# Assessing Heat-Health Vulnerabilities in Urban Africa: An Interdisciplinary Study of Climatic Impacts and Public Health Resilience

## PhD proposal of Mr Craig Parker 02/08/2023

Table of Contents

[Assessing Heat-Health Vulnerabilities in Urban Africa: An Interdisciplinary Study of Climatic Impacts and Public Health Resilience 1](#_Toc142337458)

[PhD proposal of Mr Craig Parker 02/08/2023 1](#_Toc142337459)

[Summary 2](#_Toc142337460)

[Introduction: 3](#_Toc142337461)

[Background and Rationale 3](#_Toc142337462)

[Study Setting 4](#_Toc142337463)

[Aims and Objectives: 5](#_Toc142337464)

[Hypotheses: 6](#_Toc142337465)

[Data Description 6](#_Toc142337466)

[Aim 1: Mapping Intra-urban Heat Vulnerability and Exposure 6](#_Toc142337467)

[Aim 2: Exploring Lagged Impacts of Heat-health Exposures 7](#_Toc142337468)

[Aim 3: Developing a Heat-health Outcome Prediction Model 7](#_Toc142337469)

[Methods 7](#_Toc142337470)

[Quantifying Intra-Urban Socio-Economic and Environmental Vulnerability 7](#_Toc142337471)

[Utilizing Statistical and Machine Learning Methodologies to Understand Heat-Health Exposures and Their Impacts in Johannesburg 8](#_Toc142337472)

[Developing a Spatially and Demographically Stratified Heat-Health Outcome Forecast Model 9](#_Toc142337473)

[Ethical Considerations: 10](#_Toc142337474)

[Work Plan: 11](#_Toc142337475)

[Research Outputs 11](#_Toc142337476)

[POPIA compliance and protection of personal information 12](#_Toc142337477)

[Budget 13](#_Toc142337478)

[Advisors 13](#_Toc142337479)

[Conclusion 14](#_Toc142337480)

[References: 14](#_Toc142337481)

## Summary

This research proposal endeavors to deepen our understanding of the complex, spatially, and demographically stratified heat-health interactions within large African cities, specifically focusing on Johannesburg. It achieves this through three primary objectives: mapping intra-urban heat vulnerability and exposure, using statistical and machine learning methodologies to analyze the lagged impacts of heat-health exposures, and developing a spatially and demographically stratified heat-health outcome prediction model.

The first aim integrates socio-economic survey data with spatial proxies from Copernicus ERA5 reanalysis and Landsat imagery to map intra-urban heat vulnerability and exposure. We employ Principal Component Analysis (PCA) to pinpoint key predictors, then use multi-level clustering to group urban regions with similar vulnerability profiles based on these socio-economic and satellite-derived factors.

The second aim delves into analyzing heat-health exposures and their impacts, using weather and climate data and existing longitudinal health datasets collected in Johannesburg. Statistical techniques and machine learning methodologies are leveraged to capture the complex, non-linear associations between temperature and morbidity outcomes while incorporating lag-response relationships. The Distributed Lag Non-linear Model (DLNM) is a primary tool in thisendeavor.

Lastly, the third aim is to create a prediction model for heat-health outcomes. Insights gained from previous stages will be utilized to develop groups capturing unique combinations of socio-economic and environmental conditions associated with heightened risk. Supervised machine learning techniques, such as decision trees and random forests, are employed for predictive modeling. The performance of these models will be assessed using various metrics like predictive accuracy and AUC-ROC, in addition to calibration and decision-curve analysis.

By combining these methodologies, this proposal aims to generate a comprehensive and nuanced understanding of how heat impacts health across different groups in Johannesburg, thereby providing critical guidance for necessary interventions and resource allocation.

## Introduction:

The acceleration of urbanization, combined with the intensifying impacts of climate change, has sharply magnified the heat-related health risks in large African cities, a prime example being Johannesburg[1]. Despite the growing urgency, there is a conspicuous dearth of comprehensive data and understanding of heat-health outcomes, the varying degrees of exposure and vulnerability, and the array of potential interventions tailored to these unique African urban contexts[1-4].

This PhD study is deeply intertwined with the HE2AT Center's Research Project 2 (RP2). Sponsored by the US National Institutes of Health (NIH), the HE2AT Center is a vanguard initiative committed to mitigating the health repercussions of climate change in Africa through data science and interdisciplinary collaboration. RP2, in particular, ventures into the depths of spatial and demographic nuances underlying heat-health interactions in Johannesburg. Anchored in a robust empirical foundation, the primary objective of this study transcends mere understanding, aspiring to harness this knowledge in formulating predictive models that can accurately gauge risk and recommend proactive mitigation measures.

## Background and Rationale

Climate change has brought steadily rising temperatures worldwide, severely escalating the public health threats posed by high ambient temperatures and heat waves[5, 6]. These temperature increases are more pronounced in the tropics, particularly in rapidly urbanizing African cities, which are projected to bear a substantial burden of increased heat-related mortality and morbidity[7, 8]. Unique vulnerabilities arise in these settings due to high prevalence rates of diseases like HIV and TB, limited access to cooling resources, and large populations residing in informal settlements[6, 9].

The Urban Heat Island(UHI) effect further compounds these vulnerabilities. This phenomenon, caused by urban development and the lack of vegetation, causes cities to experience considerably higher temperatures than their surrounding rural areas[10]. Structures within these cities – such as informal dwellings, low-cost government housing, school classrooms, and clinics – are often inadequately designed for these conditions, having indoor temperatures that are 3-4°C warmer than outside, thereby exacerbating heat exposure[1, 8].

