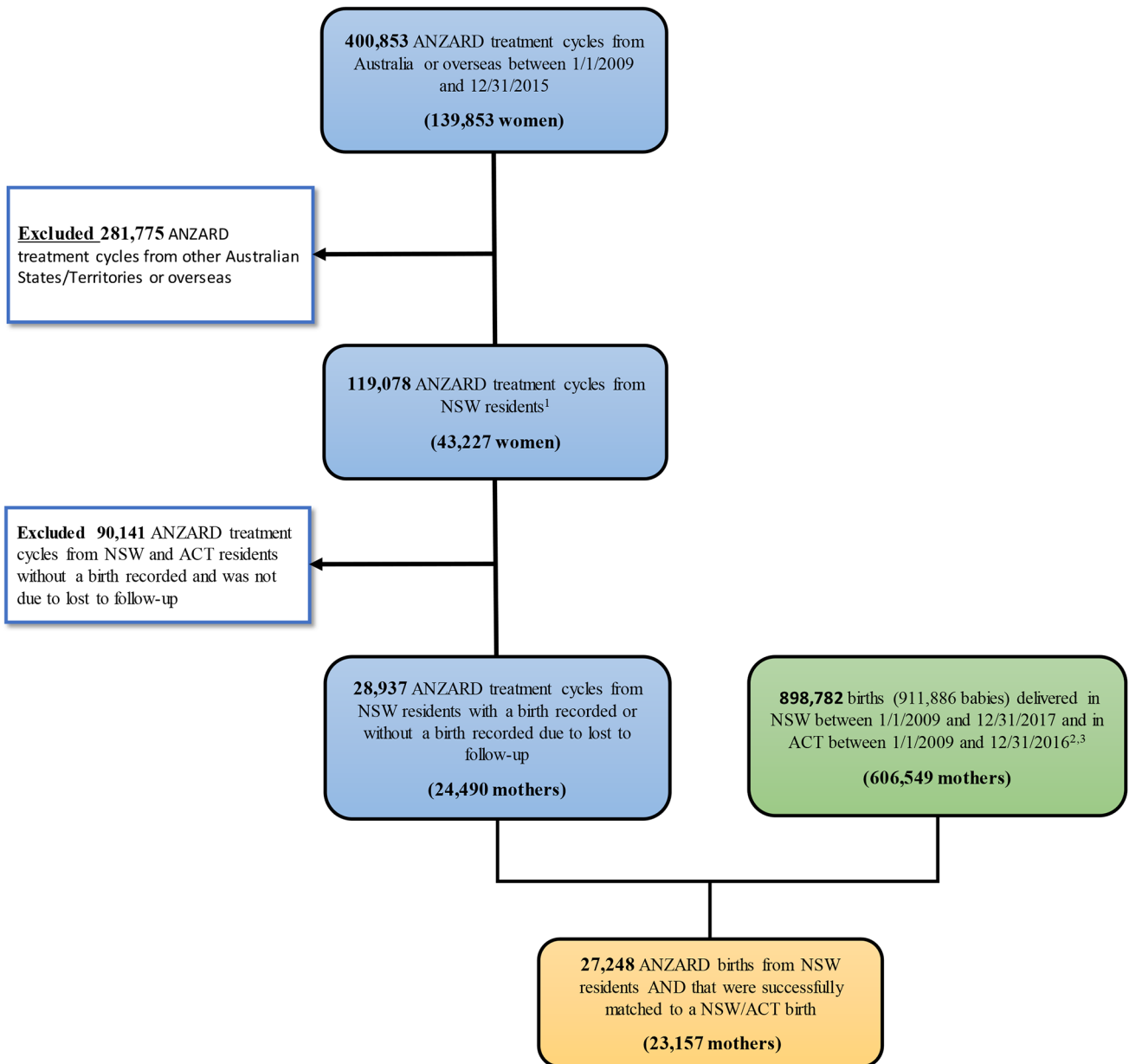


NSW = New South Wales; ACT = Australian Capital Territory; ANZARD = Australia and New Zealand Assisted Reproduction Database; ART = Assisted Reproductive Technologies; PDC = Perinatal data collections; DI = Donor insemination.

[1] We identified NSW residents based on residential postcode in ANZARD data.

[2] We included the ACT PDC data when matching the ANZARD births to a PDC births for accounting the cross-state delivery between NSW and ACT.

