Government of **Western Australia**
Department of **Health**

For Office Use Only	
DL#	201704.02
HREC#	
Version	V.1
Date	17/04/2018

APPLICATION FOR DATA

This form should be used for *all* requests for unit record data, linked data, and use – but not necessarily release – of Department of Health data collections, where Department of Health Human Research Ethics Committee (HREC) approval is required.

Double click appropriate box and select 'checked' option

REQUEST DETAILS

Data Linkage Branch services	**Other application for data**
☒ includes linked and geocoded data and sample selections (see Data Linkage WA)	☐ includes requests for unlinked data

SUBMISSION DETAILS

☐ **Draft application**

This step must be completed before applying to DoH HREC.

Email dataservices@health.wa.gov.au and attach Word versions of:

- the "Application for Data" form*
- data services forms (e.g. extraction, linkage, geocoding, family connections)
- variable lists for all datasets requested
- research protocol
- other supporting documentation

* no signatures or confidentiality agreements required for draft applications

☒ **Data application (final copy) for DoH HREC approval**

Application for Data form to accompany other HREC application forms. For submission instructions refer to the DoH HREC website.

☐ **Data application (final copy) *not* requiring DoH HREC approval**

Email dataservices@health.wa.gov.au and attach PDF versions of:

- signed Application for Data form
- data Services forms (e.g. cohort specifications, linkage, geocoding)
- variable lists for all datasets requested
- research protocol
- other supporting documentation.

☐ **Application for data amendment/update for an existing project**

Projects for which the most recent approval was granted more than a year ago, must complete this form in conjunction with the relevant amendment form.

1. Project title

Heritable and environmental determinants of hospitalisation for common childhood illnesses - Study 2: association between monozygotic and dizygotic twins, their sibs and their parents in hospital admissions.

2. Contact details

2.1 Principal Investigator

The principal investigator is the person with overall responsibility for management of the project; He/she must *not* be a student. The principal investigator must read and sign the legal declarations at the end of this form.

Title and name	Professor Nick de Klerk			
Position	Head of Biostatistics			
Organisation	Telethon Kids Institute			
Address	100 Roberts Road			
	Subiaco, WA 6008			
Phone	(w)	(08) 9489 7735	(m):	0400110814
Email	Nick.deKlerk@telethonkids.org.au			

2.2 Project Contact

The contact person for queries regarding the project ☒ As above

Title and name				
Organisation				
Address				
Phone	(w)		(m)	
Email				

2.3 Student Details

☒ Not applicable

Title and name				
Organisation				
Degree Course				
Supervisor				
Phone	(w)		(m)	
Email				

2.4 Student training and experience
Outline supervision to be provided, the proportion of work to be carried out by the student and the student's previous experience relevant to the proposed research and ethics application.
Not applicable.

3. ORGANISATION RESPONSIBLE FOR APPLICATION

3.1 List all locations where the data will be analysed (specify departments at institutions).

1. Telethon Kids Institute: McCusker Centre for Bioinformatics and Centre for Biostatistics

   All data will reside at TKI for the duration of the project. Investigators from (2) will be given secure remote access to a server where the data is residing at TKI, for the purposes of analysis.

2. Murdoch Children's Research Institute; Infection and Immunity group

   Investigators from MCRI will be conducting analysis of WADLS data as noted above, using a secure remote access method to the TKI server.

3.2 Type of organisation

- ☐ WA Department of Health (DoH) /WA Health
- ☐ A State department or agency other than DoH (e.g. other state government department or public university
- ☒ A non-governmental organisation (e.g. private hospital, medical research institute, private university)
- ☐ Commonwealth department or agency
- ☐ Other, specify below

3.3 Does this project have any potential conflicts of interest?

☒ NO ☐ YES, please provide details below

- ☐ Commercially
- ☐ Financially
- ☐ Intellectually
- ☐ Other, specify below

3.4 Has this project received funding?

☐ NO ☒ YES, please provide details below

Funded by NHMRC Project Grant APP1065494 funded for 2014-18, held by Prof Nick de Klerk (TKI), Prof David Burgner (MCRI) and A/Prof Kim Carter (TKI).

4. PROJECT SUMMARY

4.1 Provide a lay summary (approximately 50–100 words) of your project.

Avoid highly technical terms, medical jargon and abbreviations.

Infectious diseases are the leading cause of childhood death and health service use worldwide. Why some children develop more severe infection is largely unknown. The relative contribution of genetic and environmental factors to common childhood infections severe enough to require hospitalisation is largely unexplored and is the overall aim of this project. The first study on associations between non-twin brothers and sisters is currently being analysed. In this second distinct and separate part of our study, we will examine these same questions in twin families.

4.2 Information about your project, including the lay summary above, may appear on the DoH HREC and/or Data Linkage WA website following approval to maintain public confidence in research. If you do not consent to this, please tick the box below and provide justification.

