University of the Witwatersrand

School of Public Health

Urban Heat and Health in Johannesburg

A Multidimensional Analysis of Vulnerability, Explanatory Modelling, and Predictive Outcomes

Craig Parker, MSc, MM-PDM, BS

PhD Protocol Presentation

June 12, 2025

HEat and HEalth African Transdisciplinary (HE2AT) Center
NIH Common Fund Award Number U54 TW 012083

Presentation Overview

Research journey from problem identification to methodological innovation and practical application

RESEARCH JOURNEY:

Understanding → Explaining → Predicting → Acting

RESEARCH CONTEXT:

1. **Problem Statement and Context**
   Rising temperatures and health disparities
2. **Primary Aim and Objectives**
   Three interconnected research phases
3. **Analytical Framework**
   Vulnerability components and analytical lenses
4. **Data Sources and Sample**
   Comprehensive multi-domain data ecosystem

RESEARCH IMPLEMENTATION:

5. **Methods**
   Advanced quantitative approaches and validation
6. **Timeline and Milestones**
   36-month research plan with key deliverables
7. **Expected Outcomes and Impact**
   Methodological innovations and knowledge foundation
8. **Supervision Structure**
   Multidisciplinary team and research support

KEY INNOVATION: Geographically weighted approaches for socio-spatial inequality analysis in LMIC urban contexts

CONTRIBUTION TO KNOWLEDGE:

First comprehensive heat-health vulnerability study in African megacity using novel methodological integration

Problem Statement & Context

Rising temperatures create disproportionate health risks across urban populations

THE CHALLENGE:

3-6°C
Temperature increase by 2100 under high emissions

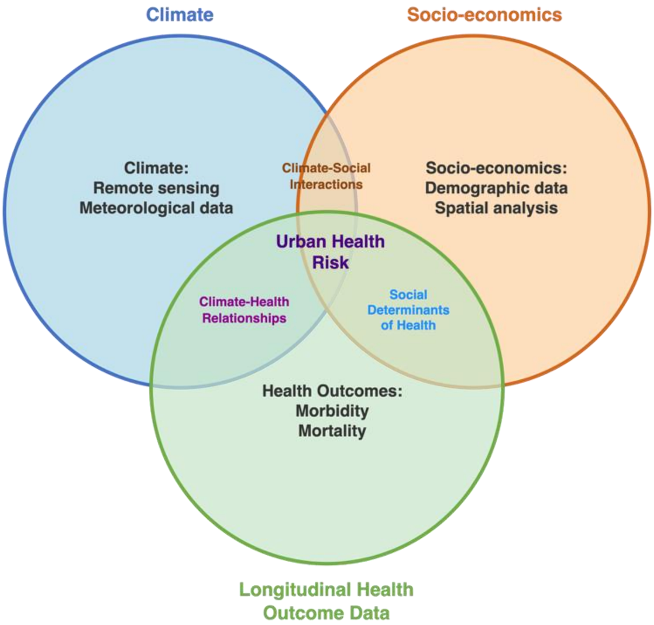
6°C
Temperature differential between townships & suburbs

0.9%
Mortality increase per 1°C above 18.7°C threshold

2.1%
Elderly mortality increase per 1°C temperature rise

RESEARCH GAP:

- Limited heat-health research in LMIC urban contexts
- Siloed approaches miss complex vulnerability patterns
- No predictive frameworks for African cities


Climate
Socio-economics
Climate:
Remote sensing
Meteorological data
Climate-Social
Interactions
Socio-economics:
Demographic data
Spatial analysis
Urban Health
Risk
Climate-Health
Relationships
Social
Determinants
of Health
Health Outcomes:
Morbidity
Mortality
Longitudinal Health
Outcome Data

Primary Aim & Objectives

Sequential approach building from descriptive mapping to predictive capabilities

PRIMARY AIM:

Analyse spatially stratified heat-health interactions in Johannesburg to inform evidence-based approaches to mitigate heat-related health risks

THREE INTERCONNECTED OBJECTIVES:

1 Vulnerability Mapping

(Descriptive)

Comprehensive spatial vulnerability assessment using geographically weighted approaches

- Geographically Weighted PCA
- Local Spatial Association
- Non-stationarity Analysis

2 Heat-Health Dynamics

(Explanatory)

