**From:** BMJ Open <[onbehalfof@manuscriptcentral.com](mailto:onbehalfof@manuscriptcentral.com)>   
**Sent:** Wednesday, February 21, 2024 12:32 PM  
**To:** Craig Parker <[cparker@wrhi.ac.za](mailto:cparker@wrhi.ac.za)>; [craigparker6@gmail.com](mailto:craigparker6@gmail.com)  
**Subject:** BMJ Open - Decision on Manuscript ID bmjopen-2023-077529

bmjopen-2023-077529 - "LEVERAGING DATA SCIENCE AND MACHINE LEARNING FOR URBAN CLIMATE ADAPTATION IN AFRICA: A HE2AT CENTER STUDY PROTOCOL"  
  
Dear Mr. Parker,  
  
Following review of your article to BMJ Open, we invite you to submit a major revision.  
  
The review comments can be found at the end of this email, together with any comments from the Editorial Office regarding formatting changes or additional information required to meet the journal’s policies at this time.  
  
Please note that your revision may be subject to further review and that this initial decision does not guarantee acceptance.  
  
To submit your revised article please click this link: \*\*\* PLEASE NOTE: This is a two-step process. After clicking on the link, you will be directed to a webpage to confirm. \*\*\*  
  
[https://mc.manuscriptcentral.com/bmjopen?URL\_MASK=bdc9615eb6fe4e94b9c8f78af5292785](https://eur02.safelinks.protection.outlook.com/?url=https%3A%2F%2Fmc.manuscriptcentral.com%2Fbmjopen%3FURL_MASK%3Dbdc9615eb6fe4e94b9c8f78af5292785&data=05%7C02%7Ccparker%40wrhi.ac.za%7C91426eb16d724583a25408dc32ea68e8%7Ca9cf2b0981114caebb79bd8ff37647da%7C0%7C0%7C638441229569029740%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C0%7C%7C%7C&sdata=uxpipHoSMnVfVJ2VM7jbSZVXRtgJcl3zk6LzhkcNXMg%3D&reserved=0). Alternatively, you can log on to your Author Dashboard in ScholarOne and under "Action" click "create a revision".  
  
Please read and respond to all of the peer review comments. You should provide a point-by-point response to explain any changes you have (or have not) made to the original article and be as specific as possible in your responses.  
  
The original files will be available to you when you start your revision. Please delete any files that you intend to replace with updated versions and upload the following using the appropriate file designation:  
- “Main Document’’ - This is a clean copy (without tracked or highlighted changes) of your revised article. Please delete your original submission file.  
- “Main Document - marked copy” - This is the edited version of your original article, including edits to address the peer review comments. Any changes have been highlighted using a track change function or bold or coloured text.  
Please replace any other files that have been updated e.g. Images, forms.  
  
The reviewers' comments, your response, and the previous versions of your article will be published as supplementary information alongside the final version of your article.  
  
Information relating to your article, including author names and affiliations, title, abstract and required statements (e.g. competing interests, contributorship, funding) will be taken directly from the information held in ScholarOne, and not from the article file. Please check that this information has been entered correctly and has been updated as appropriate. If your revised article is accepted, you will only be able to make minor changes (e.g. correction of typesetting errors and proof stage) prior to publication.  
  
To ensure the system doesn't lock you out, please submit your revised manuscript at least one calendar day before 23-Mar-2024. If you are locked out and unable to submit your revision, please email our Editorial Office ([info.bmjopen@bmj.com](mailto:info.bmjopen@bmj.com)) and they will reopen it for you.  
  
Thank you for submitting your article to BMJ Open; we look forward to receiving your revision.  
  
If you have any queries, please contact the Editorial Office at [info.bmjopen@bmj.com](mailto:info.bmjopen@bmj.com).  
  
Kind regards,  
  
Thomas Phillips, PhD  
Research Editor  
BMJ Open  
  
Editor(s)' Comments to Author (if any):

* Please revise the ‘Strengths and limitations of this study’ section of your manuscript (after the abstract). This section should contain up to five short bullet points, no longer than one sentence each, that relate specifically to the methods. The novelty, aims, results or expected impact of the study should not be summarised here.
* Please include the planned start and end dates for the study in the methods section.
* We have noticed that you have used a map in figure 2. As BMJ Open publishes material under a creative commons licence, it is problematic to include copyrighted material, and we would therefore ask that if you own the copyright for this map and if not for it to be removed and/or replaced.

