# Title: Leveraging data science and machine learning for urban climate adaptation in Africa: a HE2AT CENTER STUDY protocol

Authors: Christopher Jack1\*, Craig Parker2\*, Yao Etienne Kouakou3,4, Bonnie R. Joubert5, Kimberly A. McAllister5, Maliha Ilias6, Gloria Maimela2, Matthew Francis Chersich2, Sibusisiwe Makhanya7, Stanley Luchters8,10, Prestige Tatenda Makanga8,11, Etienne Vos7 Kristie Ebi9, Brama Kone3, Gueladio Cisse3 on behalf of the HE2AT Center  
\*Equal first authors

HE2AT Center Group (alphabetical): Abdoulaye Tall, Adja Ferdinand Vanga, Christopher Jack, Craig Mahlasi, Iba Dieudonné Dely, James Mashiyane, Lisa van Aardenne, Madina Doumbia, Nicholas Brink, Pierre Kloppers, Piotr Wolski, Sibusisiwe Makhanya, Tamara Govindasamy, Toby Kurien

Author Affiliations:

1. Climate System Analysis Group, University of Cape Town
2. Wits RHI, University of the Witwatersrand, Johannesburg, South Africa
3. University Peleforo Gon Coulibaly, Korhogo, Côte d’Ivoire
4. Centre Suisse de Recherches Scientifique, Côte d’Ivoire
5. National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Durham, North Carolina, United States of America
6. National Heart Lung and Blood Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland, United States of America
7. IBM Research Africa, Johannesburg
8. Centre for Sexual Health and HIV & AIDS Research (CeSHHAR), Zimbabwe
9. The University of Washington, Seattle, United States of America
10. Liverpool School of Tropical Medicine, Liverpool, UK; Department of Public Health and Primary Care, Ghent University, Belgium
11. Surveying and Geomatics Department, Midlands State University, Gweru, Zimbabwe

Correspondence to Dr Christopher Jack, [cjack@csag.uct.ac.za](mailto:cjack@csag.uct.ac.za)

(Word count: 3846)

ABSTRACT

**Introduction:** African cities, particularly Abidjan and Johannesburg, face challenges of rapid urban growth, informality, and strained health services, compounded by increasing temperatures due to climate change. This study aims to understand the complexities of heat-related health impacts in these cities. The objectives are: 1) mapping intra-urban heat risk and exposure using health, socio-economic, geospatial climate, and satellite imagery data; 2) creating a stratified heat-health outcome forecast model to predict adverse health outcomes; and 3) establishing an Early Warning System for timely heatwave alerts. The ultimate goal is to foster climate-resilient African cities, protecting disproportionately affected populations from heat hazards.

**Methods and Analysis:** The research will acquire health-related datasets from eligible adult clinical trials or cohort studies conducted in Johannesburg and Abidjan between 2000 and 2022. Additional data will be collected, including socio-economic, geospatial climate datasets and satellite imagery. These resources will aid in mapping heat hazards and quantifying heat-health exposure, the extent of elevated risk, and morbidity. Outcomes will be determined using advanced data analysis methods, including statistical evaluation, machine learning, and deep-learning techniques.

**Ethics and Dissemination:** The study, approved by the Wits Human Research Ethics Committee (reference no: 220606), adheres to relevant guidelines and legislation. Data management will follow approved procedures. Results will be disseminated through workshops, community forums, conferences, and publications. Data deposition and curation plans will be established in line with ethical and safety considerations.

(Word count: 228 )

**Keywords:** urban, heat, heatwaves, health, Early Warning Systems, intra-urban, socio-economics and environment, exposure mapping, hazard mapping, heat-related health impacts, African cities, data science and machine learning, temperature

Article Summary

The strengths and limitations of this study

Strengths and Limitations

1. Employs comprehensive data collection from clinical, socio-economic, and remote sensing sources, ensuring a multidimensional analysis of urban heat exposure.
2. Leverages state-of-the-art machine learning techniques for predictive modelling of heat-health outcomes, advancing the field of environmental health research.
3. Cross-disciplinary approach enriches the interpretation of data, linking climate science with public health implications.
4. Risk of sampling bias due to secondary data utilisation, which may influence the representativeness of findings.
5. The spatial resolution of datasets, particularly those capturing microclimatic urban variations, could limit the granularity of exposure assessments.

# 1 Introduction

The study forms parts of the HE2AT Center (HEat and HEalth African Transdisciplinary Center) which is a consortium of partners from South Africa (Universities of Cape Town and Witwatersrand, and IBM- Research Africa), Côte d’Ivoire (University Peleforo Gon Coulibaly), Zimbabwe (CeSHHAR), and the United States (Universities of Michigan and Washington). The Center is funded through the United States NIH “Harnessing Data Science for Health Discovery and Innovation in Africa” (DS-I Africa) program[1].

The study constitutes one of two Research Projects (RPs) within the NIH-funded HE²AT Center. It specifically addresses the complexity of urban spaces about heat-health impacts and the appropriate responses for some particularly disproportionately affected groups. Such groups may encompass residents of impoverished areas, the elderly, individuals with pre-existing health conditions, children, outdoor workers, and inhabitants of densely populated or informal settlements in urban regions[2]. The study recognises that heat-related health risks are not evenly distributed across different socio-demographic groups[3, 4].

