Map

Description automatically generated with medium confidence

**Overall Project title:**

**Developing data science solutions to mitigate the health impacts of climate change in Africa: the HE²AT Center**

***The Heat and Health African Transdisciplinary Center   
(HE²AT Center)***

Study Protocol

**Innovative machine learning and multi-source data analysis towards development of an urban heat-health Early Warning System for African cities**

**Protocol number: HEAT002**

**Study Sponsored by NIH Common Fund Program**

**Harnessing Data Science for Health Discovery and Innovation in Africa**

06 December 2023

Protocol version 3.0

Graphical user interface, application

Description automatically generated

Table of Contents

[Tables and Figures 3](#_Toc105168503)

[Acronyms 4](#_Toc105168504)

[Study partners 6](#_Toc105168505)

[Study investigators 6](#_Toc105168506)

[Principal Investigators 6](#_Toc105168507)

[Co-Principal Investigators 6](#_Toc105168508)

[Co-Investigators 6](#_Toc105168509)

[Synopsis 8](#_Toc105168510)

[1) Background and rationale 9](#_Toc105168511)

[2) Aims and Objectives 13](#_Toc105168512)

[Funding acknowledgement: 13](#_Toc105168513)

[3) Summary of overall study design 14](#_Toc105168514)

[4) Methodology 15](#_Toc105168515)

[4.1. Study sites 15](#_Toc105168516)

[4.2. Major steps and integration 16](#_Toc105168517)

[4.3. Data description and acquisition 16](#_Toc105168518)

[4.3.1. For Aim 1a: Quantify intra-urban socio-economic and environmental vulnerability 17](#_Toc105168519)

[4.3.2. For Aim 1b: Develop high resolution urban temperature hazard maps 17](#_Toc105168520)

[4.3.3. For Aim 2: Develop a spatially and demographically stratified heat-health outcome forecast model 19](#_Toc105168521)

[4.4. Data pre- and post-processing, formatting 20](#_Toc105168522)

[4.4.1 For Aim 1a: Quantify intra-urban socio-economic and environmental vulnerability 20](#_Toc105168523)

[4.4.2 For Aim 1b: Develop high resolution urban temperature hazard maps 21](#_Toc105168524)

[4.4.3 For Aim 2: Develop a spatially and demographically stratified heat-health outcome forecast model 21](#_Toc105168525)

[4.5. Data analyses and outputs 22](#_Toc105168526)

[4.5.1. Imputation of Missing Data **Error! Bookmark not defined.**](#_Toc105168527)

[4.5.2. For Aim 1a: Quantify intra-urban socio-economic and environmental vulnerability 22](#_Toc105168528)

[4.5.3. For Aim 1b: Develop high resolution urban temperature hazard maps 23](#_Toc105168529)

[4.5.4. For Aim 2: Develop a spatially and demographically stratified heat-health outcome forecast model 23](#_Toc105168530)

[5) Special data management considerations 24](#_Toc105168531)

[5.1. Data types 24](#_Toc105168532)

[5.1.1. Health related data 24](#_Toc105168533)

[5.1.2. Climate/weather data 25](#_Toc105168534)

[5.1.3. Remote sensing data excluding climate data 25](#_Toc105168535)

[5.1.4. Areal/Geospatial socio-economic data 25](#_Toc105168536)

[5.2. Data indexing and processing (harmonization, homogenization, normalization) 26](#_Toc105168537)

[5.3. Data management, documentation and curation 26](#_Toc105168538)

[5.4. Data Security and Confidentiality of Potentially Sensitive Information 27](#_Toc105168539)

[5.4.1. De-identification 27](#_Toc105168540)

[5.4.2. Network firewall and Virtual Private Network 28](#_Toc105168541)

[5.4.3. Local authentication and authorization 28](#_Toc105168542)

[5.5. Procedure for making data available to qualified individuals 28](#_Toc105168543)

[5.6. Principles of data sharing and open access 30](#_Toc105168544)

[5.6.1. Restrictions to data sharing 30](#_Toc105168545)

[5.6.2. Discoverability 31](#_Toc105168546)

[6) Ethical and Privacy considerations 31](#_Toc105168547)

[7) Training, engagement, and capacity building 40](#_Toc105168548)

[8) Dissemination and publication plan 41](#_Toc105168549)

[9) Collaborations 43](#_Toc105168550)

[10) Main activities, milestones and timelines 44](#_Toc105168551)

[11) References 44](#_Toc105168552)

[Annexes 49](#_Toc105168553)

[Annex 1: Databases available to the HE²AT Center 49](#_Toc105168554)

[Annex 2: Characteristics of the studies 51](#_Toc105168555)

[Annex 3: Individual-level variables 72](#_Toc105168556)

[Annex 4: Land Surface Temperature retrieval workflow 73](#_Toc105168557)

# Tables and Figures

[Table 1: Summary of ethical, legal and data management considerations that underpin study activities 32](#_Toc105426291)

[Table 2: Ethical and legal considerations for different types of data 37](#_Toc105426292)

[Table 3: GANTT chart: milestones and timelines 44](#_Toc105426293)

[Table 4 Overview of Databases to be used in RP2 49](#_Toc105426294)

[Table 5 Variable List for RP2 72](#_Toc105426295)

[Figure 1 Flow of activities from Aim 1 to Aim 3, starting from an analysis of past observations, to the implementation of forecasting and downscaling methods to generate an Urban Heat-Health Early Warning System. 14](#_Toc105426320)

[Figure 2 Steps for achieving the aims and objectives of the study 16](#_Toc105426321)

[Figure 3 Overview of the data and workflow for downscaling and how it will be used in the Urban Heat-Health Early Warning System 22](#_Toc105426322)

[Figure 5 Levels of stakeholder engagement 41](#_Toc105426323)

[Figure 6 Partners collaborating on the HE²AT study from across Africa and the United States. 43](#_Toc105426324)

# Acronyms

|  |  |
| --- | --- |
| **Acronym** | **Definition** |
| ACT | Antenatal Corticosteroids Trial |
| AES | Advanced Encryption Standard |
| *API* | *Application Programming Interface* |
| *ART* | *Anti Retroviral Treatment* |
| BP | Blood Pressure |
| CDC | Centers for Disease Control |
| Co-PI | Co Principal Investigators |
| CSAG | Climate Systems Analysis Group |
| DASH | Data and Specimen Hub |
| DMAC | Data Management and Analysis Core |
| DMP | Data Management Plan |
| DoH | Department of Health |
| DRS | Data Reference Syntax |
| DSA | Data Sharing Agreements |
| ECMWF | European Center for Medium range Weather Forecasts |
| EDCTP | European and Developing Countries Clinical Trials Partnership |
| EWS | Early Warning System |
| EWSs | Early Warning Systems |
| FIC | Fogarty International Center |
| FTP | File Transfer Protocol |
| GANTT | Generalized Activity Normalization Time Table |
| GCM | Global Climate Model |
| GCRO | Gauteng City Region Observatory |
| GSOD | Global Summary Of the Day |
| HDSS | Health and Demographic Surveillance System |
| HIV | Human Immune deficiency Virus |
| HPC | High Performance Computing |
| LST | Land Surface Temperature |
| MCV | Mean Cell Volume |
| MODIS | Moderate Resolution Imaging Spectroradiometer |
| MRC | Medical Research Council |
| NDVI | Normalized Difference Vegetation Index |
| NIST | National Institute of Standards and Technology |
| NIH | National Institute of Health |
| OLI | Operational Land Imager |
| PI | Principal Investigator |
| POPIA | Protection of Personal Information Act |
| RHI | Reproductive Health and HIV Institute |
| RoB | risk of bias |
| RP1 | Research Project 1 |
| RP2 | Research Project 2 |
| RPs | Research Projects |
| SC | Steering Committee |
| SCL | Scene Classification |
| TEC | Training and Engagement Core |
| TIRS | Thermal Infrared Sensor |
| TOA | Top Of Atmosphere |
| TWC | The Weather Company |
| UCT | University of Cape Town |
| UPGC | University Peleforo Gon Coulibaly |
| USAID | United States Agency for International Development |
| UW | University of Washington |
| WHC | Wits Health Consortium |
| WHO | World Health Organization |
| WMO | World Meterorological Organisation |
| WWARN | Worldwide Antimalarial Resistance Network |

**SECTION 1: OVERVIEW AND DESCRIPTION OF INVESTIGATORS AND INSTITUTIONS**

# Study partners

|  |  |  |
| --- | --- | --- |
| Name of institution | Country | Address |
| Centre for Sexual Health and HIV & AIDS Research | Zimbabwe | 4 Bath Road, Belgravia, Harare, Zimbabwe |
| Climate System Analysis Group, University of Cape Town | South Africa | Dept. Of Environmental and Geographical Science, University of Cape Town, Rondebosch, Cape Town, South Africa, 7700 |
| IBM Research Africa | South Africa | Wits Research Lab, 47 Juta Street, Braamfontein, Johannesburg, South Africa, 2001 |
| University of Washington | United States of America | Center for Health and Global Environment (CHanGE), Department of Environmental and Occupational Health Sciences, School of Public Health, University of Washington, Box 354695, Seattle, Washington, United States of America |
| University Peleforo Gon Coulibaly | Cote d'Ivoire | UPGC BP 1328 Korhogo, Côte d'Ivoire / CSRS,01 BP 1303 Abidjan 01, Adiopodoumé- Km 17-Route de Dabou, Côte d’Ivoire |
| Wits RHI, University of the Witwatersrand | South Africa | Hillbrow Health Precinct, 22 Esselen Street, Hillbrow, Johannesburg, South Africa, 2001 |

# **Study investigators**

Principal Investigators

|  |  |  |
| --- | --- | --- |
| **Name** | **Institution** | **Contact Details** |
| Guéladio Cissé | University Peleforo Gon Coulibaly | [gueladio.cisse@swisstph.ch](mailto:gueladio.cisse@swisstph.ch), +41 79 938 11 98 |
| Matthew Chersich | University of the Witwatersrand | [mchersich@wrhi.ac.za](mailto:mchersich@wrhi.ac.za) ,+27 72 752 1123 |

Co-Principal Investigators:

|  |  |  |
| --- | --- | --- |
| **Name** | **Institution** | **Address, email, phone** |
| Christopher Jack | University of Cape Town | [cjack@csag.uct.ac.za](mailto:cjack@csag.uct.ac.za) |
| [Sibusisiwe Makhanya](mailto:cjack@csag.uct.ac.za) | IBM Research Africa | [Sibusisiwe.makhanya@ibm.com](mailto:Sibusisiwe.makhanya@ibm.com) |
| Gloria Maimela | University of the Witwatersrand | [gmaimela@wrhi.ac.za](mailto:gmaimela@wrhi.ac.za) |
| Craig Parker | University of the Witwatersrand | [cparker@wrhi.ac.za](mailto:cparker@wrhi.ac.za) |

Co-Investigators:

|  |  |  |
| --- | --- | --- |
| **Name** | **Institution** | **Address, email, phone** |
| Adja Ferdinand Vanga | University Peleforo Gon Coulibaly | af\_vanga@yahoo.fr |
| Bonnie Joubert | National Institute of Health | [bonnie.joubert@nih.gov](mailto:bonnie.joubert@nih.gov) |
| Brama Koné | University Peleforo Gon Coulibaly | bramakone@gmail.com |
| Christopher Boyer | University of Washington | [cboyer10@uw.edu](mailto:cboyer10@uw.edu) |
| Etienne Vos | IBM Research Africa | [Etienne.Vos@ibm.com](mailto:Etienne.Vos@ibm.com) |
| Iba Dieudonné Dely | University Peleforo Gon Coulibaly | ibadely12@gmail.com |
| James Mashiyane | IBM Research Africa | [toby.kurien@za.ibm.com](mailto:toby.kurien@za.ibm.com) |
| Kimberley McAllister | National Institute of Health | [mcallis2@niehs.nih.gov](mailto:mcallis2@niehs.nih.gov) |
| Kouakou Etienne | University Peleforo Gon Coulibaly | [kyaoetienne@yahoo.fr](mailto:kyaoetienne@yahoo.fr) |
| Kristie Ebi | University of Washington | [krisebi@uw.edu](mailto:krisebi@uw.edu) |
| Lisa van Aardenne | University of Cape Town | [lisa@csag.uct.ac.za](mailto:lisa@csag.uct.ac.za) |
| Madina Doumbia | University Peleforo Gon Coulibaly | madinadoub@gmail.com |
| Maliha Ilias | National Institute of Health | [maliha.ilias@nih.gov](mailto:maliha.ilias@nih.gov) |
| Piotr Wolski | University of Cape Town | [wolski@csag.uct.ac.za](mailto:wolski@csag.uct.ac.za) |
| Stanley Luchters | Centre for Sexual Health and HIV & AIDS Research | stanley.luchters@cesshar.co.zw , stanley.luchters@lstmed.ac.uk |
| Relebohile Motana | University of Witwatersrand | [rmotana@wrhi.ac.za](mailto:rmotana@wrhi.ac.za) |
| Toby Kurien | IBM Research Africa | james.mashiyane1@ibm.com |
| Webster Mavhu | CeSSHAR | [webster@ceshhar.co.zw](mailto:webster@ceshhar.co.zw) |
| Zororo Mavindidze | University of Witwatersrand | [zmavindidze@wrhi.ac.za](mailto:zmavindidze@wrhi.ac.za) |

# **Synopsis**

**Rationale:** The study constitutes one of two Research Projects (RPs) within the NIH-funded HE²AT Center. It specifically addresses the complexity of urban spaces with regard to heat-health impacts and the appropriate responses for some particular vulnerable groups.

**Objectives**: The overarching goal is to advance understanding of complex spatially and demographically stratified heat-health interactions in large African cities and to apply this information to develop locally relevant and risk-stratified Early Warning Systems (EWS). The aims are three: (1) Map intra-urban heat vulnerability and exposure across urban areas in large African cities (Aim 1); (2) Develop a spatially and demographically stratified heat-health outcome forecast model in order to predict the probability of adverse health outcomes at different temperature thresholds (Aim 2); and (3) Develop an Early Warning System reflective of geospatial and individualized risk patterns (Aim 3)

**Study design:** The RP2 focuses on the conditions in two large cities in two regions of Africa (Johannesburg, South Africa, Southern Africa, and Abidjan, Côte d'Ivoire, West Africa). It adopts a transdisciplinary approach in which multidisciplinary experts will collaborate with communities, local government actors and policy makers to address heat-related complex and interconnected research gaps. Existing data from longitudinal studies (trials and cohorts) in the two cities that were performed among HIV-infected adults, HIV-uninfected adults and adults in COVID-19 prevention or treatment studies will be reanalysed, together with weather, other environmental and socio-economic and other data. Analysis will deploy a range of machine learning methods to construct an index of intra-urban socio-economic and environmental vulnerability factors. As the solutions developed by the RP2 will address two major concerns for global policymakers (how to warn people about a heatwave in urban settings in low- and middle-income countries, and then to track its impacts), local, national and international policymakers will be engaged at all stages.

**Dissemination:** Findings will be disseminated at various levels, using several channels, including workshops, policy and research fora, scientific conferences and journal publications, and towards different target groups (including policy makers, communities, specific vulnerable groups).

**Funding acknowledgement**: 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 responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health

**Keywords**: urban, heat, health, early warning systems, intra-urban vulnerability, socio-economics and environment, exposure mapping.

**SECTION 2: RESEARCH PROTOCOL**

# Background and rationale

*1.1 High ambient temperatures and heat waves*

High ambient temperatures above long-term averages during summer months and discrete heat extremes (e.g., heat waves) are associated with excess mortality and considerable morbidity [1-4] . WHO predicts that by 2030 there will be almost 92,000 deaths per year from heat waves, with sub-Saharan Africa amongst the worst affected regions [5]. Anthropogenic climate change has already resulted in a more than 1°C rise in temperature globally since the Industrial Revolution [6]. This increase is not evenly distributed across the planet, however, or even within local areas [7].

Regional differences and the effect of urban development and land use change means that many parts of Africa are experiencing higher than average temperature increases, and more frequent, intensive, and longer-lasting heat waves [8]. Work done by a team at the University of Witwatersrand estimated that temperatures in Africa will rise at about 1.5 times the global rate of increase, escalating by 4-6°C above current temperatures in the sub-tropics by 2100 [8]. Heat waves have been defined in a myriad of ways, based on static or relative thresholds (e.g. maximum or mean temperature, and taking into account humidity and even solar radiance), number of days of excessive heat, and timing during the summer season.

*1.2 Specific urban heat vulnerability*

The ‘Urban Heat Island’ is a phenomenon where concrete, non-reflective surfaces, and low greenery and wind results in temperatures considerably higher than in surrounding areas, and concomitant high levels of morbidity and mortality during heat waves [7]. This is concerning as Africa is the most rapidly urbanizing continent in the world and, by 2050, 59% of its population is expected to live in cities (up from 40% in 2018) [9].

In many African cities, nearly half of the population lives in informal settlements or “slums” that are often located in areas unsuitable or undesirable for other uses [10]. People living in these areas have high exposure to heat, as, for instance, informal settlements are often in low-lying hotter parts of cities, are typically densely settled and crowded, lack vegetation cover, have limited or no shade, and poor natural ventilation [11]. Moreover, typical housing materials in informal housing – iron metal sheeting – accentuate heat exposure, with temperatures inside these dwellings commonly 3-4°C warmer than outdoors [12, 13].

In fact, much of the urban environment in African cities offers little heat insulation, with low-cost government-built housing, school classrooms and prefabricated clinics, for example, also often exceeding outdoor temperatures by 3-4°C, or more [13-16] . People inside these structures commonly report heat-exposure symptoms such as fatigue, difficulty breathing, and headaches [14, 16-18]. Heat exposure in occupational settings such as manual labour in factories, construction sites or other outdoor activities can also reach dangerous levels [19, 20].

*1.3 Vulnerable groups and spaces in urban contexts*

Urban poor are most vulnerable to heat exposure for a range of factors related to heightened sensitivity (such as high rates of HIV, malnutrition, and non-communicable diseases) and to lowered adaptive capacity (such as lack of access to cool water, air-conditioned spaces, health services, and occupational protections). Taking all these factors together means that, for example, an elderly woman living in a tin informal dwelling who has respiratory compromise from previous TB will have very few or no options for protecting herself against heat stress. We need to understand the impact that a heat wave has on her health outcomes, and what more can be done to protect her and other vulnerable groups in our warming world. This is critical information as most heat-related morbidity and mortality is preventable with improved preparedness and avoidance of exposure.

Impacts of heat exposure are clearly contingent not just on temperature, but also on geographical, socioenvironmental, and demographic factors [20]. Therefore, robust quantifications of the magnitude and pattern of heat vulnerability, exposure, and impacts, need to include an assessment of the extent to which meteorological parameters (temperature, rainfall, wind direction, and humidity, for example), urban forms, population density, and other factors contribute to risk. Importantly, the unique characteristics of the environmental exposures, built environment, demographics, and spectrum of communicable and non-communicable diseases in Africa mean that findings elsewhere may have little applicability to African cities.

To date, however, none of the assessments of heat-health impacts in African cities has taken adequate account of the complexity of urban spaces, and most have relied on relatively small datasets with poor spatial and temporal coverage. Considering the spatial heterogeneity of living conditions in cities in Africa, mapping intra-urban heat vulnerability and exposure across urban areas is of utmost importance.