Identify causal mechanisms through advanced explanatory modeling

- Two-stage Analysis
- SHAP/LIME Interpretation
- Causal Machine Learning

3 Predictive Modeling

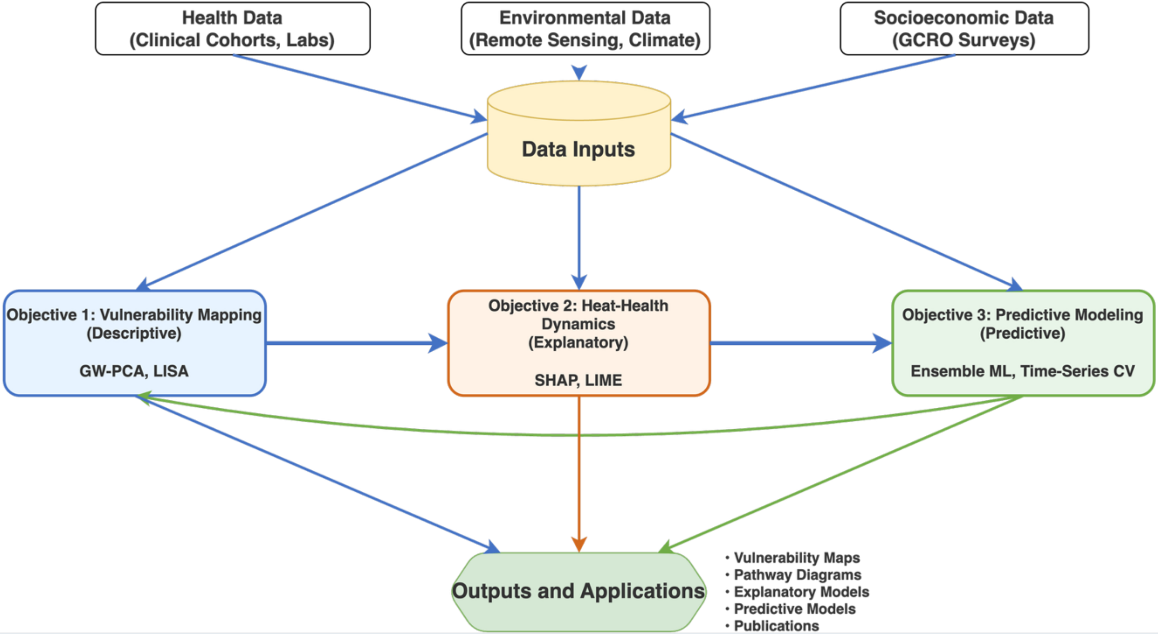
(Predictive)

Develop 24-72 hour heat-health risk forecasting capabilities

- Ensemble Machine Learning
- Uncertainty Quantification
- Spatial Validation

Working Definition of Vulnerability

Vulnerability refers to the **degree to which populations are susceptible to and unable to cope with adverse heat-health effects**, determined by **exposure** to hazards, **sensitivity** to impacts, and **adaptive capacity** influenced by socioeconomic, environmental, and historical factors IPCC (2022)


Health Data
(Clinical Cohorts, Labs)
Environmental Data
(Remote Sensing, Climate)
Socioeconomic Data
(GCRO Surveys)
Data Inputs
Objective 1: Vulnerability Mapping
(Descriptive)
GW-PCA, LISA
Objective 2: Heat-Health
Dynamics
(Explanatory)
SHAP, LIME
Objective 3: Predictive Modeling
(Predictive)
Ensemble ML, Time-Series CV
Outputs and Applications
• Vulnerability Maps
• Pathway Diagrams
• Explanatory Models
• Predictive Models
• Publications

Methods

Comprehensive quantitative research design across three interconnected analytical phases

STUDY DESIGN:

Quantitative research design with spatiotemporal analysis (2000-2022) • Multi-cohort data harmonization • Ward-level spatial units (n=135) • Individual participant data (n=7,000-9,000)

METHOD 1: VULNERABILITY MAPPING

Geographically Weighted Approaches:

- Geographically Weighted PCA (GW-PCA)
- Local Indicators of Spatial Association (LISA)
- Spatial bandwidth optimization
- Non-stationarity analysis

Accounts for spatial heterogeneity in vulnerability relationships across Johannesburg

METHOD 2: HEAT-HEALTH DYNAMICS

Two-Stage Explanatory Framework:

Stage 1: Hypothesis Generation
- Random Forest and XGBoost algorithms
- SHAP and LIME interpretation

Stage 2: Targeted Testing
- Causal machine learning (Double ML)
- Distributed Lag Non-linear Models (DLNM)
- Mediation and effect modification analysis

METHOD 3: PREDICTIVE MODELING

Ensemble Machine Learning:

- Calibrated ensemble combining multiple algorithms
- Forward-chaining time-series cross-validation
- Uncertainty quantification (conformal prediction)
- 24-72 hour forecast window

Features: Environmental data, vulnerability indices, meteorological forecasts

DATA INTEGRATION AND VALIDATION:

Data Integration Process:

