University of the Witwatersrand

School of Public Health

Urban Heat and Health in Johannesburg: A Multidimensional Analysis of Vulnerability, Explanatory Modelling, and Predictive Outcomes

PhD Protocol

Version v1.0

Last updated: April 22, 2025

Submitted by:

Craig Parker, MSc, MM-PDM, BS

Wits Planetary Health Research

Supervisors:

Dr. Admire Chikandiwa

Professor Matthew Chersich

Professor Akbar Waljee

Dr. Christopher Jack

April 22, 2025

Contents

Abstract

This research proposal investigates the complex relationship between urban heat and health in Johannesburg, South Africa. As climate change drives increasing temperatures globally, urban populations face heightened health risks, with vulnerable communities disproportionately affected. The study employs a multidimensional approach through three primary objectives: (1) mapping intra-urban heat vulnerability by integrating environmental, socio-economic, and health data; (2) delineating heat-health dynamics through a two-stage explanatory modeling approach that combines hypothesis generation with targeted testing to uncover physiological pathways and temporal effects; and (3) developing a stratified predictive model for heat-related health outcomes. Drawing on clinical trial data, satellite imagery, climate records, and socioeconomic surveys, this research will apply advanced statistical and machine learning techniques to create vulnerability maps, explain complex relationships, and generate predictive models. The findings aim to inform targeted public health interventions, urban planning decisions, and climate adaptation strategies to protect vulnerable populations from increasing heat exposure in Johannesburg and potentially other African urban centers.

Keywords: urban heat, health outcomes, vulnerability mapping, machine learning, climate change, Johannesburg

Acknowledgements

This research was conducted as part of the HEat and HEalth African Transdisciplinary (HE^{2}AT) Center initiative, supported by the NIH Common Fund under Award Number U54 TW 012083.

1 Introduction and Background

1.1 Climate Change and Heat-Health Impacts in the Johannesburg Context

Climate change has driven a global temperature increase of over 1.2°C since the Industrial Revolution, with African regions experiencing higher-than-average temperature increases (Intergovernmental Panel on Climate Change, 2022). Urban areas are particularly affected due to development patterns and land use changes, with high temperatures increasingly linked to mortality and morbidity, especially during heatwaves (Gasparrini et al., 2015; Analitis et al., 2018).

Heat-related risks are rising in Johannesburg at 1753 meters elevation with over 5.87 million inhabitants (Worldometer, 2023). The city recorded its highest temperature of 38°C in January 2016, breaking the previous record of 36.5°C from November 2015 (Strydom, 2016). Studies show that above a threshold of approximately 18.7°C apparent temperature, all-cause mortality rises by 0.9% per 1°C increase and by 2.1% per 1°C among seniors (Wichmann, 2017; Scovronick et al., 2018). While official records counted only 11 direct heat-related deaths in a recent assessment, researchers estimate the actual heat burden to be much higher (Chersich, 2023).

Climate models project substantial warming for Johannesburg. By late-century under a high-emissions scenario, mean annual temperatures in interior South Africa could rise 6–7°C above late 20th-century baselines (Engelbrecht and et al., 2015). Even by 2050, Johannesburg may warm by approximately 2°C if global emissions remain high (Engelbrecht and et al., 2015; Souverijns et al., 2022), with hot nights (minimum temperature >20°C) projected to quadruple from about 10 per year to approximately 40 per year – and up to 100 in the city's most heat-prone neighbourhoods (World Bank Cities Support Program, 2024). Researchers estimate an additional 3–4 weeks of very hot days per year by mid-century (Garland and et al., 2015), and the IPCC warns that beyond +2°C of global warming, heat-attributable mortality and morbidity in Africa will sharply rise (Intergovernmental Panel on Climate Change, 2022).

1.2 Socio-Spatial Inequity and Heat Vulnerability

Johannesburg's subtropical highland climate features hot summer days (October-April), often with afternoon thundershowers, and dry, sunny winter days with cold nights (May-September) (Tyson and Preston-Whyte, 2000). However, the city's socio-spatial layout – largely a legacy of apartheid-era planning – significantly shapes heat vulnerability patterns today. Historically, policies created wealthy, low-density suburbs with ample green spaces alongside dense townships with minimal vegetation or services (Giombini and Thorn, 2022; Venter et al., 2020). This has resulted in stark contrasts in urban heat exposure, with lush neighbourhoods enjoying cooler microclimates, while nearby townships can be approximately 6°C hotter than the surrounding countryside (World Bank Cities Support Program, 2024; U.N. Habitat, 2023).

Housing quality differences further exacerbate exposure: informal housing can become significantly hotter, with indoor temperatures up to 15°C higher than in modern housing during the day (Naicker and et al., 2017). Apartheid geography has effectively embedded vulnerability into Johannesburg's landscape, clustering heat risk in marginalized communities (Strauss, 2019).

1.3 Research Context and Positionality

This research builds upon ongoing work at the HE^{2}AT Center, which has established baseline heat vulnerability assessment frameworks for South African cities (Jack et al., 2024). As a researcher within this initiative, I bring expertise in data science and machine learning methods to extend these frameworks through advanced statistical techniques. My position at the intersection of public health and climate science informs a multidisciplinary approach that aims to translate complex data into actionable health interventions. This work specifically addresses gaps in the Heat Center's existing vulnerability assessments by developing dynamic predictive models that can account for temporal changes in vulnerability patterns across Johannesburg's diverse socioeconomic landscape.

1.4 Conceptual Framework and Research Gaps

This research employs a comprehensive framework for heat vulnerability encompassing three interconnected dimensions: exposure, sensitivity, and adaptive capacity (Intergovernmental Panel on Climate Change, 2022). Figure 1 illustrates these relationships and their determinants within Johannesburg's urban context.

