Combining climate and health data:

challenges and opportunities for longitudinal population studies.

Hannah Nissan[1], Peter J Diggle[1,2] and Claudio Fronterre[2]

[1] Independent consultant

[2] CHICAS, Lancaster University Medical School, Lancaster LA1 4YB, UK

Team

Peter Diggle

Peter Diggle is Distinguished University Professor of Statistics in the Faculty of Health and Medicine, Lancaster University. He also holds Adjunct positions at Johns Hopkins and Yale Universities, and was president of the Royal Statistical Society between 2014 and 2016.

Between 1974 and 1983 Peter was a Lecturer, then Reader in Statistics at the University of Newcastle upon Tyne. Between 1984 and 1988 he was Senior, then Principal, then Chief Research Scientist and Chief of the Division of Mathematics and Statistics at CSIRO, Australia. He has worked at Lancaster University since 1988, and held a joint appointment with the University of Liverpool from 2012 to 2015. Between 2004 and 2008 he held a UK Engineering and Physical Sciences Senior Fellowship.

Peter's research involves the development of statistical methods for spatial and longitudinal data analysis, motivated by applications in the biomedical and health sciences. He has published 12 books and more than 300 articles in open literature.

He was awarded the Royal Statistical Society's Guy Medal in Silver in 1997 and is a former editor of the Society's Journal, Series B. In 2001 he was elected as a Fellow of the American Statistical Association. He was founding co-editor of the journal "Biostatistics" between 1999 and 2009, and is a Trustee for Biometrika. He has served on the UK Medical Research Council's Population and Systems Medicine Research Board, Training and Careers Group and Population Health Group.

Claudio Fronterre

Claudio is a spatial statistician interested in disease mapping in Low-Resource Settings with a focus on the epidemiology and control of neglected tropical diseases (NTDs) in sub-Saharan Africa.

He obtained his Bsc in Economics and Management and his Msc in Finance from the University of Trento (Italy). He undertook his PhD in Statistics at the Department of Statistical Sciences of the University of Padova (Italy). During this period he also spent over a year at CHICAS, Medical School, Lancaster University under the supervision of Prof. Peter J. Diggle. His research was mainly focused on investigating the effects of and finding corrections for positional errors in geostatistical and point pattern analysis.

Hannah Nissan

Hannah Nissan is an Associate Director in Willis Towers Watson's Climate and Resilience Hub. At the time of writing this report, she was a Research Fellow at the Grantham Research Institute for Climate Change and the Environment at the London School of Economics. She has held academic positions at the Centre for Analysis of Time Series at LSE, the International Research Institute for Climate and Society (IRI) at Columbia University and Imperial College London. From 2018-2020 she was Director of the WHO Collaborating Centre for Early Warning Systems for Malaria and Other Infectious Diseases at IRI, Columbia University, where she was also a Lloyds' of London - Fulbright Scholar and an Earth Institute Fellow. Hannah serves on the American Meteorological Society's Board on Environment and Health and was a founding Steering Committee member of the Global Heat Health Information Network until 2021.

Hannah works on adaptation to climate variability and change, particularly in the health and humanitarian sectors. She specialises in the evaluation, translation and application of climate information for policy- and decision-making and in developing strategies to manage climate uncertainty. She has worked with national and regional meteorological services in the Caribbean and South Asia to tailor and translate weather and climate information for practical use and with health and disaster risk practitioners to incorporate it effectively into practice. She has published widely in the climate science literature and in the cross-disciplinary literature relating to climate services for public health and disaster risk management.

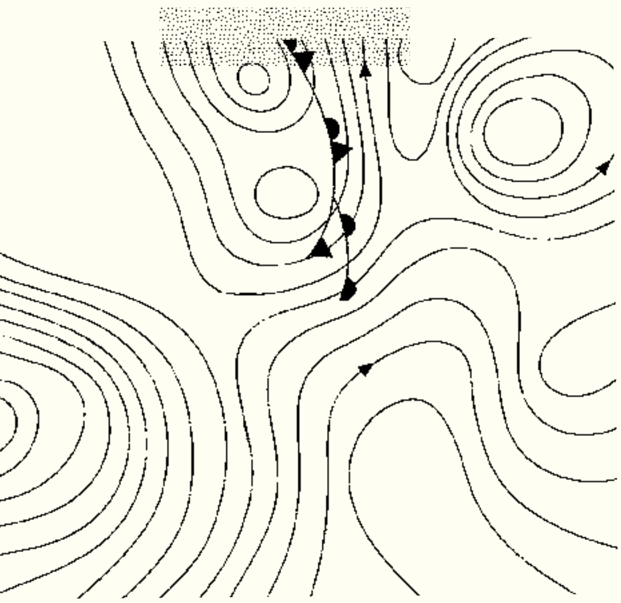
She holds a PhD in Climate Modelling from Imperial College London, a Postgraduate Diploma in Economics from the University of Cambridge and a BSc in Physics from the University of Bristol.

Index

Executive summary

This report examines the feasibility and potential value of longitudinal population health studies (LPS) to address the climate-health research challenge. The term LPS embraces a variety of study-designs, including cohorts, panel surveys and biobanks. Collectively, these provide information ranging in scale from molecular-level biomarkers to population-wide social, environmental and health outcome data. They have a global reach, albeit with under-representation of low and middle income countries (LMICs) whose populations are among the most vulnerable to the impacts of climate on population health. Climate data exist in a range of formats, including direct measurements of key climate variables (principally temperature and precipitation) from networks of weather stations and a variety of gridded data products derived from statistical and physics-based models, some of which are updated in near real-time. Our central proposition is that there is a clear and increasingly urgent need to combine health and climate data to better effect than is currently the case by stimulating closer collaboration between the two research communities.

In the remainder of the report, we first ask what data and methods are required to capture climate exposures and their health effects? We then consider how existing data sources and methods can best be used for this purpose and what changes could be implemented in new data collections. Finally, we offer conclusions and recommendations for how an ambitious climate and health research strategy, and the data to support it, could capitalise and improve upon existing datasets.



What data and methods are required to capture climate exposures and their health effects?

The climate exhibits variability across a wide range of temporal and spatial scales, from hours to centuries and from streets to continents. Climate can affect health and its upstream determinants across the full range of eco-epidemiological levels of organisation, from the molecular to the individual, community and population levels. Thus, research at the climate-health interface requires a multi-scale, multi-variate and multi-disciplinary approach: different aspects of climate affect different health outcomes both directly and, indirectly, via a combination of biological, environmental and socioeconomic factors.

These multiple scales of variation determine the structure of the data required for a comprehensive study of climate and health interactions. They also offer opportunities to build a deep knowledge of climate's role as a driver of health outcomes by adopting a whole-system approach, using evidence from multiple sites and across different timescales to capture the extent and magnitude of the climate's impact on health outcomes, corroborate findings, mitigate problems of inadequate data and support practical adaptation and resilience.

How can existing data sources and methods best be used?

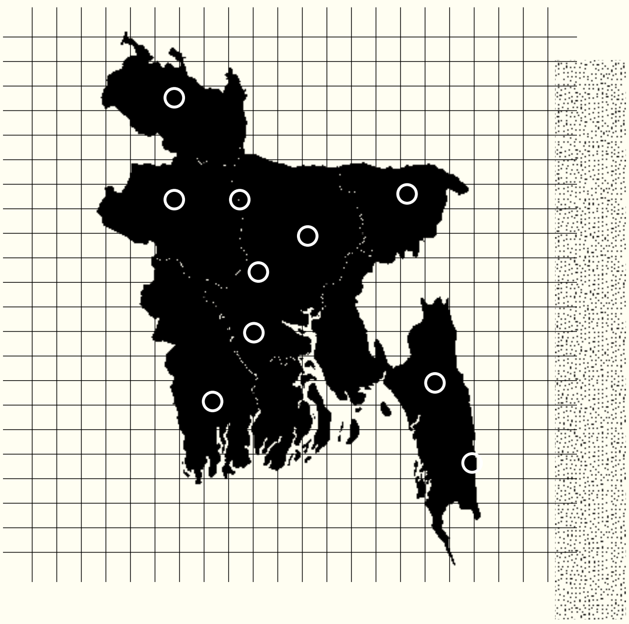
There is considerable potential for combining data from existing longitudinal population studies with existing climate data and data products for two purposes: 1) exploratory, hypothesis-generating research; 2) identification of the most suitable sites for the design and execution of new, confirmatory and in-depth studies focused on specific scientific hypotheses.

These aims can best be achieved through LPS that are both large (to achieve statistical precision) and geographically extensive (to span a wide range of climate exposures and capture their full variability).

The need for new studies, or for introducing changes in data collection protocols of existing LPS, follows primarily from the mismatch between study-designs from the two domains, especially with regard to their temporal and spatial coverage. Most existing longitudinal population studies currently collect data from individual participants at time-intervals of a year or more, whereas many of the most important climate-related exposures that threaten population health occur at seasonal or higher time-frequencies. Conversely, most climate data, i.e. direct measurements of climate variables such as temperature or precipitation, are collected at high frequencies in near-real-time but from spatially sparse networks of weather stations. To address this challenge, the epidemiological community has defaulted to using the outputs of statistical or physics-based models (including remote sensing and reanalysis) which may be combined with station data. These data products may have improved spatial coverage (including global) and resolution, and are often more readily accessible to researchers. However, their use in research is often undertaken without consideration of their inherent uncertainty or suitability for the specific health research question or operational function under investigation.

Moreover, epidemiological modelling usually employs average or accumulated values as covariates (such as daily or monthly average temperature or accumulated precipitation), which can mask many of the exposures most closely related to health outcomes, such as extreme weather events or seasonality (depending on the averaging period). Close collaboration between epidemiologists and climate scientists is needed to construct climate metrics that reflect those aspects of weather and climate most relevant for particular health impacts, and to reach a shared understanding of when and why existing data products are inadequate for the task in hand.

Novel study-designs and statistical methods will be needed to enable studies to synthesise information from multiple sources that record health and climate data at different spatial and/or temporal resolutions. For these reasons, it is essential that each such study be conducted from the outset by teams whose expertise spans the climate, health, environmental and statistical sciences. This would be greatly facilitated by the establishment of a multi-disciplinary centre of excellence in climate-health research with a global reach and a focus on policy-directed research questions.



Conclusions

There are opportunities to leverage both the general framework of longitudinal population studies and the information from existing LPS, which collectively have a global reach and provide multidimensional data from micro (subcellular) to macro (socioeconomic) levels, for climate-health research. However, to realise fully these opportunities require further investment to understand what changes can be made to existing data collections and how new LPS should be designed for better alignment between the spatial and temporal scales of the climate hazards and individual health outcomes. Specifically, understanding acute health outcomes requires the annual follow-up schedules typical of existing longitudinal population studies to be supplemented by intra-annual data. Capturing the spatial variability in climate-related exposures requires data at finer resolutions, in key locations, than are currently provided by existing meteorological infrastructure. Existing data-collection protocols from both domains need to be supplemented by data from a series of factorial experimental designs that collectively cover the important dimensions of variability in exposure and, consequently, health outcomes.

As well as investment in data collection infrastructure there is a need for more researchers with the expertise to handle the information effectively. Advancing a deep understanding of climate-health interactions and, crucially, using this understanding to generate policy-relevant and operationally-relevant research, requires more highly trained – we suggest at least to PhD level – experts in the health, climate, environmental and statistical sciences.

The socioeconomic factors that mediate climate's effects on health outcomes are a major gap in current knowledge. Whilst we have been unable in this report to investigate the social dimension in detail, we believe that an equally high level of social science expertise is needed to understand the role of socioeconomic factors in mediating climate-health pathways and thus to plan interventions at community-population levels; see Appendix C by Prof Stephen Reicher.

Finally, low-and-middle-income countries (LMICs), which mostly fall within the tropics, are simultaneously among the most vulnerable to the effects of climate change on health and the least able to afford the necessary policy responses. Priority should therefore be given to supporting research with an LMIC focus. This research can benefit from the higher levels of climate predictability (associated with the El Niño Southern Oscillation) that are found closer to the equator.

6.2. Recommendations

To work towards these ambitions, we identify below a set of specific recommendations for activities that the Wellcome Trust could undertake.

Immediate: Assess Existing Health Datasets For Climate Analysis

A metadata analysis is the first step in determining the suitability of existing health datasets for climate analysis. Given the complexity of the data required to capture climate-health effects on different spatial and temporal scales, datasets which at first appear suitable for epidemiological research can transpire to be incompatible upon further examination. The survey of LPS presented in Section 4 provides an initial assessment of the potential for using existing LPS for climate analysis, but a deeper exploration is required, which would be greatly assisted by a digital platform to visualise key metadata.

Below, we list some of the key metadata to be collected. Note that, although data are often aggregated in space and time to achieve adequate sample sizes, data collection is usually staggered and more precise temporal and geo-referencing of each observation may sometimes be available.

General attributes

- LPS type: cohort, panel, repeated cross-section
- Are the timing and location of data recorded?
- Sampling design
- Number of participants
- Number of sites/study regions

Spatio-temporal attributes

- Temporal coverage
- Frequency of surveillance
- Precision of temporal referencing
- Timing of surveillance during the year
- Geographical coverage
- Spatial resolution
- Precision of geo-referencing

Recommendations

Short term

1. Fund proposals on the following topics under existing grant and fellowship schemes:

   a) secondary analyses of existing LPS and climate data to develop hypotheses and inform the design of studies on specific climate-health interactions;

   b) projects that capitalise on opportunities for the integrated analysis of data from multiple LPS;

   c) development of novel statistical and computational methods for inferentially robust combined analysis of multiple health and climate data-sources;

   d) projects to support better understanding of the indirect drivers in climate-health pathways and better linkage with relevant data types e.g. socio-economic census data;

   e) projects to construct new retrospective cohorts and corresponding climate data and metrics;

   f) projects that engage with national health and meteorological agencies to enable all relevant data from both domains to be harnessed for climate-health research at local scales.
2. Commission selected LPS consortia to consider how they could re-orient some of their work towards climate-health research, engage directly with climate data owners and scientists and develop specific proposals accordingly. Candidates could include the Hundred thousand Plus Cohorts Consortium, AGRICOH, HELIX and the successful bidder for the African Population Cohorts Consortium.
3. Engage in discussion with Brazil 100M and INPE with a view to developing an exemplar real-time climate and health surveillance system based on country-wide, routinely collected health information.

Medium term

1. Encourage new consortium-based approaches that integrate climate data and health data across wide-ranging geographies and are co-designed by experts from both domains.
2. Engage the disaster risk community to develop funding opportunities for pilots to explore forecast-based surveillance as a means of studying the causal pathways between extreme weather or climate events and health outcomes, and to evaluate the effectiveness of interventions.
3. Advocate to the World Meteorological Organisation for the inclusion of the health sector as a priority user for the services of national meteorological agencies and propose specific meteorological data and services requirements needed to address climate-health priorities (e.g. in urban areas).

Long term

1. Develop a vision for a Wellcome Trust Climate and Health Institute with global reach but a particular focus on policy-directed research in LMIC settings. The INDEPTH network would be a useful starting point for this activity; complementing its health data system with an equally rich climate data system would create a formidable resource for climate-health research rooted in LMICs.
2. Develop a strategy for the generation and use of routine health information systems to capture and analyse real-time or near-real-time health data in lower income countries.
3. Develop the design for a network of sentinel sites taking frequent health, socio-economic and climate measurements across representative climatic regions/ exposures and socioeconomic contexts, with a view to this platform serving a multi-disease research and operations agenda.

Introduction

Introduction

Climatic processes can affect population health in different ways. Extreme climatic events, such as floods or heat-waves, can have acute impacts that disproportionately affect low-latitude countries (Harrington et al, 2016, 2017) and, within those countries, disproportionately affect the most vulnerable sectors of the population (Green, 2016). Changes in climate can also have long-term indirect impacts on health, for example through their effects on food production or on the geographical range of vector-borne diseases, that are harder to spot in the short-term but potentially more profound in their consequences for global population health (Thomson and Mason, 2019). Furthermore, these consequences extend far beyond their increasingly recognised effects on mortality (Peng et al, 2011). We believe that in order to gain a deeper understanding of present and future climate-related threats to human health there is both a need and an opportunity to build closer working relationships between two research communities: climate scientists who study the physical processes that drive the earth's climate; and epidemiologists who conduct longitudinal studies of population health outcomes.

This report, commissioned by the Wellcome Trust, focuses on the feasibility and potential value of marrying climate science and longitudinal population health studies (LPS) so as better to assess the impacts of short- to long-term climatic processes on a variety of health outcomes that operate on a range of temporal and spatial scales (e.g. extreme weather events, seasonality, interannual variability, multi-decadal cycles and long-term trends 30 to 80 years into the future). It is intended to be used as a high-level guide for two audiences: the Wellcome Trust, to develop its overall climate-health research strategy, draft funding calls and evaluate proposals; and researchers writing climate and health proposals.

Most current research at the intersection of climate and population health involves climate and health scientists working independently in a chain-like modality. Undesirable consequences of this include the following:

Health scientists typically use data and forecasts/projections from climate scientists as unquestioned inputs to health modelling, failing to recognise their inherent and often unquantifiable uncertainty, the nuances of the different types of climate data available and the different methods of preparing these data for analysis with health data. There is something to be said for reserving the term data in its standard statistical usage to refer to observed quantities; "data" that are derived from observations, for example raster images of model-based predictions, might better be called predictions, or products.

Climate scientists build products for general use without considering the specific aims of the health studies that will use these products.

Health data are rarely collected in a way that facilitates analysis with climate data, and vice versa. In particular, careful consideration of the spatial and temporal scales for data-collection are critical in both domains, and for practical reasons are often incommensurate.

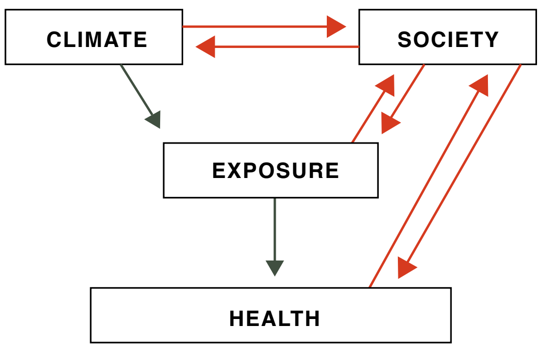
Standard methods of time series analysis require climate and health time series to be regularly structured in time, with sampling at the particular frequencies relevant to the climate event/health outcome of interest. Equally, the required spatial resolution for data-collection is context-dependent, as both climate and health phenomena typically display variation at a range of spatial scales, not all of which are relevant for every health-climate research question.

The advancement of new technologies and methods in artificial intelligence/data science creates opportunities but also risks. Such methods offer efficient means of “mining” massive datasets, such as electronic health records, to detect associations between variables. They are increasingly employed to conduct climate analyses across a range of sectors including health. However, in the absence of a strong hypothesis regarding the pathway(s) by which climate could affect human health, the complexity and high dimensionality of health and climate data increase the likelihood of detecting spurious associations that fail to address causality (Shmueli, 2010). Furthermore, the multiple temporal and spatial scales of climate variability require an understanding of climate science to ensure that epidemiological models either account for the full range of climate variability, where data allow or, failing this, at least acknowledge their limitations, particularly when they are used to inform policy. These considerations favour adopting a hypothesis-driven approach to exploring climate impacts on data from longitudinal population studies (LPS), which in turn emphasises the need to draw on a combination of expertise from the health, climate, environmental and social sciences (Figure 1).

For these reasons, it is far from straightforward to make use of existing LPS and existing climate science products fully to understand the role of climate in determining health outcomes. Nevertheless, existing LPS should be utilised where possible but we should also consider how we could modify and/or combine existing studies to achieve better temporal and spatial coverage of important health outcomes. Making use of existing datasets is important, not least because resource constraints are particularly severe in LMIC settings where, as noted above, the impacts of climate change on population health are likely to be most severe. However, we also need to ensure that new studies are designed in such a way as to deliver the best possible understanding of health-climate interactions.

We note also a wider issue that we have not been able to investigate in detail, but highlight as a key priority area for research, namely the importance of social factors involved in both the causes of climate change and its consequence for the health of populations. Figure 1 contrasts the conceptual simplicity of the direct causal pathway from climate to exposure to health outcome (black arrows) with the complexity of interactions involving society at large (red arrows). Understanding the direct causal pathway may be sufficient for aetiological research but understanding the interactions with society is critical to the success or failure of particular adaptation/mitigation strategies suggested by policy-directed research.

Figure 1
Climate, health and society form a single, inter-linked system containing multi-directional causal pathways.



The remainder of our report is structured as follows:

In sections 2 and 3 we address the following overarching questions:

1. What attributes of the climate are important to understand for health research?
2. What data are required to capture climate exposures and their effects on health outcomes?
3. What methodological challenges are involved in combining climate and health data for research?

In Section 2, we describe the key attributes of the climate, and climate data, that one must understand in order to assess the feasibility of combining longitudinal population studies with climate data to address a range of research questions relating to climate and health.

