# ACTG Data Analysis Concept Sheet (DACS)

# INSTRUCTIONS AND Proposal format

Instructions

*General*

Per SOP ACTG-106, a DACS is a proposal that describes a research project and requests data from one or more existing ACTG studies (including ACTG protocols, Data Requests [DRs] and New Works Concept Sheet [NWCS]) for use in completing that research project. Unless the ACTG has designated a protocol as concluded, the objectives of the DACS should not overlap with objectives specified in the ACTG protocol(s) or defined as a secondary analyses for the ACTG protocol(s). The objectives should also not overlap with objectives specified in an ACTG-approved but not yet concluded DR or NWCS. Requests for data to be used in analyses to be undertaken by, and without the financial support of, the Statistical and Data Analysis Center (SDAC) should be submitted using the Data Request (DR) proposal format. Refer to SOP ACTG-106 and ACTG-158 for more information about the DACS and DR processes, respectively.

The DACS proposal must be no longer than 5 pages and in Arial 11-point font (preferably in Microsoft Word), and must be submitted electronically via the ACTG Proposal Submission System at <https://submit.mis.s-3.net>/. All essential materials for review must be included within the page limit. DACS reviewers may or may not choose to review supporting materials (e.g., published articles or data) attached as appendices.

*Language Preference*

The term “participant” is to be used instead of “subject” or “patient” in all network studies, including DACSs.

*Letters of Support for NIH Grants*

If this DACS will be submitted for consideration for an NIH grant and a letter of support (LOS) from the network is required, please submit a LOS request at least 6 weeks prior to the grant application deadline. Submit your request and accompanying documents to the ACTG NCC Leadership Group to [LCSG@dlhcorp.com](mailto:LCSG@dlhcorp.com). Refer to SOP ACTG-152, ACTG-Related Grant Applications and Letters of Support, for the documents required to be submitted along with the request.

Proposal Format

STUDY TITLE:

Descriptive title of the proposed analysis.

The HE2AT Center Research Project 2: Advancing Understanding of Heat-Health Interactions in Large African Cities and Developing Locally Relevant and Risk-Stratified Early Warning Systems

VERSION NUMBER:

Indicate whether this is the original submission (version 1) or a revision (version 2, etc.).

Version 1

SUMMARY OF MAJOR REVISIONS:

If the proposal is a revision (version 2 or higher), include a brief narrative of the major changes from the prior version (e.g., new relevant studies, new or modified objectives, statistical resources, funding, etc.). Any new language in the subsequent sections should be in bold text to facilitate review by the ACTG.

NA

PROPOSING STUDY CHAIR, VICE CHAIR(S), INVESTIGATOR(S) AND INSTITUTION(S):

Identify at least one investigator to serve as study chair. Indicate if the study chair is an ACTG investigator of a non-ACTG investigator. A vice chair(s) may also be identified, if desired. Provide the name, title, institution/ACTG CRS, address, telephone, and e-mail address of the proposing investigator(s).

Study Chair: Prof. MF Chersich (a non-ACTG Investigator). Wits RHI, University of the Witwatersrand, South Africa. Address: Esselen Street Hillbrow, Johannesburg, 2000, South Africa. Phone: +27727521123. Email: [mchersich@wrhi.ac.za](mailto:mchersich@wrhi.ac.za)

Vice Chair: Prof. Lee Fairlie (ACTG Investigator). Wits RHI, University of the Witwatersrand, South Africa. Address: Esselen Street Hillbrow, Johannesburg, 2000, South Africa. Phone: +27727521123. Email: LFairlie@wrhi.ac.za

Investigators:

Dr Gloria Maimela. (a non-ACTG Investigator). Wits RHI, University of the Witwatersrand, South Africa. Address: Esselen Street Hillbrow, Johannesburg, 2000, South Africa. Phone: +27727521123. Email: GMaimela@wrhi.ac.za

Mr Craig Parker. (a non-ACTG Investigator). Wits RHI, University of the Witwatersrand, South Africa. Address: Esselen Street Hillbrow, Johannesburg, 2000, South Africa. Phone: +27727521123. Email: cparker@wrhi.ac.za>

Collaborating Statistical and Data Analysis Center (SDAC) Statistician:

Name, telephone number, and e-mail address of the collaborating SDAC statistician in those situations in which SDAC will be performing the analysis and/or will have substantial scientific involvement.

