Piloting the collection of geographic information from Longitudinal Population Studies (LPS) for future climate-health research

Tejedor-Garavito, N., Bonnie, A., Bondarenko, M., Abbott, T. and Tatem A.J.

Report version 1.0

Table of Contents

DOI 10.5258/SOTON/WP00766
How to cite this work: Tejedor-Garavito, N., Bonnie, A., Bondarenko, M., Abbott, T. and Tatem A.J. Piloting the collection of geographic information from Longitudinal Population Studies (LPS) for future climate-health research. WorldPop, School of Geography and Environmental Science, University of Southampton, UK and Wellcome Trust, March 2023.

Executive summary

Background

As recently highlighted by the Intergovernmental Panel on Climate Change (IPCC) (IPCC, 2022), populations are highly vulnerable to the effects of climate change, in particular those living in low-resource settings (Romanello et al, 2021). Quantifying the specific health and livelihood outcomes associated with climate-related risks, however, remains challenging, and is largely an unexplored area of research (Nissan et al. 2022). Moreover, a fundamental challenge is the need for more representative and localized data that reflects the spatial scale at which health data is collected and allows an in-depth understanding of health-climate interactions.

Longitudinal population studies (LPS) are a means of studying how health outcomes are influenced over time by biological, social and environmental factors. Health and Demographic Surveillance Systems (HDSS), a type of LPS, monitor demographic and health characteristics of a population living in specific geographic areas. They are increasingly used to assess health outcomes and determinants in low- and middle-income countries (LMICs) in place of national civil registration and vital statistics, which are often incomplete and ineffective, and have been used to study associations between climate and health. LPS have great potential to explore the impacts of climate hazards on human health, however locating and measuring the exposure at each LPS site is often challenging. To be able to link LPS data with climate data, it is important to understand the geography that they represent.

Central to the Wellcome Trust's Climate and Health strategic program (https://wellcome.org/what-we-do/climate-and-health), is putting health at the heart of climate action and providing funding for research to understand the health impacts of climate change. The use of LPS for research at the health-climate interface was identified in a recent report commissioned by Wellcome (Nissan et al., 2022) as a key element in the alignment between the spatial and temporal scales of the climate hazards and individual health outcomes. Moreover, it was emphasized that data availability for research would benefit from a digital platform to visualise key metadata (such as total LPS population, date of establishment, and frequency of data collection). This project, entitled Piloting the collection of geographic information from Longitudinal Population Studies (LPS) for future climate-health research – funded by the Wellcome Trust in collaboration with WorldPop (www.worldpop.org) at the University of Southampton – builds on this recommendation, producing openly-available geographical information for 64 HDSS sites across two key LPS networks visualized through an interactive online portal, actively enhancing LPS geographic data accessibility and quality.

Methods

In this pilot study, HDSS sites were selected based on previous work carried out by WorldPop and funded by the Wellcome Trust . The 64 selected sites span two LPS networks – The International Network for the Demographic Evaluation of Populations and their Health (INDEPTH) and The Child Health and Mortality Prevention Surveillance (CHAMPS) network.

Initial datasets compiled from previous work consisted of the following:

- 53 INDEPTH site boundaries mapped and made available in previous studies by WorldPop, (Tatem et al., 2006; Jia et al., 2015; Utazi et al., 2016; Utazi et al., 2018);
- Coordinates of INDEPTH sites with supporting site information (including population size, main outcomes of study and rural urban classification) Fronterre, C. (personal communication, 2022; Appendix 1).

These two datasets provided a foundation for the project and formed the basis of our initial data review. Comparison of the coordinate points and site boundaries within a geographic information system (GIS) showed a degree of disagreement for several sites, instigating a rigorous literature review of all selected sites across official sources. These sources included the official INDEPTH and CHAMPS network websites and related academic publications.

The purpose of this literature review was to compile site information and identify the most up-to-date official site boundaries, and in all cases, boundaries provided in official materials were prioritized over those provided in our initial datasets. These identified boundaries were then digitized to generate an updated LPS dataset, incorporating site information. In some cases, the site boundaries originally produced as part of previous WorldPop work matched those in the official materials and did not require updating. In other cases, a boundary was neither provided in the foundational data or the official materials, and alternative sources were adopted.

To enhance user access and data visualization, a web portal was generated to map the sites and include some interactive features. The sequential pipeline used during the project's execution is shown in Figure 1 below.

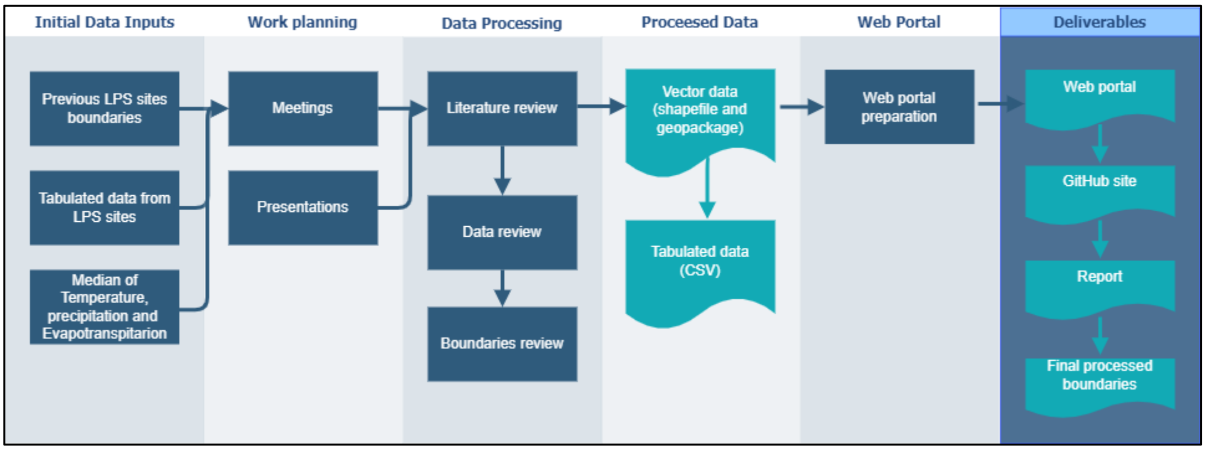

Figure 1: Flowchart illustrating the project pipeline.

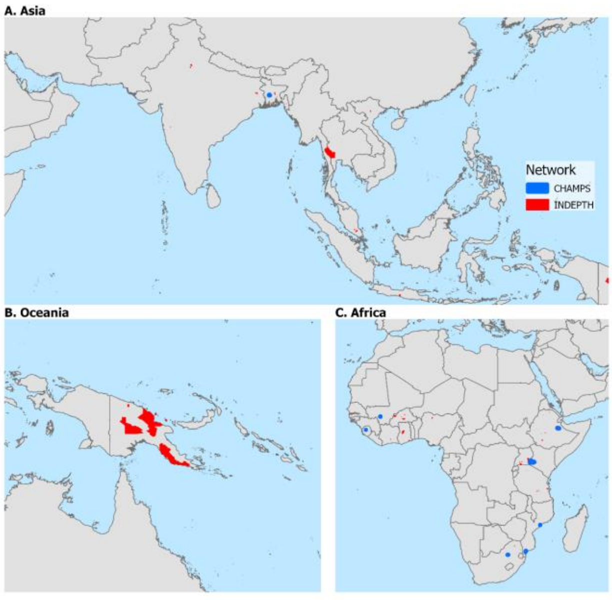

Figure 2: LPS sites displayed by network name across the continents, CHAMP sites (in blue) and in INDEPTH (in red) (Bonnie et al, 2023). Underlying data source: GADM, 2022.

Outputs

The final output (Figure 2) from this project consists of 64 HDSS site boundaries and compiled supporting site information (including, total population of HDSS site, date of establishment, frequency of data collection, rural urban classification, and basic demographic statistics, where available). This final output is provided as standalone GIS files (geopackage and shapefile) and embedded within the online portal.