# Study setting

Abidjan, located in Côte d'Ivoire, and Johannesburg, in South Africa, are cities experiencing rapid urbanisation—defined as the population shift from rural to urban areas along with the corresponding change in land use—compounded with stress on health services and increasing temperatures owing to climate change[5, 6]. Johannesburg, a diverse metropolis of 6.1 million with a blend of modern and traditional architecture, grapples with health ailments like HIV, tuberculosis, and non-communicable diseases. These are intensified by urbanisation, socio-economic disparities, and broader social determinants of health (SDOH) like education and employment[7-9]. Areas with less vegetation and higher levels of poverty face greater heat impacts, a reflection of the 'Green Apartheid' that characterizes the city's urban forest and its accessibility​​. Similarly, Abidjan, an economic centre with a population of 6.3 million, struggles with diseases such as malaria and non-communicable diseases driven by urbanisation, socio-economic factors, and wider SDOH[10-12].

Both cities represent Urban Heat Islands (UHI), a phenomenon where urban areas exhibit higher temperatures than their rural surroundings due to human activities. Johannesburg's extensive urban forest offers some respite, yet densely populated areas with scarce vegetation endure more severe heat impacts[13]. Conversely, Abidjan's Cocody district is increasingly experiencing the UHI effect due to accelerated urbanisation and land use modifications. These evolving urban landscapes underscore the requirement for holistic health strategies in both cities[14].

Abidjan and Johannesburg were selected due to their unique geographical locations, symbolising West and Southern Africa, respectively, and their varying socio-economic, environmental conditions, and SDOH. This facilitates a comparative analysis offering insights into the diverse health impacts of heat across various African contexts.

# Background/Rationale

Understanding key terms such as disproportionately affected , exposure, and adaptive capacity are crucial in the context of heat-related health risks. Table one defines these concepts and how they are used in this paper.

|  |  |
| --- | --- |
| Concept | Description |
| **Disproportionately affected** | **Sensitivity to heat exposure, particularly notable in urban poor due to factors such as age, health status, and use of heat-amplifying materials [15] [16] [17]** |
| **Exposure** | **Degree of heat contact, affected by geography, urbanization, occupation, and SDOH including living/working conditions [18]. The Urban Heat Island (UHI) effect is a key determinant 19], especially relevant in rapidly urbanizing Africa [19].** |
| **Adaptive Capacity** | **Population's ability to adjust to heat, linked with socio-economic factors, resource access, institutional support, and SDOH [20]. Often diminished in urban poor due to limited access to cooling resources and health services [21] 23].** |

Table 1: Key Concepts and Definitions in Heat Exposure Studies

Research has established a clear link between prolonged ambient temperatures above long-term averages, short-duration heat extremes, and increased mortality and morbidity rates[22, 23]. The World Health Organization projects that by 2030, heat waves will cause nearly 92,000 deaths annually, with sub-Saharan Africa being one of the most affected regions[24].

Anthropogenic climate change has led to a global temperature increase of more than one °C since the pre-industrial period (1850-1900) ) [25]. However, this rise is not uniform globally or even within local areas[19]. Factors such as regional climate variations and land use changes have resulted in parts of Africa experiencing higher-than-average temperature increases and more frequent, intense, and prolonged heat waves[26].

Occupational settings, such as manual labour in factories, construction sites, or other outdoor activities, can also result in dangerous levels of heat exposure[27]. Growing evidence supports the effectiveness of Early Warning Systems in mitigating the impact of climate extremes, including heat extremes[28]. However, challenges remain in improving these systems' effectiveness, particularly in communication, action linkage, and response pre-identification. Currently, urban heat health Early Warning Systems in Africa are virtually non-existent, and there is limited policy or planning engagement with heat-health risks. These gaps pose a significant risk of extensive human health impacts in African cities in both the immediate and longer-term as temperatures continue to rise[29].

# Aims and objectives

This study aims to create effective, locally relevant urban heat health Early Warning Systems in African cities using data science and machine learning innovations. To build more climate-resilient cities in Africa and protect disproportionately affected populations from heat hazards, the project addresses climate change, urbanisation, and health.

The project has three specific objectives as part of the research plan from Figure 1:

1. Map intra-urban heat risks and exposure in large African cities, integrating health, socio-economic, geospatial climate, and satellite imagery data to understand heat-related health impacts.
2. Create a geographically and demographically stratified heat-health outcome forecast model to predict adverse health outcomes at varying temperature thresholds for various populations and neighbourhoods over a daily and weekly basis.
3. Establish an Early Warning System to deliver timely alerts to individuals, city planners, public health officials, and community leaders, thereby assisting in preparation for and response to heatwaves in African cities and mitigating heat-related health risks.