While the health impacts of heat are widely recognized, there is a critical knowledge and methodological gap concerning the assessment and prediction of heat-related health risks within the African city context[11, 12]. Existing heat-health impact assessments frequently fail to account for the complexity of urban spaces, unique environmental exposures, and the specific demographic and disease spectrums in Africa[13]. For example, research in northern Ghana has demonstrated a significant association between high daily temperatures and all-cause mortality, yet comprehensive research investigating heat-related morbidity in Africa remains scarce[9].

A similar lack of information characterizes South Africa, where about 11% of the urban population live in high-poverty informal settlements[14]. Preliminary research in Johannesburg has indicated that indoor classroom temperatures often exceed comfortable levels, causing students to experience symptoms such as fatigue and difficulty breathing[14]. Additionally, temperatures in waiting rooms in rural clinics exceeded 38°C, suggesting that patients seeking treatment could experience considerable discomfort and health impacts due to heat exposure[15].

To address these challenges and gaps in knowledge, this PhD proposal aims to explore the unique socio-economic and environmental vulnerability at an intra-urban level in Johannesburg. Advanced statistical and machine learning methodologies will be employed to uncover the nuanced and potentially lagged impacts of heat-health exposures. A critical part of this research will be developing a prediction model that accounts for spatial and demographic differences, allowing for estimating adverse health outcomes at various temperature thresholds.

This comprehensive exploration will significantly enhance our understanding and prediction of heat-related health risks in African cities. The insights gained can subsequently inform the design of suitable interventions, such as Early Warning Systems that account for the vulnerability of different population groups and long-term structural interventions to reduce heat exposures in urban areas. This research will pave the way for creating more resilient urban environments in Africa, facing escalating challenges due to climate change.

## Study Setting

Nestled on the Highveld plateau at an elevation of 1,753 meters, the vibrant city of Johannesburg forms the centerpiece of our research. As the largest city in South Africa and the 26th largest globally, Johannesburg's population exceeds 5.635 million inhabitants[16]. This bustling metropolis, characterized by its unique subtropical highland climate, provides a compelling backdrop for exploring urban heat health impacts[17].

Johannesburg's distinct weather patterns follow a bifurcated climate cycle. Summer months, extending from October to April, are marked by hot days often followed by refreshing afternoon thundershowers, subsequently transitioning into cooler evenings. The winter period from May to September offers a contrasting spectacle of dry, sunny days leading into cold nights. Due to the city's high elevation, the climate remains generally mild, with average maximum daytime temperatures oscillating between 25.6 °C (78.1°F) in January and 16 °C (61°F) in June.

Johannesburg's socio-economic canvas is marked by stark disparities, with impoverished urban communities shouldering the disproportionate burden of climate change impacts on health and well-being[18]. Inadequate housing, limited access to resources, and poverty elevate these communities' vulnerability to heat-related health effects, a situation worsened by infrastructural deficiencies[17].

From a health standpoint, a unique set of risk factors shapes the relationship between heat and health in Johannesburg. Adverse health outcomes linked to heat exposure include high blood pressure, respiratory stress, and cardiac conditions, all further aggravated by the prevailing socio-economic and infrastructural conditions[1, 19, 20]. A critical health consideration in Johannesburg is the prevalence of communicable diseases, notably HIV, Tuberculosis, and COVID-19 [21, 22]. These diseases add complexity to the health landscape, with heat exposure potentially affecting the health status and disease progression in affected individuals.

Heatwaves pose a significant public health risk in Johannesburg[17]. Research has revealed temperature thresholds associated with a heightened risk of mortality. Such insights emphasize the importance of characterizing past and future heatwaves to enhance heat health warning systems and inform health-centric policy-making[23].

Against the backdrop of the Urban Heat Island (UHI) phenomenon, the myriad socio-economic inequalities, infrastructural challenges, and health-related considerations in Johannesburg make it an ideal study site for our research[24]. By dissecting these complexities in the context of climate, we aim to enhance our understanding of the multi-layered relationships between urban heat exposure, population vulnerability, and health outcomes. These insights will serve as valuable inputs for the evolution of heat health warning systems and policies designed to safeguard the most vulnerable from the health impacts of heat exposure.

## Aims and Objectives:

The primary objective of this research study is to deepen our understanding of the complex, spatially, and demographically stratified heat-health interactions common in large African cities.

1. Map intra-urban heat vulnerability and exposure across urban areas in large African cities, quantifying the intra-urban socio-economic and environmental vulnerability (Aim 1).

2. Utilize statistical and machine learning methodologies to elucidate the lagged impacts of heat-health exposures in Johannesburg and improve our understanding of the associations between heat and health in this context (Aim 2).

3. Develop a spatially and demographically stratified heat-health outcome prediction model that can predict the probability of adverse health outcomes at different temperature thresholds (Aim 3).