Of the final 64 site boundaries, 32 INDEPTH boundaries were updated from the foundational dataset from WorldPop based on official materials. Site boundaries were generated for all 8 CHAMPS sites, where previously there were none. Combined, these 64 HDSS sites cover an estimated total population of 5,732,391, spanning 23 different countries across Africa, Asia, and Oceania.

Recommendations

This pilot study enhances the accessibility and user-friendliness of valuable LPS spatial data, allowing for increased use across climate-health research. This study describes a suggested methodology to update and incorporate LPS data into a geospatial setting that can be implemented systematically to open up opportunities to incorporate further data and analysis and address the complex challenges presented by the climate and health crisis. Throughout the project, limitations and areas for further development were identified, and recommendations for future developments in the short-, medium- and long-term have been outlined.

Short-term

- Continued maintenance and development of the web portal, including addition of supporting datasets - such as hydrological data, urbanisation trends, vegetation indices - and inclusion of feedback forms.

Medium-term

- Addition of further HDSS sites to the web portal to increase geographic coverage.
- Increased capability and depth of web portal, such as additional data filters and basic in-portal data analysis.

Long-term

- Capacity strengthening in GIS for LPS site leaders to enable them to upload and update site data through one cohesive interface.
- Harmonization of all HDSS sites and LPS networks across the web portal for optimized data use.

Short-term

- Continued maintenance and development of the web portal, including addition of supporting datasets - such as hydrological data, urbanisation trends, vegetation indices - and inclusion of feedback forms.

Introduction

The Climate crisis is a health crisis

The threat of the climate crisis is increasing all the time (Kadandale et al., 2020, Watts et al., 2019). Affecting livelihoods, well-being and health, the short- and long-term effects of climate change are a burden for the most vulnerable populations who often live in low resource settings (Romanello et al, 2021). An area in need of further research is the impact of climate change on health (Nissan et al. 2022) and the interconnectivity between these two fields has been widely established. For example, pollution from fossil fuel emissions or wildfire smoke, can lead to an increase in cases of short and long-term respiratory related illness; warmer land-surface temperatures increase risks of heat stroke and skin cancer, and shifts geographic coverage of vector-borne diseases; extreme weather events such as flooding can result in increased cases of water-borne diseases such as cholera, as well as disrupting food production impacting nutrition (Rocque et al., 2021). The impacts of climate on health vary temporally, ranging from short- to long-term, highlighting the complexities involved in understanding the breadth of effects.

Despite there being a wide understanding of the impact climate change has on health, accurate and comprehensive quantification of these effects is lacking. This is important for informing health and social policies to mitigate against the effects of climate change on health. The 2023 Lancet Countdown report (Romanello et al., 2023) emphasizes that mitigating against the worst impacts of climate change on health requires "profound and immediate systemic changes" (pg. 4). The report goes on to highlight the need to focus on primary prevention and the acceleration of mitigation efforts across all sectors, to ensure health systems are able to adapt to the impacts of climate change. As Kandale et al. (2020) support, fortification of primary health care currently frequently fails to consider these impacts.

As the 2023 Lancet Countdown report (Romanello et al., 2023) also highlights, no region is unaffected by the increasing impacts of climate change on physical and mental health – however, due to "structural injustices, and harmful power dynamics, both between and within countries", the impacts are most acutely felt by vulnerable populations in resource-constrained settings despite contributing the last to climate change. Representative and localized data to reflect the spatial scale at which health data is collected, will allow for a more in-depth understanding of health-climate interactions to ensure optimization of healthcare within these constrained settings.

The value of digital tools

One such means of enhancing our capacity to combat the climate crisis is the use of digital tools. A particular value of digitized initiatives within a sustainability context is the lack of constraint by geographic boundaries, and the increased ability to scale up and widen impact (George et al., 2021). Raleigh & Urdal (2007) note the value of geospatial data formats, and the ability to apply data - generally collected at a local level - to any level, from household to global.

Geographic information system (GIS) tools have been widely used across both the health and environmental sectors. GIS-based approaches offer flexibility (Tomlinson et al., 2011), and the ability to draw on multiple data sources such as a remotely sensed data and local point data to carry out comprehensive analyses and identify spatial patterns at multitemporal scales. The impact of climate change on health is a complex challenge, and requires study at local and global levels, which GIS tools allow with relative ease (Raleigh et al., 2007).

LPS and HDSS

Longitudinal Population Studies (LPS) refers to the systemic collection of population-related data over time. LPS are carried out through cohorts, panel surveys and biobanks (Wellcome Trust, 2017). The Wellcome Trust has invested extensively in supporting LPS, citing their importance to the scientific community as “there is no other way of understanding how biological, social and environmental factors interact over time in a population to produce health outcomes”.

Health and Demographic Surveillance Systems (HDSS) are a type of LPS and were developed in response to a lack of systemic and comprehensive demographic and health data capture within developing countries (Islam et al., 2020). In such resource-scarce settings, where populations are often significantly more vulnerable to diseases, viruses and other health-threats, accurate health and demographic data are crucial for the development of appropriate management and prevention strategies. At their core, HDSS sites systemically survey key demographic indicators (births, deaths, cause of death) for “relatively small geographically defined populations” (Hinga et al., 2021). Despite the limited spatial coverage of HDSS sites, observations from these specific geographic areas do play a key role in informing appropriate health and social policy decisions at a much wider scale, although it is important to be cautious when drawing generalizations based on them (Herbst et al., 2021). Furthermore, despite LPS being spatial in nature, there is no spatial data (in GIS data format) currently openly available associated with LPS.

A significant strength of HDSS data is its temporal resolution, with data-collection rounds on average every 3-4 months (Bos, 2004), offering a significant advantage over other demographic surveys such as censuses and other LPS. Higher temporal resolution of health and demographic indicators allows them to be temporally matched with climate data at a finer temporal scale, offering an opportunity to understand the impacts of climate change on health. A recent report commissioned by Wellcome highlights that even this relatively high temporal resolution is much coarser in comparison to the resolution of many climate datasets, meaning that fine-scale interactions between climate and health often cannot be accurately captured (Nissan et al., 2022). However, the authors highlight the value of LPS in secondary exploratory analyses on the influence of climate on health, and the opportunity to work with and reorientate selected LPS to better capture these interactions.

Previous work

The WorldPop research programme at the University of Southampton has worked on several studies that have gathered data on HDSS sites, mapped their extents, and assessed their environmental and socioeconomic representativeness (Tatem et al., 2006; Jia et al., 2015; Utazi et al., 2016; Utazi et al., 2018). In 2006, Tatem et al. examined the environmental coverage of The International Network for the Demographic Evaluation of Populations and their Health (INDEPTH) Network in rural Africa, which consisted of 21 HDSS sites at the time of analysis. Using satellite imagery, they defined an environmental signature for each INDEPTH site, analyzed similarities to other surrounding areas, identified sites that formed environmentally cohesive groups, and highlighted areas where the location of new INDEPTH sites would significantly increase the environmental coverage of the network.

In 2015 Jia et al, expanded the work to include 43 sites within the INDEPTH network: 30 in Africa, 12 in Asia, and one in Oceania. This widened the geographical extent of study and added new global environmental and socioeconomic datasets. Further work by Utazi et al. (2016;2018) reviewed 39 sites in sub-Saharan Africa using socioeconomic and environmental data and developed a full Bayesian methodology to assess the representativeness and consequently, identification of potential future sites to increase the representativeness of the network of sites.