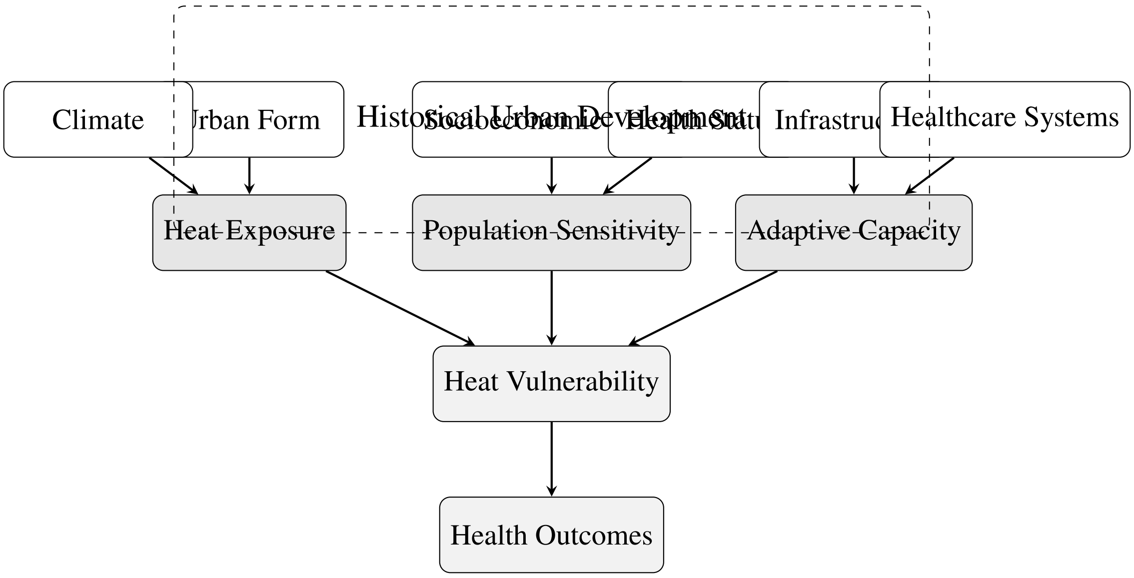

Figure 1: Conceptual framework of heat vulnerability in Johannesburg, highlighting the three primary dimensions (exposure, sensitivity, adaptive capacity) and their determinants. The dashed outline indicates components influenced by historical urban development patterns including apartheid-era planning.

These components interact as follows:

- *Exposure* refers to heat stress degree and duration, with dense urban areas showing significantly higher surface temperatures (up to 5°C) compared to well-vegetated neighbourhoods (Li and et al., 2017; Santamouris, 2015).
- *Sensitivity* reflects population susceptibility influenced by socio-economic conditions and health status, with chronic conditions significantly increasing heat-related health risks (Watts and et al., 2023; Khosla et al., 2021; Souverijns et al., 2022).
- *Adaptive capacity* depends on access to healthcare, cooling infrastructure, and social support systems (Ansah and et al., 2024), with limited access to healthcare affecting vulnerability to heat, as demonstrated by increased heat-related mortality in areas with restricted medical services (Murage and et al., 2020).

1.5 Recent Evidence on Heat Mortality

Recent studies have strengthened our understanding of heat-mortality relationships in African urban contexts. Parker et al. (Parker et al., 2023) demonstrated that in Johannesburg, heat vulnerability clusters in historically disadvantaged areas, with principal component analysis revealing environmental exposure explaining 31.5% of variance, followed by health status (12.8%) and socioeconomic conditions (12.3%). Roffe et al. (Roffe et al., 2023) projected intensification of heat stress conditions across southern Africa, with urban areas experiencing disproportionate impacts. The health impacts are substantial, with an estimated 5 million people globally dying annually from suboptimal temperatures, and approximately 37% of heat-related mortality attributable to human-induced climate change (The Lancet, 2024). These findings underscore the urgency of developing locally calibrated heat-health early warning systems for rapidly urbanizing African cities.

1.6 Research Gaps in Urban African Contexts

Despite growing climate health research worldwide, significant gaps remain in understanding the dynamics of heat health in urban African contexts. Most existing studies focus on high-income regions, overlooking the distinct characteristics of African cities (Khine and Langkulsen, 2023; Pasquini et al., 2020). Key limitations include:

1. **Scarcity of African Urban Heat-Health Studies:** Limited research examining the impacts of heat on health in African urban settings hinders the development of region-specific interventions (Ncongwane et al., 2021; Wright et al., 2019).
2. **Siloed Disciplinary Approaches:** Lack of interdisciplinary collaboration results in fragmented insights that fail to capture the multifaceted nature of heat-health challenges (Jack et al., 2024).
3. **Unique Urban Challenges:** African cities face distinct issues stemming from historical development patterns, unique disease profiles, and diverse urban morphologies requiring tailored research approaches (Giombini and Thorn, 2022; Venter et al., 2020).

Johannesburg exemplifies these challenges through its historical development and urban disparities (Strauss, 2019), high disease burden from both communicable and non-communicable diseases (Wright et al., 2021), and accelerated warming projections (Engelbrecht and et al., 2015; Souverijns et al., 2022). Addressing these gaps is crucial for developing heat health strategies tailored to African urban contexts.

2 Aims and Objectives

2.1 Primary Aim

This research aims to deepen our understanding of spatially and demographically stratified heat-health interactions in Johannesburg, developing evidence-based approaches to mitigate health risks from rising urban temperatures. By examining the complex interplay between urban heat, socioeconomic factors, and health outcomes, this project seeks to bridge critical knowledge gaps in urban climate-health research within the African context, with potential applications for similar urban environments globally.