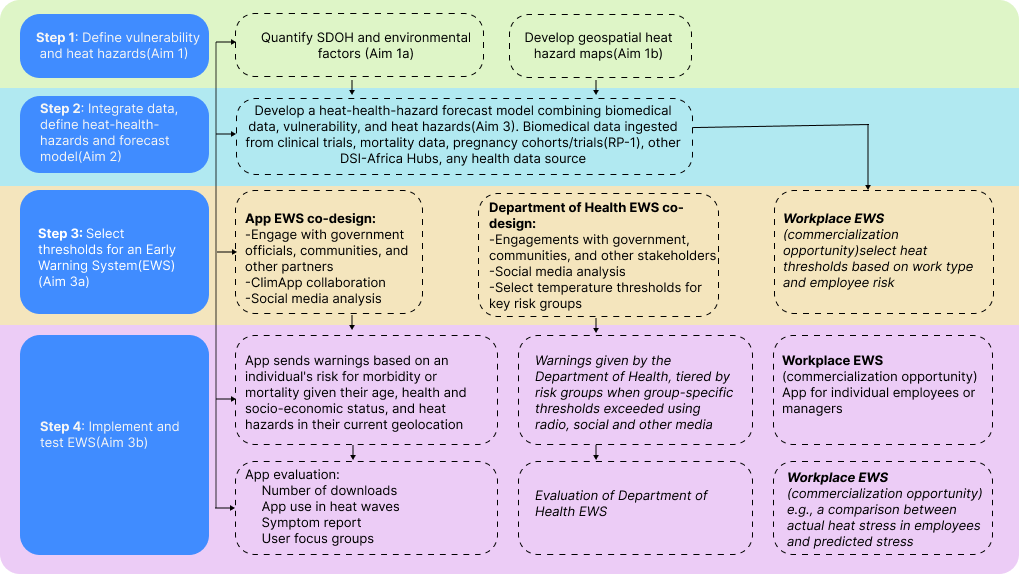


Figure 1: Research Plan for Research Project 2

# Methods

## Data sources/measurement

The study plans to combine datasets from a multitude of sources that encompass various fields—health, climate, environment, and Social Determinants of Health (SDOH). This multi-faceted approach will aid in building more thorough and locally pertinent models of heat-related health outcomes. These models will consider the diverse range of day-to-day realities and experiences encountered by inhabitants within each city, capturing how these realities impact their health in the context of heat [30]. In this study, 'lived experiences' refers to the unique conditions, challenges, and opportunities individuals encounter daily, shaped by their specific SDOH and environmental circumstances. Additionally, multiple datasets within a particular domain (e.g., multiple health trial datasets) both increase the statistical sample sizes for more robust modelling, as well as enable a rigorous quantification of key uncertainties (e.g., multiple climate datasets)[31, 32].

## Socio-economic and environmental data

This research will collect socio-economic geospatial data, which includes information on household economic conditions, service availability, and residential characteristics—referring to factors like housing type, construction materials used, and the quality and condition of living spaces[33]. The data will include national census records, specialised household, and demographic surveys. It will encompass details about individual and household income, education, occupation, living circumstances, and the accessibility to healthcare, education, and transportation services. For the Johannesburg-specific segment, many key variables will be provided by the Gauteng City-Region Observatory (GCRO) datasets[34].

Remote sensing data will be retrieved from satellite sensors, including optical images and indicators of physical aspects such as land surface temperature, soil moisture, vegetation condition, and land use and coverage [35]. In the case of Johannesburg, where such data are available, researchers will amalgamate data from current sensor networks with urban land use and building density details to create a model of urban land use heat[33, 34]. The objective of amalgamating climatic and socio-economic data is to generate maps of cities that will illustrate regions where individuals and households are most at risk for health impacts from heat[25]. These maps will be valuable for public health officials and policymakers to identify areas of need and to formulate focused interventions and policies to mitigate these risks[36].

Climate-associated data will be sourced from open data repositories, such as the Copernicus Climate Data Store (CDS) and Earth System Grid Federation (ESGF), offering observational-based datasets, historical re-analyses, and climate simulations[37]. Additionally, the IBM-PAIRS platform will be employed as an exhaustive and reliable source of climate data, inclusive of data from climate models, weather stations, and satellite observations[38]. This will furnish a comprehensive snapshot of Africa's past and future climate conditions, including the frequency, duration, and intensity of heat waves.

Geo-location data plays a pivotal role in our research, bridging the gap between individual health data and their respective socio-economic and environmental contexts within Johannesburg and Abidjan. This data allows us to spatially align health outcomes with their unique urban surroundings. By doing so, we can observe correlations between location-based socio-economic conditions, environmental factors, and health impacts.

## Health trials and cohort data

The health data for this study will be collected from clinical trials and cohort studies, such as HIV drug trials and COVID-19 vaccine trials. These studies typically involve many participants (hundreds to thousands) and are conducted over an extended period (multiple years) within a specific geographical area, providing detailed longitudinal individual health data for building statistical models relating time-varying predictors to health outcomes. Potential outcomes of interest include cardiovascular events, respiratory issues, kidney conditions, and mental health impacts, which may all be exacerbated by heat exposure in urban environments[39].

More specifically, the health cohort data integrated into the study will be identified based on the availability of three classes of variables within each study:

* Clinical variables: These encompass vital signs (e.g., body temperature, blood pressure, and heart rate), indicators of heat-related illness (e.g., headache, dizziness, fatigue, and nausea), and details on pre-existing medical conditions (e.g., hypertension, diabetes, and cardiovascular disease) that could increase the risk of heat-related illness, and documentation of adverse events potentially related to heat exposure.
* Laboratory variables: These comprise blood tests (e.g., electrolyte levels, liver and kidney function tests), markers of inflammation and oxidative stress, as well as HIV tests, including viral load and CD4 count, and COVID-19 test results.
* Demographic and SDOH variables: These involve basic demographic information (e.g., age, sex, race, and ethnicity), socio-economic factors (e.g., education, income, and occupation), and data on housing and urban infrastructure (e.g., air conditioning availability, ventilation, and shading) that could influence heat exposure and the degree to which individuals and households are at an increased risk.