The previous work assembled and digitized the boundaries of many HDSS sites across the world. This was undertaken through contacting HDSS site leads within the INDEPTH health and population research network and working with the network coordinators themselves, as well as digitizing online and journal publication documents. This work was undertaken many years ago; however, the datasets have not been updated recently, nor have they been supplemented with location data on any additional new sites, or metadata on existing sites. Moreover, they have not been made publicly available to benefit the wider community.

Project rationale

As the recent report commissioned by Wellcome highlights, understanding the interactions between climate and health is a priority area of research (Nissan et al., 2022) – both in assessing current impacts on health, and how these impacts may evolve over time, which is crucial in developing combative healthcare policies and structures. The authors emphasize the importance of developing collaborations between health and climate scientists to tackle the climate crisis, which is continually developing.

HDSS sites provide substantial data at a range of temporal and spatial scales on a range of health outcomes that have the potential to be used in research that can be connected with climate-related exposures. However, previous HDSS site research has been based on digitized site boundaries that are not available as an open resource. This presents a significant barrier to enable collaborations between the climate and health sectors.

Therefore, this project aims to develop the foundation on which these collaborations could occur, by providing a reproducible methodology to incorporate and update the geographical boundaries of LPS data and, to provide open-source and accessible spatial data for selected HDSS, while also exploring the possibilities of combining these with climatic data. This project also provides a user-friendly web portal to display the LPS sites data and enable end users to reproduce, update and upload similar LPS data to build on the existing data framework and expand the capabilities of using LPS spatial data. By establishing an accessible methodology, an open platform for data sharing, and promoting user involvement, this project aims to revolutionize data analysis and visualization using LPS data, while facilitating an open collaborative environment wherein researchers can collectively advance their insights and knowledge, ultimately leading to more informed decisions and enhanced research outcomes.

Data and Methods

HDSS networks – INDEPTH and CHAMPS

Established in 1998, the INDEPTH network is comprised of 49 HDSS sites across 42 health research centres, and 7 associate members, totaling 56 HDSS sites at the time of writing. Member sites span Africa, Asia and Oceania across 21 counties, with site populations ranging from 8,200 to 260,000, and a total observed population of around 4.5 million. INDEPTH's vision is "to be a trusted source for evidence supporting and evaluating progress towards health and development goals", drawing attention to the need for reliable health and demographic data for policy makers in low- and middle-income countries (LMICs) and the value of longitudinal population studies (http://www.indepth-network.org/about-us).

CHAMPS was established in 2015 and focuses on child health and mortality. The network consists of 8 sites, across 7 countries in sub-Saharan Africa and South Asia, identified as having notably high child mortality rates (population greater than 100,000 and under-5 mortality greater than 50 per 1000 live births). The network aims to identify and track causes of neonatal and child mortality to support the development of health and policy interventions (https://www.cdc.gov/ncezid/stories-features/global-stories/champs-program.html). A summary of key data for the two networks of sites is presented in Table 1.

Network	**Number of sites**	**Total population**	**Date Network established**	**Earliest site start date**	**Focus**
INDEPTH	49	4,500,000	1998	1961	Health & demographic
CHAMPS	8	1,170,000	2015	2015	Child mortality

Table 1: Tabular summary of key characteristics for HDSS networks. Note: Total population refers to the total population as established in official HDSS site sources.

Updating boundaries

The initial datasets available consisted of 53 digitized HDSS boundaries out of 56 total current INDEPTH sites from previous work led by WorldPop, and cartesian coordinates for all INDEPTH sites provided by Fronterre, C. (personal communication, 2022; Appendix 1), including site metadata with population size, main study outcomes and rural/urban classification. No spatial data was given for any CHAMPS sites.

The first step in the methodology was to carry out initial data exploration, where both the boundaries (taken from previous work by WorldPop) and coordinates (Fronterre, C. personal communication, 2022; Appendix 1) were compared spatially (see an example in Figure 4 and additional site comparisons in Appendix 1) and noted several cases where the two datasets did not align.

This led to the second stage of our methodology in the form of a literature review. This primarily involved reviewing materials linked on the official INDEPTH website. INDEPTH provides a webpage for each HDSS site of which the majority contain links to official site reports or publications. The majority of these reports/publications contain site maps, we could take as reference. In some cases, the linked publications either contained no site map or the site map was unclear; in other cases, the INDEPTH site page did not contain links to any reports/publications. In these cases, secondary or related publications were identified through Google Scholar, using site names and relevant terms (e.g., HDSS, INDEPTH). Where these secondary sources contained clear site maps, these were taken as reference; otherwise, site descriptions provided in primary or secondary sources were used to identify location (see Appendix 2b for publication and source of reference boundaries).

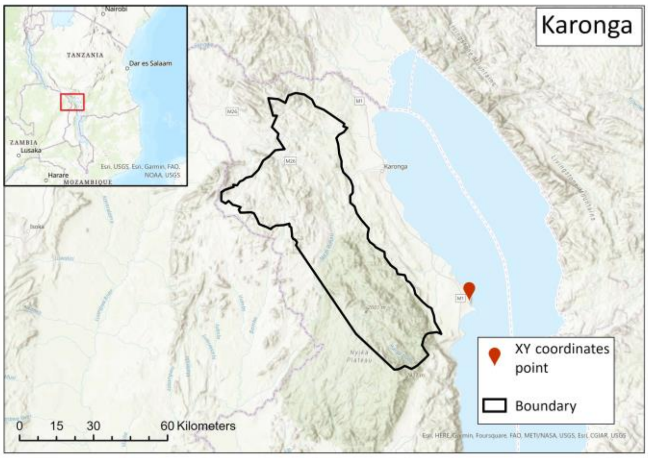

Figure 3: Karonga HDSS site as delineated (in black) from original WorldPop shapefile in comparison with the XY point provided (in red) by Fronterre C. (personal communication 2022). Basemap data licensed by ESRI (2023).

The literature review indicated a need to update all boundaries, given the persistent differences between the cartesian coordinates provided by Fronterre, C. (personal communication, 2022; Appendix 1),the digitized boundaries provided by WorldPop, and boundaries identified from the literature review. In all cases, boundaries identified in primary and secondary materials (Appendix 2b) from the literature review were prioritized in this update over those produced in previous work by WorldPop. Where the boundaries for a given site either closely aligned or matched, no update was made; otherwise, the site boundaries were updated by digitizing the boundary provided in the official material.

Boundaries were updated through digitization – taking the site map/boundary from a given publication and laying this over an informative basemap within ArcGIS Pro – for example, satellite imagery featuring Maxar imagery, or the topographic basemap which includes features such as roads, rivers and national parks. Such features provide a basis for lining up with the HDSS site maps from the publications. This is showcased well in the example of Agincourt (Figure 4), where the extent of a local natural park was used to appropriately digitize the site boundary provided in the publication.

In cases where official publications included no boundary maps, or included maps of poor quality that were not possible to digitize, boundaries were derived from supporting text descriptions or site names. This information was used to select appropriate Global Administrative Area (GADM) (GADM, 2022) boundaries, a well-known and reputable source for administrative boundaries. If there existed a GADM boundary with the same name as a given site, this was adopted unless otherwise stated in official materials. Otherwise, text descriptions were used to identify the smallest suitable administrative unit boundary that aligned with the site description. If the publication contained no or poor site maps or site descriptions, then the original boundary from the initial work by WorldPop was kept. Appendix 2 provides further detail on this – stating whether a boundary was updated or not, justification of this, and the source of the boundary update if applicable.