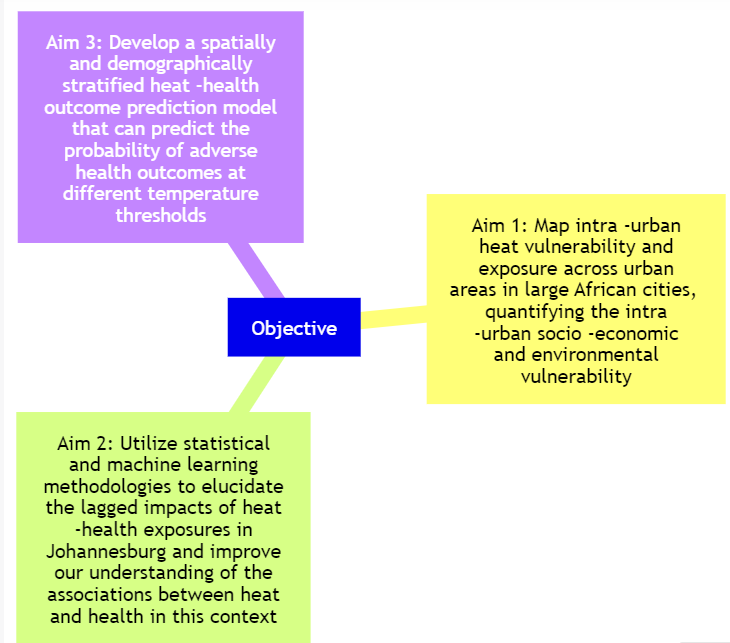


Figure 1: Summary of study aims.

## Hypotheses:

1. Socio-demographic attributes of inhabitants in African urban landscapes significantly contribute to their vulnerability to heat-related health hazards.

2. Statistical and machine learning methodologies, including Distributed Lag Non-linear Models (DLNMs), can be harnessed to illuminate the lagged impacts of heat-health exposures in African urban environments such as Johannesburg, thus enhancing our understanding of the associations between heat and health.

3. It is possible to develop a spatially and demographically stratified heat-health outcome prediction model, enabling us to forecast the probability of adverse health outcomes at different temperature thresholds.

## Data Description

We will leverage an extensive range of data sources to meet the stated objectives. All of this data will be stored and managed within the high-performance computing data hub at the University of Cape Town's Climate System Analysis Group (UCT CSAG). This hub is a key resource developed under the auspices of the HE2AT center project. This system's comprehensive data management plan is available and can be provided upon request.

### Aim 1: Mapping Intra-urban Heat Vulnerability and Exposure

The first data category encompasses socio-economic survey data, providing insight into household economic status, living conditions, and demographic and environmental features. Notable data sets include the Gauteng City Region Observatory (GCRO) survey data, including the Quality-of-Life survey (GCRO QoL survey 2020/21), Population Censuses, and community surveys like Census 2011 and the Community Survey 2016.

The second data category employs spatial proxies of socio-economic conditions derived from satellite image-based estimations of land use and cover types, access to services like electricity, proximity to healthcare facilities, and access to green spaces. Data like Google Street View analysis and SEDAC gridded human population data are also included.

### Aim 2: Exploring Lagged Impacts of Heat-health Exposures

Aim 2 necessitates the utilization of weather and climate data, including retrospective or historical datasets. This includes model-based re-analysis data that combines physics and numerical weather simulations with observations from a global network of sources such as weather stations, satellites, and simulation models. Examples of utilized data include MODIS and LANDSAT Land Surface Temperature (LST), UK Met Office/IMPALA CP4-A (4km resolution 10-year climate simulation over Africa), TAHMO weather stations, WMO Global Summary Of the Day (GSOD) weather stations, ECMWF ERA-5 Re-analysis of regional circulation fields and ECMWF ERA5-Land.

Simultaneously, this aim will integrate human health data drawn from existing longitudinal datasets collected in Johannesburg. This includes data from trials and cohorts among HIV-infected and HIV-uninfected adults and adults participating in longitudinal studies related to COVID-19 prevention or treatment.

### Aim 3: Developing a Heat-health Outcome Prediction Model

In Aim 3, the previously described human health data will be revisited, contributing to developing a heat-health outcome prediction model. The data, obtained through collaborations with the Principal Investigators of potentially eligible datasets and relevant data repositories, will provide a rich tapestry of information for mapping against the weather and other data collated in Aims 1 and 2.

Variable categories drawn from these studies span from social, economic, and demographic characteristics to geospatial location of participants, self-reported health status, medical examination findings, and laboratory outcome data. Including all participants 18 years and older at the time of enrolment in the primary studies ensures a comprehensive scope, eliminating the need for sampling to select particular subsets of individuals.

## Methods

### Quantifying Intra-Urban Socio-Economic and Environmental Vulnerability

To quantify the intra-urban socio-economic and environmental vulnerability, we will apply an integrated methodological approach that synthesizes multiple statistical techniques, encapsulating the diverse elements of urban vulnerability into a cohesive understanding[25].

We initiate our approach with a process of dimensionality reduction. Due to the high dimensionality of socio-economic and environmental factors, we need to identify those variables that significantly contribute to intra-urban socio-economic variability[26].

To achieve this, we will implement a PCA-based automated variable selection methodology to identify candidate inputs to a geodemographic classification from a collection of variables[33]. This method, exemplified in a study on the UK 2011 Census, improves cluster assignment quality as manifested by a lower total within-cluster sum of square score.

In this way, PCA will be utilized to emphasizes variation and identify strong patterns in a dataset, thereby effectively reducing its dimensionality[27]. Using this approach, we can recognize those key predictors or features that account for the majority of the variability within the data[28].

After identifying the principal components, our next step is the synthesis of a composite indicator[29]. This indicator integrates these components into a comprehensive metric representing the combined socio-economic and environmental vulnerability[30]. Though this indicator could be applied to many climate risks, our primary focus remains heat vulnerability.