[3] There were 2,323 mothers gave births in both NSW and ACT. The total unique number of mothers gave births in NSW or ACT were 606,549, which is different from the 606,658 PPNs that CHeReL sent to AIHW (Figure 1). The number of births and babies included raw records (pre-cleaning) received from CHeReL and ACT health.

vs. 94.2%, at birth-level) [37]. The MAR linkage also shows a comparable rate of agreement of plurality status (>99%) and birth outcome (≥99%) but a higher sensitivity (71.2–99.9% vs. 27.1–47.4%) and PPV (87.0–99.9% vs. 41.3–69.2%) for live birth/fetal death outcomes between the ANZARD and the PDC data [37, 39, 40].

The MAR linkage contains up to 10.25 years of follow-up allowing the assessment of the short-term and long-term health risks for women and ART conceived children.

Additionally, the prognostic value of the type of ART treatment performed, (e.g., use of fresh or frozen embryo, sperm injection, extended embryo culture) will be able to be assessed. Furthermore, the health of children born from non-ART treatments will be evaluated using national medicines and medical services claims data to identify children conceived using ovulation induction and ovarian stimulation. Moreover, because of the longitudinal nature of the datasets, women with a history of subfertility, but who

Table 3: Agreement measures of each birth outcome between ANZARD and PDC data[1,2], for New South Wales (NSW) residents[3], by plurality status

| Birth outcomes | Agreement (95% CI) | Sensitivity (95% CI) | Specificity (95% CI) | PPV (95% CI) | NPV (95% CI) | AUC (95% CI) | Kappa statistics[4] (95% CI) |
|---|---|---|---|---|---|---|---|
| **Singleton births (N = 25,758)** | | | | | | | |
| Live birth | 99.8% | 99.9% | 89.3% | 99.9% | 87.0% | 0.95 | 0.88 |
| | (99.72, 99.83) | (99.8, 99.9) | (84.5, 93.0) | (99.8, 99.9) | (82.0, 91.1) | (0.93, 0.97) | (0.84, 0.91) |
| Perinatal death | 99.8% | 92.6% | 99.9% | 87.0% | 99.9% | 0.96 | 0.90 |
| | (99.72, 99.83) | (88.3, 95.7) | (99.8, 99.9) | (82.0, 91.1) | (99.8, 100) | (0.95, 0.98) | (0.87, 0.93) |
| **Plural births (N = 1,412)** | | | | | | | |
| Live birth | 99.0% | 99.8% | 69.6% | 99.2% | 88.9% | 0.85 | 0.78 |
| | (98.5, 99.3) | (99.5, 99.9) | (57.3, 80.1) | (98.8, 99.5) | (77.4, 95.8) | (0.79, 0.90) | (0.69, 0.86) |
| Perinatal death | 99.0% | 71.2% | 99.7% | 87.0% | 99.3% | 0.85 | 0.78 |
| | (98.6, 99.3) | (58.7, 81.7) | (99.5, 99.9) | (75.1, 94.6) | (98.9, 99.6) | (0.80, 0.91) | (0.70, 0.86) |

CI = confidence intervals.
ANZARD = Australian and New Zealand Assisted Reproduction Database; PDC = Perinatal Data Collection
PPV = positive predictive value; NPV = negative predictive value; AUC = area under the receiver operating characteristic curve.
[1]NSW residents who have undergone an ART or DI Treatment with a known birth outcome or an unknown birth outcome due to loss to follow-up and an agreement of birth recorded in PDC.
[2]For births with an agreement in plurality status between ANZARD and PDC data.
[3]This agreement analysis is only conducted for births resulting from ART treatment by NSW residents and birthing in NSW or ACT. The ACT PDC data only covers births delivered in ACT public hospitals; thus, births to women residing in the ACT and who gave birth in a private hospital are missing from the ACT PDC, estimated to be about 20–25% of ACT births (Australian Institute of Health and Welfare, 2018).
[4]The 95% confidence intervals were constructed by the bias-corrected bootstrap method with 2000 replicates (Efron, 1987).

subsequently conceived naturally can be identified, allowing the role of subfertility in health outcomes to be assessed – a confounder that is often elusive in studies of ART conceived children. Finally, sibship studies will also be possible because children born to the same mothers (including siblings from plural births or from singleton births since 1994) can be identified.

# Conclusions

The MAR data linkage demonstrates that very high linkage rates can be achieved with partially identifiable data, and that a population spine such as the CHeReL's MLK can be successfully used as a bridge between clinical registries and administrative datasets. The high concordance between births recorded in ANZARD and perinatal data collections provides reassurance about the accuracy of ART treatment outcomes recorded in ANZARD. The MAR data linkage will provide invaluable information on the safety and effectiveness of ART and non-ART treatment, and the possible effect of subfertility when advising patients, clinicians, and policymakers on fertility treatments for Australia and beyond.