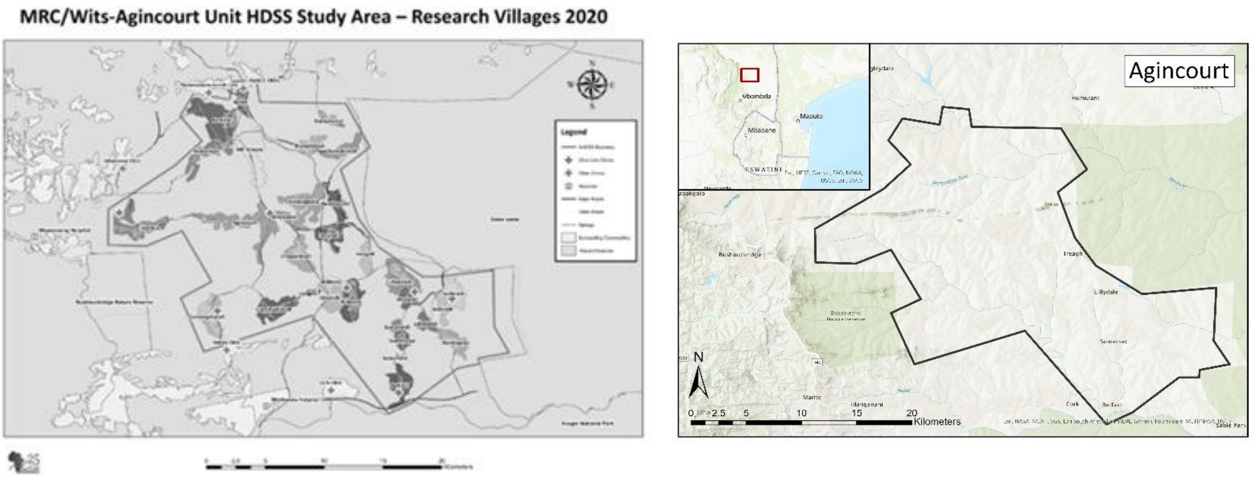

Figure 4: Agincourt HDSS site as delineated in the publication (left) and comparison with reviewed site (right) with the original site in green and the updated site in blue. Publication site map converted to black and white from original source to improve accessibility. Source: https://www.agincourt.co.za/?page_id=1896 and basemap data licensed by ESRI (2023).

In the case of the HDSS sites for Wosera (Papua New Guinea) (Figure 5) and Dikgale (South Africa) (Figure 6), the updated boundaries remain simplified, to cover the extent of the area displayed in the publication. Site extents were provided in their respective publications (see Appendix 2b for publication and source of reference boundaries), and although this would indicate the need to adopt a more appropriate boundary from GADM, it was found upon investigation that these were no more suitable (see Figure 7 for example). As such, the site boundaries provided in the output dataset are accurate to official materials, they are just more simplified in comparison to other site boundaries.

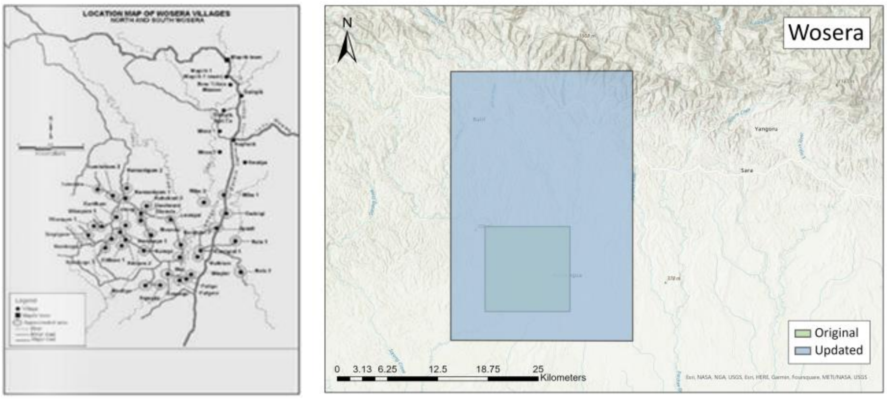

Figure 5: Wosera HDSS site as delineated in the publication (left) and comparison with reviewed site (right) with the original site in green and the updated site in blue. Publication site map converted to black and white from original source to improve accessibility. Source: https://www.yumpu.com/en/document/read/50735893/wosera-hdss-indepth-network and basemap data

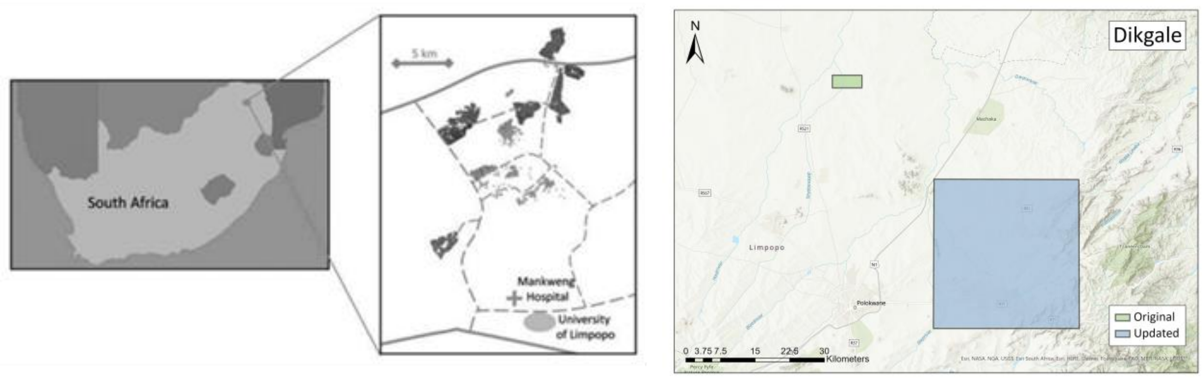

Figure 6: Dikgale HDSS site in South Africa as delineated in the publication (left) and comparison with reviewed site (right) with the original site in green and the updated site in blue. Publication site map converted to black and white from original source to improve accessibility. Source: Alberts et al. (2015) and basemap data licensed by ESRI (2023).

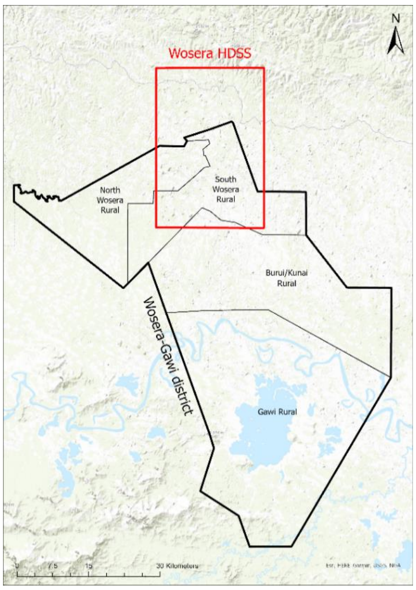

Figure 7: Comparison of updated boundary for INDEPTH site Wosera (red) and GADM administrative level 2 and 3 boundaries (black) for district Wosera-Gawi. Basemap data licensed by ESRI (2023) licensed by ESRI (2023).

CHAMPS provide names of study sites and text descriptions rather than boundary data. To facilitate inclusion of these studies in the web interface, boundary data provided by GADM were adopted where CHAMPS site names matched GADM data. In the two instances where names and data did not match (Soweto, South Africa and Makeni, Sierra Leone), boundaries were derived from alternative sources. In the case of Soweto, a map of the township from a recent publication was used (Bridger, E. 2021), and for Makeni, a boundary of the city was digitized using Maxar satellite imagery for 2021 available in the in-built ESRI imagery basemap in ArcGIS Pro (Figure 8).

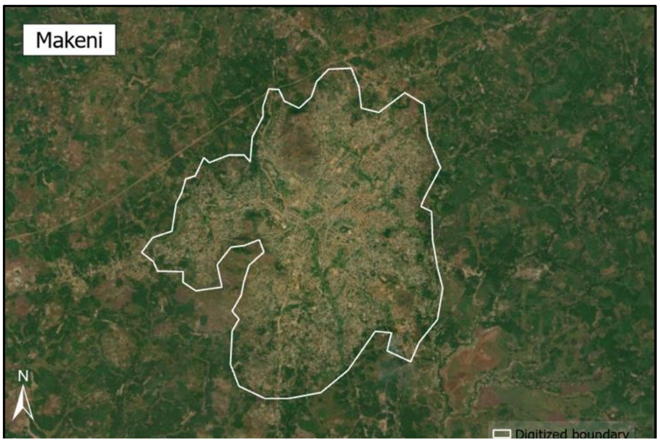

Figure 8: Makeni HDSS site digitized from satellite imagery licensed by ESRI (2023).