Finally, to visualize the varied vulnerability profiles across Johannesburg, we generate a final output map using a spatial multi-criteria analytic method[29]. This map serves as a valuable tool in identifying areas of high vulnerability, thus providing critical guidance for necessary interventions and resource allocation.

A diagram of a company

Description automatically generated

Figure 2Quantifying Intra-Urban Socio-Economic and Environmental Vulnerability: A Methodological Approach

### Utilizing Statistical and Machine Learning Methodologies to Understand Heat-Health Exposures and Their Impacts in Johannesburg

In Aim 2, we extend the scope of our investigation by examining the effects of intra-urban socio-economic and environmental vulnerability on heat-health interactions in Johannesburg, with a specific focus on health outcomes obtained from clinical cohort datasets . This aim builds on the groundwork established in Aim 1, which quantifies these vulnerabilities at an intra-urban level.

Our approach adopts an integrated methodology that combines advanced statistical techniques with machine learning. Our primary tool for this aim is the Distributed Lag Non-linear Model (DLNM), developed by Gasparrini [31]. This technique allows us to capture the complex comprehensively, non-linear associations between temperature and health outcomes while also incorporating lag-response relationships[32]. We will implement the DLNM in a quasi-Poisson regression framework, using a flexible natural cubic spline function to represent the temperature-health associations, and an unconstrained parameterization to capture the lag-response dimension of the DLNMs[33]. This setup allows us to assess both immediate and up to three-day lagged health outcomes, offering an opportunity to illuminate the temporal spectrum of heat-health impacts.

Our approach also extends to controlling for confounding variables that might affect morbidity rates. We integrate an indicator for the day of the week, capturing weekly variations in health impacts. Additionally, a natural cubic spline function of the day of the year with four degrees of freedom is included to control for the annual cyclical pattern of the morbidity rate. Long-term trends in health outcomes are addressed through an indicator for the year, and an interaction term between this indicator and the day-of-year spline function allows us to capture any changes in seasonal health patterns over time[34].

Techniques such as decision trees, random forests, gradient boosting, deep learning models, model-based regression trees (MOB), multivariate adaptive regression splines (MARS), the patient rule-induction method (PRIM), and adaptive index models (AIM) will be utilized to identify intricate patterns and interactions within the data[35]. While the traditional models like decision trees and random forests capture general patterns, the specialized methods like MOB, MARS, PRIM, and AIM focus on finding relevant splits in indicator data, doing so in different fashions. Sensitivity analyses will be conducted, and potential confounders will be adjusted to ensure the robustness and validity of our findings[36].

In terms of model evaluation, we will use metrics such as mean squared error (MSE), receiver operating characteristic (ROC) curves, and area under the curve (AUC). Overfitting, a common concern in machine learning models, will be mitigated through techniques like cross-validation and regularization, thereby ensuring the robustness and generalizability of our models[36, 37].

Through this cohesive strategy, we aim to illuminate nuanced heat-health relationships, particularly delayed impacts of heat exposure, and enrich our understanding of these dynamics in Johannesburg. Our comprehensive view integrates social and environmental factors, enhancing our understanding of how they collectively contribute to heat-related health risks in this context. By weaving together the findings from Aims 1 and 2, we will provide a robust and comprehensive analysis of heat-related health morbidity risks in Johannesburg.

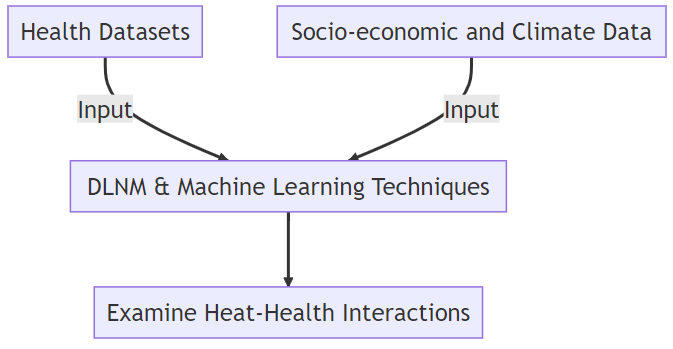


Figure 3: Statistical and Machine Learning Methodologies for Understanding Heat-Health Exposures and Their Impacts

### Developing a Spatially and Demographically Stratified Heat-Health Outcome Forecast Model

With the insights gleaned from the spatial and demographic vulnerability analysis in Aim 1 and the heat-health relationships identified in Aim 2, we are well positioned to develop a sophisticated forecasting model.

To begin, we will take the heat vulnerability maps and principal components produced in Aim 1 as a basis for stratifying our data. Leveraging these key vulnerability features, we will develop groups, or strata, that capture unique combinations of socio-economic and environmental conditions associated with heightened risk.

Next, we will integrate the heat-health relationships established in Aim 2 into this framework. The identified non-linear and lagged relationships between temperature and health outcomes will serve as the backbone of our predictive model. These complex relationships will be incorporated into our models, helping to stratify the risk groups based on potential health outcomes.

Given these stratifications, we will employ supervised machine learning techniques, like decision trees and random forests, to develop predictive models for each stratum[37]. These models will be tasked with forecasting specific health outcomes such as high blood pressure, respiratory stress, cardiac conditions, and specific HIV-related blood test results, based on a given set of predictor variables relevant to each stratum[35].