In Section 3 we present a number of challenges (and solutions) for effective research at the climate-health interface.

In Section 4 we summarise the results of a survey of existing LPS intended to assess the suitability of existing datasets for addressing a range of climate-health research questions across multiple spatial and temporal scales.

Section 5 is a discussion section, in which we take a broader perspective and argue for a multi-scale, systems approach to developing a climate-health research and data strategy.

In Section 6 we present our conclusions and put forward recommendations for activities that the Wellcome Trust could engage with in the short, medium and long term.

Three appendices are attached to this report.
Appendix A is a list of people we have consulted individually while working on this report. We thank all of them, and the participants at a virtual workshop organised by the Wellcome Trust on 4th August 2021, for taking the time to share their insights with us. We apologise in advance for any unwitting misrepresentations on our part.
Appendix B is a summary of the climate-health research priorities identified by the people we spoke with. We also thank Prof Stephen Reicher (University of St Andrews) for his initial thoughts on societal issues, which constitute **Appendix C** of our report.

2

Climate as a health exposure: what attributes of climate are important for health research?

Climate as a health exposure: what attributes of climate are important for health research?

2.1 Temporal scales of climate

In the context of this report, climate can perhaps best be understood as a process that generates the statistics of weather. Average seasonal rainfall is a weather statistic that measures one aspect of climate; others include the frequency of extreme rainfall events in a season, or over a few years. In this sense, climate captures the slowly varying parts of the climate system, but it is inextricably linked to the weather that we experience, which fluctuates from hour to hour and from day to day. Health impacts are generally associated with variability in weather and climate, which exists on multiple timescales or frequencies. These timescales provide an indication of the frequency of data required to detect the impacts of weather and climate variability on human health (Table 1). The tails of the distributions are particularly important as extreme events like storm surge, heat waves, droughts and floods can have severe consequences for human health. These events are similarly associated with characteristic durations and spatial extents (Table 2); we discuss spatial scales of climate in Section 2.2.

In most of the world, the most important timescale of climate variability is the seasonal cycle, which makes it critically important to conduct health surveillance at appropriate and consistent times of the year. In the longer term, large swings in the prevailing climate occur from year to year as well as still slower cycles over multi-decadal (10-30 year) periods.

Table 1

Timescales of climate variability and associated data requirements for studying climate-health associations

Timescale of variability	**Weather**	**Sub-seasonal**	**Seasonal**	**Inter-annual**	**Multi-decadal**	**Long-term trends**
Time period over which variations occur	Hours to several days	1-4 weeks	1-12 months	1-10 years	10-30 years	>30 years
Frequency of health data required	Sub-daily to daily	Pentads-weekly	1-3 monthly	Annual, at the same time each year	Annual, at the same time each year	Annual, at the same time each year
Length of climate and health timeseries required for analysis	Depends on variability in weather and strength of signal in health data. Extreme weather events require longer time series to obtain adequate sample sizes.	Several years.	At least several years depending on strength of seasonal signal vs. interannual variability.	At least 10 years, more for causal modelling depending on strength of signal.	At least 30 years, more for causal modelling depending on strength of signal (usually unfeasible in practice).	At least 50-60 years to detect trends. Causal modelling of trends likely unfeasible given data availability.
Considerations	Diurnal weather variability means that exposure data (and potentially health data for highly variable outcomes) must be at the appropriate time of day. Associations may differ in different seasons, or during different phases of climate variability (e.g. during ENSO events or different phases of decadal oscillations).	Several years required to account for inter-annual variations.	Several years required to account for inter-annual variations. Causal modelling to explain seasonal variability in health data would require many years of sub-annual data but average seasonal cycles could be estimated with fewer years.	Seasonality means that the timing of data collected during the year matters: health data must be at the appropriate time of year to capture the impacts of climate, and successive years require observations at the same times of year.	Annual data needed to avoid confounding by inter-annual variability. Frequent measurements are also needed to discern when different phases begin and end. More years are required to detect decadal variations in regions with high interannual variability.	Annual data needed to avoid confounding by inter-annual variability. More years required to detect trends in regions with high interannual and decadal variability.

Climate-health research can capitalise on these swings in climate that expose large areas and multiple regions to anomalous meteorological conditions at similar times, driven (at least in part) by physical processes in the climate system that are predictable to varying degrees. The most important physical driver of interannual climate variability is the El Niño Southern Oscillation (ENSO), the periodic warming and cooling of the eastern and central equatorial Pacific Ocean and accompanying changes in atmospheric circulation. An El Niño event occurs when there is anomalous warming across this region of the Pacific, while a cooling episode is called a La Niña event. El Niño events typically start in April/May and last 9-12 months, though they can occasionally persist up to 2 years. They recur approximately every 3-10 years, but this frequency itself has varied over recent decades and even over centuries (Thomson and Mason, 2018). The changes in atmospheric circulation accompanying El Niño and La Niña events result in anomalous rainfall across the tropics, particularly but not exclusively in the areas surrounding the Pacific Ocean. Less is known about how regional temperature is affected by ENSO events, but a measurable increase in global average temperature is observed during El Niño events, and a cooling is observed during La Niña (Thomson and Mason, 2018).

Natural cross-timescale variability is superimposed over a background of rising atmospheric concentrations of anthropogenic greenhouse gases and aerosols, which translate into background, non-linear trends in climate (Figure 2). To date, seasonal and interannual variability have dwarfed climate change trends in rainfall in much of the world. Long-term temperature trends are much stronger than rainfall trends and are detectable above background variability in most regions, but the seasonal and interannual timescales remain the dominant source of temperature variability in most places. Society experiences the impacts of long-term trends via the attendant shorter-term fluctuations in weather and climate, so climate change and climate variability are inextricably connected. Warming trends alter the baseline, and therefore the intensity of heat waves and anomalously warm years (Lyon et al, 2017). Long-term trends in sea level impinge on coastlines gradually, but the greatest impacts are experienced when storm surge occurs on top of this elevated base sea level, bringing more intense floods and salinisation than previously experienced. Climate change can also change the frequency and duration of weather and climate events as well as their timing during the year (Seneviratne et al, 2021), which can alter health impacts because of the seasonal nature of many socioeconomic drivers of health, such as employment and food systems.

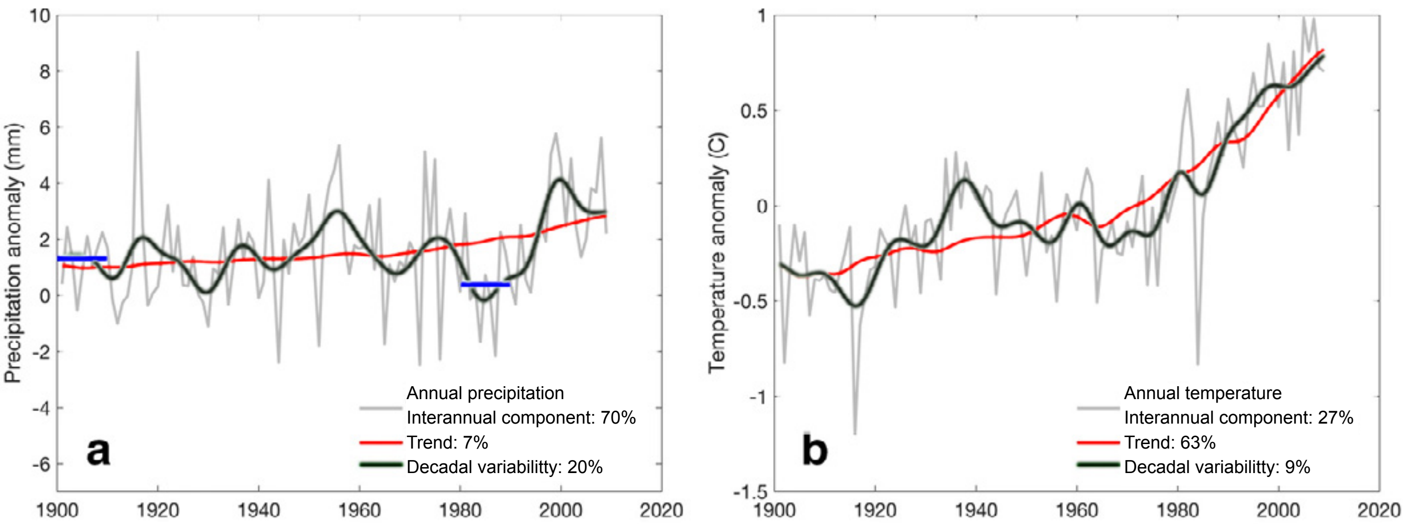
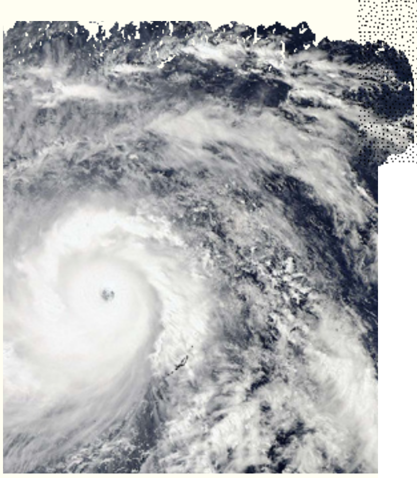


Figure 2.
Timescales of variability for global average annual precipitation (A, mm) and temperature (B, °c) anomalies. Raw annual averages are shown in grey, fitted decadal cycles in black and the long-term trend in red. In (a), the horizontal blue lines show 10-year averaged precipitation anomalies from 1900-1910 and 1980-1990, illustrating that interannual and multi-decadal variability can confound trend detection if not accounted for. The legends indicate the proportion of total variance (%) in precipitation and temperature explained by each timescale (note seasonality is not included). Reproduced from Nissan, Ukawuba and Thomson (2021). For the methodology and data, see http://iridl.ldeo.columbia.edu/maproom/global/time_scales/index.html .

Table 2
Characteristic durations and spatial extents of weather and climate events

Weather / climate event	Duration	Spatial extent
Strong winds	Hours - days	Wide variation depending on the driving weather system / mechanism.
Tornado	Hours, but they move quickly so can pass over any one location in a very short time.	Tens of m to 3 km
Floods (riverine, flash, coastal)	Hours - days	Flash floods: < a few hundred m Riverine/coastal floods: < a few hundred m inland from coastline or river banks.
High temperatures / heat waves	< 2 weeks	Absolute values vary on small scales according to land cover (urban areas are hotter) and elevation (high altitudes are cooler); temperature anomalies correlated over hundreds of km up to about 1500 km.
Drought	Months to years	Hundreds to thousands of km
Extreme winter conditions (cold, ice, snow, wind)	Days to months	Varies by parameter
Heavy precipitation	Hours to days depending on the driving weather system.	Rain occurrence: a few hundred m to a few hundred km, depending on the driving weather system. Amount of rain/snow: much less spatially correlated.
Tropical storms (cyclones & hurricanes cause strong winds, heavy rain and storm surge)	The whole system will last several days. Associated rain and wind are experienced for shorter periods depending on how quickly the storm moves once it makes landfall.	Weather varies dramatically over small distances within a tropical storm.



The path of typhoon Haiyan

Typhoon Haiyan was a tropical cyclone that affected the Philippines in South East Asia in November 2013. It was one of the strongest tropical cyclones ever recorded with winds of 313 km/h. In some areas, 281.9 mm of rainfall was recorded, much of which fell in under 12 hours. Waves of up to 7 m in height battered the coast. The whole system travelled approx 4500 km over 8 days (BBC Bitesize).

Nov. 2:

The storm is detected as a low-pressure area in Micronesia.

Nov. 4:

The system is upgraded to a tropical storm and named Haiyan.

Nov. 6:

It hits Palau and parts of Micronesia. After growing in intensity for days, Typhoon Haiyan became a Category 5 storm, with wind speeds above 157 mph.

Nov. 7:

Haiyan enters the Philippines area; alerts, preparations, and evacuations intensify.

Nov. 8:

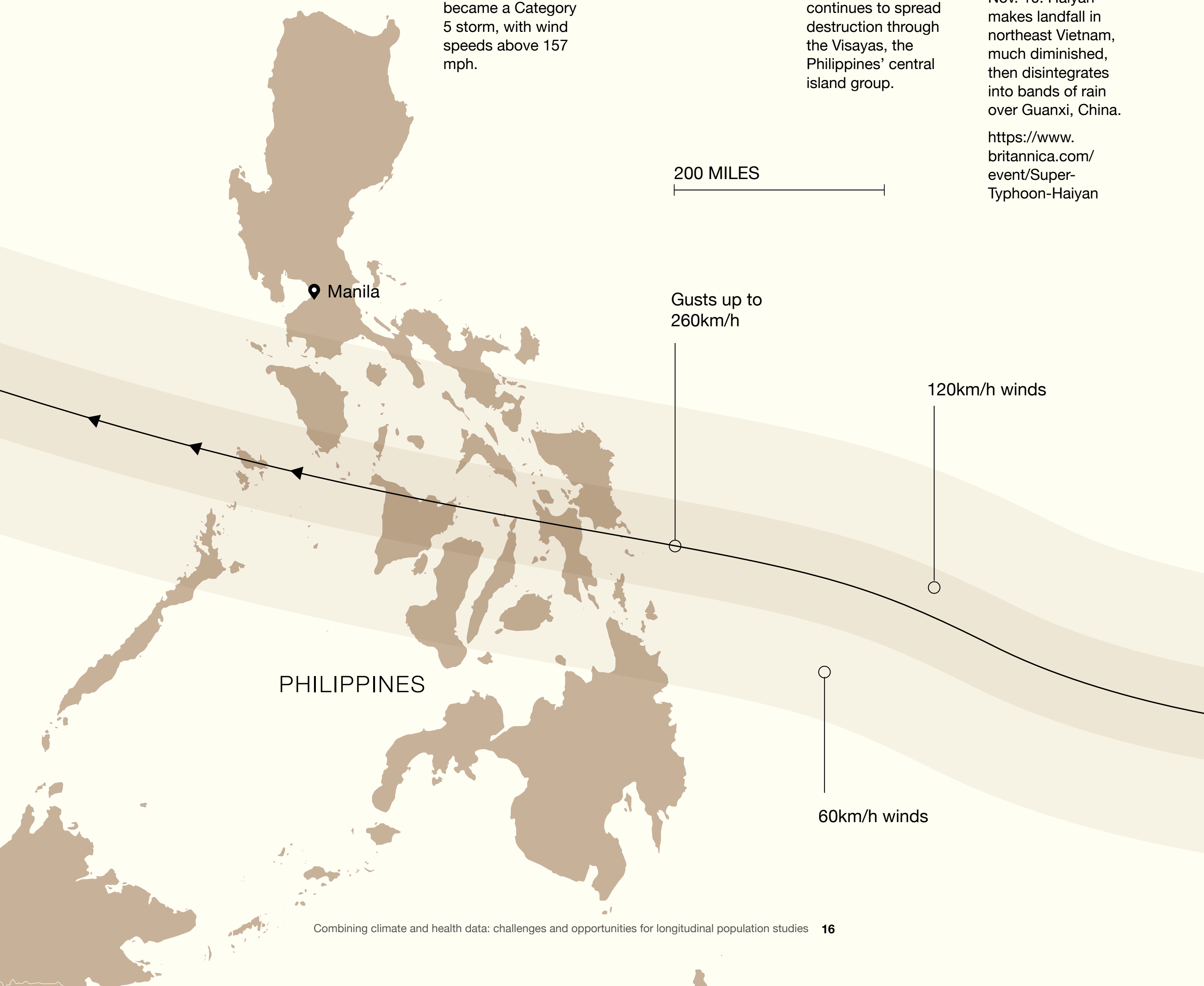
At 4:40 a.m., Haiyan makes landfall in Eastern Samar at peak capacity. It continues to spread destruction through the Visayas, the Philippines' central island group.

Nov. 9:

The storm moves out into the South China Sea, heading toward Vietnam.

Nov. 10: Haiyan makes landfall in northeast Vietnam, much diminished, then disintegrates into bands of rain over Guanxi, China.

https://www.britannica.com/event/Super-Typhoon-Haiyan



2.2. Spatial scales of climate

Spatial variations in climate are driven primarily by geography. It follows that in regions with little geographic variation, the climate also tends to be fairly homogeneous, and vice versa. Specifically, altitude, latitude, land-sea contrasts and land-cover differences are the key factors driving spatial variations in climate (Thomson and Mason, 2018). To take high temperatures as an example, major differences in heat exposure occur between areas of high and low elevation, between the tropics and the extra-tropics, between coastal regions and continental interiors and between urban centres and rural areas. These differences are not simply a matter of different average readings on a thermometer (Laaidi et al, 2012) but of differences in temporal variation that could alter the physiological impact of heat as well as the methods we would choose to combat the problem.

Major factors driving spatial variation in climate.

As a case study, the text below each driver describes how that factor influences heat exposure.

Altitude:

Temperature decreases rapidly with altitude at a rate of about 1° per 100m. Given the sparse network of weather stations in much of the world, this heterogeneity poses a major challenge to exposure estimation for epidemiological modelling in mountainous areas; for example, Figure 3 illustrates the sparse and very uneven distribution of weather stations across Kenya. More complex, but equally rapid, changes in precipitation are observed with increasing altitude.

Latitude:

Tropical regions are hotter year-round than the extra-tropics. However, tropical countries also experience much smaller ranges in temperature, both from day to day and across the year – there is very little seasonal variation in temperature close to the equator (Figure 4). Background conditions in the tropics are consistently warm, with heat waves characterised by relatively small increases in temperature, perhaps accompanied by surges in humidity. At higher latitudes there are marked seasons, and the weather fluctuates much more from day to day, resulting in heat waves that expose people to heat levels far outside the comfortable background conditions they are used to, although the absolute temperature during heat waves may remain below the extremes seen in the tropics and sub-tropics.

Land-sea contrast:

The presence of water restricts the range of temperatures experienced in coastal areas and islands, but continental interiors can see huge swings in temperature.

Land-cover type:

Differences in land surface drive smaller-scale variations in temperature between large cities and their rural surroundings and within cities themselves. People who live in urban areas are exposed to higher temperatures than those who do not. The difference is observable year-round, but is most marked at night – and hot nights have consistently been associated with heat-related mortality (Laaidi et al, 2012; Rooney et al, 1998; Nissan et al, 2017; Karl and Knight, 1997).

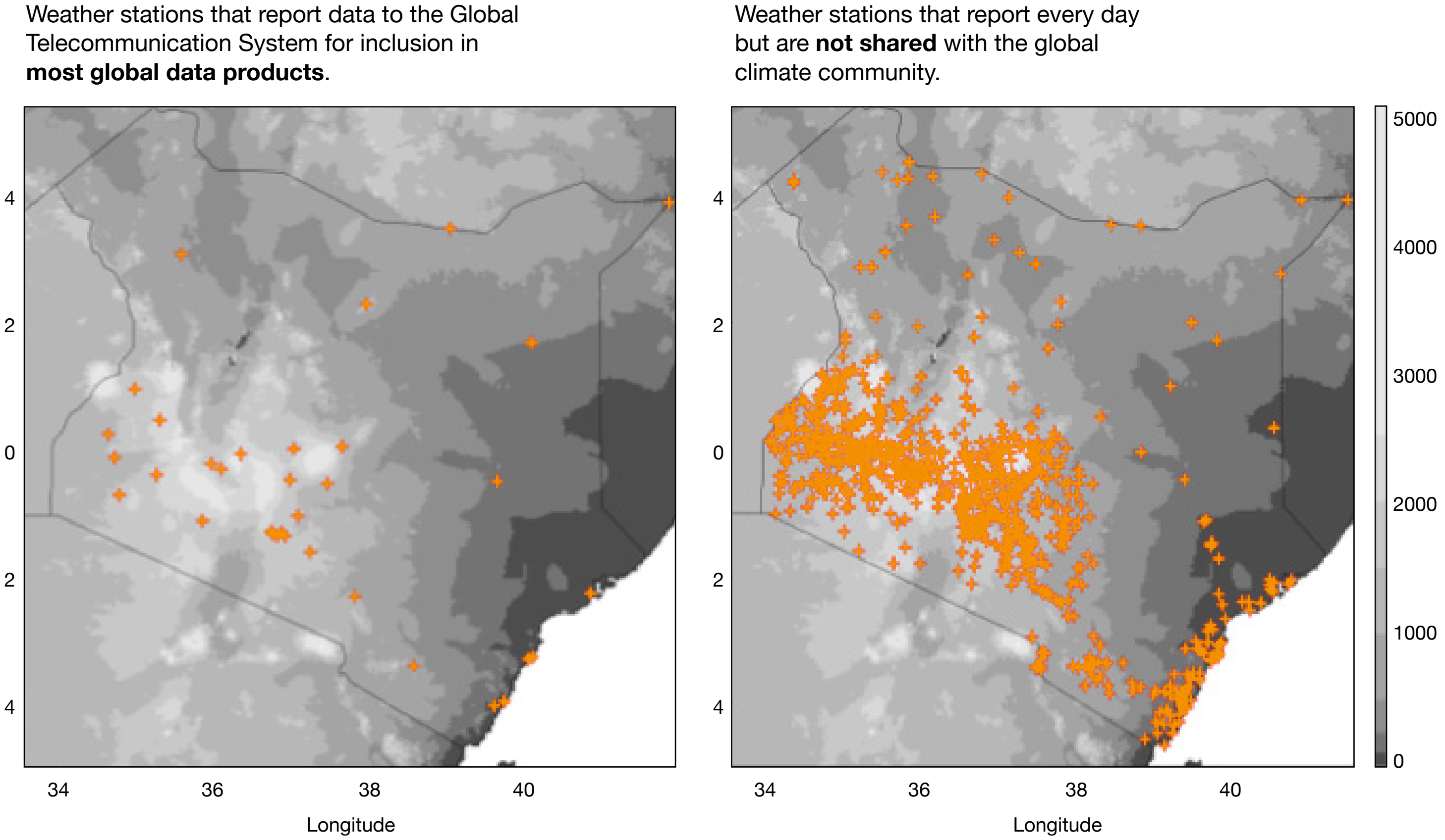
The degree of spatial variability in meteorological parameters determines the resolution of meteorological and health data required to detect relationships between climate and health. Temperature can vary by several degrees over very short distances because of changes in land-cover (cities are hotter than the countryside) and altitude (temperature decreases as you climb a mountain). For precipitation, the extent of spatial correlation depends on the driving weather system. Rain and snow can be caused by local heating (convective precipitation), which is highly localised and occurs year-round in the tropics and during the summer in mid-latitudes. Precipitation in the high-latitudes and in mid-latitude winters is mostly caused by the large-scale movement of air, which produces fronts along which precipitation is heavy and widespread (large-scale precipitation), as well as pushing air over topography to produce rain and snow on the windward sides of mountains (orographic precipitation). Precipitation amount and intensity are highly variable within rain systems (Thomson and Mason, 2019).