2.2 Specific Objectives

This research is structured around three interconnected objectives that progress from descriptive analysis to explanatory insights and ultimately to predictive capabilities:

2.2.1 Objective 1: Intra-urban Heat Vulnerability Mapping

The first objective centres on comprehensive spatial vulnerability assessment to identify patterns of heat risk across Johannesburg's diverse urban landscape. This mapping process will examine how historical urban development patterns, particularly those influenced by apartheid-era planning, continue to shape contemporary heat vulnerability. Through this assessment, we will identify priority intervention areas and quantify multidimensional vulnerability patterns with particular emphasis on socio-spatial inequities. This objective will result in detailed vulnerability characterization that can directly inform targeted urban planning and public health interventions.

2.2.2 Objective 2: Heat-Health Dynamics Exploration

Building on the vulnerability assessment, the second objective aims to unravel complex heat-health relationships by investigating the underlying physiological mechanisms through which heat exposure affects human health. Through a two-stage modelling approach, we will first generate hypotheses about key physiological pathways and then systematically test these hypotheses using targeted feature engineering. This component will focus particularly on metabolic, renal, and inflammatory responses to heat stress, examining how these systems interact across multiple temporal scales.

The research will explore non-linear climate-health interactions across different temperature thresholds while distinguishing between immediate (0-24 hours), short-term (1-7 days), and medium-term (7-30 days) physiological effects. By applying interpretable machine learning techniques with clinically validated pathway diagrams, we seek to establish causal mechanisms that explain differential vulnerability between population groups and identify critical intervention points for health protection during heat events. This explanatory approach moves beyond simple statistical associations to provide mechanistic insights to inform targeted clinical and public health responses.

2.2.3 Objective 3: Stratified Heat-Health Prediction Modeling

The final objective leverages insights from the previous components to develop predictive frameworks for heat-related health outcomes in Johannesburg. These predictive capabilities will be stratified both geographically and demographically to account for differential vulnerability across the city's diverse population. The objective seeks to identify specific risk conditions at various temperature thresholds, enabling the development of targeted early warning systems. Special attention will be given to demographic-specific risk stratification for vulnerable populations, including the elderly, those with pre-existing conditions, and communities with limited adaptive capacity. The resulting predictive tools will provide actionable intelligence for healthcare systems and emergency response planning.

Together, these three objectives form an integrated research framework that addresses the spatial dimension (identifying vulnerable areas), the physiological dimension (understanding

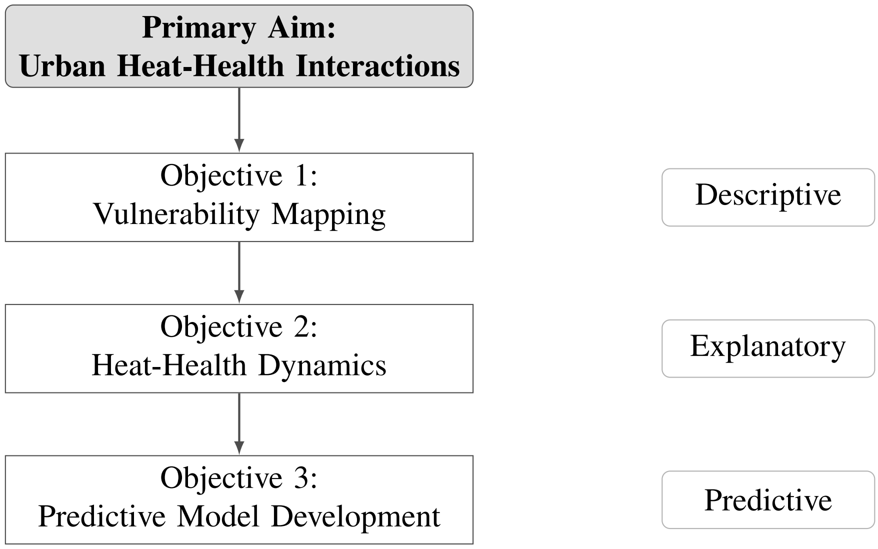

Figure 2: Hierarchical framework of research objectives showing progression from descriptive to predictive approaches.

underlying mechanisms), and the temporal dimension (developing predictive capabilities). This multi-dimensional approach will provide a comprehensive understanding of heat-health interactions in Johannesburg, with implications for urban planning, public health interventions, and climate adaptation strategies. The research design intentionally builds each objective upon the findings of the previous one, creating a cohesive analytical pathway from vulnerability characterization to mechanistic understanding to practical solution development. This progression from descriptive to explanatory to predictive aims reflects the holistic nature of the heat-health challenge and enables development of targeted, evidence-based interventions.

3 Methodology

3.1 Study Design and Setting

This research employs a multi-method approach integrating geospatial analysis, statistical modelling, and machine learning to investigate heat-health relationships in Johannesburg, South Africa. Johannesburg presents an ideal study setting due to its subtropical highland climate, significant socioeconomic disparities, pronounced Urban Heat Island effects, diverse urban morphology, and high prevalence of both communicable diseases and non-communicable conditions that may interact with heat vulnerability patterns.

3.2 Data Sources and Collection

Health data spanning 2000–2022 will be drawn from clinical trials and cohort studies conducted in Johannesburg that meet defined inclusion criteria: cohorts or trials with 200 adult participants, prospectively collected data, comprehensive clinical and laboratory variables, and appropriate ethics approvals. The dataset will include clinical indicators and laboratory tests covering renal, metabolic and inflammatory markers and demographic/socioeconomic factors. Given potential confounding effects, 2020–2022 data will undergo stratified analysis to account for pandemic-related influences on healthcare utilization and outcomes. A comprehensive overview of all data sources is provided in Appendix Table 1.

3.2.1 Environmental and Socioeconomic Data

Environmental parameters will be derived from multiple validated sources, including Landsat 8 (USGS and NASA, 2013), Sentinel-2 (European Space Agency, 2015), ERA5 reanalysis (Hersbach et al., 2017), and the Copernicus Climate Data Store (Copernicus, 2020). These data streams will provide high-resolution measurements of Land Surface Temperature, vegetation indices, air temperature, and derived heat metrics. Socioeconomic determinants will be incorporated through ward-level data from the Gauteng City-Region Observatory Quality of Life Surveys (Gauteng Observatory, 2021), capturing housing quality, infrastructure adequacy, and healthcare access. The data processing and integration workflow is detailed in Appendix Table 2, with specific security measures outlined in Appendix Table 3.

