**COMBINED PROTOCOL & APPLICATION FOR DATA TEMPLATE**

**v.8 December 2021**

This template has been designed for population health research utilising and/or linking routinely collected health data held by the NSW Ministry of Health or the Cancer Institute NSW. This template combines the NSW PHSREC research protocol template and CHeReL data request form.

**This research protocol template must be used in conjunction with, and complement, the HREA.**

All forms and further information for submission to the NSW Population and Health Services Research Ethics Committee (PHSREC) can be found at:

**https://www.cancer.nsw.gov.au/research-and-data/nsw-population-health-services-research-ethics-com/how-to-apply**

All forms and further information for submission to the NSW Centre for Health Record Linkage (CHeReL) can be found at: [**https://www.cherel.org.au/apply-for-linked-data**](https://www.cherel.org.au/apply-for-linked-data)

# PROJECT DETAILS

|  |
| --- |
| Title: Australian Urban Health Indicators (AusUrb-HI): Urban Heat, Liveability and Health study |
| Short Title: AusUrb-HI Heat, Liveability and Health study |
| CHeReL Ref: 2022.12 |
| NSW PHSREC REGIS Ref: 2022/ETH00905 |
| Other (e.g., Sax Institute, SURE workspace name, AIHW, ACT Health and/or your ref): |
| Contact Person: Tracy Baylis  Email and Phone no.: [tracy.baylis@unimelb.edu.au](mailto:tracy.baylis@unimelb.edu.au) 0417333673 |
| Is this a data linkage Project:  **🗸** Yes  ☐ No (Please speak to PHSREC before proceeding) |
| Is this a cross or multi-jurisdictional Project  ☐ Yes Please provide PHRN number (if applicable)  List jurisdictions (if applicable)  **🗸** No |
| Are you seeking PHSREC approval under the National Mutual Acceptance (NMA) Scheme?  ☐ Yes  **🗸** No |

# VERSION CONTROL (MAKE CONSISTENT WITH FOOTER)

|  |  |  |  |
| --- | --- | --- | --- |
| Version | Date | Amendment (brief description) | Amendment date  (as per amendment form) |
| 1.0 | 30/03/2022 | Original application | N/A |
| 1.1 | 27/04/2022 | Analysis plan, references, data collections | 27/04/2022 |
| 1.2 | 22/07/2022 | Revisions, additional supporting documents | 22/07/2022 |
| 1.3 | 15/12/2022 | Revisions, updated analysis plan and risk vs benefit documents | 15/12/2022 |
| 1.4 | 30/03/2023 | Centre for Epidemiology and Evidence as curators of the SURE gateway | 30/03/2023 |
| 1.5 | 05/10/2023 | Added Dr Aiden Price to the researcher list (p4 and p5). Added new datasets (p24 and p25). | 05/10/2023 |

S

# INVESTIGATORS AND PARTICIPATING INSTITUTIONS

|  |  |
| --- | --- |
| 2BPrincipal Investigator | |
| Name & Title | Associate Professor Melanie Davern |
| Institution | RMIT University |
| Division | Australian Urban Observatory (AUO) |
| Position | Principal Research Fellow |
| Mailing address | Australian Urban Observatory,  Building 8, Level 11, RMIT University City Campus, GPO Box 2476  Melbourne VIC 3001  Australia |
| Email address | [melanie.davern@rmit.edu.au](mailto:melanie.davern@rmit.edu.au) |
| Phone number | [+61 399253090](tel:+61399253090) |
| 2BCOORDINATING Principal Investigator | |
| Name & Title | Tracy Baylis |
| Institution | University of Melbourne |
| Division | Australian Urban Research Infrastructure Network (AURIN) |
| Position | Strategic Planning & Implementation Manager |
| Mailing address | AURIN,  Level 3, Thomas Cherry Building, The University of Melbourne  Melbourne VIC 3010  Australia |
| Email address | tracy.baylis@unimelb.edu.au |
| Phone number | +61 417333673 |
| 4BInvestigator / Researcher | |
| Name & title | Dr Hao Chen |
| Institution | University of Melbourne |
| Division | Australian Urban Research Infrastructure Network (AURIN) |
| Position | Spatial Database Software Developer |
| Email address | chen.h@unimelb.edu.au |
| Phone number | +61 481380464 |
| 4BInvestigator / Researcher | |
| Name & title | Flavia Barar |
| Institution | University of Melbourne |
| Division | Australian Urban Research Infrastructure Network (AURIN) |
| Position | Data Science Officer |
| Email address | bararf@unimelb.edu.au |
| Phone number | +61 426103723 |
| 4BINVESTIGATOR / RESEARCHER | |
| Name & title | Dr Aiden Price |
| Institution | University of Melbourne |
| Division | Australian Urban Research Infrastructure Network (AURIN) |
| Position | Environmental Health Domain Specialist |
| Email address | [aiden.price@unimelb.edu.au](mailto:aiden.price@unimelb.edu.au) |
| Phone number | +61 403 704 156 |
| 4BINVESTIGATOR / RESEARCHER | |
| Name & title | Dr Derrick Lopez |
| Institution | The University of Western Australia |
| Division | School of Population and Global Health |
| Position | Research Fellow |
| Email address | Derrick.Lopez@uwa.edu.au |
| Phone number | +61 8 6488 1276 |
| 4BINVESTIGATOR / RESEARCHER | |
| Name & title | Dr Fadhillah Norzahari |
| Institution | RMIT University |
| Division | Australian Urban Observatory (AUO) |
| Position | Spatial Research Manager |
| Email address | fadhillah.norzahari@rmit.edu.au |
| Phone number | 03 9925 7438 |
| 4BINVESTIGATOR / RESEARCHER | |
| Name & title | Professor David Preen |
| Institution | University of Western Australia |
| Division | School of Population and Global Health |
| Position | Chair in Population Health |
| Email address | david.preen@uwa.edu.au |
| Phone number | +61 8 6488 1251 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 4BInvestigator / Researcher Accessing Linked Unit-record Data Please list all the project personnel, including students who will have access to linked data. | | | | |
| Researcher Name | Email address | Site of data access (e.g., *Institution* via *the SURE, Institution* on *secure file server*) | Accessing NSW Linked Data | Accessing Linked Data (other Jurisdiction(s)) |
| Flavia Barar | bararf@unimelb.edu.au | University of Melbourne via SURE |  |  |
| Hao Chen | chen.h@unimelb.edu.au | University of Melbourne via SURE |  |  |
| Aiden Price | [aiden.price@unimelb.edu.au](mailto:aiden.price@unimelb.edu.au) | University of Melbourne via SURE |  |  |
| Melanie Davern | Melanie.Davern@rmit.edu.au | RMIT University via SURE |  |  |
| Fadhillah Norzahari | Fadhillah.norzahari@rmit.edu.au | RMIT University via SURE | ☒ |  |
| Derrick Lopez | Derrick.Lopez@uwa.edu.au | The University of Western Australia via SURE | ☒ |  |
| David Preen | david.preen@uwa.edu.au | The University of Western Australia via SURE | ☒ |  |