Formatting amendments (where applicable):  
  
  
Reviewer: 1  
Dr. Niels  Souverijns, Flemish Institute for Technological Research  
Comments to the Author:  
This paper describes in detail the approach the HE2AT project will tackle to find the relation between heat stress and health issues. It is a very nice project, and I am looking forward to the interesting results. Being a climate scientist myself, I limit my comments to the relevant sections in the paper.  
  
General comments:  
- 2 papers discussing heat-related health risks distribution among different socio-economic groups in Johannesburg are [https://www.sciencedirect.com/science/article/pii/S0169204620303947](https://eur02.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.sciencedirect.com%2Fscience%2Farticle%2Fpii%2FS0169204620303947&data=05%7C02%7Ccparker%40wrhi.ac.za%7C91426eb16d724583a25408dc32ea68e8%7Ca9cf2b0981114caebb79bd8ff37647da%7C0%7C0%7C638441229569044175%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C0%7C%7C%7C&sdata=NPjzB9jBkghNbaVxh1mRBsYjRxBkeT1jWBdyeIbiw%2F8%3D&reserved=0) & [https://www.sciencedirect.com/science/article/pii/S2212095522002498](https://eur02.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.sciencedirect.com%2Fscience%2Farticle%2Fpii%2FS2212095522002498&data=05%7C02%7Ccparker%40wrhi.ac.za%7C91426eb16d724583a25408dc32ea68e8%7Ca9cf2b0981114caebb79bd8ff37647da%7C0%7C0%7C638441229569055698%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C0%7C%7C%7C&sdata=zdM8TRKpqsPyTdfQwvUC4hhb27RhaS6IS4DtVwWPO1A%3D&reserved=0). They could be referenced in a few instances in the paper and offer latest insights.  
- To characterize heatwaves, the authors limit themselves to temperatures. However, in general, heat stress experienced by humans involves a combination of temperature, humidity, and radiation, such as encompassed in variables as the Wet Bulb Globe Temperature or Heat Index. I would advise the authors to look at these variables and not only temperature.  
- With respect to a heat health early warning system in a developing country, I can refer to the heat action plan in Ahmedabad (India), which offers detailed actions in case of high heat stress and the interactions with different institutional instances [https://www.nrdc.org/sites/default/files/ahmedabad-heat-action-plan-2019-update.pdf](https://eur02.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.nrdc.org%2Fsites%2Fdefault%2Ffiles%2Fahmedabad-heat-action-plan-2019-update.pdf&data=05%7C02%7Ccparker%40wrhi.ac.za%7C91426eb16d724583a25408dc32ea68e8%7Ca9cf2b0981114caebb79bd8ff37647da%7C0%7C0%7C638441229569063277%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C0%7C%7C%7C&sdata=4jlhrAs7PYG8lv4fgMzmOJBa5f8Ef1WEBgkyvALp0DI%3D&reserved=0). Maybe some inspiration can be obtained from their approaches.  
- P7 L18: it must be noted that land surface temperatures are decoupled from heat stress & are probably not the best tool to measure or quantify heat stress, see e.g. [https://ams.confex.com/ams/103ANNUAL/meetingapp.cgi/Paper/412571](https://eur02.safelinks.protection.outlook.com/?url=https%3A%2F%2Fams.confex.com%2Fams%2F103ANNUAL%2Fmeetingapp.cgi%2FPaper%2F412571&data=05%7C02%7Ccparker%40wrhi.ac.za%7C91426eb16d724583a25408dc32ea68e8%7Ca9cf2b0981114caebb79bd8ff37647da%7C0%7C0%7C638441229569069621%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C0%7C%7C%7C&sdata=zg2hDqtADy%2B2%2F45CA%2F3JFO%2B0Qx3TKq0fpMa14xb0%2FlQ%3D&reserved=0). It would be better to use a combination of temperatures and humidity.  
- P7 L23: It is not clear where the ‘geospatial climate’ data will come from yet from the paper. The data that is mentioned (CDS & ESGF) is not spatially detailed enough to distinguish different parts within the city? Recently some studies with WRF & UrbClim offer very detailed results on heat stress for cities that could potentially be used.  
- P9 L53: the following datasets might be of interest as input for the study: [https://viewer.esa-worldcover.org/worldcover/](https://eur02.safelinks.protection.outlook.com/?url=https%3A%2F%2Fviewer.esa-worldcover.org%2Fworldcover%2F&data=05%7C02%7Ccparker%40wrhi.ac.za%7C91426eb16d724583a25408dc32ea68e8%7Ca9cf2b0981114caebb79bd8ff37647da%7C0%7C0%7C638441229569075330%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C0%7C%7C%7C&sdata=qnQ8OyxHXiqCdT0ScnnIvbLKWYx5FT9lgw744ANV5Ls%3D&reserved=0) & [https://ghsl.jrc.ec.europa.eu/datasets.php](https://eur02.safelinks.protection.outlook.com/?url=https%3A%2F%2Fghsl.jrc.ec.europa.eu%2Fdatasets.php&data=05%7C02%7Ccparker%40wrhi.ac.za%7C91426eb16d724583a25408dc32ea68e8%7Ca9cf2b0981114caebb79bd8ff37647da%7C0%7C0%7C638441229569081175%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C0%7C%7C%7C&sdata=GzdojSdzMGeJRJdnsCXktSEplNgcpjYqSmno5PGPROU%3D&reserved=0);  
- P10 L37: again, it must be noted that Landsat and MODIS are land surface temperature tools and should not directly be used for air temperature retrieval (see above)  
- The idea for the forecast model is still vague, but the general ideas stated offer possibilities to perform the analysis.  
  
