The PFAS Health Study: Data Linkage Study

Research Protocol

Rosemary Korda[1], Hsei Di Law[1], Bruce Armstrong[2,3], Catherine D'Este[1], Kayla Smurthwaite[1], Susan Trevenar[1], Robyn Lucas[1], Adrian Miller[4], Martyn Kirk[1]

1. National Centre for Epidemiology and Population Health, Research School of Population Health, The Australian National University, Canberra, ACT, Australia.
2. Sydney School of Public Health, University of Sydney, NSW, Australia
3. Faculty of Health and Medical Sciences, University of Western Australia, WA, Australia
4. Office of Pro Vice-Chancellor (Indigenous Engagement), Central Queensland University, Qld, Australia

Australian National University

For the Commonwealth Department of Health

Contents

Abbreviations

Abbreviation	
AARNET	Australian Academic and Research Network
ABS	Australian Bureau of Statistics
ACCHO	Aboriginal Community Controlled Health Organisation
ACD	Australian Cancer Database
ACT	Australian Capital Territory
ADF	Australian Defence Force
AEDC	Australian Early Development Census
AFFF	Aqueous film forming foams
AIHW	Australian Institute of Health and Welfare
AMI	Acute myocardial infarction
ANU	Australian National University
APDC	Admitted Patient Data Collections
APGAR	Appearance, Pulse, Grimace, Activity, and Respiration
APP	Australian Privacy Principles
ARIA	Accessibility and Remoteness Index of Australia
AvEDI	Australian version of the Early Development Instrument
COPD	Chronic obstructive pulmonary disease
CVD	Cardiovascular disease
DDVA	Departments of Defence and Veteran's Affairs
EPA	Environmental Protection Authority
HR	Hazard ratio
HREA	Human Research Ethics Application
HREC	Human Research Ethics Committee
ICD-10	International Classification of Diseases 10th Revision
ICD-10-AM	International Classification of Diseases 10th Revision, Australian Modification
IMA	Investigation or Management Area
IRSD	Index of Relative Socioeconomic Disadvantage
LGA	Low for Gestational Age
MEF	Medicare Enrolment File
NDI	National Death Index
NMA	National Mutual Acceptance
NSW	New South Wales
NT	Northern Territory

Abbreviation	
NTDoHMSHR	Northern Territory Department of Health and Menzies School of Health Research
OR	Odds ratio
PDC	Perinatal Data Collections
PFAS	Per- and polyfluoroalkyl substances
PFHxS	Perfluorohexane sulfonic acid
PFOA	Perfluorooctanoic acid
PFOS	Perfluorooctane sulfonic acid
PMKeyS	Personnel Management Key Solution Database
PR	Prevalence Ratio
Qld	Queensland
RR	Relative Rates
SA	South Australia
SEIFA	Socio-Economic Indexes for Areas
SES	Socio-economic status
SGA	Small for Gestational Age
SIR	Standardised incidence ratio
SURE	Secure Unified Research Environment
Tas	Tasmania
Vic	Victoria
VII	Voluntary Indigenous Indicator
WA	Western Australia

1 Project summary

This data linkage study is one component of The Per- and Poly-Fluoroalkyl Substances (PFAS) Health Study: Phase II, which is investigating the exposure to and potential health effects of PFAS in areas of known contamination in the communities of Williamtown in New South Wales, Oakey in Queensland and Katherine in the Northern Territory, Australia.

The primary goal of the data linkage study is to examine whether adverse health outcomes potentially associated with PFAS exposure are more common among people who have lived in the PFAS Investigation and Management Areas of Williamtown, Oakey and Katherine (exposed), than among people who have lived outside these areas (non-exposed), after accounting for sociodemographic characteristics.

The study will use linked routinely-collected data to estimate rates[a] of candidate health outcomes in the exposed and non-exposed populations and compare them, adjusting for sociodemographic characteristics including age, sex, socioeconomic status and remoteness. Candidate outcomes are those potentially associated with PFAS exposure, as identified from a systematic review of the literature. In addition, the study will investigate selected health outcomes for which there is no known evidence of an association with PFAS exposure—referred to as control outcomes.

The data linkage study is expected to be completed by December 2020.

We expect the study to produce knowledge that will assist community members in understanding the impact on their health, if any, of living in a PFAS Investigation and Management Area, and that will assist policy makers in responding to PFAS contamination issues.

[a] Most outcomes are measured as rates, although some outcomes are measured in terms of prevalence and a few as continuous measures. In this context, the term "*rates*" will be used for brevity, rather than referring to all outcome types separately each time.

2 General information

2.1 Protocol title

The PFAS Health Study: Data Linkage Study Research Protocol

2.2 Protocol date

24 May 2019

2.3 Project funding

The Australian Government Department of Health[1] has commissioned the Australian National University (ANU) to undertake the PFAS Health Study.

[1]Department of Health
GPO Box 9848
Canberra ACT 2601, Australia

2.4 Investigators

PFAS Health Study Team

Principal Investigator

Professor Martyn Kirk, Australian National University[1]	T: +61 2 6125 5609

Co-Investigators

Associate Professor Rosemary Korda, Australian National University[1]	T: +61 2 6125 5583
Professor Robyn Lucas, Australian National University[1]	T: +61 2 6125 3448
Professor Adrian Miller, Central Queensland University[2]	T: +61 7 4726 5382
Professor Jochen Mueller, University of Queensland[3]	T: +61 7 3443 2450
Professor Archie Clements, Curtin University[4]	T: +61 8 9266 7466
Professor Catherine D'Este, Australian National University[1]	
Emeritus Professor Bruce Armstrong, University of Sydney[5]	T: +61 4 0349 6404

Associate Investigators

Professor Cathy Banwell, Australian National University[1]	T: +61 2 6125 0016
Associate Professor Philip Batterham, Australian National University[6]	T: 61 2 6125 1031
Dr Jennifer Bräunig, University of Queensland[3]	T: +61 7 3446 1899
Dr Tambri Housen, Australian National University[1]	T: +61 2 6125 0460
Dr Aparna Lal, Australian National University[1]	T: +61 2 6125 2309
Dr Katherine Todd, Australian National University[1]	T: +61 3 9096 0339
Dr Miranda Harris, Australian National University[1]	

Research Officers

Ms Susan Trevenar, Australian National University[1]	T: +61 2 6125 6079
Ms Hsei Di Law, Australian National University[1]	T: +61 2 6125 0547

Research Assistants

Ms Kayla Smurthwaite, Australian National University[1]	T: +61 2 6125 7840
Ms Eva Smeets, Australian National University[1]	T: +61 2 6125 0547
Ms Anna Rafferty, Australian National University[1]	T: +61 2 6125 7840

PFAS Data Linkage Study Team

PFAS Data Linkage Study Team
Principal Investigator
Associate Professor Rosemary Korda, Australian National University
Co-Investigators
Professor Martyn Kirk, Australian National University
Professor Robyn Lucas, Australian National University
Professor Adrian Miller, Central Queensland University
Professor Jochen Mueller, University of Queensland
Professor Archie Clements, Curtin University
Professor Catherine D'Este, Australian National University
Emeritus Professor Bruce Armstrong, University of Sydney
Ms Hsei Di Law, Australian National University
Associate Investigators
Dr Jennifer Bräunig, University of Queensland
Dr Miranda Harris, Australian National University
Research Officer
Ms Susan Trevenar, Australian National University
Research Assistants
Ms Kayla Smurthwaite, Australian National University
Ms Eva Smeets, Australian National University
Ms Anna Rafferty, Australian National University

[1]National Centre for Epidemiology and Population Health
Research School of Population Health
ANU College of Health and Medicine
Building 62, Cnr of Eggleston and Mills Roads
The Australian National University
Acton ACT 2601

[2] Office of the Pro Vice-Chancellor (Indigenous Engagement)
Central Queensland University
538 Flinders Street
Townsville Qld 4810

[3] Queensland Alliance for Environmental Health Sciences
Pharmacy Australia Centre of Excellence (PACE) Building
20 Cornwall Street
The University of Queensland
Woolloongabba Qld 4072

[4] Office of the Pro Vice-Chancellor Health Sciences
Curtin University
Kent St
Bentley WA 6102

[5] School of Public Health
Edward Ford Building A27
The University of Sydney
NSW 2006

[6] Centre for Mental Health Research
Research School of Population Health
ANU College of Health and Medicine
Building 62, Cnr of Eggleston and Mills Roads
The Australian National University
Acton ACT 2601

2.5 Study management

Professor Martyn Kirk
National Centre for Epidemiology and Population Health
Research School of Population Health
The Australian National University
ACT 2601 AUSTRALIA
Telephone: +61 2 6125 5609
Mobile: +61 4 2613 2181
Email: pfas.health.study@anu.edu.au

3 Rationale and background information

Per- and polyfluoroalkyl substances (PFAS) are a diverse family of fluorinated organic chemicals that have been produced commercially since the 1950s. [1] Due to their stability and hydrophilic and lipophilic properties, PFAS have been used in a wide variety of consumer products, including surface treatments for textiles, non-stick coatings for cookware, grease-repellent food packaging and paints; and in industrial applications, such as in the metal plating industry, in hydraulic fluids and as key ingredients in aqueous film-forming foams (AFFF). The extensive use of these chemicals and their persistence has led to concerns about environmental and human health impacts.

In recent years, historical firefighting activities on Australian Defence Force (ADF) Bases have been linked to environmental PFAS contamination in nearby areas of Williamtown in New South Wales (NSW), Oakey in Queensland (Qld) and Katherine in the Northern Territory (NT). Use of AFFF containing PFAS as the main components have been associated with elevated PFAS concentrations in ground water, soil and biota. [2-4] Members of these communities have probably been exposed to PFAS through the ingestion and use of contaminated bore water, with exposure occurring from the 1970s onwards. [5-7]

Concerns over the potential for PFAS to adversely affect human health arise from their ease of absorption into and distribution through the body, and their prolonged half-life in humans. In 2004, the ADF began phasing out the use of AFFF concentrates that contained PFOS and PFOA as active ingredients, switching to alternative chemicals with similar properties. [8] However, the chemicals in the formerly used long-chain fluorinated foams (≥C7) are likely to persist in the environment, particularly in ground water, sediment and soil, of locally contaminated areas. [8]

Information from animal studies, which generally use high oral doses of PFAS, have indicated potential effects on pre- and postnatal development and the liver and immune system. [9] However, the toxicity of PFAS in humans is poorly understood. There have been a range of proposed mechanisms for possible adverse health effects of PFAS, many of which relate to endocrine disruption potentially affecting male and female reproduction and thyroid function. [10, 11]

Exposure to PFAS chemicals has raised health concerns for people in the affected communities. There is considerable anxiety for current and past residents of these communities, about the risk of developing disease, especially cancer, as a result of living in these areas. The literature to date examining the health effects of PFAS exposure has not provided definitive answers regarding the risk of adverse health outcomes for those who have been exposed.

3.1 The PFAS Health Study

The public health significance of PFAS exposure is unclear. Risk assessments of contaminated areas have evaluated likely exposure but have been unable to quantify associated health risks due to inconsistent evidence about the health outcomes from exposure. To date, there have been no population-based epidemiological studies examining the health effects of PFAS in Australia as the affected communities are often small and the levels of exposure highly variable.