# TABLE OF CONTENTS

[1. PROJECT DETAILS 1](#_Toc89186188)

[2. VERSION CONTROL (MAKE CONSISTENT WITH FOOTER) 2](#_Toc89186189)

[3. INVESTIGATORS AND PARTICIPATING INSTITUTIONS 3](#_Toc89186190)

[4. TABLE OF CONTENTS 5](#_Toc89186191)

[5. BACKGROUND / RATIONALE 6](#_Toc89186192)

[6. AIMS AND OBJECTIVES 7](#_Toc89186193)

[7. METHODS 7](#_Toc89186194)

[8. DATA GOVERNANCE 9](#_Toc89186195)

[9. ANALYSIS PLAN 12](#_Toc89186196)

[10. PROJECT FUNDING / SUPPORT 13](#_Toc89186197)

[11. REFERENCES 14](#_Toc89186198)

[12. DATA LINKAGE 15](#_Toc89186199)

[13. SECTION C - FAMILY LINKAGE 19](#_Toc89186200)

# BACKGROUND / RATIONALE

|  |
| --- |
| *Provide an abstract/summary of your project (max 100 words)* |
| The Australian Urban Health Indicators (AusUrb-HI) project will develop new indicator data assets to improve our understanding of the health of urban populations and identify incidence patterns and key risk factors across the population. The project includes three case studies: Cancer Determinants, Heat Health Vulnerability and Urban Liveability & Health, which will integrate health, socio-economic, environmental, climate and built environment datasets to provide a holistic spatially-explicit understanding of urban population health. These indicators will allow health, urban and social infrastructure planners and policy makers to develop targeted policies and actions, and the outcomes will be shared with the research community.  Note: We are seeking a combined application for the Heat and Liveability & Health case studies due to similarities between the two projects. The Cancer Determinants case study has already received ethics approval and completed a separate data access application form. |

|  |
| --- |
| Provide a background to the study including a brief literature review, outline of knowledge gaps, how the study will address these, and the intended contribution to the field (750 - 1000 words) |
| A review of existing heat health studies, vulnerability indicators, and heat health plans was conducted to understand the adverse effects of extreme heat on human health, affected population profiles and associated factors that are likely to increase heat health vulnerability. Extreme heat events can affect all population segments (4) (7) (17), increasing emergency department presentations for all age groups during heatwaves (23) and posing an increased risk to population based on age, underlying health conditions, low socioeconomic status, and isolated living conditions (4) (7) (17). Heatwaves are Australia's deadliest natural hazard (5) and the occurrence and severity of is predicted to increase through climate change (4), further exacerbating negative effects of extreme heat in densely built urban areas with low vegetation through the urban heat island effect (4) (20) in Australian cities and towns.  While several indicators have described the phenomenon of heat vulnerability, health outcomes data (emergency department presentations, hospitalisations and deaths) have been not been linked with social, environmental, climate and health data at a fine spatial scale. The urban fabric has not been captured at an adequate level of detail and the relationship between built environment parameters has not been linked to health outcomes. More detailed analysis is required at a smaller aggregation level (ABS ASGS SA1) to fully understand these complex relationships and generate a new understanding of population vulnerability to extreme heat events in New South Wales.  This study will integrate detailed health outcomes data provided by CHeReL with existing AURIN data to produce an improved understanding of heat health vulnerability. Emergency department presentations, hospitalisations, and deaths will be examined in relation to extreme heat data and analysed in a detailed context that includes environment determinants (access to vegetation and water), underlying health conditions (diabetes, cardiovascular disease, mental illness, decreased mobility, health risk factors), built environment (road and building density, sky view factor, surface roughness, land use) and socioeconomic determinants (unemployment, education, culturally and language diverse population, housing stress, family composition, social isolation).  The heat vulnerability study will generate heat health vulnerability indicators at a fine spatial scale, facilitating use by researchers, policy, and decision-makers to inform future planning decisions, adapt to the impact of extreme heat, and improve health and habitat conditions for climate sensitive cities. This and other assets including algorithms will be made available via AURIN to enable further research. |

# AIMS AND OBJECTIVES

|  |
| --- |
| Provide a statement of primary and secondary aims/objectives, key research questions, and/or a clearly defined hypothesis. |
| **Aims**  To provide new insights into the relationships between local liveability, public health outcomes, health inequalities and environmental vulnerability within Australian cities. The relationship between public health outcomes and the engineered, social, economic and environmental fabric of the built environment is relatively poorly understood. To address this, a Liveability Index based on access to social infrastructure, walkability, public transport, public open space, housing affordability and local employment will be spatially combined with public health outcomes associated with the environmental vulnerability of the urban populous. The utility of spatially integrating liveability with environmentally determined public health outcomes will be demonstrated via an urban heat health vulnerability case study that spatially integrates health-related heat vulnerability with key liveability metrics to understand the potential of different urban planning and design options for positive environmental public health outcomes.  **Objectives**  (i) To develop a new SA1 level population health heat vulnerability indicator for New South Wales that captures population heat exposure, heat sensitivity and adaptive capacity.  (ii) To undertake spatial ‘hotspot’ analysis to statistically determine significant locations of heat health vulnerability for New South Wales major urban conurbations.  (iii) To quantify the relationship between urban liveability and population heat vulnerability for the major urban areas of New South Wales.  (iv) To evaluate the relationship between local liveability and key public health outcomes across the neighbourhoods and suburbs of New South Wales major urban conurbations.  (v) To compare the results of (iii) and (iv) with demographic data from the ABS Socioeconomic Index for Area (Index of Relative Socio-economic Disadvantage) for 2016 and 2021 to investigate how socio-economic disadvantage, liveability and environmental health inequities have changed.  (vi) To construct a new integrated data asset available via AURIN that explicitly links neighbourhood level liveability, aggregated area-based incidence of key public health outcomes and socio-economic disadvantage. It also contributes to sustainable development and achieve Sustainable Development Goals (SDGs).  (vii) To provide implications for policy development and practical intervention to tackle heat issue in NSW as well as the national level. |