# Acknowledgement

# Statement on conflicts of interest

# Ethics statement

# References

1. Inhorn MC, Patrizio P. Infertility around the globe: New thinking on gender, reproductive technologies and global movements in the 21st century. Hum Reprod Update. 2015;21:411–26. https://doi.org/10.1093/humupd/dmv016.

2. Newman JE, Paul RC, Chambers GM. Assisted reproductive technology in Australia and New Zealand 2018. 2020. https://npesu.unsw.edu.au/data-collection/australian-new-zealand-assisted-reproduction-database-anzard.

3. Sun H, Gong T-T, Jiang Y-T, Zhang S, Zhao Y-H, Wu Q-J. Global, regional, and national prevalence and disability-adjusted life-years for infertility in 195 countries and territories, 1990-2017: Results from a global burden of disease study, 2017. Aging (Albany NY). 2019;11:10952–91. https://doi.org/10.18632/aging.102497.

4. De GC, Calhaz-Jorge C, Kupka MS, Wyns C, Mocanu E, Motrenko T, et al. ART in Europe, 2014: Results generated from European registries by ESHRE: The European IVF-monitoring Consortium (EIM) for the European Society of Human Reproduction and Embryology (ESHRE). Hum Reprod 2018. https://doi.org/10.1093/humrep/dey242.

5. Kulkarni AD, Jamieson DJ, Jones HW, Kissin DM, Gallo MF, Macaluso M, Adashi EY. Fertility treatments and multiple births in the United States. N Engl J Med. 2013;369:2218–25. https://doi.org/10.1056/NEJMoa1301467.

6. Beaujouan E. Latest-Late Fertility? Decline and Resurgence of Late Parenthood Across the Low-Fertility Countries. Popul Dev Rev. 2020;46:219–47. https://doi.org/10.1111/padr.12334.

7. Levine H, Jørgensen N, Martino-Andrade A, Mendiola J, Weksler-Derri D, Mindlis I, et al. Temporal trends in sperm count: A systematic review and meta-regression analysis. Hum Reprod Update. 2017;23:646–59. https://doi.org/10.1093/humupd/dmx022.

8. Schmidt L, Sobotka T, Bentzen JG, Nyboe Andersen A. Demographic and medical consequences of the postponement of parenthood. Hum Reprod Update. 2012;18:29–43. https://doi.org/10.1093/humupd/dmx022.

9. Talmor A, Dunphy B. Female obesity and infertility. Best Pract Res Clin Obstet Gynaecol. 2015;29:498–506. https://doi.org/10.1016/j.bpobgyn.2014.10.014.

10. TSEVAT DG, WIESENFELD HC, Parks C, PEIPERT JF. Sexually Transmitted Diseases and Infertility. Am J Obstet Gynecol. 2017;216:1–9. https://doi.org/10.1016/j.ajog.2016.08.008.

11. Adamson GD, Mouzon J de, Chambers GM, Zegers-Hochschild F, Mansour R, Ishihara O, et al. International Committee for Monitoring Assisted Reproductive Technology: World report on assisted reproductive technology, 2011. Fertil Steril. 2018;110:1067–80. https://doi.org/10.1016/j.fertnstert.2018.06.039.

12. Australian Institute of Health and Welfare. Australia's mothers and babies 2018: in brief. Canberra:Australia; 2020.

13. Wang YA, Dean JH, Grayson N & Sullivan EA. Assisted reproduction technology in Australia and New Zealand 2004. Sydney; 2004.

14. Palomba S, Homburg R, Santagni S, La Sala GB, Orvieto R. Risk of adverse pregnancy and perinatal outcomes after high technology infertility treatment: A comprehensive systematic review. Reprod Biol Endocrinol. 2016;14:76. https://doi.org/10.1186/s12958-016-0211-8.

15. Zhao J, Yan Y, Huang X, Li Y. Do the children born after assisted reproductive technology have an increased risk of birth defects? A systematic review and meta-analysis. J Matern Fetal Neonatal Med. 2018:1–12. https://doi.org/10.1080/14767058.2018.1488168.

16. Giorgione V, Parazzini F, Fesslova V, Cipriani S, Candiani M, Inversetti A, et al. Congenital heart defects in IVF/ICSI pregnancy: Systematic review and meta-analysis. Ultrasound Obstet Gynecol. 2018;1 (51):33–42. https://doi.org/10.1002/uog.18932.