☐ **If you do *not* consent,** please provide justification below

5. PROJECT OUTLINE

5.1 Background

Provide an overview demonstrating the need for this study with appropriate references.

Infection is the leading cause of global child mortality and morbidity[11], and in Western Australia the commonest reason for childhood hospitalisation[12]. Strategies to identify those at highest risk, to prevent severe infection and death, and which actively survey the occurrence of disease are the centrepiece of many public health measures nationally and globally. Investigation of hospitalisation rates with severe infections in twin-families is an innovative and informative approach to identify the contribution of host genetic and environmental determinants of increased risk. Complications of infection-related conditions lead to some of the commonest surgical procedures in children[13]. Although all children are repeatedly exposed to life-threatening pathogens, only a minority develop severe infection. Understanding the basis for this differential susceptibility is critical to reduce the huge infectious disease burden. Host genetic factors contribute to infection-related mortality, but most data relate to susceptibility to specific, largely tropical diseases, rather than infections that are common in Australian children. Therefore the genetic and environmental determinants of hospitalisation with common infections in industrialised settings such as Australia are unexplored and will be addressed in this NHMRC-funded project. Strategies to identify those at highest risk, to prevent severe infection and death, and which actively survey the occurrence of disease are the centrepiece of many public health measures nationally and globally. Investigation of hospitalisation rates in siblings of children with severe infections is an innovative and informative approach to identify the contribution of host genetic and environmental determinants of increased risk.

The study, and hence the request for data, has two separate parts:

(i) For the first part of this study (Project 201505.02), we aimed to estimate the genetic and environmental effects, estimated from sibling analyses, on risk of hospitalisation for common childhood infection and related procedures at a WA population level. These analyses are ongoing.

(ii) The second study (and the subject of this application) will also examine the contributions of genetic and shared and unshared environmental effects on infection risk, using zygosity data on the >5000 twin pairs in the Western Australian Twin Register, as well as the unlike sex twin pairs (who must definitely be dizygous) not on the Register, but with a different underlying approach, utilising data from the twins and their sibs and parents.

5.2 Research Aim(s)

These should directly relate to the requested datasets, variables and timeframe.

Aim 1: To estimate the genetic and shared and non-shared environmental contributions to the overall risk of infection-related hospitalisation, and to individual risks of specific infection groups.

Aim 2: To estimate the genetic and shared and non-shared environmental contributions to infection-related surgical procedures, such as myringotomy and ventilation tube insertion, and appendectomy.

5.3 Design

e.g., retrospective cohort study, case-control study, single group pre/post

This is a twin-family study based on all twins born alive in WA from 1980 onwards. A subset of these were recruited into the WA Twin Register (WATR). The WATR is a data collection obtained from earlier studies, with identifying data and research data stored separately. The data are managed at Telethon Kids Institute by Michelle de Klerk who has been responsible for collecting, collating and storing the paper and telephone questionnaire data. She also forwards de-identified data to researchers, including Prof Nick de Klerk, the data steward, for analysis. The only variable from the WATR that is required for this study is the twins' zygosity, which will be needed for most analyses described below (under Methodology). Data for the WATR twins will be supplemented by unlike sex twins (who therefore must be dizygous) not on the Register, and their families, to increase precision in some estimates. The analyses are based around families, their degrees of relatedness, and the levels of association for the different diseases. In the first part of the study we will use hospitalisation with infection (and specific infections and groups of infections, as described further below) as the primary outcome measure, as this is a robust and well-accepted measure of severity. Its use as a marker of infectious disease severity is less influenced by health-seeking behaviour and social disadvantage than emergency department or primary care presentations, and by practitioner-related variation in management[10]. We will subsequently use the same analysis methodology using operative procedures as the outcome measure. Data from the MNS (such as birthweight, gestational age, complications and conditions etc. as marked on the variable list) will be used as adjustment factors in the analyses described below. Data from the Deaths register (date and cause of death) will be used for accurate apportioning of person-time and, where the cause of death is related to the specific cause or cause-grouping being investigated, using it as the more severe, and competing, outcome measure.

In addition, with the linked data on all the WA twins in this MNS based study cohort, we will be able to make the necessary comparisons between all twins on the Register and all those not on the Register (on demographic and health-related variables) in order to address issues of comparability and generalisability.

Indigenous status will be used only as an adjusting factor (or exclusion criterion) and so no

WAAHEC approval will be required for this Stage.