Lastly, to ensure that our models are not only accurate but also interpretable, we will employ methods like SHapley Additive exPlanations (SHAP) to understand the contribution of each variable to the prediction. Our models' performance will be assessed using metrics like predictive accuracy, AUC-ROC, among others, and the models will be cross-validated to ensure robustness and avoid overfitting[38].

Thus, through Aim 3, we will generate a suite of predictive models that, while stratified according to spatial and demographic vulnerabilities, can accurately forecast heat-related health outcomes. This methodology will result in a comprehensive and nuanced understanding of how heat impacts health across various groups in Johannesburg.

A diagram of a flowchart

Description automatically generated

Figure 4: Stratified Heat-Health Outcome Forecast Model Development Process

|  |  |  |
| --- | --- | --- |
| Aim | Description | Key Methods |
| 1 | Quantifying Intra-Urban Socio-Economic and Environmental Vulnerability | Principal Component Analysis (PCA) to identify key predictors, Spatial Principal Component Analysis for spatial variability, Multi-level clustering to group regions with similar vulnerability profiles, Spatial multi-criteria analytic method to generate the final output map |
| 2 | Utilizing Statistical and Machine Learning Methodologies to Understand Heat-Health Exposures and Their Impacts in Johannesburg | Distributed Lag Non-linear Model (DLNM) to assess lag–response relationships, Machine learning techniques (decision trees, random forests, gradient boosting, deep learning models) for pattern recognition, Time-series and survival analysis for temporal dynamics, Cross-correlation analysis for lagged effects |
| 3 | Developing a Spatially and Demographically Stratified Heat-Health Outcome Forecast Model | Supervised machine learning techniques (regression, decision trees, random forests, K-means clustering, support vector machines) for predictive modeling, Stratification methods for sub-group discovery, Selection of optimal algorithms based on engagement with health practitioners and researchers, Use of Aim 1 and Aim 2 outputs as inputs for model prediction |

Table 1: This table provides an overview of the specific techniques and processes to be used in each aim.

## Ethical Considerations:

Ethics and Privacy Considerations for the Study on Heat-Health Exposures in Johannesburg

This doctoral research, examining the impacts of heat-health exposures in Johannesburg, is committed to respecting and protecting the rights of the individuals represented in our secondary health datasets. The study design acknowledges and addresses several ethical considerations and legal requirements, including the use of secondary data for new research purposes, the handling of potentially identifying information, and the necessity for robust data management.

We have obtained ethical clearance from the University's ethical committee (HE2AT Center HEAT002: 220606 Wits Human Research Ethics Committee) and will continue to strictly adhere to the University's ethical approval procedures throughout the course of this research. Our commitment to ethical excellence ensures that any data collected from human participants is gathered, analyzed, stored, and shared securely and ethically.

Guided by globally recognized guidelines, including the Declaration of Helsinki, ICH Good Clinical Practice, and the Ethics in Health Research (Department of Health South Africa), as well as the Protection of Personal Information Act (POPIA), our research aims to safeguard the interests of the individuals while serving the public interest.

The objectives of the research are transparently defined, and the reasons for involving personal data are clearly elucidated. To honour the principles of informed consent, we are respecting the conditions of data use as per the original data collection. We are prepared to seek re-consenting or apply for an informed consent waiver from the ethics committee for 'narrow' or 'tiered consent' datasets, if necessary.

In compliance with the provisions set forth by POPIA and other data security laws, a detailed data management plan will be implemented. This plan will cover secure data transfer, storage, and access, with strict controls limiting access to necessary personnel only. Data will be anonymized when required, and our commitment extends to ensuring that publications from this research will not contain any potentially identifiable information from participants.

In conclusion, we believe these measures not only meet ethical and legal requirements but also respect and protect the rights of the individuals represented in our datasets. Our commitment to maintaining high ethical standards throughout this research underscores our dedication to both scientific rigor and the welfare of the public.

## Work Plan:

Year 1: Undertake a comprehensive literature review, establish the research protocol in alignment with the study objectives, and prepare the first paper on intra-urban heat vulnerability and exposure across urban areas in large African cities. This will lay the groundwork for creating high-resolution urban temperature hazard maps.

Year 2: Analyze GCRO and climate data to elucidate socio-demographic characteristics contributing to heat vulnerability in Johannesburg. These findings will inform the second paper, which focuses on quantifying intra-urban socio-economic and environmental vulnerability.

Year 3: Employ advanced statistical and machine learning techniques to investigate the lagged impacts of heat-health exposures in Johannesburg. This knowledge will improve our understanding of the associations between heat and health in this unique urban context and form the basis for the third paper.

Year 4: Construct a spatially and demographically stratified heat-health outcome prediction model, capable of forecasting the probability of adverse health outcomes at different temperature thresholds. This model, which incorporates an LLM approach, will be thoroughly evaluated and iteratively improved upon. The successful development and evaluation of this model will be documented in the fourth paper. The final year will also be dedicated to writing the PhD thesis, bringing together all the research insights garnered over the course of the PhD study.

## Research Outputs

Our research will culminate in four seminal papers, each showcasing a distinct aspect of our investigation into the heat-health outcomes in Johannesburg. These papers, central to our scholarly contribution, will be leveraged for maximum impact through our comprehensive dissemination strategy.