17. Dayan N, Joseph KS, Fell DB, Laskin CA, Basso O, Park AL, et al. Infertility treatment and risk of severe maternal morbidity: A propensity score–matched cohort study. CMAJ. 2019;191:E118-27. https://doi.org/10.1503/cmaj.181124.

18. Lacamara C, Ortega C, Villa S, Pommer R, Schwarze JE. Are children born from singleton pregnancies conceived by ICSI at increased risk for congenital malformations when compared to children conceived naturally? A systematic review and meta-analysis. JBRA Assist Reprod. 2017;21:251–9. https://doi.org/10.5935/1518-0557.20170047.

19. Messerlian C, Gaskins AJ. Epidemiologic Approaches for Studying Assisted Reproductive Technologies: Design, Methods, Analysis and Interpretation. Current epidemiology reports. 2017;4:124–32. https://doi.org/10.1007/s40471-017-0105-0.

20. Reigstad MM, Larsen IK, Myklebust TÅ, Robsahm TE, Oldereid NB, Brinton LA, Storeng R. Risk of Cancer in Children Conceived by Assisted Reproductive Technology. Pediatrics. 2016;137:e20152061. https://doi.org/10.1542/peds.2015-2061.

21. Palomba S, Santagni S, Daolio J, Gibbins K, Battaglia FA, La Sala GB, Silver RM. Obstetric and perinatal outcomes in subfertile patients who conceived following low technology interventions for fertility enhancement: A comprehensive review. Arch Gynecol Obstet. 2018;297:33–47. https://doi.org/10.1007/s00404-017-4572-9.

22. Australian Govement National Health and Medical Research Council. Ethical guidelines on the use of assisted reproductive technology in clinical practice and research. Canberra: National Health and Medical Research Council; 2017.

23. Australian Bureau of Statistics. National, state and territory population. June 2020. https://www.abs.gov.au/statistics/people/population/national-state-and-territory-population/jun-2020.

24. Irvine K, Hall R, Taylor L. Centre for Health Record Linkage. International journal of population data science 2019. https://doi.org/10.23889/ijpds.v4i2.1142.

25. Taylor LK, Irvine K, Iannotti R, Harchak T, Lim K. Optimal strategy for linkage of datasets containing a statistical linkage key and datasets with full personal identifiers. BMC Med Inform Decis Mak. 2014;14:85. https://doi.org/10.1186/1472-6947-14-85.

26. SaxInstitute. The Secure Unified Research Environment. Sydney: Australia: SaxInstitute.

27. Cohen J. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement. 1960;20:37–46. https://doi.org/10.1177/001316446002000104.

28. Fleiss JL, Levin B, Paik MC, Shewart WA, Wilks SS. Statistical Methods for Rates and Proportions 2003. https://doi.org/10.1002/0471445428.

29. Coulson TG, Bailey M, Reid C, Shardey G, Williams-Spence J, Huckson S, et al. Linkage of Australian national registry data using a statistical linkage key. BMC Med Inform Decis Mak. 2021;21:37. https://doi.org/10.1186/s12911-021-01393-1.

30. Karmel R, Anderson P, Gibson D, Peut A, Duckett S, Wells Y. Empirical aspects of record linkage across multiple data sets using statistical linkage keys: The experience of the PIAC cohort study. BMC Health Serv Res. 2010;10:41. https://doi.org/10.1186/1472-6963-10-41.

31. Taylor LK, Irvine K, Iannotti R, Harchak T, Lim K. Optimal strategy for linkage of datasets containing a statistical linkage key and datasets with full personal identifiers. BMC Med Inform Decis Mak. 2014;14:85. https://doi.org/10.1186/1472-6947-14-85.

32. Centre for Health Record Linkage. Master Linkage Key (MLK). 2021. https://www.cherel.org.au/master-linkage-key.

33. Reproductive Technology Accreditation Committee. Code of Practise for Assisted Reproductive Technology Units in Australia and New Zealand. 2017. https://www.fertilitysociety.com.au/wp-content/uploads/2017-RTAC-ANZ-COP-FINAL-1.pdf.

34. Sunderam S, Schieve LA, Cohen B, Zhang Z, Jeng G, Reynolds M, et al. Linking birth and infant death records with assisted reproductive technology data: Massachusetts, 1997-1998. Matern Child Health J. 2006;10:115–25. https://doi.org/10.1007/s10995-005-0013-7.

35. Zhang Y, Cohen B, Macaluso M, Zhang Z, Durant T, Nannini A. Probabilistic linkage of assisted reproductive technology information with vital records, Massachusetts 1997-2000. Matern Child Health J. 2012;16:1703–8. https://doi.org/10.1007/s10995-011-0877-7.