The PFAS Health Study has five main components, over two phases. During Phase I a systematic review was conducted to examine the evidence on health effects of PFAS in humans as reported in published literature. Phase II is an epidemiological study of the PFAS contamination in three Australian communities, Williamtown (NSW), Oakey (Qld) and Katherine (NT). This phase comprises four component studies:

I. Component 1 – Focus Group Study

A focus group study to determine the concerns that individuals living in the vicinity of Williamtown, Oakey and Katherine have in relation to exposure to PFAS and their health.

II. Component 2 – Cross-sectional Survey

A cross-sectional survey to investigate PFAS exposure and risk factors for high serum PFAS levels, including sociodemographic (e.g. age, sex, location) and other factors (e.g. duration of residence in the area, water source), and associations of high serum PFAS levels with self-reported common symptoms, signs and diagnosed illnesses in the Williamtown, Oakey and Katherine communities.

III. Component 3 – Blood Serum Study

A blood serum study to define the serum concentrations (mean and range) of PFAS in Williamtown, Oakey and Katherine residents living in the PFAS Investigation and Management Areas and to compare these levels to those of people residing in non-contaminated areas.

IV. Component 4 – Data Linkage Study

A data linkage study, hereafter referred to as the PFAS Data Linkage Study, to examine whether adverse health outcomes potentially associated with PFAS are more common among people who have lived in the Williamtown, Oakey and Katherine PFAS Investigation and Management Areas than among people who have not lived in these areas, after accounting for sociodemographic characteristics.

This document outlines a research proposal for Component 4 –Data Linkage Study.

4 Study goals and objectives

4.1 Goals

The primary goal of the PFAS Data Linkage Study is to examine whether adverse health outcomes potentially associated with PFAS exposure are more common among people who have lived in the PFAS Investigation and Management Areas of Williamtown, Oakey and Katherine (the exposed populations) than among people who have lived outside of these areas in Australia (the non-exposed populations). The study will compare rates of candidate outcomes for which routinely-collected data are available in the exposed and non-exposed populations. Candidate outcomes are those potentially associated with PFAS exposure, as identified from a systematic review of the literature. In addition, the study will investigate health outcomes for which there is no known evidence of an association with PFAS exposure—referred to as control outcomes.

4.2 Research questions

The research questions of the study are:

1. What are the relative rates of candidate outcomes in relation to community PFAS exposure in Williamtown, Oakey and Katherine, after adjusting for sociodemographic characteristics?
2. What are the relative rates of control outcomes in relation to community PFAS exposure in Williamtown, Oakey and Katherine, after adjusting for sociodemographic characteristics?

4.3 Hypotheses

If living in a PFAS Investigation and Management Area where there is known PFAS contamination in the local environment is harmful to health, we would expect to observe:

1. Higher rates of candidate outcomes in people who have lived in a PFAS Investigation and Management Area in Australia compared to those who have not i.e. relative rates of candidate outcomes > 1[b], after accounting for sociodemographic characteristics.
2. Similar rates of control outcomes in people who have ever lived in a PFAS Investigation and Management Area in Australia compared to those who have not i.e. relative rates of candidate outcomes ≈ 1[c], after accounting for sociodemographic characteristics.

[b] Or prevalence ratio, risk ratio, odds ratio or hazard ratio >1; or absolute difference > 0

[c] Or prevalence ratio, risk ratio, odds ratio or hazard ratio ≈ 1; or absolute difference ≈ 0.

5 Study Design

5.1 Type of study

This will be a retrospective cohort study using linked routinely-collected data.

5.2 Study population

The study population will include both exposed and non-exposed populations. For most outcomes, the study population will be assembled from the Australian Government Department of Human Services Medicare Enrolment File (MEF). The exposed population will include all people on the MEF who registered an address in the PFAS Investigation and Management Areas of Williamtown, Oakey and Katherine any time between 1984 and 2018[d]. The PFAS Investigation and Management Areas have been determined by the Australian Government Department of Defence in Oakey and Katherine and the NSW Environmental Protection Authority (EPA) in Williamtown based on environmental sampling and are outlined in Attachment 1. The non-exposed population will include a contemporaneous frequency-matched sample of all people on the MEF who registered an address in comparison areas outside of the PFAS Investigation and Management Areas between 1984 and 2018. The comparison areas are postcodes selected by ANU researchers based on similarities to the PFAS Investigation and Management Areas, in terms of state, area-level socioeconomic status and remoteness. More details on the selection of comparison areas are available in section 6.4.

For the group of neonatal, infant and maternal outcomes only, the study population will be assembled from the NSW, Qld and NT Perinatal Data Collections (PDCs). The exposed population will include all mothers and babies in these data who had an address in the PFAS Investigation and Management Areas at any time between inception of the data collection[e] and 2018. The non-exposed population will include a contemporaneous frequency-matched sample of mothers and babies in these data who had an address in comparison areas.

5.3 Sampling frame

The sampling frame includes all individuals on the MEF or PDCs with an address in the PFAS Investigation and Management Areas or a comparison area between 1984[f] and 2018. All Australian citizens and permanent residents are eligible for registration with Medicare. All

[d] While the population of interest is anyone who has ever lived in the PFAS Investigation and Management Area since PFAS exposure was first noted in the 1970s, there are no databases dating back to that period that include the names and address of all residents in Williamtown, Oakey or Katherine (both adults and children).

[e] Or when mother's address was first collected as a data item in each state's PDC.

births that occur in-state, whether in public hospitals, private hospitals and homebirths, are within the scope of the jurisdictional PDCs.

5.4 Inclusion and exclusion criteria

Individuals are eligible for inclusion in the study if they ever had an address on the MEF or PDCs between 1984[f] and 2018 that was either in:

a) the PFAS Investigation and Management Areas of Williamtown, Oakey and Katherine; or
b) one of the comparison areas.

Individuals will be excluded from the study population if they have:

a) missing data for their date of birth or sex;
b) invalid dates, such as those whose recorded date of birth follows entry into the study[g], or whose recorded date of death precedes entry into the study; and/or
c) less than a minimum follow-up time (to be defined), to allow for a minimum latency period between exposure and outcome.

5.5 Recruitment

There is no active recruitment or participation in this study as it only uses routinely-collected data.

5.6 Expected duration

The data linkage study is expected to be completed by December 2020, but is subject to timely approvals and linkage from external agencies.

[f] Or when mother's address was first collected as a data item in each state's PDC.
[g] For the study population assembled from the MEF, entry into the study is defined as the date of his or her first registration with Medicare, regardless of which state/territory they registered in.

6 Methods

6.1 Study procedure

Individual-level data from the MEF will be linked to the PFAS Management and Investigation Areas Address Database, the Personnel Management Key Solution (PMKeyS) database and multiple routinely-collected health databases: the Australian Cancer Database (ACD), state-based Admitted Patient Data Collections (APDC), the Australian Early Development Census (AEDC) and the National Death Index (NDI). For neonatal, infant and maternal outcomes, individual-level data from the NSW, Qld and NT PDCs will be linked to the PFAS Management and Investigation Areas Address Database. Figure 1 shows an overview of the study design. The analytical approach for the linked data is discussed in section 9.2.

6.2 Data Sources

I. Medicare Enrolment File (1984-2018)

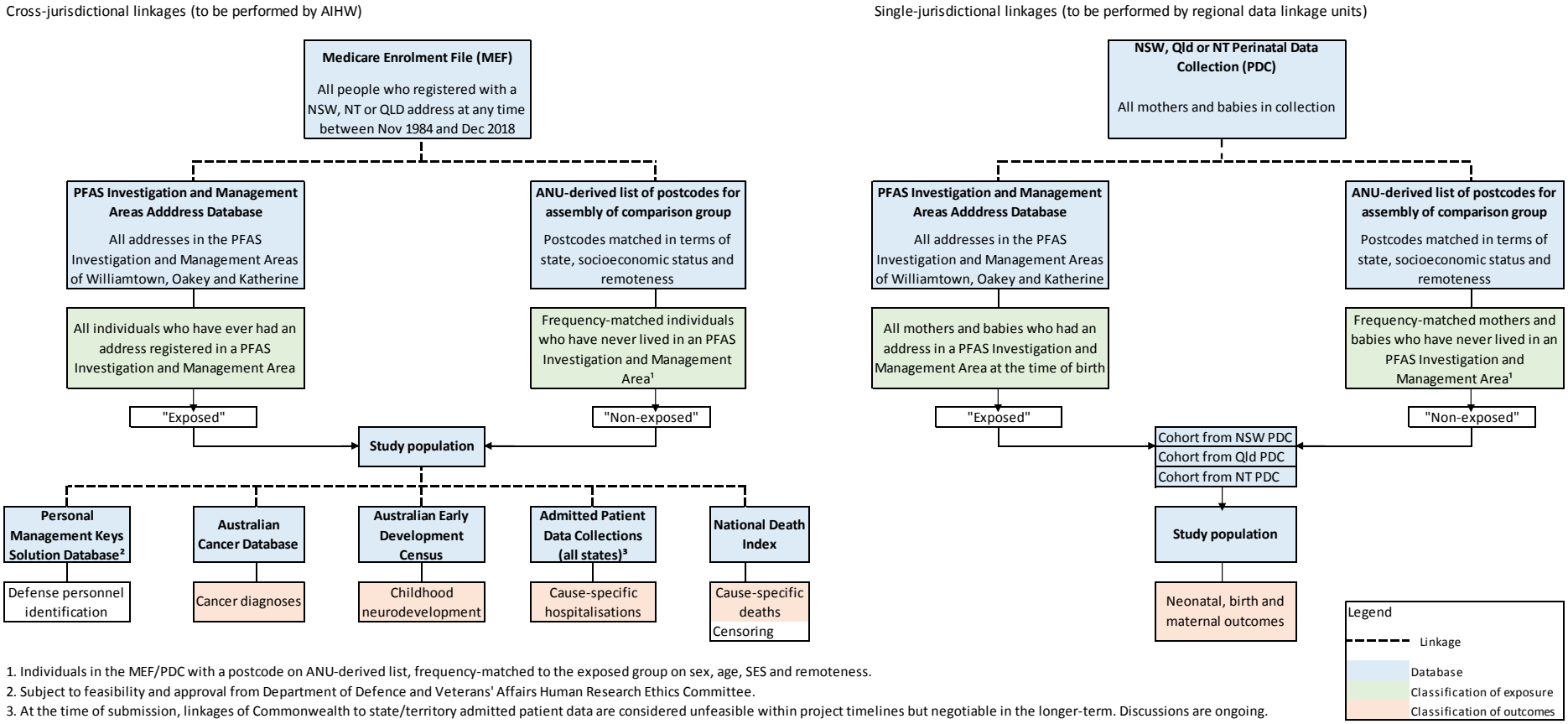
Medicare is Australia's universal health insurance provider, which is open to all Australian and New Zealand citizens living in Australia and permanent residents of Australia. Medicare is administered by the Australian Government Department of Human Services, which collects and stores personal details—including name, sex, date of birth and address—for each registered individual.