5.4 Methodology

Specify the data extraction process, proposed data analyses and how these will achieve your research aims

As in Stage 1 of this study there is no control group as such, and 'exposed' subjects will be those who have a family member with a hospital admission for the diagnosis or group of diagnoses under study at a particular time, 'unexposed' subjects will be those who don't, at that same time. Depending on the outcome and the study time, people can be 'cases' or 'controls' and can also be 'exposed' or 'unexposed'. Families with twins not on the WA Twin Register (see Linkage form) will be incorporated into these main analyses only if they have opposite sex twins (who therefore must be dizygous). However, all the unlinked families will enable comparison with the linked families for any major differences, so as to enable estimation of measures of generalisability and applicability to the whole population. After checking, cleaning and editing, as in work from our previous studies, we will use a combination of ICD9 (for admissions 1980-1987), ICD-9-CM (1987-1999), and ICD-10-AM (1999 onwards) for diagnostic codes and Australian Classification of Health Interventions (ACHI) for procedure codes. For infectious disease diagnoses, we will use (as in previously published studies) individual ICD codes and clinical groupings of disease (such as lower respiratory tract infection and infectious gastroenteritis) using *a priori* modifications based on the Clinical Classifications Software system.

In order to distinguish genetic and environmental determinants of disease susceptibility, the most efficient study design involves nuclear families and among these, one of the most powerful and broadly applicable is the *twin family* design consisting of nuclear families with childhood twins[1-3]. Usually in variance components analysis, the decomposition of residual variation into genetic and environmental components has been used to generate a combination of four different components of variance: σ^2_A (additive polygenic effects = cumulated linear effects at many individual loci), σ^2_D (genetic dominance = cumulated non-linear effects at many individual loci), σ^2_C (common family environment), σ^2_{Cs} (common sibling environment). These components may be used to determine whether disease clustering within families is best explained by genes, environment or a combination of both, and precisely what type of genetic or environmental exposure should be sought. They can also be used to investigate sharing of genetic or environmental determinants between different components of a complex phenotype and therefore have major implications for aetiological studies. A standard nuclear family design allows up to three of these components to be resolved: σ^2_A, σ^2_C, and (σ^2_{Cs} or σ^2_D). A traditional twin design (MZ or DZ twins reared together) can be used to estimate two components. However, a twin family design not only increases the statistical power to resolve the individual components of variance, but also allows all four components to be estimated uniquely[1,2]. Recent refinements to the standard twin design allow more complex familial correlations to be estimated when other children are also incorporated[4]. However, as we will compare our twin family analyses with results from Study 1, and focus our results on public health issues, most of our planned analyses will involve censored survival-type data (either from birth, or from index sibling's or twin's admission) as described in Study 1, but incorporating different family relationships and covariates. This approach will also enable translation of the relative risks into population attributable fractions, as we have done previously for lower respiratory tract infections in the WA population[5].

We will also employ a Weibull regression-based cumulative risk model[6] for different family relationships, and then the methods from Thomsen et al[7] for joint survival data. Any remaining within-family correlation after including the different relationships for each member, will be allowed for by using robust variance estimation.

As confirmation of the significance of our findings, we will repeat our analyses using both the mixed model formulations of Rabe-Hesketh[8] and Wang *et al*[9], which allow extension to

censored response variables, which will form the principal analyses. To confirm these effects, further analyses will use the structural equation models in Open MX, described above.

1. Clifford CA, Hopper JL, Fulker DW, Murray RM. A genetic and environmental analysis of a twin family study of alcohol use, anxiety, and depression. Genet Epidemiol. 1984;1:63-79.
2. Hopper JL. Variance components for statistical genetics: applications in medical research to characteristics related to human diseases and health. Stat Methods Med Res. 1993;2:199-223.
3. Weiss ST. Association studies in asthma genetics. Am J Respir Crit Care Med. 2001;164:2014-5.
4. Maes HH, Neale MC, Medland SE, Keller MC, Martin NG, Heath AC, et al. Flexible Mx specification of various extended twin kinship designs. Twin Res Hum Genet. 2009;12:26-34.
5. Moore HC, de Klerk N, Richmond P, Lehmann D. A retrospective population-based cohort study identifying target areas for prevention of acute lower respiratory infections in children. BMC Public Health. 2010;10:757.
6. de Klerk N, Alfonso H, Olsen N, Reid A, Sleith J, Palmer L, et al. Familial aggregation of malignant mesothelioma in former workers and residents of Wittenoom, Western Australia. International Journal of Cancer. 2013;132:1423-8.
7. Thomsen SF, Duffy DL, Kyvik KO, Backer V. Genetic influence on the age at onset of asthma: a twin study. J Allergy Clin Immunol. 2010;126:626-30.
8. Rabe-Hesketh S, Skrondal A, Gjessing HK. Biometrical modeling of twin and family data using standard mixed model software. Biometrics. 2008;64:280-8.
9. Wang X, Guo X, He M, Zhang H. Statistical inference in mixed models and analysis of twin and family data. Biometrics. 2011;67:987-95.
10. Petersen L, Andersen PK, Sørensen TIA. Genetic influences on incidence and case-fatality of infectious disease. PLoS ONE 2010;5:e10603.
11. Murray CJ, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet. 2013;380(9859):2197-223.
12. Carville KS, Lehmann D, Hall G, Moore H, Richmond P, de Klerk N, et al. Infection is the major component of the disease burden in aboriginal and non-aboriginal Australian children: a population-based study. Pediatr Infect Dis J. 2007;26:210-6.
13. Kubba H, Pearson JP, Birchall JP. The aetiology of otitis media with effusion: a review. Clin. Otolaryngol. Allied Sci. 2000;25:181-194