Climate data

In line with the project aim to develop a foundation for the study of climate-health interactions, example climate data were extracted for each HDSS site. The purpose of this was to demonstrate the functionality of the site boundaries and the value these can bring to climate-health studies. As the purpose of this was to simply demonstrate this functionality, limitations within the climate data itself were not explored in any depth, such as low spatial resolution, or climate anomalies due to influencing factors such as a La Niña phase, for example.

A selection of climate data - total annual precipitation (mm), total annual evapotranspiration (mm), and mean annual temperature (°C) - were extracted for each HDSS site. The most recent annual interpolated meteorological station data for 2021 from the University of East Anglia Climate Research Unit were used (University of East Anglia Climatic Research Unit, 2021). Processing was carried out in QGIS 3.22.9 (Figures 9 and 10). Climate data were provided in raster format; centroids were generated for each polygon within each HDSS site (so sites with multiple areas of study generated multiple centroids), and the underlying raster cell value was then extracted to each point. These point values were then exported to comma separated value (csv) format, and further processed in Excel so that, where sites consisted of multiple point values, a median value was generated. The processed csv file was then joined to the original HDSS shapefile and exported to a new output.

An alternative method would be the production of zonal statistics, where the mean (for example) of all raster cells that fall within a defined boundary (e.g., HDSS site) is generated; however, this is reliant upon one or more raster cell centroids sitting within the polygon boundary, and since the resolution of the climate data used here was quite coarse (0.5°), some HDSS sites were too small or did not incorporate any raster cell centroids, so no statistics were generated. Hence, the outlined process of using site centroids was adopted.

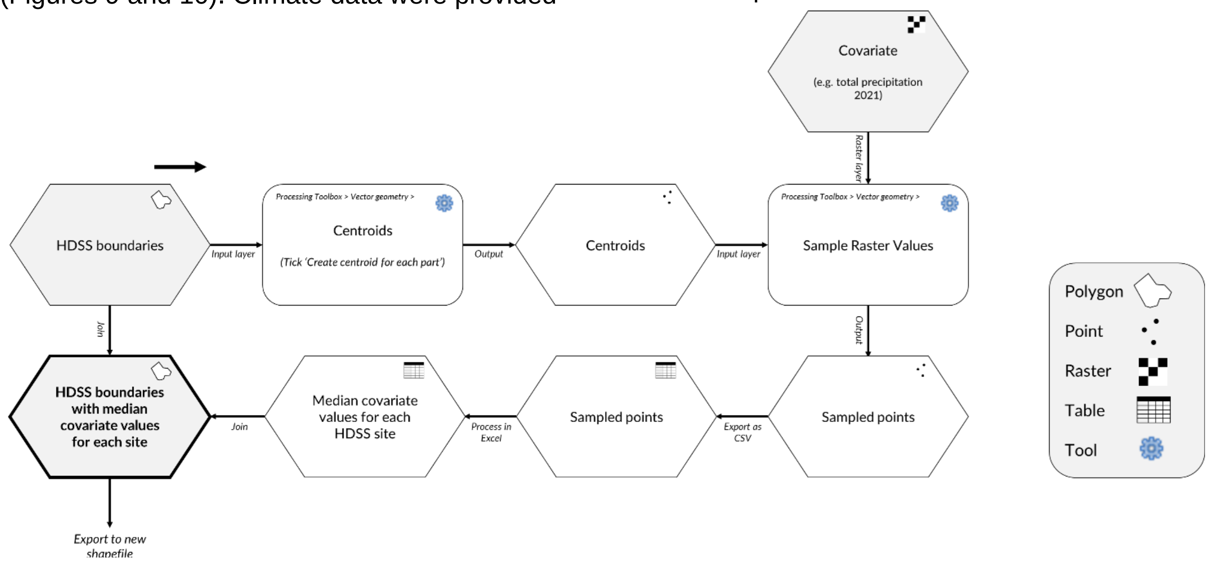

Figure 9: Flow chart illustrating the process of extracting climate data for HDSS sites.

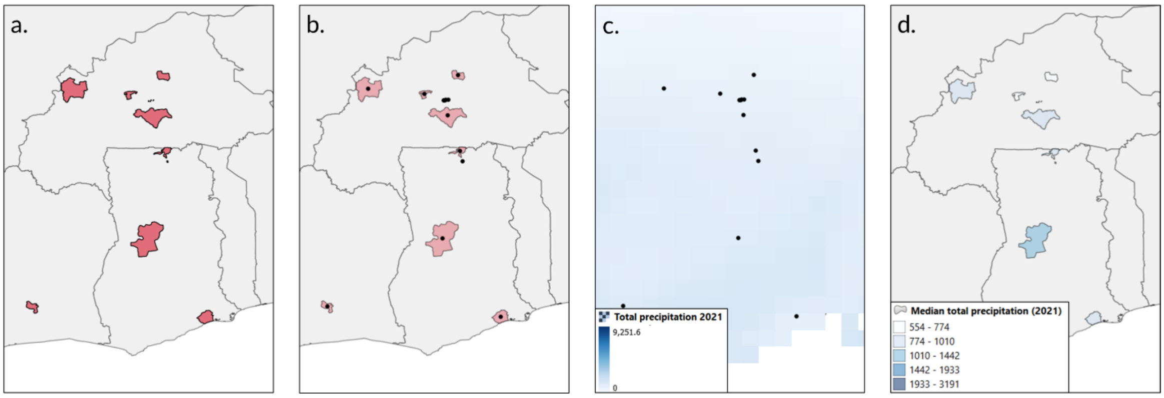

Figure 10: Example of climate data extraction processing in QGIS 3.22.9-Białowieża; a) HDSS site boundaries, b) generation of centroids for all parts of each HDSS site, c) input of climate raster data and extraction of values to centroids, d) joining median of extracted raster values back to the HDSS site boundaries and visualization of these values. Underlaying data source: GADM, 2022.

Shapefile preparation

Once the boundaries and data for all the sites were complete, they were prepared as a shapefile using ArcGIS Pro 3.0.2 and QGIS 3.22.9-Białowieża and published with open access (Bonnie et al, 2023) available at https://data.worldpop.org/repo/prj/LPS/v1_0/. This was then the base for the web portal production.

Construction of the portal

An interactive online portal was created (Figure 11) to visualize the HDSS site, allowing end users to explore the data for the different networks. The main objectives of the portal are:

- To display the HDSS locations for the two networks (INDEPTH and CHAMPS) in the different countries
- To provide options to filter the HDSS on the map by country, Network and LPS
- To display more information and metadata for each HDSS

The workflow for the portal production is outlined in Figure 12. In brief, once the vector data were finalised, these were saved in a format that can be visualized in any software that handles geospatial data, in this case shapefile and geopackage. GIS software was used to convert the data into a GeoJSON format for it to be incorporated into the portal.