*1.4 Health outcomes: heat related morbidity and mortality*

Worldwide, the majority of studies of heat hazards in cities quantify environmental vulnerability and develop temperature hazard maps without linking these to actual health outcomes (see for example: [21-23]). Equally, studies that focus primarily on health outcome data such as excess mortality during heat waves seldom integrate data on vulnerabilities from the built environment or urban space [24, 25]. Many studies use summary health outcome data for a district, or hospital record data [25, 26], and assume that the temperatures that patients experienced in the district and their living conditions are uniform. Coarse location information constrains the ability to map health outcomes, and limits the extent to which outcomes may be linked with environmental stressors and population activities such as work and residential location.

Studies of heat impacts in cities in high-income countries have found that risk for heat-related morbidity and mortality are highest among the elderly, those with chronic cardiovascular or respiratory conditions, and people who are homeless or in lower socio-economic groups [19]. Some studies in Africa have documented higher heat health risks among the elderly [23, 24, 26-29], pregnant women [27-29], and children under five years [24, 30-34]. Most studies, regardless of setting, have not considered vulnerability as determined by geospatial factors representing the built environment form.

It is also important to note that the majority of heat impact studies in Africa and worldwide relied on analyses of temperature associations with mortality because these data are often most easily accessible. However, from a public health perspective, it is more important to understand heat-related *morbidities* as these account for a considerably greater burden of disease than heat-related mortality. Additionally, from an evidence viewpoint, there are major gaps in evidence on heat-related morbidities than mortality.

*1.5 People with particular health problems*

Little evidence is available on whether high temperatures pose particular risks for some of the key vulnerable population groups in sub-Saharan Africa, such as those living with HIV. This is a critical question: there are an estimated 26 million people living with HIV on the subcontinent, up to 30% of adults are infected in some countries on the sub-continent, and some 440,000 deaths are due to AIDS annually [35, 36]. Of note, some recent studies have found that rates of HIV are highest in areas most vulnerable to heat impacts, such as informal settlements [35] This is a critical question: there are an estimated 26 million people living with HIV on the subcontinent, up to 30% of adults are infected in some countries, and some 440,000 deaths are due to AIDS annually [35, 36]. Of note, some recent studies have found that rates of HIV are highest in areas most vulnerable to heat impacts, such as informal settlements [35] . In a PubMed (MEDLINE) search on 7/11/2020, we located three studies that had documented the impacts of heat exposure (measured through actual temperatures) on disease patterns in HIV-infected people [37], and 13 that had identified increased levels of adverse conditions during the summer or the ‘warm’ seasons [38-50]. Increased rates of adverse health conditions during the warm seasons strongly suggests that more nuanced analyses of temperature exposures will demonstrate sizable temperature-outcome relationships and identify actionable information.

Reducing risks to health from current and projected high temperatures depends not only on physiological acclimatization, but also on planned adaptation by the health and other sectors. Early Warning Systems, one of the key adaptative interventions applied across the globe, is discussed hereafter.

*1.6 Early warning systems (EWS)*

Early Warning Systems (EWSs) play a central role in the health sector response to heat waves*[51-57]*. These EWS form the central element in the health system’s response to heat waves [58, 59].

In these systems, once a predefined temperature threshold is exceeded, warnings trigger a set of actions intended to lower exposure to harmful temperatures and to assist people to take protective actions. Common interventions in high-income countries include opening of public cooling shelters, targeted messaging, wellness checks in the community at people’s homes or in institutions such as homes for the elderly, distribution of cold water, and surveillance systems [58, 60]. EWSs have been shown to reduce morbidity and mortality during heat waves [52, 61], including in low- and middle-income countries [62].

EWSs in Ahmedabad, India for example, led to a dramatic decline in heat wave deaths in 2015 compared to a similar heat wave in 2010[62]. Over longer time scales, changes to the built environment such as green roofs, shade structures, reflective surfacing that can reflect infrared and UV wavelength and reduce heat transfer, as well as thermal emittance. Reflective surfacing includes painting surfaces white can reduce heat exposures in urban areas [63, 64]Reflective surfacing includes painting surfaces white can reduce heat exposures in urban areas [63, 64]. However, these structural interventions are unlikely to be feasible at scale in most African cities for some time. Well-functioning, adequately sensitive EWSs accompanied by preventive measures such as dedicated cool spaces for high-risk groups, community health worker outreach, water supplies, and closing of schools and workplaces will be the mainstay of heat wave responses in the short- to medium-term in African cities. Over longer time scales, changes to the built environment such as green roofs, shade structures, reflective surfacing, and painting surfaces white can reduce heat exposures in urban areas [63, 64]. However, these structural interventions are unlikely to be feasible at scale in most African cities for some time. Well-functioning, adequately sensitive EWSs accompanied by preventive measures such as dedicated cool spaces for high-risk groups, community health worker outreach, water supplies, and closing of schools and workplaces will be the mainstay of heat wave responses in the short- to medium-term in African cities [61].

Most EWSs have cut-off thresholds directed to the entire population and do not take the level of vulnerability of different groups into consideration [65]. There is growing evidence that most heat-related illnesses and deaths occur on dates without heat warnings [66], suggesting that current thresholds could be improved to take particularly vulnerable populations into account.

*1.8 Research gaps and needs of a multi-disciplinary approach*

Within Africa there is a pressing need to identify:

* the aspects of socio-economics and demographics, such as housing types and density, commuting distances, and working conditions, that contribute to heat vulnerability and exposure
* the types of heat exposure that are most dangerous for people living in different conditions in African cities (e.g., maximum in night-time or daily temperatures, heat waves or heat spells of long duration combined air quality & heat/cold snaps)
* geospatially detailed heat-hazard conditions across cities related to altitude, vegetation, water bodies, and risk groups under different large-scale weather conditions
* people’s and in particular most vulnerable groups’ perception and understanding of heat as a health hazard and how these risks are framed

This research study will contribute to providing answers to these questions that would constitute a major leap towards developing interventions to reduce vulnerability and build resilience against extreme heat in African cities. It will adopt a transdisciplinary approach, drawing together knowledge in the fields of biomedical and climate science, as well as geospatial analysis and computer science.

# Aims and Objectives

The overarching objective of this Research Study is to advance understanding of complex spatially and demographically stratified heat-health interactions in large African cities and to apply this information to develop locally relevant and risk-stratified Early Warning Systems(EWSs).

More specifically, this overarching objective will be pursued through three aims:

1. Map intra-urban heat vulnerability and exposure across urban areas in large African cities (Aim 1)
   1. Quantify intra-urban socio-economic and environmental vulnerability (Aim 1a)
   2. Develop high resolution urban temperature hazard maps (Aim 1b)
2. Develop a spatially and demographically stratified heat-health outcome forecast model to predict the probability of adverse health outcomes at different temperature thresholds (Aim 2)
3. Develop an Early Warning System reflective of geospatial and individualized risk patterns (Aim 3)
   1. Develop an EWS based on a digital App (Aim 3a)
   2. Develop a “tiered” EWS in collaboration with the Ministries of Health, National Meteorological and Hydrological Services, and other relevant stakeholders (Aim 3b)

The methods for Aim 3 are not presented in this protocol and will be elaborated in a subsequent protocol.

# Funding acknowledgement:

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 responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health

# Summary of overall study design

This study adopts a transdisciplinary approach in which experts in the fields of computer science, healthcare, climate science, statistics, physics, physiology, environmental epidemiology, and public health will collaborate with communities, local governance actors and policy makers to address heat-related complex and interconnected research challenges, in African urban contexts.

The study will pursue the three aims in an integrated and a consistent way with a logical flow of activities from understanding the local urban contexts to the development of forecasting models intended for the development of an effective and appropriate urban early warning system for heat.

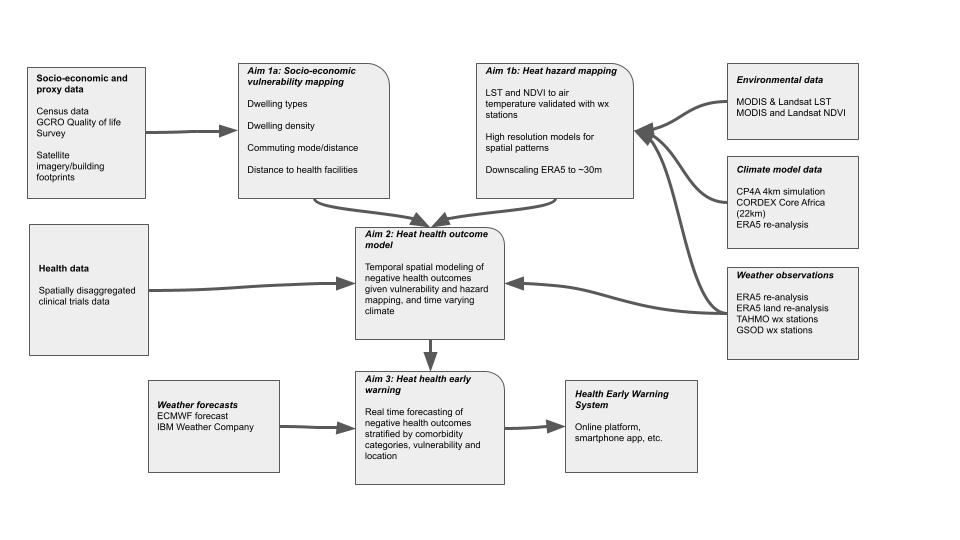


Figure 1 Flow of activities from Aim 1 to Aim 3, starting from an analysis of past observations, to the implementation of forecasting and downscaling methods to generate an Urban Heat-Health Early Warning System.

The study will make use of existing, prospectively collected, high-quality health data from longitudinal datasets (see in Annex 1 Table 4 the description of health outcomes, environmental and other datasets). Additionally, the study will make use of a multitude of existing weather data, both retrospective or historical, as well as weather and seasonal prospective or forecast data (see table4 in Annex 1).

The study will depend on a DMAC (Data Management and Analysis Core) hosted at the University of Cape Town. DMAC will implement a remote access analysis platform based on Python Jupyter Lab which is a collaborative data science coding platform enabling access by study partners to the High-Performance Computing (HPC) and large storage provided by UCT. Climate System Analysis Group (CSAG) at UCT hosts a HPC cluster and large storage system that already holds many of the relevant climate datasets with a number of new geospatial datasets being made available in support of the HE²AT study.

The study will also draw on the resources of IBM-PAIRS (Physical Analytics Integrated Repository and Services), which is an extensive data store of aligned, pre-processed geo-spatial data such as satellite and aerial images, Light Detection and Ranging (LiDAR) data, land use surveys, numerical weather prediction model outputs, weather reanalysis and measurement, and other environmental data, Internet of Things(IoT) sensor data, anonymized cell phone tracking based human mobility and social-economic data such as population, age-distribution, capacities of health infrastructure, and road networks.

To date, PAIRS has curated more than 6 petabytes(PB) of data at a daily ingestion rate of about 20 terabytes(TB). The technology is based on a large-scale Hadoop/HBase system for distributed data storage and processing. IBM-PAIRS leverages efficient data indexing methods, which result in spatially and temporally linked data layers, both for data from 2Dimensional grids (e.g. satellite images) and from point locations (e.g. measurements from distributed sensor networks).

Access to PAIRS will either be directly through the various PAIRS interfaces which will be made available by providing authentication to identified individuals, or through integration into the DMAC data analysis platform (Jupyter Hub) which will also involve authentication for identified individuals. Full details of data access and safety are in Annex 6.

# Methodology

## 4.1. Study sites

The study covers two large cities in two regions of Africa (Johannesburg, South Africa, Southern Africa, and Abidjan, Côte d'Ivoire, West Africa).

Johannesburg is the largest city in South Africa, classified as a megacity, the 26th-largest in the world, with 14,167,000 inhabitants in 2021[67]. Johannesburg is in the eastern plateau area of South Africa known as the Highveld, at an elevation of 1,753 metres. It has a subtropical highland climate, with the summer months (October to April) characterised by hot days followed by afternoon thundershowers and cool evenings, and the winter months (May to September) by dry, sunny days followed by cold nights. Temperatures in Johannesburg are usuall mild due to the city's high elevation, with an average maximum daytime temperature in January of 25.6 °C (78.1°F), dropping to an average maximum of around 16 °C (61°F) in June. The city is a Köppen Climate Classification subtype Cwb, with January the warmest month, with an average temperature of 20.6°C (69.0°F).

Abidjan is the largest city in Ivory Coast, with a population of 4.7 million (estimation in 2014), which is 20 percent of the overall population of the country, and this also makes it the sixth most populous city proper in Africa. Abidjan lies on the south-east coast of the country, on the Gulf of Guinea. The city experiences a tropical wet and dry climate. In Abidjan, the wet season is warm and mostly cloudy, the dry season is hot and partly cloudy, and it is oppressive year-round. The hot season lasts for 5.7 months, from November 22 to May 13, with an average daily high temperature above 30.5°C (87°F). The hottest month of the year in Abidjan is April, with an average high of 31°C (88°F) and low of 25.5°C (78°F). The cool season lasts for 2.5 months, from July 3 to September 19, with an average daily high temperature below 28°C (82°F). The coldest month of the year in Abidjan is August, with an average low of 23°C (73°F) and high of 27°C (81°F). Abidjan is a Köppen Climate Classification subtype Am: Tropical Monsoon Climate, with February the warmest month with an average temperature of 28.3°C (83.0°F)[68].

Given the marked differences between these two cities, the study will be able to undertake cross-region comparisons and assess whether variations between sites can be explained by differences in vulnerability, heat hazards or in underlying disease patterns between the two areas. Having one coastal, tropical city, and one inland, high elevation, sub-tropical city will provide a valuable basis to test the transferability of models and methods in different geographic settings.

## 4.2. Major steps and integration

The study will be conducted following four major steps presented in the Figure 2 below. These steps correspond to the study aims.

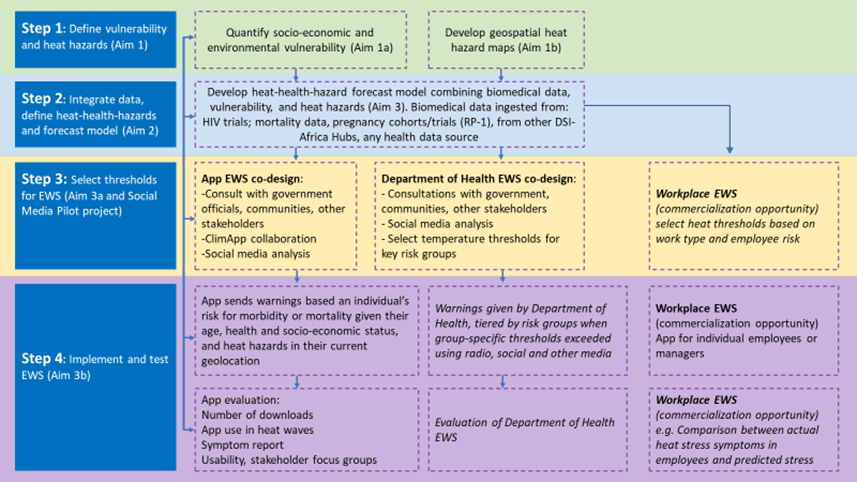


Figure 2 Steps for achieving the aims and objectives of the study

The research activities over the two cities will be conducted in a collaborative and an integrated way, including the teams from the two cities working together, sharing experiences and expertise in team building and capacity-building, and attending workshops organised in one setting or the another.

Details of the work from data acquisition to data processing and data analyses are described below.

## 4.3. Data description and acquisition

The different data types of data and datasets are presented in Annex 1

, and described below under each aim and sub-aim, as relevant.

### 4.3.1. For Aim 1a: Quantify intra-urban socio-economic and environmental vulnerability

*For Aim 1a, none of the data that will be used are identifiable human health data. All social or demographic and health data that will be used are available publicly and only at an aggregated level. The data privacy and related ethical concerns that apply to analysis of human health data therefore do not apply to data used in this aim.*

Two categories of data will be used.

The first category is socio-economic survey data. These data will provide information on household economic status or living standards, demographic and environmental characteristics.

* City observatory surveys data (e.g., in South Africa, Gauteng City Region Observatory (GCRO) survey data including the Quality-of-Life survey (GCRO QoL survey 2020/21), and in Abidjan the Bureau National d'Etudes Techniques et de Development Reports including the Urban Slums Reports and similar data sets)
* Population Censuses and other community surveys (e.g., in South Africa, Stats SA census and community surveys such as Census 2011 and the Community Survey 2016, and similar datasets in Ivory Coast)

The second category consists of proxy spatial measures of socio-economic conditions:

* Satellite image-based estimation of land use, land cover types, building and roof types (GCRO/Terra Image 2012).
* Access to services including electricity
* Proximity to health care centres and services (e.g., Johannesburg’s public hospital and clinics directory, and similar data sets in Abidjan, Ivory Coast)
* Access to green spaces (e.g., City Parks directory, and similar data sets in Abidjan, Ivory Coast))
* Other (e.g., Google Street View analysis, distributions of human population, such as SEDAC-gridded human population data)

The satellite imagery and other geospatial data acquired in Aim 1a is used to characterize urban vulnerabilities to heat and for mapping heat- or temperature-related hazards. For example, the satellite image views of the city will be classified to capture different typologies (e.g., slum, industrial, green areas, and suburban residences), where this information will then be used to account for differences in the diffusion of heat (in Aim 1b). The Stats SA small area layer from the 2011 Census is an example of socio-economic geospatial dataset that may be used for identifying urban residential areas and their characteristics such as dwelling types and the type of access to services electricity and water, which may be useful when quantifying and mapping heat hazards.

The data for achieving Aim 1a is freely available, or already curated within the existing IBM or University of Cape Town databases.

### 4.3.2. For Aim 1b: Develop high resolution urban temperature hazard maps

*For Aim 1b, only weather and climate data will be used for this aim (no human data or measurements). The data privacy and related ethical concerns that apply to analysis of human health data therefore do not apply to data used in this aim.*

For urban temperature, several datasets will be used for downscaling which will subsequently also be used for developing an urban-level Early Warning System. The weather datasets we will use can be summarised into two categories.

The first category, **retrospective or historical data**, will serve as proxy ground truth data entailing observations or proxy observations for temperature across the case study urban areas (e.g., Johannesburg and Abidjan). Included in this are model-based re-analysis data, which use physics and numerical weather simulations in combination with observations from a global network of sources (e.g., weather stations, satellites, and simulation models). Examples of these datasets include:

* MODIS and LANDSAT Land Surface Temperature (LST)
* UK Met Office/IMPALA CP4-A (4km resolution 10-year climate simulation over Africa)
* TAHMO weather stations (https://tahmo.org/)
* WMO Global Summary Of the Day (GSOD) weather stations
* ECMWF ERA-5 Re-analysis of regional circulation fields (zonal and meridional winds, temperature, humidity near surface (10m and 2m above the surface) as well as on relevant pressure levels)
* ECMWF ERA5-Land which provides the output of a higher resolution (10km) resolution land surface model forced by ERA5 and includes variables such as LST

This category of data will serve as a reference of what happened in the past in terms of weather and climate.  Reanalysis datasets, such as ECMWF’s ERA5 and ERA5 Land have a global spatial extent with hourly temporal resolution and coverage starting from 1950 to the present, which is extremely valuable for machine-learning studies which require sufficient temporal coverage to perform model training and evaluation. ERA5 reanalysis data consists of several climate variables, such as 2m temperature, sea-surface temperature, relative humidity, soil temperature, and total precipitation, all of which can be used as additional sources of predictability when post-processing and de-biasing physics-based weather/climate forecasts with machine learning.

