

2022/ETH00905 HREA Review Response Cover Letter

30/04/2023

Dear HREC Committee,

I am writing to respond to the comments and concerns raised by the Ethics Committee regarding our recent submission. We appreciate the committee's detailed review and feedback. Please find below our responses to your queries of 21st April 2023.

1. **As stated in the analysis plan of the protocol, heat vulnerability and liveability will be “determined from literature review using climate data from the Bureau of Meteorology”. It was unclear to the Committee how this stated method would develop the indicators of heat and vulnerability, therefore please update the protocol to elaborate upon and provide further details around how this will be done.**

Response: The sentence from the protocol has been modified to: “The exposures are heat and liveability which are determined using climate data. We will acquire land surface temperature (LST) data from both the Bureau of Meteorology (BOM) and MODIS (using the Google Earth Engine API) to minimize data gaps. The well-established Excess Heat Factor (EHF) algorithm proposed in the literature will be applied to determine heatwave periods during high-risk months (December to February end) for studying heat exposure and liveability.”

Then, as described in the detailed analysis plan, the retrospective cohort analysis of patients in selected NSW regions in the study period will be conducted, examining health outcomes during heatwave and non-heatwave periods. Using these insights, a heat vulnerability index will be created combining environmental and social factors.

2. **The Committee noted that the COD URF has been omitted from Section A of the Data Request/Application for Data. Please update.**

Response: In this study, we will only be using the death registry, alongside EDDC and APDC, for cohort identification. The COD URF is not required for our cohort selection, as deaths have been determined from the death registry.

3. **The Committee also noted that the study period includes the bushfire season of 2019-2020 and the lockdown periods during the COVID-19 pandemic. Please make comment on how the study will control for these potentially confounding factors.**

Response: The study area was not significantly impacted by COVID lockdown in general. COVID restrictions on movement in NSW occurred only from late March to June 2020 and not during summer or periods of greatest heat, which are January and February.

Health outcomes collected over the study period will be used to investigate large trends in results across this period. The bushfire period was relatively brief within this extended multi-year period and even shorter during the summer months. However, we acknowledge that the impact of air quality on deaths was significant in Sydney, with 109, or 38% of total deaths, brought forward by short-term exposure to PM2.5. The additional deaths will be considered in mortality analysis with COD URF data. Alternatively, if we think it is difficult to identify bushfire-related mortality records during the analysis when examining the data, we will exclude this year. We will also perform sensitivity analyses excluding the bushfire period to determine if the results are consistent.

We will engage with epidemiologists on the AusUrbHi expert advisory panel and seek advice on the most appropriate statistical method to account for the anomaly.

- 4. The Committee queried whether the NSW APDC would be a reliable data source for information on underlying chronic health conditions. Please comment on whether this is intended to be the direct source for this information, or if it will be drawn from other datasets. Further, please elaborate upon how it will be determined whether an admission, presentation or death is heat related.**

Response: During the study period, information on chronic health conditions will be supplemented using PHIDU datasets, specifically the Prevalence of Chronic Diseases and the Prevalence of Selected Health Risk Factors datasets. These datasets have been collected and are listed in the protocol. Nevertheless, we will continue to use the NSW APDC datasets for two reasons: 1) evaluating the insights this data can offer is a study objective, and 2) we will perform sensitivity analyses (cross-validation) to compare findings derived from APDC against PHIDU chronic health conditions, ensuring data validity and reliability.

- 5. It was understood that the Spatial Smoothing Technique will be applied to reduce the risk of identifying communities or individuals, however it was unclear what factors will determine when it is appropriate to employ it. Subsequently, please update the protocol with information on the criteria that will be used to determine when it is appropriate to employ the Spatial Smoothing Technique.**

Response: This study applies spatial smoothing to target identifiable regions and adjust values based on neighbouring areas. It is utilized in scenarios like those in the Statistical Spatial Framework (SSF) guidance by ABS for confidentiality and privacy protection (e.g., regions should

have enough statistical units). No standard numeric thresholds exist for using spatial techniques; they usually rely on professional experience and empirical adjustments.

For example, the paper "An Evaluation of Kernel Smoothing to Protect the Confidentiality of Individual Locations" demonstrates how kernel smoothing can reduce the risk of individual location disclosure. Another paper "Spatial Smoothing and Statistical Disclosure Control" describes how spatial smoothing can be applied to aggregate data into larger geographic units to reduce identification risk.

In this study, spatial smoothing may not be employed if target regions don't arise during analysis. The protocol's spatial smoothing section has been updated to offer further clarification according to the explanation above.

- 6. Under Data Governance on page 13 of the protocol, the flow diagram indicates that in Stage 3 a risk analysis will occur. Please clarify whether this is the role of the expert panel on the AusUrb-HI project advisory committee (as referred to on page 18 of the protocol), as well as more broadly what it will entail.**

Response: We acknowledge that the reference to "risk analysis" in the protocol's flow diagram is oversimplified and confusing. There are four layers of risk analysis in this study: 1) Health data custodian will be joining SURE and get involved in the data import and export stages to monitor risk (**medium risk**); 2) CHReL and Health data custodian has done risk analysis before handling the data (**medium risk**); 3) SURE workspace custodians monitor any data transactional risk during data placement, processing, and extraction (**medium risk**); 4) advisory committee members only review the analysis plan as well as aggregated results upon extraction, since they have no access to SURE (**low risk**).

The reference to "risk analysis" in the diagram for Stage Three from the protocol has been changed to "Overseen by SURE and Data Custodians."

- 7. Multiple supporting documents are referred to throughout the application, however were unfortunately not included as part of the submission:**

A) M5.1 HREA: please see attachment titled Health Liveability Study Areas with map and list of identified areas.

B) Page 10 of protocol: excel file "study area geocodes".

C) Page 10 of protocol: 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**

D) Page 18 of protocol: 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

E) Variables spreadsheet: In addition to the above, we also require SA1 codes for this study. The justification for requiring SA1 codes is provided in the attachment titled "Justification for SA1 data.docx"

Please provide copies of these aforementioned documents.

Response: Items A, B, and C refer to the same document. A new file called *study_area_geocode.csv* has been added to REGIS. For item C, the original sentence discussing ABS's different versions of significant urban area geometry files has been deemed unnecessary since these files can be downloaded from the ABS website. To avoid confusion, the sentence has been revised as follows: "ABS defines a Significant Urban Area as a region with more than 10,000 residents. We have provided a file named *study_area_geocode.csv* that includes SA1 codes and names defining the study regions."

For items D and E, two new files named *Analysis plan.docx* and *Justification for SA1 data.docx* have been added to REGIS respectively.

- 8. The Committee acknowledged that the work from this study will result in an enduring data asset, and requested confirmation that any future research using the data asset will not proceed before ethics approval is obtained.**

Response: We confirm that any future research using the data asset will not proceed before ethics approval is obtained.

- 9. Please provide more information on the Cancer Determinants case study that has already received ethics approval (as referenced on page 7 of the protocol), including a copy of the protocol and approval from the reviewing HREC.**

Response: The reference to the Cancer study was included in error and the two studies are completely separate. The sentence has been removed from the protocol.

Thank you for your time and consideration. Please contact us if you have any queries.