## Integration of datasets

The successful execution of our study hinges upon the effective amalgamation of various data sources. To establish a comprehensive and nuanced understanding of heat's impact on health in African cities, we aim to create an integrated database that merges socio-economic, clinical, environmental, and geospatial data.

The process begins with cross-referencing the geolocation details of participants involved in the health trials and cohort studies with the socio-economic and environmental data. This will provide us with a spatial context to our health data, allowing us to correlate health outcomes with specific environmental and socio-economic conditions in Johannesburg and Abidjan. To preserve participant privacy and confidentiality during this process, we will employ spatial jittering techniques. Spatial jittering is a method of adding controlled 'noise' to the data to slightly alter the geolocations, thereby ensuring individuals cannot be identified based on their location, while still retaining the overall spatial distribution and trends in the data[40].

Additionally, our analysis will leverage remote sensing data and climate-associated data. By overlaying these datasets with our integrated health and socio-economic data, we can build a more granular understanding of how changes in environmental and climatic conditions influence health outcomes, particularly in relation to heat exposure.

## Trials and cohort identification

In pursuit of our research objective to explore the correlation between heat and health within the urban environments of Johannesburg and Abidjan, we have developed a comprehensive strategy to systematically identify relevant clinical trials and cohort studies. This strategy involves searching key databases using a combination of MeSH (Medical Subject Headings) terms and free-text terms, including those relating to study location, diseases of interest, the number of participants, study type, collected data, and the timeframe of study conduction. Our targeted search terms are designed to retrieve studies providing robust clinical, laboratory, and demographic data relevant to the impact of heat on health outcomes.

Post identification of potentially relevant studies through the search strategy, a two-step process of dual independent review will be employed. Initially, studies will be screened based on their titles and abstracts. Subsequently, studies deemed potentially eligible will be procured in their full-text format for a more thorough assessment against our pre-defined selection criteria (Table 1).

The quality of the selected studies will be evaluated by health researchers through a peer-reviewed tracking tool to ensure their scientific soundness and reliability. The collected data will be collated, synthesized, and discrepancies, if any, will be addressed and resolved through consensus discussions among team members.

A research project must meet the criteria outlined in Table 1 to be considered for inclusion in our study.

|  |  |
| --- | --- |
| Criteria | Description |
| Study type | Cohort or trial with at least 200 adult participants |
| Study location | Johannesburg or Abidjan, or both cities |
| Study design | Randomized or non-randomized clinical trial, or observational or interventional cohort with prospectively collected data |
| Data collected | At least two of the clinical or lab variables |
| Ethics approval | Local ethics approvals obtained |

**Table 1: Eligibility Criteria for Research Project 2**

## Managing bias

Managing potential biases is a critical aspect of ensuring the integrity and robustness of our study. To this end, we have outlined a strategy that is geared towards mitigating these biases.

Primarily, our approach will involve careful selection of health data sources, ensuring they meet established quality criteria and represent diverse demographic and geographic segments within our target cities of Johannesburg and Abidjan. This strategy will assist us in avoiding selection bias that could skew our findings[41].

In cases where potential biases are identified, we will undertake adjustments in the analysis phase. Specific statistical methods like propensity score matching, inverse probability weighting, and stratification will be applied. These methods help to control for confounding variables and reduce bias in observational studies, increasing the validity of our outcomes[42].

Ultimately, our goal is to minimize the impact of bias on our modelling process and the ensuing Early Warning System. By carefully managing potential biases, we aim to boost the reliability of our results and increase the practical applicability of our findings in the real-world context[43].

## Assessing the degree of increased risk within cities

The primary goal of our research is to assess and illustrate the socio-economic and environmental factors that increase risk, in conjunction with the exposure to heat hazards, in two African cities. Here, 'risk' is defined as the probability of adverse health outcomes arising from the interplay of these SDOH and environmental determinants and the exposure to elevated heat stress[44]. We will utilize the comprehensive datasets described earlier, including geospatial data from OpenStreetMaps[45], satellite imagery from Sentinel, land use maps, and socio-economic data[46]. These resources will provide a detailed picture of city characteristics such as building density, the presence of green spaces, and the socio-economic status of residents[47].

To distill this wealth of data into a single, actionable indicator of increased risk, we will employ dimensionality reduction methods like Principal Component Analysis (PCA) [48]. This technique will help us identify the most significant correlations among the various factors. To account for spatial variability in the data, we will use spatial analysis techniques. The resulting output will be a detailed map, highlighting the regions most at increased risk. This map will serve as a crucial tool for directing targeted interventions to mitigate heat-related health risks[49].

## Creating a geographically and demographically stratified heat-health outcome forecast model

The second objective of this study is to construct a geographically and demographically stratified heat-health outcome forecast model. This model will be designed to predict adverse health outcomes at varying temperature thresholds for different populations and neighborhoods.