  
Reviewer: 2  
Dr. Jérémie Boudreault, INRS  
Comments to the Author:  
Review of : LEVERAGING DATA SCIENCE AND MACHINE LEARNING FOR URBAN CLIMATE ADAPTATION IN AFRICA: A HE2AT CENTER STUDY PROTOCOL  
  
This protocol proposes to study heat-related health risks in two major African cities, namely Johannesburg and Abidjan, over the 2000-2022 period. The project has three components: 1) mapping the exposure and sensitivity to urban heat, 2) building a heat-health outcome forecast model, and 3) developing an early warning system for heatwaves. The paper describes in detail the data that will be used, the proposed methods and the ways in which bias and consent will be handled. Although the objectives of this research are considerable, including mapping heat risks, modelling heat-related health outcomes and developing an early warning system, this research is highly relevant in the context of the lack of studies in the African region on heat-related health effects. However, there are some important points that need to be corrected in the paper prior to acceptance. Firstly, the overall structure of the paper needs to be improved by numbering sections (or using different fonts for subsections) and avoiding repeating information. Secondly, the introduction needs to be deepened to describe how this study differs from past work in Africa and worldwide. Thirdly, regarding the datasets, some details on the potential use of administrative health data (instead of clinical trials) should be explained, as well as the level of confidence that the authors will gain access to clinical trials data and its usefulness for heat-health studies. Finally, the methods, particularly the type of machine learning models employed, should be better explained. I have provided the authors with the following constructive and specific comments so that they can improve their paper.  
  
Major points  
  
1. Improving the structure of the paper  
  
Please add numbered sections and subsections (or use more clearly different font types). Also, please review the entire structure of the document. The three different objectives should always be identified in subsections (e.g., regarding the data and models used). Finally, continuous line numbering in future submissions would greatly facilitate the work of the reviewers.  
  
2. Deepening the introduction  
  
The introduction section is way too short (i.e., only two paragraphs). It lacks introducing the past work that has been conducted in Africa (and, more specifically, in the two studied cities) to tackle heat-related health effects. A quick search of ‘heat health Africa’ on Google Scholar already highlighted some research that could be of interest :  
  
Ncongwane, K. P., Botai, J. O., Sivakumar, V., & Botai, C. M. (2021). A literature review of the impacts of heat stress on human health across Africa. Sustainability, 13(9), 5312.  
  
Pasquini, L., van Aardenne, L., Godsmark, C. N., Lee, J., & Jack, C. (2020). Emerging climate change-related public health challenges in Africa: A case study of the heat-health vulnerability of informal settlement residents in Dar es Salaam, Tanzania. Science of the total environment, 747, 141355.  
  