The portal is based on JavaScript using the following libraries: “Leaflet”(https://leafletjs.com/) for mapping of geographical data with some interactive features, and “JQuery” (https://jquery.com/) for easy HTML manipulation. Leaflet is a comprehensive open-source JavaScript library for friendly interactive maps. Leaflet can handle different basic JavaScript tasks like converting data to map layers and mouse interactions. Leaflet also works across most types of devices. Additional information can be found in Appendix A4d,

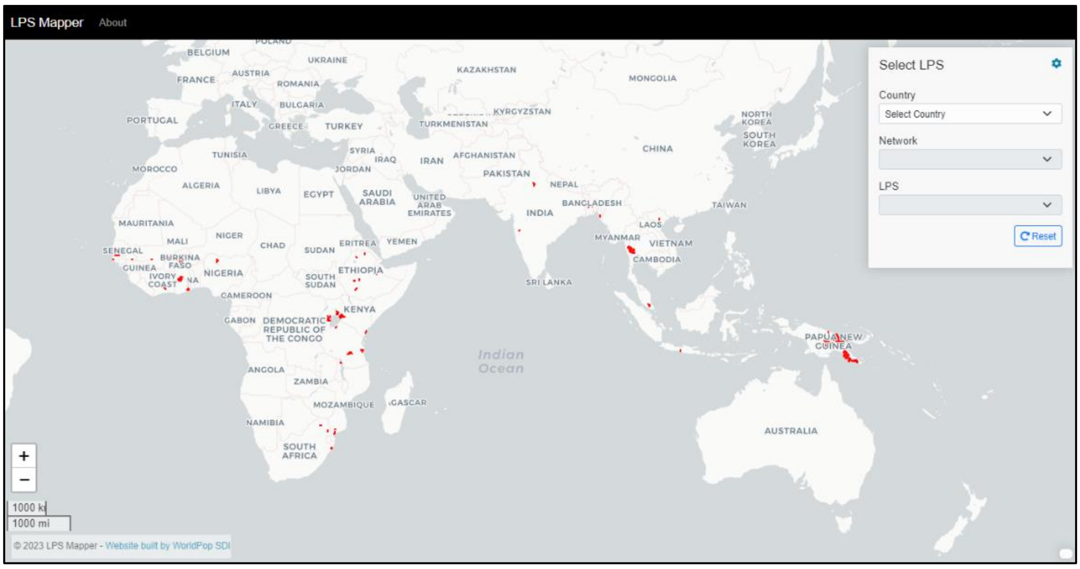

Figure 11: Screenshot of the interactive online portal available at https://wpgp.github.io/lpsmapper/.

where the folder structure behind the portal and additional modules used can be found.

Three basemaps were included in the portal, OpenStreetMap (OSM) (https://www.openstreetmap.org/), ESA WorldCover 10 m 2021 v200 (https://esa-worldcover.org/) and Global Grid of Probabilities of Urban Expansion to 2030 (https://doi.org/10.7927/H4Z899CG). These datasets were used to exemplify the potential of including datasets via an API to display in the background of the portal. The source files are available at https://github.com/wpgp/lpsmapper, and a mirror portal can be accessed at https://wpgp.github.io/lpsmapper/. For maintenance, update, and further development, basic knowledge and experience of those libraries will be necessary.

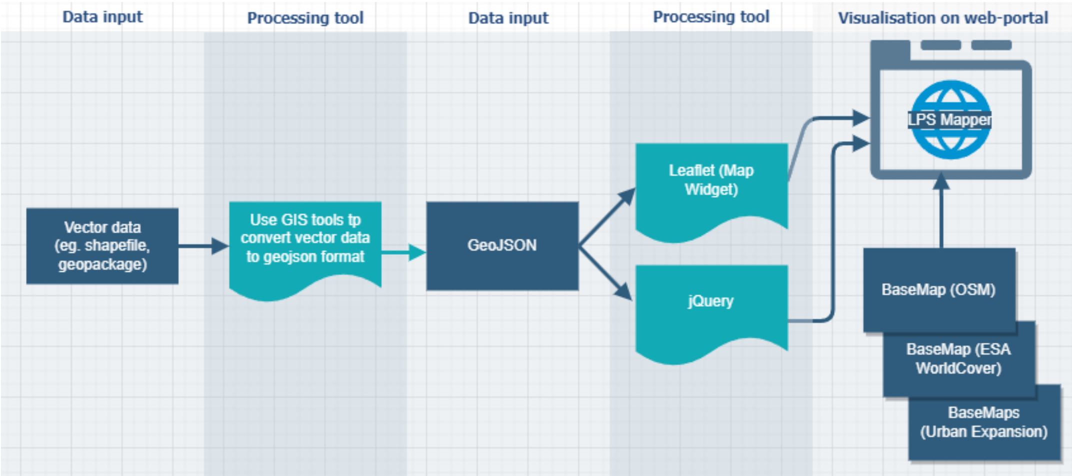

Figure 12: Workflow diagram of the web portal production

Outputs

64 sites (Figure 13) were collated from the CHAMPS and INDEPTH networks. These include updated boundaries for 32 out of 56 INDEPTH sites and GADM/digitized boundaries for all 8 CHAMPS sites. HDSS site information was based on data from previous work (Nissan et al., 2022), and where possible, information provided on websites and in publications was included to complement data for the sites. Key statistics (titled 'fast facts') were provided for the CHAMPS sites on the official CHAMPS website, including Still Birth Rate, Maternal Mortality Rate, Neonatal Mortality Rate, Infant Mortality Rate and Under-5 Mortality Rate (CHAMPS, 2022), which have been included as site information. Additionally, key climate data were extracted for each site. An example of the information included for each of the sites is shown in Figure 14.

A total of 32 boundaries out of 56 sites total were updated. Appendix 1 provides a comparison with initial XY coordinates and boundaries and Appendix 2 provides the revised polygons for the 56 sites.

Figure 13: Overview of final LPS dataset (Bonnie et al, 2023). Underlaying data source: GADM, 2022.

Name	Country	Population Size	Number of Households	Number of Villages	Type	Enroll Start	Temporal resolution (months)	Spatial resolution
Agincourt DSS	South Africa	115000	16000	27	Rural	1992	12	Household
AHRI	South Africa	168000	11000		Rural	2000	4	Household
Arba Minch DSS	Ethiopia	68802			Rural	2009		
Ballagbargh DSS	India	90240	9584	28	Urban and Rural	1961	12	Household
Bandafassi DSS	Senegal	13373		42	Rural	1970	12	Village
Bandarban DSS	Bangladesh	19403			Rural			
Bandim DSS	Guinea-Bissau	105000		182	Urban and Rural	1978	6	Village
Birbhum DSS	India	59395	12557	315	Rural	2008	12	Household
Butajira DSS	Ethiopia	78000		10	Urban and Rural	1990	12	Village
Chokwe DSS	Mozambique	99834	21498	15	Urban and Rural	2010	6	Household
Chakaria DSS	Bangladesh	120000	19847	49	Rural	1999	12	Household

Figure 13: Figure 14: Example information for all INDEPTH sites. Data was reviewed and adapted from the original table provided by Fronterre, C. (personal communication, 2022) (Appendix 1), information supplied from the INDEPTH website, and related publications. For the full table, including information for CHAMPS sites, see Appendix 3.

Shapefile output and web portal

Once the data were compiled, a shapefile was updated (Figure 10) to prepare the web portal. Appendix 4 contains the shapefile and metadata for the column headers included in the shapefile and the web portal (see Section 6).

Limitations and data gaps

HDSS boundaries

During this pilot study, key limitations and data gaps were noted. Primarily, these limitations derive from a lack of data homogeneity between sites, within and between networks. This was particularly the case with INDEPTH sites, where available site information varied in quality and quantity, and time since the latest update was unclear. Official publication materials were provided for INDEPTH sites; however, these varied in terms of author and source (Appendix 2b), and thus lack consistency in what information was provided, with limited useability in this format. This study compiles the available information into a consistent and usable format. However, the information available for each site is varied. Contrastingly, the CHAMPS sites are more consistent in the site information provided on the official website, but these arguably lack depth when compared to the INDEPTH sites. Site information between the two networks varies, as well as between sites within the same network.

