**ANALYSIS PLAN**

**Introduction**

Extreme heat poses a significant threat to population health, with heatwaves resulting in the greatest amount of preventable morbidity and mortality compared to any other natural hazard in Australia 1. Extreme heat has adverse effects on population health, as well as adds pressure on health services due to sudden surges in demand 2.

**Indicator generation framework**

The AURIN team will develop a SA1 level Heat Vulnerability Indicator (HVI) pilot project based on the Intergovernmental Panel on Climate Change (IPCC) 2014 framework 3 to assess the risks of extreme events driven by climate change on population health in the five largest cities in the state of New South Wales: Sydney, Wollongong, Newcastle – Maitland, Tweed Heads and Albury.

HVI is the sum of weighted scores for sensitivity, exposure and adaptive capacity to heat as sub-indicators; meaning that the outcome of a heatwave is a function of these three factors (Figure 1). Sensitivity refers to characteristics of the population that influence risk of suffering from heat exposure such as socio-economic (living conditions, employment, education), demographic (e.g., age, sex) and health conditions (e.g., comorbidities obtained from linked data). Exposure refers to characteristics of the built environment like building height and density, impervious surfaces that can exacerbate the response to heatwaves. Adaptive capacity refers to the characteristics of the natural environment that mitigate the effects of heat such as vegetation and water coverage.

**Identifying components and variables for sub-indicators**

Each sub-indicator is made up of several components. The sensitivity sub-indicator consists of the following components: health (including emergency department presentations, hospitalisations, and deaths obtained from linked administrative data), social isolation, tenure, age groups, housing composition, employment, education, SEIFA and restricted mobility (obtained from the Australian Bureau of Statistics 2011, 2016 and 2021 Census). For each component, we will create one or more variables. For example, the health component will consist of variables such as “proportion of people in the geographical area (i.e., SA1) with history of heart disease” or “proportion with >3 hospitalisations in the previous year” determined from linked emergency department (ED), inpatient and mortality data.

As a result, for this study we require linked health data including all conditions leading to hospital presentation or death. Heat-related illness is a spectrum of conditions rather than just heatstroke.4 Recording of these conditions in administrative data can vary depending on coding guidelines in each jurisdiction. In addition, deaths may not be reported as heat-related if the cause of death was attributed to an underlying health condition that worsened during a heatwave. 5 For this study, cause-specific outcomes are conditions that are over-represented in the principal discharge diagnosis field during the heatwave (e.g. myocardial infarction, stroke, cardiovascular disease (CVD), renal disease, diabetes and general non-external causes 2).

**Identification of heatwave periods**

The Bureau of Meteorology defines heatwaves as unusually high minimum and maximum temperatures that persist for 3 or more consecutive days at a certain location. We will use daily temperature data, obtained from the Bureau of Meteorology and Google Earth Engine, at a fine spatial resolution to identify heatwaves and categorise different responses and health outcomes to heatwaves in New South Wales. We will use linked health data (ED presentations, hospitalisations, and deaths) to examine population sensitivity to heat during spring-to-summer months when extreme heat is more likely to occur. We will compare health outcomes and conditions during heatwave days compared to non-heatwave days (Figure 2).

**Identification of health outcomes from heatwaves**

The study requires person-level deidentified linked health data to ascertain the number and cause of ED presentations, hospitalisations, and deaths during the heatwave. The methodology is described as follows:

1. We will use linked health data to establish the health component of the sensitivity sub-indicator. Variables will include clinical history of people from each SA1 who presented to ED or were hospitalized during a comparable non-heatwave period or the cause-of-death during this period. We will determine the Charlson comorbidity score 6 and specific medical conditions (e.g., proportion with heart disease, lung disease, cancer) using a 5-year look-back for patients hospitalised during the comparable non-heatwave period (Figure 2). We will use statistical methods (e.g., latent class analysis, principal component analysis) to identify variables that are correlated with one or multiple other independent variables in order to reduce the amount of collinearity. This process will potentially reduce the number of variables used for each sub-indicator, keeping only the most influencing and dependent variables as possible (Step 1, Figure 3).