Thiaw, W. M., Bekele, E., Diouf, S. N., Dewitt, D. G., Ndiaye, O., Ngom Ndiaye, M. K., ... & Diouf, A. (2022). Toward Experimental Heat–Health Early Warning in Africa. Bulletin of the American Meteorological Society, 103(8), E1843-E1860.  
  
Chapman, S., Birch, C. E., Marsham, J. H., Part, C., Hajat, S., Chersich, M. F., ... & Kovats, S. (2022). Past and projected climate change impacts on heat-related child mortality in Africa. Environmental Research Letters, 17(7), 074028.  
  
Wright, C. Y., Dominick, F., Kapwata, T., Bidassey-Manilal, S., Engelbrecht, J. C., Stich, H., ... & Matooane, M. (2019). Socio-economic, infrastructural and health-related risk factors associated with adverse heat-health effects reportedly experienced during hot weather in South Africa. Pan African medical journal, 34(1).  
  
In addition, some international contexts on the past heat-health studies to highlight the current state-of-the-art methods should be added regarding 1) heat risks mapping, 2) heat-health outcome models and 3) early warning systems. Finally, the last paragraph should clearly state the contribution of this study compared to the existing literature and its objectives.  
  
3. Discussing the reliability/availability of clinical trial datasets for heat-health studies and considering administrative health data  
  
The dataset section suggests looking at past literature to find adequate clinical trials datasets for this research. First, there is a risk that such data will not be available/shareable for this research. How will this risk be mitigated ? Have potential data sources already been identified ? Second, there is also a risk of using a dataset collected for another purpose as the 1st law of medical informatics states : “Data shall be used only for the purpose for which they were collected”. I have not seen much of clinical trial datasets be used for heat-health studies. This information should be provided in the Introduction and/or Method sections of the paper. Finally, and most importantly, most heat-health studies use administrative datasets such as daily counts of mortality or health services use (ED visits, hospitalizations, etc.). Is that data available for the two cities considered ? If so, it should be considered, especially for building the forecast model (objective II).  
  
4. Motivating the choice of the models, their advantages/drawbacks and how they will be validated  
  
For the first objective, it is mentioned that PCA will be used. There are many other methods that exist for dimensionality reduction. Why PCA in particular? For the forecast model (objective 2), deep recurrent neural networks are proposed. While these models are state-of-the-art tools in computer science, the study lacks identifying why these models should be used in the current context. In fact, the literature cited for deep learning applications is mostly out-of-scope. The following papers may help understand the current use of deep learning in heat-health studies :  
  
J Boudreault\*, C Campagna, F Chebana (2024). Revisiting the importance of temperature, weather and pollution variables in heat-mortality relationships with machine learning. Environmental Science and Pollution Research.  
  
Wang, C., Qi, Y., & Chen, Z. (2023). Explainable Gated Recurrent Unit to explore the effect of co-exposure to multiple air pollutants and meteorological conditions on mental health outcomes. Environment International, 171, 107689.  
  
J Boudreault\*, C Campagna, F Chebana (2023). Machine and deep learning for modelling heat-health relationships. Science of the Total Environment, 682(164660): 1-12.  
  
Lee, W., Lim, Y.-H., Ha, E., Kim, Y., & Lee, W. K. (2022). Forecasting of non-accidental, cardiovascular, and respiratory mortality with environmental exposures adopting machine learning approaches. Environmental Science and Pollution Research, 29(58), 88318–88329.  
  
Nishimura, T., Rashed, E. A., Kodera, S., Shirakami, H., Kawaguchi, R., Watanabe, K., Nemoto, M., & Hirata, A. (2021). Social implementation and intervention with estimated morbidity of heat-related illnesses from weather data: A case study from Nagoya City, Japan. Sustainable Cities and Society, 74, 103203.  
  
Wang, C., Feng, L., & Qi, Y. (2021). Explainable deep learning predictions for illness risk of mental disorders in Nanjing, China. Environmental Research, 202, 111740.  
  
Usmani RSA, Pillai TR, Hashem IAT, Marjani M, Shaharudin R, Latif MT (2021) Air pollution and cardiorespiratory hospitalization, predictive modeling, and analysis using artificial intelligence techniques. Environ Sci Pollut Res 28(40):56759–56771  
  
Wang, C., Qi, Y., & Zhu, G. (2020). Deep learning for predicting the occurrence of cardiopulmonary diseases in Nanjing, China. Chemosphere, 257, 127176.  
  