1. **Research Protocol Documentation:** The first paper will detail the research protocol for the study. This protocol will be publicly accessible, encouraging replication of our methodological framework in other settings. This will not only reinforce scientific rigor but also catalyze further research in this critical field.
2. **Socio-economic and Climate Vulnerability Analysis:** The second paper will leverage socio-economic and climate data to dissect the vulnerability traits of the Johannesburg population. By sharing these findings through scientific conferences and open-access publications, we aim to trigger dialogues within the scientific community and beyond, fostering a greater understanding of the socio-economic implications of climate change.
3. **Heat-Health Correlations and Lagged Impact Analysis:** The third paper will employ sophisticated statistical and machine learning methodologies to explore the correlation between temperature and health outcomes. It will also scrutinize the lagged impacts of heat-health exposures in Johannesburg. By sharing these nuanced findings, we aim to highlight the hidden complexities of heat-health interactions, informing future research and public health policies.
4. **Heat-Health Outcome Prediction Model:** The fourth paper will document the development and evaluation of a spatially and demographically stratified heat-health outcome prediction model for Johannesburg. This paper will not only outline the predictive power of our model but also underscore its potential for informing risk mitigation strategies. Through its dissemination, we aspire to catalyze proactive and data-informed public health interventions in Johannesburg and similar settings.

These research papers will serve as the basis for scientific presentations and community and policy engagement. They will be critical in not only contributing to academic discourse but also in shaping health policies, informing public awareness, and influencing future climate change and health adaptation planning.

|  |  |  |
| --- | --- | --- |
| Year | Activities | Outcome |
| 1 | Comprehensive literature review, establish research protocol, prepare first paper on intra-urban heat vulnerability. | First paper detailing research protocol and groundwork for high-resolution urban temperature hazard maps. |
| 2 | Analyze GCRO data, prepare second paper on quantifying intra-urban socio-economic and environmental vulnerability. | Second paper providing a deeper understanding of vulnerability characteristics in Johannesburg. |
| 3 | Apply advanced statistical and machine learning techniques, prepare third paper on lagged impacts of heat-health exposures. | Third paper presenting the correlation between temperature and health impacts in Johannesburg. |
| 4 | Construct and improve a prediction model, write fourth paper, complete the writing of the PhD thesis. | Fourth paper documenting the prediction model and PhD thesis incorporating all research insig |

Table 2: PhD Research Timeline, Activities and Outcomes

## POPIA compliance and protection of personal information

In alignment with the Protection of Personal Information Act of South Africa (POPIA, 2013), our research meticulously attends to data security and confidentiality. POPIA sets limitations on personal information processing but simultaneously allows for its use in scientific research. Our study is cognizant of this, alongside other governing legal frameworks like the National Health Act No 61 of 2003, the Constitution of the Republic of South Africa, and Department of Health guidelines on Ethics in Health Research.

Our research strategy includes processing de-identified health databases in which re-identification is virtually impossible. Where personal information has not been de-identified, we comply with the relevant sections of POPIA, allowing us to process health data for historical, statistical, and research purposes.

The information gathered and processed by our team will only be used for research and statistical purposes, which directly relate to addressing the major public interest of understanding and mitigating the health implications of rapidly escalating temperatures and heat waves, particularly in Africa. This processing of data is deemed necessary and justified as it serves to inform strategies to combat one of the greatest health threats of the 21st century – climate change.

Security measures will be implemented to prevent unlawful access or processing of personal information, while the operators involved in the data handling process will be bound by a written contract, ensuring accountability. This approach aligns with Sections 19, 20, and 21 of POPIA, demonstrating our commitment to preserving the rights of individuals and upholding the highest ethical standards in scientific research.

## Strengths and Weaknesses

### Strengths

1. **Multi-Dimensional Approach**: The proposal addresses complex spatial and demographically stratified heat-health interactions through a combination of socio-economic survey data, satellite imagery, statistical techniques, and machine learning methodologies. This multifaceted approach adds depth to the research.
2. **Focus on Vulnerability Mapping**: By emphasizing the identification and grouping of urban regions with similar vulnerability profiles, the proposal targets an area of critical importance, especially in the context of climate change.
3. **Use of Existing Datasets**: Leveraging existing longitudinal health datasets and weather data provides a foundation for insights, potentially ensuring that the research builds on existing knowledge and connects to real-world contexts.
4. **Predictive Modeling**: Developing a predictive model for heat-health outcomes represents a concrete, actionable output that could be translated into interventions and policy decisions.

### Weaknesses

1. **Data Dependence**: The study relies on the availability and quality of multiple datasets, including socio-economic survey data, spatial proxies, and health datasets. Issues with data quality, accessibility, or integration could pose significant challenges.
2. **Complexity**: The proposal's multifaceted approach, while a strength, may lead to a complexity that could be challenging to manage, requiring a high degree of methodological rigor and clear articulation of each stage of the research.
3. **Generalizability**: With a focus on Johannesburg, the findings' applicability to other African cities or varying contexts might be limited. While this specific focus enables in-depth analysis, a clear rationale for the transferability of insights would strengthen the proposal.

## Budget

The financial plan for this research is comprehensive and carefully curated to ensure that every requirement for the project is catered to. A detailed budget will include:

1. **Data Collection**: As the backbone of our research, the financial provision for data collection is paramount. This encompasses the procurement of secondary datasets, possible field surveys, and necessary tools or services for data gathering.
2. **Software Licenses**: Our research will leverage several specialized software for data analysis and model building. The budget covers the licensing costs of necessary software including, but not limited to, statistical software, machine learning libraries, geographic information system (GIS) software, and data visualization tools.
3. **Hardware**: To ensure efficient analysis of extensive datasets, high-performance computing resources are required. We will allocate funds for acquiring suitable hardware or subscribing to cloud-based computational services.
4. **Publication Fees**: As part of the research dissemination, the budget accounts for possible charges associated with publishing in open-access, peer-reviewed scientific journals.
5. **Training**: To keep abreast with the fast-evolving data science landscape and emerging tools and techniques in climate-health research, funds will be set aside for continuous training and capacity building for the research team.