NA

COORDINATING SDAC Statistician (this may be blank at initial submission):

SDAC will provide the name, telephone number, and e-mail address of an SDAC statistician who will coordinate with relevant parties regarding the approved DACS, if one is not among the proposing team members.

NA

STUDY RATIONALE:

Complete, concise discussion of the analysis rationale, including sufficient background information (e.g., data, references) to support the scientific merits and importance of the analysis.

Global temperatures have already risen about 1.2C, and the world is on track for an increase of 1.5C within the next decades. People living in large and rapidly growing African cities face significant health risks from observed past and projected future temperature increases. Many such cities constitute ‘Urban Heat Islands’, where concrete or asphalt surfaces absorb and retain heat, and ‘cooling’ areas, such as parks are limited. Data and understanding on heat-health outcomes, exposure, vulnerability, and potential solutions in African urban contexts are a major public health priority. In particular, additional research is necessary to understand patterns of heat impacts in detail, and to define temperature sensitivity thresholds for Early Warning Systems according to the specific vulnerabilities of the local population and setting, and microclimates within different parts of a city.

Understanding heat-health linkages allows one to identify high-risk groups and settings, as well as to document the burden of heat-related conditions. Quantifying current heat burdens and projecting future burdens can inform resource prioritisation and allocation. Additionally, understanding patters of heat impacts helps to develop personalised Early Warning Systems. These Warning Systems aim to capture unique geospatial and individualized heat risk patterns in order to warning individuals at high risk during periods of high temperatures, but also to assist key stakeholders, including the public, government, workplaces, and sporting organizations, in preparing for heat waves or brief periods of extreme heat.

The study forms part of the HEat and HEalth in Africa Transdisciplinary Center (HEAT Center; <https://heatcenter.wrhi.ac.za/>). The HEAT Center is one of the Research Hubs in the NIH DS-I Africa initiative (<https://dsi-africa.org>), which is a major 5-10 year initiative across Africa, aiming to maximise use of existing data on the continent and find actionable solutions to the most pressing health issues facing people living in Africa, especially vulnerable populations.

This research was made possible by support from the Fogarty International Center, the National Institute of Environmental Health Sciences (NIEHS), and OD/Office of Strategic Coordination (OSC) of the National Institutes of Health, under Award Number U54TW012083. Although this study was supported by the NIH, the content and responsibility of the authors are not necessarily representative of the official views of the National Institutes of Health.

STUDY OBJECTIVE(S):

A clear and thorough description of all study objectives of the analysis.

The study aims to analyze heat-health vulnerability and exposure in Johannesburg, South Africa. The objectives are as follows:

1. Map heat vulnerability and exposure across urban areas in African cities using a combination of health, socioeconomic, geospatial climate and satellite imagery data.
2. Develop a heat-health outcome forecast model using statistical, machine learning and deep learning techniques to predict the probability of adverse health outcomes at different temperature thresholds, stratified by geography and demographics.
3. Investigate the impact of socio-economics and demographics such as housing types and density, commuting distances, and working conditions on heat exposure and heat-health vulnerability.
4. Determine the most dangerous types of heat exposure for people living in different conditions in African cities, including night-time temperatures, daily maximums, extremes, or long-term accumulated heat burden.
5. Create an app-based Heat-Health Early Warning System to provide timely warnings to city planners, public health officials, and community leaders. The system will reflect the unique risk patterns identified through mapping and forecasting.

STUDY DESIGN:

In summary, the study involves applying data science techniques to document the impacts of heat exposure on health outcomes using geospatially and temporally linked data. We apply data science analyses to document the impact of heat-related health hazards in Johannesburg, South Africa. Identical research activities are taking place in Abidjan, Côte d'Ivoire, but are not part of this data request. The study has received approval from the Human Research Ethics Committee of the Faculty of Health Sciences, University of the Witwatersrand HREC approval reference number 220606: HEAT002).