Although absolute temperature is very heterogeneous in space, temperature anomalies, which can be the more relevant parameter for health impacts (Vaidyanathan et al, 2016), are fairly uniform over large distances. This assumption holds whether we are talking about a heat wave that lasts a few days or a particularly hot season. The exception is the marked temperature changes observed along weather fronts in the extra-tropics, which only last a few days. Precipitation anomalies are much more localised than temperature anomalies, but their spatial coherence does increase at seasonal timescales; if a particular summer is very wet locally, it is likely also very wet regionally (Thomson and Mason, 2019).

In general, weather and climate events that last longer tend to be spatially more extensive (Table 2). This heuristic can be a helpful guide to the spatial resolution of data required to detect the health impacts of meteorological events. However, it does not tell the whole story because it also depends how we choose to define climate exposures (Section 2.3). Differences between definitions can significantly affect the relevant spatial scale of analysis as well as the frequency with which that event is experienced. Furthermore, weather systems are usually mobile: clouds move with the wind, so small-scale convective showers can affect areas far larger than their size; tornadoes can have an extremely small diameter but travel at great speed, devastating everything in their path. Large storm systems like tropical cyclones contain many different types of weather within them, with high spatial variability in rainfall, wind intensity and storm surge. Nonetheless, some very general rules of thumb are provided in Table 2.

Figure 3.

Density of reporting weather stations in Kenya, with grey shading indicating elevation. The orange crosses indicate the locations of the weather stations. The left panel shows stations that report data to the Global Telecommunication System for inclusion in most global data products. The right panel shows operational stations that report every day but are not shared with the global climate community. Reproduced from Dinku et al (2016)



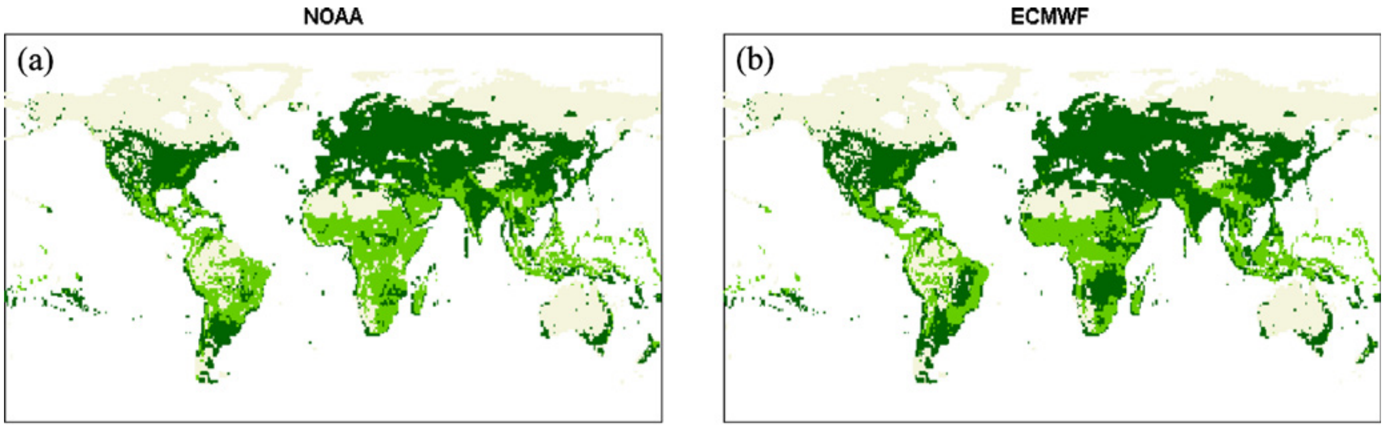


Figure 4.
Seasonality of heatwaves in areas with exposure estimated from two different reanalysis datasets from (a) the US National Oceanic and Atmospheric Administration and (b) the European Centre for Medium Range Weather Forecasting. Dark green indicates regions that show distinct seasonality in heat waves, while light green areas do not have this seasonality. Note that the two models do not show the same results over southern Africa, Latin America and the southern 'Stans. Cream colored areas have low human exposure to heatwaves. Reproduced from Coughlan de Perez et al (2018).

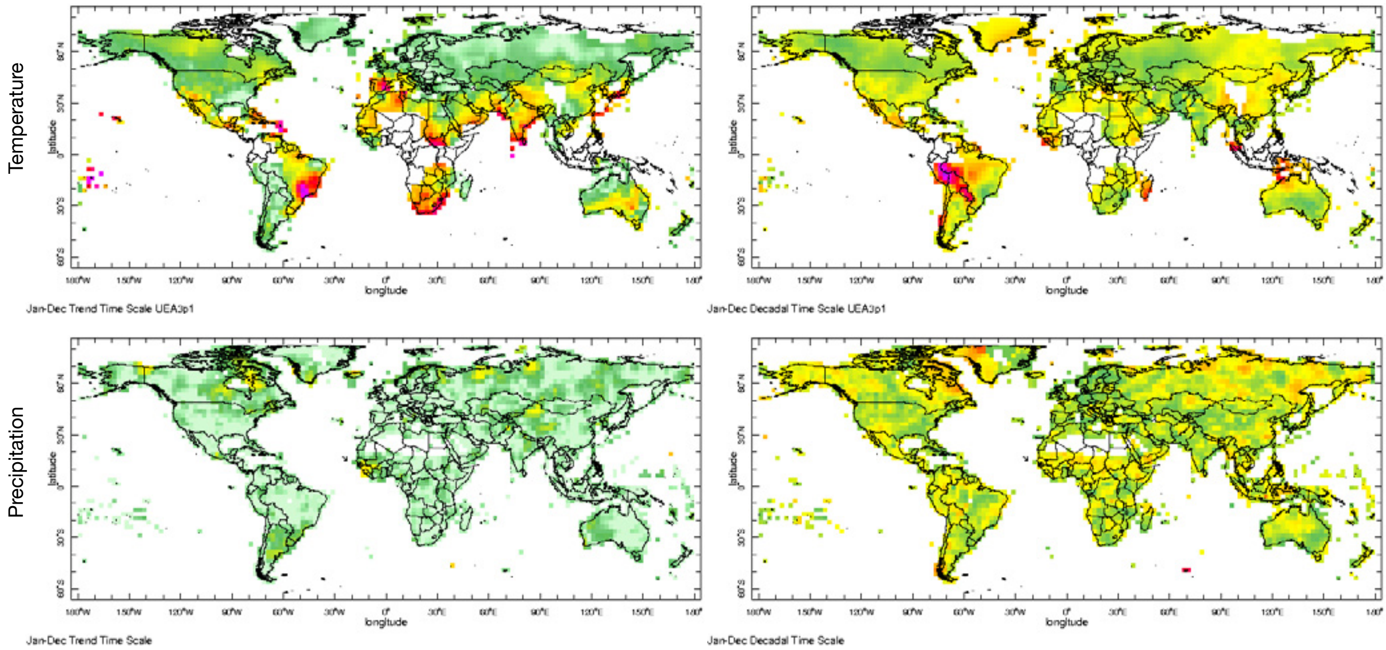


Figure 5.
Timescale decomposition for annual average temperature (top, A & B) and accumulated precipitation (bottom, C & D). Each map shows the percent of total variance in annual temperature/precipitation which can be explained by the long-term trend (left, A & C) and by decadal variability (right, B & D). White areas indicate places where insufficient data were available for robust analysis. See http://mbell.maproomdev.iri.columbia.edu/maproom/global/time_scales/index.html for the methodology, Source: IRI.

Looking beyond the local climate, there exist macro-scale differences in climate between regions of the world. As well as disparities in average climatic conditions, spatial variations in climate can take the form of differences in temporal variability between regions (Figure 5). Climate change trends are not occurring uniformly around the world: instead, we see marked differences, with some regions hardly warming at all while others are heating up at alarming rates. Similarly, decadal fluctuations are only a minor component of the total variation in climate in much of the world, but they play a significant role in others, particularly the Sahel (Pomposi, Kushnir and Giannini, 2014). Although the maps in Figure 5 are for annual temperature and precipitation, in most of the world seasonality is the most important timescale of climate variability. Even in the tropics, where temperature is broadly consistent year-round, precipitation still exhibits distinct seasons that are important for understanding the climate's impact on health.

Modes of climate variability (such as ENSO and the Madden Julian Oscillation, a key driver of intra-seasonal climate variability in the tropics and extra-tropics) affect the regions of the world in different ways, with some areas particularly affected while others show little signal at all. Rainfall anomalies during ENSO events have complex regional and seasonal patterns depending on the large-scale atmospheric circulation. The map in Figure 6 depicts the average rainfall anomalies over many events, but no two events are the same and the changes experienced can differ markedly from these averages.

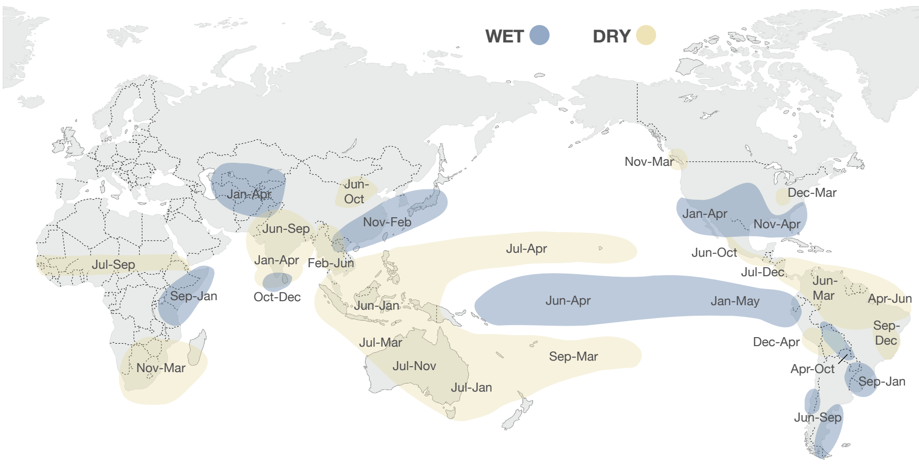
Figure 6

Average regional changes in rainfall by season, during El Niño (top) and La Niña (bottom) events.

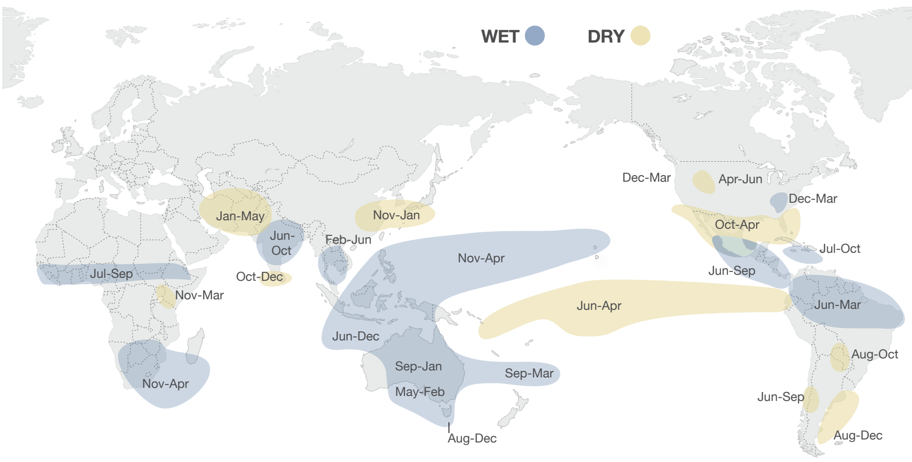
Source: IRI https://iri.columbia.edu/our-expertise/climate/enso/

IRI International Research Institute for Climate and Society
Earth Institute | Columbia University

El Niño and Rainfall



La Niña and Rainfall



2.3. From climate to exposure

Our experience of the climate is through our exposure to the elements – primarily temperature, precipitation, humidity, wind speed, solar radiation and air pressure. All climate and weather exposures, from heat waves to monsoons to the harsh Mongolian winters ('Dzud'), can be described using basic meteorological parameters. Temperature and precipitation are the most widely available variables in meteorological datasets, but other measures can be more important for particular health impacts. For example, an individual's heat load and consequent risk of heat stress involve far more than just temperature: humidity, wind speed and solar radiation all contribute strongly.

Downstream environmental factors, including flooding, air quality, wildfire, dust storms, saltwater intrusion and ocean acidification, which themselves are driven by the climate, can also be directly responsible for health impacts. For example, rainfall is a primary driver of riverine and flash floods, but the hydrology of the area – how water moves through the landscape – ultimately determines whether it will flood when it rains, as well as where, when and how quickly the flood waters will build up. If hydrology were to remain constant, rainfall would be the causal driver, but changes in land use like urbanisation and deforestation alter the relationship between rainfall and flooding, and thus between rainfall and health. Similarly, urban air quality is primarily the result of emissions of particulate matter but is worsened by particular circulation patterns. Finally, the exposures that a population experiences in response to a specific climate event are inherently uncertain; for example, hot and dry conditions promote, but do not inevitably result in, wildfires.

Whether or not meteorological parameters are a reasonable proxy for downstream environmental exposures is highly dependent on the context. Local rainfall, for example, may be predictive of flash flooding, but not of river flooding because the rainwater flowing into river systems comes from a much wider catchment area. Downstream environmental exposures are measured using variables that are not available in climate datasets and their interpretation for use in health research requires domain expertise.

A dzud occurs when a severe winter follows a dry summer, making it difficult for the country's livestock to feed. The animals, already weakened by insufficient summer grazing, and unable to reach sparse grassland buried under snow and ice, risk starving or freezing to death. The catastrophic dzud of 2010 wiped out as many as 11m animals, over 20% of the country's total population.

(Mongolia's deadly winters are becoming more frequent, The Economist 2020)

2.4. Climate data and data products

There are many types of meteorological data products, each with its own set of strengths and drawbacks that have to be weighed up when deciding which to use for health research. The choice involves trade-offs in precision, geographical and temporal coverage, spatial resolution, frequency, cost and accessibility. Table 3 lists the main types of meteorological data product and highlights some of the important considerations relating to their use for health research. A more thorough discussion of climate data and its uses for health applications is provided by Mason et al (2018).

Direct observations from weather stations are considered the "ground truth" of meteorological data, but station datasets have several deficiencies (Table 3). Chiefly, the low density of weather stations in much of the world, inconsistent reporting and data quality problems pose challenges for climate and health research. Temporal inconsistencies in station data make them problematic for research on longer timescales and thus quality control by a meteorologist or a climatologist is important.

To address this deficiency, a variety of *gridded data products* also exist: some are produced by interpolating station data, others blend different data types such as station and satellite observations, and a third category are produced by dynamical (physics-based) models of the climate system. The complete spatial coverage offered by gridded data products is attractive, but the resolution of the grids is often too low to be useful. Regardless of the resolution, the quality of gridded data products in regions with few stations is extremely poor. Moreover, global products tend to include only a fraction of existing weather stations; far more data exist, housed by national meteorological services (NMS), which often do not report all stations for a variety of reasons (Figure 3, Dinku et al, 2016). Working with NMS to use the best available data can pay dividends, but can be costly as many NMS charge for their data. Initiatives that work with NMS to improve data availability, access and use are making climate-health research and surveillance feasible in an increasing number of LMICs (Dinku et al, 2018). Satellites provide the geographical coverage that station data lack. Products that combine satellite data with station observations provide some of the most accurate datasets available.

Reanalyses provide gridded predictions of the past weather, produced by a weather forecast model fed with observations (from stations, weather balloons, ships, buoys and satellites, among other sources). They provide predictions for a huge range of variables (a significant advantage over direct observations, which are often limited to temperature and precipitation) and offer the best 3D estimate of the circulation of the atmosphere, but they tend to be inaccurate for the near-surface variables that are relevant for most health impacts. Figure 4, for example, demonstrates the disagreement between two different model reanalyses on the seasonality of heat waves. Reanalysis data may be more useful for air quality, as it is dependent on circulation at higher levels of the atmosphere. Improvements have been made to the quality of surface variables in some recent reanalysis datasets such as ERA5 (Tarek, Brissette and Arsenault, 2020) but they remain problematic, particularly in data-sparse regions (Gleixner, Demissie and Diro, 2020) and their accuracy is often overstated.

Table 3
Characteristics of different types of meteorological dataset

Type of dataset	**Description**	**Characteristics**	**Strengths, weaknesses and considerations**
Station data	Observations measured at fixed weather stations over land	Parameters: usually total precipitation and daily minimum, maximum and/or average temperature. Other parameters available at some stations. Frequency: monthly accumulations or averages; daily data are also available from many stations. Temporal coverage: varies. Most stations have fewer than 50 years but some have much longer records. Resolution: huge differences in density of stations and consistency of data reporting between countries. Geographical coverage: stations exist in all land areas but stations are most sparse in LMICs	Most accurate observations available but highly sensitive to position of equipment (e.g. shielding by trees, buildings etc). Data are often incomplete, especially during severe weather events and in LMICs. Temporal inconsistencies occur when stations are moved or sensors are changed. Data require quality control by a meteorologist before use. Data are usually owned by the national meteorological services and may not be freely available. Many daily datasets are not yet digitised, especially in LMICs.
Gridded station data	Station data interpolated onto a regular grid	Parameters: generally precipitation and temperature but other parameters may be available. Frequency: monthly or daily summaries. Temporal coverage: varies between datasets. Resolution: varies between datasets. Geographical coverage: Global and regional products are available from different sources.	Very inaccurate in places with low density of stations and during periods with low data reporting rates. Global gridded datasets are suitable for large-scale analyses at sub-continental scales but not for estimating local exposures. National gridded datasets may be suitable for national-scale or smaller analyses, depending on the density of reporting stations included.

Type of dataset	Description	Characteristics	Strengths, weaknesses and considerations
Satellite data	Environmental data estimated from satellites	Parameters: varies between satellites. Can include rainfall and near-surface temperature, estimated indirectly, and other parameters e.g. NDVI (a measure of vegetation). Frequency: varies, but higher frequency data are generally available for more recent years. Temporal coverage: some satellite datasets are available from the late 1970s, others more recently. Resolution: varies, but higher resolution data are generally available for more recent years. Geographical coverage: complete for most of the world.	Best for a broad overview rather than for precise local data. Accuracy varies considerably according to weather conditions. Environmental data are inferred indirectly from observations higher in the atmosphere, so have to be calibrated using field observations. Rainfall and near-surface air temperature estimates are not very accurate. Estimating near-surface air temperature from surface temperature observations is very complex. Some data are freely available without cost. Some data are available in real-time.
Blended datasets	Combined datasets incorporating station and satellite data on a regular grid	Parameters: generally precipitation and temperature. Frequency: monthly or daily summaries. Temporal coverage: varies between datasets. Resolution: varies between datasets. Geographical coverage: global and regional products exist.	Quality is best where station density is high, and is usually improved, relative to gridded station data, where few stations exist.
Reanalysis data	Estimates of historical weather conditions produced by incorporating observations into a weather forecast model	Parameters: Many meteorological parameters are available at surface level and throughout the atmosphere. Frequency: 6-hourly or even hourly. Temporal coverage: several decades. Resolution: most current global datasets have at least 2° spatial resolution, with some much higher; for example, ERA5 has 31km resolution (Hersbach, et al. 2020). Some national reanalysis datasets may have higher resolution. Geographical coverage: global.	Poor accuracy for near-surface variables that are often important for health impacts, like 2m temperature, rainfall and humidity. Can be useful for air pollution studies which require data on atmospheric circulation. Not usually updated in real time. Best estimate available of the 3D circulation of the atmosphere.
Index datasets	Indices for key large-scale modes of variability in the climate system, such as ENSO, constructed from station data	Characteristics depend on the index NINO3.4 index, a measure of the ENSO, is a 5-month running mean of sea-surface temperatures in a particular region of the Pacific Ocean. All-India Monsoon Rainfall index is an annual areal average of rainfall across all Indian districts from June-September.	Index datasets are reliable as they have undergone extensive quality control.