# METHODS

|  |
| --- |
| 5BStudy Design |
| Describe the type of study (e.g., retrospective cohort study, case control study). |
| This is a retrospective cohort study of all patients living in selected regions of NSW who presented to a NSW Emergency Department, admitted to hospital or died between 2016 and 2021. We will compare the number and type (i.e. diagnosis, cause of death) of events during periods of heatwave and non-heatwave periods. |
| 6BCohort/study population |
| Please describe your cohort/study population, specifying any inclusion /exclusion criteria. |
| The cohort is population of all ages living in major urban areas in New South Wales (please use excel file “study area geocodes” to select areas of interest) who presented to a NSW Emergency Department, admitted to hospital or died between 2016 and 2021. For each identified patient, we require all their Emergency Department and hospitalisation data dating back to 2011 in order for us to determine the medical history with a minimum of a 10-year look-back.  Heatwaves are the deadliest natural hazard in Australia and while the elderly population is more vulnerable to the risk of heat related illness, all age groups are susceptible to illness and death during heatwaves. It is important to include population of all ages in this study to complete an accurate heat vulnerability indicator based on health outcomes.  We used the Australian Bureau of Statistics (ABS) definition of Greater Capital Cities Statistical Areas, Significant Urban Areas and Sections of State with Major Urban/Other Urban to define urban areas in this study. The NSW study area includes Albury, Newcastle-Maitland, Sydney, Wollongong and Tweed Heads.  ABS defines a “Significant Urban Area” as regions with more than 10,000 residents. We have provided two files derived from the ABS that provide SA1 codes defining these urban regions:  (i) 2011 NSW significant urban area.xlsx  (ii) 2016 and 2021 NSW significant urban area.xlsx  Note: “Significant Urban Area” for 2021 is not currently available, therefore we are using the 2016 classification for 2021. |

|  |  |  |  |
| --- | --- | --- | --- |
| 7BDATA COLLECTION | | | |
| Please identify the nature of the data to be collected (multiple options may be selected). | | | |
| Primary data collection (e.g., original data from surveys, interviews, and/or focus groups etc.)  **Please provide a description of primary data sources below.**  **Please specify the names of the sites for primary data collection.** | | | |
|  | | | |
| Secondary data collection (e.g., routinely collected data)  Please provide details for all datasets including those not to be linked by CHeReL | | | |
| **Dataset Name**  (e.g., NSW Admitted Patient Data Collection, Medicare Benefits Schedule, Victorian Admitted Episodes Dataset) | **Data Collection Site/Agency**  (e.g., NSW Dept. Education) | **To be Linked By**  (e.g., CHeReL/AIHW/DVL) | **Agency type for secondary data** |
| NSW Admitted Patient Data Collection (APDC) | NSW Department of Health | CHeRel | State / Territory  Commonwealth  Private Sector |
| NSW Cause of Death Unit Record File (CODURF) | NSW Department of Health | CHeRel | State / Territory  Commonwealth  Private Sector |
| NSW Emergency Department Data Collection (EDDC) | NSW Department of Health | CHeRel | State / Territory  Commonwealth  Private Sector |
| NSW Registry of Births Deaths and Marriages - Death Registrations (RBDM – death registrations) | NSW Department of Health | CHeRel | State / Territory  Commonwealth  Private Sector |
| *Please note: The AURIN datasets we are using include aggregated area-level social, demographic, and environmental data in SA1, SA2, PHA level, or in other format (e.g., raster). These datasets are non-person identifiable, and the integration (i.e., joining tables) between AURIN data and NSW health data will only happen at these aggregated area levels.* | | | |
| 8BConsent | | | |
| Outline the consent process(es) to be used e.g.   1. **Informed consent** 2. **Opt-out consent** 3. **Request a waiver of consent – with strong justifications** | | | |
| We are seeking waiver of consent for both cohorts based on the following justifications:  (i) there is no direct involvement of patients in the research, but we will use data that were derived from their previous encounters with health service providers dating back to 2001 (depending on the data source);  (ii) since the names or addresses of people are not included in the data extract provided to the research team, any requirement to seek consent will increase the threat to privacy due to the need to identify these people for the purpose of seeking consent;  (iii) it is reasonable to assume that most patients would have consented to our research proposal if they had been asked;  (iv) there will be an estimated 10,000+ patients, making it very difficult to obtain informed consent from this large number of patients, some of whom would have died while others may be in poor health as a result of their illness. Excluding patients who have died would seriously affect the scientific merit of the study;  (v) attempts to collect consent from patients (or their carers) who have experienced adverse events could induce unnecessary anxiety or distress;  (vi) de-identified data provided to the research team to be used in the analyses provides no potential threat to the well-being of the patients concerned because treatments have already occurred. The public benefit of this research is the generation of an improved understanding of health, heat and liveability that will lead to better planning and policies to improve social, economic, health and habitat conditions for climate sensitive cities.;  (vii) measures to ensure privacy and the confidentiality of information during the data linkage will follow well-established procedures developed by CHeReL. These methods are now well established as being protective of the privacy of individuals. | | | |