The budget has been structured within the limits of the National Institutes of Health (NIH) grant, which covers the PhD research. However, given the magnitude and scope of the project, additional funding might be sought through various avenues such as additional grants, research partnerships, or institutional collaborations.

## Advisors

This research will be supported by an outstanding team of advisors, each bringing their vast knowledge and expertise in the intertwined disciplines of health and climate science.

Professor Matthew Chersich, based at Wits RHI, offers a wealth of experience in public health research that is invaluable to our study, particularly the health-related aspects. His career, spanning over two decades, has been focused on medical and public health research in Africa, particularly on maternal health and HIV, and recently on climate change and health. He has an extensive academic background in clinical medicine and public health, contributing to 14 WHO guidelines or monologues and serving as a contributing author to the Africa chapter of the 6th Intergovernmental Panel on Climate Change report. He has published more than 175 papers in peer-reviewed journals and has a significant H-Index of 48.

Professor Akbar Waljee of the University of Michigan brings crucial experience in statistical modelling and machine learning, essential for our data analysis and predictive modelling. Born in Kenya and educated in the United States, Prof. Waljee leads several key data and healthcare initiatives at the University of Michigan and the VA Ann Arbor Healthcare System. His work primarily involves utilizing machine learning and deep learning techniques to enhance healthcare access, quality, and efficiency, particularly in resource-constrained settings. His innovative work in decision support systems and personalized care is set to revolutionize patient care in gastroenterology and liver disorders in under-resourced settings globally.

Dr. Christopher Jack from the University of Cape Town strengthens the climate aspects of our study with his extensive knowledge in climate science, ensuring a well-rounded and sophisticated understanding of the climate-health nexus. With a background in computer science and ocean/atmospheric science, Dr. Jack possesses a unique blend of skills in high performance computing, modeling, analysis, science-society engagement, and decision-making under uncertainty. His current research activities are concentrated on the intersection of urban contexts and climate risk, leveraging his comprehensive expertise in climate science and modeling, and his proficiency in decision support and capacity development. His passion lies in working with and across diverse disciplines in complex problem spaces, making him especially interested in urban climate resilience in developing contexts.

Together, these advisors contribute a multidisciplinary perspective to our research, enriching its depth and breadth, and enhancing its potential impact.

## Conclusion

This research project seeks to explore the intricate relationship between urban heat exposure, population vulnerability, and health outcomes within the unique socio-economic, environmental, and climatic context of Johannesburg. Utilising advanced statistical techniques, machine learning methods, and a variety of robust data sources, the research aims to establish a nuanced understanding of heat-health effects in the city. This will culminate in the development of a spatially and demographically stratified heat-health outcome prediction model, which will enhance the city's readiness and response to heat-related health risks, ultimately contributing to the wellbeing of Johannesburg's inhabitants. As global temperatures continue to rise, the insights generated from this study could provide pivotal contributions to climate science, public health, AI, and the broader interdisciplinary field of climate and health.

## References

1. Wright, C.Y., et al., *Socio-economic, infrastructural and health-related risk factors associated with adverse heat-health effects reportedly experienced during hot weather in South Africa.* Pan Afr Med J, 2019. **34**: p. 40.

2. Kapwata, T., M.T. Gebreslasie, and C.Y. Wright, *An analysis of past and future heatwaves based on a heat-associated mortality threshold: towards a heat health warning system.* Environmental Health, 2022. **21**(1): p. 112.

3. Opoku, S.K., et al., *Climate Change and Health Preparedness in Africa: Analysing Trends in Six African Countries.* Int J Environ Res Public Health, 2021. **18**(9).

4. Ellena, M., M. Breil, and S. Soriani, *The heat-health nexus in the urban context: A systematic literature review exploring the socio-economic vulnerabilities and built environment characteristics.* Urban Climate, 2020. **34**: p. 100676.

5. Harmer, A., et al., *WHO should declare climate change a public health emergency.* Bmj, 2020. **368**: p. m797.

6. Chersich, M.F., et al., *Impacts of Climate Change on Health and Wellbeing in South Africa.* Int J Environ Res Public Health, 2018. **15**(9).

7. Wright, C.Y. and M. Norval, *Climate change: One of the greatest threats to public health in the 21st century.* S Afr Med J, 2014. **104**(8): p. 578.

8. Chapman, S., et al., *Past and projected climate change impacts on heat-related child mortality in Africa.* Environmental Research Letters, 2022. **17**(7): p. 074028.

9. Adams, I., S. Ghosh, and G. Runeson, *Access to Early Warning for Climate Change-Related Hazards in Informal Settlements of Accra, Ghana.* Climate, 2022.

10. Pandey, R.K. *Urban Heat Island Effect and Climate Change: An Assessment of Interacting and Attainable Variations in Indian Cities: Study of Gorakhpur*. 2018.

11. Mhedhbi, Z., et al. *Mining the Web of Science for African cities and climate change (1991–2021)*. in *Frontiers in Sustainable Cities*. 2023.