6. PROJECT DURATION

6.1 Entire period spanning research design, approval, implementation, analysis to publication. Note that delivery of linked data can take several months depending on request complexity.

Expected start date	14/05/2018
Expected end date	23/04/2021

6.2 Do you have project deadlines to bring to our attention?

Every effort will be made to deliver your data within the requested timeframe, pending project

complexity and existing workloads
☐ NO ☒ YES, please provide details, including specific dates, below
The end date for the NHMRC grant is now extended to April 2019, so receipt of data within the next 6 months would help considerably in getting the grant competed on time. Given recent changes to NHMRC regulations, the grant period itself can't be extended any further, however analysis and reporting will be continued until finished (certainly by April 2021).

7. PERSONNEL

List all personnel, describing qualifications, expertise, project role and institutional email.

Projects seeking linked data must have at least one team member based at a WA institution.

Non-WA Public Sector employees must sign a Declaration of Confidentiality.

Where multiple positions are held across organisations, list only the institution relevant to this application.

Title, full name, qualifications, employing institution and email *e.g.* *Prof Albert Smith, MBBS, The University of Western Australia, a.smith@uwa.edu.au*	Expertise and role in the project	Access to data required	Confidentiality Agreement submitted
Prof Nick de Klerk, PhD, Telethon Kids Institute, nick.deklerk@Telethonkids.org.au	Principal Investigator: advising on data analysis, interpretation of results, report writing	☒ Yes ☐ No	☒ Yes ☐ No
Prof David Burgner, PhD, FRACP, Murdoch Childrens Research Institute, david.burgner@mcri.edu.au	Co-investigator: infectious disease expertise, interpretation of diagnostic codes and results, report writing	☒ Yes ☐ No	☒ Yes ☐ No
Dr Jessica Miller, PhD, Murdoch Childrens Research Institute, jessica.miller@mcri.edu.au	Analyst: statistical and epidemiological analysis, linked data expertise, report writing	☒ Yes ☐ No	☒ Yes ☐ No
A/Prof Kim Carter, PhD, Telethon Kids Institute, kim.carter@telethonkids.org.au	Co-investigator: data handling and analysis, interpretation of results, report writing	☒ Yes ☐ No	☒ Yes ☐ No
Ms Michelle de Klerk, Telethon Kids Institute, michelle.deklerk@Telethonkids.org.au	Research assistant/data manager	☐ Yes ☒ No	☐ Yes ☒ No

8. ETHICS REVIEW AND APPROVALS

8.1 Does your project require review by the DoH HREC? Refer to DoH HREC website

☐ NO ☒ YES

8.2 Does your project require approval by any other ethics committee?

This could include WA Health ethics committees and/or external ethics committees. Refer to DoH Research Development Unit information links:

- WA Health Research Governance Policy and Procedures 2012
- WA Health Research Ethics

☐ NO ☒ YES, please attach a copy of each ethics committee approval granted and provide details of the current status of all applications

Murdoch Children's Research Institute, HREC

WAAHEC Approval. We already have approval for Stage 1, and since Stage 3 of the project requires separate analysis by indigenous status and will require results from both Stage 1 and Stage 2, we will be seeking WAAHEC approval before commencing Stage 3 (if Stage 3 is feasible).

TKI and UWA recognise HREC approval from DOHWA, so no further approvals are required.

8.3 Does your project require other approvals?

☐ NO ☒ YES, please attach a copy of each approval granted and provide details of the current status of all applications

ACR approval for Code of Death and Country of Birth data.

9. PROJECT DATA

Data Services

9.1 Specify each data service you seek.

Corresponding forms must be attached for each data service requested, available at the Data Linkage WA website.