The only open-source boundaries available for the INDEPTH sites were provided in the related publications/official documents, and significantly varied in legibility. The quality of the digitized boundary was limited to the quality of the boundary provided in each publication. Additionally, it was not always clear from the publication how accurate the boundaries were, i.e., whether they referred to the wider area, and many of the site boundaries were from outdated publications with no recent alternative, so it is not known whether these may have changed since. Consultation with site leads would be required to check this, and this was beyond the scope of the work.

Where site boundaries were non-existent, or not usable from the official documentation, GADM boundaries were used, which introduces further uncertainty. The highest resolution suitable boundaries were used in these cases. However, these boundaries may encompass a larger area than the DSS site itself. This accommodation compromises the accuracy of the climate data. For example, where the GADM boundaries are larger than the site, climate centroid data may not be truly accurate to the site. This is a more significant issue for sites that consist of multiple ‘sub-sites’ - the PiH boundaries, for example, are spread out across the entire country of Papua New Guinea so there is likely to be much higher uncertainty in the climate data here, in terms of it representing conditions at the sites themselves. To improve the delimitation of site boundaries, alternative sources of digital boundaries could be explored, such as Second Administrative Level Boundaries (https://salb.un.org/en) or Geoboundaries, which are also open datasets of political administrative boundaries (https://www.geoboundaries.org/).

Conclusion

The outputs of this study provide a foundation on which climate-health research can be carried out, by producing a reproducible methodology and a standardized open-source set of LPS site boundaries under one dataset. Moreover, the production of a reproducible web portal ensures these data are easily accessible and demonstrates an alternative, more visual approach to HDSS data storage and presentation, in comparison to current HDSS network websites, enabling an inclusive environment for interdisciplinary collaboration.

This study covers two HDSS networks, presenting a methodology that supports a maintained level of quality of data standardization that ensures comparisons can be made across sites and reduces processing time by end-users. Key to this project is harnessing the existing value of LPS (i.e. their high temporal coverage) in place of more optimized studies, recognizing the potential value they have for understanding complex climate impacts on health and developing this into actuality. This is outlined by the Wellcome Trust's combining climate and health data report (Nissan et al., 2022), which encourages combining health and climate data, specifically for secondary analysis of existing LPS and climate data to develop hypotheses and inform the design of studies on specific climate-health interactions). Furthermore, key in facilitating these secondary analyses is ensuring data are open source and accessible. Prior to this study, HDSS spatial data were not openly available. Site boundaries, if published at all, were confined to papers or source materials of varying levels of quality. This presented a significant barrier to vital climate-health research.

There are limitations and areas for improvement within this initiative, as outlined in this report, but this project has produced a solid foundation on which these can take place. Through continued vital support from Wellcome, the increasing accessibility of current and future LPS datasets will allow researchers to better understand climate-health interactions and in turn inform healthcare and social policy decision makers to ensure optimized responses to the developing climate-health crisis.

Recommendations

The work presented in this report provides an initial detailed methodology to create geospatial datasets for LPS sites with selected climate data. Additionally, the datasets have been published along with an online portal allowing different stakeholders to explore the data and for those with additional skills to recreate the web portal and add any further data. This initial work is a pilot example which could easily be expanded to other LPS. Future work could explore the inclusion of additional site data as well as different environmental and socio-economic variables in the web portal and shapefiles, including climate data that coincides with the time points of the LPS. Examples of potential additional research work could include:

Short-term

- The inclusion of additional geospatial layers into the portal and their harmonization potentially enables analyses and improved understanding on cross-site spatiotemporal trends in demographic shifts, urbanization, infrastructure development and land cover/use changes, as well as in surrounding areas. This could include calculation of urban/rural area within sites and integrating data on building footprints and road networks.
- The assembly of LPS data and linkage with geospatial and other datasets to enable multi-factor spatiotemporal analyses of the impacts of climate change, urbanization, poverty alleviation and infrastructure improvements, among other factors, on population health across LMICs.

Medium-term

- Improved functionality of the web portal – for example, embedding links to LPS reports or facilitating basic data analysis in the portal, such as generating measurements or producing simple statistics from geospatial layers included within portal.
- The addition, harmonization, and linkage of geolocated national household survey data from, for example, the Demographic and Health surveys or Multiple Indicator Cluster Surveys, to enable drawing on the strengths of each type of demographic data source (i.e. the wide spatial coverage of national household surveys, vs the detailed temporal coverage of LPS) and bring new insights.
- Re-examination of how representative the current distribution of LPS sites are of the range of environmental, climatic, demographic, and socioeconomic factors that exist across the low and middle income regions of the world. If key factors are missed or poorly represented, insights from the full set of sites will still represent an incomplete picture. Previous methods have shown how this can be measured and recommendations for establishment of additional monitoring made (e.g. Tatem et al., 2006; Jia et al., 2015; Utazi et al., 2016; Utazi et al., 2018).
- The development of an R package to optimize data querying (such as spatial analysis, attribute filtering, comparison with other datasets).

Long-term

- Standardization of LPS datasets under one web portal for optimized data use. This could involve capacity strengthening in GIS for HDSS site leads to enable them to upload and update site data through one cohesive interface.
- The regular population enumerations undertaken for LPS offer possibilities for using the multi-site data to train and validate population estimation models (e.g. https://www.worldpop.org/methods/population-estimation-for-sustainable-development/) to improve inter-censal population estimates, especially for settings where denominators for health interventions are outdated and unreliable.

About WorldPop

WorldPop (www.worldpop.org) is a research and implementation group at the University of Southampton, UK. First established in 2005 to construct high resolution population maps to support disease mapping activities, WorldPop aims to provide an open access archive of spatial demographic datasets, primarily for low- and middle-income countries, to support development, disaster response and health applications. Through integration of 'traditional' datasets such as census and surveys, with more novel 'big' data from sources such as satellites and cell phones, the group aims to provide timely and high-resolution data on demographics, health and development. The methods used are designed with full open access and operational application in mind, using transparent, fully documented and peer-reviewed methods to produce easily updateable maps with accompanying metadata and measures of uncertainty.

WorldPop works closely with governments, UN agencies and other international organizations to undertake analyses and strengthen capacity. The group currently consists of over 30 data scientists, statisticians, demographers, geographers, ecologists, epidemiologists and support staff, and receives funding from the Bill and Melinda Gates Foundation, UK Foreign, Commonwealth and Development Office, World Bank, WHO, GAVI, Clinton Health Access Initiative, CIFF and Wellcome Trust, among others. Key long-term partnerships and collaborations include UNFPA, UNICEF, Google, Facebook, Microsoft, Vodafone, ESRI, WHO, FAO, UN-OCHA and WFP.

Acknowledgements

The authors would like to thank Andy South and Anelda van der Walt for their initial input on this project and Claudio Fronterre for his support. Additionally, the authors would like to thank the support received by the PMO team at WorldPop and Wellcome for the funding received for this project, especially to Isabel Fletcher and Talia Caplan.

References

Alberts, M., Dikotope, S. A., Choma, S. R., Masemola, M. L., Modjadji, S. E., Mashinya, F., & Byass, P. (2015). Health & demographic surveillance system profile: the Dikgale health and demographic surveillance system. International journal of epidemiology, 44(5), 1565-1571.

Alex Nginyo Hinga, Sassy Molyneux, and Vicki Marsh (2021) 'Towards an appropriate ethics framework for Health and Demographic Surveillance Systems (HDSS): learning from issues faced in diverse HDSS in sub-Saharan Africa', BMJ Global Health, 6(1), p. e004008. Available at: https://doi.org/10.1136/bmjgh-2020-004008.