The second category of data consists of outputs from dynamical physics-based **forecast or prospective data** emanating from global or regional ocean-atmosphere coupled climate models:

* ECMWF sub-seasonal to seasonal forecasts
* NOAA Climate Forecast System (v2) sub-seasonal to seasonal forecasts
* IBM The Weather Company seasonal forecasts
* NOAA NCEP GFS short-range forecasts

Dynamical short and medium-range forecasts, like the ones listed above, will assist in developing an urban heat-health early warning system based on sound physical properties. ECMWF issues 1–46-day ensemble forecasts at daily intervals for horizons of 1-14 days ahead, and twice a week for horizons from 15-46 days ahead. NOAA CFSv2 issues a 90 day ahead horizon forecast every day. IBM’s The Weather Company issues a 1-7 month ahead daily forecast every month. All of these forecast products are available operationally from the IBM PAIRS platform at various spatial resolutions and are accessible via an API. These forecast datasets can be pre-processed on demand in PAIRS on using map-reduce functions in order to be harmonized with climate datasets hosted by CSAG.  The spatial resolution of most of these gridded forecast products varies between 10km to 50km.

The data for achieving Aim 1b is already curated within the existing IBM or University of Cape Town databases.

### 4.3.3. For Aim 2: Develop a spatially and demographically stratified heat-health outcome forecast model

*Aim 2 incudes reanalysis of existing data on human health. Data privacy and related ethical concerns that apply to analysis of human health data are considered below and further unpacked in Section 6 below.*

The datasets used in study activities for this aim are drawn from existing longitudinal datasets collected in the studies conducted in Johannesburg, South Africa or in Abidjan, Côte d'Ivoire. This encompasses data from trials and cohorts among HIV-infected adults, HIV-uninfected adults, and adults participating in longitudinal studies related to COVID-19 prevention or treatment. We will include data from all participants 18 years and older at the time of enrolment in the primary studies, and not use sampling to select particular sub-sets of individuals.

Data acquisition will occur through:

* 1. approaching the Principal Investigators of potentially eligible datasets to request data sharing and their participation in the research project
  2. approaching data repositories that store the relevant data
  3. completion of legal agreements on data transfer between the study team, and the data provider or data repository, as relevant

The following databases will be included:

* 1. datasets from studies done within research facilities within the City of Johannesburg or Abidjan geographical area, regardless of whether participants live outside the city boundaries
  2. databases that contain the population groups of interest, without restrictions on the size of database
  3. studies that enrolled participants since the year 2000
  4. the investigators have given permission for use of the data

Though the geolocation data on the area where individual participants reside is the optimal level of geolocation that we require, we will not exclude databases where participant geolocation data is not available. In those instances, we will use the geolocation of the research site where participants attended study visits as a proxy for their geolocation.

We will include both interventional and non-interventional studies. Longitudinal studies include cohorts where participants are enrolled and then followed over time.

We will select datasets to include based on convenience sampling, where we balance considerations around:

* 1. the number of datasets we can analyse within the available financial and human resources available for the study
  2. the size of the study and duration of follow-up in a study
  3. the number and types of health variables collected
  4. the quality of the study and the data
  5. the availability of participant geolocation data
  6. data sharing stipulations within the study (e.g., whether data sharing was explicitly included in informed consent, or if data sharing was mandated by study funders)
  7. interest of the study investigators in sharing data

We will include data from three categories of trials/cohorts, each done in different study populations. Drawing data from three different categories of studies aims to ensure that the size of the datasets are sufficiently large for the training of machine learning models, but another valuable reason for the inclusion of data from different population groups is that this provides an opportunity to examine sub-group level associations, and thus compare heat-health outcomes across groups.

The first category of studies is those done among people with HIV infection. Most of these studies were done to assess the effectiveness or efficacy of different antiretroviral treatment drugs. These data will allow us to investigate specific associations between heat and other environmental exposures and health outcomes in HIV-infected people. The associations between heat and health outcomes in HIV-infected people are likely to differ in size and nature from heat-health linkages in other population groups.

The second category of trials or cohorts is those done among populations at risk for acquiring HIV infection (HIV-negative adults). These are mostly trials that were done to assess the effectiveness or efficacy of interventions to prevent HIV acquisition. Interventions include vaginal microbicides and pre-exposure prophylaxis with antiretroviral drugs.

The third category of studies that will provide data for re-analysis will be sought from COVID-19 prevention or treatment trials conducted in primary study sites of Johannesburg[69-71]. These may include trials of COVID-19 vaccines. Studies done on COVID-19 followed large cohorts of people both with and without HIV, providing a rich data set to map against the weather and other data collated in Aims 1a and 1b. Further, these longitudinal COVID-19 studies have been done in recent years when there have been clearer provisions for data sharing.

Variable categories which will be drawn from these studies are presented in Annex 3 (Table 5). These include variables related to social, economic and demographic characteristics, geospatial location of participants, self-reported health status, medical examination findings, laboratory outcome data and outcomes of other types of investigations are presented in Annex 3 (Table 5).

The use of the personal information of study participants places this study under the restrictions and permissions described by the Protection Of Personal Information Act of South Africa (POPIA 2013). The requirements of the act with regards to processing of personal information, accountability, security and quality, informed consent, and subject notification, are detailed in the associated Data Management Plan[72].

## 4.4. Data pre- and post-processing, formatting

The following sections describe the data processing steps specific to this research study. The over-arching data management plan is described in detailed in the associated DMP document.

### 4.4.1 For Aim 1a: Quantify intra-urban socio-economic and environmental vulnerability

* The data will be aggregated/interpolated onto the same spatial units (such as the Stats SA Census 2011 small area units, or GCRO QoL survey spatial units), and saved as ESRI Shapefile(s) or equivalent (GeoJSON/PostgreSQL-postGIS)
* Image-based data aggregation to spatial units will include mean values and counts, or proportions, as well as variance to represent the heterogeneity within each spatial unit (e.g., mix of *township formal* and *township informal*)

*No identifiable human data are used in Aim 1a activities (only data aggregated to small levels, and not human data at an individual level).*

### 4.4.2 For Aim 1b: Develop high resolution urban temperature hazard maps

* Data will be processed to align with the CSAG Data Reference Syntax (DRS) which is based on international DRS such as CORDEX and CMIP.  This provides a uniform structure for directories and filenames.
* Climate and Forecasting (CF) conventions will be applied to ensure file level meta-data conformity.  Data will be stored primarily in NetCDF (NetCDF4-HDF5) format through subsetssoftware and will enable access to particular data in other formats (e.g., CSV).

*There is no human data involved in Aim 1b activities.*

### 4.4.3 For Aim 2: Develop a spatially and demographically stratified heat-health outcome forecast model

* The health datasets will be harmonised into a single coherent dataset through codebook remapping on the variables listed in Annex 3.
* Codebook remapping entails:
  + Step 1 taking the variables across the different datasets
  + Step 2 feeding the data into the study database
  + Step 3 defining the variables on a continuum which allows for all data to be

For example, pre-term delivery might be translated from “number of days” to an ordinal categorical variable such as the trimester of preterm birth.  Another example occurs when measuring levels of education. In some cases, the number of years attended may be the measure as opposed to situations where total subjects passed would need to be put on the same continuum to be useful for analysis.

* Throughout the pre-processing process, steps will be taken to deidentify data to ensure patient privacy is adhered to. These steps are described in full in section and the associated DMP .
* We will undertake post-processing based on machine-learning and a forecast skill evaluation as described in the next sub-sections.

#### 4.4.3.2 Machine-Learning Based Post-Processing

* It has been shown by various studies that machine learning has the ability to improve the forecast skill of dynamical models at a global and regional scale[73-75].  In order to obtain the best possible forecast of temperature and other relevant climate variables, machine learning will be used to post-process and blend various dynamical forecasts.  This approach will aim to de-bias dynamical forecasts to better account for rare or extreme events that will be of particular interest for the urban early warning system.  Approaches that can be implemented include algorithms like Xgboost or Ngboost, convolutional, and recurrent neural networks.

#### 4.4.3.3 Forecast Skill Evaluation

Evaluating the skill of a forecast product in the vicinity of the regions of interest is necessary to establish what the expected performance will be in an operational setting, especially as a function of horizon (lead time).  The following baseline forecasts are normally used to evaluate the skill of a new forecast product:

* Persistence - which is obtained by simply taking the most recent available observation at a given location
* Climatology - which is obtained by taking a long-term average (~30 years) of a given variable, at a specified location/pixel and for a given time-of-year. This represents the climate that can be expected at a certain point on the globe for the considered time-of-year.

More advanced climatologies that take trends into account (as a result of climate change) also serve as a good forecast baseline. The goal for a new forecast product is to improve on these baselines. For Machine Learning-based post-processed forecasts, the goal will be to improve on the skill of the baselines as well as the base dynamical model(s) on which post-processing methods are being applied, as well as climatology. Emphasis will be given to evaluating forecasts for their ability to predict extreme events and events categorized in e.g., the 70th, 80th, 90th percentile ranges. Evaluation metrics that will be used will include Mean Absolute Error (MAE), Anomaly Correlation Coefficient (ACC), and Ranked Probability Skill Score (RPSS), among others.

Once an ML-based post-processed GCM forecast product has been developed for lead times of between 1-30 days with sufficient skill, downscaling approaches will be applied to obtain urban-level estimates of temperature and other relevant variables, conditioned on the coarse GCM forecast. This forecast will be an integral component for the Urban Heat-Health EWS. Figure 2 shows an overview of the components related to downscaling of temperature maps and how the resulting high-resolution maps will interface with the EWS.

Diagram

Description automatically generated

Figure 3 Overview of the data and workflow for downscaling and how it will be used in the Urban Heat-Health Early Warning System

## 4.5. Data analyses and outputs

### 4.5.2. For Aim 1a: Quantify intra-urban socio-economic and environmental vulnerability

* Dimensionality reduction methods will be implemented, for example the use of Principal Component Analysis (PCA)to identify dominant correlation structures across the variables. This will produce candidate predictors or features that best explain intra-urban socio-economic variability as well as a way to combine the components into a single indicator that confers combined socio-economic and environmental vulnerability that may be associated with a wide range of climate risks, but in our contexts, this would be primarily heat. Hierarchical or multi-level clustering methods can also be applied to identify regions similar in terms of vulnerability. A critical consideration in terms of methods, will be accounting for spatial variability, hence the appropriate approach would be to implement techniques such as spatial principal component analysis and geodemographic clustering methods. Aggregation into a single output map will be implemented using spatial multi-criteria analytic method.

### 4.5.3. For Aim 1b: Develop high resolution urban temperature hazard maps

The analyses will mainly include downscaling.

* MODIS and LANDSAT LST data will be processed to fill in gaps resulting from cloud cover on a day-to-day basis.  Different gap filling approaches will be explored including auto-regression and multi-variate regression imputation using coarse scale temperature estimates from ERA5 and/or ERA5-Land.
* Land surface temperatures are a derived product from the Landsat multi-spectral band information.  Landsat 8 Top Of Atmosphere (TOA) 16-bit images which are available at a 16-day temporal interval. The Operational Land Imager (OLI) sensor provides multispectral bands 1-7 and 9 with 30 m spatial resolution and Thermal Infrared Sensor (TIRS) which collects data at a 100 m spatial resolution (bands 10 and 11). One of the main algorithms that will be implemented makes use of the thermal infra-red band and band 5 & 7 [76].   Typically, the resulting resolution using this algorithm is 100 m for LST, but with the requirement to produce higher spatial resolution temperature grids for correlation with health outcomes, more complex approached will be pursued aimed at 30 m resolution LST.  One of these will build upon the work of Mushore et al, 2022, who developed a multivariate regression method that used several land cover indices (e.g., Normalized Vegetation Index (NDVI)) and topographical variables (elevation and slope) to downscale LST to 30 m resolution.  Based on this, a statistical downscaling model will be developed, trialling a number of machine learning approaches, to predict sub-daily (night time minimum and day time maximum) spatially coherent local scale (~30 m resolution) near surface air temperature patterns across the city.  Recent methods such as the random forest regression kriging or quantile random forest regression kriging methods [77] will be investigated. Generative models, specifical variational auto-encoders are candidates that will also be considered for this task.
* Physics based models will be used to derive near surface air temperature from LST data based on the physical disaggregation of latent and sensible heat fluxes.  Machine learning approaches will be compared to the physical approach.

### 4.5.4. For Aim 2: Develop a spatially and demographically stratified heat-health outcome forecast model

* Data analysis done as part of Aim 1a will be extended under Aim 2, in the development of the heat-health outcome model, where geolocated socio-economic variables will be used as predictors and machine learning models (e.g., the quantile random forest model which enables identification of important variables and a measure of uncertainty) will identify the strongest predictor variables from the suite of socio-economic variables
* Machine learning models will be developed and tuned (various supervised approaches will be trialled, such as regression, decision trees, random forest, K-Means, support vector machines) that capture the predictive relationship (associative) between negative health outcomes and temperature.
* Examples of health impacts that are of interest include:
  + Impact on other health outcomes: High blood pressure, respiratory stress, cardiac conditions. HIV related blood tests such CD4 counts, basophils PRT, eosinophils PRT, MCV (Mean Cell Volume), lymphocytes PRT
* Predictors that will be considered for this model will be required to be geographically coincident (to the limits of the health geolocation data) and will include downscaled near surface air temperature estimates, estimates of socio-economic conditions (e.g., dwelling type, commuting conditions, etc.) and more preferably the indicator of socio-economic vulnerability to heat.
* An important component to consider in the models that will be developed will the discovery of sub-groups or risk strata. As an initial step, different manual aggregations will be explored based on inputs from biomedical or epidemiological experts.
* Participants in the databases will be stratified into different sub-groups, such as age category and socio-economic status. As a follow-up step, the applicability of automatic sub-group discovery methods will be investigated, and these include the use of open-source tools such as Pysubgroup and auto-stratification tools that have been developed by IBM.
* Models will be trained to maximize predictive skill relative to different groups in order to identify models that provide the most skill for outcomes of the greatest value.
* The results of machine learning analyses need to be considered in terms of existing evidence of plausible biological or social pathways, for example. Engagement with health practitioners and researchers will be done to provide feedback on the relative value of different models, and selection of optimum algorithms.

# Special data management considerations

## 5.1. Data types

Multiple categories of data will be used across the project divided into four categories.

* + 1. Health related data

Health related data for this study will focus on clinical trials datasets generated by clinical trials undertaken in the case study cities (Johannesburg and Abidjan).  Trials datasets will be identified based on the scale of the trial as well as the availability of geospatial variables (e.g. clinic locations or other geospatial information) in order to allow spatial mapping of health outcomes and the intersection with socio-economic spatial mapping and climate variable spatial mapping.

**Health related data is generally considered to be personal data** as it pertains to individuals medical records, diseases history and health events.  In some cases health related data is also considered *personally identifiable data* where individuals can be identified either directly through names, ID numbers, etc. or indirectly deduced from address, GPS locations, or cross referencing with other identifiers.  Many countries now have specific laws dictating conditions for sharing and processing of such data and associated protective measures to avoid unauthorized access and potential harm to individuals involved in the studies.  In South Africa the relevant legislation is the Protection of Personal Information Act (POPIA 2013).  This necessitates processing in compliance with the specific requirements of the POPIA act.  These requirements and compliance with them is outlined in detail in the section below: POPIA compliance  
  
Due to above mentioned privacy, ethical and legal requirements, health related data acquisition almost always involves an associated DSA between the data provider and the HEAT Center (see data management workflow description below).  In particular, where health data is traveling across international borders, compliance with the relevant national laws, permissions, and associated requirements for the sharing and transfer of health data will be followed.

Once data is in South Africa, local ethical clearance and, where data constitutes personal information, POPIA compliance apply.

* + 1. Climate/weather data

These data include both observational-based datasets (weather station observations, satellite proxy observations, and processed/gridded observations).  Gridded climate data produced from atmospheric re-analysis and climate simulations will form historical gridded climate observations and forecasts.

Climate-related data will in almost all cases involve accessing open data repositories such as Copernicus Climate Data Store (CDS), or Earth System Grid Federation data systems.  Climate-related data will be stored on IBM PAIRS data storage and/or CSAG/UCT data storage systems. however, CSAG/UCT will manage and update the primary data index for climate-related data available to the consortium.

The large majority of climate datasets used are available through open data policies with no restrictions on non-commercial research use.  In some cases a processing fee may be charged. We will cite the original source.

* + 1. Remote sensing data excluding climate dataThe focus is on data derived from satellite sensors, mainly optical imagery (e.g. satellite images of urban centers) as well as indicators of physical measures such as land surface temperature, soil moisture estimates, vegetation condition, land use and cover, etc.) for the purposes of estimating environmental quantities, and land use/building density etc.  Remote sensing data will not be used to identify individuals in any way and does not constitute sensitive or personally identifiable data. Climate data is often collected through remote sensing. However, for the purposes of this study, we will classify climate data collected through remote sensing as in the first category of data (Climate/weather data), and not in this category.

Remote sensing related data will in most cases involve accessing open data repositories such as Copernicus Climate Data Store (CDS), Sentinel data systems.  Remote sensing related data will be stored on IBM PAIRS data storage and/or CSAG/UCT data storage systems. CSAG/UCT will manage and update the primary data index for remote sensing related data.

Most remote datasets used are available through open data policies with no restrictions on non-commercial research use.  We will cite the original source.

* + 1. Areal/Geospatial socio-economic dataThese data represent measures of socio-economic and related conditions such as household economic status, access to services such as water and sanitation, dwelling type, etc.  Typical sources include national census data, and more focused household & demographic survey data. These data are not sourced through remote sensors.

Socio-economic data will be sourced from both open data repositories as well as restricted access repositories (e.g. South African census data, and GCRO Quality of Life Surveys).  Primary copies will be indexed and stored on CSAG/UCT data storage but versions may already exist or will be uploaded on IBM PAIRS to enable analysis through IBM PAIRS.

South African census data is already available through the UCT DataFirst data repository and GCRO QoS data is available through the GCRO open data platform as well as through direct queries with GCRO.

South African census data is aggregated up to small areas and so does not constitute personally identifiable data.  GCRO Quality of Life survey data is likewise aggregated to small areas and does not constitute personally identifiable sensitive data.

## Data indexing and processing (harmonization, homogenization, normalization)

Data currently available on the CSAG/UCT data storage system as well as data available on IBM PAIRS will be indexed using an appropriate pragmatic meta-data standard and this index will be made available by CSAG/UCT data platform (Gitlab).

CSAG currently implements a **Data Reference Syntax (DRS)** which is a structured mapping from a controlled vocabulary of meta-data elements to a directory and file naming syntax. This is standard practice within climate data management and we will continue to implement this approach for the climate and remote sensing datasets.

The current DRS is documented on the CSAG Gitlab Wiki which is accessible only to team members with requisite credentials.

**Integration with DSI-Africa Open Data Science Platform** will ensure that meta-data propagates to the ODSP system.

Environmental data (climate data, remote sensing data, socio-economic mapping data) will be homogenized to align with existing meta-data and storage standards (CF conventions and OGC standards) as documented on the Gitlab Wiki.

**Health data harmonization and de-identification** will require meta-data/codebook mapping from source codes through to common codes (to be developed).  Codebook mapping will involve both RP and DMAC participation but will be technically implemented by DMAC.  The codebook remapping will be developed and documented separately to this DMP.