# DATA GOVERNANCE

|  |
| --- |
| **1A. Data Flow Diagram** |
|  |

|  |
| --- |
| **1B. Data Flow:**  Please outline the movement of data within and between sites including the methods of transfer used. This section should correspond with the flow diagram. |
| Data will be linked by the CHeReL Data Linkage Unit in accordance with standard models. Content variables and the PPN will be securely transferred to investigators by the data custodian or the CHeReL Data Integration Unit on the custodian’s behalf. NSW health Datasets will be uploaded from CHeReL to the Secured Unified Research Environment (SURE) (<https://www.saxinstitute.org.au/our-work/sure/using>), and AURIN datasets will be uploaded from AURIN data repositories to SURE by Dr Hao Chen as the project database developer. The upload process will be via the SURE secured Curated Gateway by a nominated Data Custodian from SURE. All inputs and outputs are vetted through the gateway for compliance and the SURE system records and archives all transactions for future reference, and the data cannot be copied, downloaded or transmitted by email or other means. Each workspace also has at least one nominated Curator from the Centre for Epidemiology and Evidence whose role is to review inbound files containing unit record data entering SURE, and approve (or reject) these files before they are made available to users within SURE. |
| **2. Data Storage, Access and Security:**  Where will the data be stored? (e.g., SURE, Local Server, etc).  Provide details of the security measures and access governance in place at each site. |
| All computer files containing original data used in the study will be kept in SURE virtual workspaces accessible only by the named investigators. To log into SURE, a username, passphrase and token is required. These are supplied to a user by the SURE team. Users also need to install a personal digital certificate on each computer from which they access SURE. The computer used to access SURE should be an employer-provided desktop or laptop computer. Furthermore, the local computer used to access SURE also needs to have a recent version of an approved anti-virus product installed and running.  Only researchers that are listed in this application and ethically approved to have access to linked unit record data will have access to the linked unit record data via SURE.  The SURE Workspace will be only be curated by the Centre for Epidemiology and Evidence (CEE) to ensure that no information is reported or released at the SA1 level. The Curators of the workspace are also responsible for reviewing and approving outgoing files from the SURE workspace. SA2 level is the lowest geographical breakdown that aggregate data will be approved to be released from the SURE workspace by the CEE. |
| **3. Use and Disclosure**  How and in what format will the data outputs be disseminated and used. How is confidentiality upheld. |
| Findings based on aggregated area-based health indicators from this AURIN project will be disseminated on the AURIN and RMIT Australian Urban Observatory (AUO) websites.  (i) Sensitive data will be stored within SURE and accessible only by the named investigators. Aggregated results of statistical analyses will be securely kept in AURIN and the AUO.  (ii) All publications and presentations arising from this project will not contain any identifying information, and no individual, medical practice or hospital will be identified or identifiable in such material.  (iii) Several mechanisms will be implemented to protect individuals and communities if necessary, i.e., there is a significant risk of re-identification. A smoothing algorithm will be applied to the generated indicators considering not only each studied area but also neighbouring areas. The resulting indicators will also be represented by intervals rather than exact values.  (iv) The data will be used in a manner consistent with its use by Government agencies when they report on patient health outcomes. However, our team has the technical expertise for using advanced statistical methods and providing insights beyond what is available in Government reports. The results of these enhanced analyses will be published in national or international peer reviewed journals, reported at scientific conferences, and be translated into lay language, with assistance from our consumer representatives, to be disseminated through the websites of the collaborating organisations.  Furthermore, the CEE will have sole curation of the SURE Workspace and curated gateway to ensure that no information is reported and released at SA1 level. The Curators of the workspace are responsible for reviewing and approving outgoing files from the SURE workspace and will only approve the release of SA2 level data as the lowest geographical breakdown aggregate data. |

|  |
| --- |
| **4. Data Retention:**  Specify the method and duration of data retention following completion of the project. |
| All unit record data obtained for this project will be archived in SURE and destroyed within 7 years after publication of the final output. This is in accordance with the NHMRC’s guidelines on data retention in their Australian code for the responsible conduct of research, which recommends a minimum of five years from the date of publication. Specifically, the project workspace in SURE will be removed from the server and backed up onto physical tape after the project, for the duration of 7 years.  On the other hand, Information generated as part of this project (aggregated results of statistical analyses or area-based aggregated health indicators) will be stored indefinitely as the only data that will be held at AURIN/UOM and RMIT Australian Urban Observatory (AUO). |
| **5. Data Disposal:**  Specify the method the information will be destroyed. |
| Once the 7 years has elapsed, all archived original unit record data files will be deleted from the SURE. After this time, all project data will be destroyed, and there will be no backup drives containing those original files. |