- Multi-source harmonization workflow
- Spatial and temporal alignment
- Quality-weighted modeling approach
- POPIA-compliant de-identification
- Missing data imputation strategies

Validation Strategy:

- Spatially stratified cross-validation
- Out-of-distribution detection
- Model calibration assessment
- Sensitivity analysis across urban morphologies
- Clinical expert validation of pathways

ANALYTICAL SOFTWARE:

R (GWmodel, causal-learn), Python (econml, scikit-learn, tensorflow), QGIS (spatial analysis)

References:

Praharaj et al. (2024), Gasparrini et al. (2015), Velasquez et al. (2022), Kinney et al. (2020), GWmodel R package documentation

REGULATORY COMPLIANCE:

✓ Wits Human Research Ethics Committee (HREC) approval - Reference: 220606
✓ U.S. Department of Health and Human Services regulations (45 CFR 46)
✓ Declaration of Helsinki principles for medical research
✓ South African Ethics in Health Research guidelines (2015)
✓ NIH Common Fund ethical standards and oversight

DATA PRIVACY AND PROTECTION:

Technical Safeguards:

- AES-256 encryption for all data storage
- Secure restricted access servers (UCT)
- Geographic jittering and aggregation
- Data minimization principles
- Automated de-identification protocols
- Regular security audits and updates

POPIA Compliance Framework:

- Lawful processing for legitimate research purposes
- Purpose limitation and data minimization
- Designated Information Officer oversight
- Processing records documentation
- Privacy impact assessments
- Data subject participation rights protection

SECONDARY DATA USAGE:

- Contractual guarantees from all data providers confirming appropriate consent practices
- Broad consent waivers for research approved by ethics committees
- Strict compliance with institutional data transfer agreements
- Quality-weighted modeling approach respecting data source limitations
- Community engagement through HE2AT Center collaborative framework

ETHICAL RISK MITIGATION:

Risk Management:
- Re-identification prevention through aggregation
- Community stigmatization mitigation
- Transparent research communication
- Regular ethical review processes

Ongoing Oversight:
- Quarterly compliance audits
- Annual ethics committee reviews
- Continuous stakeholder engagement
- Regular supervisor oversight meetings

References: Protection of Personal Information Act (POPIA) 2013, Declaration of Helsinki (2013), South African Ethics in Health Research (2015)

Timeline & Milestones

36-Month PhD research timeline with clear deliverables and publications

36-MONTH PhD TIMELINE:

Phase	Timeline	Key Deliverable	Publication	Milestone
Protocol Development	Months 1-3	M1: Protocol finalization	Protocol Paper (Published)	M1
Data Acquisition	Months 4-9	M2: Data integration complete	-	M2
Vulnerability Mapping	Months 10-18	M3: Vulnerability maps & Paper 2	Vulnerability Paper	M3
Explanatory Modeling	Months 19-30	M4: Causal models & Paper 3	Explanatory ML Paper	M4
Predictive Modeling	Months 31-33	M5: Predictive framework & Paper 4	Predictive Modeling Paper	M5
Thesis Writing	Months 34-36	M6: Thesis submission	Final Thesis Document	M6

PUBLICATIONS PLAN: 3 papers aligned with each objective plus protocol paper

KEY SUCCESS FACTORS:

- Early stakeholder engagement and data access secured
- Flexible analytical approach to accommodate data limitations
- Strong supervision team with complementary expertise
- Realistic timelines with built-in contingency periods

Expected Outcomes & Impact

Multi-dimensional outcomes advancing methodology and informing practice

RESEARCH OUTPUTS:

- Spatially explicit heat vulnerability index
- Comprehensive explanatory models of heat-health relationships
- Validated predictive framework for heat-health outcomes
- 3 peer-reviewed publications

Publications targeting high-impact journals in public health, climate science, and urban planning disciplines

METHODOLOGICAL INNOVATIONS:

- Integration of GW-PCA with causal machine learning
- Context-specific approaches for socio-spatial inequality
- Dynamic vulnerability assessment over 22 years (2000-2022)

Novel methodological framework applicable to other LMIC urban contexts facing extreme inequality

KNOWLEDGE FOUNDATION FOR:

Evidence-based Adaptation Strategies

Spatial vulnerability patterns inform targeted intervention design and resource allocation decisions

Climate-resilient Urban Planning

Infrastructural justice framework guides equitable urban development and adaptation planning

Health System Preparedness

Predictive capabilities enable proactive health system responses and early warning systems

BROADER IMPACT:

First comprehensive heat-health vulnerability study in African megacity with novel methodological integration for LMIC contexts, providing foundation for climate adaptation across similar urban environments

Note: Development and implementation of interventions falls beyond PhD scope

Supervision Framework

Multidisciplinary supervisory team with complementary expertise and structured mentorship