3.2.2 Limitations of Data Sources

While the Gauteng City-Regional Observatory (GCRO) Quality of Life Surveys provide comprehensive socioeconomic indicators, they have notable limitations including temporal gaps between survey waves, varying sample sizes across wards, and potential response biases in informal settlements. These data were selected due to their unparalleled ward-level resolution and longitudinal consistency compared to alternatives such as census data (which lacks recency) or satellite-derived proxies (which lack socioeconomic depth). To mitigate these limitations, we will implement three strategies: (1) temporal harmonization through interpolation techniques for missing years, (2) confidence-weighted analysis based on ward-level sample sizes, and (3) triangulation with alternative data sources where available.

3.2.3 COVID-19 Considerations

To account for COVID-19's potential confounding effects on health systems and data quality during 2020-2022, we will employ a multi-faceted approach. First, interrupted time series analysis will quantify pandemic-related discontinuities in healthcare utilization patterns. Second, stratified analysis will separately examine pre-pandemic, pandemic, and post-pandemic periods to identify divergent heat-health relationships. Third, sensitivity analyses will test model stability with and without pandemic-period data. This approach allows us to isolate pandemic effects while maintaining analytical integrity across the full study period, addressing potential biases in both environmental exposure assessment and health outcome recording during this anomalous period.

3.2.4 Sample Size Justification

The study's statistical power is ensured through comprehensive data coverage and adequate sample sizes. For vulnerability mapping, data from all 135 wards provide complete spatial coverage. For the explanatory and predictive modelling components, the combined dataset of approximately 5,000–7,000 records substantially exceeds the minimum requirement of 10-20 observations per predictive feature (with 20-25 features anticipated). This sample size provides >80% power to detect clinically meaningful effects, calculated as:

$$n = \frac{2(Z_{\alpha/2} + Z_{\beta})^2 \sigma^2}{\Delta^2} \tag{1}$$

where $Z_{\alpha/2}$ and Z_{β} are standard normal deviates for type I and II errors, σ^2 is the expected variance, and Δ is the minimal detectable effect size.

3.3 Objective 1: Vulnerability Mapping

Geospatial data will undergo rigorous preprocessing, including normalization, completeness assessments, and spatial harmonization. The methodology employs Geographically Weighted

Principal Component Analysis (Harris et al., 2011) to develop a comprehensive Heat Vulnerability Index that accounts for spatial non-stationarity in vulnerability relationships. The local GW-PCA model at location i can be formulated as:

$$X_i = V_i \Lambda_i V_i^T + \varepsilon_i \tag{2}$$

where X_i represents the local data matrix, V_i contains the local eigenvectors, Λ_i is the diagonal matrix of local eigenvalues, and ε_i is the error term.

Spatial clustering analysis using Local Indicators of Spatial Association will identify significant (p<0.05) vulnerability hot spots and outliers (Anselin, 1995). The local Moran's I statistic will be calculated as:

$$I_i = z_i \sum_j w_{ij} z_j \tag{3}$$

where z_i and z_j are the standardized values of the vulnerability index, and $w_i j$ represents the spatial weight between locations i and j. Given the sensitive nature of location data, geographic de-identification techniques outlined in Appendix Table 4 will be employed to protect privacy while maintaining analytical validity.

3.3.1 Mixed Socioeconomic Areas

To accurately identify areas with mixed socioeconomic classes, particularly relevant in Johannesburg's complex urban landscape, we will employ multi-scale analysis techniques that account for within-ward heterogeneity. Recent research on Johannesburg's heat vulnerability (Parker et al., 2023) demonstrates that ward-level aggregation can mask significant internal variations in both environmental exposure and socioeconomic conditions. Our methodology addresses this through: (1) sub-ward disaggregation using high-resolution (30m) satellite data to identify microscale heat exposure patterns; (2) socioeconomic inequality indices calculated at multiple spatial scales; and (3) local indicators of spatial association to detect statistically significant socioeconomic clustering patterns within larger administrative units. This approach allows detection of vulnerable pockets within otherwise affluent areas and resilient communities within generally vulnerable regions.

3.4 Objective 2: Explanatory Modeling

The explanatory modelling framework employs a two-stage clinical-computational approach examining both physiological pathways and temporal effects of heat exposure. This framework harmonizes data from multiple cohort studies to investigate underlying mechanisms across different physiological systems.

3.4.1 Clinical Pathway Validation

Working with clinical experts, we will develop comprehensive physiological pathway diagrams representing hypothesized heat-health mechanisms at multiple biological levels. These diagrams will be rigorously validated through expert consensus panels and systematic literature review, establishing physiologically plausible causal pathways before computational analysis. This clinical foundation ensures that subsequent machine learning insights remain grounded in biological reality while allowing for novel relationship discovery.

3.4.2 Multi-System Physiological Analysis

Physiologically, the analysis will systematically investigate how heat stress affects renal function through creatinine, electrolytes, and GFR measurements; metabolic pathways including glucose dysregulation, insulin sensitivity changes, and mitochondrial function; and inflammatory responses through markers such as CRP, IL-6, and other cytokine profiles (Desai et al., 2023). Of particular focus will be the metabolic stress pathway through which heat exposure influences glucose metabolism and overall metabolic health. Appendix F.7 (Metabolic Pathway Analysis) provides visualisations of these pathways.