# ANALYSIS PLAN

|  |
| --- |
| Outcomes/exposures and covariates |
| *Describe the study outcome measures (primary and secondary) and include information on study exposure/s, covariates, and other factors. How are these defined based on the data and how do they reflect the aims? Please provide sufficient detail (200 word minimum).* |
| The primary outcomes of this study include:  - Spatial layers of heat vulnerability and liveability indicators  - New data assets (e.g. built environment dataset)  - Publications on the AURIN (aurin.org.au) and AUO (auo.org.au) platforms (e.g. visualisation platform for the indicators).  The process of this pilot project will be documented by Jupyter notebooks and made available to researchers for future reproduction. Access to the AURIN/AUO infrastructure data assets will be made available to registered users.  The exposures are heat and liveability which are determined from literature review using climate data from the Bureau of Meteorology.  Covariates are patient age, sex, comorbidities (determined from all diagnosis fields in Admitted Patient Data Collection using a minimum lookback period of 10 years), as well as socioeconomic, underlying chronic health conditions (diabetes, mental health, mobility, obesity, cardiovascular and respiratory disease), socioeconomic factors (social isolation, financial stress, housing stress, unemployment, education), environmental (lack of access to vegetation and water), as well as built environment (urban heat island).  Consultations with the Public Health Research Network (PHRN), School of Population of Global Health (The University of Western Australia), NSW Health Department, the Australian Digital Health Agency and the AusUrb-HI Advisory Committee have been held during development of the project methodology. This has also included guidance on the preparation and use of emergency department, hospitalisations and deaths data, as well as compare use of codes to record emergency department data: ICD and Snomed.  Ms Aileen dela Pena from the Australian Digital Health Agency has indicated that her agency will provide resources and technical support to this case study for working with Snomed codes used in the Emergency Department data. The Australian Digital Health Agency will not be directly involved with the analysis and will not have access to the data.  Outcomes are number and type (i.e. diagnosis, cause of death) of related emergency department presentations, hospitalisations and deaths related to heat exposure; number of hospitalisations and deaths in a year related to liveability |
| statistical analysis |
| *Provide a statistical analysis plan outlining how the aims/objectives will be met, the statistical methods to be used, and who will be carrying out the analysis. Provide sufficient detail (200 word minimum).* |
| The heat health vulnerability (HVI) and liveability indicators will integrate health outcomes data with individual datasets from environment, climate, built environment, underlying health conditions, as well as socioeconomic and demographic parameters from 2011, 2016 and 2021 ABS Census. Both indicators follow a framework approach based on weighted exposure, vulnerability, and adaptive capacity spatial layers. The figure below shows the four main steps of data processing and data analysis approaches, and each of the steps will be explained in detail afterwards.    In the first step, the collected raw datasets from AURIN and CHeReL will be pre-processed. Again, any processing involving the linked health datasets will be conducted in the SURE environment. Data cleansing will be done to identify and resolve any potential data gap, duplication, or inconsistency. Each of the dataset (which will be called component below) will be refined to match the study area, changed to grid-based format, and normalized by a numeric-based representation with value between 0 and 1 for each grid cell. For instance, a normalized income indicator value will be calculated based on the maximum income of the study area cells (as 1) as well as the lowest value (as 0). The derived numeric values will be used in the next step for generating sub-indicators (for Sensitivity, Adaptive capacity, and Exposure). Certain components may require more complex and/or potentially non-linear approaches to be normalized. For example, the relationship between daily health outcomes data provided by CHeReL will be examined in relation to extreme heat events during summer months, observing an increase in ED presentations, hospitalisations, and deaths during heatwaves days (+3 days post-heatwave) compared to non-heatwave days during the spring-summer period (September-March).  In the second and third step, the pre-processed datasets as components will be used to derive thematic layers called sub-indicators, which will then be used to derive health indicators in the next step. After the first step, the normalized grid-based values as input will now be tested for normality and dimensionality reduction approach, in order to identify how determinants correlate to each other and understand the amount of collinearity (25). The strongest features will be selected and combined into weighted spatial layers of Sensitivity, Adaptive Capacity, and Exposure. A Poisson regression approach will be employed with determinant weights derived using a pairwise comparison matrix (26). Finally, map algebra will be used to integrate the selected, weighted components to derive the sub-indicators using the formula . The result will be normalized to be between 0 and 1 again.  In the following steps, the resulting sub-indicators will be integrated once again, but this time to derive the output health indicators in the form of normalized heat vulnerability and liveability layers to express the “relative” vulnerability of locations across the chosen study areas as health indicators. Hotspot analysis will be used to investigate statistically significant locations of heat health vulnerability and liveability. A Z-score and corresponding P-value spatial fields will be generated, to which significance testing will be applied (27).  Finally, the result will be presented and communicated in various ways. Other than the traditional way of publishing research papers, a web application where users can adjust parameters such as weights and age groups, and view maps visualizing the output indicators (including health indicators and all sub-indicators) will additionally be developed. Since only aggregated data is used, the risk of re-identification is minimized. We keep in mind the key issues of maintaining data privacy and statistical stability, and therefore spatial smoothing method is leveraged to further reduce any risk of identifiability for specific individuals or communities.  The above introduced steps will be done by the AURIN and RMIT investigators/researchers listed above using QGIS, Python, R. Split analysis tasks will be carried out within and outside of SURE facility - e.g., processing of AURIN, environmental, climate data before loading into SURE facility where it will be analysed together with sensitive health data, which will only be used in SURE. Sensitive data is aggregated so that there are no individual health records identifiable by researchers.  A data verification and validation process will be conducted as part of QA/QC to ensure the accuracy and quality of the data, appropriate analysis techniques and adequate data management. An expert panel on the AusUrb-HI project advisory committee will supervise the process and establish a re-identification risk analysis procedure to ensure data privacy and security matters are addressed. The process also involves extensive consultations with independent external researchers and health data professionals to review and support the project methods and outcomes.  Note 1: the detailed analysis methodology plan of all the datasets and their integration into the overall health indicator is described in detail in the separately attached Analysis Plan document.  Note 2: 2021 Census data will be integrated in the study once it is made available by the Australian Bureau of Statistics. The expected publication date is August 2022, based on previous releases. |

# PROJECT FUNDING / SUPPORT

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Funding** | **Confirmed or Sought?** | | | **Amount of funding $** |
| **External Competitive Grant** | Confirmed | Sought | Not Sought | $400,000- |
| **Internal competitive grant** | Confirmed | Sought | Not Sought |  |
| **Sponsor** | Confirmed | Sought | Not Sought |  |
| **Researchers’ department or organisation** | Confirmed | Sought | Not Sought | $757,000- |

|  |  |
| --- | --- |
| **External Competitive** |  |
| Name of Grant/Sponsor | Australian Research Data Commons |
| **Internal Competitive** |  |
| Name of Grant/Sponsor |  |
| **Sponsor** |  |
| Name of Grant/Sponsor |  |
| **Researchers Department or Organisation** |  |
| Name of Grant/Sponsor | Australian Urban Research Infrastructure Network |
| Name of Grant/Sponsor | Population Health Research Network |

# REFERENCES

1 Barnett, G, Beaty, RM, Chen, D, McFallan, S, Meyers, J, Nguyen, M, Ren, Z, Spinks, A & Wang, X 2013, Pathways to climate adapted and healthy low income housing, National Climate Adaptation Research Facility, Gold Coast, 95 pp.

2 Bodilis, C., Yenneti, K., Hawken, S. (2018): Heat Vulnerability Index for Sydney. Faculty of Built Environment, UNSW Sydney.