The involves the creation of high-resolution urban temperature hazard maps (see Figure 2). We will utilize techniques such as remote sensing, statistical downscaling, and combined modeling to derive near-surface air temperatures from Landsat and MODIS data[50].

These temperatures will then be validated using weather station records and land-use maps. The resulting heat hazard maps will serve as a critical input for the subsequent stages of our machine learning pipeline.

In addition to temperature data, it is crucial to incorporate humidity measurements to accurately assess heat stress on human health. The Universal Thermal Climate Index (UTCI) is a prominent heat stress index that considers the combined effects of temperature, humidity, wind speed, and solar radiation on human comfort and health. By integrating UTCI or similar indices into our model, we can provide a more comprehensive evaluation of heat-related health risks. This will enable us to better understand the interactions between climatic factors and their impact on different population groups and neighborhoods.

Figure 2: Mapping of areas at increased risk and heat hazard in Johannesburg, South Africa

Once the temperature hazard maps are generated, they will be integrated with health datasets. This combined dataset will then undergo the feature engineering stage. Feature engineering is a crucial step in machine learning as it involves the selection and transformation of relevant predictors that better represent the underlying data patterns[51]. In our case, the features will be derived from the high-resolution temperature hazard fields and spatially disaggregated variables from the health datasets.

With the features engineered, we will apply various standard machine learning models, such as decision trees, linear and quantile regression trees, support vector machines, and logistic regression[52]. These models are chosen for their proven effectiveness in capturing relationships in complex datasets. We will use robust techniques like random forests and cross-validation to identify influential predictor variables and assess model performance.

Recognizing the potential for further performance improvement, we will also explore various deep learning architectures, including recurrent neural networks (RNNs), long short-term memories (LSTMs), and gated recurrent units (GRUs) [53-55]. Deep learning models can capture complex, non-linear relationships and dependencies within our data, potentially leading to more accurate predictions of the health effects of extreme heat[56].

Throughout this process, we will assess the significance of predictors for different populations within the cities under investigation. This will allow us to identify varying susceptibility levels to heat-induced health conditions based on demographics and risk factors. Potential health co-morbidities to be explored include cardiovascular disease, respiratory disease, renal disease, and HIV status[18].

Finally, to ensure the ongoing relevance of our model, we will regularly update it with new data during the the life-cylce of the project. This iterative process of model refinement and validation will enable us to continually improve the model's performance and maintain its applicability to the evolving urban heat-health landscape[57].

## Sample size considerations for health datsets

In order to ensure the robustness and accuracy of our models, we will implement a tailored approach to determine the sample size specifically for the health datasets. This four-step procedure takes into account factors such as the number of events, participants, outcome proportion, and expected model performance. This approach is applicable for continuous, binary, or time-to-event outcomes and is designed to minimise potential overfitting and target precise parameter estimates[58, 59]. By applying this procedure specifically to the health data, we ensure that our models are representative and reliable and that they accurately reflect the health conditions and risks of the populations under study.

A picture containing text, screenshot, font, diagram

Description automatically generated

Figure 3: Methodology to create a spatially and demographically stratified heat-health outcome forecast model

## Develop an Early Warning System reflective of geospatial and individualised risk profiles.

The third objective of our study is to devise an Early Warning System (EWS) that encapsulates the geospatial and individualised risk profiles of heat-related health impacts in Abidjan and Johannesburg as outlined in figure 3[60, 61]. This system aims to equip various working partners, including community health workers, clinic managers, urban planners, and at-risk individuals, with actionable insights.

The EWS development process involves integrating risk profile data from our geospatial heat hazard maps and health outcome forecast model[60, 61]. This integration will enable the system to reflect the specific risks associated with different demographic groups, health conditions, locations, and socio-economic statuses. Furthermore, we plan to develop a user-friendly digital tool, such as an app, to present the EWS data in an accessible and actionable format. This tool will facilitate users, including affected communities, in identifying the groups most at risk and planning appropriate responses.

Beyond the aforementioned factors, the Early Warning System (EWS) will incorporate heat hazard predictions, granting users a proactive comprehension of potential risks in the forthcoming days and weeks. This feature facilitates timely interventions across the health system. Furthermore, the EWS will provide guidance tailored to individuals deemed at risk, suggesting effective risk mitigation strategies such as adequate hydration, appropriate scheduling of activities, and signals for when medical help should be sought[62]. Through our methodological approach, we aim to construct an EWS that encapsulates both geospatial and individual risk profiles while also being actionable, user-friendly, and flexible enough to adapt to the ever-evolving landscape of heat-health dynamics.

## Patient and Public Involvement Statement

Public and patients are integral to our study, informing our Early Warning System design. Their input will guide risk mitigation strategies and the development of user-friendly, actionable digital tools.

## 5.12 Project timeline

# Ethical approval and protection of human subjects

This research study received ethical approval from both the Wits Human Research Ethics Committee in Johannesburg (reference number 200606) on June 30, 2022, and the National Ethics Committee for Life and Health Sciences, Côte d'Ivoire, on November 25, 2022 (reference number 176-22/MSHPCMU/CNESVS-kp) and will follow the, and the United States Department of Health and Human Services regulations for the protection of human subjects in research (45 CFR 46). There are two critical ethical and legal considerations for our research protocol: informed consent for secondary data usage and the protection of potentially identifiable information.