We draw on satellite image analysis, socio-economic data, and open mapping data. These geospatial data will be brought together with geo-located health data to identify the size and patterns of temperature impacts on health. Dates of health events are linked temporally with the prevailing weather and other environmental exposure on that date. Information of the geolocation of participants will be obtained directly from the site investigators. Depending on what information is available on geolocation, we will use the location of the research site where activities took place, or household addresses, if available, such as from patient tracking. We have several means of safeguarding and anonymizing household addresses and other potentially identifiable data.

The study analysis involves deploying a range of machine learning methods to construct an index of intra-urban socio-economic and environmental vulnerability factors (e.g. housing types, formal versus informal areas, green compared to built-up areas, population mobility, commuting conditions, and distance from health services).

The HE2AT Center will then utilize a vulnerability-heat-hazard model to provide operational heat health risk forecasts for daily and seasonal time frames, differentiated by susceptible populations as determined through demographics and specific geographic locations. The forecasts will trigger early warnings when certain risk thresholds are reached. The thresholds for when warnings are sent will be selected based on an analysis of heat health outcomes in this study, inputs from experts in the consortium, and external specialists in physiology and clinical medicine.

In this paragraph we describe the analysis methods in more detail. Data analysis methods include natural language processing, geospatial analysis, and predictive analytics to identify individuals at high risk for heat-related conditions. The statistical components of the analysis include applying existing image processing techniques to satellite images, combining data from high-resolution climate models and meteorological station observations, superimposing health outcomes on vulnerability-hazard maps, and using traditional statistical and machine learning methods, including deep learning techniques, to develop heat-health outcome forecast models. The study will also evaluate the efficacy and accuracy of different machine learning techniques, such as recurrent neural networks (RNNs), long short-term memories (LSTMs), gated recurrent units (GRUs), Multi-Layer Perceptron (MLP), Bayesian Neural Network (BNN), Radial Basis Functions (RBF), K-Nearest Neighbor regression (KNN), and Gaussian Processes (GRU), to build the best possible model for predicting the health effects of extreme heat.

In conclusion, the study design and analysis methods aim to provide a comprehensive understanding of heat-related health hazards in Johannesburg, South Africa to develop an Early Warning System to prepare the public, government, and stakeholders for heatwaves or brief periods of extreme heat.

STUDY DURATION:

An estimated timeline to complete the new work.

The study duration is anticipated to be 3 years (the HEAT Center is a 5 year project, with the project currently in its second year).

A picture containing text

Description automatically generated

RELEVANT ACTG STUDIES:

List of all ACTG protocols, DACSs, DRs, and/or NWCSs pertinent to the research questions.

Please note that we only require data from the Johannesburg study sites of the below listed studies. All study activities described above take place in Johannesburg only.

* Pilot Study On Effect Of Antiretroviral Therapy (Art) On Hepatitis B Viral (Hbv) Infection (AACTG.51.PEPFAR.01)
* Effect of Pitavastatin on Kidney Function in HIV-Infected Persons (CHRU - ACTG - 5R01DK108438-04)
* AIDS Clinical Trials Group for Research on Therapeutics for HIV and Related Infections [ACTG LOC: PF Chair A5356] (CHRU - UCLA - NIH - 2UM1AI068636-15)
* KARABELO Clinical Trials Unit for NIAID Networks: Soweto Clinical Trials Unit - ACTG CRS, HVTN CRS & IMPAACT CRS, MEDUNSA CRS, HPTN CRS, KARABELO (NIH - U01AI 069453-13 - PHRU)
* Soweto Clinical Trials Unit - ACTG CRS, HVTN CRS & IMPAACT CRS, MEDUNSA CRS, HPTN CRS, KARABELO (NIH - U01AI 069453-01 - PHRU)
* Leadership and Operations Center(LOC), AIDS Clinical Trials Group(ACTG);LOC 1/[Protocol Funds Mentorship] (PHRU - UCLA - UMA1AI068636 - 13R)
* PHRU - Setshaba Clinical Trials Unit - Clinical Trials Unit for NIAID Networks: Soweto Clinical Trials Unit - ACTG CRS, HVTN CRS & IMPAACT CRS, MEDUNSA CRS, HPTN CRS, KARABELO (NIH - U01AI 069453-15 - PHRU)
* ACTG PROTOCOL FUNDS SUBAWARD AGREEMENT - AIDS Clinical Trials Group for Research on Therapeutics for HIV and Related Infections (PHRU - UCLA - NIH - 2UM1AI068636-15)

DATA MANAGEMENT AND DATA ANALYSIS:

Responsible party for both data management and analysis (SDAC, drug company, institution), and specific variables and associated case report forms (CRFs) required for the analysis.