3 Centers for Disease Control and Prevention. (2020). National Environmental Public Health Tracking Network. CDC’s National Environmental Public Health Tracking Program. https://ephtracking.cdc.gov/

4 Climate Council. (2016). The silent killer: Climate change and the health impacts of extreme heat. Retrieved from https://www.climatecouncil.org.au/uploads/b6cd8665c633434e8d02910eee3ca87c.pdf

5 Coates, L., Haynes, K., O’Brien, J., McAneney, J., & De Oliveira, F.D. (2014). Exploring 167 years of vulnerability: An examination of extreme heat events in Australia 1844–2010. ScienceDirect.com | Science, health and medical journals, full text articles and books. https://www.sciencedirect.com/science/article/pii/S1462901114000999

6 CRC for Water Sensitive Cities. (2016). Urban Heat, Heat Reduction and Public Health. Retrieved from https://watersensitivecities.org.au/wp-content/uploads/2016/08/Urban-Heat-Briefing\_Final.pdf

7 Department of Health and Human Services. (2021). Heat health plan for Victoria. Retrieved from https://www.health.vic.gov.au/publications/heat-health-plan-for-victoria

8 Department of Planning, Industry and Environment. (2019). NSW heat vulnerability index to ABS statistical area level 1 2016 - NSW planning Portal. Retrieved from: https://www.planningportal.nsw.gov.au/opendata/dataset/nsw-heat-vulnerability-index-to-abs-statistical-area-level-1-2016

9 Haines, A., & Ebi, K. (2019). The Imperative for Climate Action to Protect Health. New England Journal of Medicine, 380(3), 263–273

10 L. Jian, PhD, B. Scalley, MPH, A. Xiao, PhD, J. Nairn, MS, T. Spicer, BS, P. Somerford, MPH, B. Ostendorf, T. Weeramanthri, MD, Is Excess Heat Factor a Good Indicator for Assessing Heatwave Related Health Outcomes in Western Australia?., International Journal of Epidemiology, Volume 44, Issue suppl\_1, October 2015, Page i65, https://doi.org/10.1093/ije/dyv097.241

11 Loughnan M., Tapper N., Phan T., Lynch K., Mcinnes J. (2013). A spatial vulnerability analysis of urban populations during extreme heat events in Australian capital cities Final Report.

12 Loughnan, M., J. Tapper, N., Phan, T., McInnes, J. (2014). Can a spatial index of heat-related vulnerability predict emergency service demand in Australian capital cities? International Journal of Emergency Services, 3(1), 6–33. https://doi.org/10.1108/ijes-10-2012-0044

13 Nairn, J., Beaty, M., & Varghese, B. M. (2021). Australia’s Black Summer heatwave impacts. Australian Journal of Emergency Management, 36(1), 17–20.

14 Nairn, John & Fawcett, Robert. (2014). The Excess Heat Factor: A Metric for Heatwave Intensity and Its Use in Classifying Heatwave Severity. International journal of environmental research and public health. 12. 227-53. 10.3390/ijerph120100227.

15 NCEconomics. (2018). Heatwaves in Victoria: A Vulnerability Assessment. Department of Environment, Land, Water and Planning

16 Nicholls L., McCann H., Strengers Y. & Bosomworth K. (2017). Heatwaves, Homes & Health: Why household vulnerability to extreme heat is an electricity policy issue. Centre for Urban Research. RMIT University, Melbourne.

17 NSW Health. (2020, December). Heat is a health risk - beat the heat. https://www.health.nsw.gov.au/environment/beattheheat/Pages/default.aspx

18 Ogie, R., & Pradhan, B. (2020). Social vulnerability to natural hazards in Wollongong: comparing strength-based and traditional methods. Australian Journal of Emergency Management, 60–68.

19 Park, C., Ha, J., & Lee, S. (2017). Association between Three-Dimensional Built Environment and Urban Air Temperature: Seasonal and Temporal Differences. Sustainability, 9(8), 1338. https://doi.org/10.3390/su9081338

20 Pfautsch, S., Rouillard, S. (2019) Benchmarking heat in Parramatta, Sydney's Central River City. Western Sydney University, 56 p.

21 Physical Environment Research Network. (2021). Reducing Illness and Lives Lost from Heatwaves: Project Report. Canberra, Australia: Australian Government Data Integration Partnership for Australia. Retrieved from: https://www.pean.gov.au/sites/default/files/2021-07/BOM%20%282021%29%20Heatwaves%20report.pdf

22 Sun C, Hurley J, Amati M, Arundel J, Saunders A, Boruff B, Caccetta P (2019) Urban Vegetation, Urban Heat Islands and Heat Vulnerability Assessment in Melbourne, 2018. Clean Air and Urban Landscapes Hub, Melbourne, Australia

23 Toloo, G., Yu, W., Aitken, P., FitzGerald, G., & Tong, S. (2014). The impact of heatwaves on emergency department visits in Brisbane, Australia: a time series study. Critical Care, 18(2), R69. https://doi.org/10.1186/cc13826

24 Wondmagegn, B. Y., Xiang, J., Dear, K., Williams, S., Hansen, A., Pisaniello, D., Nitschke, M., Nairn, J., Scalley, B., Varghese, B. M., Xiao, A., Jian, L., Tong, M., Bambrick, H., Karnon, J., & Bi, P. (2021). Impact of heatwave intensity using excess heat factor on emergency department presentations and related healthcare costs in Adelaide, South Australia. Science of The Total Environment, 781, 146815. https://doi.org/10.1016/j.scitotenv.2021.146815

25 Conlon, K. C., Mallen, E., Gronlund, C. J., Berrocal, V. J., Larsen, L., & O’Neill, M. S. (2020). Mapping Human Vulnerability to Extreme Heat: A Critical Assessment of Heat Vulnerability Indices Created Using Principal Components Analysis. Environmental Health Perspectives, 128(9). https://doi.org/10.1289/ehp4030

26 Guerreiro, S.B., Dawson, R.J., Kilsby, C., Lewis, E., and Ford, A. (2018). Future heat-waves, droughts and floods in 571 European cities. Environmental Research Letters, 13 (3), 034009.

27 Thacker, S., Barr, S., Pant, R., Hall, J.W., and Alderson, D. (2017) Geographic Hotspots of Critical National Infrastructure. Risk Analysis, 37(12), 2490-2505.