A variety of models are also run in prediction mode across the full range of lead times: weather forecasts predict the weather from hours to about a week in advance, seasonal forecasts predict the general climate conditions expected over the coming months and climate projections provide scenarios of how the future climate could evolve under a variety of greenhouse gas and aerosol emission trajectories. Although climate models are evaluated extensively and their limitations are well documented in the literature, information about forecast skill is often unavailable and complex to interpret. Moreover, because forecast evaluation requires adequate samples of past forecasts and observations for comparison, our ability to quantify prediction skill is vastly superior at shorter timescales for which there are ample data available. On climate change timescales we actually cannot evaluate skill at all and so must rely on evaluating the processes in the models instead (for a fuller discussion of how to interpret climate change projections see Nissan et al, 2018; Nissan et al, 2021; Nissan and Conway, 2018). Sub-seasonal and multi-annual to decadal predictions (2-30 years) are emerging timescales for prediction and both are critical lead-times for adaptation. Sub-seasonal forecasts are of great interest for early warnings, but as skill is limited they are not yet suitable for most practical applications. Likewise, decadal predictions are an active area of research, and skill is improving, but they are not yet operationally available (Smith et al, 2019).

Within all these datasets, a major gap exists at the urban scale – a priority for climate-health research given that 55% of the world's population now lives in cities, with this figure projected to rise to 68% by 2050 (United Nations, 2018). Differences in land cover within cities and between cities and their rural surroundings lead to complex microclimates in urban areas that cannot be represented by a single weather station. Existing meteorological datasets lack the spatial resolution to detect these heterogeneities at the urban scale. Most cities have only one weather station (if they have one at all), usually situated at the airport, which can be some distance from the city centre. Monthly and seasonal averages and accumulations from a single site in the area could be predictive of climate conditions in the city itself, but for research on shorter timescales exposures in the city cannot be determined accurately from sparse station data. This gap poses a particular problem for research on heat-health thresholds to inform adaptation measures like early warning systems, research to determine causal pathways of impact in urban settings and for operational disease control programmes that rely on environmental surveillance to determine when and where to focus resources. Urban flooding is also highly heterogeneous, in part due to rainfall patterns, but differences in land use, elevation and draining infrastructure are also key determinants of local flooding.

Climate and environmental data or model outputs are widely used as inputs to epidemiological models, and operationally as indicators of environmental suitability for certain diseases and health outcomes. It is common practice for climate data to be used unquestioningly in epidemiological modelling when in fact, as with any dataset, climate data and models have errors and uncertainties and require expertise for their correct interpretation. Failure to recognise the uncertainties in input data runs into the well-known errors-in-variables problem, the simplest version of which is the following. Suppose that an outcome variable Y is related to an input variable X by the simple linear regression model, $Y = \alpha + \beta X + Z$, where Z represents the random variation of Y about the true regression line. Suppose however, that we can only measure $X^* = X+W$, where W represents measurement error in X*, and write the respective variances of X and W as σ_X^2 and σ_W^2 Then, naïve fitting of the regression model estimates not the target parameter β but the shrunken parameter $\beta^* = \beta/(1+ \sigma_W^2/\sigma_X^2)$, thus under-estimating the strength of the relationship between Y and X. This phenomenon is exacerbated when the degree of imprecision in X* is unknown and not easily estimated.

2.5 Climate metrics for health research

The large mass of available meteorological data products descibed in Section 2.4 is a huge resource for research on climate and health. However, data products must first be processed into covariates for epidemiological modelling that reflect aspects of weather and climate that are relevant for health impacts.

It may seem that with only two variables widely available (temperature and precipitation) there would be few candidates for climate covariates to use in health research, but there are infinite ways to summarise temperature and precipitation data. Continuous meteorological parameters can be recorded as instantaneous values or as accumulations, averages, minimum or maximum values over a period of time, which in practice is usually twenty-four hours. The choice makes a material difference because there are large variations in weather during the course of the day (Figure 7). Temperature has pronounced, predictable diurnal cycles, with the largest diurnal range observed in areas inland from the coast. Temperature extremes (daily minimum or maximum temperature) are more useful than average daily temperature for most types of health impact, including heat and cold exposure, as well as mosquito-borne diseases like malaria, where the parasite has critical temperature thresholds for survival. The rainfall rate can also vary dramatically during the course of one day and daily accumulations are generally used for most applications. However, some exposures, like flash flooding, are associated with peak rainfall intensity, which is estimated from rainfall accumulations over short periods, often just a few minutes. The appropriate metric may also differ according to the target population of the research. For example, several studies have shown that hot nights (high minimum temperature) are a strong determinant of heat-related mortality among the elderly (Laaidi et al, 2012; Burkart et al, 2014), but may not be the most important factor in outdoor occupational settings, where hot days (high maximum temperature) are likely to be a better predictor of heat-related morbidity and mortality.

For analyses on climate timescales, meteorological data must be aggregated into climate metrics. Typically, temperature is averaged over monthly or seasonal periods, while rainfall is accumulated, and these two simple metrics tend to be used most in climate-health studies on longer timescales. Monthly data are widely available in climate datasets, but they may not be closely related to some health outcomes because they cannot capture rapid-onset, short duration events like heat waves or flash floods. Events lasting a week or more, such as extended cold snaps, might be discernible in monthly data if they are particularly extreme. For slower-onset exposures like drought, or when health impacts are mediated by agriculture, monthly or seasonally accumulated rainfall could be appropriate measures.

However, climate metrics need not be constrained to simple averages and accumulations. Event-based metrics are more suitable for exploring the health effects of weather and climate extremes. Some examples of event-based metrics include: binary indicators (e.g. whether maximum temperature is greater than a threshold); frequency of hot days; frequency of rainy days; date of monsoon onset/cessation during the year; length of monsoon season; ENSO index (e.g. NINO3.4); seasonal total rainfall as a fraction of rainy-day frequency; maximum duration of heat wave conditions; maximum dry spell duration. On daily timescales, event-based metrics are binary indicators (e.g. it is raining or it is not raining; there is a heat wave or there is not a heat wave; the monsoon has started or it has not), but on longer timescales event-based climate metrics capture the statistics of weather events and extremes over a period of time. They can be expressed as frequencies, durations or even dates, if the timing of exposure during the year is important for a particular pathway (e.g. date of monsoon onset).

To construct event-based metrics, we first have to define what we mean by an event. How deep do waters have to be to constitute a flood? How hot is a heat wave, and for how many days/nights in succession? Should drought be defined when accumulated rainfall is below average, or below a particular amount, and for how long? For some events, there are standard definitions used by the meteorological

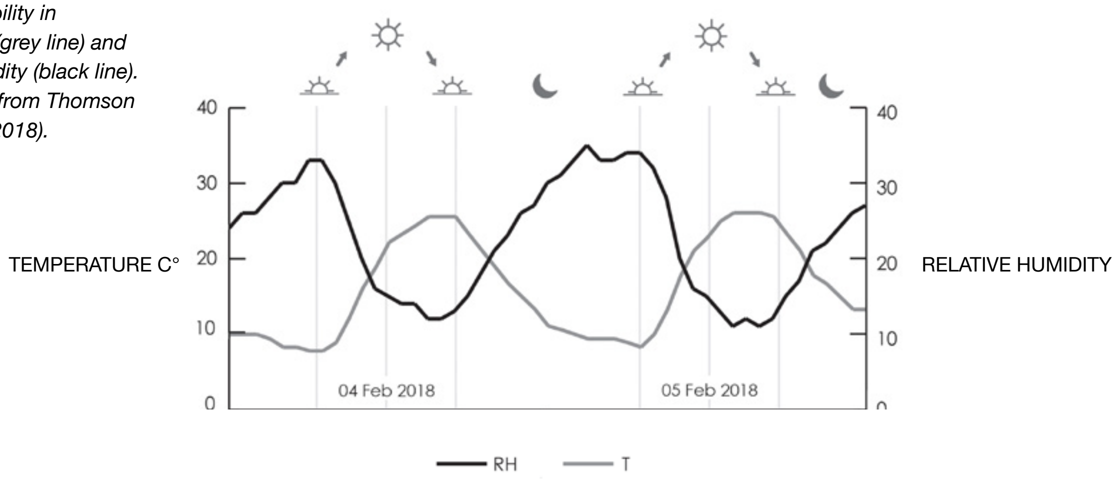


Figure 7.
Diurnal variability in temperature (grey line) and relative humidity (black line). Reproduced from Thomson and Mason (2018).

community, but for many there are not. In either case, the choice of event definition for use in health research should be informed by the health impacts we seek to investigate.

Identifying meteorological thresholds for health impacts should be a priority for climate-health research because these thresholds are the building blocks for a lot of other climate-health research, surveillance and operations (see Appendix B). Health-relevant event definitions are needed to construct suitable climate metrics for research on seasonal and longer timescales, for detection and attribution to climate change of the health impacts of extreme events, for health early warning systems and for guiding operational resource allocations.

Except where hard biological limits exist (e.g. survival potential in extreme temperatures) there are no universal thresholds. Different demographic groups, or those with medical conditions, may experience adverse health effects at different levels of exposure than others. Between regions, populations are not equally vulnerable to the same exposures, because of differences in the socioeconomic and ecological contexts that mediate climate-health pathways. For temperature, physiological adaptation can also be a reason for different thresholds of vulnerability between regions as well as during the course of a season or even a particular weather or climate event (McGregor et al, 2015). The appropriate threshold also depends on its intended use: research focused on the most extreme and rare impacts will need a higher threshold, while research looking at gradual impacts at lower exposure levels will need a more moderate threshold. Thresholds can be defined by an absolute value (e.g. 2°C) or a value relative to the local climate (e.g. 5th percentile of local winter-time temperatures over the last 30 years) and may need to be continuously or periodically adjusted to account for climate change.

Lastly, seasonality in climate means that identifying the relevant timing of exposure during the year is critical when choosing a climate metric. Timing of exposures during the year can mediate their effects on health, particularly for vector-borne diseases (as there are seasonal components to animal and insect behaviours and population dynamics) or in agricultural economies, which depend on the weather conditions at particular points in the agricultural calendar. The timing of exposures is also important because of lags in climate-health associations, which mean that a health impact may not be experienced immediately after an initiating meteorological exposure. Lags between exposure and outcome vary according to the pathway in question, which may be mediated by natural or socioeconomic systems (Table 4). Depending on the hypothesised pathway of impact, covariates may also need to be lagged in order to detect associations.

Some sense can be made of this complexity by considering two broad categories of climate-health research: hypothesis-generating, exploratory research and research to test a specific hypothesis about a particular climate-health pathway.

- *For hypothesis-generating, exploratory research* on climate-health associations, monthly or seasonal average temperature or accumulated precipitation are an appropriate place to start but can be supplemented with climate metrics that provide general measures of the "weather within climate", which may be more relevant for health impacts. For example, generic metrics could include frequencies or intensities of weather extremes, either defined relative to the local climatology (e.g. as percentile values) or in reference to any existing literature on health-relevant thresholds. Climate metrics associated with the indirect effects of climate on health could also be used, such as via agriculture, tourism or vector ecology, where evidence on these connections exists within the literature.
- *When testing a specific hypothesis*, the choice of climate metric can be informed by the hypothesised pathway. Important factors to consider include any meteorological thresholds relevant for health impacts, the timing of exposure during the year and any lags in the pathway of impact (Table 5). For example, temperature extremes above 30-32°C can result in declines in wheat yields when they occur during particular crop growth stages (Arshad et al, 2017). A plausible hypothesis could test the effect of heat on nutrition that operates via poor crop yields at harvest and spikes in food prices. A suitable climate covariate for this pathway could be the number of days on which daily maximum temperature exceeds crop tolerance levels during the critical window of the growing season.

3

Combining climate and health data for research

Combining climate and health data for research

Some of the statements in this section will draw on the results of our investigations into current LPS, which we will describe in Section 4 below.

The challenges of bringing LPS and climate datasets together begin with the differences in the spatial and temporal characteristics of these data sources, as summarized broadly in Table 4. High-quality meteorological datasets generally lack sufficient spatial resolution to estimate individual exposures accurately, given the spatial heterogeneity in weather and climate. Conversely, the infrequent follow-up protocols typical of most longitudinal population studies are inadequate to detect the acute impacts of climate variability and extreme events (Table 4).

Table 4.
Typical characteristics of longitudinal population studies and climate data

Longitudinal Population Studies	Climate data
Blended datasets	Network of monitoring sites or grids
Large number of individuals	Long time series at each site
Small number of follow-up times	High-frequency data: monthly, daily or sub-daily
Yearly or longer follow-up intervals	Small number of variables
Many variables	Global coverage, but locally sparse
Local geographical coverage	

Below, we describe some basic requirements for overlaying health and climate datasets to address research across the breadth of space and time scales. This exercise is guided by two questions, the answers to which depend on the scale of analysis concerned:

1. What climate data are required to characterise the exposure?
2. What health data are required to capture the effects of that exposure?

In the following discussion, we outline these requirements and highlight the situations for which existing datasets generally fail to meet them. We also summarise some methodological approaches to addressing incommensurate spatio-temporal characteristics between datasets.

3.1. Where, and at what spatial resolution, should data be collected?

To capture the effects of weather and climate exposures we require both health and meteorological data at resolutions comparable to the relevant spatial scale of weather or climate variability.

The requisite spatial resolution depends on both the timescale of analysis and the variable of interest. For example, on timescales of days, absolute temperatures vary over smaller spatial scales than temperature anomalies. The spatial resolution of meteorological observational data from weather stations and most gridded data products is usually inadequate to capture fine-scale gradients in weather exposures that are needed, for example, to build effective local early warning systems for flooding. Also, most gridded data products are derived from a combination of spatially sparse direct measurements and modelling assumptions, typically without accompanying information on their uncertainty limits (see Section 2.4).

Since climate variables on longer timescales tend to be spatially correlated over larger distances (see Section 2.2), coarser-resolution meteorological data may be adequate to characterize exposures from longer-term climate variability. During El Niño or La Niña events a whole region can be affected by changes in weather over the course of several months. While any individual weather or climate event will be limited in temporal and spatial scale, together they can expose large areas over the course of the season, which is reflected in a greater degree of spatial coherence in seasonal averages and other seasonal climate statistics.

Adverse health impacts from climate are often associated with exceedance or non-exceedance of a critical threshold. Where the climate of a region is very heterogeneous, large-scale anomalies in climate can disproportionately affect the health of people whose local climate hovers close to these important thresholds. For instance, malaria cannot be transmitted if temperature drops below the minimum threshold for parasite development. In mountainous areas, the sharp variation of temperature with altitude means that populations living at elevations close to this threshold can be exposed to malaria during warmer periods, when the parasite is able to go through its development cycle at higher altitudes (Lyon et al, 2017). Thus, even if abnormally warm temperatures are seen across a large area, finer-scale data may be needed to capture the exposure and impacts of that temperature anomaly on highland populations compared with lowland populations.

Even when fine-scale spatial resolution is needed in principle, it may be inefficient to apply this requirement to the whole of a study-area. For the efficient detection of effect thresholds, the spatial frequency of sampling needs to be high in areas close to the critical threshold. One way to address this question is through adaptive design (Chipeta et al, 2016). An adaptive design is one in which data are collected in batches, and the data accrued in early batches inform the sampling design for the next batch. For example, when designing a study on a blank spatial canvas the first batch in an adaptive design will collect data over the whole of the study area but later batches may concentrate on sub-areas where more information is needed to answer the question of primary interest; see Figure 8.

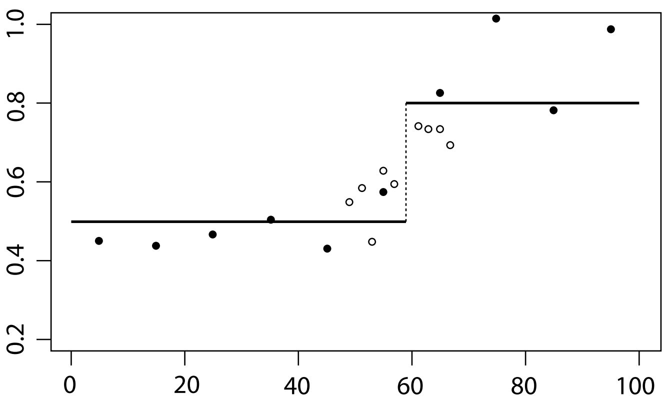


Figure 8.
Adaptive design. A hypothetical threshold effect (solid line) and a two-stage adaptive design to detect the threshold. Solid dots show locations and measured values from ten stage 1 samples along an east-west transect, open circles show locations and measured values from a further ten stage 2 samples chosen after inspection of the stage 1 data.

3.2. When, and at what temporal resolution, should data be collected?

The weather is highly variable on very short timescales. A snapshot of the weather taken once a month or every season is therefore a very poor indicator of the prevailing climate. High-frequency (usually daily) weather station data are needed, even for investigations examining the effects of slower-moving components of climate variability. For example, drought is the result of accumulated precipitation, which requires high-frequency measurements to calculate because the rainfall rate is so variable. However, though they are often spatially sparse, directly measured climate data are typically temporally dense. Weather stations routinely record directly measured meteorological data at daily or sub-daily frequency (Table 4). Longer-term averages or other metrics can be constructed from these as required, although free access to daily data from weather stations can be a challenge in much of the world.

How frequently *health data* must be collected and the overall temporal coverage required are functions of the type of health outcome, the type of weather or climate exposure concerned and the lags between the timing of exposure and health impacts, which can be substantial (Table 5). The key overarching message is that climate-health research requires multiple repeated health observations to capture the impacts of weather and climate variations over time. Sources lacking frequent samples cannot be used to address most questions of interest regarding climate and health (e.g. Demographic Health Surveys, or longitudinal studies with very few follow-ups many years apart). Health outcomes that are highly variable over time (e.g. blood pressure) will require more frequent sampling to detect any changes caused by a meteorological exposure, while slower-varying health outcomes can be sampled less frequently. Analyses exploring health associations with short-term weather fluctuations or with seasonal, inter-annual or longer timescales of climate variation have corresponding frequency requirements, as outlined in broad terms in Table 5. The short temporal scales associated with the health effects of extreme events like floods, heat waves or droughts present a substantial challenge given the infrequent follow-ups typical of longitudinal population studies (Table 4).

In much of the world, the presence of strong and dynamic seasonal cycles, both in climate and in some aspects of health, mean that the timing of health data collection during the year is critical for research questions at all timescales, not just for those directly concerned with investigating seasonal effects. Furthermore, in many parts of the world the magnitude of the seasonal cycle is much larger than longer-term variations, including climate change trends. Thus seasonality can lead to incorrect estimates of long-term trends and variability when multi-year health data are measured at different points during the seasonal cycle. A suitable response to this challenge is a rolling sampling design for the collection of health data whereby, for example, each cohort member is sampled annually but different sub-cohorts are sampled each month. When data cannot be collected throughout the year, the timing of data collection should be informed by plausible hypotheses regarding aspects of the climate thought to be relevant for health in the local context. For example, to explore the influence of flooding on diarrheal disease, data would be needed during or shortly after the flood-prone season. Finally, the effects of the El Niño Southern Oscillation on regional climate vary according to the season (Figure 6), so annual data or data in the wrong season may not be useful for exploring the health impacts of El Niño or La Niña events.

Studies aimed at understanding the causal mechanisms driving health impacts at an individual level are the most data-hungry. They need health and climate exposure data to be collected at frequencies at least as high as the timeframe of the pathways in question, which could vary from hours to months (Table 5). For studies of this kind, personal monitoring of climate exposure and health biomarkers may be needed to address both the spatial and temporal data requirements (see below, Section 3.3). To understand the steps along these pathways, data would also be required on any mediating factors in the climate-health relationship relating to the social, economic or ecological contexts. Variables with high temporal variance will need more frequent observations (e.g. food prices), while those which are more stable could be sampled less often (e.g. socioeconomic status). Ideally, studies to explore the acute impacts of weather or climate events would be planned in advance and mobilized at short notice when climate forecasts indicate an increased risk, learning from the experience of communities skilled at developing early warning systems, particularly the operational health and disaster management communities.