Regarding informed consent for secondary data usage, we will critically examine the consent procedures that were initially put in place. If a participant has previously provided "broad consent", permitting the use of their data in future research endeavours, we can proceed with sharing their data without additional ethical approvals. For participants who granted "narrow consent", which restricts data sharing beyond the original study purpose, we will give the situation careful deliberation. If obtaining renewed consent is unfeasible or involves a disproportionate effort, we will seek an informed consent waiver from the appropriate ethics committee.

As for the protection of potentially identifiable information, we are committed to minimising any privacy risks. The data we collect may contain indirect identifiers like geographical data. To counter this risk, we will employ several protective measures: we will not collect participant names, we will restrict the publication of identifiable data, and we will store data in a password-protected server with limited access. Additionally, we will follow data minimisation principles, retaining only the data essential for achieving our study objectives. When applicable, we will anonymise data through geographical aggregation and jittering, especially when home addresses are used.

Finally, we acknowledge the specific legislative requirements for using health data in different countries, including the laws surrounding the cross-border transfer of such data. We will therefore require data providers to provide a contractual guarantee, as part of the data sharing agreement, that all original studies followed appropriate informed consent procedures and that the sharing of this data complies with all relevant data protection laws.

## Study oversight

Prof. Chersich, Prof. Luchters, and the Hub Administrator direct the overall HE2AT Center. Steering Committee members represent six South, East, and West African institutes. This study is led by Prof. Cisse of Ivory Coast's Peleforo Gon Coulibaly University and co-led by Dr Christopher Jack of the University of Cape Town.

# Dissemination

To maximise HE2AT Center's effectiveness, prompt dissemination of research findings is crucial. We have devised a publication strategy detailing publication types, authors, and release dates. Our findings will be shared with research partners and relevant working partners to inform various levels of activities and update recommendations as needed. Timely dissemination is vital to HE2AT Center's success and mission.

# Study status

Ongoing.

# Author Contributions**:**

GC, MC, CJ, and SL were involved in the conception and design of the research. CP, MC and GM obtained ethics approval. CP, MC and SL prepared the figures. CP, CJ drafted the manuscript. All authors were involved in design and discussions to implement the protocol. All authors also edited and revised the manuscript. All authors approved the final version of the manuscript.

# Funding statement:

Research reported in this publication was supported by the Fogarty International Center and National Institute of Environmental Health Sciences (NIEHS) and OD/Office of Strategic Coordination (OSC) of the National Institutes of Health under Award Number U54 TW 012083. The content is solely the authors' responsibility and does not necessarily represent the official views of the National Institutes of Health.

# Competing interests

The authors declare potential competing interests: MF, GM, and CP have pension fund investments in the fossil fuel industry. The University of the Witwatersrand holds endowments and financial reserves invested in the same industry.

Data sharing statement:  In accordance with the NIH data sharing policy, data from the HEAT002 study will be made available to the scientific community. Researchers interested in accessing this data should submit a detailed request to Chris Jack at [cjack@csag.uct.ac.za](mailto:cjack@csag.uct.ac.za). The request will be reviewed and data will be shared subject to the approval of the request, ensuring the purpose aligns with ethical standards and the privacy of the participants is protected.

# References

1.  *Harnessing Data Science for Health Discovery and Innovation in Africa (DS-I Africa). Retrieved from* [*https://commonfund.nih.gov/AfricaData*](https://commonfund.nih.gov/AfricaData)*.*

2. 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.

3. Jung, J., et al., *Heat illness data strengthens vulnerability maps.* BMC Public Health, 2021. **21**(1): p. 1999.

4. Xu, R., et al., *Socioeconomic level and associations between heat exposure and all-cause and cause-specific hospitalization in 1,814 Brazilian cities: A nationwide case-crossover study.* PLoS Medicine, 2020. **17**(10): p. e1003369.

5. Lwasa, S., *Managing African Urbanization in the Context of Environmental Change.* Interactions, 2014. **2**.

6. Wang, Y.P. and K. Kintrea, *Urban Expansion and Land Use Changes in Asia and Africa.* Environment and Urbanization Asia, 2021. **12**: p. S13 - S17.

7. Rees, H.V., et al., *At the Heart of the Problem: Health in Johannesburg’s Inner-City.* BMC Public Health, 2017. **17**.

8. Abrahams, C. and D. Everatt, *City Profile: Johannesburg, South Africa.* Environment and Urbanization Asia, 2019. **10**: p. 255 - 270.

9. *Macrotrends. (2023). Johannesburg, South Africa Metro Area Population 1950-2023. Retrieved May 23, 2023, from* [*https://www.macrotrends.net/cities/22486/johannesburg/population*](https://www.macrotrends.net/cities/22486/johannesburg/population)*.*

10. Granado, S., et al., *Appropriating "malaria": local responses to malaria treatment and prevention in Abidjan, Cote d'Ivoire.* Med Anthropol, 2011. **30**(1): p. 102-21.

11. Djomand, G., et al., *Virologic and immunologic outcomes and programmatic challenges of an antiretroviral treatment pilot project in Abidjan, Côte d'Ivoire.* Aids, 2003. **17 Suppl 3**: p. S5-15.