If an individual changes their address they are required to notify the Department of Human Services of the change by phone, online or in person. A history of these changes is stored in the MEF. Therefore, multiple address records are held for registered individuals who have moved. A *start date* is associated with every address registered with Medicare. However, the start date is the date the Department was notified of the change. There is often a delay between actual change of address and this change being recorded in the MEF.

The Department of Human Services collects both residential and mailing addresses for the MEF. However, residential addresses are non-mandatory. Only mailing addresses are provided to the Australian Institute of Health and Welfare (AIHW) for data linkage. While mailing and residential addresses are the same for the vast majority of Australians, a proportion of addresses on the AIHW MEF are non-residential, including post office box addresses[h].

[h] See section 9.2(ii) regarding sensitivity analyses that will be conducted to account for this discrepancy.

Figure 1. PFAS Data Linkage Study: Overview of data linkage showing data sources and study population, Australia, 1984-2018.



1. Individuals in the MEF/PDC with a postcode on ANU-derived list, frequency-matched to the exposed group on sex, age, SES and remoteness.
2. Subject to feasibility and approval from Department of Defence and Veterans' Affairs Human Research Ethics Committee.
3. At the time of submission, linkages of Commonwealth to state/territory admitted patient data are considered unfeasible within project timelines but negotiable in the longer-term. Discussions are ongoing.

The MEF will be used to assemble a study population (for analysing all outcomes except neonatal, infant and maternal outcomes[i]) and classify members as exposed or non-exposed. The exposed cohort will be identified via linkage of address variables in the MEF to the PFAS Investigation and Management Areas Address Database held by the ANU (described below). The non-exposed cohort will be selected from those in the MEF who registered with a postcode in comparison areas outside of PFAS Management and Investigation Areas. Subsequently, personal-identifying information from the MEF for the study population—including full name, sex, date of birth and postcode—will be required for linkages to the ACD, APDC, AEDC and the NDI to ascertain their health outcomes (and death dates for censoring). The researchers will not have access to personal-identifying information, only a categorical variable supplied by the AIHW data linkage unit that indicates which area the individual is from.

II. PFAS Investigation and Management Areas Address Database

The PFAS Investigation and Management Areas Address Database list the addresses of 5,137 properties in the PFAS Investigation and Management Areas of Williamtown (757), Oakey (1,755) and Katherine (2,625). The boundaries of the PFAS Investigation and Management Areas have been determined by the Australian Government Department of Defence in Oakey and Katherine and the NSW EPA in Williamtown, based on environmental sampling. A list of addresses included in these areas was compiled by the Australian Government Department of Defence and the NSW EPA and shared with the ANU PFAS Health Study Team in June 2018. This list of addresses will be used identify all individuals in the MEF and jurisdictional PDCs who have had an address in PFAS Investigation and Management Areas i.e. the exposed populations in the study.

III. The Personnel Management Key Solution database (2001-latest)

The PMKeyS database is a Defence staff and payroll management system that contains information on all people with ADF service on or after 1 January 2001, when the system was first introduced. This database contains personal identifiers, demographic and service information.

Defence personnel and their families are thought to be among those exposed to PFAS due to the historical usage of AFFF on ADF bases. However, it is anticipated that PFAS exposure levels among defence personnel differ from those in the community. Therefore, we are proposing a linkage between PMKeyS and our study population in order to derive a binary indicator that flags defence personnel. This linkage is subject to approval from the Departments of Defence and Veteran's Affairs Human Research Ethics Committee (DDVA HREC).

[i] A separate study population will be assembled for analysing neonatal, infant and maternal outcomes using jurisdictional PDCs.

IV. The Australian Cancer Database (1982-2018)

The ACD is a data collection of primary malignant cancers diagnosed in Australia since 1982. Reporting of newly diagnosed cancers has been mandatory in most but not all jurisdictions since at least 1982[j]. The ACD is compiled at the AIHW from cancer data provided by state and territory cancer registries through the Australasian Association of Cancer Registries.

Personal-identifying information for the study population assembled from the MEF will be used to link to the ACD to ascertain cancer outcomes. This data linkage will follow standard AIHW methods using full name, sex, date of birth and postcode. The following data items will be required for analysis:

- International Classification of Diseases 10th Revision (ICD-10) diagnosis codes
- Age at diagnosis
- Date of diagnosis
- Date of death
- Underlying cause of death
- Indigenous status (subject to approval)

Further information on the ACD is available at: http://www.aihw.gov.au/australian-cancer-database.

V. The National Death Index (1980–2018)

The NDI is housed at the AIHW and contains records of all deaths that have occurred in Australia since 1980. The data are provided by the Registries of Births, Deaths and Marriages, the Australian Bureau of Statistics and the National Coroners Information System.

Personal-identifying information for the study population assembled from the MEF will be used to link to the NDI to ascertain death status and cause of death. This data linkage will follow standard AIHW methods using full name, sex, date of birth, postcode and where possible, address. The following data items will be required for analysis:

- Underlying cause of death
- Other causes of death
- Date of death (for censoring).

Further information on the National Death Index can be found at: http://www.aihw.gov.au/national-death-index.

VI. The Australian Early Development Census

The AEDC is a full population census of children's health and development in their first year of formal full-time schooling. It provides a comprehensive map of early developmental

[j] Mandatory reporting in: ACT—1994; NSW—1972; NT—1991; Qld—1982; SA—1977; Tas—1992; Vic—1982; WA—1981

outcomes across Australia, however only neurodevelopmental outcomes are of interest for this study. The first AEDC was administered in 2009 and a collection occurs every three years. The AEDC datasets have near-total national coverage of school entrants, and provide data on 97.5 per cent of the estimated 5-year-old population in 2009 and 96.5 per cent of children enrolled to start school in 2012.

Personal-identifying information for the study population assembled from the MEF will be used to link to the 2009, 2012, 2015 and 2018 collections of the AEDC following standard AIHW linkage methods. The following groups of variables will be required for analysis:

- Demographic variables
- Geographic variables
- AEDC Domain variables

Further information on the AEDC can be found at: https://www.aedc.gov.au/.

VII. NSW, Qld and NT Perinatal Data Collections

Jurisdictional PDCs include information about pregnancy care, services and outcomes. The scope of each PDC includes all live births and stillbirths of at least 400 grams birthweight or at least 20 weeks gestation. Data items related to the mother include demographic characteristics and factors relating to the pregnancy, labour and birth. Data items related to the baby include sex, birth status, gestational age at birth, birthweight, and neonatal morbidity and fetal deaths.

The following jurisdictional PDCs will be used to assemble a separate study population for analysing perinatal outcomes:

- The NSW Perinatal Data Collection (January 1994–2016)
- The Qld Perinatal Data Collection (July 2007–2017)
- NT Perinatal Trends (January 1986–December 2015)

Jurisdictional PDCs collect mother's address of usual residence at time of birth that may be available for data linkage purposes, however these data are at varying levels of detail e.g. street name and number, or simply postcode or Statistical Local Area. Subject to approval and technical feasibility, the exposed population will be identified via linkage of address variables in each PDC to the PFAS Investigation and Management Areas Address Database. The non-exposed cohort will be selected from those in the PDC with a postcode in comparison areas. The following data items will be required for analysis (where available in each state):

- Birth weight
- Gestational age
- Birth length
- Head circumference
- Baby: sex
- Method/type of birth
- Fetal death
- Still birth

- APGAR score at 5 min after birth
- Febrile
- Meconium liquor
- Rupture of membranes
- Placental abruption
- Placenta previa
- Complications of labour
- Precipitate delivery
- Cord prolapse
- Fetal distress in labour
- Eclampsia
- Gestational hypertension
- Gestational diabetes
- Congenital anomaly
- Mother: Indigenous status (subject to approval)
- Baby: Indigenous status (subject to approval)

Potential confounders:

- Mother: date of birth
- Baby: date of birth
- Mother: weight
- Mother: body mass index
- Mother: country of birth
- Area of usual residence
- Mother: usual address
- Alcohol in pregnancy
- No. cigarettes daily
- Smoking before 20 weeks
- No. cigarettes first 20 weeks
- No. cigarettes after 20 weeks
- Smoking after 20 weeks
- Pre-existing diabetes
- Parity
- Marital status
- Pre-existing hypertension
- Plurality

VIII. The Admitted Patient Data Collections

At the time of submission, we were advised by the Population Health Research Network that linkage of Commonwealth to state/territory admitted patient data is not feasible, at least not within the timeframe of the researchers' contractual agreement with the funder. We will continue to negotiate these linkages as a longer-term outcome, however it is unlikely that we will include hospitalisation outcomes during our December 2020 delivery.

The purpose of the APDC is to collect information about hospital care provided to admitted patients in Australian hospitals. While state-based specifications may differ slightly, these collections record data from all inpatient separations (including discharges, transfers or deaths) from recognised public and private hospitals (except in South Australia where private hospital separations are currently not available for data linkage). The collections comprise demographic, clinical and administrative data items and are generally based on standard definitions that comply with the Admitted Patient National Minimum Data Set.

We are proposing linkages of the study population assembled from the MEF to all jurisdictional APDCs from their inception (except for WA and Qld from 1995, which is the

earliest year data are available in the main states of NSW, QLD and NT), to the most recent dates for which data are available:

- ACT Admitted Patient Care (2004–2016)
- NSW Admitted Patient Data Collection (2001–2017)
- NT Inpatient Activity (2000–2017)
- Qld Hospital Admitted Patient Data Collection (1995–2018)
- SA Inpatient Hospital Separations (2001–2017)
- TAS Public Hospital Admitted Patients (2000–2015)
- VIC Admitted Episodes Dataset (1995–ongoing)
- WA Hospital Morbidity Data Collection (1995–ongoing)

We will require the following data items for analysis (where available in each state):

- Hospital type (public/private)
- Age
- Sex
- Date of admission
- Date of separation
- Mode of separation
- ICD-10 Australian Modification (ICD-10-AM) diagnosis codes
- Procedure codes
- Indigenous status (subject to approval)

Only records relating to admissions with specified ICD or procedure codes (see Appendix 1) are required, including those relating to:

- Congenital abnormalities
- Chronic kidney disease
- Liver disease
- Acute myocardial infarction
- Stroke
- Major cardiovascular disease (CVD)
- Hip fractures
- Chronic obstructive pulmonary disease (COPD)

6.3 Data linkage and integration

Formal guidelines for integrating Commonwealth data for research projects were endorsed by the Commonwealth Secretaries Board in 2010. Full details, including how to apply for access to Medicare data for research purposes, are available on the Australian Government

National Statistical Service website at: http://statistical-data-integration.govspace.gov.au. A map detailing the process is shown in Appendix 2.

The Data Integration Services Centre at the AIHW, which is a Commonwealth-accredited data integrating authority, will be the appointed authority for this project. They will facilitate access to the data and perform linkages between the MEF to:

1. The PFAS Investigation and Management Areas Address Database;
2. The Personnel Management Key Solution database;
3. The Australian Cancer Database;
4. Jurisdictional Admitted Patient Data Collections;
5. The Australian Early Development Census; and
6. The National Death Index.