12. Mir Alvarez, C., et al., *A Scoping Review on Air Quality Monitoring, Policy and Health in West African Cities.* Int J Environ Res Public Health, 2020. **17**(23).

13. Rohat, G., et al., *Projections of human exposure to dangerous heat in African cities under multiple socioeconomic and climate scenarios.* Earth's Future, 2019. **7**(5): p. 528-546.

14. Weimann, A. and T. Oni, *A Systematised Review of the Health Impact of Urban Informal Settlements and Implications for Upgrading Interventions in South Africa, a Rapidly Urbanising Middle-Income Country.* Int J Environ Res Public Health, 2019. **16**(19).

15. Wright, C.Y., et al., *Indoor Temperatures in Patient Waiting Rooms in Eight Rural Primary Health Care Centers in Northern South Africa and the Related Potential Risks to Human Health and Wellbeing.* Int J Environ Res Public Health, 2017. **14**(1).

16. *Worldometer. (2023, January 1). World population by country. Retrieved from* [*https://www.worldometers.info/world-population/population-by-country/*](https://www.worldometers.info/world-population/population-by-country/)*.*

17. Souverijns, N., et al., *Urban heat in Johannesburg and Ekurhuleni, South Africa: A meter-scale assessment and vulnerability analysis.* Urban Climate, 2022. **46**: p. 101331.

18. Khine, M.M. and U. Langkulsen, *The Implications of Climate Change on Health among Vulnerable Populations in South Africa: A Systematic Review.* Int J Environ Res Public Health, 2023. **20**(4).

19. Gronlund, C.J., *Racial and socioeconomic disparities in heat-related health effects and their mechanisms: a review.* Curr Epidemiol Rep, 2014. **1**(3): p. 165-173.

20. Johnson, D.P., J.S. Wilson, and G.C. Luber, *Socioeconomic indicators of heat-related health risk supplemented with remotely sensed data.* International Journal of Health Geographics, 2009. **8**(1): p. 57.

21. *Centers for Disease Control and Prevention. (n.d.). South Africa. Retrieved August 3, 2023, from* [*https://www.cdc.gov/globalhealth/countries/southafrica/default.htm*](https://www.cdc.gov/globalhealth/countries/southafrica/default.htm)*.*

22. López-Carr, D., et al., *Extreme Heat and COVID-19: A Dual Burden for Farmworkers.* Front Public Health, 2022. **10**: p. 884152.

23. Nana, M., K. Coetzer, and C. Vogel, *Facing the heat: initial probing of the City of Johannesburg’s heat-health planning.* South African Geographical Journal, 2019. **101**: p. 1-16.

24. Hardy, C.H. and A.L. Nel, *Data and techniques for studying the urban heat island effect in Johannesburg.* Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., 2015. **XL-7/W3**: p. 203-206.

25. Linder, S., et al., *A Population-Based Approach to Mapping Vulnerability to Diabetes.* Int J Environ Res Public Health, 2018. **15**(10).

26. Southworth, J., et al., *Using a coupled dynamic factor - random forest analysis (DFRFA) to reveal drivers of spatiotemporal heterogeneity in the semi-arid regions of southern Africa.* PLoS One, 2018. **13**(12): p. e0208400.

27. Wolf, T. *What are the real limitations of Heat Vulnerability Mapping support to decision-making?* 2016.

28. Liu, X., et al., *Mapping Urban Heat Vulnerability of Extreme Heat in Hangzhou via Comparing Two Approaches.* Complex., 2020. **2020**: p. 9717658:1-9717658:16.

29. Qureshi, A.M. and A. Rachid, *Heat Vulnerability Index Mapping: A Case Study of a Medium-Sized City (Amiens).* Climate, 2022.

30. Abrar, R., et al., *Assessing the Spatial Mapping of Heat Vulnerability under Urban Heat Island (UHI) Effect in the Dhaka Metropolitan Area.* Sustainability, 2022.

31. Gasparrini, A., B. Armstrong, and M.G. Kenward, *Distributed lag non‐linear models.* Statistics in medicine, 2010. **29**(21): p. 2224-2234.

32. Ruuhela, R., et al., *Temperature-Related Mortality in Helsinki Compared to Its Surrounding Region Over Two Decades, with Special Emphasis on Intensive Heatwaves.* Atmosphere, 2020.

33. Armando, C.J., et al., *Climate variability, socio-economic conditions and vulnerability to malaria infections in Mozambique 2016-2018: a spatial temporal analysis.* Front Public Health, 2023. **11**: p. 1162535.

34. Achebak, H., D. Devolder, and J. Ballester, *Heat-related mortality trends under recent climate warming in Spain: A 36-year observational study.* PLoS Med, 2018. **15**(7): p. e1002617.

35. Masselot, P., et al., *Machine learning approaches to identify thresholds in a heat-health warning system context.* Journal of the Royal Statistical Society Series A: Statistics in Society, 2021. **184**(4): p. 1326-1346.

36. Ghahramani, M.B., et al., *Intelligent Geodemographic Clustering Based on Neural Network and Particle Swarm Optimization.* IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2022. **52**: p. 3746-3756.

37. Kumar R, S., et al., *Analysis of classification and clustering techniques for ambient AQI using machine learning algorithms.* 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT), 2022: p. 902-908.

38. Yuan, H., et al., *An empirical study of the effect of background data size on the stability of SHapley Additive exPlanations (SHAP) for deep learning models.* arXiv preprint arXiv:2204.11351, 2022.