A dedicated team in the HEAT Center will oversee the management of the databases and ensure data security and protection. We will store and curate the data, with a machine-learning-based data ingestion and curation engine, a massive distributed compute and data store, an analytics and data platform, and an interface to interact with the system.

Data security and storage are top priorities for the project, and all data will be stored in secure databases hosted at different institutions. Access to the data will be controlled through strict security protocols, and all personal and sensitive information will be de-identified and treated in accordance with relevant privacy laws and regulations.

The HE2AT Center project (HREC approval reference number 220606: HEAT002) is focused on creating a comprehensive database for clinical trial and cohort studies by gathering data from various sources. The variables thus include biomedical data, demographic data, and clinical data such as laboratory results. The aim is to analyze this information to understand the relationship between human health and wellness and the effects of climate and weather conditions. As with all data science projects, the richer and more

The following table outlines the categories of variables that will be analyzed in the project:

| **Category** | **Variables** |
| --- | --- |
| Biomedical Data | Medical History, Laboratory Results |
| Demographic Data | Age, sex, Socio-economic Status, Education Level |
| Climate Data | Heat, Humidity, Satellite imagery |
| Geographical Information | Address |

In conclusion, the HE2AT Center project places a high emphasis on data security, storage, and privacy, and the DMAC team will play a critical role in ensuring that these priorities are met at all times. Data integration will involve transforming the data into standardized datasets and integrating data from different domains to enable robust analysis.

POSTING OF GWAS DATA TO THE NIH DATABASE OF GENOTYPES AND PHENOTYPES (dbGaP):

Any DACS involving genome-wide association studies (GWAS) data must contain a section addressing NIH Policy NOT-OD-07-088 and related policies, including a statement in the DACS that ACTG GWAS will not be posted to dbGaP or a similar repository by the DACS investigator. Posting will have been done by the ACTG or the investigator who generated the GWAS data with non-ACTG funding, if required and consistent with the signed informed consent forms under which the specimens were obtained.

NA

RESOURCES:

Estimated SDAC statistical, programming, and data management time required to provide the data and, when relevant, conduct the analysis. Where relevant, identify time required at ACTG sites (e.g., for abstracting site-level data).

We do not require SDAC statistical, or programming time for conducting the analyses.The analysis will be done by the relevant HEAT Center data science staff who are employed at Wits RHI, the University of the Witwatersrand, South Africa, IBM, and the University of Cape Town, South Africa.

The HEAT Center staff are based in Johannesburg at the University of the Witwatersrand. This is the sites where the studies were done, all of which fall within the Wits Health Consortium. If required, we will cover any costs related to abstracting site-level data if this requires activities at the research sites.

An approximate estimate for a scenario with moderate-sized and moderately complex data sets, with a high level of data quality required, could be several weeks to a few months.

We have resources in the project to cover the SDAC data management expenses if required.

The estimated time required for SDAC statistical, programming, and data management to provide the 5-10 clinical trial datasets would depend on several factors such as:

1. The size of the data sets: If the data sets are large, it will take more time to curate and transfer them.
2. The complexity of the data: If the data are complex, with multiple variables and nested data structures, it may take more time to curate and transfer it.
3. The required data quality checks: If a high level of data quality is required, this will add to the time required for data management.
4. The tools and techniques used: The time required may also depend on the tools and techniques used for data management and transfer.

SIGNATURES:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Proposing Investigator DATE

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

ACTG CRS Leader DATE

(required only if the Proposing Investigator is

receiving funding from an ACTG CTU/CRS)