Bondarenko, M., Bonnie, A., Tatem A.J. and Tejedor-Garavito, N. lpsmapper: Portal - Piloting the collection of geographic information from Longitudinal Population Studies (LPS) for future climate-health research (v1.0). Zenodo. https://doi.org/10.5281/zenodo.8256800 (2023)

Bonnie, A., Bondarenko M., Tatem A.J. and Tejedor-Garavito N., 2023. Geographic information from Longitudinal Population Studies for INDEPTH and CHAMP sites. Version 1.0. WorldPop, University of Southampton. doi: 10.5258/SOTON/WP00760. https://data.worldpop.org/repo/prj/LPS/v1_0/.

Bos, E. (2004) 'Population and Health in Developing Countries. Volume 1, Population, Health, and Survival at INDEPTH Sites. INDEPTH Network. Ottawa: International Development Research Centre, 2002, pp. 356, (PB) $40.40 ISBN: 0-88936-948-8.', International Journal of Epidemiology, 33(4), pp. 916–917. Available at: https://doi.org/10.1093/ije/dyh220.

Bridger, E. (2021). Map of Soweto, in Young Women against Apartheid: Gender, Youth and South Africa's Liberation Struggle (p. Xiv). Woodbridge, UK: Boydell & Brewer.

Child Health and Mortality Prevention Surveillance (CHAMPS) network. https://champshealth.org/.
GADM. GADM data. https://gadm.org/data.html. Accessed 2022.

George, G., Merrill, R. K., & Schillebeeckx, S. J. D. (2021). Digital Sustainability and Entrepreneurship: How Digital Innovations Are Helping Tackle Climate Change and Sustainable Development. Entrepreneurship Theory and Practice, 45(5), 999-1027. https://doi.org/10.1177/1042258719899425

Herbst, K. et al. (2021) 'Health and demographic surveillance systems in low- and middle-income countries: history, state of the art and future prospects', Global Health Action, 14(sup1), p. 1974676. Available at: https://doi.org/10.1080/16549716.2021.1974676.

IPCC (2022). Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [H.-O. Pörtner, D.C. Roberts, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, B. Rama (eds.)]. Cambridge University Press. Cambridge University Press, Cambridge, UK and New York, NY, USA, 3056 pp., doi:10.1017/9781009325844.

Islam, M.M. et al. (2020) 'Health and Demographic Surveillance System Sites: Reflections on Global Health Research Ethics', Wārasān prachākọn læ sangkhom = Journal of population and social studies, 28, pp. 265–275. Available at: https://doi.org/10.25133/JPSSv28n3.018.

Jia, P., Sankoh, O. & Tatem, A. J. (2015). Mapping the environmental and socioeconomic coverage of the INDEPTH international health and demographic surveillance system network. Health Place. Nov;36:88-96. doi: 10.1016/j.healthplace.2015.09.009. Epub 2015 Oct 27. PMID: 26454610.

Kadandale S, Marten R, Dalglish SL, Rajan D, Hipgrave DB (2020). Primary health care and the climate crisis. Bull World Health Organ. 2020 Nov 1;98(11):818-820. doi: https://doi.org/10.2471%2FBLT.20.252882
Nissan, H., Diggle, P.J. & Fronterre, C. (2022). Combining climate and health data: challenges and

opportunities for longitudinal population studies. Wellcome Trust Report.

Raleigh, C & Urdal, H. (2007) Climate change, environmental degradation and armed conflict. Political Geography, 26(6). 674-694. https://doi.org/10.1016/j.polgeo.2007.06.005

Rhea J Rocque et al. (2021) 'Health effects of climate change: an overview of systematic reviews', BMJ Open, 11(6), p. e046333. Available at: https://doi.org/10.1136/bmjopen-2020-046333.

Romanello, M. et al. (2023) 'The 2023 report of the Lancet Countdown on health and climate change: the imperative for a health-centred response in a world facing irreversible harms', The Lancet [Preprint]. Available at: https://doi.org/10.1016/S0140-6736(23)01859-7.

Romanello M, McGushin A, Di Napoli C, et al. (2021). The 2021 report of the Lancet countdown on health and climate change: code red for a healthy future. Lancet, 2021;398:1619-1662.

Sankoh, O., & Byass, P. (2012). The INDEPTH Network: filling vital gaps in global epidemiology. International Journal of Epidemiology, 41(3), 579-588.

Tatem, A. J., Snow, R. W. & Hay, S.I. (2006). Mapping the environmental coverage of the INDEPTH demographic surveillance system network in rural Africa. Tropical Medicine & International Health, 11: 1318-1326. doi.org/10.1111/j.1365-3156.2006.01681.x

Tomlinson, C.J., Chapman, L., Thornes, J.E. et al. Including the urban heat island in spatial heat health risk assessment strategies: a case study for Birmingham, UK. Int J Health Geogr 10, 42 (2011). https://doi.org/10.1186/1476-072X-10-42

University of East Anglia Climatic Research Unit; Harris, I.C., Jones, P.D. & Osborn, T. (2022). CRU TS4.06: Climatic Research Unit (CRU) Time-Series (TS) version 4.06 of high-resolution gridded data of month-by-month variation in climate (Jan. 1901- Dec. 2021). NERC EDS Centre for Environmental Data Analysis, 2023. https://catalogue.ceda.ac.uk/uuid/e0b4e1e56c1c4460b796073a31366980

Utazi, C. E., Sahu, S. K., Atkinson, P. M., Tejedor, N., & Tatem, A. J. (2016). A probabilistic predictive Bayesian approach for determining the representativeness of health and demographic surveillance networks. Spatial Statistics, 17, 161-178. doi.org/10.1016/j.spasta.2016.05.006

Utazi, C. E., Sahu, S. K., Atkinson, P. M., Tejedor-Garavito, N., Lloyd, C. T., & Tatem, A. J. (2018). Geographic coverage of demographic surveillance systems for characterising the drivers of childhood mortality in sub-Saharan Africa. BMJ Global Health, 3(2), e000611. doi:10.1136/bmjgh-2017-000611

Wellcome Trust (2017), 'Longituinal Population Studies Strategy', Wellcome's Longitudinal Population Studies Working Group. https://wellcome.org/sites/default/files/longitudinal-population-studies-strategy_0.pdf

Glossary

GIS: Geographic Information System
Vector: Geographic data represented by points, lines and polygons
Raster: Geographic data represented by grid cells
GeoJSON: Open standard format to represent simple geographical features, along with their non-spatial attributes.
Shapefile: A common data format for storing vector GIS data
Geopackage: A common data format for storing GIS data
Github: A platform and cloud-based service for software development and version control using Git, allowing developers to store and manage their code
HTML: HyperText markup language
LPS: Longitudinal Population Studies
INDEPTH: The International Network for the Demographic Evaluation Populations and their Health
CHAMPS: The Child Health and Mortality Prevention Surveillance (CHAMPS) network
HDSS: Health Demographic Surveillance Site
Basemap: GIS data and/or imagery that provides contextual background detail for a map
JQuery: JavaScript library
Leaflet: JavaScript library for mobile-friendly interactive maps

Appendices

Appendix 1

- A1a: Original table provided by Claudio Fronterre
- A1b: Maps comparing original digitized site boundaries against the XY coordinates

Appendix 2

- A2a: Boundaries comparisons and updated boundaries.
- A2b: Table summarizing boundary updates and justification.

Appendix 3

- A3a: A csv/excel file with the whole data

Appendix 4

- A4a: LPS Release shapefile first version (Bonnie, et al 2023)
- A4b: Metadata for shapefile columns (Bonnie, et al 2023)
- A4c: Metadata of the headings to include in the portal
- A4d: Documentation detailing the production of the web-portal.

Wellcome supports science to solve the urgent health challenges facing everyone. We support discovery research into life, health and wellbeing, and we're taking on three worldwide health challenges: mental health, global heating and infectious diseases

Wellcome Trust, 215 Euston Road, London NW1 2BE, United Kingdom
T +44 (0)20 7611 8888, E contact@wellcome.org, wellcome.org