## **Data management, documentation and curation**

Data will be managed by a team of data specialists within the study and ***primary data management*** will take place on the CSAG/UCT data platform.

IBM PAIRS will continue to implement their own internal data management; however, this will include sharing of meta-data indexing with CSAG/UCT for data related to NIH HE²AT Center and DS-I Africa.

Individual partner institutions will have their own data and computing platforms ranging from central institutional platforms through to personal computing devices.  Data management within these partner institutions does not fall under this data management plan and we refer to this as ***secondary data management***.  However, partners will be encouraged to regularly consider what locally managed data should be integrated into the study wide data management and made available to the broader DSI-Africa programme and beyond.

Primary data management will involve:

* Homogenization to agreed data archive standards (see above for each class of data)
* Version control of datasets that are regularly updated to ensure prior version remain accessible
* Harmonizing of health-related data
* Meta-data indexing within the CSAG/UCT DMP as well as DS-I Africa ODSP
* Documentation of data on CSAG/UCT GitLab wiki
* Sharing of data management code through CSAG/UCT GitLab code repositories

## Data Security and Confidentiality of Potentially Sensitive Information

The use of health datasets requires careful consideration of data security and confidentiality.  This is strongly guided by the relevant legislative context for each datasets including the specific country legislation around the use of personal/sensitive data and the cross-border transfer of such datasets.  The development and negotiation of these data sharing agreements.

In order to support the development of the Data Sharing Agreements and ensure compliance with legislation around access to and protection of personal/sensitive information the following security measures will be implemented:

### De-identification

POPIA Section 10 prescribes the principle of Minimality which means that only information relevant to the purpose of processing should be processed.  Where personal information is acquired that is required to fulfill the research purposes described by the relevant research project protocols, de-identification will be implemented according to the following protocol which is guided by US Department of Human and Health Services (HSS) guidelines[78].

1. Street addresses may be aggregated into geographical regions such that many records map to the same region and it is no longer possible to derive individual residential locations.  For example, in RP2, where the highest spatial granularity is needed to map urban heat health outcomes, census small areas or wards with spatial scales of the order of 2 to 5 km will be used to aggregate records.  Aggregation will also consider the number of records that map to the same geographical area.
2. Latitude/longitude coordinates may be “jittered” which involves adding random values to each coordinate such that the exact location is lost[79] assesses different approaches to location masking and jittering), while retaining sufficient geographical information to support the analysis.  This will likely be the case for RP1 where jittering will be used to shift latitude/longitude locations in the order of 10s of kilometers, adequate to prevent locating individual residential locations.

Geo-location masking/jittering and aggregation will be considered adequate through expert determination involving experts from UCT, IBM and NIH.

#### Data encryption

Primary data (before de-identification) will be encrypted using 256bit AES (Advanced Encryption Standard) a standard established by the US NIST (National Institute of Standards and Technology) with encryption keys only available to the minimum number of people required to implement the de-identification processing once data is transferred to the UCT HPC system.

The use of cryptographic modules validated to NIST [FIPS 140-2](https://en.wikipedia.org/wiki/FIPS_140-2) is required by the United States Government for encryption of all data that has a classification of Sensitive but Unclassified (SBU) or above. We will use these standards when transferring and storing data.

For datasets that include personal identifiers (ie. personally identifiable data), datasets will be encrypted by the original data provider and transferred to UCT through a secure data transfer that uses TLS encryption.

#### Storage isolation

Primary encrypted data will be stored in an isolated virtual storage server that is only accessible to private data team.

### 5.4.2. Network firewall and Virtual Private Network

CSAG compute infrastructure sits within the UCT intranet and falls under the broader UCT security policies and services.

UCT broader security policy: <http://www.icts.uct.ac.za/sites/default/files/image_tool/images/286/UCT_Information_Security_Policy_PC03_2020.pdf>

UCT ICTS implements a Cisco firewall and Cisco Virtual Private Network service to ensure that intranet access is limited to authorized users.

UCT firewall policy  
<http://www.icts.uct.ac.za/uct-perimeter-firewall-policy>

These network level security measures ensure that access to the UCT intranet is controlled to authorized UCT users.  Furthermore, CSAG server access is controlled by CSAG authentication and authorization as described below.

### 5.4.3. Local authentication and authorization

While CSAG compute servers sit within UCTs intranet, the CSAG/UCT platform (storage and compute infrastructure) implements an independent authentication and authorization service (Linux filesystem and LDAP authentication).  Access to restricted datasets will therefore be ensured through both UCT authentication protocols and internal CSAG DMP authentication and authorizations.

## Procedure for making data available to qualified individuals

The Steering Committee (SC) will develop a Data Access Request Form which people requesting data need to complete prior to consideration of their request. The form will include a proposal outline of the intended research, and the procedures to maintain data confidentiality and security.

The SC will review and approve or reject all requests from the research community, which includes scientists or medical professionals working at academic, non‐profit or government institutions, or commercial companies. We will ensure that all requests conform with NIH policies and procedures including compliance with informed consent procedures if relevant, and any limitations stipulated by the institutions/investigators who contributed data to either of the RPs. The study will comply with the principles of the Data Protection Act of the country of the participating site. Some countries may have restrictions on data sharing outside of the host country, which we will abide by. Researchers who request to share the resources of the HE²AT Center will need to agree to not seek to identify individuals within the dataset, not distribute the data to any other entity, keep the data secure, and acknowledge the HE²AT Center and DS-I Africa as appropriate in publications and presentations (the exact acknowledgement text to be agreed among the DS-I Africa Program). Further, researchers who share the Center’s resources will be strongly encouraged to collaborate with and train African investigators as part of the work they conduct with the resources shared.

There may be exceptions to the resource sharing plans outlined above. Firstly, there may be considerations around intellectual property protections for the research products that the consortium aims to commercialize. Decisions about resource sharing in these circumstances will follow the NIH policies in this regard, including those on resources sharing, disseminating unique research resources, and program income. The NIH will be provided with a copy of documents or sample of these products developed under the grant award (e.g. the Digital App). A large portion of the data we will use in the Project activities will be drawn from the IBM-PAIRS platform, which contains several datasets that require a license for access. We will thus not be able to share those datasets without permission from the licensee.

Clearly, the rights and privacy of individuals who participate in research must be protected at all times. Thus, data intended for broader use should be free of identifiers that would permit linkages to individual research participants and variables that could lead to deductive disclosure of the identity of individual subjects. All data shared beyond the HEAT Center will be de-identified following the procedures described above. Requests for original data containing personally identifiable or sensitive information will be referred back to the original study.

The HE²AT Center team will especially endeavour to make the unique research resources that we develop readily available to other researchers working on climate and health. There is a pressing need within the field to expedite the translation of data into knowledge and, in particular, into interventions that protect people against extreme heat and other manifestations of climate change. We thus feel obligated to expedite our responses to such data requests. Fellow researchers wishing to access the data will thus not have to wait for our research findings to be accepted for publication or for the ‘final research data” prior to data sharing, provided the research questions they are exploring do not directly overlap with the specific question we are addressing.

The study involves ‘data sharing’ on a large scale, where we are reliant on the willingness and ability of data owners to share. In that spirit, as the holders of the database it will be beholden on us to share the data with other groups, provided the original data providers agree to sharing and the procedures of HE²AT Center data sharing processes are followed. We will develop formal collaboration agreements around data sharing with the data owners who contribute data to the database, which set out the terms and conditions for data reuse. This is important, as for example, some of these investigators may not wish their data to be shared with third parties, for example. While a willingness to share data forms the basis of this project and most often reflects a desire of investigators to collaborate, this may not necessarily translate into a willingness to then share data beyond the Center. Sharing of the datasets gathered into the database will thus require signed agreement from the original data owners and may require an application to the local IRB where the original study was done. People who make use of the database will have to undertake to adhere to the authorship guidelines stipulated in the collaboration agreement signed between the HE²AT Center and the research groups who contributed their data.

Of note, much of the data and resources generated by the Hub will be useful for a range of other disciplines. The research resources generated in this study such as vulnerability-heat-health data visualization will also potentially have a wide range of applications, among people working on urban geography or planning, for example. Additionally, the datasets have tremendous potential to answer a wide range of questions, outside of environmental health. We will facilitate access to these resources by interested parties. We also undertake to share data and resources generated by the HE²AT Center with other Hubs or components in the DS-I Africa program, wherever possible. In particular, the DS-I Africa Open Data Science Platform will be used as a mechanism for making relevant datasets from the HE²AT Center available across the program, and beyond.

Data and data resources (primarily software/code) will be shared through standard protocols such as File Transfer Protocol (FTP) servers and the UCT instance of Gitlab (software version control system). UCT also hosts an open data portal which will be used to make more final datasets such the urban heat vulnerability maps available following FAIR principles. Data that is made publicly or otherwise available through data use and sharing agreements will be accompanied by metadata such as data dictionaries and data descriptors such as principal investigator, funding sources, data collector, project description, sample and sampling procedures, temporal and geographic coverage of the data collection, variables, technical information on files (file formats, linking, etc.), interviewer guides and coding instruments.

Analytic data sets will be provided as de-identified data files that can be read by common statistical package software, such as SAS or STATA. De-identified data sets will have names and other personal health identifiers removed (see section 5). Data and other resources will be transferred to others under the terms of a DSA to ensure that the data will be used for the proposed purpose and that no attempts will be made to identify participants. We will maintain records of all researchers who have been given access to the research resources. Results of the HE²AT Center activities will be shared with the research community and the public through conference presentations, publication in peer-reviewed journals and media interactions, as described in the Training and Engagement Core(TEC).

According to the Research Data Management policy, ‘publicly funded research data are a public good, produced in the public interest, which should be made openly available with as few restrictions as possible in a timely and responsible manner’. Data is therefore open by default, closed by exception (e.g. privately-funded research, or research with commercialisation possibilities).

## 5.6. Principles of data sharing and open access

### 5.6.1. Restrictions to data sharing

According to Section 4.6 of the [UCT Research Data Management policy](http://www.digitalservices.lib.uct.ac.za/dls/rdm-policy),

 "[n]ecessary constraints on the availability of data include the protection of personal data; the protection of intellectual property; the protection of commercial interests of study partners; and security concerns."

Strategies to limit restrictions may include data being anonymised or aggregated; gaining participant consent for data sharing; and gaining copyright permissions.

### 5.6.2. Discoverability

The HEAT Center and DMAC will implement FAIR principles:

* Data will be **findable** through publicly accessible and searchable meta-data indexes (need to decide if the DS-I Africa ODSP is going to be the primary mechanisms for meta-data search, UCT also hosts the ZivaHub repository)
* Data will be **accessible** either openly through a public facing component of the DMP data repository, or through a data access request to DMAC (where a Data Sharing Agreement is required)
* **Inter-operability** will be enabled through the strict adherence to established data and meta-data standards (see above)
* **Reuse** will be supported through rigorous documentation of the data including limitations and guidance for reuse.

# Ethical and Privacy considerations

**Ethical considerations and legal requirements**

**Summary of general ethical and legal requirements**

The project raises several ethical and legal considerations. This includes 1) the use of secondary data for research purposes that may differ from the original purpose of the study, 2) sharing data across countries in Africa, and 3) the presence of potential indirectlyidentifying information.

The project will endeavour to ensure the protection of human subjects and their data through procedures that adhere to the following legal and ethical guidelines:

* Declaration of Helsinki
* ICH Good Clinical Practice
* Ethics in Health Research (Department of Health South Africa) (109)
* Protection of Personal Information Act (POPIA)
* US Department of Health and Human Services, Protection of Human Subjects in Research, 45 CFR 46
* Any additional African country-specific protection of data legislation

Table 1 below summarises some of the ethical and legal considerations for the project, and how these will be addressed. Additional details describing these procedures are described thereafter and further detail included in the DMP in Annex 6.

Table 1: Summary of ethical, legal and data management considerations that underpin study activities

|  |  |  |
| --- | --- | --- |
| Ethical considerations | Legal requirements | Data management activities |
| * The purposes of the research are clearly defined in the study protocol and the reasons that the study needs to involve secondary personal data are provided * Processing of historical data for research purposes in the study serves the public interest * Specific concerns of vulnerable populations such as pregnant women, neonates, and children have been considered * We assess ethics concerns in the studies that contribute data, including information on the ethics committee that provided approval for the original study, securing a copy of the informed consent of the study, if possible, and clarifying any concerns about the study with the data provider, where feasible * Participant informed consent has been addressed, separated into two scenarios:   + Participant has signed a ‘’broad consent’’ for use of their data in future research projects, even though the precise nature of the future research is unclear, this will allow data sharing without further ethical approvals * Reconsenting will not be required for use of secondary data in the HE2AT Center, as conforming to standard practices and relevant ethics guidelines in South Africafor use of secondary anonymised data Ethics committee has approved the conduct of the study and are provided with information on each of the studies that contribute data * Risks to individuals’ privacy and potential for identification are limited as far as possible * The data provider will bear the responsibility of ensuring their local ethical procedures are fulfilled before sharing the data, as will be stipulated in the data sharing agreement | * POPIA requirements are met * The data provider has determined that the country’s legal requirements for personal data security are adhered to * Data sharing agreements are in place for each dataset from data provider and their legal department(if applicable) to UCT/Wits RHI * Data sharing agreements specify the use of the data through UCT’s database by other partners like IBM, CeSSHAR and Wits RHI * Any time limitations on use of data are adhered to | * A data management plan clearly defines how the data is transferred, stored and secured * Clearly defined and limited personnel will have access to data through access-controlled platforms, with specific individuals authorized for different levels of data access (e.g. only specific named individuals have access to encryption codes, and a log of all those who have access to the data for analysis) * Only data that is necessary for the conduct of the study will be collected and data will be anonymised, when required * Publications will not include any potentially identifiable information from participants, and only aggregated data will be published or presented publicly |

**Informed consent process in primary study**

Waivers of informed consent are relatively common in public health settings. For example, in instances where data are used on individuals or groups to prevent the spread of communicable diseases, as occurred in the COVID-19 pandemic. Oftentimes health information is collected without consent to determine the incidence of various health conditions. This is analogous to our study where we wish to calculate the incidence of adverse health outcomes during hot periods, compared to cold periods. The justification for these breaches of the normal right of individuals to informed consent is that this right can be subordinated to the needs of the community if the two are in conflict. Naturally, the extent to which this subordination applies to research needs to be determined for each research proposal.

The informed consent requirements for participation in research are stricter than those for both individual and public health interventions because the research participants are exposed to the risks of research but often do not receive any benefits even when the research is successful. In contrast, public health interventions, such as programmes to reduce the impacts of extreme heat on health, often do benefit the recipients of the intervention. In the absence of a waiver, the research would be severely impeded, or not possible in fact.

We acknowledge that for research using identifiable human data, researchers must normally seek consent for the collection, analysis, storage and/or reuse. The Declaration of Helsinki, paragraph 25 notes that it is preferable that potential research participants be informed in advance about any possible future uses of their data so that it will not be necessary to re-contact them when another such use is proposed. When the information in databases are anonymized and aggregated, it is often impossible to identify the source persons from whom to ask informed consent for new uses. This is especially true for databases that were established in the past when informed consent for future uses was not sought. Research ethics committees may decide to waive the informed consent requirement by considering the Declaration of Helsinki's exception to the requirement as stated in paragraph 25: “There may be situations where consent would be impossible or impractical to obtain for such research or would pose a threat to the validity of the research. In such situations the research may be done only after consideration and approval by a research ethics committee.” The CIOMS International Ethical Guidelines for Epidemiological Studies (commentary on Guideline #4) also provides support on this exception.

No reconsenting of study participants will be conducted for use of secondary data in the HE2AT Center, which meets the requirements of ethical guidelines, listed above, that are being adhered to.

**POPIA compliance and protection of personal information**

The use of health datasets requires careful consideration of data security and confidentiality.  This is strongly guided by the relevant legislation context for each datasets including the specific country legislation around the use of personal/sensitive data and the cross-border transfer of such datasets.

The Protection of Personal Information Act of South Africa (2013) institutes limitations on the processing of personal information, but also provides the legal basis for the use of personal information for the purpose of scientific research.  POPIA functions alongside other legislation and regulatory structures governing research in South Africa, such as the Constitution of the Republic of South Africa, the National Health Act No 61 of 2003 and regulations as prescribed by the Minister of Health, such as the South African Department of Health guidelines on Ethics in Health Research Principles, Processes and Structures 2015 (“DOH Guidelines”). The law which takes precedent will be that which provides the most comprehensive protections to the rights of individuals in South Africa.

Section 6 of the Act indicates that the Act does not apply in cases where personal information has been de-identified to the extent that it cannot be re-identified again.  Where this is the case, the information can be used without restriction or conditions of the Act.  It is envisaged that a number of health databases acquired for the project will have been de-identified to the extent that re-identification is virtually impossible.  However, where this is not the case, the following sections of POPIA provide a basis for processing health data.

While Section 15(1) requires that personal information processing must be compatible with the purpose of collection, Section 15(3)(e) allows for processing for the purposes of historical, statistical and research purposes regardless of the original purpose of collection. As the study will be using health datasets collected prior to the project, we will never be processing personal information compatible with the original purpose of collection, unless specific consent for further processing was agreed to by the data subjects.  Section 15(3)(e) therefore provides the basis for processing information where this specific consent was not obtained.

Furthermore, while Section 18(1) lays out requirements for informing the data subject that their personal information is being processed for a particular purpose, Section 18(4)(f) provides for an exemption from these requirements where the information will be used for historical, statistical or research purposes.

Furthermore, Section 14(2) allows for the retention of personal information for research purposes as long as safeguards are in place to prevent the use of the personal information for other purposes

Section 19 requires that security measures are implemented to prevent the unlawful access to or processing of personal information.  We will implement extensive data security measures. Section 20 describes requirements around individuals operating on or processing personal information.  We have identified the individuals who will be responsible for this processing and will continually update this list.  Access to personal information is restricted through passwords and other security measures so that only individuals authorized by the responsible party have access.

Section 21 requires a written contract between the responsible party and the operators implementing processing.  This contract will be signed by all individuals processing personal information and is presented in the data management plan and includes agreement that the operator will inform the responsible party where there is any basis for believing that personal information has been accessed by an unauthorized person.

As noted in Section 27 of the POPIA Act, in instances where the study investigators are requesting access to data for a purpose which was not originally agreed upon by the participant in the informed consent, we will only use the data for research and statistical purposes.

We believe that the inclusion of such data for the research purposes set out in this protocol is justified as it serves a major public interest, and the processing of data is necessary for the purposes concerned. Climate change, and in particular, rapidly escalating temperatures and heat waves, are a major threat to health across the world, and especially in Africa. Many parts of Africa are warming at twice the average warming rate. Additionally, there is little research on the impacts of extreme heat on pregnant women and children in Africa and this information is vital for developing services to assist these population groups as temperatures escalate. There is thus a major public interest to understand the implications of high temperatures on health in Africa. This information will help to inform the allocation of funding to ameliorate these impacts and assist in selecting appropriate interventions in this regard. Climate change is one of the greatest threats to health in the 21st century, and indeed is an existential threat to the whole globe. Studies such as that described here are essential for addressing this threat.

**Data sharing compliance**

The data provider will be required to provide a contractual assurance in the data sharing agreement that informed consent procedures were followed for the original studies and that they have consent to share the data for further scientific research of this nature, in compliance with all ethical and data protection legislation applying in the host country.