2. We will explore normalising each variable to a value between 0 and 1 or categorise based on distribution within standard deviation (e.g., within -1 SD, 2 SD, +1 SD, +2 SD), which will allow us to compare multiple types of variables with different scales of measurement. The normalised values or categories will be summated to generate a sub-indicator score, which will be normalised to a value between 0 and 1 (Step 2, Figure 3).

3. The three sub-indicator scores together with heat intensity will be entered in a Poisson multivariable regression where the outcome is number of ED presentations, hospitalisations or deaths from the linked health data (Step 3, Figure 3).

4. The sub-indicator weights from the Poisson regression will be used to calculate the HVI score at the SA1 level (Step 4, Figure 3):

HVI score = β1(sensitivity) + β2(adaptive capacity) + β3 (exposure)

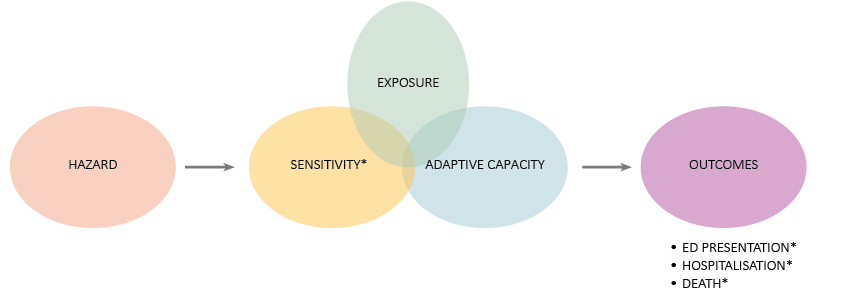
5. We will repeat Steps 1 to 4 using different combinations of sub-indicators and components in Step 1. The empirically selected appropriate combination will be selected based on model fit statistics in the Poisson regression.

6. The reporting of statistical data by small geographical areas needs to consider two key issues: (i) data privacy and (ii) statistical stability. Data privacy relates to the responsibility to protect the identity of individuals in the data, and ensure that this is not compromised by the release of that data for reporting purposes. Statistical stability relates to the inherent random error that occurs more commonly with small numbers of cases; the smaller the numbers, the more they fluctuate, potentially leading to incorrect interpretation. When working with geographical data where spatial location is additionally relevant, these issues may become more significant. To address both these issues for geographical data, we will use a specific statistical method known as “spatial smoothing” (Step 6 in Figure 3). While standard methods typically only adjust for age and sex in each area, spatial smoothing recognises the geographical structure of the data and includes data from the neighbouring geographical areas when calculating the spatial estimates. This additional data provides greater stability to the estimates. In addition, because the spatial estimates are modelled, rather than observed, spatial smoothing reduces any risk of identifiability for specific individuals. Smoothed estimates are designed to reflect the real differences in the underlying rate or risk between areas without presenting actual personalised data.

For this study, the spatial smoothing will be adjusted for age, sex and comorbidities (determined from the principal and secondary hospital discharge diagnosis fields). Figure 4 is an example of spatial smoothing at two different levels of distance function (in this example, s=0.1 and s=0.5 are standardized values calculated based on the distance function used). The final sensitivity to heat indicators will be reported as summary statistics (e.g., quartiles, quintiles) or range of values rather than absolute numbers.

7. The resulting heat health vulnerability indicator will express the 'relative' vulnerability of locations across the chosen study area. To investigate statistically significant locations of heat health vulnerability, hotspot analysis will be performed to generate a Z-score and corresponding P-value spatial fields, to which significance testing can be applied.