The relevant regional data linkage units (CHeReL, Data Linkage Queensland and SA-NT DataLink) will facilitate access to the respective jurisdictional PDC and perform linkages between these data and the PFAS Management and Investigation Areas Address Database (see Figure 1).

All proposed linkages at this stage are subject to approvals from the relevant state linkage nodes, ethics committees and data custodians.

The data will be linked probabilistically based on relevant variables in the various datasets — including full name, sex, date of birth, and address/postcode. Importantly, a separation principle will be in place. Use of the separation principle ensures that no one working with the data will be able to view both the linking (identifying) information (such as name, address or date of birth) together with the merged analysis (content) data (such as clinical information, medical or pharmaceutical details) in an integrated dataset. More information on the separation principle is included in Appendix .

The AIHW will also provide secure storage of the data and make them available in de-identified form for analysis in the Sax Institute's Secure Unified Research Environment (SURE). More details on the data management during the linkage and analysis processes are available in section 9.1.

6.4 Measurement of exposure

The MEF and jurisdictional PDCs will be used to assemble the exposed and non-exposed populations. The *exposed group* is all people on the MEF or PDC who had an address matching any of those in the PFAS Investigation and Management Areas Address Database anytime between 1984[k] and 2018. The *non-exposed group* (or the comparison group) is a frequency-

[k] Or when mother's address was first collected as a data item in each state's PDC.

matched sample of people on the MEF or PDC who had an address with a postcode in comparison areas between 1984[11] and 2018.

The comparison areas will be selected by ANU researchers based on similarities to each of the PFAS Investigation and Management Areas, including state or territory, area-level socioeconomic status (SES) and remoteness. The ANU will select these areas using Australian Bureau of Statistics (ABS) Census data and correspondence files that map postcodes to standard SES (SEIFA) and remoteness (ARIA+) indices[I]. ANU researchers will supply a list of relevant postcodes to the AIHW and regional data linkage units from which the comparison group will be assembled. A contemporaneous comparison group will be frequency-matched to the exposed group on area-level SES, area-level remoteness and Indigenous status (access to the Indigenous status variable on the PDCs and the Voluntary Indigenous Indicator (VII) on the MEF are subject to approval). The comparison group will be sampled at a ratio of up to 10:1 (non-exposed: exposed), to be determined in conjunction with considerations of maximising study power and frequency of suitable individuals.

Apart from Williamtown, Oakey and Katherine, there are multiple other sites under investigation for environmental contamination with PFAS. The ANU will undertake an extensive search of public domain and government documents to identify potential sites of PFAS exposure across Australia and geographically code them at the postcode level. ANU researchers will supply this list to the AIHW in order to identify participants who have ever lived in these potentially exposed areas. We will censor participants at the start date of their first registered residence in these areas.

In this study, exposure will be treated as a binary variable (exposed/non-exposed) even though duration of exposure, not just fact of exposure, is of interest. However, because Medicare data are left-truncated— there are no data before 1984—and PFAS exposure has been noted since the 1970s, it is not possible to estimate duration of exposure for all individuals. This is particularly so for those who were already exposed at the start of the study period i.e. those who registered an address in a PFAS Investigation or Management Area at the inception of Medicare in 1984. Therefore, any health effects observed will be presented relative to the fact of exposure, not duration. However, if sufficient data, we will perform supplementary analyses limiting the exposed population to those with at least 10 years of exposure. Note that it is not possible to estimate duration of exposure for perinatal outcomes as jurisdictional PDCs do not collect historical addresses.

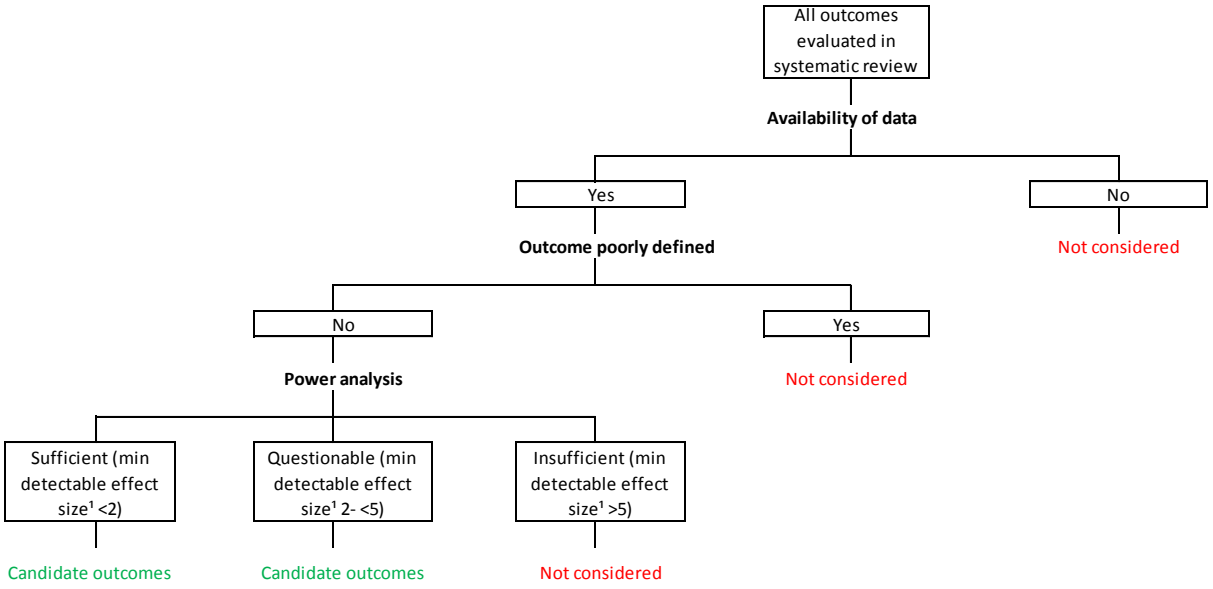
[I]The ABS correspondence files used are Postal Area 2017 to SEIFA 2016 and Postcode 2017 to Remoteness Area 2016. The indices used are the Index of Relative Socioeconomic Disadvantage (IRSD) from the Socio-Economic Indexes for Areas (SEIFA) suite of indexes for SES, and Accessibility and Remoteness Index of Australia (ARIA+) for remoteness. The IRSD is categorised into deciles and we will obtain postcodes within one decile of the exposed areas (and within the same ARIA+ score).

6.5 Measurement of health outcomes

I. Candidate outcomes

The selection of candidate outcomes for this study was based on the findings of the *PFAS Health Study Systematic Literature Review* conducted in Phase I of the study. [12] For each outcome that was evaluated in the systematic review, for our study we considered the availability of administrative data to allow appropriate measurement of the outcome, whether the outcome was meaningfully defined, and whether there would be sufficient statistical power[m] to detect a statistically significant difference in outcomes between the exposed and non-exposed groups. The power estimates are based on separate analyses for each of the three study areas; further details on the approach for the power calculations are provided in Section 9.2(ii). The selection criteria for candidate outcomes are shown below (Figure 2.)

Figure 2. PFAS Data Linkage Study: Selection criteria for binary/time-to-event candidate outcomes



¹ Effect size = relative measure of association (including prevalence/odds/risk/hazard ratio)

We have included several outcomes where *a priori* calculations showed the power to detect a difference in outcomes between the exposed and non-exposed group is questionable. We categorised binary and time-to-event outcomes as having 'questionable' power where the minimum detectable relative effect size (prevalence ratio (PR), odds ratio (OR), relative risk (RR) or hazard ratio (HR)) is between 2.0 and 5.0. For these outcomes, it is anticipated that

[m] The statistical power for a given outcome is a function of both an estimated baseline prevalence/incidence of the outcome in the general population and an estimated sample size of the population relevant for the outcome in the study.

there will be considerable uncertainty in the results, reflected in wide confidence intervals. Because of this, it will be necessary to interpret the results for these outcomes with caution, particularly where the point estimates of effect sizes are small.

For continuous outcomes, the minimum detectable difference between groups is expressed in terms of standard deviation. It is not meaningful to discretise this measure as it is only informative when the standard deviation of the outcome of interest is known. *A priori* calculations showed that the largest minimum detectable difference is 0.2 standard deviations[n]; we have included all continuous outcomes on this basis, but will individually consider the absolute units of difference during analysis of the data.

The proposed candidate outcomes are shown below in Table 1 (relevant codes in Appendix 1). This list is based on the results of power analysis in Katherine (see Table 4), which has the largest population size thus allowing for the maximally powered analysis for any given outcome. Some candidate outcomes included in this list will be excluded from analysis in the smaller areas of Williamtown and Oakey due to insufficient power to detect appropriate differences between groups. Additionally, in interpreting the results, we will take into account multiple testing (i.e. the probability of chance findings). The full list of all outcomes considered and their reasons for exclusion is shown in Attachment 2.

Note that for outcomes involving hospitalisations, we are primarily interested in quantifying time to first event i.e. the first hospitalisation with diagnosis of the health outcome after an individual's entry into the study. However, due to few data linkage studies reporting the rates of first hospitalisations for specific outcomes (except for acute myocardial infarction and stroke), we were unable to estimate an expected number of first hospitalisation events. [13] For hospitalisation outcomes besides acute myocardial infarction and stroke, power analysis was based on an expected total number of hospitalisations. Total hospitalisations potentially include multiple episodes per individual; the number of first hospitalisation events will be smaller and statistical power will be reduced. If there are too few first hospitalisation events in the data, we will quantify rates of total hospitalisations, where appropriate.

Note power calculations for all candidate outcomes are indicative only. Prior to analyses we will check event numbers to confirm sufficient statistical power before proceeding.

[n] A minimum detectable difference of 0.2 standard deviations represents a minimum detectable between-group difference of 2 units if the standard deviation of the outcome is 10.

Table 1. PFAS Data Linkage Study: Selected candidate outcomes[o]

Health outcome	Source	Outcome Definition[‡]; Outcome Type
Neonatal, infant and maternal outcomes		
Birthweight	PDC	Weight in grams; continuous
Small for gestational age (SGA)	PDC	Birthweight below 10th percentile for gestational age; binary
Large for gestational age (LGA)	PDC	Birthweight above 90th percentile for gestational age; binary
Birth length*	PDC	Length in cm; continuous
Head circumference at birth*	PDC	Length in cm; continuous
APGAR score at 5 minutes	PDC	Score; continuous
Preterm birth	PDC	Born before 37 weeks gestation; binary
Gestational age	PDC	Number of weeks; continuous
Stillbirth	PDC	Stillbirth; binary
Mode of delivery	PDC	Caesarean; binary Caesarean/assisted vaginal; binary
Delivery complications	PDC	Any complication; binary
Gender outcomes of pregnancy	PDC	Male; binary
Eclampsia*	PDC	Diagnosis of outcome; binary
Pregnancy induced hypertension	PDC	Diagnosis of outcome; binary
Congenital abnormalities	APDC	First hospitalisation with any congenital condition; binary
Congenital cryptorchidism	APDC	First hospitalisation with outcome; binary
Congenital hypospadias	APDC	First hospitalisation with outcome; binary
Kidney function		
Chronic kidney disease incidence	APDC	First hospitalisation with outcome; time to event
Chronic kidney disease mortality	NDI	Death with outcome; time to event
Liver function		
Liver disease incidence	APDC	First hospitalisation with outcome; time to event
Liver disease mortality	NDI	Death with outcome; time to event
Childhood neurodevelopment		
Physical health and wellbeing	AEDC	Vulnerability in domain; binary
Social competence	AEDC	Vulnerability in domain; binary

[o] All proposed candidate outcomes at this stage are subject to approvals from the relevant state linkage nodes, ethics committees and data custodians.