**Considerations specific to data source**

As described in Section 4.3, data will be sought from various sources, such as data repositories, from research sites run by Wits University, from other universities or research organisations in South Africa and across sub-Saharan Africa. The ethical and legal considerations for each study may vary by data source, as presented in Table 8 below

Table 2: Ethical and legal considerations for different types of data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Types of data sources | | | Approach | |
| Type | **Description** | **Example** | **Ethical** | **Legal** |
| Open source | Available online for download, no data sharing agreements required | Reassessing pregnancy intention and its relation to maternal, perinatal and neonatal outcomes in a low-income setting: A cohort study. <https://rdr.ucl.ac.uk/articles/dataset/Pregnancy_intention_and_perinatal_outcomes/8949050> | * No further ethics submissions or approvals required * Data will be treated with equal protection standards as other datasets | * Assess study individually for any terms and conditions of sharing (may have embargo’s on duration of access etc.) |
| In data repository (including from a network) | Access restricted and study staff need to request data, which is reviewed prior to approval | Worldwide Antimalarial Resistance Network, IMPAACT, ACTG | * The data provider will be responsible for their own ethics committee submission, if required | * Complete data requests forms as required by data repository * Complete DSA if required * Adhere to terms and conditions of sharing (e.g., time embargo) * Data transfer is subject to country specific legal requirements |
| Wits University (including Wits Health Consortium) | Participants have signed broad or narrow informed consent | ADVANCE study(Ezintsha) | * To submit study synopsis, informed consent forms, and ethical approval letter to Wits HREC to assess and grant approval for use in the HEAT Center, following a review of the study synopsis, informed consent form, ethics approval letter, and data sharing agreement, where available | * DSA to be completed and shared with Wits HREC * POPIA compliance |
|
| South African university/organisation (other than Wits) |  | SAFE Passage – large cohort from Stellenbosch University | * In the DSA, the data provider, with their legal representative, will provide assurance that the relevant local ethical procedures were followed | * DSA to be completed and shared with HREC * POPIA compliance |
| Study and/or PI is from another country | Study was conducted in an African country that is not South Africa – can be single- or multisite study | Kesho Bora study: Maternal ART during pregnancy and breastfeeding prevents more infections than short-course prophylaxis. Study conducted in Kenya, Burkina Faso, and South Africa | * In the DSA, the data provider, with their legal representative, will provide assurance that the relevant local ethical procedures were followed | * DSA, to be shared with Wits HREC * POPIA compliance * Compliance with personal data legislation in other countries (data provider is responsible), including cross-border data transfer laws , assurance is provided in DSA |

This DMP (Annex 6) provides additional detail on the steps in which ethical and legal standards will be upheld in the acquisition, handling, and safekeeping of data.As additional databases become available to the study team and DSAs are agreed to, we will notify the Wits Human Research Science Faculty Ethics Committee in writing of the new studies that have agreed to contribute data to the study every six months. We will provide information about the database prior to actual data transfer. We will inform the committee of the name of the study, study acronym, contact details of the data providers, details of the ethics approval of the original study including informed consent parameters, where available, study country and provide a copy of the DSA. We will make additional information available on the study where requested by the ethics committee.

**Risks to the participants in the primary studies**

Risks to a participant in this study are minimal as there are no new procedures involved and steps will be taken to protect their personal information and to avert risks of identification, as described in the data management plan.

# Training, engagement, and capacity building

Coordinated and consistent implementation of health system operations can be effective and cost-efficient but are labour and resource intensive. For the early warning system to decrease the number of heat-health risks among vulnerable populations, there needs to be a parallel effort to ensure response capabilities are developed that take advantage of the alerts. Working with the ministries of health to build a response system that is useful, intuitive, and capable of being implemented for long-term use is critical. High-risk maps can be translated and shared with partner countries to be integrated into appropriate communication and adaptation planning mechanisms.

Through the Training and Engagement Core (TEC), the HE²AT Center will serves as a Center of Excellence and resource for the DS-I Africa Research Training Program. The HE²AT Center will engage with the communities in which the research is conducted, to increase acceptability of the study and its objectives. These activities use existing engagement mechanisms, engaging with local Community Advisory Boards and media and distributing newsletters. Because the solutions developed by the Center address two major concerns for policymakers (how to warn people about a heatwave and then to track its impacts), we anticipate that local, national and international policymakers will be willing to engage.

The specific aims of the TEC are to:

* Advance knowledge and skills of early career researchers and scientists of the HE²AT Centre and other Research Hubs on transdisciplinary heat-health research using data science approaches
* Build knowledge, skills, and competencies in solutions for addressing heat-health impacts, particularly early warning and response systems
* Develop and implement a systematic strategy for forming, maintaining, and strengthening engagement with governments, industry, non-governmental and related organizations, and communities
* Disseminate results through scientific publications and presentations at scientific meetings, community engagement, policy engagement, and reports to NIH

***Approach***

The TEC engagement plan will take a phased approach to prioritize different stakeholder groups at different stages of the program lifecycle. Given some of the potential research outputs, it will be important to think about engagement across the implementation phases (e.g., pre- mid – post) – with a focus on the pre-implementation stage in Year 1 and 2, with mid-level implementation starting in Year 3. Across RP2 there are several points of engagement needed, ranging from global to local levels (Figure 4). The TEC will aim to address needs across multiple scales during the lifespan of the development of the heat-health early warning system.

Utilizing the working group for policy and stakeholder engagement the TEC will provide targeted support to the RP2 team to help meet the specific needs of end-users, including linking research outputs to policy interventions and building awareness through the documentation of health impacts of extreme heat. A secondary goal of the engagement working group will be to foster learning opportunities across HE²AT Centre teams to strengthen knowledge engagement in the context of climate change and health. This may include inclusion of presentations to the technical talk series focusing on engagement approaches, as well as journal clubs within/across working groups to discuss relevant literature.

Chart, sunburst chart

Description automatically generated

Figure 4 Levels of stakeholder engagement

The TEC and engagement working group will provide technical support to the activities of research project 2.

* Conducting a stakeholder mapping exercise and workshop to identify and assess needs of decision-makers and end-users involved in the development and implementation of the heat-health early warning system
* Conducting a heat-health policy review and analysis for selected study locations
* Conducting a technical needs assessment and training for health workers to build awareness of heat-health risks, as well as identify, prioritize, and implement intervention strategies based on forecasts and warnings.
* Supporting the engagement with communities and decision-makers to describe the study goals, and gather inputs on the development of the heat-health early warning, including potential intervention strategies.
* Conducting capacity building exercises with key stakeholders including end-users and decision-makers to ensure core competencies are established and knowledge increased to utilize the heat-health early warning system information and implement response actions

# Dissemination and publication plan

The HE²AT Center will disseminate study findings and other learnings in a timely manner across our consortium, partners, policymakers, the scientific community, other stakeholders, and the general public to maximise its impact. Our engagement plan includes:

* Scientific publications and presentations to scientific conferences
* Community engagement; and
* Policy engagement

Scientific publicationsfrom the Research Studies will be published in open access journals with data sharing.

The final protocol for Research Project 2 will be made available on open access to allow replication of the approach in other settings and will be registered with the appropriate registry. It is anticipated that the study will generate high quality data sufficient for at least four publications in high-impact journals. A publication plan will be developed within the consortium that will define the potential publications, authorship, and timeline for manuscript production, as set out in the Administration Core.

Each year there will be *presentations to scientific conferences*, such as the International Society for Environmental Epidemiology, the NIH biannual meetings, the EDCTP Forum, and others. Presentations will also be made upon request to meetings organised by the African Regional Office of the World Health Organization, the United Nations Environment Programme, the United Nations Development Programme, and others. Further, study findings and communication materials may be shared via the Global Heat Health Information Network. The emphasis will be on promoting junior researchers to increase their confidence and skill in preparing and delivering scientific presentations. Costs for these presentations are included in the Research Studies.

The HE²AT Center will promote the Studies and their findings among the *communities* in which the research will be conducted, to ensure increased acceptability of the Studies and their objectives, as well as promote awareness of heat-health risks. Dissemination tools such as newsletters, brochures, and introductory study posters, may be distributed among the community. We will emphasise the guiding principles of Good Participatory Practice, including respect, mutual understanding, integrity, transparency, accountability, and autonomy. Some of the team members already engage frequently with the media about climate or extreme heat concerns, which opens up further avenues for the dissemination of findings.

Study results will be disseminated to the *local, provincial, and national authorities*, to inform their Climate Change and Health Adaptation Planning. Very often these authorities do not receive sufficient technical support for the design and implementation of the types of solution that will be developed in Research Study 1 and Research Study 2 (Overall Hub Aim 4).

The Africa CDC will also play a major role in dissemination activities of the HE²AT Center and are fully committed to participate actively in the activities of the Center, including hosting joint dissemination meetings and/or stakeholder consultations. Timely dissemination and clear communication of results is key for updating guidance and for informing sub-national, national, and international policies. The results will contribute to the development and implementation of heat-health early warning and integrated surveillance systems and associated policies, strategies, and interventions to reduce the health risks of extreme heat in Cote d'Ivoire and South Africa, with the potential of broader application for low- and middle-income countries.

# Collaborations

Partners in the study are drawn from South Africa, Zimbabwe, Côte d'Ivoire and the United States. The institutions names and locations are presented below in Figure 5.

The RP2 brings together all the institutions of the HE²AT Center in efficient and complementary roles. To get the most out of the collaboration each aim will be run as a collaboration between the different partners.

Map

Description automatically generated

Figure 5 Partners collaborating on the HE²AT study from across Africa and the United States.

# Main activities, milestones and timelines

Table 3: GANTT chart: milestones and timelines

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Sept 2021-June 2022** | | | | **July 2022-June 2023** | | | | **July 2023-Julne 2024** | | | | **July 2024-June 2025** | | | | **July 2025-June 2026** | | | | **Co\*** |
| **Milestones Study quarters** | **1** | **2** | **3** | **4** | **1** | **2** | **3** | **4** | **1** | **2** | **3** | **4** | **1** | **2** | **3** | **4** | **1** | **2** | **3** | **4** | **1-4** |
| Research Project 2 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Initial and continuing IRB and ethics compliance (M6) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Training of HE²AT Center on Research Project 2 methods (M6) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Protocol paper published (M12) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Data analysis plan written (M6) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Data harmonization across DSI-Africa platform and other grantees(M12) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Biomedical data recoded and merged into one database (M8) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Database prepared for vulnerability-heat-health analysis (M12) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Publications submitted on primary and secondary study outcomes |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Engagement with national government for EWS development |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Stakeholder engagements (community and health departments) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Publications submitted on Methodological aspects (M38) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Publication on vulnerability-heat-health data |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Review of District-level surveillance systems performance (M38) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| PhD enrolment (M7) until award (M54) of one student |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Dissemination meetings and conferences |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Final report submitted (M58) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

\* Close-out, M=Month

# References

1. Analitis, A., et al., *Synergistic Effects of Ambient Temperature and Air Pollution on Health in Europe: Results from the PHASE Project.* Int J Environ Res Public Health, 2018. **15**(9).

2. Chersich, M.F., et al., *Associations between high temperatures in pregnancy and risk of preterm birth, low birth weight, and stillbirths: systematic review and meta-analysis.* BMJ, 2020. **371**: p. m3811.

3. Gasparrini, A., et al., *Mortality risk attributable to high and low ambient temperature: a multicountry observational study.* Lancet, 2015. **386**(9991): p. 369-75.

4. Manyuchi A, et al., *Extreme heat events, high ambient temperatures and human morbidity and mortality in Africa: A systematic review.* Environmental Research (under review), 2020.

5. 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.

6. Watts, N., et al., *The 2020 report of The Lancet Countdown on health and climate change: responding to converging crises.* Lancet, 2020.

7. Debela, N., et al., *Perception of climate change and its impact by smallholders in pastoral/agropastoral systems of Borana, South Ethiopia.* Springerplus, 2015. **4**: p. 236.

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

9. 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.

10. Kupika, O.L., et al., *Local Ecological Knowledge on Climate Change and Ecosystem-Based Adaptation Strategies Promote Resilience in the Middle Zambezi Biosphere Reserve, Zimbabwe.* Scientifica (Cairo), 2019. **2019**: p. 3069254.

11. Scott, A.A., et al., *Temperature and heat in informal settlements in Nairobi.* PLoS One, 2017. **12**(11): p. e0187300.

12. Chersich, M.F., et al., *Impacts of climate change on health and wellbeing in South Africa.* International journal of environmental research and public health, 2018. **15**(9): p. 1884.

13. Naicker, N., et al., *Indoor temperatures in low cost housing in Johannesburg, South Africa.* International journal of environmental research and public health, 2017. **14**(11): p. 1410.

14. Bidassey-Manilal, S., et al., *Students’ perceived heat-health symptoms increased with warmer classroom temperatures.* International journal of environmental research and public health, 2016. **13**(6): p. 566.

15. Scott, A.A., et al., *Temperature and heat in informal settlements in Nairobi.* PloS one, 2017. **12**(11): p. e0187300.

16. Wright, C.Y., et al., *Indoor temperatures in patient waiting rooms in eight rural primary health care centers in northern South Africa and the related potential risks to human health and wellbeing.* International journal of environmental research and public health, 2017. **14**(1): p. 43.

17. Codjoe, S.N., et al., *Impact of extreme weather conditions on healthcare provision in urban Ghana.* Social Science & Medicine, 2020. **258**: p. 113072.

18. Dapi, L.N., et al., *Heat impact on schoolchildren in Cameroon, Africa: potential health threat from climate change.* Glob Health Action, 2010. **3**.

19. 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).

20. Sera, F., et al., *How urban characteristics affect vulnerability to heat and cold: a multi-country analysis.* International journal of epidemiology, 2019. **48**(4): p. 1101-1112.

21. Chen, W., et al., *Impacts of Building Features on the Cooling Effect of Vegetation in Community-Based MicroClimate: Recognition, Measurement and Simulation from a Case Study of Beijing.* International Journal of Environmental Research and Public Health, 2020. **17**(23): p. 8915.

22. Li, B., et al., *Analysis of the relationship between urban landscape patterns and thermal environment: a case study of Zhengzhou City, China.* Environmental Monitoring and Assessment, 2020. **192**(8): p. 1-13.

23. Yang, Q., X. Huang, and J. Li, *Assessing the relationship between surface urban heat islands and landscape patterns across climatic zones in China.* Scientific reports, 2017. **7**(1): p. 1-11.

24. Scovronick, N., et al., *The association between ambient temperature and mortality in South Africa: A time-series analysis.* Environmental research, 2018. **161**: p. 229-235.

25. Wichmann, J., *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, 2017. **587**: p. 266-272.

26. Chersich, M.F., et al., *Associations between high temperatures in pregnancy and risk of preterm birth, low birth weight, and stillbirths: systematic review and meta-analysis.* bmj, 2020. **371**.

27. Asamoah, B., T. Kjellstrom, and P.O. Ostergren, *Is ambient heat exposure levels associated with miscarriage or stillbirths in hot regions? A cross-sectional study using survey data from the Ghana Maternal Health Survey 2007.* Int J Biometeorol, 2018. **62**(3): p. 319-330.

28. Bakhtsiyarava, M., K. Grace, and R.J. Nawrotzki, *Climate, birth weight, and agricultural livelihoods in Kenya and Mali.* American Journal of Public Health, 2018. **108**(S2): p. S144-S150.

29. MacVicar, S., et al., *Whether weather matters: Evidence of association between in utero meteorological exposures and foetal growth among Indigenous and non-Indigenous mothers in rural Uganda.* PLoS One, 2017. **12**(6): p. e0179010.

30. Azongo, D.K., et al., *A time series analysis of weather variability and all-cause mortality in the Kasena-Nankana Districts of Northern Ghana, 1995-2010.* Glob Health Action, 2012. **5**: p. 14-22.

31. Diboulo, E., et al., *Weather and mortality: a 10 year retrospective analysis of the Nouna Health and Demographic Surveillance System, Burkina Faso.* Global Health Action, 2012. **5**(1): p. 19078.

32. Egondi, T., et al., *Time-series analysis of weather and mortality patterns in Nairobi's informal settlements.* Global health action, 2012. **5**(1): p. 19065.

33. Musengimana, G., et al., *Temperature variability and occurrence of diarrhoea in children under five-years-old in Cape Town metropolitan sub-districts.* International journal of environmental research and public health, 2016. **13**(9): p. 859.

34. Sewe, M., et al., *The association of weather variability and under five malaria mortality in KEMRI/CDC HDSS in Western Kenya 2003 to 2008: a time series analysis.* International journal of environmental research and public health, 2015. **12**(2): p. 1983-1997.

35. UNAIDS, *Seizing the moment. Global AIDS update. Tackling entrenched inequalities to end epidemics.* <https://www.unaids.org/en/resources/documents/2020/global-aids-report>, 2020.

36. UNAIDS, *UNAIDS data 2020.* <https://www.unaids.org/en/resources/documents/2020/unaids-data>, 2020.

37. Lubis, N., et al., *Prospective cohort study showing changes in the monthly incidence of Pneumocystis carinii pneumonia.* Postgraduate medical journal, 2003. **79**(929): p. 164-166.

38. Bacchetti, P., *Incidence of HIV-related deaths in the United States: seasonality and trend.* Stat Med, 1997. **16**(6): p. 645-52.

39. Bogaerts, J., et al., *AIDS-associated cryptococcal meningitis in Rwanda (1983–1992): epidemiologic and diagnostic features.* Journal of infection, 1999. **39**(1): p. 32-37.

40. Bulterys, P.L., et al., *Environmental predictors and incubation period of AIDS-associated Penicillium marneffei infection in Ho Chi Minh City, Vietnam.* Clinical infectious diseases, 2013. **56**(9): p. 1273-1279.

41. Chariyalertsak, S., et al., *Seasonal variation of disseminated Penicillium marneffei infections in northern Thailand: a clue to the reservoir?* Journal of Infectious Diseases, 1996. **173**(6): p. 1490-1493.

42. de Oliveira-Silva, M.B., et al., *Seasonal profile and level of CD4+ lymphocytes in the occurrence of cryptosporidiosis and cystoisosporidiosis in HIV/AIDS patients in the Triângulo Mineiro region, Brazil.* Rev Soc Bras Med Trop, 2007. **40**(5): p. 512-5.

43. Dietz, V., et al., *Active, multisite, laboratory-based surveillance for Cryptosporidium parvum.* The American journal of tropical medicine and hygiene, 2000. **62**(3): p. 368-372.

44. Djawe, K., et al., *Environmental risk factors for Pneumocystis pneumonia hospitalizations in HIV patients.* Clinical infectious diseases, 2013. **56**(1): p. 74-81.

45. Inungu, J.N., A.A. Morse, and C. Gordon, *Risk factors, seasonality, and trends of cryptosporidiosis among patients infected with human immunodeficiency virus.* Am J Trop Med Hyg, 2000. **62**(3): p. 384-7.

46. Le, T., et al., *Epidemiology, seasonality, and predictors of outcome of AIDS-associated Penicillium marneffei infection in Ho Chi Minh City, Viet Nam.* Clin Infect Dis, 2011. **52**(7): p. 945-52.

47. Miller, R., et al., *Seasonal variation in mortality of Pneumocystis jirovecii pneumonia in HIV-infected patients.* International journal of STD & AIDS, 2010. **21**(7): p. 497-503.