**Figure 1: Relationship between exposure, explanatory factors to heatwave and outcome health sub-indicator will be determined from linked health data; \* determined from linked health data**



**Figure 2: Scheme illustrating use of linked health data to determine (i) outcomes during heatwave and (ii) background medical history during a comparable non-heatwave period.**

**Heatwave period**

Use linked health data to determine outcomes of ED presentations, hospitalisations and/or deaths

5-year look-back

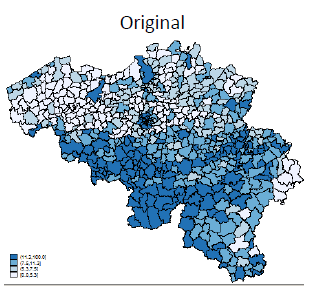
**Comparable non-heatwave period**

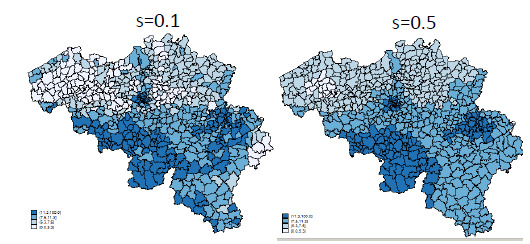
Use linked health data to determine background medical history of patients presenting to ED, admitted to hospital (with 5-year look-back) or deaths

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  | **SUB-INDICATORS:** |  |  |  |  |
| Step |  |  |  | **SENSITIVITY** |  | **ADAPTIVE CAPACITY** |  | **EXPOSURE** |  | **HAZARD** |
|  |  |  |  |  |  |  |  |  |  |  |
| 1 |  | Normalise each sub-indicator |  | Health (***linked data***) Social isolation Tenure Age groups Housing composition Employment Education SEIFA Restricted mobility |  | Land use Built environment  Road density |  | Vegetation coverage Water coverage |  | Heat intensity *(no heatwave, low intensity, moderate intensity, high intensity*) |
|  |  |  |  |  |  |  |  |  |  |  |
| 2 |  | Summate (∑) and normalise to generate indicator score |  | ∑ normalised sub-indicators to generate 'sensitivity' score. Normalise summated score |  | ∑ normalised sub-indicators to generate 'adaptive capacity' score. Normalise summated score |  | ∑ normalised sub-indicators to generate 'exposure' score. Normalise summated score |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| 3 |  | Obtain weights (β1, β2, β3) for each indicator |  | Model number of ED presentation/hospitalisation/death (***linked data***) as a function of ‘sensitivity', 'adaptive capacity', 'exposure' summated scores and ‘hazard’ using Poisson regression | | | | | | |
|  |  |  |  |  |  |  |  |  |  |  |
| 4 |  | Heat vulnerability score |  | Summate using weights from Step 3 to generate 'Heat vulnerability score' i.e. Heat vulnerability score = β1(sensitivity) + β2(adaptive capacity) + β3 (exposure) | | | | | | |
|  |  |  |  |  |  |  |  |  |  |  |
| 5 |  | Repeat Steps 1-4 |  | Iterative process using different combinations of sub-indicators in Step 1. The appropriate combination will be selected based on model fit statistics in the Poisson regression | | | | | | |
|  |  |  |  |  |  |  |  |  |  |  |
| 6 |  | Spatial smoothing and categorization |  | Indicator and heat vulnerability scores will be spatially smoothed and presented as categories rather than specific values. | | | | | | |
| 7 |  | Hotspot analysis and significance testing |  | Hotspot analysis will be used to investigate statistically significant locations of heat health vulnerability. A Z-score and corresponding P-value spatial fields will be generated, to which significance testing will be applied | | | | | | |

**Figure 3: Flow diagram of the analysis plan**

**Figure 4: Example of spatial smoothing - adapted from Deschacht 2016 7**





**Results**

The output of the analysis will be the Heat Vulnerability Indicator and the three weighted subindicator layers of Sensitivity, Adaptive Capacity and Exposure (for each year within the study period 2016 to 2021), which will allow us to understand the relationship between subindicators that contributed to a specific vulnerability outcome and measures. The results will be expressed in values between 0 and 1 (0 = low vulnerability and 1= high vulnerability).