DATA SERVICE		REQUEST
Extraction	Data extraction from one or more data collections	☒
Linkage	New data to be linked to one or more datasets	☒
Geocoding	New addresses requiring geocoding	☐
Sample Selection	Population sample to be selected from the WA Electoral Roll	☐
Genealogical Data	Family relationships or data for related individuals	☒
Study Recruitment	Use of DoH data to contact persons for research purposes	☐
Indigenous Status Flag	YES/NO flag that can be included in record-level data provided for data linkage projects. Created from a validated algorithm for each individual from one or multiple data records held in the WA Data Linkage System. All available linked records are used. Note, it may contradict specific records in any single data collection or other independently collected data.	☒

Data Collections			
9.2 Complete a separate form for each dataset requested, available at the DLB website. If a required variable list is unavailable, please contact dataservices@health.wa.gov.au. It is strongly recommended that requests are discussed with the corresponding data custodians prior to submission. For other datasets, attach a separate Word document with details including dataset name, data custodian with contact details, and data variables required.			
DATASET	**SELECT**	**FROM** *e.g. Jan 1984*	**TO** *e.g. latest available*
Birth Registrations (since 1974)	☐		
Emergency Department Data Collection (since 2002)	☐		
Electoral Roll (since 1988)	☐		
Hospital Morbidity Data System (since 1970)	☒	Jan 1970	Most recent
Mental Health Information System (since 1966)	☐		
Midwives Notification System (since 1980)	☒	Jan 1980	Most recent
Mortality Register (since 1969)	☒	Jan 1969	Most recent
WA Cancer Register (since 1982)	☐		
WA Notifiable Infectious Diseases Data (since 1988)	☐		
WA Register of Developmental Anomalies (since 1980) Birth Defects Cerebral Palsy	 ☐ ☐		
Other datasets (please list below)	☒		
WA Twins Register (see separate document attached)		Jan 1980	Most recent

10. PRIVACY AND CONSENT

Personal Health Information

Personal health information is defined by the Practice Code for the Use of Personal Health Information from the Department of Health Data Collections (herein referred to the Practice Code) Section 3 as information about an individual or institution where the identity is apparent or can be reasonably ascertained from the information itself. If it is reasonably possible for a data recipient to identify individuals using other linked information they possess then this is also considered personal information.

Personal Information Variables

10.1 Specify if you require any of the following information in your data.

Note this does *not* apply to data used by DLB for linkage purposes or to contact people.

Participant names	☐ YES ☒ NO
Participant addresses	☐ YES ☒ NO
Participant *full* dates of birth (i.e., ddmmyyyy)	☒ YES ☐ NO
Patient identifiers (e.g. UMRN)	☐ YES ☒ NO
Clinician or health service provider identifications	☐ YES ☒ NO
Individual hospital or healthcare institution identifications	☐ YES ☒ NO
Geo-coded points (longitude and/or latitude)	☐ YES ☒ NO

10.2 Are you applying for the release of personal information?

If you ticked YES to any item in Question 10.1, you must answer YES to this question.

☐ NO ☒ YES, please explain why non-identifiable information cannot be used

Timing is critical for our study, and we need the order of events and an exact DOB (coupled with full date of admission for HMDS data) to determine when critical events have occurred and the exact ages of occurrence The precise age at birth and hospitalization is essential for addressing large, previously under-appreciated risk groups, such as 'late preterm infants' (a group defined in gestational ***days*** and not weeks i.e. 34^{+0} to 36^{+6}). Furthermore, derived variables such as age in days at admission or death are not applicable for the comparison children who do not die or are not admitted.

Otherwise, we are not requesting identifying variables, and will not be publishing any individual level data – only aggregated statistics and broad population trends.

Although we are not applying for personal information from DOH, data relating to twin families on the Twin Register (WATR) will be potentially re-identifiable, as the details of cohort members, including personal information, is held by us. The WATR was established by us and has been maintained by us for nearly 20 years. However, all analyses will be conducted using coded data, as has been the case for all previous work. Identifying data, for use in linkages, is stored in a separate table, with separate security rights, and linked only by a record number. These record numbers are sequential and contain no imbedded meaning i.e. you cannot identify a person, or any details about the person's identity, from a record number without access to the identification table itself.

The data manager who works on updating and maintaining the database is not involved in data analyses and data analysts only receive coded data files, without personal details, for analyses. When results of our research are presented or published, individuals are not

identified. All staff will comply with DOH, TKI, and MCRI confidentiality rules.

In terms of privacy, (as noted below in Section 12) all project data will be stored on a secure Bioinformatics server located at the Telethon Kids Institute, in the secured computer room. The server itself is not directly externally accessible, and all project data will be encrypted (and password protected) when stored on the server. The named project investigators will have password access to decrypt the data for the purposes of this project – no other individuals will be able to access the server or data. The MCRI named investigators will be given encrypted VPN access to TKI, to allow them to login and perform data analyses on the data at TKI as described above.

Privacy sensitive variables (i.e. full date of birth) will be used to ascertain exact age, timings and orders of events, and as such will be used to derive new analytical variables, which will form part of analytical datasets (i.e. the variables themselves are not directly used in analyses).