48. Ndirangu, J., et al., *Decline in early life mortality in a high HIV prevalence rural area of South Africa: evidence of HIV prevention or treatment impact?* AIDS (London, England), 2010. **24**(4): p. 593.

49. Tuli, L., et al., *Correlation between CD4 counts of HIV patients and enteric protozoan in different seasons - an experience of a tertiary care hospital in Varanasi (India).* BMC Gastroenterol, 2008. **8**: p. 36.

50. Wang, Y.-F., et al., *Serological surveillance for Penicillium marneffei infection in HIV-infected patients during 2004–2011 in Guangzhou, China.* Clinical Microbiology and Infection, 2015. **21**(5): p. 484-489.

51. Astrom, C., et al., *Developing a heatwave early warning system for Sweden: evaluating sensitivity of different epidemiological modelling approaches to forecast temperatures.* Int J Environ Res Public Health, 2014. **12**(1): p. 254-67.

52. Ebi K, et al., *Heat watch/warning systems save lives: Estimated costs and benefits for Philadelphia 1995-98.* Bull. Am. Meteorol. Soc, 2004. **85**(8): p. 1067–73.

53. Ebi, K.L. and J.K. Schmier, *A stitch in time: improving public health early warning systems for extreme weather events.* Epidemiologic reviews, 2005. **27**(1): p. 115-121.

54. Hess, J.J., et al., *Building Resilience to Climate Change: Pilot Evaluation of the Impact of India's First Heat Action Plan on All-Cause Mortality.* J Environ Public Health, 2018. **2018**: p. 7973519.

55. Lowe, D., K.L. Ebi, and B. Forsberg, *Heatwave early warning systems and adaptation advice to reduce human health consequences of heatwaves.* International journal of environmental research and public health, 2011. **8**(12): p. 4623-4648.

56. Morin, C.W., et al., *Unexplored opportunities: use of climate-and weather-driven early warning systems to reduce the burden of infectious diseases.* Current Environmental Health Reports, 2018. **5**(4): p. 430-438.

57. Zhang, K., et al., *Comparing exposure metrics for classifying ‘dangerous heat’in heat wave and health warning systems.* Environment international, 2012. **46**: p. 23-29.

58. Lee, V., F. Zermoglio, and K. Ebi, *Heat waves and human health–emerging evidence and experience to inform risk management in a warming world.* United States Agency for International Development, 2019.

59. WMO and WHO, *Heatwaves and Health: Guidance on Warning-System Development.* 2015. [**https://www.who.int/globalchange/publications/WMO\_WHO\_Heat\_Health\_Guidance\_2015.pdf**](https://www.who.int/globalchange/publications/WMO_WHO_Heat_Health_Guidance_2015.pdf).

60. Casanueva, A., et al., *Overview of existing heat-health warning systems in Europe.* International journal of environmental research and public health, 2019. **16**(15): p. 2657.

61. Lee V, et al., *Heat waves and human health: emerging evidence and experience to inform risk management in a warming world.* USAID, 2019.

62. Hess, J.J., et al., *Building resilience to climate change: pilot evaluation of the impact of India’s first heat action plan on all-cause mortality.* Journal of environmental and public health, 2018. **2018**.

63. Mandal, J., et al., *Hierarchically porous polymer coatings for highly efficient passive daytime radiative cooling.* Science, 2018. **362**(6412): p. 315-319.

64. Vanosac JK, Herdt AJ, and Lochbaum MR, *Effects of physical activity and shade on the heat balance and thermal perceptions of children in a playground microclimate.* Building and Environment, 2017(126): p. 119-131.

65. Hess, J.J. and K.L. Ebi, *Iterative management of heat early warning systems in a changing climate.* Annals of the New York Academy of Sciences, 2016. **1382**(1): p. 21-30.

66. Wu, Y., et al., *Performance of heat-health warning systems in Shanghai evaluated by using local heat-related illness data.* Science of the total environment, 2020. **715**: p. 136883.

67. Treivish, A., *The Largest, Metropolitan, and Global Centers in the World Arena: Matching Networks.* Regional Research of Russia, 2021. **11**(4): p. 543-554.

68. *Weatherbase*. 2022 [cited 2022; Available from: <https://www.weatherbase.com/>.

69. Madhi, S.A., et al., *Safety and immunogenicity of the ChAdOx1 nCoV-19 (AZD1222) vaccine against SARS-CoV-2 in people living with and without HIV in South Africa: an interim analysis of a randomised, double-blind, placebo-controlled, phase 1B/2A trial.* The Lancet HIV, 2021. **8**(9): p. e568-e580.

70. Makoni, M., *COVID-19 vaccine trials in Africa.* The Lancet Respiratory Medicine, 2020. **8**(11): p. e79-e80.

71. Vaccines, J. and P. BV, *A study of Ad26. COV2. S for the prevention of SARS-CoV-2-mediated COVID-19 in adult participants (ENSEMBLE).* ClinicalTrials. gov: NCT04505722, 2020.

72. JASINSKAITE, E. and Ø.Y. SKJEI, *Combining Deep Privacy with an Attribute-driven Generative Adversarial Network to Preserve Gender and Age in De-identified CCTV Footage*. 2021, University of Agder.

73. Dueben, P.D. and P. Bauer, *Challenges and design choices for global weather and climate models based on machine learning.* Geoscientific Model Development, 2018. **11**(10): p. 3999-4009.

74. Ham, Y.-G., J.-H. Kim, and J.-J. Luo, *Deep learning for multi-year ENSO forecasts.* Nature, 2019. **573**(7775): p. 568-572.

75. Weyn, J.A., D.R. Durran, and R. Caruana, *Can machines learn to predict weather? Using deep learning to predict gridded 500‐hPa geopotential height from historical weather data.* Journal of Advances in Modeling Earth Systems, 2019. **11**(8): p. 2680-2693.

76. Avdan, U. and G. Jovanovska, *Algorithm for automated mapping of land surface temperature using LANDSAT 8 satellite data.* Journal of sensors, 2016. **2016**.

77. Gao, B., A. Stein, and J. Wang, *A two-point machine learning method for the spatial prediction of soil pollution.* International Journal of Applied Earth Observation and Geoinformation, 2022. **108**: p. 102742.

78. *Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule*. [cited 2022 09/05/2022]; Available from: <https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html>.

79. Zandbergen, P.A., *Ensuring Confidentiality of Geocoded Health Data: Assessing Geographic Masking Strategies for Individual-Level Data.* Advances in medicine, 2014. **2014**: p. 567049-567049.

80. *DoH 2015 Ethics in Health Research - Principles, Processes and Structures 2nd Ed*.

81. Boulle, A., et al., *Data Centre Profile: The Provincial Health Data Centre of the Western Cape Province, South Africa.* Int J Popul Data Sci, 2019. **4**(2): p. 1143.

82. Sterne, J.A.C., et al., *RoB 2: a revised tool for assessing risk of bias in randomised trials.* BMJ, 2019. **366**: p. l4898.

83. Sterne, J.A., et al., *ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions.* BMJ, 2016. **355**: p. i4919.

84. Davidge, A., et al., *Heat Alert and Response Systems to Protect Health: Best Practices Guidebook.*

# Annexes

## Annex 1: Databases available to the HE²AT Center

Table 4 Overview of Databases to be used in RP2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Name and source of dataset** | **Description** | **UCT /PAIRS** | **Key variables** | **Spatio-temporal coverage** | **Relevance** |
| Biomedical data | | | | | |
| HIV and COVID 19 databases | Pooled health database from multiple large HIV and COVID vaccine trials conducted among adults in Johannesburg, South Africa | WHC studies | Participants are followed up every 3-months for several years, with a multitude of physical measurements, laboratory tests, images and health questionnaires |  | Research Project 2: The study population has high rates of co- morbidities and adverse health outcomes |
| Climate/ weather data | | | | | |
| European Centre for Medium-Range Weather Forecasts (ECMWF) -https://[www.ecmwf.i](http://www.ecmwf.i/) nt/en/forecasts/data sets/set-i | Outputs from a numerical weather prediction system, run twice daily, designed to produce state-of-the-art medium (10 days) global forecasts (contains only the latest forecast) | PAIRS | Temperature (Ground, Min, Max) at 2 m above ground; Solar irradiance; Wind speed (toward east, north) at 10 m above ground; Daily precipitation (total, rate); Dewpoint; Pressure | Spatial: Global coverage, 0.065536 deg.Temporal: 3 – 6 hourly & daily res.; Jan 2014 – Oct2019 | Determination of heat hazard; Thermal comfort metrics; combined climate exposures (forecasts) |
| IBM TWC (TheWeather Company) Current and historical weather | Data layers from The Weather Company, an IBM Business | PAIRS | Temperature (Change, Min, Max, Feels like); Solar irradiance; Wind (speed, gust & dir.), Rel.Humidity, Daily precipitation (total, rate); Dewpoint; 3-hrlyPressure Change | Spatial: Global coverage, 4km landmass and coastal waterways (hourly & daily res from 2015) | Determination of heat hazard; Thermal comfort metrics; combined climate exposures (historical) |
| Fifth-generation ECMWF high-res. Reanalysis (ERA5) https://cds.climate.c opernicus.eu | A global reanalysis dataset combining observed data with the output of meteorological models. Contains data for 2000, 2005 and from 2009 onwards. | PAIRS | Temperature (2 m above ground, Min, Max); Wind speed (toward east, north); Daily precipitation (total, rate, type); Atmospheric water/ water vapor content, Thermal radiation; Soil temperature; Vegetation types and cover (high, low) | Spatial: Global coverage, 0.131072 degrees PAIRS resolution (raw: 0.25 deg.) Temporal: hourly; coverage from Jan1980 – Jun 2019 | Determination of heat hazard; Thermal comfort metrics; combined climate exposures (historical) |
| Copernicus S2S seasonal forecast data | Model outputs forecasting climate conditions over the three months following the forecast initialization | UCT | Temperature 2m above ground (min, max), Daily precipitation (total) | Temporal: daily | Seasonal (weeks to 3 months) time horizon forecasting of relevant weather conditions (heat hazard) for early warning |
| CP4-A (NERC JASMIN) | Very high resolution (4km) simulations of historical and future climate over Africa | UCT | Temperature 2m above ground (min, max), daily precipitation (total), multi-level circulation | Temporal: daily | Dynamical downscaling to support sub-urban temp hazard mapping |
| CORDEX Africa (ESGF) | Ensemble of dynamically downscaled simulations of African climate to 50km,25km, and 10km resolution | UCT | Temperature 2m above ground (min, max), daily precipitation (total), multi-level circulation fields | Temporal: daily and sub-daily (6 hourly) | Dynamical downscaling of climate to support sub-urban temperature hazard mapping |
| GHCN station data (NOAA GHCN) | Global archive of daily weather station data | UCT | Temperature 2m above ground, daily precipitation (total) | Temporal: daily, station locations | To support statistical downscaling of temperature hazard |
| Other relevant data domains | | | | | |
| 30 m res Elevation (SRTM) (NASAhttps://www2.jpl.nasa.gov/srtm/) | Global elevation data from the Shuttle Radar Topography Mission (SRTM). | PAIRS | Elevation | Released in 2013 | Determination of heat hazard; Urban heat Island Effect |
| High res imagery (ESA Sentinel 2) European Space Agency https://sentinel.esa.i nt/web/sentinel/sentinel-data-access | Images from the Sentinel 2 satellite pair which view land surface regions in 13 spectral bands. | PAIRS | Urban land cover – vegetation coverage, morphological features, possibly pollution levels (AOT). Bands 4 (red), 8 (NIR)and SCL (Scene Classification); Aerosol Optical Thickness; NDVI- sh layer. | Spatial: Global coverage; 0.000064 deg res Temporal: every 5 days or faster; from Aug 2015 –Nov 2020. | if there is requirement to control for pollution effects or to look at combined heat-pollution exposures |
| Global population (SEDAC) - Gridded Population of the World (GPW), v4 https://sedac.ciesin.columbia.edu/data/c ollection/gpw-v4 | Distribution of human population (counts and densities) on continuous global raster surface. Input data are extrapolated toproduce population estimates for 5-year intervals | PAIRS | Population counts and density estimates | Spatial: Global coverage, 1km grid res Temporal: 5- yearly; coverage from Jan 2000 to Jan 2020 | Accounting for the population exposed |
| News coverage (https://www.gdeltproject.org/) | GDELT; Portion of news coverage about specific area and time related to Covid-19. | PAIRS | Global events derived from worldwide news coverage. | Spatial: Global coverage,0.008192 deg. res | Example for production of spatial data layer for news events |

## Annex 2: Characteristics of the studies

These are the characteristics of each of the trials or cohorts that are included in the study

Revised Cochrane risk-of-bias tool for randomized trials (RoB 2) [82]

Revised Cochrane risk-of-bias tool for randomized trials (RoB 2)

TEMPLATE FOR COMPLETION

Edited by Julian PT Higgins, Jelena Savović, Matthew J Page, Jonathan AC Sterne  
on behalf of the RoB2 Development Group

**Version of 22 August 2019**

The development of the RoB 2 tool was supported by the MRC Network of Hubs for Trials Methodology Research (MR/L004933/2- N61), with the support of the host MRC ConDuCT-II Hub (Collaboration and innovation for Difficult and Complex randomised controlled Trials In Invasive procedures - MR/K025643/1), by MRC research grant MR/M025209/1, and by a grant from The Cochrane Collaboration.



This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](http://creativecommons.org/licenses/by-nc-nd/4.0/).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Study details**   |  |  | | --- | --- | | **Reference** |  |   **Study design**   |  |  | | --- | --- | | X | Individually-randomized parallel-group trial | | £ | Cluster-randomized parallel-group trial | | £ | Individually randomized cross-over (or other matched) trial |   **For the purposes of this assessment, the interventions being compared are defined as**   |  |  |  |  | | --- | --- | --- | --- | | Experimental: |  | Comparator: |  |  |  |  | | --- | --- | | **Specify which outcome is being assessed for risk of bias** |  |  |  |  | | --- | --- | | **Specify the numerical result being assessed.** In case of multiple alternative analyses being presented, specify the numeric result (e.g. RR = 1.52 (95% CI 0.83 to 2.77) and/or a reference (e.g. to a table, figure or paragraph) that uniquely defines the result being assessed. |  |   **Is the review team’s aim for this result…?**   |  |  | | --- | --- | | £ | to assess the effect of *assignment to intervention* (the ‘intention-to-treat’ effect) | | £ | to assess the effect of *adhering to intervention* (the ‘per-protocol’ effect) |   **If the aim is to assess the effect of *adhering to intervention***, select the deviations from intended intervention that should be addressed (at least one must be checked):  £ occurrence of non-protocol interventions  £ failures in implementing the intervention that could have affected the outcome  £ non-adherence to their assigned intervention by trial participants  **Which of the following sources were obtained to help inform the risk-of-bias assessment? (tick as many as apply)**  £ Journal article(s) with results of the trial  £ Trial protocol  £ Statistical analysis plan (SAP)  £ Non-commercial trial registry record (e.g. ClinicalTrials.gov record)  £ Company-owned trial registry record (e.g. GSK Clinical Study Register record)  £ “Grey literature” (e.g. unpublished thesis)  £ Conference abstract(s) about the trial  £ Regulatory document (e.g. Clinical Study Report, Drug Approval Package)  £ Research ethics application  £ Grant database summary (e.g. NIH RePORTER or Research Councils UK Gateway to Research)  £ Personal communication with trialist  £ Personal communication with the sponsor |

**Risk of bias assessment**

Responses underlined in green are potential markers for low risk of bias, and responses in red are potential markers for a risk of bias. Where questions relate only to sign posts to other questions, no formatting is used.

**Domain 1: Risk of bias arising from the randomization process**

|  |  |  |
| --- | --- | --- |
| **Signalling questions** | **Comments** | **Response options** |
| **1.1 Was the allocation sequence random?** |  | Y / PY / PN / N / NI |
| **1.2 Was the allocation sequence concealed until participants were enrolled and assigned to interventions?** | Y / PY / PN / N / NI |
| **1.3 Did baseline differences between intervention groups suggest a problem with the randomization process?** |  | Y / PY / PN / N / NI |
| **Risk-of-bias judgement** |  | Low / High / Some concerns |
| Optional: What is the predicted direction of bias arising from the randomization process? |  | NA / Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable |

Domain 2: Risk of bias due to deviations from the intended interventions (*effect of assignment to intervention*)

|  |  |  |
| --- | --- | --- |
| **Signalling questions** | **Comments** | **Response options** |
| **2.1. Were participants aware of their assigned intervention during the trial?** |  | Y / PY / PN / N / NI |
| **2.2. Were carers and people delivering the interventions aware of participants' assigned intervention during the trial?** | Y / PY / PN / N / NI |
| **2.3. If Y/PY/NI to 2.1 or 2.2: Were there deviations from the intended intervention that arose because of the trial context?** |  | NA / Y / PY / PN / N / NI |
| **2.4 If Y/PY to 2.3: Were these deviations likely to have affected the outcome?** |  | NA / Y / PY / PN / N / NI |
| **2.5. If Y/PY/NI to 2.4: Were these deviations from intended intervention balanced between groups?** |  | NA / Y / PY / PN / N / NI |
| **2.6 Was an appropriate analysis used to estimate the effect of assignment to intervention?** |  | Y / PY / PN / N / NI |
| **2.7 If N/PN/NI to 2.6: Was there potential for a substantial impact (on the result) of the failure to analyse participants in the group to which they were randomized?** |  | NA / Y / PY / PN / N / NI |
| **Risk-of-bias judgement** |  | Low / High / Some concerns |
| Optional: What is the predicted direction of bias due to deviations from intended interventions? |  | NA / Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable |

Domain 2: Risk of bias due to deviations from the intended interventions (*effect of adhering to intervention*)

|  |  |  |
| --- | --- | --- |
| **Signalling questions** | **Comments** | **Response options** |
| **2.1. Were participants aware of their assigned intervention during the trial?** |  | Y / PY / PN / N / NI |
| **2.2. Were carers and people delivering the interventions aware of participants' assigned intervention during the trial?** | Y / PY / PN / N / NI |
| **2.3. [If applicable:] If Y/PY/NI to 2.1 or 2.2: Were important non-protocol interventions balanced across intervention groups?** |  | NA / Y / PY / PN / N / NI |
| **2.4. [If applicable:] Were there failures in implementing the intervention that could have affected the outcome?** |  | NA / Y / PY / PN / N / NI |
| **2.5. [If applicable:] Was there non-adherence to the assigned intervention regimen that could have affected participants’ outcomes?** |  | NA / Y / PY / PN / N / NI |
| **2.6. If N/PN/NI to 2.3, or Y/PY/NI to 2.4 or 2.5: Was an appropriate analysis used to estimate the effect of adhering to the intervention?** |  | NA / Y / PY / PN / N / NI |
| **Risk-of-bias judgement** |  | Low / High / Some concerns |
| Optional: What is the predicted direction of bias due to deviations from intended interventions? |  | NA / Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable |

Domain 3: Missing outcome data

|  |  |  |
| --- | --- | --- |
| **Signalling questions** | **Comments** | **Response options** |
| **3.1 Were data for this outcome available for all, or nearly all, participants randomized?** |  | Y / PY / PN / N / NI |
| **3.2 If N/PN/NI to 3.1: Is there evidence that the result was not biased by missing outcome data?** |  | NA / Y / PY / PN / N |
| **3.3 If N/PN to 3.2: Could missingness in the outcome depend on its true value?** |  | NA / Y / PY / PN / N / NI |
| **3.4 If Y/PY/NI to 3.3: Is it likely that missingness in the outcome depended on its true value?** | NA / Y / PY / PN / N / NI |
| **Risk-of-bias judgement** |  | Low / High / Some concerns |
| Optional: What is the predicted direction of bias due to missing outcome data? |  | NA / Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable |

Domain 4: Risk of bias in measurement of the outcome

|  |  |  |
| --- | --- | --- |
| **Signalling questions** | **Comments** | **Response options** |
| **4.1 Was the method of measuring the outcome inappropriate?** |  | Y / PY / PN / N / NI |
| **4.2 Could measurement or ascertainment of the outcome have differed between intervention groups?** |  | Y / PY / PN / N / NI |
| **4.3 If N/PN/NI to 4.1 and 4.2: Were outcome assessors aware of the intervention received by study participants?** |  | NA / Y / PY / PN / N / NI |
| **4.4 If Y/PY/NI to 4.3: Could assessment of the outcome have been influenced by knowledge of intervention received?** |  | NA / Y / PY / PN / N / NI |
| **4.5 If Y/PY/NI to 4.4:** **Is it likely that assessment of the outcome was influenced by knowledge of intervention received?** | NA / Y / PY / PN / N / NI |
| **Risk-of-bias judgement** |  | Low / High / Some concerns |
| Optional: What is the predicted direction of bias in measurement of the outcome? |  | NA / Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable |

Domain 5: Risk of bias in selection of the reported result

|  |  |  |
| --- | --- | --- |
| **Signalling questions** | **Comments** | **Response options** |
| **5.1 Were the data that produced this result analysed in accordance with a pre-specified analysis plan that was finalized before unblinded outcome data were available for analysis?** |  | Y / PY / PN / N / NI |
| **Is the numerical result being assessed likely to have been selected, on the basis of the results, from...** |  |  |
| **5.2. ... multiple eligible outcome measurements (e.g. scales, definitions, time points) within the outcome domain?** |  | Y / PY / PN / N / NI |
| **5.3 ... multiple eligible analyses of the data?** |  | Y / PY / PN / N / NI |
| **Risk-of-bias judgement** |  | Low / High / Some concerns |
| Optional: What is the predicted direction of bias due to selection of the reported result? |  | NA / Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable |

Overall risk of bias

|  |  |  |
| --- | --- | --- |
| **Risk-of-bias judgement** |  | Low / High / Some concerns |
| Optional: What is the overall predicted direction of bias for this outcome? |  | NA / Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable |



This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](http://creativecommons.org/licenses/by-nc-nd/4.0/).