12. *World Population Review. (2023). Abidjan Population 2023. Retrieved May 23, 2023, from* [*https://worldpopulationreview.com/world-cities/abidjan-population*](https://worldpopulationreview.com/world-cities/abidjan-population)*.*

13. *UN-Habitat. (n.d.). South Africa. Retrieved May 22, 2023, from* [*https://unhabitat.org/south-africa*](https://unhabitat.org/south-africa)*.*

14. Dongo, K., M. Kablan, and F. Kouamé, *Mapping urban residents’ vulnerability to heat in Abidjan, Côte d’Ivoire.* Climate and Development, 2018. **10**: p. 1-14.

15. Bakhtsiyarava, M., et al., *Modification of temperature-related human mortality by area-level socioeconomic and demographic characteristics in Latin American cities.* Soc Sci Med, 2023. **317**: p. 115526.

16. Voelkel, J., et al., *Assessing Vulnerability to Urban Heat: A Study of Disproportionate Heat Exposure and Access to Refuge by Socio-Demographic Status in Portland, Oregon.* Int J Environ Res Public Health, 2018. **15**(4).

17. Metzger, K.B., K. Ito, and T.D. Matte, *Summer heat and mortality in New York City: how hot is too hot?* Environ Health Perspect, 2010. **118**(1): p. 80-6.

18. Arsad, F.S., et al., *The Impact of Heatwaves on Mortality and Morbidity and the Associated Vulnerability Factors: A Systematic Review.* Int J Environ Res Public Health, 2022. **19**(23).

19. United Nations Department of Economic and Social Affairs Population Division, *World Urbanization Prospects: The 2018 Revision (ST/ESA/SER.A/420). New York: United Nations.* <https://population.un.org/wup/>, 2019.

20. Rathi, S.K., et al., *A Heat Vulnerability Index: Spatial Patterns of Exposure, Sensitivity and Adaptive Capacity for Urbanites of Four Cities of India.* Int J Environ Res Public Health, 2021. **19**(1).

21. *World Health Organization. (n.d.). Social Determinants of Health. Retrieved July 4, 2023, from* [*https://www.who.int/health-topics/social-determinants-of-health#tab=tab\_1*](https://www.who.int/health-topics/social-determinants-of-health#tab=tab_1)*.*

22. Petkova, E.P., A. Gasparrini, and P.L. Kinney, *Heat and mortality in New York City since the beginning of the 20th century.* Epidemiology, 2014. **25**(4): p. 554-60.

23. Rahman, M.M., et al., *The Effects of Coexposure to Extremes of Heat and Particulate Air Pollution on Mortality in California: Implications for Climate Change.* Am J Respir Crit Care Med, 2022. **206**(9): p. 1117-1127.

24. WHO, *Quantitative risk assessment of the effects of climate change on selected causes of death, 2030s and 2050s.* <https://www.who.int/globalchange/publications/quantitative-risk-assessment/en/>, 2014.

25. Adamo, N., N. Al-Ansari, and V.K. Sissakian, *Climate Change and the Need for Future Research.* IOP Conference Series: Earth and Environmental Science, 2022. **1120**.

26. Engelbrecht F, et al., *Projections of rapidly rising surface temperatures over Africa under low mitigation.* 10, 2015. **8**.

27. International Labour Organization, *Working on a warmer planet: The effect of heat stress on productivity and decent work.* 2019. [**https://www.ilo.org/global/publications/books/WCMS\_711919/lang--en/index.htm**](https://www.ilo.org/global/publications/books/WCMS_711919/lang--en/index.htm).

28. Bhardwaj, J., et al., *Climate Risk and Early Warning Systems: Adaptation Strategies for the Most Vulnerable Communities.* Climate Change Solutions, 2021.

29. Kapwata, T., et al., *Towards the development of a heat-health early warning system for South Africa.* Environmental Epidemiology, 2019.

30. Wolf, S.T., D.J. Vecellio, and W.L. Kenney. *Adverse heat-health outcomes and critical environmental limits (PSU HEAT Project)*. 2022.

31. Schubert, S. *An Update on Experimental Climate Prediction and Analysis Products Being Developed at NASA's Global Modeling and Assimilation Office*. 2011.

32. Riedel, M., S.E. Dosso, and L. Beran, *Uncertainty estimation for amplitude variation with offset (AVO) inversion.* Geophysics, 2003. **68**(5): p. 1485-1496.

33. Alonso, L. and F. Renard, *A Comparative Study of the Physiological and Socio-Economic Vulnerabilities to Heat Waves of the Population of the Metropolis of Lyon (France) in a Climate Change Context.* International Journal of Environmental Research and Public Health, 2020. **17**(3): p. 1004.

34. *Gauteng City-Region Observatory (2019). Quality of life in the Gauteng city-region: A report on key indicators. Retrieved from* [*https://www.gcro.ac.za/about/annual-reports/*](https://www.gcro.ac.za/about/annual-reports/)*.*

35. Hofierka, J., M. Gallay, and K. Onačillová, *Physically-based land surface temperature modeling in urban areas using a 3-D city model and multispectral satellite data.* urban climate, 2020. **31**: p. 100566.