Health outcome	Source	Outcome Definition‡; Outcome Type
Emotional maturity	AEDC	Vulnerability in domain; binary
Language and cognitive skills	AEDC	Vulnerability in domain; binary
Communication skills and general knowledge	AEDC	Vulnerability in domain; binary
One or more of the above	AEDC	Vulnerability in domain; binary
Cancer		
Bladder cancer	ACD	Diagnosis of outcome; time to event
Kidney cancer	ACD	Diagnosis of outcome; time to event
Liver cancer	ACD	Diagnosis of outcome; time to event
Prostate cancer	ACD	Diagnosis of outcome; time to event
Pancreatic cancer	ACD	Diagnosis of outcome; time to event
Colorectal cancer	ACD	Diagnosis of outcome; time to event
Breast cancer	ACD	Diagnosis of outcome; time to event
Testicular cancer	ACD	Diagnosis of outcome; time to event
Thyroid cancer	ACD	Diagnosis of outcome; time to event
Oesophageal cancer	ACD	Diagnosis of outcome; time to event
Stomach cancer	ACD	Diagnosis of outcome; time to event
Laryngeal cancer	ACD	Diagnosis of outcome; time to event
Lung cancer	ACD	Diagnosis of outcome; time to event
Bone cancer	ACD	Diagnosis of outcome; time to event
Hodgkin lymphoma	ACD	Diagnosis of outcome; time to event
Non-Hodgkin lymphoma	ACD	Diagnosis of outcome; time to event
Leukaemia	ACD	Diagnosis of outcome; time to event
Brain cancer	ACD	Diagnosis of outcome; time to event
Head and neck cancer†	ACD	Diagnosis of outcome; time to event
Ovarian cancer†	ACD	Diagnosis of outcome; time to event
Uterine cancer†	ACD	Diagnosis of outcome; time to event
Diabetes		
Gestational diabetes	PDC	Diagnosis of outcome; binary
Cardiovascular effects		
Acute myocardial infarction (AMI)	APDC, NDI	First hospitalisation with outcome or death from outcome; time to event
Stroke	APDC, NDI	First hospitalisation with outcome or death from outcome; time to event
Major cardiovascular disease (CVD)	APDC, NDI	First hospitalisation with outcome or death from outcome; time to event
Bone fractures		

Health outcome	Source	Outcome Definition‡; Outcome Type
Hip fracture	APDC	First hospitalisation with outcome; time to event
Respiratory outcomes		
Chronic obstructive pulmonary disease (COPD)	APDC	First hospitalisation with outcome; time to event

*These health outcomes are not available for all relevant states (NSW, Qld and NT)

†These health outcomes were included due to community interest

‡All relevant ICD/ICD-10-AM/ACHI codes used to define outcomes are available in Appendix 1

II. Control outcomes

The control outcomes for this study were selected on the basis of having no known evidence of association to PFAS (Table 2).

Table 2. PFAS Data Linkage Study: Control outcomes

Health outcome	Source	Outcome Definition‡; Outcome Type
Diseases of the nervous system	APDC	First hospitalisation with outcome; time to event
Diseases of the skin and subcutaneous tissue	APDC	First hospitalisation with outcome; time to event
Land transport accidents mortality	NDI	Death with outcome; time to event
Accidental falls mortality	NDI	Death with outcome; time to event

‡All relevant ICD/ICD-10-AM/ACHI codes used to define outcomes are available in Appendix 1

III. Potential confounders

Age, sex, Indigenous status (subject to approval) and calendar year will be included as covariates in the analysis. Age and calendar year will be treated as grouped categorical variables with 5-year brackets, given sufficient sample size in each stratum. Separate variables for SES and remoteness categories will not be created as these will be controlled through matching of comparison areas with the exposed areas based on area-level SES (SEIFA) and remoteness characteristics (ARIA+).

For perinatal outcomes, the following variables will be used for adjustments in the analysis, where relevant: maternal age at delivery, parity, maternal weight, maternal morbid obesity, maternal country of birth, smoking during pregnancy, maternal pre-existing diabetes and/or hypertension, SES (derived from postcode of usual residence), marital status, gestational age, and baby's sex.

7 Safety considerations

The study will use routinely-collected administrative data and will not involve direct contact with any participants. Additional information on privacy and ethical considerations is included in section 15.

8 Follow-up

This study will not include prospective participant follow-up.

9 Data management and statistical analysis

9.1 Data management

I. Data storage and handling

All data linked to the MEF will be stored, accessed and analysed in the Secure Unified Research Environment (SURE) computing environment through the Sax Institute. SURE will be accessed via the AARNET (Australian Academic and Research Network) or the internet using an encrypted connection from researchers' local computers, which must meet security requirements. Descriptive data and analysis results only will be downloaded from SURE under curator surveillance and stored on secure, password-protected networks at the ANU. Data from jurisdictional PDCs will be stored as per the jurisdictional requirements (either through a secure virtual laboratory such as SURE or on ANU secure servers). Only approved members of the PFAS Health Study team will have access to the data.

II. Data analysis

All statistical analyses will be performed using Stata version 15 (StataCorp) and/or SAS version 9.4.

III. Data verification

Data checking for incorrect/unusual values and outliers, investigation of missing values, assessment of distributions and exploratory data analysis will be undertaken for all variables.

Linked data will be verified for consistency of similarly defined variables for a given entity across datasets, for example sex. We will also verify the data for temporal consistencies for a given entity during the follow-up period, for example:

- Date of birth must strictly precede the date of entry into the study
- Date of diagnosis and date of hospitalisation(s) must strictly follow date of birth
- Date of death must strictly follow all other dates including date of birth, date of entry into the study, date of diagnoses and date of hospitalisation(s)

IV. Missing data

Individuals will be excluded from the study population if they have missing data for sex or date of birth. Further exclusions may occur for outcome-specific analyses, for example where there is missing data on date of diagnosis, date of admission (hospitalisation outcomes) or date of death, such that time to event cannot be calculated or censoring cannot be performed.

9.2 Statistical analysis

I. Analysis of outcomes

A range of different statistical approaches will be used to compare candidate outcomes in the exposed and non-exposed groups, depending on the type of outcome. Three types of outcomes will be considered in this study:

1. Continuous outcomes
2. Binary outcomes
3. Time-to-event outcomes

All outcomes will be analysed separately for each of the three study areas: Williamtown, Oakey and Katherine. For some outcomes where there is inadequate statistical power for area-specific analysis, we will perform a combined analysis with adjustment for area. This will be reviewed on a case-by-case basis for appropriateness and statistical validity. Outcomes will also be analysed separately by sex where sample sizes allow, otherwise adjustment will be made for sex in the models. Where appropriate, adjustments will be made for age and calendar period, and where possible Indigenous status. All results will be reported as point estimates with 95% confidence intervals. The choice of model for each outcome, as outlined below, assumes model assumptions are met.

Continuous outcomes

For continuous outcomes, means will be estimated for the exposed and non-exposed groups and linear regression used to estimate the difference in means between the exposed and non-exposed groups.

Binary outcomes

Prevalence will be calculated for all binary outcomes as the total number of events divided by the total number of persons in the relevant population. For example, the number of births where neonates were defined as small for their gestational age, divided by the total number of births. Regression methods will be used to estimate relative effect sizes, expressed as either prevalence ratios (PR) (using modified Poisson regression) or odds ratios (OR) (using logistic regression).

Time-to-event outcomes (rates)

Crude rates will be calculated as the total number of events divided by total person-time at risk. Where there are large numbers of events, regression methods will be used to estimate relative effect sizes, expressed as either relative rates (RR) (using Poisson or negative binomial regression) or hazards ratios (HR) (using Cox regression).

Standardised incidence ratios (SIR) will be calculated for rare time-to-event outcomes (low number of events), as the total number of observed cases in the exposed group divided by the total number of expected cases in the exposed group. The SIR for each outcome will be estimated using indirect standardisation. To do this, age-sex-calendar period specific rates will be calculated for each outcome in the non-exposed group. These rates will then be applied to the exposed group to generate the number of expected cases per outcome.

For all relative effects (PRs, ORs, RRs, HRs, SIRs): interpretation is as follows:

- effect size >1 means that the prevalence/rate of the outcome is higher in the exposed group than the non-exposed group
- effect size <1 means that the prevalence/rate of the outcome is lower in the exposed group than the non-exposed group
- effect size = 1 means that there is no difference in prevalence/rate of the outcome between the exposed group and the non-exposed group

Total person-years will be calculated for all rate-based measures as the sum of individual person-years at risk, which is the time from entry into the study (with or without a lag applied) until either:

a) The date of diagnosis or hospitalisation for the outcome;
b) The date of first registered residence in other exposed areas apart from Williamtown, Oakey and Katherine (see section 6.4);
c) Death from any cause;
d) The individual's age is 85 years old; or
e) The date of 31 December 2018 or last available data.

An individual's entry into the study will be defined as the start date of their first Medicare registration, regardless in which state or territory of Australia they were registered. All person-years at risk will be classified as exposed or non-exposed.

- *For individuals who have never lived in a PFAS Investigation and Management Area*, all person-years will be classified as non-exposed.
- *For individuals who have ever lived in a PFAS Investigation and Management Area*, any person-years before the start date of their first registered residence in a PFAS Investigation and Management Area will be classified as non-exposed (after taking

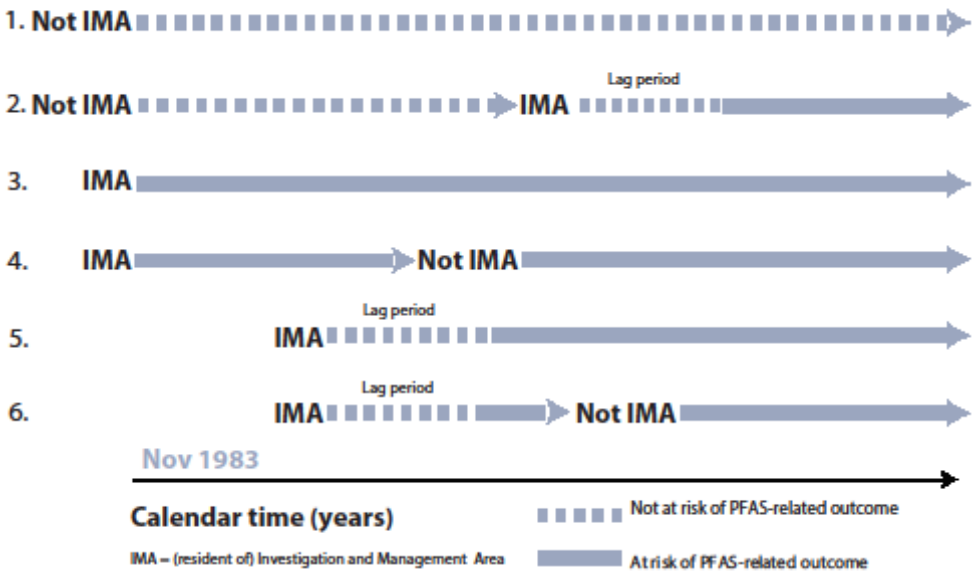
into account lag periods, defined below). The remainder of person-years after this will be classified as exposed, regardless of whether the individual moved out of the Investigation Area thereafter.