Supervisor	Affiliation	Expertise
Dr. Admire Chikandiwa	Wits University	Clinical epidemiology
Prof. Matthew Chersich	Trinity/Wits	Climate and health
Prof. Akbar Waljee	Michigan	Statistical modeling, ML

MEETING STRUCTURE:

- Biweekly primary supervision meetings
- Monthly team meetings with all supervisors
- Quarterly progress reviews with formal documentation
- Annual assessment reviews with external input

ADDITIONAL RESEARCH SUPPORT:

- HE2AT Center collaborative research environment with international expertise
- Access to high-performance computing resources and specialized software
- Data partnerships with GCRO, UCT Climate Systems Analysis Group, and Wits Health Consortium

References: HE2AT Center collaborative framework, Multidisciplinary supervision structure, Jack & Parker et al. (2024)

Thank You

Questions & Discussion

craig.parker@witsphr.org

Cutter, S.L., Boruff, B.J., and Shirley, W.L. (2003). Social vulnerability to environmental hazards. Social Science Quarterly, 84(2), 242-261.

Declaration of Helsinki (2013). World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects.

Engelbrecht, F., et al. (2015). Projections of rapidly rising heat extremes in South Africa. Journal of Climate, 28(6), 2434-2451.

European Space Agency (2015). Sentinel-2 Mission. Copernicus Programme. Accessed: 2025-03-30.

Gasparrini, A., 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. Ward-level data. Gauteng City-Region Observatory.

Hersbach, H., et al. (2017). ERA5 Reanalysis Data. Copernicus Climate Change Service Climate Data Store.

Intergovernmental Panel on Climate Change (2022). Climate Change 2022: Impacts, Adaptation and Vulnerability. Cambridge University Press.

Intergovernmental Panel on Climate Change (2024). Climate Change 2024: Synthesis Report. IPCC, Geneva, Switzerland.

Jack, C., Parker, C., et al. (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.

Kinney, P.L., et al. (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., 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.

Niu, S., et al. (2021). A systematic review of heat vulnerability index development and validation. Environmental Research Letters, 16(5), 053007.

Parker, C.J., et al. (2025). Quantifying intra-urban socio-economic and environmental vulnerability to extreme heat events in Johannesburg, South Africa. Research Square. Preprint.

Praharaj, S., Choi, M., and Han, H. (2024). A novel urban heat vulnerability analysis: Integrating machine learning and remote sensing for enhanced insights. Remote Sensing, 16(16), 3032.

ADDITIONAL KEY REFERENCES:

Protection of Personal Information Act (POPIA) (2013). Act No. 4 of 2013. Government of South Africa.

Romanello, M., et al. (2023). The 2023 report of the Lancet Countdown on health and climate change: The imperative for a health-centred response in a world facing irreversible harms. The Lancet, 402(10417), 2022-2060.

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.

South African Ethics in Health Research (2015). Department of Health Guidelines. Government of South Africa.

USGS and NASA (2013). Landsat 8 Data. United States Geological Survey. Accessed: 2025-03-30.

Velasquez, D., et al. (2022). Machine learning-aided causal inference framework for environmental justice. Environmental Science and Technology, 56(4), 2126-2138.

Watts, N., 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.

Note: Complete bibliography with additional methodological and theoretical references available in full protocol document

Data Sources & Sample

Comprehensive data ecosystem spanning health, environment, and socioeconomic domains

HEALTH DATA

- Clinical trials & cohort studies (2000-2022)
- n=7,000-9,000 participants
- Cardiovascular, metabolic, renal, inflammatory markers

Biomarker Categories:
- Heat-sensitive physiological indicators
- Chronic disease markers
- Inflammatory responses

ENVIRONMENTAL DATA

- Landsat 8, Sentinel-2, ERA5, MODIS
- Land surface temperature
- Vegetation indices

Remote Sensing Metrics:
- Urban Heat Island patterns
- Thermal stress indicators
- Built environment characteristics

SOCIOECONOMIC DATA

- GCRO Quality of Life Surveys
- All 135 wards (complete coverage)

Key Indicators:
- Housing quality
- Healthcare access
- Demographics
- Infrastructure availability

↓

INTEGRATED SPATIOTEMPORAL DATABASE

- Harmonized across spatial and temporal dimensions (2000-2022)
- Ward-level aggregation with sub-ward analysis capability
- Quality-weighted modeling with completeness assessment
- POPIA-compliant de-identification and security protocols

COVID-19 CONSIDERATION:

Stratified analysis with pandemic-period adjustments and healthcare utilisation corrections

STATISTICAL POWER:

>80% power to detect clinically meaningful effects with 20-25 predictive features across 135 wards