Risk of Bias in Non-randomized Studies – of Interventions (ROBINS-I) assessment tool [83]

The Risk Of Bias In Non-randomized Studies – of Interventions (ROBINS-I) assessment tool

(version for cohort-type studies)

**Version 19 September 2016**



This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](http://creativecommons.org/licenses/by-nc-nd/4.0/).

ROBINS-I tool (Stage I): At protocol stage

**Specify the review question**

|  |  |
| --- | --- |
| Participants |  |
| Experimental intervention |  |
| Comparator |  |
| Outcomes |  |

**List the confounding domains relevant to all or most studies**

|  |
| --- |
|  |

**List co-interventions that could be different between intervention groups and that could impact on outcomes**

|  |
| --- |
|  |

ROBINS-I tool (Stage II): For each study

**Specify a target randomized trial specific to the study**

|  |  |
| --- | --- |
| Design | Individually randomized / Cluster randomized / Matched (e.g. cross-over) |
| Participants |  |
| Experimental intervention |  |
| Comparator |  |

**Is your aim for this study…?**

|  |  |
| --- | --- |
| £ | to assess the effect of *assignment to* intervention |
| £ | to assess the effect of *starting and adhering to* intervention |

**Specify the outcome**

Specify which outcome is being assessed for risk of bias (typically from among those earmarked for the Summary of Findings table). Specify whether this is a proposed benefit or harm of intervention.

|  |
| --- |
|  |

**Specify the numerical result being assessed**

In case of multiple alternative analyses being presented, specify the numeric result (e.g. RR = 1.52 (95% CI 0.83 to 2.77) and/or a reference (e.g. to a table, figure or paragraph) that uniquely defines the result being assessed.

|  |
| --- |
|  |

Preliminary consideration of confounders

Complete a row for each important confounding domain (i) listed in the review protocol; and (ii) relevant to the setting of this particular study, or which the study authors identified as potentially important.

“Important” confounding domains are those for which, in the context of this study, adjustment is expected to lead to a clinically important change in the estimated effect of the intervention. “Validity” refers to whether the confounding variable or variables fully measure the domain, while “reliability” refers to the precision of the measurement (more measurement error means less reliability).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **(i) Confounding domains listed in the review protocol** | | | | |
| Confounding domain | Measured variable(s) | Is there evidence that controlling for this variable was unnecessary?\* | Is the confounding domain measured validly and reliably by this variable (or these variables)? | OPTIONAL: Is failure to adjust for this variable (alone) expected to favour the experimental intervention or the comparator? |
|  |  |  | Yes / No / No information | Favour experimental / Favour comparator / No information |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **(ii) Additional confounding domains relevant to the setting of this particular study, or which the study authors identified as important** | | | | |
| Confounding domain | Measured variable(s) | Is there evidence that controlling for this variable was unnecessary?\* | Is the confounding domain measured validly and reliably by this variable (or these variables)? | OPTIONAL: Is failure to adjust for this variable (alone) expected to favour the experimental intervention or the comparator? |
|  |  |  | Yes / No / No information | Favour experimental / Favour comparator / No information |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

\* In the context of a particular study, variables can be demonstrated not to be confounders and so not included in the analysis: (a) if they are not predictive of the outcome; (b) if they are not predictive of intervention; or (c) because adjustment makes no or minimal difference to the estimated effect of the primary parameter. Note that “no statistically significant association” is not the same as “not predictive”.

Preliminary consideration of co-interventions

Complete a row for each important co-intervention (i) listed in the review protocol; and (ii) relevant to the setting of this particular study, or which the study authors identified as important.

“Important” co-interventions are those for which, in the context of this study, adjustment is expected to lead to a clinically important change in the estimated effect of the intervention.

|  |  |  |
| --- | --- | --- |
| **(i) Co-interventions listed in the review protocol** | | |
| Co-intervention | Is there evidence that controlling for this co-intervention was unnecessary (e.g. because it was not administered)? | Is presence of this co-intervention likely to favour outcomes in the experimental intervention or the comparator |
|  |  | Favour experimental / Favour comparator / No information |
|  |  | Favour experimental / Favour comparator / No information |
|  |  | Favour experimental / Favour comparator / No information |
|  |  | Favour experimental / Favour comparator / No information |

|  |  |  |
| --- | --- | --- |
| **(ii) Additional co-interventions relevant to the setting of this particular study, or which the study authors identified as important** | | |
| Co-intervention | Is there evidence that controlling for this co-intervention was unnecessary (e.g. because it was not administered)? | Is presence of this co-intervention likely to favour outcomes in the experimental intervention or the comparator |
|  |  | Favour experimental / Favour comparator / No information |
|  |  | Favour experimental / Favour comparator / No information |
|  |  | Favour experimental / Favour comparator / No information |
|  |  | Favour experimental / Favour comparator / No information |

Risk of bias assessment

Responses underlined in green are potential markers for low risk of bias, and responses in red are potential markers for a risk of bias. Where questions relate only to sign posts to other questions, no formatting is used.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Signalling questions** | **Description** | **Response options** |
| **Bias due to confounding** | | | |
|  | 1.1 Is there potential for confounding of the effect of intervention in this study?  **If N/PN to 1.1:** the study can be considered to be at low risk of bias due to confounding and no further signalling questions need be considered |  | Y / PY / PN / N |
| **If Y/PY to 1.1**: determine whether there is a need to assess time-varying confounding: |  |  |
| 1.2. Was the analysis based on splitting participants’ follow up time according to intervention received?  **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6)  **If Y/PY**, go to question 1.3. |  | NA / Y / PY / PN / N / NI |
| 1.3. Were intervention discontinuations or switches likely to be related to factors that are prognostic for the outcome?  **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6)  **If Y/PY**, answer questions relating to both baseline and time-varying confounding (1.7 and 1.8) |  | NA / Y / PY / PN / N / NI |

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Questions relating to baseline confounding only** | | |
| 1.4. Did the authors use an appropriate analysis method that controlled for all the important confounding domains? |  | NA / Y / PY / PN / N / NI |
| 1.5. **If Y/PY to 1.4**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study? |  | NA / Y / PY / PN / N / NI |
| 1.6. Did the authors control for any post-intervention variables that could have been affected by the intervention? |  | NA / Y / PY / PN / N / NI |
|  | **Questions relating to baseline and time-varying confounding** | |  |
| 1.7. Did the authors use an appropriate analysis method that controlled for all the important confounding domains and for time-varying confounding? |  | NA / Y / PY / PN / N / NI |
| 1.8. **If Y/PY to 1.7**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study? |  | NA / Y / PY / PN / N / NI |
|  | **Risk of bias judgement** |  | Low / Moderate / Serious / Critical / NI |
| Optional: What is the predicted direction of bias due to confounding? |  | Favours experimental / Favours comparator / Unpredictable |

|  |  |  |  |
| --- | --- | --- | --- |
| **Bias in selection of participants into the study** | | | |
|  | 2.1. Was selection of participants into the study (or into the analysis) based on participant characteristics observed after the start of intervention?  **If N/PN to 2.1:** go to 2.4 |  | Y / PY / PN / N / NI |
| 2.2. **If Y/PY to 2.1**: Were the post-intervention variables that influenced selection likely to be associated with intervention?  2.3 **If Y/PY to 2.2**: Were the post-intervention variables that influenced selection likely to be influenced by the outcome or a cause of the outcome? |  | NA / Y / PY / PN / N / NI  NA / Y / PY / PN / N / NI |
| 2.4. Do start of follow-up and start of intervention coincide for most participants? |  | Y / PY / PN / N / NI |
| 2.5. **If Y/PY to 2.2 and 2.3, or N/PN to 2.4**: Were adjustment techniques used that are likely to correct for the presence of selection biases? |  | NA / Y / PY / PN / N / NI |
| **Risk of bias judgement** |  | Low / Moderate / Serious / Critical / NI |
| Optional: What is the predicted direction of bias due to selection of participants into the study? |  | Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable |

|  |  |  |  |
| --- | --- | --- | --- |
| **Bias in classification of interventions** | | | |
|  | 3.1 Were intervention groups clearly defined? |  | Y / PY / PN / N / NI |
| 3.2 Was the information used to define intervention groups recorded at the start of the intervention? |  | Y / PY / PN / N / NI |
| 3.3 Could classification of intervention status have been affected by knowledge of the outcome or risk of the outcome? |  | Y / PY / PN / N / NI |
| **Risk of bias judgement** |  | Low / Moderate / Serious / Critical / NI |
| Optional: What is the predicted direction of bias due to classification of interventions? |  | Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable |

|  |  |  |  |
| --- | --- | --- | --- |
| **Bias due to deviations from intended interventions** | | | |
|  | **If your aim for this study is to assess the effect of assignment to intervention, answer questions 4.1 and 4.2** | |  |
| 4.1. Were there deviations from the intended intervention beyond what would be expected in usual practice? |  | Y / PY / PN / N / NI |
| 4.2. **If Y/PY to 4.1**: Were these deviations from intended intervention unbalanced between groups *and* likely to have affected the outcome? |  | NA / Y / PY / PN / N / NI |
| **If your aim for this study is to assess the effect of starting and adhering to intervention, answer questions 4.3 to 4.6** | |  |
| 4.3. Were important co-interventions balanced across intervention groups? |  | Y / PY / PN / N / NI |
| 4.4. Was the intervention implemented successfully for most participants? |  | Y / PY / PN / N / NI |
| 4.5. Did study participants adhere to the assigned intervention regimen? |  | Y / PY / PN / N / NI |
| 4.6. **If N/PN to 4.3, 4.4 or 4.5**: Was an appropriate analysis used to estimate the effect of starting and adhering to the intervention? |  | NA / Y / PY / PN / N / NI |
| **Risk of bias judgement** |  | Low / Moderate / Serious / Critical / NI |
| Optional: What is the predicted direction of bias due to deviations from the intended interventions? |  | Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable |

|  |  |  |  |
| --- | --- | --- | --- |
| **Bias due to missing data** | | | |
|  | 5.1 Were outcome data available for all, or nearly all, participants? |  | Y / PY / PN / N / NI |
| 5.2 Were participants excluded due to missing data on intervention status? |  | Y / PY / PN / N / NI |
| 5.3 Were participants excluded due to missing data on other variables needed for the analysis? |  | Y / PY / PN / N / NI |
| 5.4 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Are the proportion of participants and reasons for missing data similar across interventions? |  | NA / Y / PY / PN / N / NI |
| 5.5 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Is there evidence that results were robust to the presence of missing data? |  | NA / Y / PY / PN / N / NI |
| **Risk of bias judgement** |  | Low / Moderate / Serious / Critical / NI |
| Optional: What is the predicted direction of bias due to missing data? |  | Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable |

|  |  |  |  |
| --- | --- | --- | --- |
| **Bias in measurement of outcomes** | | | |
|  | 6.1 Could the outcome measure have been influenced by knowledge of the intervention received? |  | Y / PY / PN / N / NI |
| 6.2 Were outcome assessors aware of the intervention received by study participants? |  | Y / PY / PN / N / NI |
| 6.3 Were the methods of outcome assessment comparable across intervention groups? |  | Y / PY / PN / N / NI |
| 6.4 Were any systematic errors in measurement of the outcome related to intervention received? |  | Y / PY / PN / N / NI |
| **Risk of bias judgement** |  | Low / Moderate / Serious / Critical / NI |
| Optional: What is the predicted direction of bias due to measurement of outcomes? |  | Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable |

|  |  |  |  |
| --- | --- | --- | --- |
| **Bias in selection of the reported result** | | | |
|  | Is the reported effect estimate likely to be selected, on the basis of the results, from... |  |  |
| 7.1. ... multiple outcome *measurements* within the outcome domain? |  | Y / PY / PN / N / NI |
| 7.2 ... multiple *analyses* of the intervention-outcome relationship? |  | Y / PY / PN / N / NI |
| 7.3 ... different *subgroups*? |  | Y / PY / PN / N / NI |
| **Risk of bias judgement** |  | Low / Moderate / Serious / Critical / NI |
| Optional: What is the predicted direction of bias due to selection of the reported result? |  | Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable |

|  |  |  |  |
| --- | --- | --- | --- |
| **Overall bias** | | | |
|  | **Risk of bias judgement** |  | Low / Moderate / Serious / Critical / NI |
| Optional: What is the overall predicted direction of bias for this outcome? |  | Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable |



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International Lic

## Annex 3: Individual-level variables

These are the variables that were collected on individual participants at the enrolment or follow-up visits.