Studies comparing different models in the literature have shown that deep learning was not always superior to other methods such as tree-based approaches or non-linear statistical models (e.g., Boudreault et al., 2023; 2024; Lee et al., 2022). Will more simplistic statistical models be considered as well ? Furthermore, it needs to be more clearly mentioned how the models will be implemented, regarding the hyperparameters optimization, training and calibration datasets, types of validation (e.g., during heatwaves versus classical performance criteria, see e.g. Boudreault et al. 2023), etc. Finally, the methodology for the heat warning system should be better explained (will it be using the forecast model developed in objective 2 or both maps and models of objectives 1 and 2 ?).  
  
Other comments  
  
P2 L22 : How will data during the 2020-2022 COVID-19 pandemic will be handled ? There were some major shifts/changes in mortality/morbidity during that period. Please add this information in the “Data description section”.  
P4 L19 : Elaborate on the different socio-demographic groups affected by heat.  
P4 L5-L21 : Please see Major Comment #2 above.  
P4 L47 : It would be interesting to add some contexts why these two cities were selected for this study. By reading the title, we can think the study will be over the entire African continent, but it is limited to two cities. Is it because they have health data that other cities do not have ? Do they have higher population size compared to other cities ?  
P4 L55 : The background/rationale section should be presented before the “Study setting”. Please provide numbered sections (1, 2, 3) and subsections (2.1, 2.2.) to ease understanding the structure of the document. In addition, I think “Study setting” should belong to the “Method” section. (see Major Comment #1)  
P5 L3-21: How do these definitions relate to the hazard-vulnerability-exposure framework of IPCC ? If the framework is different than IPCC, please explain why and cite the appropriate references.  
P5 L41-L44 : This should be mentioned in the Introduction while adding past studies conducted in Africa. (see Major Comment #2)  
P5 L49 : This section may be included as the final paragraph of a broader Introduction. (see Major Comment #2)  
P6 L38 : Using numbered sections and subsections will help the reader understand the structure of the paper. (see Major Comment #1)  
P6 L40 : A summary table with all data sources (including health data) related to each of the three objectives would be useful here.  
P6 L50-L51 : I am curious if there is any administrative health database available (e.g., mortality, health services use). These are often the types of datasets used for heat-health studies. (See Major Comment #3)  
P7 L10 and L14 : What about the other city ?  
P7 L38-39 : If health trials and cohort data are used, they need to be introduced in the Introduction and how these can be used for heat-health studies. (see Major Comment #3)  
P8 L37-52 : I understand the point of looking for available datasets, but what if the trials and cohort data are not available/shareable for the current project ? Does this pose a risk to the achievability of the further steps ? (see Major Comment #3)  
P9 L47-55 : This is somewhat of a repetition of what was introduced in the Data section of the method. A better structure of the paper will help not repeating information in different sections. (see Major Comment #1)  
P9 L44 : Please add “First objective” to this section name.  
P9 L58 : Please elaborate on why only PCA will be used. What are the advantages/drawbacks of this method ? (See Major Comment #4)  
P10 L7-10 : Please add “Second objective” to this section name.  
P11 L3-4 : More details should be given related to the cross-validation and how models will be trained/calibrated/validated. (see Major Comment #4)  
P11 L9 : References 53 and 56 are out-of-scope for the proposed application. In addition, Ke et al. used an XGBoost model, which is not considered in the deep learning family. I recommend that the authors look at up-to-date literature for the use of deep learning applications in heat-health studies. (see Major Comment #4)  
P12 L8-L11 : Add “Objective 3” to this section title.  
P12 L19-L26 : Please develop here how the models developed in objectives I and II can be used jointly for the third objective. The maps of the first objective will be at much higher resolution than the forecast model of objective II, so how these divergent resolutions be handled for objective III ?  (see Major Comment #4)  
P12 L31 : Projection weeks in advance seems optimistic. Please nuance.  
   
Reviewer: 1  
Competing interests of Reviewer: None  
  
Reviewer: 2  
Competing interests of Reviewer: None to declare.