36. Mneimneh AS, Boulet SL, Sunderam S, Zhang Y, Jamieson DJ, Crawford S, et al. States Monitoring Assisted Reproductive Technology (SMART) Collaborative: Data collection, linkage, dissemination, and use. J Womens Health (Larchmt). 2013;22:571–7. https://doi.org/10.1089/jwh.2013.4452.

37. Kotelchuck M, Hoang L, Stern JE, Diop H, Belanoff C, Declercq E. The MOSART database: Linking the SART CORS clinical database to the population-based Massachusetts PELL reproductive public health data system. Matern Child Health J. 2014;18:2167–78. https://doi.org/10.1007/s10995-014-1465-4.

38. Kissin DM, Adamson GD, Chambers GM, Geyter C de. Chapter 9 Monitoring Long-Term Outcomes of ART: Linking ART Surveillance Data with Other Datasets from Assisted reproductive technology surveillance. Cambridge: Cambridge University Press; 2019.

39. Stern JE, Gopal D, Liberman RF, Anderka M, Kotelchuck M, Luke B. Validation of birth outcomes from the Society for Assisted Reproductive Technology Clinic Outcome Reporting System (SART CORS): Population-based analysis from the Massachusetts Outcome Study of Assisted Reproductive Technology (MOSART). Fertil Steril. 2016;106:717–722.e2. https://doi.org/10.1016/j.fertnstert.2016.04.042.

40. Cohen B, Bernson D, Sappenfield W, Kirby RS, Kissin D, Zhang Y, et al. Accuracy of assisted reproductive technology information on birth certificates: Florida and Massachusetts, 2004-06. Paediatr Perinat Epidemiol. 2014;28:181–90. https://doi.org/10.1111/ppe.12110.