3.3. Accounting for mobility

People are not stationary; they move around outside and between indoor and outdoor settings. In some regions, temporary (e.g. seasonal) and permanent migration over larger distances are also important considerations. To account for movement, conventional meteorological datasets must be transformed from a set of time series at fixed points in space (stations or grids) to a set of trajectories corresponding to the movements of individuals through exposure space. Closed cohort studies, which exclude individuals who move in or out of the study area, can avoid these complications if the study area is sufficiently small that the climate can be treated as homogeneous. However, even within a homogeneous climatic environment, indoor and outdoor conditions can differ substantially and there is little understanding of how the two are related. The association between indoor and outdoor conditions will vary according to building materials, features (such as whether windows can be opened for ventilation) and amenities like heating and air conditioning, which can change over time and are likely to vary between individuals within the same study. Studies are needed to understand indoor temperature and humidity exposures during heat and cold extremes in a range of building types and geographies.

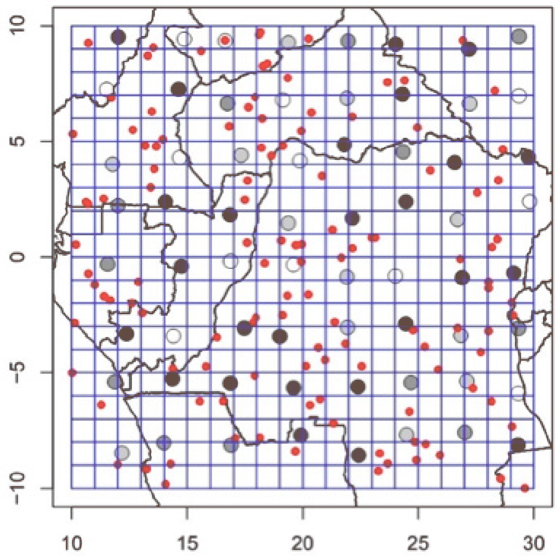
Fine-scale monitoring of individuals is increasingly feasible. For long-running cohort and panel studies, individual monitoring can be both expensive and intrusive, but the very wide global penetration of mobile phones and, in wealthy countries, the growth in use of fitness bracelets and smart watches may alleviate these concerns. If such methods are not practical for extended time periods, they could still be employed for short periods as a complement to longer-term studies that rely on traditional climate data sources, for example through the use of forecast-based surveillance (see Section 3.2).

3.4. Methodology

As is by now evident, climate data and health outcome data are generally not available at the same set of sampling units. Further processing or modelling is usually required to make the data suitable for analysis using routine statistical methods. There is no one-size-fits-all approach; there are advantages and disadvantages to all options and climate expertise is required to make informed choices about how best to prepare the data for analysis with health data. It is vital to use a consistent dataset and method across the length of the analysis, which poses a particular challenge for analyses on longer timescales.

In the spatial dimension, there is a wide range of possible data-formats. Figure 9 shows a hypothetical scenario containing four datasets in different formats: a point process of health events; a raster image of a gridded climate data product; direct measurements from a weather station network; and socio-economic data for a partition of the study-area into administrative districts.

Figure 9.
A hypothetical scenario involving four datasets in different formats. The red dots are the locations of individual cases of a particular disease. The grid squares represent a gridded data-product. The grey-shaded circles represent measurements made at a fixed set of weather stations. The study-area (here, western equatorial Africa) is partitioned into sub-areas (here, countries) on each of which aggregated socio-economic data are available.



Most currently used methods of dealing with data that are collected from incommensurate sets of spatial or temporal sampling units operate by transforming the data. For example, in the spatial dimension, the modifiable areal unit problem (MAUP) arises in human geography when there is a need to understand the relationship between two quantities for which data covering the whole of the study-area are only available as aggregated values derived from different partitions of the study-region into sub-areas. Openshaw and Taylor (1979) coined the term MAUP and showed empirically how the correlation between two aggregated values can be highly sensitive to the choice of sub-areas over which to aggregate the data.

With regard to the time dimension, most phenomena of scientific interest evolve as a continuous-time process, S(t) say, from which a discrete time series of data can be extracted either by *sampling* S(t) at specified set of times or *aggregating* S(t) over a specified set of time-intervals. A third format of temporal data is a point process, consisting of the actual times at which an event of interest occurs and which, by aggregation, can be converted into a time series of incident or prevalent counts.

When multiple datasets are recorded at different spatial and/or temporal resolutions, one approach is to aggregate all of them to the coarsest resolution. For example, using this approach, if the town in which a subject currently lives is recorded at the time of data collection, but only their district is known at the time of a flood that is hypothesized to have impacted their health, then the analysis can only proceed at the district level. A related problem is data-misalignment. For example, health data may be available for a town or area with no weather stations in it. Options for analysis then include using the nearest, or most representative, weather station available, interpolating meteorological data from multiple nearby stations or using a gridded data-product rather than direct meteorological measurements.

From a statistical perspective, a more principled approach is first to specify a model for the process of scientific interest, then to specify a joint model for all of the data that arise from incomplete or imperfect observation of the process. Figure 10 is a schematic representation of this approach in its simplest form. The process of scientific interest is the causal effect of temperature on the spatially and temporally continuous variation in the risk of a particular disease. Health outcome data are available from a randomised prevalence survey of communities within the study-area. Temperature data are recorded at a network of weather stations within the study-area that do not align with the locations of the communities.

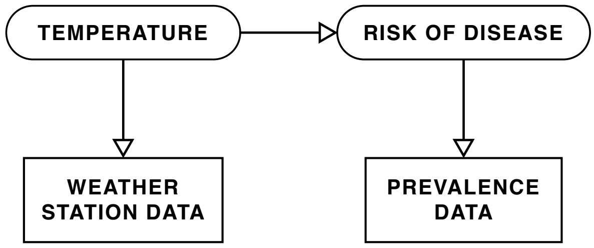
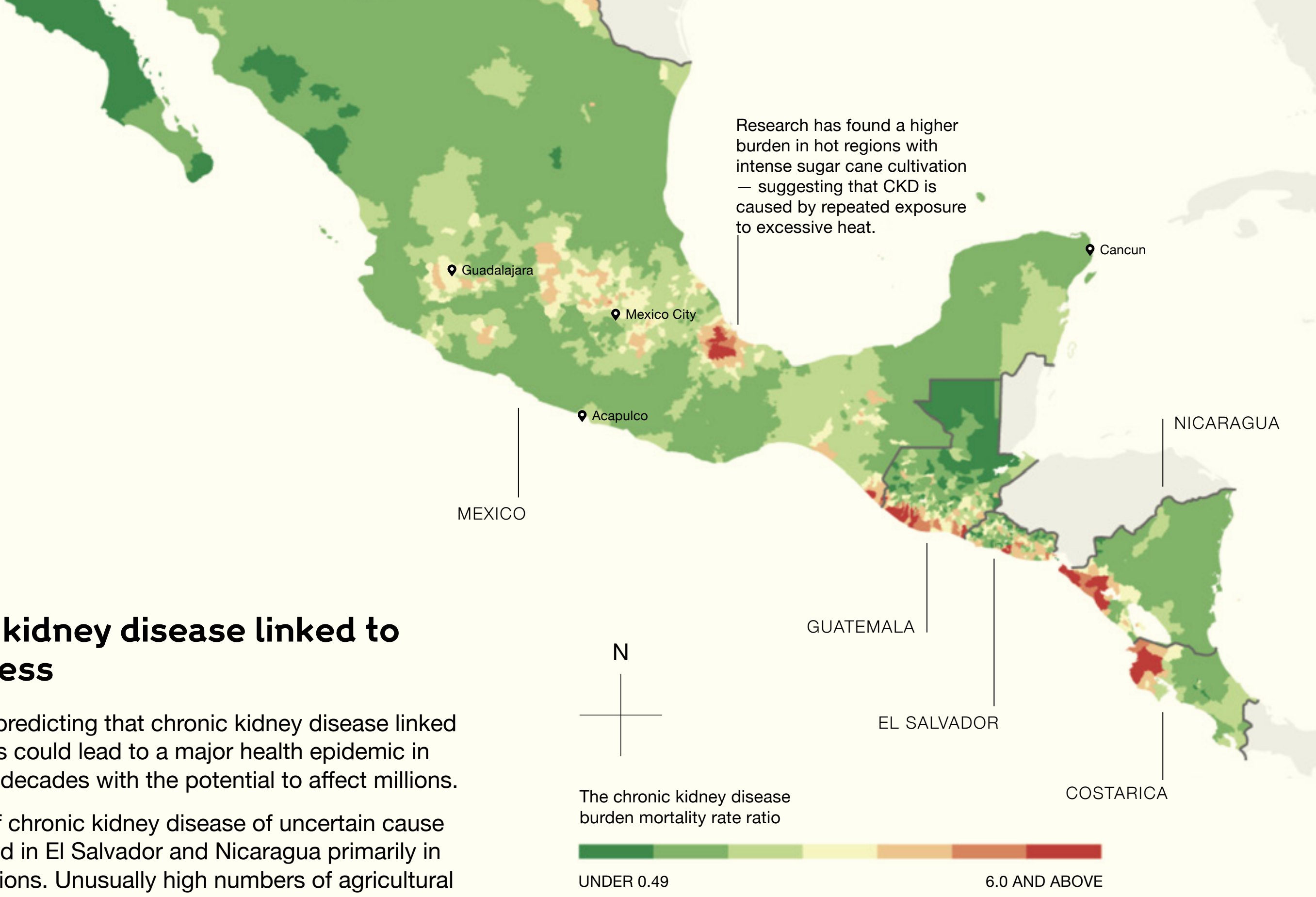


Figure 10.
Diagrammatic representation of a hypothetical model for the relationship between temperature and disease risk. Rectangles and ovals represent observed and unobserved spatio-temporally varying quantities, respectively. The relationship of interest, which in this case is between temperature and disease risk, cannot be measured directly.

The diagrammatic representation of Figure 10 translates to a set of conditional probability distributions that collectively define the joint model for process and data. Expressed in mathematical notation, the process model is [T][R|T] where T is temperature, R is disease risk, each set of square brackets is to be read as "the probability distribution of" and a vertical bar denotes conditioning. The two data-models are: [W|T], the probability distribution of the weather station data (in whatever format they have been collected) conditional on the true but unobserved temperature field; and [P|R], the probability distribution of the prevalence data conditional on the true but unobserved risk field. Inference about the probability distribution of scientific interest, [R|T], follows by application of Bayes' theorem. This bare formalism conceals a growing body of contemporary statistical methodology that allows each data-element to contribute its full information content to the problem in hand, rather than being degraded by unnecessary aggregation to obtain a neat alignment of all data-elements; see, for example, Congdon (2021) or Nicholson et al (2022). At least as importantly, each of the conditional distributios that collectively define the model can benefit from a combination of scientific insight into the underlying mechanisms that generate the data, and statistical expertise to ensure that modelling assumptions are validated empirically and model parameters estimated efficiently. Put another way, information is derived from both data and scientific knowledge. Sophisticated statistical methods of this kind are extremely powerful, but are underutilised in both the climate and health domains.



Chronic kidney disease linked to heat stress

Doctors are predicting that chronic kidney disease linked to heat stress could lead to a major health epidemic in the next few decades with the potential to affect millions.

Epidemics of chronic kidney disease of uncertain cause have emerged in El Salvador and Nicaragua primarily in hot, rural regions. Unusually high numbers of agricultural workers have begun dying from irreversible kidney failure. In other parts of the world with hot temperatures, such as in India, a large number of people involved in heavy manual labour have started to be affected by this disease. Kidneys are particularly sensitive to extreme temperatures as they are responsible to maintaining fluid balance in the body.

Consensus is emerging that chronic kidney disease of uncertain cause should be recognised as a heat stress-related injury. Subtle damage to worker's kidneys occurs each day while they are in the field and this in turn can develop into severe kidney disease or complete renal failure over time (The Guardian, 2021).

Map: The chronic kidney disease burden mortality rate ratio. The measures of CKD burden shown are the proportional mortality odds ratio (Nicaragua), mortality ratio (Guatemala, Mexico, Costa Rica), and hospital admissions rate ratio (El Salvador). Source: BMC Public Health, Hansson, E, Mansourian, A, Farnaghi, M et al

Table 5.
Durations and relevant timescales of variability for a selection of plausible (hypothesised) climate-health pathways

Health-related outcome or disease	Plausible climate pathway(s)	Duration of pathway(s) including any lags	Relevant timescales of variability	Considerations
Health systems (Nissan, Ukawuba and Thomson, 2021)	Extreme weather and climate events can disrupt access to vulnerable populations, impairing routine and emergency healthcare provision and disease control programs.	Days to weeks according to the timescale of different extreme events.	Many disease control programs (e.g. malaria) are seasonal.	
	Economic impacts of reduced agricultural productivity in some countries may lead to reduced household income, lower government revenue and decreased availability of funding for health care.	Months (the effects of climate exposures during the growing season are not realized until harvest).	Impacts will depend on inter-annual climate variability, which impacts production in key seasons.	
Chronic kidney disease (CKD) (Friel et al, 2011)	Recurrent heat stress-causes dehydration and osmolarity resulting in CKD onset and accelerated progression.	Heat waves typically last 1-10 days. Health impacts from heat occur during the event itself but disease onset may require repeated exposure via many heat events.	Heat waves are seasonal in most places but not all. Long-term trends in heat wave frequency and severity are already detectable.	Length of heat wave season is increasing with global warming –heat waves are occurring outside the normal (historical) summer season.
Epilepsy (Sisodiya et al, 2019; Gulcebi et al, 2021)	Exposure to extreme weather events, especially heat waves, increases risk of fever, stress and sleep deprivation, any of which can precipitate seizures. Genetic mutations involved in some forms of epilepsy code for proteins that are highly sensitive to ambient temperature	Heat waves typically last 1-10 days and health impacts are fairly immediate. Effects of changes in molecular-level biochemistry can be short-term acute (e.g. fatal seizure) or long-term accumulative (e.g. more frequent non-fatal seizures)	Heat waves are seasonal in most places but not all. Long-term trends in heat wave frequency and severity are already detectable.	

Health-related outcome or disease	**Plausible climate pathway(s)**	**Duration of pathway(s) including any lags**	**Relevant timescales of variability**	**Considerations**
Cardio-vascular disease (CVD) (Friel et al, 2011)	High atmospheric pressure systems associated with heat waves increase air pollution which directly contributes to CVD.	Heat waves typically last 1-10 days and health impacts are fairly immediate, but recurrent exposure could play a role in CVD.	Pollution surges likely to become more frequent in line with heat waves.	
	Dry conditions enhance air pollution which directly contributes to CVD.	Dry conditions develop over weeks to months. Uplift of dust and particulate matter enhanced by windy conditions occurring over a timeframe of hours to days. Lag to impacts are not well understood.	Rainfall is seasonal in most locations. Strong rainfall variability often exists on inter- to multi-annual timescales, with ENSO associations across much of tropics. Long-term drying trends in places, increasing precipitation in others. Many trends undetectable to date. Trend detection is complicated by decadal variability.	Also occurring over background trends in air pollution, which will be affected by climate policies and adaptation. Hot and dry conditions often occur together. Compound events could become more likely and more severe.
	Wildfires occur in hot and dry places, causing air pollution and CVD.	Days to seasons (dry pre-conditions for fire develop over months, but wildfires are triggered during periods of hot weather).	Strong interannual variability in rainfall alters risk of wildfires, with ENSO links in many places. Increasing frequency and intensity of heat waves as the climate warms.	
	Extreme heat stresses the cardiovascular system leading to heat illness and death.	Heat waves typically last 1-10 days.	Heat waves are seasonal in most places but not all. Long-term trends already detectable in most locations.	Length of heat wave season is increasing with global warming. The risk of mortality can be worse early than late in the season due to physiological adaptation and seasonal behavioural factors (Sheridan and Kalkstein, 2004; Koppe and Jendritzky, 2005)

Health-related outcome or disease	Plausible climate pathway(s)	Duration of pathway(s) including any lags	Relevant timescales of variability	Considerations
Respiratory illness (Friel et al, 2011)	Higher temperatures associated with an increase in tropospheric ozone cause respiratory illness and exacerbate asthma.	General shift towards warmer weather raises background ozone levels across the full distribution of temperatures and timescales. Heat waves (1-10 days) cause surges in ozone exposure.	Long-term trends are detectable in average temperature and in heat wave frequency and intensity.	
	Wildfires occurring in hot and dry conditions cause air pollution and respiratory illness.	Wildfires last hours to days. Dry pre-conditions for wildfires develop over months.	Strong interannual variability with ENSO links across the tropics.	
Mental health (Vins et al, 2015)	Extreme weather events have immediate mental health impacts through displacement, injury, death of friends/family members, trauma -> stress, anxiety, depression.	Hours to weeks depending on the hazard concerned. Flood waters can take time to recede. PTSD may take time to onset. Pathways and associated lags are not well understood.	Many extreme events are seasonal. Impacts on indirect determinants of health (e.g. livelihoods, food systems) may be season-dependent. Long-term trends occurring in some extreme weather event characteristics.	Compound events could be important, as could successive exposure to multiple events.
	Climate (especially drought) affects the socioeconomic determinants of mental health (e.g. poverty, inequality, community, competition for natural resources).	Drought develops over weeks to months. Lags between drought and full mental health impacts could be long as takes time for socioeconomic effects to materialise – likely months.	Drought is typically seasonal. The effect of climate exposures on the wider economy could depend on their timing during the year, especially in agricultural economies.	

Health-related outcome or disease	Plausible climate pathway(s)	Duration of pathway(s) including any lags	Relevant timescales of variability	Considerations
Undernutrition/ acute malnutrition; (many plausible pathways exist – see Davis, Downs and Gephart, 2021; Fanzo et al, 2018)	Climate shocks and stressors (e.g. drought, flood, heat, and changes in seasonality of normal, (non-extreme) climate occurrences) affect multiple components of the food system including agricultural productivity, supply chains and distribution networks, affecting diet quantity and quality via changes in the availability and affordability of nutritious food. Poor diet can also lead to CVD.	Drought is a slow-onset hazard that lasts several months. Floods last from days up to a couple of weeks, depending on the driving mechanism, but can be months in particular locations (e.g. Bangladesh). Heat extremes last 1-10 days. Effects take time to filter through the food system (lags largely unknown and highly context dependent).	Impacts will depend on timing of hazards during the agricultural calendar (seasonality). Strong interannual variability with drought, flood and potentially heat linked to ENSO in many agricultural economies. Compound events (e.g. heat concurrent with drought, or drought followed by flood)	
	Heat waves and droughts decrease crop yields and livestock productivity, affecting access to high-quality diets and decreasing nutritional status (Fanzo et al, 2018)	Months overall. Heat extremes last 1-10 days, while drought conditions can last months. Impacts on crop yields may be realised months after exposure.	Heat waves and drought are mostly seasonal. Trends in heat wave frequency are already detectable (Alexander et al, 2006; Donat et al, 2013a, b) Droughts are becoming more frequent and severe in some places (Seneviratne, 2021).	
	Hot and damp conditions lead to food-borne illnesses, e.g. via aflatoxin and bacterial contamination (Davis, Downs and Gephart, 2021). Disruption to food distribution from extreme weather events can result in delays and food spoiling.	Days to weeks, depending on storage and distribution times.	Weather to seasonal. Long-term trends in humidity (in line with temperature increases) and in frequencies of extreme events will increase risks over time.	

4

Longitudinal population studies

Longitudinal population studies

Wellcome uses the term *longitudinal population studies* (LPS) to include "cohorts, panel surveys and biobanks" (https://wellcome.org/what-we-do/our-work/longitudinal-population-studies).

In a *cohort* study, a set of individuals who share a common characteristic, for example birth date or area of residence, are followed up over time, with measurements taken periodically on aspects of their health considered relevant to the research questions in hand. A cohort is *closed* if follow-up is restricted to an initial set of recruits, *open* if newcomers can be recruited over time.

A *panel* study is similar in concept to a closed cohort. The distinction is that in a panel study recruits are selected to span a range of characteristics, for example their birth date or area of residence. In the authors' opinion the distinction is largely semantic; a panel is a closed cohort to which recruitment from the target population uses a sampling design that is stratified by one or more characteristics, presumably to guarantee balance across the said characteristics.

A *biobank* is a repository for biological samples that are collected and stored with a view to their being used in multiple future research studies; a biobank may or may not include multiple samples collected at different times on the same individual; UK Biobank (https://www.ukbiobank.ac.uk/) is an example of a biobank whose design is an open cohort with a target recruitment of 500,000 individuals aged between 40 and 69 years at recruitment.

Finally, a *repeated cross-sectional study* is one in which one or more groups of individuals are followed up over time, but at each follow-up time measurements are taken on a differen sample within each group. A topical example is the REACT study of Covid-19 prevalence UK-wide (Riley et al, 2021; see also https://www.imperial.ac.uk/medicine/research-and-impact/groups/react-study/), in which data are collected from residents in each of England's 315 Lower Tier Local Authorities by random sampling at approximately monthly intervals.