10.3 Describe all measures to protect privacy. Refer to the Practice Code (Section 3.1).

All the information we get will only be used for the purposes described and only by those listed in the application. We will be using the data security measures described further below and data will not be passed on to anyone else or to any other institutions, nor will data merged with any other information without prior approval from DoH HREC and the Data Custodians. We are not requesting identifying variables, and will not be contacting any individual nor publishing or identifying any individual. No information will be retained after the approved retention date, when it will be disposed of securely and completely.

Cell sizes less than 5 will not be published

Consent

10.4 Will consent be sought from participants for the use and disclosure of their information from the data collections?

☒ NO → go to 10.5

☐ YES → describe below your consent procedure and attach contact letters, information sheets and consent forms.

10.5 If consent will not be sought, explain why it would be impracticable to obtain and provide details below.

☒ The size of the population involved in the research.

☒ The proportion of individuals likely to have moved or died since the health information was originally collected.

☒ The risk of introducing bias into the research.

☒ The risk of creating additional threats to privacy.

☒ The risk of inflicting psychological, social or other harm by contacting individuals.

☒ The difficulty of contacting individuals directly when there is no existing or continual relationship between the organisation and the individuals.

☒ The difficulty of contacting individuals indirectly through public means.

All the above itemised reasons contribute in some way to the complete impracticability of obtaining consent. There are no existing resources to do so and the most appropriate methods to use, given the need for the linked information obtained through Family Connections, are not at all clear. Bias is almost certain to result from non-random agreement to participate and privacy rights could well be compromised by attempts to contact study participants, especially those with no previous connection to the WATR.

11. OTHER SOURCES OF INFORMATION FOR THE PROJECT

11.1 Indicate other sources of information to be used in this project and provide details below.

☐ Information will be collected directly from participants

☐ Information will be collected about participants from another person (e.g. carers, doctors)

☐ Information will be used from existing records held by individuals or organisations other than the DoH

☒ Information previously collected by you or your organisation for other purposes

☐ Other DoH data collections

☐ Other

11.2 Describe below the source and nature of the information and specify whether your project involves the matching of records to these data.

Attach a separate Word document detailing each dataset name, the variables sought and the data custodian's name and contact details.

To avoid conflicts of interest, if the data custodian is one of the named investigators, then written approval is required from that person's manager for the release of the corresponding data. Provide details below.

Western Australian Twin Registry

Data Steward: Prof Nick de Klerk, nick.deklerk@telethonkids.org.au, 9489 7735.

Only data regarding the twins' zygosities will be merged with DOHWA data.

Prof de Klerk's manager is Prof Stephen Zubrick: Stephen.Zubrick@telethonkids.org.au

Please see further details in attached document.

12. SECURITY PLAN

12.1 Provide a detailed security plan for the information provided by DoH or collected as a result of DoH's actions. Your security plan will be assessed against the Practice Code.

Technological Security (Practice Code Section 4.3)

All project data will be stored on a secure Bioinformatics server located at the Telethon Kids Institute, in the secured computer room. The server itself is not directly externally accessible, and all project data will be encrypted (and password protected) when stored on the server. Remote access to the TKI server environment is a two-stage login security process. In the first stage, a remote user logs into a secure gateway server using the Secure Shell (SSH) protocol. The gateway server is a tightly secured and highly monitored login gateway that only allows connections from a strict list of highly randomised usernames and password that only exist on the gateway server. Any invalid username login attempts result in an instant IP address ban. Once a user has established an initial SSH login to the gateway, a second SSH connection can be connected through the gateway server to the project Bioinformatics server on the internal TKI network, which requires a second, separate username and password combination, again issued only to valid users. Once logged into the project server, the MCRI users can decrypt the project data and perform any necessary analyses on the server.
All laptops at TKI are password protected (password unique to each individual) and have automatic locking after 5mins. Server storage files are restricted to authorized users of the data. Data files being stored or archived will be encrypted. The servers and computers are firewall protected. All computers and servers have virus and spyware protection (Trendmicro).

Physical Security (Practice Code Sections 4.2 and 4.6)

Entry to the TKI building during office hours is only through continuously manned reception. Only TKI staff have access to any of the offices or computers using individually signed electronic keys. Offices are left locked when unattended. After-hours access to any part of the building is by the same keys and building is alarmed in sections. Security guards and Building Emergency Response Team are on 24-hour alert.

Transport (Practice Code Sections 4.4 and 4.7)

DoH prefers to send and receive data via secure online data transfer, such as MyFT or SUFEX.

WATR data for linkage will be sent to the DLB using SUFEX, by Michelle de Klerk. Once the encrypted data extracts are received from the DLB via secure electronic transfer (MyFT), data will be kept on the TKI servers for analyses. After this, no further transport of data will occur in any direction.