3.4.3 Two-Stage Modeling Approach

Our analytical strategy follows a structured two-stage process that begins with hypothesis generation and concludes with targeted hypothesis testing. In the initial hypothesis generation stage, we will apply interpretable machine learning techniques to identify potential heat-health relationships without imposing strong a priori assumptions (Schwartz et al., 2022). This first stage employs Random Forest models with permutation importance metrics to identify key predictive factors, XGBoost models with SHAP (SHapley Additive exPlanations) values to quantify feature contributions and interactions, and LIME (Local Interpretable Model-agnostic Explanations) to decompose predictions for individual cases. The SHAP value for feature j is defined as:

$$\phi_j = \sum_{S \subseteq N \setminus \{j\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [f_x(S \cup \{j\}) - f_x(S)] \tag{4}$$

where N is the set of all features, S is a subset of features, $f_x(S)$ is the model prediction for feature set S, and the summation is over all possible feature subsets.

The second stage focuses on hypothesis testing through targeted feature engineering. Based on insights from clinical pathways and initial model interpretations, we will develop composite variables representing specific physiological mechanisms such as cumulative heat stress indices, dehydration risk scores, and inflammatory burden metrics. Additional engineered features will include integrated heat-humidity stress indices capturing physiological strain, time-weighted exposure variables reflecting adaptation mechanisms, interaction terms between existing comorbidities and heat metrics, and system-specific vulnerability scores derived from multiple biomarkers. These engineered features will be tested in subsequent models to evaluate their physiological relevance and predictive power, creating an iterative feedback loop between clinical understanding and data-driven discovery.

3.4.4 Validation Framework

The hypothesized relationships and mechanisms identified through this process will undergo formal validation through multiple complementary approaches. We will employ K-fold cross-validation stratified by demographic and clinical subgroups to assess model stability across different populations. Confidence intervals will be established through bootstrapping with 1,000 resamples, while sensitivity analyses across varying model parameters and threshold definitions will test the robustness of modelling assumptions. Permutation testing will help distinguish true effects from chance correlations, and distributed lag modelling will validate temporal relationships. This comprehensive validation framework ensures findings reflect genuine heat-health relationships rather than statistical artefacts. Notably, the final DLNM models serve as a rigorous statistical validation of the mechanistic hypotheses generated through the interpretable machine learning and feature engineering process(Gasparrini et al., 2015).

This iterative clinical-computational approach explicitly incorporates expert knowledge into the machine learning pipeline while allowing the data to reveal relationships that may confirm, refine, or contradict clinically-derived hypotheses. The resulting insights will establish a foundation for targeted interventions based on validated physiological mechanisms rather than purely statistical associations.

3.4.5 Causal Machine Learning Approaches

To enhance the explanatory power of our models and address the limitation of conventional predictive approaches, we will incorporate causal machine learning methods that go beyond correlative relationships to identify potential causal mechanisms. Following approaches developed by Runge et al. (2019), we will employ causal discovery algorithms to remove spurious links in our predictive models, focusing on "climate invariant" causal drivers that remain robust across different temperature regimes (Beucler et al., 2021). This methodology allows for more reliable identification of intervention points by distinguishing between direct causal effects and mere statistical associations.

Specifically, we will implement three causal learning approaches: (1) causal structure learning using PC algorithm and conditional independence tests to construct directed acyclic graphs of heat-health relationships; (2) causal effect estimation techniques including doubly robust estimation and targeted maximum likelihood estimation to quantify the magnitude of causal impacts; and (3) causal mediation analysis to identify pathways through which heat exposure influences health outcomes. This causal framework will be particularly valuable for distinguishing between factors that mediate heat vulnerability (potential intervention targets) versus those that merely indicate vulnerability (monitoring targets), informing more effective climate adaptation strategies.

3.5 Objective 3: Predictive Modeling

The predictive framework integrates environmental, socioeconomic, and health data to forecast heat-related health outcomes within a 24–72 hour horizon. The methodology employs advanced feature engineering combining domain knowledge with statistical selection techniques including mutual information analysis and recursive feature elimination with cross-validation to identify optimal predictive variables while controlling for collinearity.

Rather than relying on a single algorithm, the approach employs a carefully calibrated ensemble that combines gradient boosting methods, random forests, and other algorithms through a meta-learning framework. For the gradient boosting component, the objective function being optimized is:

$$\mathcal{L}(\phi) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{5}$$

where l is the loss function comparing the prediction $\hat{y}_i$ with the target y_i, and Ω represents the regularization term controlling model complexity.

The validation strategy employs forward-chaining time-series cross-validation to realistically simulate forecasting scenarios, with performance evaluated using metrics including area under the receiver operating characteristic curve (AUC-ROC):

3.5.1 Model Accuracy and Reliability Measures

To ensure predictive model accuracy and reliability, particularly across Johannesburg's diverse urban contexts, we will implement a comprehensive validation framework. This includes: (1) spatially stratified cross-validation to test model performance across different urban morphologies; (2) out-of-distribution detection methods to identify and address anomalous predictions; (3) uncertainty quantification through conformal prediction intervals; and (4) regular recalibration procedures to maintain model accuracy over time. Our model evaluation will employ metrics beyond traditional accuracy measures (AUC-ROC, F1) to include reliability diagrams, expected calibration error, and variable-specific sensitivity analyses. This approach ensures that predictions remain valid across the city's diverse built environments and socioeconomic contexts, with explicit reliability assessments for informal settlements and areas undergoing rapid urban transformation.

3.6 Forecast Window Selection

The 24–72 hour forecast window was selected based on three key considerations. First, physiological evidence indicates that significant heat-health impacts manifest within 1-3 days of exposure (Gasparrini et al., 2015; Kinney et al., 2020), making this the critical intervention window for preventative healthcare measures. Second, meteorological forecast accuracy significantly degrades beyond 72 hours, particularly for urban microclimates, limiting the actionable reliability of longer-term predictions. Third, stakeholder consultations with healthcare providers indicated that a 1-3 day window optimally balances advance warning with operational response capabilities in Johannesburg's healthcare system. Sensitivity analysis will test alternative windows (12–24 hours and 72–120 hours) to assess the stability of predictive relationships across different temporal scales.

$$\text{AUC-ROC} = \int_0^1 TPR(FPR^{-1}(t))dt \tag{6}$$

where TPR is the true positive rate and FPR is the false positive rate at different classification thresholds.