When discussing the role of LPS for research at the health-climate interface, an important scientific distinction is between aetiological questions concerning an individual's molecular-level responses to particular climate-related exposures and policy-relevant questions concerning population-level effects.

4.1. Survey of existing LPS

The number of current individual LPS must run into the thousands world-wide. Rather than attempt a comprehensive survey, we have searched primarily for collaborative activities such as networks or consortia that have a wide geographical reach.

Table 6 summarises the results of this search. In this table, *enrolment start* indicates when the "baseline" data were collected for the first time, it does not indicate whether the cohort is still open or if it is now closed. It is difficult to provide this information because a single network/consortium usually contains both open and closed individual cohorts; similarly, the term *variable* for either enrolment or follow-up refers to variation over the individual cohorts within the network. Under the *georeferencing* heading, the stated resolution (region, town, house etc...) is the lowest level of spatial resolution across all individual studies.

Table 6.
Summary descriptions of some current LPS

	Type of study	Substantive focus	Geographical coverage	Enrollment start	Follow-up interval (years)	Georeferencing	Size	Link
CHPT	Collection of cohorts	Etiology of cancer and other major chronic diseases	12 Developed countries in North America and Europe)	1976 - 2012	1 – 5	Unknown	2,45M	More
EPIC	Prospective cohort	Cancer and nutrition	10 western European countries	1992 to 1999	3 – 5	Town	521K	Mor
ATHLOS	Collection of cohorts	Ageing	Global (38 countries)	1992 - 2012	1 – 10	Unknown	444K	More
IHCC	Consortium of LPS	Wide-ranging	Global (110 cohorts in 48 countries)	Variable	Variable	Unknown	52M	More
MORGAM	Collection of cohorts	Cardiovascular diseases	9 European countries plus Russia and Australia	1982 - 2014	1/3 – 5	Town	367K	More
Closer	Access point for multiple LPS	Wide-ranging	UK	1931 - 2013	Variable	Unknown	774K	More
ALPHA	Consortium	HIV	Sub-Saharan Africa	Variable	1 – 2	Household/ Village	645K	More
INDEPTH	Network of health surveillance systems	Wide-ranging	Africa, Asia and Oceania (21 countries)	1961 - 2011	¼ – 2	Household/ Village	4.5M	More
ISAAC	Multi-centre	Asthma	Global (105 countries)	1994 - 1995	4	Town	2M	More
IALSA	Research network	Aging and dementia	Global (38 countries)	Variable	Variable	Unknown	1.4M	More
AGRICOH	consortium	Agricultural exposures	Global (13 countries)	1969 - 2013	1 – 3	Unknown	1.3M	More
BioS-HaRE-EU	Biobank consortium	Multifactorial diseases	Europe plus Canada	Variable	Variable	Unknown	1M	More
Cosmic	Cohort consortium	Cognitive aging	Global (33 countries)	1948 - 2010	Variable	Unknown	121K	More
HELIX	Cohort consortium	Environmental exposure in early life	Europe (5 countries)	Variable	Variable	Town / Mobility data	31K	More

Here, we describe a selection of existing activities that seem to us to offer particular promise as vehicles for investment in climate-health research by virtue of their wide geographical span and collection of data on a range of health outcomes that have a potential connection with climate-related exposures. Before doing so, we make a number of general observations:

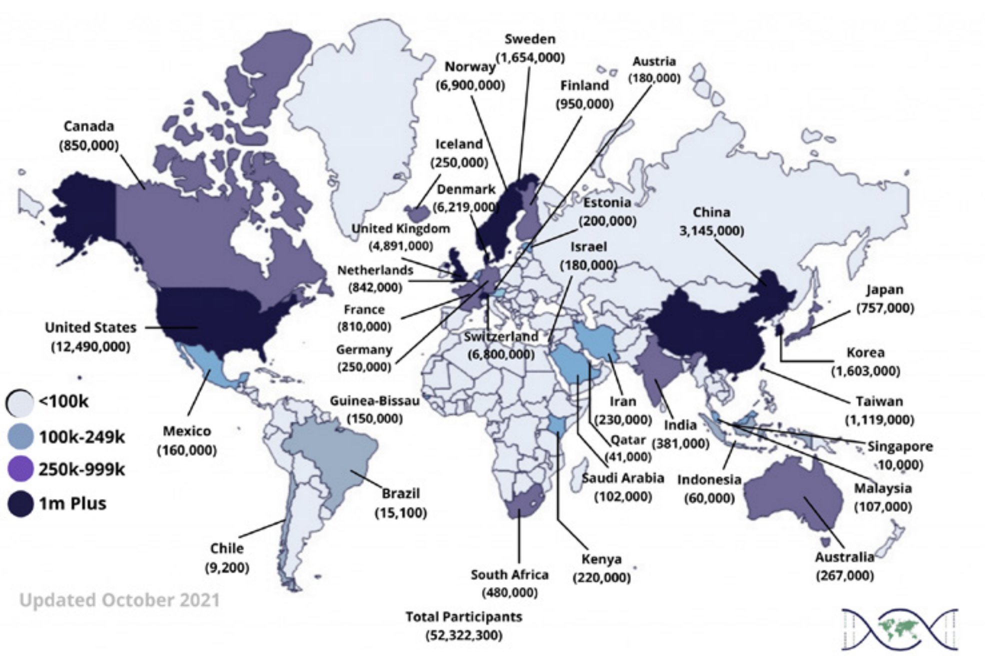
Firstly, a limitation of all of the LPS that we have found is that they collect health outcome data using follow-up intervals of the order of years rather than months or shorter. This restricts the range of climate-health interactions that they can investigate; see Section 2.1.

The above observation notwithstanding, secondary analysis of data from existing LPS can be a useful starting point for exploratory research, with the potential to lead to closer long-term collaborations on specific research questions using bespoke study-designs. Existing networks or consortia that span long total follow-up periods and extensive geographies are likely to be the most useful for this kind of research.

Thirdly, we have found a number of web-sites that gather information about multiple LPS and provide searchable electronic data catalogues. These include maelstrom (https://www.maelstrom-research.org/), Synchros (https://synchros.eu/), Closer (https://www.closer.ac.uk/) and the LMIC LPS Directory (https://ifs.org.uk/tools_and_resources/longitudinal).

The attractions of investing in a consortium of ongoing LPS are that the infrastructure for collection and collation of health data is already in place, and the consortium as a whole can achieve both a scale and a geographical (and hence climatic) reach greater than any single LPS. The major challenges facing this approach are the need to engage, potentially across multiple countries, with the national or sub-national agencies that hold climate and health data to secure their participation, and to fund infrastructure for the collection and curation of more localised climate data to capture the small-scale variations in micro-climate, for example between rural and urban settings or ambient and indoor exposures, that can affect multiple health outcomes but are not captured by global-scale climate maps.

Figure 11.
Total number of participants in IHPCC member cohort studies, by country as of February 2020 (from https://ihccglobal.org/, downloaded 9 August 2021).



LPS consortia

The International Hundred Thousand Plus Cohort Consortium (IHCC, https://ihccglobal.org/) was established in 2018. It seeks to bring together the data-resources and intellectual strengths of multiple LPS world-wide; as of May 2020 it included 110 cohorts spanning 48 countries across all five continents (Manolio, Goodhand and Ginsburg, 2020), albeit with an understandably greater concentration of members in high-income than in low and middle income countries; for example, it embraces almost 12 million individual study-participants in the USA but less than one million Africa-wide (Figure 11). The consortium is open to the possibility of re-directing some of its future activities towards environmentally sensitive health outcomes.

The proposed African Population Cohort Consortium (https://wellcome.org/what-we-do/our-work/longitudinal-population-studies) presents an opportunity to develop the consortium model within a newly established consortium in a region of the world that is both under-served by current LPS and particularly vulnerable to the effects of climate change on human health (Harrington et al, 2016, 2017).

Electronic health records

Electronic health record (EHR) systems offer the potential to harness comprehensive health data from whole populations in real-time, rather than from limited samples collected periodically. Systems of this kind have the greatest potential in countries that operate single, national health services. However, to varying degrees in different countries, concerns about individual privacy can limit access to EHR data for research purposes. Also, reference to "whole populations" is somewhat misleading, as even in wealthy countries the most disadvantaged sectors of society can experience difficulty in accessing health care, and other forms of bias can arise from variation in individuals' or communities' propensity to engage with their national health services. For these reasons, considerable value can be added to EHR data by investing in relatively small, randomised LPS, perhaps targeted at particular sub-populations or particular health outcomes, to enable exploitation of the much more extensive EHR data whilst correcting for their inherent biases. Nicholson et al (2021) give an example of this in the context of the current Covid-19 epidemic in the UK, using data from monthly randomised prevalence surveys to de-bias prevalence estimates derived from routine testing.

EHR systems with national reach are found predominantly, but not exclusively, in the wealthiest countries. The 100 Million Brazilian Cohort (Brazil 100M, https://cidacs.bahia.fiocruz.br/en/platform/cohort-of-100-million-brazilians/) incorporates comprehensive health outcome and other data from the more than 100 million Brazilian residents who have been in receipt of social benefits, hence predominantly those in the poorer half of the country's overall population in excess of 200 million. Individual-level data accrue continuously in time and are georeferenced to small-area (typical population around 600 people) census units. De-identified research-quality data-sets are extracted periodically using probabilistic linkage methods (Pita et al, 2018); the potential to combine these with real-time health outcome and climate data is as yet untapped but could be transformative, especially if integrated with real-time climate information such as is generated by Brazil's Ministry of Science, Technology and Innovations (INPE,)

Mobile phone technology also presents an opportunity to supplement data from randomised studies with much more extensive, albeit potentially biased, health outcome data accruing in real-time. A second example from the current Covid-19 epidemic in the UK is the ZOE Covid-19 Symptom App (Menni et al, 2020), whereby participants are invited to submit daily symptom reports and, in the event of their being classified by an algorithm as a possible Covid-19 case, are offered a diagnostic test. Smart-phone-based systems are not yet feasible for use in poor countries, but basic mobile phone technology has reached very wide penetration even in the poorest countries and could form the basis for the development of remote triage systems such as the UK's former nationwide NHS Direct service (Munro et al, 2000), now replaced by 44 local NHS111 services (Pope et al 2017).

National-level data infrastructure initiatives

Population research UK (PRUK) is a newly established initiative being conducted by Health Data Research UK (https://www.hdruk.ac.uk/) on behalf of the UK's Economic and Social Research Council, Medical Research Council and the Wellcome Trust. Its declared vision is that "bringing studies and data together, PRUK will enable a greater understanding of the complex interplay between biological, social, economic and environmental determinants of health, social and economic outcomes, and address high-impact research questions that single studies cannot address alone" https://www.hdruk.ac.uk/population-research-uk/). It will seek to "maximise the use, innovation and benefit from the UK's rich collection of LPS across social and economic, and biomedical science." (HDRUK, 2021). PRUK is therefore a large, UK-wide data infrastructure project that aims to facilitate cross-LPS collaborative working, rather than to initiate substantive research projects in its own right. In this sense, it mimics the role that biobanks serve as underpinning infrastructure for molecular-level research; see, for example, the UK Biobank (https://www.ukbiobank.ac.uk/). The prospect of being able to combine cohort, biobank and climate data is tantalising but, for the foreseeable future, may be confined to high-income-country settings.

The INDEPTH Network

INDEPTH (Sankho and Osman, 2012, http://www.indepth-network.org/) is a global network of Health and Demographic Surveillance Systems (HDSSs) that collect wide-ranging longitudinal health and demographic data on populations from 56 HDSS sites spread across 21 countries in Africa, Asia and Oceania (Figure 12). The main goal of INDEPTH is to create a rich and, to some extent, harmonised dataset of reliable population-based data on health across many LMICs and thereby to foster research aimed at filling knowledge gaps in the epidemiology of these countries. Linkage of population and health facility data is facilitated by use of the Comprehensive Health and Epidemiological Surveillance System (CHESS), whose ambition is to "use technological solutions to establish an integrated electronic surveillance system combining all relevant data sources and allowing for appropriate response" (Sankho, 2015).

Unlike a classic cohort study, an HDSS follows up the entire population of its geographical catchment. The population monitored at a site varies between 8,000 and 262,000 (Figure 12) with a total of approximately 4.5 million. Most HDSSs (58%) are located in predominantly rural settings, 6% are in urban centres and the remaining 36% cover a mix of urban and rural areas. Collectively, they span a range of ecological zones and collect data on many different health outcomes and socio-economic indicators (Figure 13). In all sites, data on births, pregnancy outcomes, deaths and migration are collected with follow-up intervals between 3 and 24 months. Additional data on morbidities, vaccination, HIV, malaria and non-communicable diseases are also collected at some of the sites either routinely or through specific embedded research projects. Auxiliary demographic and socio-economic variables are also collected, sometimes with a coarser temporal resolution. GPS coordinates of either the house or the village of the individual are almost always available to enable spatial analyses. Although every site taken individually usually spans a limited geographic area, the whole network covers several climatic and ecological zones. A small number of the sites also conduct studies related to climate change or were set up specifically to cover different climatic areas.

INDEPTH represents a valuable and wide-ranging potential source of data for research at the climate-health interface, although Hinga, Molyneux and Marsh (2021) note that HDSSs "occupy a grey area between research, healthcare and public health practice and it is unclear how ethics guidance that rely on a research-practice distinction apply to HDSSs."

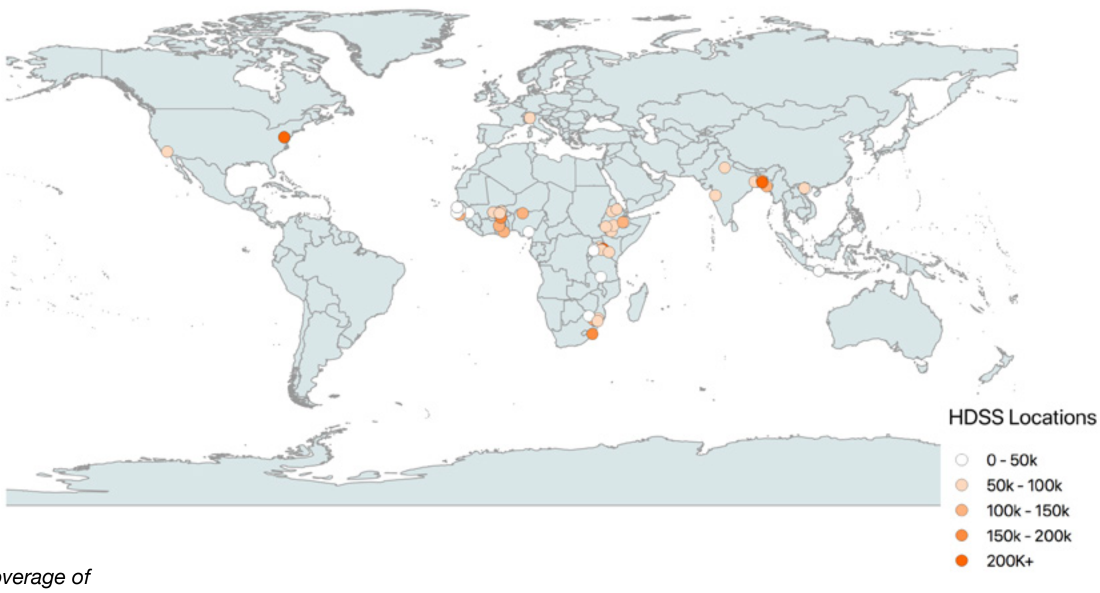
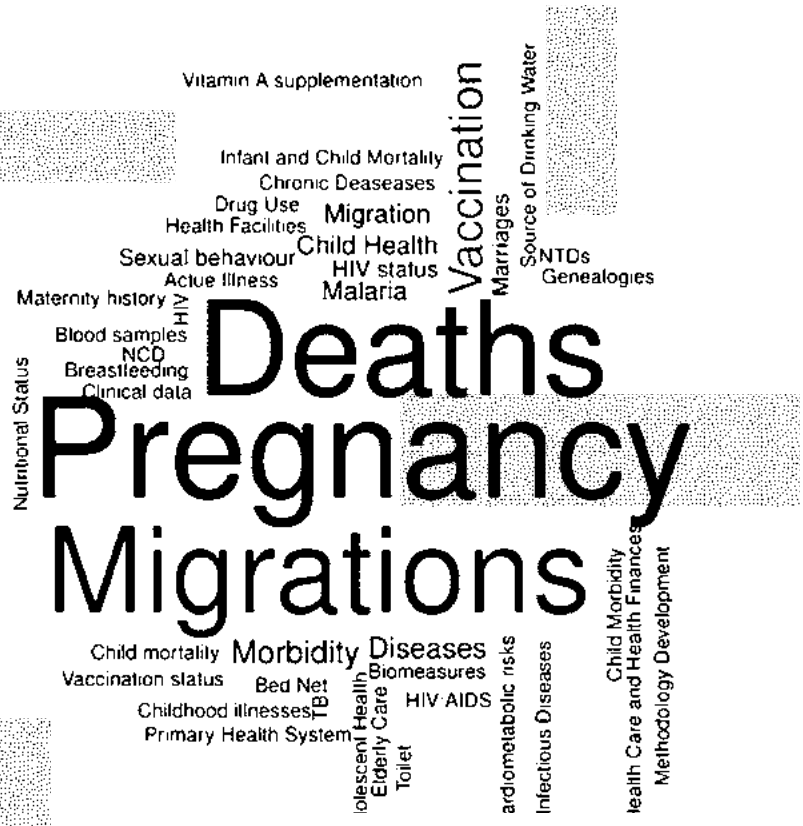


Figure 12.
Geographical coverage of the INDEPTH network.

In this word cloud, the size of each word indicates how many HDSS collect data on that domain.

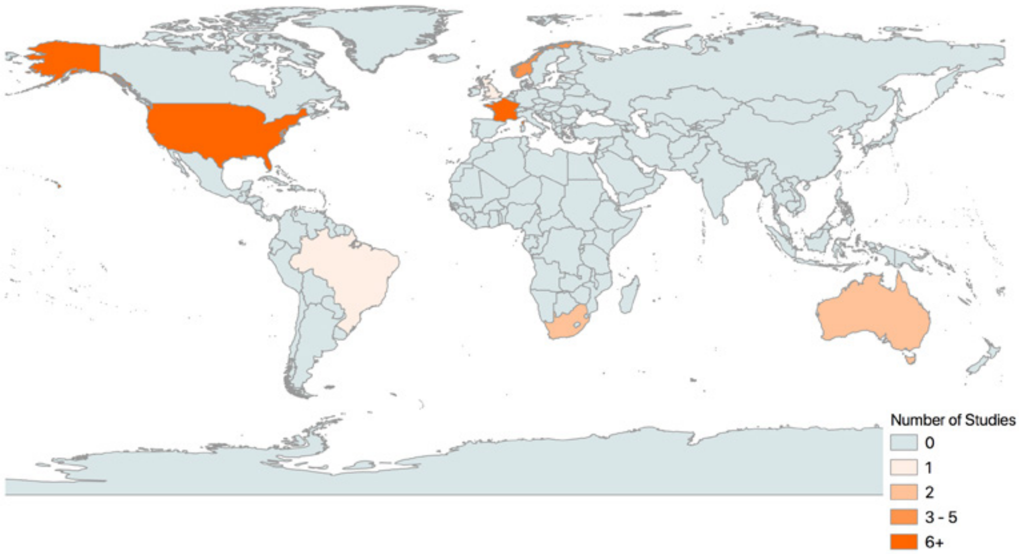


The Manhica Health Research Centre, established in 1996 in a rural area of southern Mozambique, currently follows around 92,000 individuals living in approximately 20,000 enumerated and geo-positioned households. The centre is located opposite the Manhica District Hospital.

The AGRICOH consortium

The distinctive feature of AGRICOH (https://agricoh.iarc.fr/cohorts/index.php) is its focus on the association between agricultural exposures and health outcomes, and especially on rare exposures and outcomes for which data pooling is particularly beneficial. Its 29 studies span 13 countries in both developed and low- and middle-income settings (Figure 13).

Figure 13. *Geographical coverage of the AGRICOH consortium*



The aim of the consortium is to create a harmonised dataset to extend the overall geographical coverage and power of studies that look at the relationship between multiple types of agricultural and environmental exposures and health outcomes, with a focus on relatively rare outcomes such as certain types of cancer and neurologic and auto-immune diseases (Leon et al., 2011). The individual cohorts have a broad definition of agricultural exposures. Most of them use questionnaires to gather their exposure data. The size of the population monitored per country can be as small as 300 (Uganda and Costa Rica) up to almost 900,000 individuals followed up in Norway, with a total population of 1,212,822. Biological specimens are also collected in 16 of these cohorts.