## Abbreviations

ACT: Australian Capital Territory

ACT APC: Australian Capital Territory Admitted Patient Care

AIHW: Australian Institute of Health and Welfare

ANZARD: Australia and New Zealand Assisted Reproduction Database

ART: Assisted reproductive technologies

CHeReL: New South Wales Centre for Health Record Linkage

DI: Donor insemination

DOB: Date of birth

IUI: Intrauterine insemination

IVF: In-vitro fertilisation

MAR: Medically Assisted Reproduction

MEF: Medicare Enrolment File

MLK: Master Linkage Key

NPESU: National Perinatal Epidemiology and Statistics Unit

NSW: New South Wales

NSW APDC: New South Wales Admitted Patient Data Collection

OI: Ovulation induction

PDC: Perinatal data collections

PPN: Project Person Number

RBDM: Registry of Births, Deaths and Marriages

SLK: Statistical Linkage Key

# Supplementary Appendix

Figure S1: Overall concordance rate[1] between ANZARD and New South Wales (NSW) and Australian Capital Territory (ACT) perinatal data collection data for the final linked data of the the medically assisted reproduction (MAR) data linkage for New South Wales (NSW) residents[2], by treatment year, 2009–2015[3]
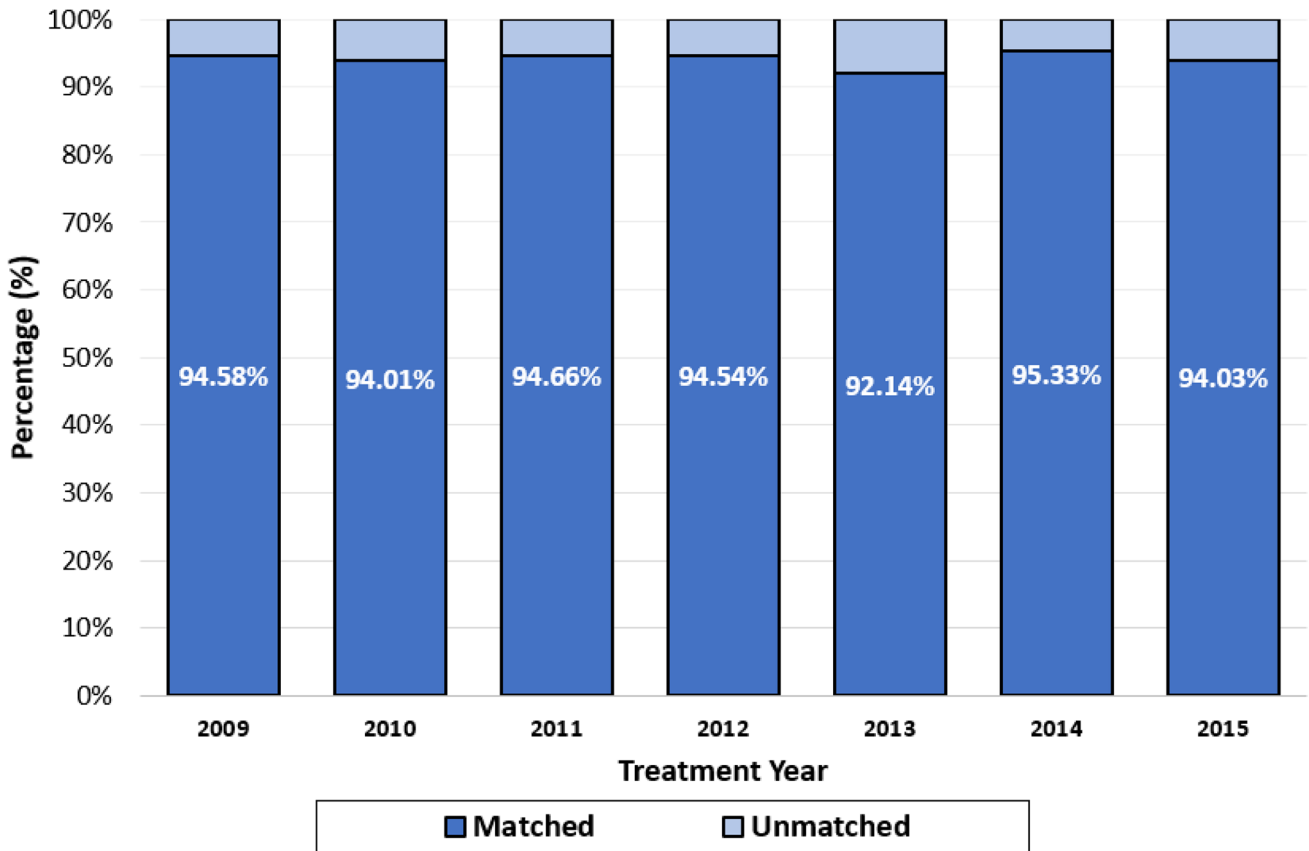


ANZARD = Australia and New Zealand Assisted Reproduction Database
[1]The data concordance rate was calculated as a total number of births in agreement between ANZARD and PDC data divided by a total number of ANZARD treatment records from NSW residents with an ANZARD birth recorded or without an ANZARD birth recorded due to loss of follow-up.
[2]This agreement analysis is only conducted for births resulting from ART treatment by NSW residents and birthing in NSW or ACT. The ACT PDC data only covers births delivered in ACT public hospitals; thus, births to women residing in the ACT and who gave birth in a private hospital are missing from the ACT PDC, estimated to be about 20-25% of ACT births.
[3]To assess the concordance rate between ANZARD and the PDC, the cohort was restricted to ANZARD's cycles performed between 2009 and 2015 (rather than 2016) because ACT PDC data was only available up to December 31, 2016, and one year of follow-up after an ART cycle was needed to capture birth in the PDCs.

Supplementary table 1: Details for each data sources in the medically assisted reproduction (MAR) data linkage, by New South Wales and Australian capital territory