By default, all person-time associated with post office box addresses (except in Katherine) will be classed as non-exposed, even though the true residence of the individual in that time period, and thus his or her exposure status, is unknown. If there is a substantial proportion of post box office addresses, we will conduct sensitivity analysis where all participants with a post office box address will be excluded from the analysis. More information is available in section 9.2(ii). Note that the whole of Katherine is classified as PFAS Investigation and Management Area, therefore we will not apply the exclusion rule to Katherine i.e. we will consider all participants as exposed even if they had at least one post office box address.

Lag periods: There is an expected delay between an individual's first exposure to PFAS and a diagnosis/hospitalisation for some outcomes potentially attributable to PFAS exposure. Therefore, statistical analyses will include lag periods to account for minimum plausible latency periods. A 'lag period' is a specified amount of time applied from the date of the individual's first exposure during which all events (outcomes) and person-years will be classified as non-exposed. This means that outcomes are not attributable to PFAS exposure until after the lag period has passed.

As there is insufficient evidence for lag periods, we propose adopting a 10-year lag period for all outcomes except for neonatal, infant and maternal outcomes and childhood neurodevelopment outcomes. Note that applying a lag period will reduce statistical power. We will vary the lag period in sensitivity analyses. More information is available in section 9.2(ii). Figure 3 describes the attribution of person-years with application of a lag, under different scenarios of household movement.

Figure 3. PFAS Data Linkage Study: *Attribution of person-years, with application of lag period*



If an individual was not living in a PFAS Investigation and Management Area initially, but later moved into one, outcomes that occur immediately after the move will not be attributable to PFAS until after the lag period has passed (Scenario 2). We propose that no lag periods will apply if an individual was already living in a PFAS Investigation and Management Area at the start of the study (i.e. those whose first Medicare registration was at an address in a PFAS Investigation and Management Area and the registration occurred before 1 January 1985), as we will assume that they have been living in the same place for at least as long as the lag period (Scenarios 3 and 4). For individuals who lived in a PFAS Investigation and Management Area at any time after the start of the study, outcomes will not be attributable to PFAS until after the lag period has passed and thereafter (Scenario 5), including if the individual moves away from the PFAS area (Scenario 6).

II. Sensitivity analyses

Variation in lag periods

We will vary the lag period from 10 years to 0 years, 5 years and 15 years for all outcomes where appropriate, except for neonatal, infant and maternal outcomes and childhood neurodevelopment outcomes.

Exclusion of participants with post office box addresses

We will exclude all participants who had at least one post office box address from the study population. This is because we do not know the true exposure status for participants whilst being registered with a post office box address. However, we will not exclude participants who have already been classified as exposed at the time of their first registration with a post office box address.

Censoring

We will censor all participants at age 100 years instead of 85 years old.

III. Power analysis

Sample size estimations

The exposed group comprises everyone who has ever had an address on the MEF or PDCs in the PFAS Investigation and Management Areas. Therefore, the sample size for the exposed group is fixed. The number of individuals in the non-exposed group will be limited only by the number of individuals that can be frequency-matched to participants in the exposed group. We will sample non-exposed participants up to a maximum ratio of 10:1 (non-exposed: exposed), to be determined in conjunction with considerations of maximising study power and frequency of suitable individuals. Generally, there is limited benefit to statistical power beyond a ratio of 4:1 (non-exposed: exposed). However, since the PFAS Data Linkage Study involves the use of linked administrative data rather than primary data collection and recruitment of participants, the marginal cost of increasing the size of the non-exposed groups is minimal, therefore the power calculations were undertaken for ratios of non-exposed participants to exposed participants ranging from 1:1 – 10:1.

The estimated number of individuals living in each PFAS Investigation and Management Area at any point in time is shown in Table 3. The estimated sample size of the total exposed group across the three study areas is 45,000, after multiplying the number of individuals by an estimated mobility factor of 3.4 (based on a previous data linkage study using similar methods). [14] The estimated total sample size including the non-exposed group sampled at 4:1 is 45,000 + 180,000 = 225,000.

Table 3. Estimated number of individuals in PFAS Investigation and Management Areas

	Williamtown	Oakey	Katherine
Number of households	757	1,755	2,625
Average number of individuals per household	2	2.6	2.8
Number of individuals (at any point in time)	1,514	4,563	7,350

Power calculations

We completed a series of power calculations given ratios of non-exposed participants to exposed participants ranging from 1:1–10:1. Estimations were calculated for a significance level of 5% for 80% power and 90% power. To account for correlation of outcomes for individuals living in the same household and for more than one pregnancy for women, estimates were obtained assuming independent observations and allowing for a design effect of 1.1.

For all power calculations we made the following assumptions:

- An estimated 30 years of data[p] are available for neonatal, maternal and birth outcomes
- An estimated 4 years of data are available for childhood neurodevelopmental outcomes
- An estimated 15 years of average follow-up time per entity for cancer and cause-specific death outcomes
- An estimated 7 years of average follow-up time per entity for hospitalisation outcomes
- An estimated mobility factor of 3.4. The population of interest at a particular point was multiplied by this factor in power calculations to allow for movement in and out of the areas over the study period

The following table lists the minimum detectable effect size for binary (PR/OR) and time-to-event (HR) candidate outcomes in the Katherine PFAS Investigation and Management Area, assuming 5% significance and 80% power in the study. The power to detect significant differences will be lower for Oakey and Williamtown due to their smaller populations and fewer years of perinatal data.

Table 4. Estimated minimum detectable effect sizes for candidate outcomes in Katherine

		Minimum detectable OR/HR*	Minimum detectable OR/HR**
Binary Outcome	**Estimated Prevalence (%)**		
Neonatal, infant and maternal outcomes			
Small for gestational age	10	1.3	1.2
Large for gestational age	10	1.3	1.2
Preterm birth	10	1.3	1.2
Stillbirth	0.7–0.9	2–2.5	1.7–2
Mode of delivery (caesarean)	30–40	1.2	1.15
Delivery complications	40–50	1.2	1.15
Gender outcomes of pregnancy (male)	52–67	1.2	1.15
Eclampsia	0.10	5.5	3.7
Pregnancy induced hypertension	10	1.3	1.2

[p] This is a conservative estimate based on the data collection period of the NT PDC. The collection periods of NSW and Qld PDCs are much shorter and therefore statistical power will be reduced in these study areas.

		Minimum detectable OR/HR*	Minimum detectable OR/HR**
Congenital abnormalities	3–4	1.4–1.5	1.3–1.4
		Minimum detectable OR/HR*	**Minimum detectable OR/HR****
Binary Outcome	**Estimated Prevalence (%)**		
Congenital cryptorchidism	No estimates	NA	NA
Congenital hypospadias	1.5–2	1.5–2	1.4–1.7
Neurodevelopmental outcomes			
Physical health and wellbeing	8.9	1.9	1.7
Social competence	8.7	1.9	1.7
Emotional maturity	6.9	2	1.8
Language and cognitive skills	5.6	2.2	1.85
Communication skills and general knowledge	8.8	1.9	1.7
One or more of the above	20.7	1.6	1.5
Diabetes			
Gestational diabetes	13	1.25	1.2
Time-to-event Outcome	**Estimated rate per 100,000**		
Kidney function			
Chronic kidney disease	1500–4000	1.03–1.06	1.03–1.05
Chronic kidney disease mortality	30–150	1.2–1.5	1.2–1.4
Liver function			
Liver disease incidence	No estimates	NA	NA
Liver disease mortality	5–10	1.5–2.5	1.3–1.8
Cancers			
Bladder cancer	10–30	2–1.5	1.7–1.3
Kidney cancer	15–30	2–1.5	1.7–1.3
Liver cancer	10–20	2–1.5	1.7–1.3
Prostate cancer	200–300	1.4	1.3
Pancreatic cancer	15–30	2–1.5	1.7–1.3
Colorectal cancer	70–140	1.3	1.2
Breast cancer	150–300	1.4	1.3
Testicular cancer	6–8	3	2.5

		Minimum detectable OR/HR*	Minimum detectable OR/HR**
Thyroid cancer	18	2–1.5	1.7–1.3
Oesophageal cancer	5–15	2–3	1.5–2

Time-to-event Outcome	Estimated rate per 100,000	Minimum detectable OR/HR*	Minimum detectable OR/HR**
Stomach cancer	10–25	1.5–2	1.5–1.7
Laryngeal cancer	2–5	3–4	2–3
Lung cancer	50–120	1.2–1.4	1.2–1.3
Bone cancer	1	5.5	3.5
Hodgkin lymphoma	2	4–5	2.5–3.5
Non-Hodgkin lymphoma	20–40	1.5–2	1.5
Leukaemia	10–30	1.5–2	1.4–1.7
Brain cancer	5–15	2–3	1.5–2
Head and neck	20–40	1.5–2	1.5
Ovarian	10–25	2–2.5	1.5–2
Uterine	25–50	1.5–2	1.5–1.7
Cardiovascular disease			
AMI	150–300	1.3	1.25
Stroke	500–800	1.2	1.1-1.2
Major CVD	700–1200	1.2	1.12
Bone fracture			
Hip fracture	190	1.6	1.5
Respiratory outcomes			
COPD	730	1.1	1.1

* (unexposed: exposed ratio of 1:1)

** (unexposed: exposed ratio of 4:1)

10 Quality assurance

The ANU is proactive and responsible in its approach to risk management. The Research Office within the College of Health and Medicine oversees all population health research within ANU. The Research Office oversees the application of research proposals and financial accountability for the conduct of research. The ANU has human and animal research ethics committees that function in accordance with National Health and Medical Research guidance. The Research Office ensures that all funded research is approved by the appropriate ethics committee and complies with University policies.

The study will have ethics approval from the Human Research Ethics Committee (HREC) at the ANU and all other relevant committees. These Committees will ensure that research is conducted according to the National Health and Medical Research Council's National Statement on Ethical Conduct in Human Research.

11 Expected outcomes of the study

We expect to generate findings regarding health effects of living in PFAS Investigation and Management Areas of Williamtown, Oakey and Katherine This information will assist community members in understanding the impact of living in a PFAS Investigation and Management Area on their health and assist policy makers in responding to the ongoing issue of PFAS contamination. This study is the first of its kind in Australia and will add to the body of international literature around the effects of environmental PFAS exposure on human health.

Dissemination of results and publication policy

The findings from this data linkage study will be detailed in a report provided to the Australian Government Department of Health. The study team will also prepare articles for publication in peer-reviewed journals. All reports and publications will acknowledge funding from the Department of Health and input from the community and other experts. Authorship of peer reviewed articles will be determined in accordance with the Uniform Requirements for Manuscripts Submitted to Biomedical Journals produced and endorsed by the International Committee of Medical Journal Editors, specifically those considerations set out under heading IIA 'Ethical Considerations in the Conduct and Reporting of Research: Authorship and Contributorship'.