**Figure 5: Example of output table for visualisation (using fictitious data and subindicator weights)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **SA1 Code** | **Heat Vulnerability Index** | **Sensitivity** | **Exposure** | **Adaptive Capacity** | **Year** |
| 21401137140 | 0.47 | 0.5 | 0.85 | 0.2 | 2016 |
| 21001123104 | 0.68 | 0.85 | 0.92 | 0.1 | 2016 |
| 20302103904 | 0.26 | 0.4 | 0.8 | 0.9 | 2016 |
| 21401937149 | 0.18 | 0.4 | 0.4 | 0.7 | 2016 |

The output table will contain SA1 codes that can be mapped to ABS boundaries to visualise the results in form of spatial layers (Figure 6).

The findings will be discussed in research papers and include basic descriptive statistics using broad levels of indicator scores (e.g. ranking, quintiles) to compare between different groups of indicators.

**Figure 6: Example of visualisation of spatial layers for Statistical Areas 1**

A picture containing text, envelope, businesscard

Description automatically generated

**Bibliography**

1. Coates L, Haynes K, O’Brien J et al. Exploring 167 years of vulnerability: An examination of extreme heat events in Australia 1844–2010. *Environmental Science & Policy* 2014; 42: 33–44. doi: doi.org/10.1016/j.envsci.2014.05.003

2. Toloo G, Yu W, Aitken P, et al. The impact of heatwaves on emergency department visits in Brisbane, Australia: A time series study. *Critical care* 2014; doi: 10.1186/cc13826.

3. Cardona, O.D., M.K. van Aalst, J. Birkmann, M. Fordham, G. McGregor, R. Perez, R.S. Pulwarty, E.L.F. Schipper, and B.T. Sinh, 2012: Determinants of risk: exposure and vulnerability. In: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation [Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor, and P.M. Midgley (eds.)]. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change (IPCC). Cambridge University Press, Cambridge, UK, and New York, NY, USA, pp. 65-108.

4. Jacobson AA, Raukar NP. The 2020 ICD-10 definitions of heat illness go beyond heatstroke 2020 [updated 16 Jun 202024 Sep 2022]. Available from: https://www.acepnow.com/article/the-2020-icd-10-definitions-of-heat-illness-go-beyond-heatstroke/2/ Accessed 24 Sep 2022.

5. Doctors for the Environment Australia 2020, Heatwaves and Health in Australia. Available from: <https://www.dea.org.au/wp-content/uploads/2021/01/DEA-Fact-Sheet_HeatwavesWEB.pdf> Accessed 26 Sep 2022.

6. Holman CD, Preen DB, Baynham NJ, et al. A multipurpose comorbidity scoring system performed better than the Charlson index. *J Clin Epidemiol* 2005;58(10):1006-14.

7. Deschacht N. Creating smoothed maps with the help of the command spmap, 2016. Available from: <https://www.stata.com/meeting/belgium16/slides/belgium16_deschacht.pdf> Accessed 24 Sep 2022.

8. Quan H, Sundararajan V, Halfon P, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Medical Care* 2005;43(11):1130-9.

9. Whitaker HJ, Hocine MN, Farrington CP. The methodology of self-controlled case series studies. *Stat Methods Med Res* 2009;18(1):7-26. doi: 10.1177/0962280208092342

10. Lopez D, Nossent J, Warren R, et al. Abstract 11765: Admission to hospital with acute gout is associated with a short term increased risk of major adverse cardiovascular events: a longitudinal, population-level linked health data study in Western Australia. *Circulation* 2021;144(Suppl\_1):A11765-A65. doi: doi:10.1161/circ.144.suppl\_1.11765

11. Lopez D, Preen D, Raymond W, et al. Risk of a major adverse cardiovascular event (MACE) following first-ever hospitalisation for acute gout: a Western Australian population-level linked data study. *International Journal of Population Data Science* 2022;7(3) doi: https://doi.org/10.23889/ijpds.v7i3.1805 [published Online First: 25 Aug 2022]