Table 5 Variable List for RP2

|  |  |  |
| --- | --- | --- |
| Variable Name | Variable Abbreviation | Description |
| [Date](https://www.ebi.ac.uk/ols4/ontologies/ncit/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FNCIT_C25164) | date | The particular day, month and year an event has happened or will happen. |
| [Time](https://www.ebi.ac.uk/ols4/ontologies/ncit/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FNCIT_C25207) | time (local) | The continuum of experience in which events pass from the future through the present to the past. |
| [Country](https://www.ebi.ac.uk/ols4/ontologies/ncit/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FNCIT_C25464) | country | A collective generic term that refers here to a wide variety of dependencies, areas of special sovereignty, uninhabited islands, and other entities in addition to the traditional countries or independent states. |
| [Patient Identifier](https://www.ebi.ac.uk/ols4/ontologies/ncit/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FNCIT_C164337) | patient\_id | An alphanumeric identifier assigned to a specific patient. |
| [Age-Years](https://www.ebi.ac.uk/ols4/ontologies/ncit/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FNCIT_C37908) | age\_enrolment | The length of a person's life, stated in years since birth. |
| [Race](https://www.ebi.ac.uk/ols4/ontologies/ncit/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FNCIT_C17049) | race | A geographic ancestral origin category that is assigned to a population group based mainly on physical characteristics that are thought to be distinct and inherent. |
| [Sex](https://www.ebi.ac.uk/ols4/ontologies/ncit/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FNCIT_C28421) | sex | The assemblage of physical properties or qualities by which male is distinguished from female; the physical difference between male and female; the distinguishing peculiarity of male or female. |
| [Height](https://www.ebi.ac.uk/ols4/ontologies/ncit/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FNCIT_C25347) | height | The vertical measurement or distance from the base to the top of an object; the vertical dimension of extension. |
| [Body Weight](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C81328) | weight | The weight of a subject. |
| [Body Mass Index](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C16358) | bmi | An individual's weight in kilograms divided by the square of the height in meters. |
| [Location](https://www.ebi.ac.uk/ols4/ontologies/ncit/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FNCIT_C25341) | study\_location | A position, site, or point in space where something can be found. |
| [Address](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C25407) | participant\_address | A standardized representation of the location of a person, business, building, or organization. |
| [Housing Type](https://www.ebi.ac.uk/ols4/ontologies/ncit/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FNCIT_C90397) | housing\_type | The classification of a residential structure |
| [Number in household (observable entity)](https://www.ebi.ac.uk/ols4/ontologies/snomed/classes/http%253A%252F%252Fsnomed.info%252Fid%252F224525003) | num\_people\_household | Number in household |
| [exposure to air conditioning unit](https://www.ebi.ac.uk/ols4/ontologies/ecto/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FECTO_1000032) | aircon\_access | A exposure event involving the interaction of an exposure receptor to the condition of air conditioning unit. |
| [Income](https://www.ebi.ac.uk/ols4/ontologies/ncit/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FNCIT_C41150) | income | A gain or recurrent benefit during a period of time, usually measured in money that derives from capital or labor. |
| [Household Income](https://www.ebi.ac.uk/ols4/ontologies/ncit/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FNCIT_C70811) | house\_income | A demographic parameter indicating the amount of earnings made by a family. |
| [Substance Abuse](https://www.ebi.ac.uk/ols4/ontologies/ncit/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FNCIT_C18272) | substance\_abuse | Maladaptive pattern of drug or alcohol use that may lead to social, occupational, psychological, or physical problems. |
| [Smoking Status](https://www.ebi.ac.uk/ols4/ontologies/ncit/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FNCIT_C19796) | smoking\_status | An indication of a person's current tobacco and nicotine consumption as well as some indication of smoking history. |
| [Alcohol Abuse](https://www.ebi.ac.uk/ols4/ontologies/ncit/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FNCIT_C20701) | alcohol\_status | The use of alcoholic beverages to excess, either on individual occasions ("binge drinking") or as a regular practice. |
| [Employment Status](https://www.ebi.ac.uk/ols4/ontologies/ncit/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FNCIT_C179143) | employment | The state of a person with regard to earning wages or salary. |
| [Education Level](https://www.ebi.ac.uk/ols4/ontologies/ncit/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FNCIT_C17953) | education\_years | An indication of the years of schooling completed in graded public, private, or parochial schools, and in colleges, universities, or professional schools. |
| [Lost To Follow-Up](https://www.ebi.ac.uk/ols4/ontologies/ncit/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FNCIT_C48227) | loss\_to\_follow\_up | The subject was not available for follow-up procedures. |
| [Hypertension](https://www.ebi.ac.uk/ols4/ontologies/ncit/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FNCIT_C3117) | hypertension | Blood pressure that is abnormally high. |
| [Diabetes Mellitus](https://www.ebi.ac.uk/ols4/ontologies/ncit/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FNCIT_C2985) | dm | A metabolic disorder characterized by abnormally high blood sugar levels due to diminished production of insulin or insulin resistance/desensitization. |
| [Number of Fetuses](https://www.ebi.ac.uk/ols4/ontologies/ncit/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FNCIT_C124626) | number\_of\_foetuses | A measurement of the total number of fetuses, which includes alive and dead fetuses, present in the uterus. |
| [Death Indicator](https://www.ebi.ac.uk/ols4/ontologies/ncit/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FNCIT_C93546) | death | Death is the permanent cessation of all biological functions that sustain a living, physical organism. |
| [Hospitalization](https://www.ebi.ac.uk/ols4/ontologies/scdo/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FSCDO_0000573) | hospital\_admission | The confinement of a patient in a hospital. |
| [Gastroenteritis](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C34632) | gastroenteritis | An inflammatory disorder that affects the upper and lower gastrointestinal tract. Most commonly, this is attributed to viruses; however bacteria, parasites or adverse reactions can also be the culprit. Symptoms include acute diarrhea and vomiting. |
| [Pneumonia](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C3333) | pneumonia | An acute, acute and chronic, or chronic inflammation focally or diffusely affecting the lung parenchyma, due to infections (viruses, fungi, mycoplasma, or bacteria), treatment (e.g. radiation), or exposure (inhalation) to chemicals. Symptoms include cough, shortness of breath, fevers, chills, chest pain, headache, sweating, and weakness.… |
| [Urinary Tract Infection](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C50791) | urinary\_tract\_infection | A bacterial infectious process affecting any part of the urinary tract, most commonly the bladder and the urethra. Symptoms include urinary urgency and frequency, burning sensation during urination, lower abdominal discomfort, and cloudy urine. |
| [Group B Streptococcus Infection](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C87168) | groupb\_strep\_infection | Group B Streptococcus, also known as Streptococcus agalactiae, colonizes the vaginal and gastrointestinal tracts of up to 45% of healthy women and may infect neonates in utero or during delivery, causing neonatal sepsis in 1-2% of colonized neonates. GBS infection may also occur in nonpregnant (particularly elderly) adults with underlying medical conditions, presenting as urinary tract infection, pneumonia, or soft-tissue infection. [ NCI ] |
| [Syphilis](https://www.ebi.ac.uk/ols4/ontologies/ncit/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FNCIT_C35055) | syphilis\_status | A contagious bacterial infection caused by the spirochete Treponema pallidum. It is a sexually transmitted disorder, although it can also be transmitted from the mother to the fetus in utero. Typically, it is initially manifested with a single sore which heals without treatment. If the infection is left untreated, the initial stage is followed by skin rash and mucous membrane lesions. A late stage follows, which is characterized by damage of the internal organs, including the nervous system. |
| [HIV Status](https://www.ebi.ac.uk/ols4/ontologies/ncit/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FNCIT_C157155) | hiv\_status | The result of testing to determine if an individual is infected with the human immunodeficiency virus. |
| [Hepatitis B Infection](https://www.ebi.ac.uk/ols4/ontologies/ncit/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FNCIT_C3097) | HepB\_status | A viral infection caused by the hepatitis B virus. |
| [Schistosomiasis](https://www.ebi.ac.uk/ols4/ontologies/ncit/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FNCIT_C35000) | schistosomiasis\_status | A parasitic infection caused by flukes of the genus Schistosoma. Signs and symptoms include fever, abdominal pain, eosinophilia and hepatosplenomegaly. If left untreated it may eventually cause liver damage leading to cirrhosis, bladder cancer and kidney failure. |
| [Tuberculosis](https://www.ebi.ac.uk/ols4/ontologies/ncit/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FNCIT_C3423) | tb\_status | A chronic, recurrent infection caused by the bacterium Mycobacterium tuberculosis. Tuberculosis (TB) may affect almost any tissue or organ of the body with the lungs being the most common site of infection. The clinical stages of TB are primary or initial infection, latent or dormant infection, and recrudescent or adult-type TB. Ninety to 95% of primary TB infections may go unrecognized. Histopathologically, tissue lesions consist of granulomas which usually undergo central caseation necrosis. Local symptoms of TB vary according to the part affected; acute symptoms include hectic fever, sweats, and emaciation; serious complications include granulomatous erosion of pulmonary bronchi associated with hemoptysis. If untreated, progressive TB may be associated with a high degree of mortality. This infection is frequently observed in immunocompromised individuals with AIDS or a history of illicit IV drug use. |
| [Hemoglobin](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C16676) | haemoglobin | The red respiratory protein of erythrocytes, consisting of approximately 3.8% heme and 96.2% globin (64.5 KD), which as oxyhemoglobin (HbO2) transports oxygen from the lungs to the tissues where the oxygen is readily released and HbO2 becomes Hb. |
| [Creatinine](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C399) | creat | The breakdown product of creatine, a constituent of muscle tissue, that is excreted by the kidney and whose serum level is used to evaluate kidney function. |
| [Creatinine Clearance](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C25747) | creat\_clearance | The determination of the clearance of endogenous creatinine, used for evaluating the glomerular filtration rate. |
| [HIV Viral Load Measurement](https://www.ebi.ac.uk/ols4/ontologies/ncit/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FNCIT_C92544) | hiv\_vl | The determination of the HIV viral load in a specimen. |
| [CD4 Expressing Cell Count](https://www.ebi.ac.uk/ols4/ontologies/ncit/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FNCIT_C103810) | cd4 | The determination of the amount of the CD4 expressing cells in a sample. |
| [Platelet Count](https://www.ebi.ac.uk/ols4/ontologies/ncit/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FNCIT_C51951) | plt | The determination of the number of platelets in a biospecimen. |
| [Aspartate Aminotransferase Measurement](https://www.ebi.ac.uk/ols4/ontologies/ncit/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FNCIT_C64467) | ast | A quantitative measurement of aspartate aminotransferase present in a sample. |
| [Alanine Aminotransferase Measurement](https://www.ebi.ac.uk/ols4/ontologies/ncit/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FNCIT_C64433) | alt | A quantitative measurement of alanine aminotransferase present in a sample. |
| [Protein to Creatinine Ratio Measurement](https://www.ebi.ac.uk/ols4/ontologies/ncit/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FNCIT_C79463) | urine\_pcr | The determination of the ratio of total protein compared to creatinine present in a sample. The measurement may be expressed as a ratio or percentage. |
| [Alkaline Phosphatase Measurement](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C64432) | alp | A quantitative measurement of alkaline phosphatase present in a sample. |
| [mean corpuscular volume](https://www.ebi.ac.uk/ols/ontologies/efo/terms?iri=http%3A%2F%2Fwww.ebi.ac.uk%2Fefo%2FEFO_0004526) | mcv | A mean corpuscular volume is the result of calculation of the mean volume of erythrocytes in a blood sample. |
| [mean corpuscular hemoglobin concentration](https://www.ebi.ac.uk/ols/ontologies/efo/terms?iri=http%3A%2F%2Fwww.ebi.ac.uk%2Fefo%2FEFO_0004528) | mchc | The mean corpuscular hemoglobin concentration is a measure of the concentration of hemoglobin in a given volume of packed red blood cell |
| [Viral Resistance Domain](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C106581) | hiv\_resistance\_mutations | A findings domain that captures information regarding the genetics of viral drug resistance. It contains the reference sequence used to validate the observed genetic mutation of interest. |
| [Drug resistance to antiretroviral therapy (disorder)](https://www.ebi.ac.uk/ols/ontologies/snomed/terms?iri=http%3A%2F%2Fsnomed.info%2Fid%2F425581000) | hiv\_drug\_resistance | A binary variable describing the presence or abscence of any drug resistance to antiretroviral therapy. |
| [Highly Active Antiretroviral Therapy](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C16165) | art\_regimen | Drug therapy which targets retrovirus function by multiple mechanisms. |
| [Noncompliance with antiretroviral medication regimen (finding)](https://www.ebi.ac.uk/ols/ontologies/snomed/terms?iri=http%3A%2F%2Fsnomed.info%2Fid%2F713017009) | art\_adherance | Noncompliance with antiretroviral therapy |
| [Study Subject Radiography Report](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C115514) | xray\_findings | Records pertaining to the findings from a study subject's radiographic images. |
| [Oligohydramnios](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C92839) | oligohydramnios | A lower than normal quantity of amniotic fluid in the amniotic sac as compared to normal values. Typically associated with an amniotic fluid index (AFI) of less than 5 cm or a single maximum vertical pocket (MVP) of less than 2 cm. |
| [thyroid stimulating hormone measurement](https://www.ebi.ac.uk/ols/ontologies/efo/terms?iri=http%3A%2F%2Fwww.ebi.ac.uk%2Fefo%2FEFO_0004748) | tsh | Is a quantification of thyroid-stimulating hormone, a glycoprotein and hormone secreted from the pituitary which regulates the thryoid. |
| [Lower Respiratory Tract Infection](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C35158) | lrti | An acute or chronic, viral or bacterial infectious process that affects the lower respiratory tract. |
| [Upper Respiratory Tract Infection](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C35650) | urti | An infectious process affecting the upper respiratory tract (nose, paranasal sinuses, pharynx, larynx, or trachea). Symptoms include congestion, sneezing, coughing, fever, and sore throat. |
| [Direct Bilirubin Measurement](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C61024) | direct\_bili | The bilirubin is bound to glucuronide to form conjugated bilirubin (direct bilirubin). Direct Bilirubin measurement is accomplished by a colorimetric method. Direct Bilirubin in biological fluids reacts with sulfanilic acid at acidic pH to produce a red colored complex. The optical density of produced color has a direct relationship with Direct Bilirubin concentration in the solution. |
| [Indirect Bilirubin Measurement](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C64483) | indirect\_bili | Bilirubin is in the insoluble form, unconjugated bilirubin (indirect bilirubin).The non-water soluble, free bilirubin does not react with sulfanilic acid at acidic pH to produce a red colored complex until an accelearator, alcohol, is added to the solution to perform a quantitative measurement of unconjugated bilirubin levels. |
| [Amylase Measurement](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C64434) | amylase | A quantitative measurement of amylase present in a sample. |
| [Lipase Measurement](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C117748) | lipase | The determination of the amount of lipase present in a sample. |
| [Cholesterol Measurement](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C105586) | cholesterol | The determination of the amount of total cholesterol present in a sample. |
| [mean corpuscular hemoglobin](https://www.ebi.ac.uk/ols/ontologies/efo/terms?iri=http%3A%2F%2Fwww.ebi.ac.uk%2Fefo%2FEFO_0004527) | mch | The MCH is the average mass of hemoglobin per red blood cell in a sample of blood and is calculated by dividing the total mass of hemoglobin by the RBC count |
| [Low Density Lipoprotein Cholesterol Measurement](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C105588) | ldlc | The determination of the amount of low-density lipoprotein cholesterol present in a sample. |
| [High Density Lipoprotein Cholesterol Measurement](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C105587) | hdlc | The determination of the amount of high-density lipoprotein cholesterol present in a sample. |
| [Glycosylated Hemoglobin Measurement](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C64849) | hba1c | A quantitative measurement of the amount of glycosylated hemoglobin present in a sample of blood. |
| [Albumin Measurement](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C64431) | albumin | A quantitative measurement of albumin present in a sample. |
| [Cortisol Measurement](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C74781) | cortisol | The determination of the amount of cortisol present in a sample. |
| [Whole Parathyroid Hormone Measurement](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C103451) | pth | The determination of the amount of the whole parathyroid hormone (consisting of amino acids 1-84) in a sample. |
| [Free Thyroxine Measurement](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C74786) | ft4 | The determination of the amount of free thyroxine present in a sample. |
| [Free Triiodothyronine Measurement](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C74787) | ft3 | The determination of the amount of free triiodothyronine present in a sample. |
| [Blood Urea Nitrogen Measurement](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C61019) | urea | A quantitative measurement of the amount of urea nitrogen present in a serum sample. |
| [Calcium Measurement](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C64488) | calcium | A quantitative measurement of the amount of calcium present in a sample. |
| [Bone Mineral Density Z-Score](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C139217) | dexa\_scan | A statistical score representing the number of standard deviations above or below what is expected for an individual's bone density based on his age, sex, weight, and race. Z-scores are most useful in evaluating low bone density in children, premenopausal women, and men younger than age fifty. |
| [Serum Uric Acid Measurement](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C61034) | uric\_acid | A quantitative measurement of the amount of uric acid present in a sample of serum. |
| [Potassium Measurement](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C64853) | potassium | A quantitative measurement of the amount of potassium present in a sample. |
| [Sodium Measurement](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C64809) | sodium | A quantitative measurement of the amount of sodium present in a sample. |
| [COVID-19 RT-PCR assay](https://www.ebi.ac.uk/ols/ontologies/cido/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FCIDO_0000019) | covid\_19\_pcr | Result of a COVID-19 PCR based assay. |
| [C-Reactive Protein Measurement](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C64548) | crp | A quantitative measurement of the amount of C-reactive protein present in a sample. |
| [Procalcitonin Measurement](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C103430) | pct | The determination of the amount of the procalcitonin in a sample. |
| [interleukin-6 measurement](https://www.ebi.ac.uk/ols4/ontologies/efo/classes/http%253A%252F%252Fwww.ebi.ac.uk%252Fefo%252FEFO_0004810) | il-6 | Is a quantification of interleukin-6, a pro-inflammatory and anti-inflammatory cytokine. |
| [International Normalized Ratio](https://www.ebi.ac.uk/ols4/ontologies/ncit/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FNCIT_C25352) | inr | A system, commonly called the INR, established by the World Health Organization (WHO) and the International Committee on Thrombosis and Hemostasis for reporting the results of blood coagulation (clotting) tests. All results are standardized using the international sensitivity index (ISI) for the particular thromboplastin reagent and instrument combination utilized to perform the test; the ratio of a patient's clotting time to the lab's mean reference value is normalized against the ISI. (from medterms.com and medicine.ucsf.edu) |
| [D-Dimer Measurement](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C82621) | d\_dimers | The determination of the amount of d-dimers present in a sample. |
| [Unipolar Depression](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C35094) | depression | A mood disorder having a clinical course involving one or more episodes of serious psychological depression that last two or more weeks each, do not have intervening episodes of mania or hypomania, and are characterized by a loss of interest or pleasure in almost all activities and by some or all of disturbances of appetite, sleep, or psychomotor functioning, a decrease in energy, difficulties in thinking or making decisions, loss of self-esteem or feelings of guilt, and suicidal thoughts or attempts. |
| [Generalized Anxiety Disorder](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C92622) | gen\_anxiety\_disorder | An anxiety disorder characterized by excessive and difficult-to-control worry about a number of life situations. The worry is accompanied by restlessness, fatigue, inability to concentrate, irritability, muscle tension, and/or sleep disturbance and lasts for at least 6 months. |
| [Adverse Event](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C41331) | adverse\_event | Any unfavorable or unintended disease, sign, or symptom (including an abnormal laboratory finding) that is temporally associated with the use of a medical treatment or procedure, and that may or may not be considered related to the medical treatment or procedure. Such events can be related to the intervention, dose, route of administration, patient, or caused by an interaction with another drug(s) or procedure(s) |
| [Heart Rate](https://www.ebi.ac.uk/ols4/ontologies/ncit/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FNCIT_C49677) | heart\_rate | The number of heartbeats per unit of time, usually expressed as beats per minute. |
| [Systolic Blood Pressure](https://www.ebi.ac.uk/ols4/ontologies/ncit/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FNCIT_C25298) | systolic\_blood\_pressure | The maximum pressure exerted into the systemic arterial circulation during the contraction of the left ventricle of the heart. |
| [Diastolic Blood Pressure](https://www.ebi.ac.uk/ols4/ontologies/ncit/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FNCIT_C25299) | diastolic\_blood\_pressure | The minimum pressure exerted into the systemic arterial circulation during cardiac ventricular relaxation and filling. |
| [Mean Arterial Pressure](https://www.ebi.ac.uk/ols4/ontologies/ncit/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FNCIT_C49679) | mean\_arterial\_pressure | The mean pressure of the blood within the arterial circulation. The arterial pressure may be directly measured by insertion of an intra-arterial catheter connected to a transducer. The mean arterial pressure (MAP) can be calculated by subsequent analysis of the waveform. MAP can be approximated without an invasive procedure using the following formula: diastolic pressure plus 1/3 of the pulse pressure, where pulse pressure is systolic pressure - diastolic pressure. |
| [Oxygen Saturation Measurement](https://www.ebi.ac.uk/ols4/ontologies/ncit/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FNCIT_C60832) | oxygen\_saturation | The measurement of the ratio of oxygenated hemoglobin to total hemoglobin in the blood |
| [Myocardial Infarction](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C27996) | myocardial\_infarct | Gross necrosis of the myocardium, as a result of interruption of the blood supply to the area, as in coronary thrombosis. |

## Annex 4: Land Surface Temperature retrieval workflow

Diagram

Description automatically generated

Figure 6 Land Surface Temperature retrieval workflow

## Annex 5: Example Impact Assessment Questions and Indicators

|  |  |  |
| --- | --- | --- |
|  | Evaluation questions | Potential indicators |
| Alert protocol | Were the alerts issued efficiently? | Frequency of partner notification and public alerts  Timeliness of alert information received  Timeliness and efficiency of message delivery to the end-users |
| Are extreme heat events forecasted and monitored accurately? | Quality of surveillance data  Frequency of warnings and alerts issued in relation to actual weather conditions occurring  Capacity of participating agencies to monitor and deliver surveillance and weather data |
| Response plan | How involved were stakeholders with implementing response measures? | Level of participation of agencies and other community groups in education activities, issuing warnings and responding to an alert  Number and types of response measures delivered by stakeholders |
| Did stakeholders follow the response plan and find it helpful? | Number and diversity of engaged stakeholders and meeting frequency Perceived importance of the heat response among stakeholders Partners’ views on the degree of coordination of activities Stakeholders’ views on the adequacy of support offered  Level of stakeholder satisfaction |
| Are response measures being used by the public (e.g. cooling centres)? | Number of at-risk people who took preventive actions  Number of people, their demographic makeup, and length of visits to cooling facilities  Number of people and their demographic makeup who took advantage of other response measures |
| Communication plan | Were key messages and services provided to the public? | Number of planned communication elements delivered  Vulnerable and general populations reached by each communication element  Number and types of inquiries received  Number and types of resources distributed  Promotion and publicity received through media activities |
| Was the target population aware of heat-health EWS and its key messages? | Number of media and information sources engaged as part of the outreach campaign  Penetration of key messages into the media  Accessibility of information to the public |
| Did the target population understand and follow key messages? | Number of at-risk people who perceive extreme heat to be a health risk  Number of at-risk people who can identify preventive measures Capacity of targeted population to recall accurate messaging |
| Operational costs | What resources are used to operate the heat-health EWS? | Resources used by each partner  Resources required for collecting and monitoring surveillance data  Staff time spent on the program at various stages  Costs to communicate messages to stakeholders and the public  Costs of maintaining the system |

**Source: Adapted from [84]**

## Graphical user interface, application, Word Description automatically generatedAnnex 7: Overall Data Management Flow