12.2 Nominated data recipient/s

Nick de Klerk

13. RETENTION AND DISPOSAL PLAN

13.1 Provide a detailed retention and disposal plan for the information provided by DOH or collected as a result of DOH's actions. This must include the period of retention beyond the study and intended date of data destruction. Your plan will be assessed against the Practice Code (Section 5).
For data described in peer reviewed publications arising from this work, we will retain data for a period of 5 years before destruction, as described in the NHMRC Code of Responsible Conduct of Research. At the conclusion of this time, data will be destroyed with due care, and any physical media returned to the DLU, DOHWA. The Principal Investigator will notify the Data Linkage Branch and DOH HREC as soon as the data have been destroyed.

Data destruction/ return date	April 2026

13.2 Does your security, retention and disposal plan comply with the Practice Code (Sections 4 and 5)?
☒ YES ☐ NO, please provide a justification below

14. DATA FORMAT

Optional (for our planning purposes only) Please describe the software packages and analytical tools you will use for this project, e.g. SPSS, SAS, Excel.
Data will be analysed and manipulated using a combination of R, PERL, SAS, Stata, and Java.

15. DISSEMINATION OF RESULTS

15.1 Explain how results will be disseminated, e.g. report, publication, conference, thesis.

Refer to the Practice Code (Section 8).

Note that final drafts of all reports, publications and presentations must be sent to the data custodians and/or Data Linkage Branch for comment at least two weeks prior to dissemination, as per Declarations in Section 17 below.

While this research involves no direct interaction with consumers / participants, the findings from this study will be directly applied to the wider community, and will be disseminated to the community via scientific publications, conferences and community engagement. If this study is successful and comes up with useful results, we plan to establish a Management Committee and Community Reference Group to advise on the appropriate ways of communicating results (before or concurrently with journal publication) and the planning of new studies. Any research outputs from the study will be sent to the DLB and Data Custodians at least 2 weeks before any wider dissemination.

15.2 Describe how confidentiality of participants will be maintained in the dissemination of results.

We are not requesting identifying variables, and will not be publishing any individual level data – only aggregated statistics and broad population trends. Any table cell counts less than 5 will be suppressed.

16. GOVERNANCE

Head of Department / School / Research Organisation

Tick the boxes to indicate you have read and understood each clause.

I/ we certify that:

☒ I/we are familiar with this project and endorse its undertaking.

☒ The resources required to undertake this project are available.

☒ The researchers have the skill and expertise to undertake this project appropriately or will undergo appropriate training as specified in this application.

☒ I/we warrant that I/we are authorised to make this application and to bind the institution below in relation to the obligations arising out of the submission of this application.

☒ The conduct of the project has been approved by…

I/we certify that	Telethon Kids Institute
	(name of institution)

accepts the legal and ethical responsibility for the conduct of this project and have adequate indemnity insurance to cover the conduct of this project and indemnifies the Minister for Health, the State of Western Australia, the Department of Health and their officers, servants, agents and contractors for any loss or damage they suffer through any breach in the conduct of this project.

FULL NAME *(PRINTED)*:	Stephen Zubrick
POSITION:	Head of Brain and Behaviour
ORGANISATION:	Telethon Kids Institute

SIGNATURE *	DATE

*** Note: if the Principal Investigator is the Head of Department / School / Research Organisation, then the next tier or authority above is required to sign the Indemnity Form. This section cannot be signed by a member of the Project Team.**

17. DECLARATIONS AND SIGNATURES

17.1 Applicant / Principal Investigator

Tick the boxes to indicate that you have read and understood each clause.

I certify that;

☒ All information in this application is truthful and as complete as possible.

☒ The project will be conducted in accordance with the ethical and research arrangements of the organisations involved.

☒ I am aware of and understand the relevant legislation and regulations, and the project will be conducted in accordance with these.

☒ I recognise that unit record data from DoH is confidential information and that I am responsible for ensuring that the information will be kept confidential.

☒ The information provided for this project by DoH will be used only for the project outlined in this application.

☒ The project will be conducted in accordance with the protocol and conditions approved for this project and in accordance with the provisions of the DoH *Practice Code for the Use of Personal Health Information from the Department of Health Data Collections.*

☒ I will make available all resulting draft manuscripts, reports or other presentations based on the analysis of linked data in this application to the relevant Data Custodians and will thereby allow DoH the opportunity to review and respond within 14 days (2 weeks).

☒ I will provide the Data Linkage Branch and/or Data Custodians with an electronic copy of all publications of results of analysis as they become publicly available.

☒ I will acknowledge the Data Linkage Branch and/or DoH in any publications, reports or presentations resulting from this application.

FULL NAME *(PRINTED)*:	Nicholas de Klerk
	26/04/2018
SIGNATURE	DATE

17.2 Supervisor/s of student/s

Tick the boxes to indicate that you have read and understood each clause.