The Human Early Life Exposome

The Human Early Life Exposome (HELIX, https://www.projecthelix.eu/) study is unique among the consortia that we have reviewed in its explicit focus on environmental exposures and their impact on health in early life. It brings together six birth cohort studies, one in each of six European countries: France, Greece, Lithuania, Norway, Spain and the UK. Its aim is to investigate the interplay between early-life environmental exposures, biomolecular markers and child health (Maitre et al, 2018). Early life exposure is measured by looking at three domains: outdoor exposure (climate, urban, environmental and societal factors), individual exposure (smoking, diet and physical activity) and an internal domain (gene expression, metabolism). Environmental exposures were estimated using geospatial models to generate gridded data products that that could be linked to existing health outcome data. Examples of the environmental exposures considered include atmospheric pollutants (NO2, Pm2.5, Pm10), vegetation, land surface temperature, population density and building density.

The six cohorts provide data on 31,472 mother-child pairs from singleton births between 1999 and 2010. For a subset of 1,301 pairs, biomarkers, omics signatures and child health outcomes were measured at age 6-11. For this sub-cohort, accurate data on children's mobility and commuting patterns were collected to allow an accurate estimate of outdoor exposures at different locations and for different time-activity patterns. Pregnant women were followed-up at least once during pregnancy, at birth and then multiple times after delivery, the exact frequency of data collection varying according to the individual cohort.

5

Discussion

Discussion

The multiple spatial and temporal scales of variation in climate can affect health and its upstream determinants across the full range of eco-epidemiological levels of organisation: from the molecular to the individual, community and population levels (Figure 14). In general, larger temporal and spatial scales of climate may be more likely to affect the upper levels of the eco-epidemiological hierarchy, with large-scale climate events such as a strong El Niño event exposing a whole region and thus affecting population health either directly or indirectly. However, one can easily think of counter-examples, such as where small-scale weather exposures in important agricultural areas could affect the health of populations across a wide geographical range through their impacts on food supply, distribution or pricing. Thus, climate and health is an inherently multi-scale problem, cutting across temporal, spatial and eco-epidemiological scales. It is also inherently multi-variate and multi-disciplinary: there are many different climate parameters that affect different aspects of health, both directly and indirectly, via the socioeconomic, behavioural, entomological, zoonotic and environmental factors that determine the health of populations.

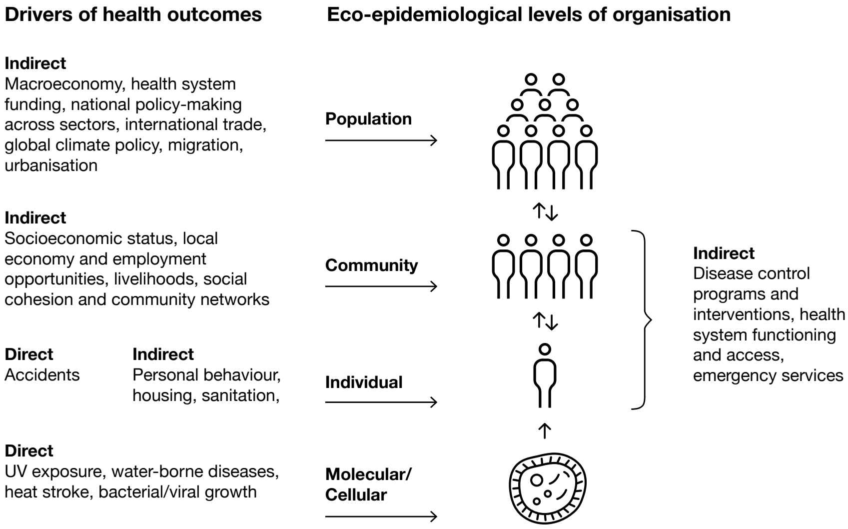


Figure 14.
Climate's effects on health across eco-epidemiological levels of organisation. Climate can either affect health directly or indirectly, when its impacts are mediated by the socioeconomic and natural systems that underpin health risk.

Such complexity necessitates a systems approach. No one model, discipline or method can tackle the full spread of interactions and scales. A systems approach can involve bringing together multiple strands of evidence from different disciplinary and inter-disciplinary methods and scales of analysis to build up an understanding of the system as a whole. A whole-system understanding is needed in order better to manage uncertainty, identify entry points for leverage and weigh up the benefits and potential disadvantages or unintended consequences of interventions to protect or promote health in a changing climate (Fiksel, 2017; Ahn et al, 2006). Consistent results from research across different climate regions and timescales of weather and climate variability can strengthen evidence of climate signals in health that are difficult to discern. In most cases, if climate is an important driver we would expect to see that association across the different timescales of climate variability. A systems approach is also a useful way to keep sight of the direct and indirect pathways of impact, many of which are currently unquantifiable due to a lack of adequate data or methods for analysis. Within such a whole-system approach, findings from climate-health research can and should inform policy-directed research but, as noted in Section 1, need to be embedded within a wider multi-disciplinary framework. Social theory, behavioural science, economics and geo-politics all have important roles to play in the translation of research findings into public health policy and practice.

Understanding the effects that climate change is already having on health and how those impacts may evolve in the future is a priority research area (see Appendix B). Causal modelling of long-term trends in climate and health requires time series spanning several decades, but these are unavailable in most of the world, and even when they are available there are too many factors at play to avoid spurious associations confidently. Research on the role of shorter-term climate variability (extreme weather, seasonality and inter-annual variability) in driving variation in health outcomes (or its drivers) can confirm associations and unpack the mechanisms involved. Shorter-term variability thus constitutes an important line of evidence in explaining the role of climate change as a long-term driver of health. Detection and attribution to climate change of the health impacts of individual extreme weather and climate events is another line of evidence (Ebi et al, 2017). When it comes to predicting the future, modelled projections of the health impacts of climate change are highly unreliable (Nissan, Ukawuba and Thomson, 2018; Nissan and Conway, 2018; Nissan et a. 2019), so research which develops an understanding of the processes via which climate influences health is the best basis from which to conjecture what may happen in the future. From this understanding, emerging and shifting risks can be managed with the help of environmental surveillance and forecasts at shorter, more skilful lead-times.

Research on shorter timescales, from weather events to interannual fluctuations, is also needed to determine health-relevant thresholds and to support planning and evaluation of adaptation measures. Some adaptation strategies only have to be implemented once, such as retrofitting infrastructure to reduce overheating in buildings. However, many other interventions will require managing the impacts of weather and climate shocks and stressors on health, for example via seasonal planning and meteorological early warning systems.

Large-scale climate variability offers a number of opportunities to address the climate-health research agenda. First, research that spans multiple climate regions (for instance, through combining multiple studies in an LPS consortium) can potentially circumvent the problem of short data series in many parts of the world by sampling a wider range of exposures across regions. Second, a major research gap exists in our understanding of the indirect pathways of climate's influence on health. Studies can address this gap through 1) comparative analyses between regions with similar climate exposures but different socioeconomic situations and 2) analyses spanning long time periods that can explore the role of changing socioeconomic contexts over time in mediating health outcomes in one or several locations.

An ambitious research agenda for climate and health requires a new conceptual model and new analytical methods to tackle the different scales of climate variability and their impacts on health. Life-course epidemiology (Ben-Shlomo and Kuh, 2002; Ben-Shlomo, Cooper and Kuh, 2016) has some attributes that may prove useful in this area. In life-course epidemiology, particular attention is given to the timing and sequencing of exposures during an individual's life course and the interrelationships among exposures, both directly and indirectly through intermediary variables. This framing could offer a means of addressing the repeated, cyclical nature of the climate, which exposes people to recurring events more often than to one-off incidents. Examples of such repeated exposures include the seasonal cycle, the extreme rainfall and flooding experienced during periodic El Niño events in some regions and the repeated, protracted episodes of drought in the Sahel. These repeated exposures could interact and combine to influence other health drivers through a 'chain of risks', a concept already espoused within 'accumulation of risk' models (Kuh et al, 2003), whereby climate shocks can tip people into a downward spiral or accelerate a decline in quality of life and health by making other adverse outcomes more likely. For example, a shock to household income following a failed harvest could lead to poor educational attainment and reduced socioeconomic status, with consequent effects on health. Sustainable development theory has long considered the effects of repeated shocks to developmental drivers such as livelihoods, housing, access to functioning social systems and financial stability, all of which underpin health risk. Epidemiologic theory can draw from ways of conceptualising climate shocks and stressors in other fields.

6

Conclusions and recommendations

Conclusions and recommendations

6.1. Conclusions

Were we to design from scratch the data infrastructure to underpin a deep investigation of the climate-health system, it would not look like existing LPS and climate datasets. For acute health outcomes, intra-annual data are needed, rather than the (at best) annual follow-up schedules of most existing LPS. To capture the spatial variability in exposures, climate data are generally needed at finer resolutions (in key locations) than existing meteorological infrastructure can provide. In our view, a deep understanding of climate-health interactions would be advanced substantially by an additional layer of data from a series of factorial experimental designs that collectively cover several key dimensions, including (but not necessarily limited to): wealthy country vs LMIC settings; within-country socio-economic variation; tropical vs temperate climatic zone; urban vs rural populations; areas that are or are not impacted by ENSO effects; infectious vs vector-borne vs non-communicable diseases; small islands and coastal locations vs continental interiors. Though global in reach, the density of this network of monitoring sites would need to be highly variable, guided not by national borders but by climatic or ecological zones. More sites are needed in regions where the climate is particularly complex, with high spatial heterogeneity caused, for example, by altitudinal and land-cover changes. In regions with less complex geography, such as large plains and deserts, fewer monitoring sites are needed.

This series of sentinel sites could record high frequency meteorological and environmental data and sub-annual (e.g. monthly or quarterly) health data. If well-designed, such a platform could support a multi-disease research agenda and would generate knowledge about climate and health across the full range of climate timescales, from extreme events, sub-seasonal and seasonal patterns to longer-term variability and trends. It would support an improved understanding of the role of indirect factors in mediating these impacts through comparative studies across multiple sites. Crucially – and often undervalued – this type of surveillance would support operational activities to protect and promote health in a changing climate (e.g. through dynamic risk mapping to target vector control measures) and to evaluate and improve upon interventions. It would also make a critical contribution to the formation of hypotheses concerning the mechanistic pathways of climate's impact on health.

Because, in many cases, there are countless potential pathways through which climate could affect health, detailed mechanistic research is not feasible until there is a plausible hypothesis to test, which then dictates the data requirements. Data from the sentinel network could then be supplemented with in-depth studies to explore these causal mechanisms by collecting additional data at higher spatial and temporal resolutions as needed, potentially making use of mobile technologies to record accurate climate exposures as people move around and/or forecast-based surveillance to observe how extreme climatic (or downstream environmental) events affect health.

Despite the challenges of mismatched characteristics and inadequate temporal and spatial coverage, there are opportunities to make advances using existing climate and health datasets. Secondary analysis of data from existing LPS and climate data products has a role to play in exploratory analysis, hypothesis formulation and identification of the most suitable sites for the design and execution of new, confirmatory and in-depth studies. We have not found any existing combination of climate and health data that is well-suited to answer questions around the causal effects of particular aspects of climate on particular health outcomes. However, we have identified a number of ongoing LPS that we feel could be reoriented in this way by supplementing existing data with in-depth, hypothesis-driven studies. These additional studies could include embedded randomised trials of specific interventions and individual exposure monitoring using wearable devices, the latter being especially important for understanding the differences between indoor and ambient exposure. Our survey was necessarily constrained to a sample of LPS and there are no doubt many other health datasets available, especially in national repositories, that would be useful for more localised research and operations.

Underlying all of the specific recommendations below is the need for deep engagement between the climate, environmental, health and (particularly for policy-directed research) social sciences. The continuing growth in the availability of massive datasets presents both opportunities and threats. The opportunities are obvious, the threats less so. One, already mentioned earlier, is the risk of finding spurious associations through the use of context-free algorithms. Another is that taking data at face value can all too easily lead to both climate and health data being used inappropriately and without due consideration of their inherent uncertainties, many of which may be unquantifiable. In the health sciences, there is a need to extend, and accelerate, current work on data harmonisation to enable the unambiguous merging of ostensibly comparable datasets from multiple studies. For example, data on the prevalence of a particular disease needs to be accompanied by meta-data including: the precise definition of the target population; the eligibility and exclusion criteria for individuals to be included in the sample; the non-compliance rate and what, if any, reasons were given for non-compliance; the diagnostic used, and its sensitivity and specificity. On the climate data side, there is a need to extend the availability to researchers of ground-truth data by engaging with in-country data-owners, such as national meteorological services, rather than restricting their attention to freely available data products (see Figure 3) and to promote the more judicious use of climate data products in health research to account fully for their uncertainties. We would also hope to see a commitment from the meteorological community (through the World Meteorological Organisation) to developing the data and climate services needed specifically to serve the health sector, which has its own requirements separate from other sectors.

We conclude that an ambitious and effective climate-health research strategy should:

- be policy-driven, aiming to influence strategies to adapt to climate variability and change and mitigate further global warming;
- adopt a systems approach to tackle the multiple temporal, spatial and eco-epidemiological scales of climate-health interactions and the indirect pathways by which climate drives health outcomes;
- develop and support initiatives for the co-production of study-design and data collection protocols by health, climate, environmental and statistical scientists, as well as other relevant disciplines;
- develop the field of climate-health research, recognising that multiple disciplines are required to work together in a deep and integrated way to address the climate-health challenge;
- recognise and seek to influence the governance structures needed for the proper integration of climate, health, socioeconomic and environmental data;
- most ambitiously, establish an international centre of excellence populated by health, climate, environmental, statistical and social scientists to undertake policy-directed research at the climate-health interface with a focus on areas of the world (primarily LMICs) that lack the necessary resources and whose populations are most vulnerable to the health effects of climate change.

6.2. Recommendations

To work towards these ambitions, we identify below a set of specific recommendations for activities that the Wellcome Trust could undertake.

Immediate: Assess Existing Health Datasets For Climate Analysis

A metadata analysis is the first step in determining the suitability of existing health datasets for climate analysis. Given the complexity of the data required to capture climate-health effects on different spatial and temporal scales, datasets which at first appear suitable for epidemiological research can transpire to be incompatible upon further examination. The survey of LPS presented in Section 4 provides an initial assessment of the potential for using existing LPS for climate analysis, but a deeper exploration is required, which would be greatly assisted by a digital platform to visualise key metadata.

Below, we list some of the key metadata to be collected. Note that, although data are often aggregated in space and time to achieve adequate sample sizes, data collection is usually staggered and more precise temporal and geo-referencing of each observation may sometimes be available.

General attributes

- LPS type: cohort, panel, repeated cross-section
- Are the timing and location of data recorded?
- Sampling design
- Number of participants
- Number of sites/study regions

Spatio-temporal attributes

- Temporal coverage
- Frequency of surveillance
- Precision of temporal referencing
- Timing of surveillance during the year
- Geographical coverage
- Spatial resolution
- Precision of geo-referencing

Short term

1. Fund proposals on the following topics under existing grant and fellowship schemes:
   aa) secondary analyses of existing LPS and climate data to develop hypotheses and inform the design of studies on specific climate-health interactions;

   b) projects that capitalise on opportunities for the integrated analysis of data from multiple LPS;

   c) development of novel statistical and computational methods for inferentially robust combined analysis of multiple health and climate data-sources;

   d) projects to support better understanding of the indirect drivers in climate-health pathways and better linkage with relevant data types e.g. socio-economic census data;

   e) projects to construct new retrospective cohorts and corresponding climate data and metrics;

   f) projects that engage with national health and meteorological agencies to enable all relevant data from both domains to be harnessed for climate-health research at local scales.
2. Commission selected LPS consortia to consider how they could re-orient some of their work towards climate-health research, engage directly with climate data owners and scientists and develop specific proposals accordingly. Candidates include the Hundred thousand Plus Cohorts Consortium, AGRICOH, HELIX and the successful bidder for the African Population Cohorts Consortium.
3. Engage in discussion with Brazil 100M and INPE with a view to developing an exemplar real-time climate and health surveillance system based on country-wide, routinely collected health information.

Medium term

1. Encourage new consortium-based approaches that integrate climate data and health data across wide-ranging geographies and are co-designed by experts from both domains.
2. Engage the disaster risk community to develop funding opportunities for pilots to explore forecast-based surveillance as a means of studying the causal pathways between extreme weather or climate events and health outcomes, and to evaluate the effectiveness of interventions.
3. Advocate to the World Meteorological Organisation for the inclusion of the health sector as a priority user for the services of national meteorological agencies and propose specific meteorological data and services requirements needed to address climate-health priorities (e.g. in urban areas).

Long term

1. Develop a vision for a Wellcome Trust Climate and Health Institute with global reach but a particular focus on policy-directed research in LMIC settings. The INDEPTH network would be a useful starting point for this activity; complementing its health data system with an equally rich climate data system would create a formidable resource for climate-health research rooted in LMICs.
2. Develop a strategy for the generation and use of routine health information systems to capture and analyse real-time or near-real-time health data in lower income countries.
3. Develop the design for a network of sentinel sites taking frequent health, socio-economic and climatic measurements across representative climatic regions/ exposures and socioeconomic contexts, with a view to this platform serving a multi-disease research and operations agenda.

References

Ahn, A.C., Tewari, M., Poon, C.-S. and Phillips, R.S. (2006). The clinical applications of a systems approach," *PLOS Medicine*, **3**, e209, 2006, doi: 10.1371/JOURNAL.PMED.0030209.

Alexander, L. V. *et al.* (2006). Global observed changes in daily climate extremes of temperature and precipitation. *Journal of Geophysical Research*, **111**, D05109, doi: 10.1029/2005JD006290.

Arshad, M. *et al.* (2017). Climate variability and yield risk in South Asia's rice-wheat systems: emerging evidence from Pakistan. *Paddy Water Environment*, **15**, 249–261, doi: 10.1007/s10333-016-0544-0.

BBC Bitesize. 2021. Tropical cyclone case study - Typhoon Haiyan - Tropical cyclones - Edexcel - GCSE Geography Revision - Edexcel - BBC Bitesize. [online] Available at: <https://www.bbc.co.uk/bitesize/guides/z9whg82/revision/4> [Accessed 12 December 2021].

Ben-Shlomo, Y., Cooper, R. and Kuh, D. (2016). The last two decades of life course epidemiology, and its relevance for research on ageing. *International Journal of Epidemiology*, **45**, 973–988, doi: 10.1093/ije/dyw096.

Ben-Shlomo, Y. and D. Kuh, D. (2002). A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. *International Journal of Epidemiology*, **31**, 285–293, doi: 10.1093/ije/31.2.285.

Burkart, K., S. Breitner, S., A. Schneider, A., M. H. Khan, M.H., A. Kramer, A. and W. Endlicher, W. (2014). An analysis of heat effects in different subpopulations of Bangladesh. *International Journal of Biometeorology*, **58**, 227–237, doi: 10.1007/s00484-013-0668-5.

Chen, Z., Chen, J., Collins, R., Guo, Y., Peto, R., Wu, F. and Li, L. on behalf of the China Kadoorie Biobank (CKB) collaborative group (2011). China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. International Journal of Epidemiology, **40**, 1652–1666

Chipeta, M., Terlouw, D.J., Phiri, K. and Diggle, P.J. (2016). Adaptive geostatistical design and analysis for sequential prevalence surveys. *Spatial Statistics*, **15**, 70-84.

Congdon, P.D. (2020). Bayesian Hierarchical Models: with applications using R (Second Edition).Boca Raton: CRC Press

Coughlan de Perez, E. *et al.* (2018). Global predictability of temperature extremes. *Environmental Research Letters*, **13**, doi: 10.1088/1748-9326/aab94a.

Davis, K.F., Downs, S. and Gephart, J.A. (2021). Towards food supply chain resilience to environmental shocks. *Nature Food*, **2**, 54–65, doi: 10.1038/s43016-020-00196-3.

Dinku, T. *et al.* (2014). Enhancing national climate services (ENACTS) for development in Africa. *Climate and Development*, **10**, 664–672, doi: 10.1080/17565529.2017.1405784.

Dinku, T. *et al.* (2016). The ENACTS Approach. Transforming climate services in Africa one country at a time. *World Policy Papers*. Available online at: http://www.worldpolicy.org/sites/default/files/The ENACTS Approach-Transforming Climate Services in Africa One Country at a Time.pdf

Doiron, D., Burton, P., Marcon, Y., Gaye, A., Wolffenbuttel, B.H.R., Perola, M., Stolk, R.P., Foco, L., Minelli, C., Waldenberger, M., Holle, R., Kvaløy, K., Hillege, H.L., Tassé, A.M. and Ferretti, V. (2013). BioSHaRE project. Emerging Themes in Epidemiology, **12**. doi:10.1186/1742-7622-10-12.

Donat, M.G. *et al.* (2013a). Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: the HadEX2 dataset. *Journal of Geophysical Research: Atmospheres*, **118**, 2098–2118, doi: 10.1002/jgrd.50150.

Donat, M.G. *et al.* (2013b). Global land-based datasets for monitoring climatic extremes. *Bulletin of the American Meteorological Society*, **94**, 997–1006, doi: 10.1175/BAMS-D-12-00109.1.

Ebi, K.L., Ogden, N.H., Semenza, J.C. and Woodward, A. (2017). Detecting and attributing health burdens to climate change. *Environmental Health Perspectives*, **125**, doi: 10.1289/EHP1509.