36. Maharjan, M., et al., *Evaluation of Urban Heat Island (UHI) Using Satellite Images in Densely Populated Cities of South Asia.* Earth, 2021.

37. Kershaw, P., et al. *Delivering resilient access to global climate projections data for the Copernicus Climate Data Store using a distributed data infrastructure and hybrid cloud model*. 2019.

38. Albrecht, C.M., et al., *Pairs (Re)Loaded: System Design & Benchmarking For Scalable Geospatial Applications.* 2020 IEEE Latin American GRSS & ISPRS Remote Sensing Conference (LAGIRS), 2020: p. 488-493.

39. Arifwidodo, S.D., P. Ratanawichit, and O. Chandrasiri. *Understanding the Implications of Urban Heat Island Effects on Household Energy Consumption and Public Health in Southeast Asian Cities: Evidence from Thailand and Indonesia*. 2020.

40. Han, S., et al., *Location Privacy-Preserving Distance Computation for Spatial Crowdsourcing.* IEEE Internet of Things Journal, 2020. **7**: p. 7550-7563.

41. Narod, S.A., *Countercurrents: The Bias of Choice.* Current Oncology, 2019. **26**(6): p. 712-713.

42. Schwartz, R., et al., *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence.* 2022.

43. Balayn, A., C. Lofi, and G.-J. Houben, *Managing bias and unfairness in data for decision support: a survey of machine learning and data engineering approaches to identify and mitigate bias and unfairness within data management and analytics systems.* The VLDB Journal, 2021. **30**(5): p. 739-768.

44. Ebi, K.L. and J.J. Hess, *Health Risks Due To Climate Change: Inequity In Causes And Consequences: Study examines health risks due to climate change.* Health Affairs, 2020. **39**(12): p. 2056-2062.

45. Zhou, M., et al., *Efficient Localisation Using Images and OpenStreetMaps.* 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2021: p. 5507-5513.

46. *European Space Agency. (n.d.). Sentinel Online: Sentinel Data Access. Retrieved from* [*https://sentinel.esa.int/web/sentinel/sentinel-data-access*](https://sentinel.esa.int/web/sentinel/sentinel-data-access)*.*

47. Ludwig, C., et al., *Mapping Public Urban Green Spaces Based on OpenStreetMap and Sentinel-2 Imagery Using Belief Functions.* ISPRS Int. J. Geo Inf., 2021. **10**: p. 251.

48. Friesen, C.E., P. Seliske, and A. Papadopoulos, *Using Principal Component Analysis to Identify Priority Neighbourhoods for Health Services Delivery by Ranking Socioeconomic Status.* Online J Public Health Inform, 2016. **8**(2): p. e192.

49. Abson, D.J., A.J. Dougill, and L.C. Stringer, *Using Principal Component Analysis for information-rich socio-ecological vulnerability mapping in Southern Africa.* Applied Geography, 2012. **35**(1): p. 515-524.

50. Janatian, N., et al., *A statistical framework for estimating air temperature using MODIS land surface temperature data.* International Journal of Climatology, 2017. **37**(3): p. 1181-1194.

51. Kelleher, J.D. and B. Tierney, *Data science*. 2018: MIT Press.

52. Xu, J., et al., *Downscaling Aster Land Surface Temperature over Urban Areas with Machine Learning-Based Area-To-Point Regression Kriging.* Remote. Sens., 2020. **12**: p. 1082.

53. Mirzaei, S., J.-L. Kang, and K.-Y. Chu, *A comparative study on long short-term memory and gated recurrent unit neural networks in fault diagnosis for chemical processes using visualization.* Journal of the Taiwan Institute of Chemical Engineers, 2021. **130**.

54. Ke, D., et al., *Effects of heatwave features on machine-learning-based heat-related ambulance calls prediction models in Japan.* Sci Total Environ, 2023. **873**: p. 162283.

55. Lee, M.-H., H.-J. Mun, and S.-K. Kang, *Prediction of Health Problems Using Deep Learning Images and Bio-Signals.* Applied Sciences, 2022. **12**(23): p. 12457.

56. Haque, E., S. Tabassum, and E. Hossain, *A Comparative Analysis of Deep Neural Networks for Hourly Temperature Forecasting.* IEEE Access, 2021. **9**: p. 160646-160660.

57. Lin, C.-Y., et al., *Impact of an improved WRF urban canopy model on diurnal air temperature simulation over northern Taiwan.* Atmospheric Chemistry and Physics, 2015. **16**: p. 1809-1822.

58. Amatya, A. and D.K. Bhaumik, *Sample size determination for multilevel hierarchical designs using generalized linear mixed models.* Biometrics, 2018. **74**(2): p. 673-684.

59. Riley, R.D., et al., *Calculating the sample size required for developing a clinical prediction model.* BMJ, 2020. **368**: p. m441.

60. Kalkstein, L.S., S.C. Sheridan, and A.J. Kalkstein. *Heat/Health Warning Systems: Development, Implementation, and Intervention Activities*. 2009.

61. Toloo, G., et al., *Evaluating the effectiveness of heat warning systems: systematic review of epidemiological evidence.* International journal of public health, 2013. **58**: p. 667-681.

62. *Bridging Earth and Health Science Communities in Environmental Health: Focus on Vector-borne Diseases, Extreme Heat, and Air Quality*. 2022.