The final report will be made publicly available via the study webpage. In addition to the final report, study results will be summarised in a format suitable for lay-people. This will be made available on the study website and sent via email to study participants who have indicated that they would like to receive a copy and have provided their email address on the survey

form. The form and timing of the communication will be determined by the Study team in consultation with the Department of Health. Additionally, after the acceptance of the report by the Department of Health, the study team will hold face-to-face community consultations in each of the three communities, including Aboriginal communities where appropriate, to present the findings.

Findings will be communicated to national and international media through a coordinated media release between the Australian National University and the Australian Government Department of Health.

Duration of the project

The PFAS Data Linkage Study is expected to be completed by December 2020.

12 Project management

The PFAS Health Study team includes seven co-investigators, six associate investigators and six research officers. Co-investigators and associate investigators provide expertise in the fields of data linkage, biostatistics, epidemiology, PFAS toxicology, medicine and Aboriginal community engagement. Experts will oversee the execution and analysis of the research, while research officers will manage daily administration of the project.

13 Ethics

13.1 Ethics applications

The PFAS Data Linkage Study involves linkage between Commonwealth data collections and between Commonwealth and state and territory government data. The MEF, NDI, ACD and AEDC are AIHW-held data collections, however state and territory governments retain ownership of their jurisdiction's data for the ACD. The APDC are held by individual states and territories. Access to the MEF for data linkage purposes also requires a Public Interest Certificate signed by the Minister for Health of the Commonwealth, which will be initiated by the AIHW at the time of ethics approval. A detailed description of this process is included in Appendix 2.

For perinatal outcomes, the study involves single-jurisdictional data linkages. For these linkages, the regional data linkage units will collaborate on this project through the Population Health Research Network (PHRN). The PHRN will formally review and coordinate this process. A letter of feasibility and conditional approval from data custodians will be required for each

jurisdiction. Prior to final approval from the data custodians, ethics approval is required for all jurisdictions and institutions involved in the research.

Additional ethical consideration must be given with the inclusion of Aboriginal persons in the study. We have addressed six core values with reference to how the Aboriginal population will benefit or be impacted by the study:

a. Reciprocity

Although this is a data linkage study and there is no direct involvement of study participants, relevant administrative data of individuals living in in PFAS Investigation and Management areas will be used, including Aboriginal and Torres Strait Islander people. The results of the data linkage analysis aim to provide the Katherine community with a clearer understanding of the health impact of PFAS exposure. Special attention was given to involvement and participation of the Aboriginal community in Katherine, Northern Territory, due to the large percentage of Aboriginal people living in the PFAS Investigation and Management Area, The research team is actively working with an Aboriginal Elder, an academic consultant in Aboriginal engagement (Professor Adrian Miller) and with the local Aboriginal community controlled health organisation (ACCHO) of Katherine (Wurli-Wurlinjang Health Services) to provide advice on Aboriginal engagement. The research team will return to Katherine to present the outcomes of the data linkage analysis to the Katherine community and they will also provide an opportunity to the community to discuss any health issues associated with the exposure. Wurli-Wurlinjang Health Services and our Aboriginal advisors will be consulted regarding how best to deliver the results of the data linkage analysis so they are easily understandable and usable in the community after the conclusion of the research.

b. Respect

The PFAS Health Study team aims to develop respectful and sustainable relationships with the local Aboriginal communities. Researchers will ensure the above mentioned Aboriginal representatives and consultants are consulted before any material is published and made available to the public. Additionally, all published material will acknowledge Aboriginal communities and their contribution to the research. We understand and acknowledge that consent for participation is important for members of the Aboriginal population. We have consulted with the Wurli-Wurlinjang Health Services, the Aboriginal Elder representing the community and the Aboriginal consultant and they provided support for the overall study, including the data-linkage study where a waiver of consent will be sought. This information has also been provided to the Aboriginal communities in Katherine.

c. Equity

The chief investigator and a co-investigator presented the overall study design to the Board, CEO and Director of Medical Services at Wurli-Wurlinjang Health Service. The data linkage study does not require direct involvement of study participants, this includes both individuals living in contaminated areas and individuals living in the control areas. The PFAS Health Study team will also ensure the research findings are presented to community in a manner that can be easily understood and usable in the community (e.g., posters and face-to-face consultations).

d. Responsibility

No direct involvement is required for participation in the data linkage study as only relevant de-identified data will be used for analysis. There is a minor risk of identification of Aboriginal participants, however all data will be presented in aggregated format to mitigate this risk. If cell sizes are <5, these data will not be reported in order to prevent identification of the individual involved. This slight risk is considered to be outweighed by the public benefit and the benefit for the Aboriginal and Torres Strait Island communities.

e. Cultural continuity

In collaboration with Professor Adrian Miller, the Aboriginal Elder and the Wurli-Wurlinjang Health Service, the PFAS Health Study team hopes to establish trust and credibility within the Katherine community. From a generational perspective PFAS contamination in Katherine, occurred in the past, but is now a concern in the present and future. The PFAS Health Study team acknowledges the impact PFAS might have on local Aboriginal communities, including their ongoing connection to country and cultural activities.

f. Spirit & Integrity

The research team will be guided by the Wurli-Wurlinjang Health Service, Professor Adrian Miller and the Aboriginal Elder on how to best inform the Aboriginal community of possible involuntary participation.

We are requesting Indigenous status as a variable in the data linkage study; this information will only be used in final analysis to adjust for possible confounding. There will be no sub-group analyses i.e. no comparisons will be made between Indigenous and non-Indigenous groups.

The protocol for the PFAS Data Linkage Study will be submitted to the following ethics committees:

1. ACT Health HREC
2. AIHW Ethics Committee

3. ANU HREC
4. Darling Downs Hospital and Health Service Committee
5. The Departments of Defence and Veterans' Affairs (DDVA) HREC
6. NSW Population and Health Services Ethics Committee
7. Aboriginal Health and Medical Research Council of New South Wales Ethics Committee
8. HREC of NTDoHMSHR
9. South Australia (SA) Health HREC
10. Human Research Ethics Committee (Tasmania) Network
11. Victoria Department of Health and Human Services HREC
12. WA Department of Health HREC
13. WA Aboriginal Health Ethics Committee

13.2 Participant consent

The study is compliant with all Australian Privacy Principles (APP) except APP6 (use or disclosure of personal information). As this project is to be conducted without consent, which would breach APP6, a waiver of consent pursuant to section 95 of the Privacy Act 1988 will be sought on the basis of the large number of people involved, and the high degree of privacy protection afforded by application of the separation principle and additional measures relating to data access and use that are aimed at minimisation of linked data records' re-identification risk.

13.3 Secure data management

The study team will adhere to strict guidelines to ensure security of data. Individual-level identifying information, required to link data, will be performed by the AIHW, a Commonwealth-accredited data integration authority, and regional data linkage units (CHeReL, Data Linkage Queensland and ST-NT DataLink). Details on the accreditation criteria, which have been fulfilled by the AIHW, can be found on the Australian Government National Statistical Service website:

http://nss.gov.au/nss/home.nsf/pages/Data%20Integration%20-%20AIHW%20accreditation%20application%20and%20audit%20summary/#CIV

Only de-identified data will be available to the researchers at the ANU who will be carrying out the analyses.

14 References

1. Kissa E: **Fluorinated surfactants and repellents**, Second edn. New York: CRC Press; 2001.
2. AECOM Services Pty Ltd: **Preliminary ecological risk assessment December 2016 - RAAF Base Williamtown Stage 2B environmental investigation**. In. Edited by Department of Defence; 2016: 1.
3. AECOM Services Pty Ltd: **Stage 2C environmental investigation - Preliminary ecological risk assessment, Army Aviation Centre Oakey**. In. Edited by Department of Defence; 2016: 2.
4. Coffey Services Australia Pty Ltd: **Detailed Site Investigation Report RAAF Base Tindal**. In. Edited by Department of Defence. Victoria, Australia; 2018: 164.
5. Coffey Services Australia Pty Ltd: **RAAF Base Tindal Human Health Risk Assessment**. In. Edited by Defence Do. Abbortsford, Victoria, Australia: Coffey Services Australia Pty Ltd; 2018: 152.
6. AECOM Services Pty Ltd: **Stage 2C Environmental Investigation - Human Health Risk Assessment, Army Aviation Centre oakey**. In. Edited by Defence Do. Fortitude Valley, Queensland, Australia: AECOM Australia; 2016: 2020.
7. AECOM Services Pty Ltd: **Off-site Human Health Risk Assessment July 2016, RAAF Base Williamtown, Williamtown, NSW**. In. Edited by Defence Do. Sydney, NSW, Australia: AECOM Services Oty Ltd; 2016: 4191.
8. **PFAS Investigation and Management Program** [http://www.defence.gov.au/Environment/PFAS/]
9. Li K, Gao P, Xiang P, Zhang X, Cui X, Ma LQ: **Molecular mechanisms of PFOA-induced toxicity in animals and humans: Implications for health risks**. *Environment international* 2017, **99**:43-54.
10. Coperchini F, Awwad O, Rotondi M, Santini F, Imbriani M, Chiovato L: **Thyroid disruption by perfluorooctane sulfonate (PFOS) and perfluorooctanoate (PFOA)**. *Journal of Endocrinological Investigation* 2017, **40**(2):105-121.
11. Joensen UN, Bossi R, Leffers H, Jensen AA, Skakkebaek NE, Jorgensen N: **Do perfluoroalkyl compounds impair human semen quality?** *Environmental health perspectives* 2009, **117**(6):923-927.
12. Kirk M, Smurthwaite K, Bräunig J, Trevenar S, D'Este C, Lucas R, Lal A, Korda R, Clements A, Mueller J *et al*: **The PFAS Health Study: Systematic literature review**. In. Canberra: The Australian National University; 2018.
13. Sarink D, Nedkoff L, Briffa T, Shaw JE, Magliano DJ, Stevenson C, Mannan H, Knuiman M, Hung J, Hankey GJ *et al*: **Trends in age- and sex-specific prevalence and incidence of cardiovascular disease in Western Australia**. *European Journal of Preventive Cardiology* 2018, **25**(12):1280-1290.
14. J Korda R, S Clements M, K Armstrong B, Di Law H, Guiver T, R Anderson P, M Trevenar S, Kirk M: **Risk of cancer associated with residential exposure to asbestos insulation: a whole-population cohort study**, vol. 2; 2017.
15. Centre for Epidemiology and Evidence: **Quality and coverage of the NSW Register of Congenital Conditions using Admitted Patient Data: A record linkage study**. In. Sydney: NSW Ministry of Health; 2016.