The modelling output translates into actionable intelligence through stratified risk profiles, geospatial risk mapping, and population-specific threshold recommendations. This translation from statistical prediction to practical application ensures research findings can directly inform heat-health interventions and resource allocation for vulnerable populations in Johannesburg.

4 Ethical Considerations

This study received ethical approval from the Wits Human Research Ethics Committee (reference 220606) and complies with U.S. Department of Health and Human Services regulations (45 CFR 46). Privacy protection includes data minimization, secure servers with restricted access, and geographical jittering/aggregation in compliance with South Africa's Protection of Personal

Information Act (POPIA, 2013).

For secondary data usage, contractual guarantees from data providers confirm appropriate consent practices. Ethical risks—such as re-identification, secondary consent issues, and community stigmatization—are mitigated through data aggregation, secure storage, broad consent waivers, and responsible community engagement.

Detailed POPIA compliance framework and data protection procedures are provided in Appendix B.

5 Expected Outcomes and Impact

5.1 Research Outputs

This doctoral research will generate several significant outputs, including a spatially explicit heat vulnerability index for Johannesburg, a comprehensive explanatory model of urban heat-health relationships, and a validated predictive framework for heat-health outcomes under varied climate scenarios. The work will also produce evidence-based policy recommendations for city planning and public health systems, culminating in at least three peer-reviewed publications addressing vulnerability mapping, explanatory modelling, and predictive applications.

5.2 Anticipated Impact and Infrastructural Justice Framing

The research is expected to yield significant impacts by enhancing understanding of fine-scale spatial patterns of heat vulnerability in Johannesburg, improving heat-health early warning systems, and identifying priority areas for targeted intervention. This will enable more efficient resource allocation and provide robust, evidence-based policy guidance for urban heat adaptation, particularly for informal settlements and densely populated areas. These methodological advances in integrated urban heat health research can be adapted for other African urban contexts.

5.2.1 Climate-Resilient Cities and Infrastructural Justice

Beyond its analytical contributions, this research serves as a platform for advancing infrastructural justice in the context of urban climate resilience. By explicitly mapping the distribution of heat vulnerability in relation to historical urban development patterns, the work highlights how infrastructure inequities—in green spaces, cooling resources, healthcare access, and housing quality—create differential climate risk exposure along socioeconomic and racial lines. This framing recognizes that vulnerability to heat is not merely a function of temperature but a manifestation of systemic infrastructure decisions that have unequally distributed adaptive capacity across the urban landscape.

The project's findings will inform a justice-oriented approach to infrastructure development that prioritizes addressing historical imbalances in resource allocation. Specifically, results will identify areas where structural interventions (green infrastructure expansion, cooling centers, healthcare facility development) are most urgently needed to reduce disproportionate vulnerability. This focus on infrastructural justice extends beyond technical solutions to incorporate community governance structures and local knowledge systems in designing context-appropriate adaptations. By framing climate resilience through an equity lens, this research contributes to transformative urban planning that addresses current vulnerability while preventing the perpetuation of historical injustices in future climate adaptation efforts.

5.3 Knowledge Translation

Research findings will be strategically disseminated through peer-reviewed publications in journals spanning public health, climate science, and urban planning disciplines. Additional dissemination channels include presentations at national and international conferences, targeted policy briefs for local and national authorities, community engagement workshops in vulnerable areas, and interactive data visualization tools to facilitate stakeholder decision-making. Detailed impact assessment metrics and comprehensive dissemination strategies are provided in Appendix D.

6 Project Timeline

This 36-month PhD research will progress through four key phases, with critical milestones at months 3 (M1: protocol finalization), 9 (M2: data acquisition completion), 18 (M3: vulnerability mapping), and 30 (M4: model validation), culminating in thesis submission at month 36. Figure 3 presents a detailed timeline of research activities including faculty assessment milestones. The research timeline includes four major publications: a protocol paper titled "Leveraging