28 Estoque, R. C., Ooba, M., Seposo, X. T., Togawa, T., Hijioka, Y., Takahashi, K., &amp; Nakamura, S. (2020). Heat health risk assessment in Philippine cities using remotely sensed data&nbsp;and social-ecological indicators. Nature Communications, 11(1). https://doi.org/10.1038/s41467-020-15218-8

|  |
| --- |
| **DATA REQUEST SECTION** |

# DATA LINKAGE

TYPE OF PROJECT

|  |
| --- |
| *Does this project involve datasets not within* [*CHeReL’s Master Linkage Key (MLK)*](https://www.cherel.org.au/master-linkage-key)*?* |
| ☐ No - Please complete Section A & B.1  **🗸** Yes - Please complete Section A & B.1 & B.2 |

|  |
| --- |
| *Does the project involve family linkage (e.g., mother-baby,* *mother-baby-sibling)?* |
| **🗸** No - Please complete Section A & B (B.1 & B.2 as required)  ☐ Yes - Please complete Section A & C (C.1 & C.2 as required) |

|  |
| --- |
| *During which calendar year do you require your first data linkage?* |
| 2022 |

|  |
| --- |
| *Will updates to your Linkage be required?* |
| No  Yes (please specify how/when *e.g., “annual extracts until 2028, one further extract in 2025”)* |
|  |

#### (SECTION A) COHORT AND RESTRICTIONS FOR LINKAGE

|  |  |  |  |
| --- | --- | --- | --- |
| *Approximately how many individuals are in the cohort(s)?* | | | |
| Estimated 10,000+ | | | |
| *List all datasets to be used to define your cohort(s) for linkage* | | | |
| ***Dataset*** | ***Start Date*** | ***End Date*** | ***Any additional information*** |
| NSW Admitted Patient Data Collection | 2016 | 2021 | The cohort is population of all ages living (at the time of ANY presentation / admission / death) in major urban areas in NSW (please see attachment titled Health Liveability Study Areas with map and list of identified areas using SA1s, see excel file “study area geocodes”) who presented to a NSW Emergency Department, admitted to hospital or died between 1 Jan 2016 and 31 Dec 2021.  For each identified patient, we require all their Emergency Department and hospitalisation data dating back to 1 Jan 2011 in order for us to determine the medical history with a minimum of a 10-year look-back. |
| NSW Registry of Births, Deaths and Marriages | 2016 | 2021 |
| NSW Emergency Department Data Collection (from 2005) | 2016 | 2021 |
| *If your cohort is defined by diagnosis/procedure codes (e.g., ICD, ACHI, SNOMED) please select:* | | | |
| Principal diagnosis/procedure codes only OR  Any of the multiple diagnosis/procedure codes within a record | | | |
| *Please describe codes required and attach a list in excel format (filename – abbreviation\_cohort\_code\_list\_date.xlsx)* | | | |
| *(e.g., from the APDC - all separations during 2002 with the following ICD diagnoses)*  *N/A* | | | |

#### (SECTION B) What data collections are being linked to your cohort?

|  |
| --- |
| *Do you require:* |
| All linked records required for these individuals (e.g., all APDC records relating to all morbidity)  Only records relating to the specified condition (e.g., APDC records relating only to CVD). Please specify:  The following data is readily available in AURIN and will be linked to the requested data (except for 2021 ABS Census data which will be ingested after the release in June 2022). *Please note: all the AURIN datasets are aggregated to some area level without any person-level record and thus are not person identifiable*   |  | | --- | | * ABS - SEIFA Index (SA1) for 2011, 2016 and 2021\* (awaiting release) | | * ABS – Census Data G01 Selected Person Characteristics by Sex (SA1) for 2011, 2016 and 2021\* (awaiting release) * ABS – Census Data G07 Indigenous Status by Age by Sex (SA1) for 2011, 2016 and 2021 \* (awaiting release) Note: used as back-up to control statistical model. * ABS - Census DataG39 Dwelling Structure by Household Composition and Family Structure (SA1) for 2011, 2016 and 2021\* (awaiting release) | | * ABS - Census DataG21 Unpaid Assistance to a Person with a Disability by Age by Sex (SA1) for 2011, 2016 and 2021\* (awaiting release) | | * ABS - Census Data G22a Unpaid Child Care by Age by Sex (SA1) for 2011, 2016 and 2021\* (awaiting release) | | * ABS - Census Data G22b Unpaid Child Care by Age by Sex (SA1) for 2011, 2016 and 2021\* (awaiting release) | | * ABS - Census Data G23a Relationship in Household by Age by Sex (SA1) for 2011, 2016 and 2021\* (awaiting release) | | * ABS - Census Data G23b Relationship in Household by Age by Sex (SA1) for 2011, 2016 and 2021\* (awaiting release) | | * ABS - Census Data G43a Labour Force Status by Age by Sex (SA1) for 2011, 2016 and 2021\* (awaiting release) | | * ABS - Census Data G43b Labour Force Status by Age by Sex (SA1) for 2011, 2016 and 2021\* (awaiting release) | | * ABS - Census Data G45a Labour Force Status by Sex of Parents by Age of Dependent Children for One Parent Families (SA1) for 2011, 2016 and 2021\* (awaiting release) | | * ABS - Census Data G45b Labour Force Status by Sex of Parents by Age of Dependent Children for One Parent Families (SA1) for 2011, 2016 and 2021\* (awaiting release) | | * ABS - Census Data G46a Non-School Qualification-Level of Education by Age by Sex (SA1) for 2011, 2016 and 2021\* (awaiting release) | | * ABS - Census Data G46b Non-School Qualification-Level of Education by Age by Sex (SA1) for 2011, 2016 and 2021\* (awaiting release) | | * ABS - Census Data G16a Highest Year of School Completed by Age by Sex (SA1) for 2011, 2016 and 2021\* (awaiting release) | | * ABS - Census Data G16b Highest Year of School Completed by Age by Sex (SA1) for 2011, 2016 and 2021\* (awaiting release) | | * ABS - Census Data G33 Tenure and Landlord Type by Dwelling Structure (SA1) for 2011, 2016 and 2021\* (awaiting release) | | * ABS - Census Data G28 Total Family Income (Weekly) by Family Composition (SA1) for 2011, 2016 and 2021\* (awaiting release) * OSM – OpenStreetMap – Lines (Australia) 2017, 2018, 2020 * OSM – OpenStreetMap – RoadNetwork (Australia) 2020 * Geoscience Australia – NEXIS – Residential Building Exposure (SA1) 2015, 2016, 2017, 2020 | | * ABS - Data by Region Income (Including Government Allowances) (SA2) 2011-2019 | | * ABS - Data by Region Family & Community (SA2) 2011-2018 | | * ABS - Data by Region Health & Disability (SA2) 2011-2018 | | * ABS - Data by Region Persons Born Overseas (SA2) 2011-2016 | | * ABS - Regional Population Population Estimates by Age and Sex (SA2) 2019, 2018, 2017 | | * NATSEM Social and Economic Indicators - Synthetic Estimates (SA2) 2016 | | * NATSEM Financial Indicators - Synthetic Estimates (SA2) 2016 * NATSEM Dependency rate (SA2) 2016 | | * PHIDU Prevalence of selected chronic diseases and conditions (estimates) (PHA) 2011-2012, 2017-2018 | | * PHIDU Prevalence of selected health risk factors (estimates) (PHA) 2014-2015, 2017-2018 * PHIDU Home and Community Care Program (PHA) 2014-2015 * PHIDU Admissions - Same-day Renal Dialysis (PHA) 2017-2018 * PHIDU Housing and Transport (PHA) 2016-2020 * PHIDU Income Support Recipients (PHA) 2017-2020 * PHIDU Personal and Financial Stressors (PHA) 2014 * PHIDU Access to Services (PHA) 2014 * PHIDU Home and Community Care Program (PHA) 2014-2015 * National Health Services Directory (NHSD) (Point) 2020 | |  |   The following datasets will be provided to AURIN from external sources and extracted to SA1 level data (e.g. raster, grid, pointcloud data)   * BOM - Australian Hydrological Geospatial Fabric (AHGF) – (2011 – 2020) * MODIS – Normalized Difference Vegetation Index (NDVI) – (2000 onwards) * MODIS – Land Surface Temperature (LST) – (2011 – 2020) * Landsat 8 – Land Surface Temperature (LST) * BOM – Daily temperature data (2011 – 2020) * Microsoft – Building footprints (Australia) – 2018 * NSW Spatial Services - AHD – Pointclouds (2011 – 2020) * Geoscience Australia – DEA Land Cover (Landsat) – (2011 – 2020) |