15 Appendices and attachments

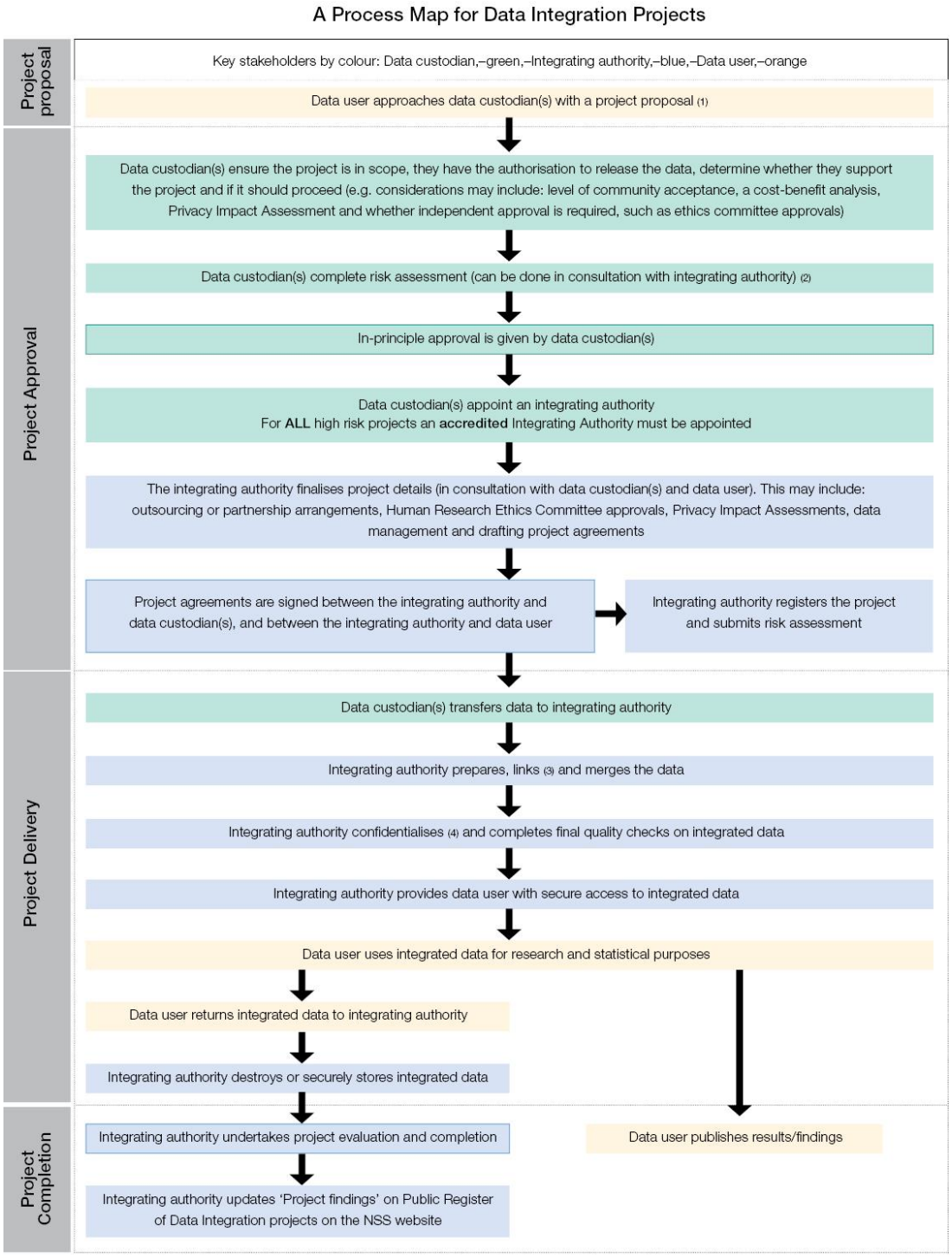
15.1 Appendix 1 – ICD/Procedure codes

Health outcome	Data	ICD-10, ICD-10-AM or ACHI (procedure) codes
Neonatal, infant and maternal outcomes		
Congenital abnormalities	APDC	Primary or additional diagnosis: D18.1, D56.0, D56.1, D82.1, E03.0, E03.1, E70.0, E70.1, E84, E84.0, E84.1, E84.8, E84.9, P83.2, all Q codes (excluding those listed in Appendix 3 of the Quality and coverage of the NSW Register of Congenital Conditions report)[15] Procedure codes: 37803-00, 37803-01, 49718-01, 49724-00, 49724-01, 49727-00, 50321-00, 50324-00, 50324-01, 50327-00
Congenital cryptorchidism	APDC	Procedure codes: 37803-00, 37803-01
Congenital hypospadias	APDC	Primary or additional diagnosis: Q54.0, Q54.1, Q54.2, Q54.3, Q54.8, Q54.9
Kidney function		
Chronic kidney disease incidence	APDC	Primary or additional diagnosis: E10.2, E11.2, E13.2, E14.2, I12, I13, I15.0, I15.1, N00–N08, N11, N12, N14, N15, N16, N18, N19, N25–N28, N391, N392, Q60–Q63, T82.4, T86.1, Z49.0, Z94.0, Z99.2
Chronic kidney disease mortality	NDI	Underlying or associated cause: E10.2, E11.2, E12.2, E13.2, E14.2, I12, I13, I15.0, I15.1, N00–N07, N11, N12, N14, N15, N18, N19, N25–N28, N39.1, N39.2, E85.1*, D59.3*, B52.0*, Q60–Q63, T82.4,T86.1
Liver function		
Liver disease incidence		Primary or additional diagnosis: K70–K76
Liver disease mortality		Underlying or associated cause of death: K70–K76
Cancer		
Bladder cancer	ACD	Site/type of cancer: C67
Kidney cancer	ACD	Site/type of cancer: C64

Health outcome	Data	ICD-10, ICD-10-AM or ACHI (procedure) codes
Liver cancer	ACD	Site/type of cancer: C22
Prostate cancer	ACD	Site/type of cancer: C61
Pancreatic cancer	ACD	Site/type of cancer: C25
Colorectal cancer	ACD	Site/type of cancer: C18-C20
Breast cancer	ACD	Site/type of cancer: C50
Testicular cancer	ACD	Site/type of cancer: C62
Thyroid cancer	ACD	Site/type of cancer: C73
Oesophageal cancer	ACD	Site/type of cancer: C15
Stomach cancer	ACD	Site/type of cancer: C16
Laryngeal cancer	ACD	Site/type of cancer: C32
Lung cancer	ACD	Site/type of cancer: C33-C34
Bone cancer	ACD	Site/type of cancer: C40-41
Hodgkin lymphoma	ACD	Site/type of cancer: C81
Non-Hodgkin lymphoma	ACD	Site/type of cancer: C82-C86
Leukaemia	ACD	Site/type of cancer: C91-C95
Brain cancer	ACD	Site/type of cancer: C71
Head and neck	ACD	Site/type of cancer: C00-C14, C30-C32
Ovarian	ACD	Site/type of cancer: C56
Uterine	ACD	Site/type of cancer: C54-C55
Cardiovascular effects		
Acute myocardial infarction (AMI)	APDC/ NDI	Primary diagnosis: I21
Stroke	APDC/ NDI	Primary diagnosis: I60–I64
Major cardiovascular disease (CVD)	APDC/ NDI	Primary diagnosis: I11–I13, I20–I25, I26–I28, I34–I36, I42, I44, I46-I51, I61–I67, I69, I70–I77, I80, G45, G46
Bone fractures		
Hip fracture	APDC	Primary diagnosis: S72.0, S72.1, S72.2
Respiratory outcomes		
Chronic obstructive pulmonary disease (COPD)	APDC	Primary diagnosis: J40-J44
Control outcomes		
Diseases of the nervous system	APDC	Primary diagnosis: G00-G99
Diseases of the skin and subcutaneous tissue	APDC	Primary diagnosis: M00-M99

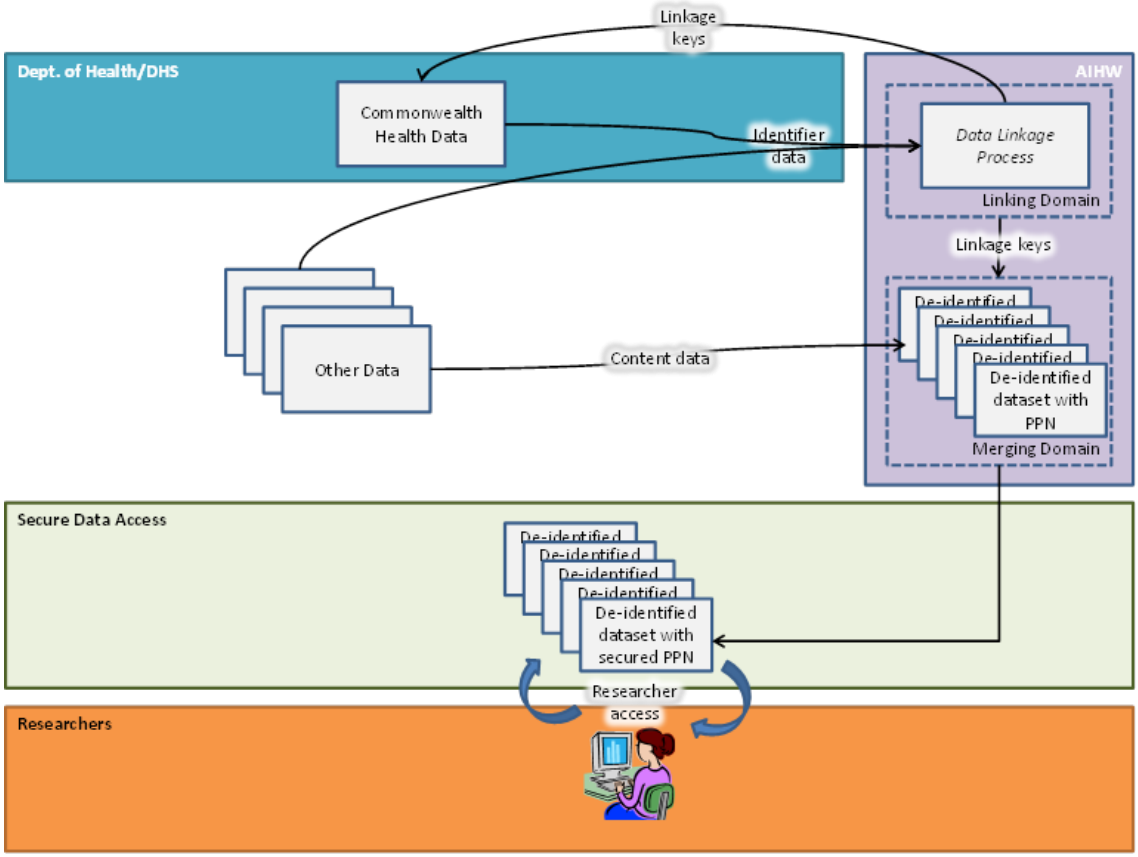
Health outcome	Data	ICD-10, ICD-10-AM or ACHI (procedure) codes
Land transport accidents mortality	NDI	Underlying cause of death: V01-V89
Accidental falls mortality	NDI	Underlying cause of death: W00-W19

15.2 Appendix 2 – Data linkage process



Source: National Statistical Service: http://statistical-data-integration.govspace.gov.au

15.3 Appendix 3 – Data separation process



Source: Australian Institute of Health and Welfare: http://www.aihw.gov.au