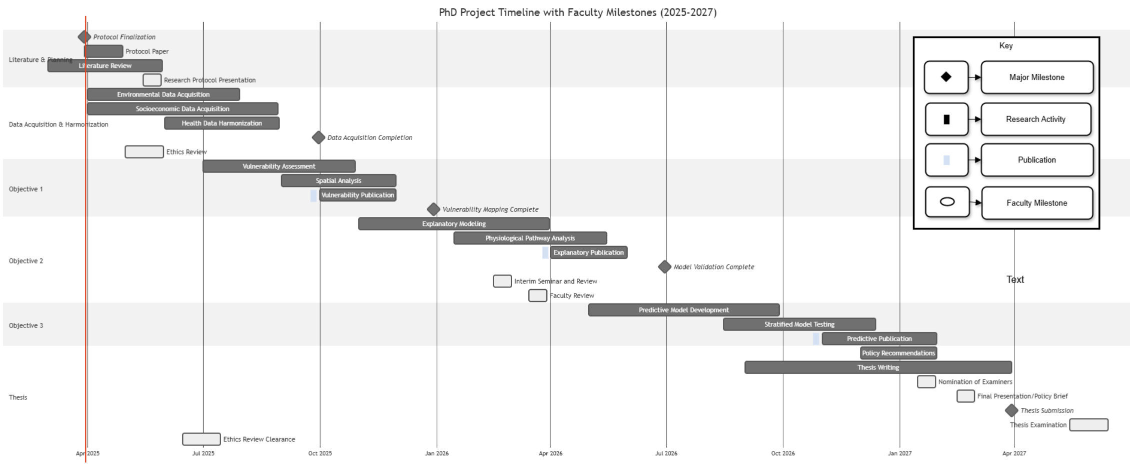

Figure 3: Project Timeline with Key Research Milestones and Publications

data science and machine learning for urban climate adaptation in two major African cities: a HE2AT Center study protocol" Jack et al. (2024) (Month 4), followed by three objective-specific publications: "Quantifying intra-urban socio-economic and environmental vulnerability to extreme heat events in Johannesburg, South Africa" (Month 9), "Uncovering Heat-Health Pathways in Johannesburg: An Explanatory Machine Learning Approach Using Harmonized Urban Cohort Data" (Month 19), and "Developing Stratified Heat-Health Early Warning Systems for Johannesburg: Integration of Vulnerability Mapping and Predictive Modeling" (Month 33). Primary risks include data accessibility challenges, computational constraints, and model performance issues. Detailed timelines, activities, and risk mitigation strategies are provided in Appendix C.

7 Supervision

This research will be conducted under the guidance of a multidisciplinary supervisory team with expertise spanning clinical epidemiology, public health, statistical modeling, and climate science. The team includes Dr. Admire Chikandiwa (Wits University), Professor Matthew Chersich (Trinity College/Wits University), Professor Akbar Waljee (University of Michigan), and Dr. Christopher Jack (University of Cape Town).

Monthly progress meetings and bi-weekly consultations will ensure consistent guidance throughout the research process. A detailed supervision framework, including supervisor profiles and advisory committee information, is provided in Appendix E.

8 References

References

Analitis, A., De' Donato, F., Scortichini, M., Lanki, T., Basagana, X., Ballester, F., Astrom, C., Paldy, A., Pascal, M., Gasparrini, A., et al. (2018). Synergistic effects of ambient temperature and air pollution on health in europe: results from the phase project. *International journal of environmental research and public health*, 15(9):1856.

Ansah, E. and et al. (2024). Health systems response to climate change adaptation: a scoping review of global evidence. *BMC Public Health*, 24(1):2015.

Anselin, L. (1995). Local indicators of spatial association—lisa. *Geographical Analysis*, 27(2):93–115.

Beucler, T., Pritchard, M., Rasp, S., Ott, J., Baldi, P., and Gentine, P. (2021). Achieving conservation of energy in neural network emulators for climate modeling. *Journal of Advances in Modeling Earth Systems*, 13(9):e2021MS002589.

Chersich, M. (2023). "africa is getting hotter" – wits university news.

Copernicus (2020). Climate Data Store. https://cds.climate.copernicus.eu/. Accessed: 2025-03-30.

Desai, Y., Khraishah, H., and Alahmad, B. (2023). Heat and the heart. *Yale J Biol Med*, 96(2):197–203.

Engelbrecht, F. and et al. (2015). Projections of rapidly rising heat extremes in south africa. *Journal of Climate*.

European Space Agency (2015). Sentinel-2 Mission. https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-2. Accessed: 2025-03-30.

Garland, R. and et al. (2015). Estimating future heat-related mortality in gauteng. *Global Health Action*, 7(1):26713.

Gasparrini, A., Guo, Y., Hashizume, M., Lavigne, E., Zanobetti, A., Schwartz, J., Tobias, A., Tong, S., Rocklöv, J., Forsberg, B., et al. (2015). Mortality risk attributable to high and low ambient temperature: a multicountry observational study. *The Lancet*, 386(9991):369–375.

Gauteng Observatory (2021). Quality of Life Survey. https://gcro.ac.za/research/project/detail/quality-of-life-survey. Ward-level data. Accessed: 2025-03-30.

Giombini, V. and Thorn, J. P. (2022). Urban green spaces in a post-apartheid city: Challenges and opportunities for nature-based solutions. pages 207–219.

Harris, P., Brunsdon, C., and Charlton, M. (2011). Geographically weighted principal components analysis. *International Journal of Geographical Information Science*, 25:1717–1736.

Hersbach, H. et al. (2017). ERA5 Reanalysis Data. https://cds.climate.copernicus.eu/. Copernicus Climate Change Service. Accessed: 2025-03-30.

Intergovernmental Panel on Climate Change (2022). Climate change 2022: Impacts, adaptation and vulnerability. contribution of working group ii to the sixth assessment report of the intergovernmental panel on climate change.

Jack, C., Parker, C., Kouakou, Y. E., Joubert, B., McAllister, K. A., Ilias, M., Maimela, G., Chersich, M., Makhanya, S., Luchters, S., Makanga, P. T., Vos, E., Ebi, K. L., Koné, B., Waljee, A. K., and Cissé, G. (2024). Leveraging data science and machine learning for urban climate adaptation in two major african cities: a he2at center study protocol. *BMJ Open*, 14(6).

Khine, M. M. and Langkulsen, U. (2023). The implications of climate change on health among vulnerable populations in south africa: A systematic review. *International Journal of Environmental Research and Public Health*, 20(4):3468.

Khosla, R., Jani, A., and Perera, R. (2021). Health risks of extreme heat. *BMJ*, 375:n2438.

Kinney, P. L., Schwartz, J., Pascal, M., Petkova, E., Le Tertre, A., Medina, S., and Vautard, R. (2020). Climate change, heat, and mortality in the tropical urban area of san juan, puerto rico. *International Journal of Environmental Research and Public Health*, 17(11):4037.

Li, X. and et al. (2017). The surface urban heat island response to urban expansion: A panel analysis for the conterminous united states. *Science of The Total Environment*, 605-606:426–435.

Murage, P. and et al. (2020). What factors increase the risk of heat-related mortality? a case-crossover study in london. *Environment International*, 134:105292.

Naicker, N. and et al. (2017). Indoor temperatures in low-cost housing in johannesburg, south africa. *International Journal of Environmental Research and Public Health*, 14(11):1410.