| Dataset | Jurisdiction | Women/mothers/ babies/siblings | Description | Key measures |
|---|---|---|---|---|
| **Australia and New Zealand Assisted Reproduction Database (ANZARD)** | Australia + overseas | Women | The ANZARD is a regional registry of all ART cycles performed in female patients who attended 90 Australian fertility clinics between $1^{st}$ January, 2009 and $31^{st}$ December, 2016. The ANZARD data contains cycle records from both Australian patients and patients from overseas. | female patient and their partner: the treatment and laboratory techniques performed, infertility information, treatment outcomes (pregnancy and birth), and infant perinatal outcomes. |
| **NSW and ACT Perinatal Data Collection (PDC)**[1] | NSW + ACT | Mothers, babies, siblings[2] | PDC encompasses all live births and stillbirths of at least 20 weeks gestation or at least 400 grams birth weight. For each birth, the attending midwife or medical practitioner completes a form (or its electronic equivalent), giving socio-demographic, pregnancy-related risk, medical and obstetric information on the mother, and information on the labour, delivery, and condition, and birth outcome of the infant. | Socio-demographic, pregnancy-related risk, medical and obstetric information on the mother, and information on the labour, delivery, and condition, and birth outcome of the infant |
| **NSW Admitted Patient Data Collection (APDC) and ACT Admitted Patient Care (APC)**[1] | NSW + ACT | Mothers, babies, siblings[2] | The APDC/APC encompass all inpatient separations (discharges, transfers, and deaths) from all public, private, psychiatric and repatriation hospitals in NSW or ACT, as well as public multi-purpose services, private day procedure centres and public nursing homes. | Date of admission and separation, diagnosis, procedure performed, length of stay, etc. |
| **NSW and ACT Registry of Births, Deaths and Marriages (RBDM) Birth Registry**[1] | NSW + ACT | Babies, siblings[2] | The RBDM birth registry encompasses all birth registrations for all births born in NSW and ACT. | Baby's sex, date of birth, Mother's Aboriginal status, plurality, and age of other parent, etc. |
| **NSW and ACT RBDM Death Registry and Cause of Death Unit Record File (COD-URF) Registry**[1] | NSW + ACT | Babies, siblings[2] | The RBDM death registry encompass all deaths that were certified by a registered medical practitioner and COD-URF registry contains the cause of death that the coroner registered after an inquiry into the death circumstances. This data source contains the age of death (in months and years) of the infants and their underlying and contributing cause. | Age of death (in months and years) of the infants and their underlying and contributing cause and the corresponding diagnostic codes |
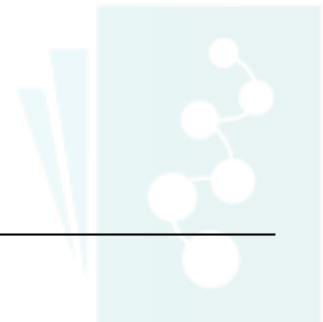
Supplementary table 1: (Continued)

| Dataset | Jurisdiction | Women/mothers/ babies/siblings | Description | Key measures |
|---|---|---|---|---|
| **The NSW Register of Congenital Conditions (RoCC) Registry** | NSW | Babies | The RoCC Registry is a population-based surveillance system established to monitor congenital conditions detected during pregnancy or at birth, or diagnosed in infants up to one year of age under the NSW Public Health Act in 1991. | Month and year of date of birth of baby, the month and year of date of birth of mother, congenital condition, and the corresponding diagnostic code |
| **Pharmaceutical Benefits Scheme (PBS)**[1] | NSW + ACT | Mothers | The PBS encompasses all drug dispensing records that qualify for a benefit under the National Health Act 1953. | Date of prescription and supply, prescription drug code (Anatomical Therapeutic Chemical code and PBS item code), patient category, quantity, dose, drug name, benefit paid by the Australia government, etc |
| **Medicare Benefits Schedule (MBS)**[1] | NSW + ACT | Mothers | The MBS encompasses all clinically relevant medical services subsidised by the Australian Government. | Item codes for fertility-related procedure, service date, fee charges, benefit paid, etc. |

NSW = New South Wales; ACT = Australian Capital Territory.

[1]All ACT linked data only covers births delivered in ACT public hospitals.

[2]We are currently requesting the siblings' data (who were born before 1st January, 2009) to add in the MAR data linkage.

Supplementary table 1: Details for each data sources in the medically assisted reproduction (MAR) data linkage, by New South Wales and Australian capital territory (Cont'd)