(SECTION B.1) EXTRACT FROM CHEREL MLK COLLECTIONS

|  |  |  |  |
| --- | --- | --- | --- |
| **MLK Data Collections** | | Dates | |
| From  (e.g., Jul 2001) | To  (e.g., Dec 2020 / Latest Available) |
|  | NSW Admitted Patient Data Collection (from Jul 2001)  Based on:  Admission date **OR**  Separation date | 2006 | latest available |
|  | NSW Emergency Department Data Collection (from 2005) | 2006 | latest available |
|  | NSW Non-Admitted Patient Data Collection (from Jul 2015) |  |  |
|  | NSW Perinatal Data Collection (from 1994) |  |  |
|  | NSW Perinatal Death Review Database (from 2000) |  |  |
|  | NSW RBDM Birth Registrations (from 1994) |  |  |
|  | NSW RBDM Death Registrations (from 1985) | 2006 | Latest available |
|  | NSW Cause of Death Unit Record File (from 1985) | 2006 | latest available |
|  | NSW Mental Health Ambulatory Data Collection (from 2001) |  |  |
|  | NSW Central Cancer Registry (from 1972) |  |  |
|  | NSW Pap Test Register (from Jul 1996) |  |  |
|  | BreastScreen NSW (from Jan 1988) |  |  |
|  | NSW Ambulance (from 2009) |  |  |
|  | The 45 and Up Study | Baseline (2006-09) SEEF (2010)  Follow-up (2012‑15)  Follow-up (2018‑20) | |
|  | NSW Notifiable Conditions Information Management System  (from 1993) |  |  |
|  | NSW Notifiable Conditions Information Management System – COVID-19 Case (from 2020) |  |  |
|  | [NSW ANZDATA](http://www.anzdata.org.au/documents/pdf/ANZDATA-data-dictionary-2016_v0.4_20170626.pdf) (from 1963) |  |  |
|  | NSW Australian Early Development Census (from 2009) |  |  |
|  | NSW Bureau of Crime Statistics and Research Reoffending Data Collection (BOCSAR) (from 1994) |  |  |
|  | NSW Controlled Drugs Data Collection (CoDDaC) (from 1985) |  |  |
|  | ACT Admitted Patient Collection (from Jul 2004)  Based on:  Admission date **OR**  Separation date |  |  |
|  | ACT Cancer Registry (from 1994) |  |  |
|  | ACT Emergency Department Data Collection (from Jul 2005) |  |  |
|  | ACT Perinatal Data Collection (from 1997) |  |  |
|  | ACT Notifiable Diseases Register (from 2000) |  |  |
|  | ACT BDM Death Registrations (from 1997) |  |  |
|  | ACT Cause of Death Unit Record File (from 2006) |  |  |
|  | ACT BDM Birth Registrations (from 1997) |  |  |
|  | ACT Australian Early Development Census (from 2009) |  |  |
|  | ACT Kindergarten Health Check (from 2014) |  |  |
|  | ACT ANZDATA (from 1963) |  |  |
|  | ACT Ambulance (from 2011) |  |  |

(SECTION B.2) OTHER COLLECTIONS (NOT HELD IN CHEREL MLK)

Please include any non-MLK datasets that **will be linked by CHeReL** (e.g., your own cohort dataset etc). Please list all other non-MLK datasets at Section 7- METHODS – Secondary Data.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Dataset 1 | | | | | | |
| Name and/or brief description of dataset |  | | | | | |
| Number of records |  | Year span of dataset | | | |  |
| Custodian name |  | Email/Phone no. | | | |  |
| Agency Type | …State / Territory | …Commonwealth | | | | …Private Sector |
| Who should CHeReL contact to source data? |  | |  | | | |
| Please list personal identifiers available for linkage |  | | | | | |
| Dataset 2 | | | | | | |
| Name and/or brief description of dataset |  | | | | | |
| Number of records |  | Year span of dataset | | | |  |
| Custodian name |  | Email/Phone no. | | | |  |
| Agency Type | …State / Territory | …Commonwealth | | | | …Private Sector |
| Who should CHeReL contact to source data? | Name | | | | Email | |
| Please list personal identifiers available for linkage |  | | | | | |
| Dataset 3 | | | | | | |
| Name and/or brief description of dataset |  | | | | | |
| Number of records |  | Year span of dataset | | | |  |
| Custodian name |  | Email/Phone no. | | | |  |
| Agency Type | …State / Territory | …Commonwealth | | | | …Private Sector |
| Who should CHeReL contact to source data? | Name | | | Email | | |
| Please list personal identifiers available for linkage |  | | | | | |