Ncongwane, K. P., Botai, J. O., Sivakumar, V., Botai, C. M., Adeyemi, A., Wernecke, B., Mphahlele, M., Khakhau, R. M., Zwane, N., Chikoore, H., et al. (2021). A literature review of the impacts of heat stress on human health across africa. *Sustainability*, 13(9):5312.

Parker, C., Chersich, M., Waljee, A., and Jack, C. (2023). Quantifying intra-urban socio-economic and environmental vulnerability to extreme heat events in johannesburg, south africa. *Environment International*, 174:107788.

Pasquini, L., van Aardenne, L., Godsmark, C. N., Lee, J., and Jack, C. (2020). Emerging climate change-related public health challenges in africa: A case study of the heat-health vulnerability of informal settlement residents in dar es salaam, tanzania. *Science of the Total Environment*, 747:141355.

Roffe, S. J., Fitchett, J. M., Engelbrecht, C. J., and Engelbrecht, F. A. (2023). Projected intensification and expansion of heat stress and related heat-attributable mortality in africa under anthropogenic climate change. *Earth's Future*, 11(4):e2022EF003338.

Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., Coumou, D., Deyle, E., Glymour, C., Kretschmer, M., Mahecha, M. D., Muñoz-Marí, J., et al. (2019). Inferring causation from time series in earth system sciences. *Nature Communications*, 10(1):2553.

Santamouris, M. (2015). Analyzing the heat island magnitude and characteristics in one hundred asian and australian cities and regions. *Science of The Total Environment*, 512-513:582–598.

Schwartz, R., Vassilev, A., Greene, K., Perine, L., Burt, A., and Hall, P. (2022). Towards a standard for identifying and managing bias in artificial intelligence. *NIST Special Publication*, 1270.

Scovronick, N., Sera, F., Acquaotta, F., Garzena, D., Fratianni, S., Wright, C. Y., and Gasparrini, A. (2018). The association between ambient temperature and mortality in south africa: A time-series analysis. *Environmental research*, 161:229–235.

Souverijns, N., De Ridder, K., and Veldeman, N. (2022). Urban heat in johannesburg and ekurhuleni, south africa: A meter-scale assessment and vulnerability analysis. *Urban Climate*, 46:101331.

Strauss, M. (2019). A historical exposition of spatial injustice and segregated urban settlement in south africa. *Fundamina*, 25(2):135–168.

Strydom, N. (2016). Johannesburg temperature hits record high as south africa drought persists. Reuters News Agency.

The Lancet (2024). Climate crisis, cities, and health. *The Lancet*, 403(10438):1659–1660.

Tyson, P. and Preston-Whyte, R. (2000). *The Weather and Climate of Southern Africa*. Oxford University Press, Cape Town, 2nd edition.

U.N. Habitat (2023). World cities report 2024 – box 4.2: Johannesburg's urban heat rise.

USGS and NASA (2013). Landsat 8 Data. https://www.usgs.gov/landsat-missions/landsat-8. Accessed: 2025-03-30.

Venter, Z. S., Shackleton, C. M., Van Staden, F., Selomane, O., and Masterson, V. A. (2020). Green apartheid: Urban green infrastructure remains unequally distributed across income and race geographies in south africa. *Landscape and Urban Planning*, 203:103889.

Watts, N. and et al. (2023). The 2023 report of the lancet countdown on health and climate change: health at the mercy of fossil fuels. *The Lancet*, 402(10397):1610–1654.

Wichmann, J. (2017). Heat effects of ambient apparent temperature on all-cause mortality in cape town, durban and johannesburg, south africa: 2006-2010. *Science of the Total Environment*, 587:266–272.

World Bank Cities Support Program (2024). Heat mapping by citizen scientists points the way to a cooler future. Business Day (11 March 2024).

Worldometer (2023). Johannesburg south africa population. Retrieved from: https://www.worldometers.info/world-population/.

Wright, C. Y., Dominick, F., Kapwata, T., Bidassey-Manilal, S., Engelbrecht, J. C., Reddy, P., Olwoch, J. M., Garland, R. M., Matooane, M., Wernecke, B., et al. (2019). Socio-economic, infrastructural and health-related risk factors associated with adverse heat-health effects reportedly experienced during hot weather in south africa. *Pan African Medical Journal*, 34.

Wright, C. Y., Kapwata, T., Wernecke, B., Garland, R. M., Nkosi, V., Abramson, M. J., Mathee, A., Kunene, Z., Balakrishna, Y., and Chersich, M. F. (2021). Major climate change-induced risks to human health in south africa. *Environmental Research*, 196:110973.

Appendices

NOTE:

All appendices and supplementary tables are provided in a separate document titled:

"Appendices: Data Management and Methodological Tables for PhD Proposal"

This separation has been implemented to comply with page count requirements.

The appendix document includes:

- **Appendix F: Supplementary Figures**
  - F.1 Project Framework: Heat Vulnerability Assessment
  - F.2 Maximum Temperature Distribution
  - F.3 Seasonal Heat Pattern Analysis
  - F.4 Global Temperature Trends
  - F.5 Urban Heat Stress Patterns
  - F.6 Heat Vulnerability Index Components
  - F.7 Heat-Health Mechanisms Pathways
  - F.8 Data Sources and Integration Framework
- **Data Management and Methodological Tables**
  - Table 1: Key Data Sources for Heat-Health Research
  - Table 2: Data Processing and Integration Workflow
  - Table 3: Data Access Levels and Security Measures
  - Table 4: Geographic De-identification Techniques
  - Table 5: Data Request and Approval Process
  - Table 6: Methodological Approaches for Research Objectives
  - Table 7: Time-Lag Analysis of Heat-Health Effects
- **Appendix A:** Roles and Responsibilities in Data Management
- **Appendix B:** POPIA Compliance Framework
- **Appendix C:** Detailed Project Timeline and Risk Assessment
  - C.1 Comprehensive Timeline of Research Activities