| Dataset | Number of records | Number of persons | Years of Coverage |
|---|---|---|---|
| **ANZARD** | Australia +overseas: 638,036 records | 195,390 women (note: the 195,490 SLKs that includes duplicates) | 1st January, 2009 and 31st December, 2016 |
| **PDC**[1] | NSW: 1,136,633 records<br>ACT: 45,052 records | NSW:<br>• 581,241 mothers<br>• 874,922 babies (born between 2009-2017; 261,711 siblings of the baby cohort (born before 2009)<br>ACT:<br>• 27,631 mothers<br>36,964 babies (born between 2009-2016); 8,088 siblings of the baby cohort (born before 2009) | In the MAR data linkage, a birth cohort compromises all NSW and ACT babies born after 1st January, 2009 to the date of the latest available data, i.e., 31st December, 2017, for NSW and 31st December, 2016 for ACT.<br>The MAR linkage also comprises the birth cohort's siblings from the same biological mothers and siblings were born from 1st January, 1994, and 1st January, 1997, to 31st December, 2008 in NSW and in ACT, respectively. |
| **NSW APDC and ACT APC**[1] | NSW: 4,180,473 records<br>ACT: 140,911 records | NSW: 579,601 mothers; 862,418 babies<br>ACT: 27,970 mothers; 35,800 babies | The MAR data linkage contains NSW and ACT hospital admission records for mothers between 1st July, 2001, and 30th June, 2018; for the babies in the birth cohort between 1st January, 2009, and 30th June, 2018; for the siblings of the birth cohort between 1st January, 2001, and 31st December, 2010. |
| **NSW and ACT RBDM**[1] | NSW: 834,702 records<br>ACT: 35,398 records | NSW: 828,084 babies<br>ACT: 35,382 babies | The MAR data linkage contains all birth registration between 1st January, 1994, and 31st December, 2017 in NSW between 1st January, 1997, and 31st December, 2018, in ACT. |
| **NSW and ACT COD-URF**[1] | NSW: 3,528 records<br>ACT: 165 records | NSW: 3,511 babies<br>ACT: 165 babies | For NSW, the MAR data linkage contains death records for the babies and their siblings between 1st January, 1994, and 30th June, 2018. It includes the cause of death information between 1st January, 1994, and 31st December, 2016, for NSW.<br>For ACT, the MAR data linkage also contains death records and cause of death for the babies and their siblings between 1st January, 1997 and 30th June, 2018. |
| **NSW RoCC** | NSW: 8,098 records | NSW: 8,090 babies | In the MAR data linkage, this data source is only available between 1st January, 2011, to 31st December, 2015, because it only allows a five-year data coverage for any linked data. |
| **PBS**[1] | NSW & ACT: 23,033,842 records | NSW & ACT: 573,045 mothers | In the MAR data linkage, this data source contained all fertility-related drug dispensing for the mothers between 1st July, 2002 to 31st December, 2006 and all drug dispensing for mothers between 1st January, 2007 to 31st March, 2019. |
| **MBS**[1] | NSW & ACT: 751,281 records | NSW & ACT: 64,386 mothers | In the MAR data linkage, this data source contained all fertility-related procedures between 1st July, 2002, to 31st March, 2019. |

NSW = New South Wales; ACT = Australian Capital Territory.<br>
[1]All ACT linked data only covers births delivered in ACT public hospitals.<br>
[2]We are currently requesting the siblings' data (who were born before 1st January, 2009) to add in the MAR data linkage.

Supplementary table 2: Sensitivity analysis of the concordance rate between ANZARD and New South Wales (NSW) and Australian Capital Territory (ACT) perinatal data collection for the final linked data of the medically assisted reproduction (MAR) data linkage, for NSW residents[1]

| | Grace period for criteria 1 (in days)[2] | Grace period for criteria 2 (in days)[2] | Total number of ART treatment cycle | Number of direct links between ART Treatment cycles and PDC births | Concordance rate |
|---|---|---|---|---|---|
| **Main Analysis** | **16** | **10** | **28,937** | **27,248** | **94.2%** |
| **Scenario #1** | 16 | 5 | 28,937 | 27,243 | 94.1% |
| **Scenario #2** | 16 | 15 | 28,937 | 27,250 | 94.2% |
| **Scenario #3** | 5 | 10 | 28,937 | 27,248 | 94.2% |
| **Scenario #4** | 5 | 5 | 28,937 | 27,243 | 94.1% |
| **Scenario #5** | 5 | 15 | 28,937 | 27,250 | 94.2% |
| **Scenario #6** | 31 | 10 | 28,937 | 27,249 | 94.2% |
| **Scenario #7** | 31 | 5 | 28,937 | 27,249 | 94.2% |
| **Scenario #8** | 31 | 15 | 28,937 | 27,250 | 94.2% |

ANZARD = Australia and New Zealand Assisted Reproduction Database.
[1]NSW and ACT residents who have undergone an ART Treatment with a known birth outcome or an unknown birth outcome due to loss to follow-up and an agreement in birth recorded in PDC.
[2] Details of criteria 1 & 2 can be found in Table 1.

Supplementary table 3: Concordance of plurality status between ANZARD and PDC data[1], for New South Wales (NSW) residents

| | | Plurality recorded in PDC | | |
|---|---|---|---|---|
| | | 1 | 2 | 3 |
| **Plurality recorded in ANZARD** | 1 | 25,758 | 33 | 0 |
| | 2 | 27 | 1,377 | <5 |
| | 3 | 0 | 0 | 35 |
| | **No delivery or Missing** | 14 | <5 | 0 |

ANZARD = Australia and New Zealand Assisted Reproduction Database; Perinatal Data Collection.
[1] NSW residents who have undergone an ART Treatment with a known birth outcome or an unknown birth outcome due to loss to follow-up and we were successfully matched the ANZARD birth to a birth in the PDC data. This analysis is only feasible to NSW residents.