I /we certify that:

☐ I/we will provide appropriate supervision to the student to ensure that the project is conducted in accordance with the undertakings above.

☐ I/we will ensure that any necessary training is provided to enable the project to be undertaken skilfully and ethically.

FULL NAME *(PRINTED)*:	
SIGNATURE	DATE

Government of **Western Australia**
Department of **Health**

For Office Use Only	
DL#	201704.02
HREC#	
Version	V.1
Date	17/4/18

Extraction

Heritable and environmental determinants of hospitalisation for common childhood illnesses - Study 2: association between monozygotic and dizygotic twins, their sibs and their parents in hospital admissions

Please specify the data to be extracted, listing all data sets and specific detail around selection of a cohort or cases and controls. Any date limits should also be given. If you require assistance with completing this form, please email DataServices@health.wa.gov.au.

Please note: DOHWA HREC approval is required for any personal health information*.

*Personal health information is information or opinions that relate to the health of a person where the identity of a person is apparent or can reasonably be ascertained from the information. For a more detailed explanation see the *Practice Code For the Use of Personal Health Information*. All project personnel who will have access to personal health information provided by the DOHWA must enter into a Confidentiality Agreement. Please contact the Manager or Project Manager of the Data Linkage Branch or the DOHWA HREC Executive Officer if you are unsure of the criteria for "personal health information".

1. COHORT/CASE GROUP

Cohort Description
Please provide a description of your cohort and how your study population is defined. E.g. all people living in Bunbury who were diagnosed with lung cancer between 1995 and 2005, defined through the WA Cancer Registry
All twins and multiples born alive in WA from 1980 onwards, according to Midwives data, form the basis of the cohort. Twins or multiples where any or all of them died before or during birth should be excluded.
For quoting purposes please provide an estimate of how many people there will be in your cohort.
25,000
Disease and Procedure Codes
If your cohort is to be selected from specific disease or procedure groupings, please specify the version of ICD codes you require (i.e. ICD9, ICD10), and whether they should be applied only to the principal disease/procedure code or to any of the multiple codes within a record. For the time periods and versions of ICD codes used in WA please see http://www.datalinkage-wa.org.au/sites/default/files/HMDS_ICD_DRG.pdf **Please attach an Excel spreadsheet of all the specific ICD codes you require. Please note that the DLB cannot provide or check ICD codes.** E.g. ICD10, diagnosis codes for lung cancer. See attached spreadsheet.
Geographical areas
If your cohort is to be restricted to a specific residential or service area, please specify the relevant postcodes or CDs that define the area required. **Please attach an Excel spreadsheet of the postcodes or CDs you require.** *SEIFA and ARIA codes are available for Emergency, Death, Hospital and Midwives data. If you require*

SEIFA or ARIA codes, please select this in the variable lists for each dataset (Module 4). E.g. People residing in Bunbury. See attached spreadsheet of postcode and collection districts.

Service data extraction

Please specify the datasets you require, the time period and any restrictions on which records you need.

WA Cancer Registry: Please refer to the variable list for record scope. If you require specific exclusions please list them in the comments section of that form.

Midwives Notification System, Birth Registrations & WA Register of Developmental Anomalies: Please specify below whether you need records for the birth of a person, records where they are the parent or both.

***Attach variable lists (Module 3)**

Dataset e.g. Hospital Morbidity	**Time period** e.g. Jan 1995 – most recent	**Restrictions** e.g. 'I wish to analyse comorbidities. Please extract all records for 5 years before and all after the index admission'
WA Twin Register	Jan 1980 – current	We wish to link zygosity records from the WATR to all twins identified in the cohort.
Midwives Notification System	Jan 1980 – current	Linked records where they are the baby for twins and multiples, except where any of them died before or during birth
Mortality	Jan 1980 – current	All linked records
Hospital Morbidity Data Collection	Jan 1980 – current	All linked records

2. COMPARISON GROUP/CONTROLS

Control Group Description

Please provide a description of your controls.
E.g. Random sample of people from the electoral roll, matched on year of birth and gender to the cohort 5:1.

This is a twin-family cohort study on all twins born in WA since 1980. The twins form the basis of the study. As in Stage 1 of this study there is no control group as such, and 'exposed' subjects will be those who have a family member with a hospital admission for the diagnosis or group of diagnoses under study at a particular time, 'unexposed' subjects will be those who don't, at that same time. Depending on the outcome and the study time, people can be 'cases' or 'controls' and can also be 'exposed' or 'unexposed'. Families with twins not on the WA Twin Register (see Linkage form) will be incorporated into the main analyses only if they have opposite sex twins (who therefore must be dizygous). However, all the unlinked families will enable comparison with the linked families for any major differences, so as to enable estimation of measures of generalisability and applicability to the whole population.