Evans, A., Salomaa, V., Kulathinal S, Asplund, K., Cambien, F., Ferrario, M., Perola, M., Peltonen, L., Shields, D., Tunstall-Pedoe, H. and Kuulasmaa, K., for the MORGAM Project. (2005). MORGAM (an international pooling of cardiovascular cohorts). *International Journal of Epidemiology,* **34**,21-27

Fanzo, J., Davis, C., McLaren, R. and Choufani, J. (2018). The effect of climate change across food systems: Implications for nutrition outcomes. *Global Food Security*, **18**, 12–19, doi: 10.1016/J.GFS.2018.06.001.

Fiksel, J. (2017). Sustainability and resilience: toward a systems approach. *Sustainability: Science, Practice and Policy*, **2**, 14–21, doi: 10.1080/15487733.2006.11907980.

Friel, S., Bowen, K., Campbell-Lendrum, D., Frumkin, H., McMichael, A.J. and Rasanathan, K. (2011). Climate change, noncommunicable diseases, and development: the relationships and common policy opportunities. *Annual Review of Public Health*, vol. **32**, 133–147, doi: 10.1146/annurev-publhealth-071910-140612

Gasparrini, A., Guo, Y., Hashizume, M., Lavigne, E., Zanobetti, A., Schwartz, J., Tobias, A., Tong, S., Rocklöv, J., Forsberg, B., Leone, M., De Sario, M., Bell, M.L., Guo, Y. L., Wu, C., Kan, H., Yi, S., de Sousa Zanotti, M., Coelho, S., Saldiva, P.H.N., Honda, Y., Kim, H. and Armstrong, B. (2015). Mortality risk attributable to high and low ambient temperature a multi-country observational study. *Lancet*, **306**, 369-375.]

Gleixner, S., T. Demissie, T. and G. T. Diro, G.T. (2020). Did ERA5 improve temperature and precipitation reanalysis over East Africa? *Atmosphere,* **11**, 996, doi: 10.3390/ATMOS11090996.

ALSPAC±The Avon Longitudinal Study of Parents and Children I. Study methodology Jean Golding a , Marcus Pembrey b , Richard Jones a,b and the ALSPAC Study Team ALSPAC±The Avon Longitudinal Study of Parents and Children

Golding, J., Pembrey, M., Jones, R and the ALSPAC Study Team (2001). ALSPAC. The Avon longitudinal study of parents and children I. Study methodology. *Paediatric and Perinatal Epidemiology*, **15**, 74-87.

Green, D. (2016). The spatial distribution of extreme climate events, another climate inequity for the world's most vulnerable people. *Environmental Research Letters*, **11**, 091002.

Gulcebi, M.I., Bartolini, E., Lee, O. et al (2021). Climate change and epilepsy: insights from clinical and basic science. *Epilepsy and Behavior*, **116**, https://doi.org/10.1016/j.yebeh.2021.107791

Hagan, J.E., Moraga, P., Costa, F., Capian, N., Ribeiro, G.S., Wunder, E.A., Felzemburgh, R.D.M., Reis, R.B., Nery, N., Santana, F.S., Fraga, D., dos Santos, B.L., Santos, A.C., Queiroz, A., Tassinari, W., Carvalho, M.A., Reis, M.G., Diggle, P.J. and Ko, A.I. (2016). Spatiotemporal determinants of urban leptospirosis transmission: Four-year prospective cohort study of slum residents in Brazil. *Public Library of Science: Neglected Tropical Diseases*, **10**, e0004275

Harrington, L.J., Frame, D.J., Fischer, E.M., Hawkins, E. and Joshi, M. and Jones, C.D. (2016 Poorest countries experience earlier anthropogenic emergence of daily temperature extremes. *Environmental Research Letters*, **11**, 055007

Harrington, L.J., Frame, D.J., Hawkins, E. and Joshi, M. (2017). Seasonal cycles enhance disparities between low and high-income countries in exposure to monthly temperature emergence with future warming. *Environmental Research Letters*, **12**, 114039

HDRUK (2021). Green paper Consultation on recommendations for developing Population Research UK. https://www.hdruk.ac.uk/wp-content/uploads/2021/07/PRUK_Green-paper.pdf

Hersbach, H. *et al.* (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, **146**, 1999–2049, doi: 10.1002/QJ.3803.

Hinga, N., Molyneux, S. and Marsh, V. (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, e004008. doi:10.1136/ bmjgh-2020-004008

Karl, T.R. and Knight, R.W. (1997). The 1995 Chicago heat wave: how likely is a recurrence? Bulletin of the American Meteorological Society, 78, 1107–1119, doi: 10.1175/1520-0477(1997)078<1107:TCHWHL>2.0.CO;2.

Koppe, C. and Jendritzky, G. (2005). Inclusion of short-term adaptation to thermal stresses in a heat load warning procedure. Meteorologische Zeitschrift, 14, 271–278, doi: 10.1127/0941-2948/2005/0030.

Kuh, D., Ben-Shlomo, Y., Lynch, J., Hallqvist, J. and C. Power, C. (2003). Life course epidemiology. *Journal of Epidemiology & Community Health*, **57**, 778–783, doi: 10.1136/JECH.57.10.778.

Laaidi, K. *et al.* (2012). The impact of heat islands on mortality in Paris during the August 2003 heat wave. *Environmental Health Perspectives*, **120**, 254–259, doi: 10.1289/ehp.1103532.

Lau, C.L., Smythe, L.D., Craig, S.B. and Weinstein, P. (2010). Climate change, flooding, urbanisation and leptospirosis: fuelling the fire? *Transactions of The Royal Society of Tropical Medicine and Hygiene*, **104**, 631–638.

Leon, M.E., Freeman, L.E.B., Douwes, J., Hoppin, J.A., Kromhout, H., Lebailly, P., Nordby, K.-C., Schenker, M., Schüz, J., Waring, S.C., Alavanja, M.C.R., Annesi-Maesano, I., Baldi, I., Dalvie, M.A., Ferro, G., Fervers, B., Langseth, H., London, L., Lynch, C.F., McLaughlin, J., Merchant, J.A., Pahwa, P., Sigsgaard, T., Stayner, L.T., Wesseling, C., Yoo, K.-Y., Zahm, S.H., Straif, K. and Blair, A. (2011). AGRICOH: A Consortium of Agricultural Cohorts. *International Journal of Environmental Research and Public Health* 2011, **8**, 1341-1357. https://doi.org/10.3390/ijerph8051341

Lyon, B., Dinku, T., Raman, A. and Thomson, M.C. (2017). Temperature suitability for malaria climbing the Ethiopian Highlands. *Environmental Research Letters*, **12**, doi: 10.1088/1748-9326/aa64e6

Maitre, L., de Bont, J., Casas, M. et al (2018). Human Early Life Exposome (HELIX) study: a European population-based exposome cohort *BMJ Open*, **8** e021311. doi: 10.1136/bmjopen-2017-021311

Manolo, T.A., Goodhand, P. and Ginsburg, G. (2020). The International Hundred Thousand Plus Cohort Consortium: integrating large-scale cohorts to address global scientific challenges. *Lancet Digital Health*, **2**, e567.

Marmot, M.G., Stansfield, S., Patel, C., North, F., Head J., White, I., Brunner, E., Feeney, A. and Davey-Smith, G. (1991). Health inequalities among British civil servants: the Whitehall II study. *Lancet*, **337**, 1387-1393

Mason, S. *et al.* (2018). Climate data. In *Climate Information for Public Health Action*, Abingdon: Routledge, 125–146.

McGregor, G.R., Bessemoulin, P., Ebi, K. and Menne, B. (2015). *Heatwaves and Health: Guidance on Warning-System Development*. Geneva: World Meteorological Organization and World Health Organization

Menni, C., Valdes, A.M., Freidin, M.B. , Sudre, C.H., Nguyen, L.H., Drew, D.A., Ganesh, S. , Varsavsky, T. , Cardoso, M.J. , Moustafa, J.S.E. , Visconti, A. , Hysi, P. , Bowyer, R.C.E., Mangino, M., Falchi, M. , Wolf, J. , Ourselin, S. , Chan, A.T., Steves, C.J. and Spector, T.D. (2020). Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nature Medicine*, **26**, 1037-1040.

Munro, J., Nicholl, J., O'Cathain, A. and Knowles, E. (2000). Impact of NHS Direct on demand for immediate care: observational study. BMJ, 321, 150-153

Nicholson, G., Lehmann, B., Padellini, T., Pouwels, K.B., Jersakova, R., Lomax, J., King, R.E., Mallon, A.-M., Diggle, P.J., Richardson, S., Blangiardo, M. and Holmes, C. (2021). Local prevalence of SARS-CoV-2 infection: an integrative causal model for debiasing fine-scale targeted testing data. Nature Microbiology.

Nicholson, G., Blangiardo, M., Briers, M., Diggle, P.J., Fjelde, T.E.,Ge, T.E., Goudie, R.J.B., Jersakova, R., King, R.E., Lehmann, B.C.L., Mallon, A-M., Padellini, T., Teh, Y.W., Holmes, C. and Richardson, S. (2022). Interoperability of statistical models in pandemic preparedness: principles and reality. Statistical Science

Nissan, H., Burkart, K., Coughlan De Perez, E., van Aalst, M. and S. Mason, S. (2017). Defining and predicting heat waves in Bangladesh. *Journal of Applied Meteorology and Climatology*, **56**, 2653–2670, doi: 10.1175/JAMC-D-17-0035.1.

Nissan, H. and Conway, D. (2018). From advocacy to action: projecting the health impacts of climate change. *PLoS Medicine*, **15**, e1002624. https://doi.org/10.1371/journal.pmed.100262

Nissan, H., Thomson, M.C., Mason, S. and Muñoz, A. (2018). Climate information for adaptation. In *Climate Information for Public Health Action*, Abingdon: Routledge, 199–233.

Nissan, H. *et al.* (2019). On the use and misuse of climate change projections in international development. *WIRES Climate Change*, **10**, e579, doi: 10.1002/wcc.579.

Nissan, H., Ukawuba, I. and Thomson, M (2021). Climate-proofing a malaria eradication strategy. *Malaria Journal*, **20**, 1–16, 2021, doi: 10.1186/s12936-021-03718-x.

Openshaw, S. and Taylor, P.J. (1979). A million or so correlation coefficients: three experiments on the modifiable areal unit problem. In N. Wrigley, ed. *Statistical Applications in the Spatial Sciences*, 127–144. London: Pion.

Peng, R.D., Bobb, J.F., Tebaldi, C., McDaniel, L., Bell, M.L. and Dominici, F. (2011). Toward a Quantitative Estimate of Future Heat Wave Mortality under Global Climate Change. *Environmental Health Perspectives*, **119**, 701-706.

Pita, R., Pinto, C., Sena, S., Fiaccone, R., Amorim, L., Reis, S., Barreto, M.L., Denaxas, S., and Barreto, M.E. (2018). On the accuracy and scalability of probabilistic data linkage over the Brazilian 114 Million Cohort. *IEEE Journal of Biomedical and Health Informatics*, **22**, 346-353.

Pomposi, C., Kushnir, Y. and Giannini, A. (2014). Moisture budget analysis of SST-driven decadal Sahel precipitation variability in the twentieth century. *Climate Dynamics*, **44**, 3303–3321, doi: 10.1007/S00382-014-2382-3.

Pope, C., Turnbull, J., Jones, J., Prichard, J., Rowsell, A. and Halford, S. (2017). Has the NHS 111 urgent care telephone service been a success? Case study and secondary data analysis in England. *BMJ Open*, **7**, e014815.

Riley, S., Ainslie, K.E.C., Eales, O., Walters, C.E., Wang, H., Atchison, C., Fronterre, C., Diggle, P.J., Ashby, D., Donnelly, C.A., Cooke, G., Barclay, W., Ward, H., Darzi, A., and Elliott, P. (2021). Resurgence of SARS-CoV-2 in England: detection by community antigen surveillance. *Science*, **372**, 990-995.

Rooney, C., McMichael, A.J., Kovats, R.S. nd Coleman, M.P. (1998). Excess mortality in England and Wales, and in Greater London, during the 1995 heatwave. *Journal of Epidemiology & Community Health*, **52**, 482–486, https://doi.org/10.1136/jech.52.8.482

Sankoh, O. and Byass, P. (2012). The INDEPTH Network: filling vital gaps in global epidemiology, *International Journal of Epidemiology*, **41**, 579-588, https://doi.org/10.1093/ije/dys081

Sankho, O. on behalf of the INDEPTH Network (2015). CHESS: an innovative concept for a new generation of population surveillance. *Lancet Global Health*, **3**, e742.

Seneviratne, S.I. et al (2021). Weather and Climate Extreme Events in a Changing Climate. In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press.

Sheridan, S.C. and Kalkstein, L.S. (2004). Progress in heat watch-warning system technology. *Bulletin of the American Meteorological Society*, **85**, 1931–1941, doi: 10.1175/BAMS-85-12-1931.

Shmueli, G. (2010). To explain or to predict? *Statistical Science*, **25**, 289-310.

Sisodiya, S.M., Scheffer, I.E., Lowenstein, D.H. and Free, S.L. (2019). Why should a neurologist worry about climate change? *Lancet Neurology*, **18**, 335-336.

Smith, D.M *et al.* (2019). Robust skill of decadal climate predictions. *Climate and Atmospheric Science* **2,** 1–10, doi: 10.1038/s41612-019-0071-y.

Tarek, M., Brissette, F.P. and Arsenault, A. (2020). Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modelling over North America. *Hydrology and Earth System Sciences*, **24**, 2527–2544, doi: 10.5194/HESS-24-2527-2020.

The Economist. 2021. Mongolia's deadly winters are becoming more frequent. [online] Available at: <https://www.economist.com/graphic-detail/2020/01/25/mongolias-deadly-winters-are-becoming-more-frequent> [Accessed 12 December 2021].

The Guardian. 2021. Global heating 'may lead to epidemic of kidney disease'. [online] Available at: <https://www.theguardian.com/global-development/2021/oct/21/global-heating-may-lead-to-epidemic-of-kidney-disease> [Accessed 12 December 2021].

Thomson, M.C. and Mason, S.J. (2018). *Climate Information for Public Health Action*. Abingdon: Routledge.

United Nations, (2018). *World Urbanization Prospects: the 2018 Revision*.

Vaidyanathan, A., Kegler, S.R., Saha, S.S. and Mulholland, J.A. (2016). A statistical framework to evaluate extreme weather definitions from a health perspective: a demonstration based on extreme heat events. *Bulletin of the American Meteorological Society*, **97**, 1817–1830, doi: 10.1175/BAMS-D-15-00181.1.

Vins, H., Bell, J., Saha, S. and Hess, J.J. (2015). The mental health outcomes of drought: A systematic review and causal process diagram. *International Journal of Environmental Research and Public Health*, **12,** 13251–13275, doi: 10.3390/ijerph121013251.

Wadsworth, M., Kuh, D., Richards, M. and Hardy, R. (2006). Cohort Profile: The 1946 National Birth Cohort (MRC National Survey of Health and Development). *International Journal of Epidemiology*; **35**, 49–54

WHO MONICA Project Principal Investigators (1988). The World Health Organization MONICA Project (monitoring trends and determinants in cardiovascular disease): a major international collaboration. *Journal of Clinical Epidemiology*, **41**, 105-114. (doi: 10.1016/0895-4356(88)90084-4)

Image Credits

Cover	Photo by Meddy Huduti on Unsplash.
Cover	Photo by Dani Franco on Unsplash.
Page 3	Photo by Malachi Brooks on Unsplash.
Page 6	Photo by Daniel Lozano Valdes on Unsplash.
Page 8	Photo by Wolfgang Hasselmann on Unsplash.
Page 8	Photo by Janko Ferlic on Unsplash.
Page 10	Extratropical cyclone near Iceland from the Aqua/MODIS satellite. Credit: 2003 NASA.
Page 10	Photo by NOAA on Unsplash.
Page 10	Photo by Mat Napo on Unsplash.
Page 12	Photo by Etty Fidele on Unsplash.
Page 16	Photo by Daria Devyatkina on Flickr.
Page 16	A visible image of Super Typhoon Haiyan approaching the Philippines takefrom the MODIS instrument aboard NASA's Aqua satellite on Nov. 7, 2013 at 04:25 UTC/Nov. 6 at 11:25 p.m by NASA Goddard MODIS Rapid Response Team.
Page 18	A PDS Automatic Weather Station at a dam located near Brisbane, Australia. Automatic Weather Station consists of a DT50 data logger, solar panels, radio modem & meteorological sensors. Source: Pacific Data Systems Pty Ltd, http://www.pacdatasys.com.au.
Page 21	Photo by Tom Schuller on Unsplash.
Page 21	Livestock covered by snow during Mongolia's 2010 dzud. Photo by UNDP Mongolia on Flickr.
Page 21	Bones litter the ground in Azraga Bag in Central Mongolia. They are the remnants of a dzud - an unusually harsh winter that kills livestock. Photo by Taylor Weidman on The Vanishing Cultures Project.
Page 22	Photo by NOAA on Unsplash.
Page 25	Photo by Luke Michael on Unsplash.
Page 28	Photo by Arun Clarke on Unsplash.
Page 28	Photo by Dani Franco on Unsplash.
Page 29	Photo by Piron Guillaume on Unsplash.
Page 29	Photo by Patrick Hendry on Unsplash.
Page 34	Photo by Ashwini Chaudhary on Unsplash.
Page 34	Photo by Jeremy Bezanger on Unsplash.
Page 35	Photo by Ben Hershey on Unsplash.
Page 37	Photo by Jasper Wilde on Unsplash.
Page 39	Photo by Beth Macdonald on Unsplash.
Page 40	Photo by Annie Spratt on Unsplash.
Page 44	Photo by Dominic Chavez on Global Financing Facility.
Page 46	Photo by Cesar Rincon on Unsplash.
Page 46	Photo by JC Gellidon on Unsplash.
Page 46	Photo by Torsten Dederichs on Unsplash.
Page 48	Photo by Dani Franco on Unsplash.
Page 48	Photo by Dani Alaez on Unsplash.

Appendix

A B C

Appendix A
Individual consultations

Kobus Herbst (Africa Health Research Institute)
Director of the Department of Science and Technology funded South African Population Research Infrastructure Network (SAPRIN).

Mayowa Owolabi (University of Ibadan)
Dean of Clinical Sciences

Ibrahim Abubakar (University College, London)
Professor of Infectious Disease Epidemiology and Director, Institute for Global Health

Sola Olapede (University of Chicago)
Professor of Medicine and Director, International Programs

Osman Samkoh (Statistics Sierra Leone)
Statistician General of Sierra Leone and former Executive Director of the INDEPTH Network (http://www.indepth-network.org/)

Dotun Bobadoye (African Technology Policy Studies, Nigeria
Chief Operating Officer of the Global Emerging Pathogens Treatment Consortium (GET, https://www.getafrica.org/)

Judit Kumuthini (Cape Town, Western Cape)
Bioinformatics Manager

Ezinne Uvere (University of Ibadan)
Faculty of Public Health, College of Medicine

Mauricio Barreto (Fiocruz, Salvador, Brazil)
Coordinator of the Center for Data Integration and Knowledge for Health (Cidacs, https://cidacs.bahia.fiocruz.br/en/)

Liam Smeeth (London School of Hygiene and Tropical Medicine)
Director of School and Professor of Clinical Epidemiology

Tord Kjellstrom (retired, New Zealand)
Climate Change Heat Impact and Prevention initiative (https://climatechip.org)

Geoff Ginsburg (Duke University)
Professor of Medicine also Director of Duke Center for Applied Genomics and Precision Medicine, also International HundredK+ Cohorts Consortium (IHCC, https://ihccglobal.org/)

Neil Pearce (London School of Hygiene and Tropical Medicine)
Professor of Epidemiology and Biostatistics, also International Study of Asthma and Allergies in Childhood (ISAAC, http://isaac.auckland.ac.nz/)

Zhengming Chen (Oxford)
Professor of Epidemiology, also China Kadoorie biobank (https://www.ckbiobank.org/site/)

Abdisalan Noor (World Health Organisation)
Team leader, Surveillance, Monitoring and Evaluation Unit at WHO Global Malaria Programme (GMP, https://www.who.int/teams/global-malaria-programme), Geneva, Switzerland

James Pickett (Health Data Research UK)
Hub Development Manager, Population Research UK (PRUK, https://www.hdruk.ac.uk/population-research-uk/)

Elizabeth Shove (Lancaster University)
Professor, Sociology Department

Stephen Reicher (St Andrews University)
Bishop Wardlaw Professor, School of Psychology and Neuroscience

Kristie Ebi (University of Washington)
Professor, Centre for Health and the Global Environment

Tilly Alcayna (Red Cross Red Crescent Climate Centre)
Senior Technical Adviser for Health and Climate

Meghan Bailey (Red Cross Red Crescent Climate Centre)
Senior Technical Adviser

Sir Andy Haines (London School of Hygiene and Tropical Medicine)
Professor of Environmental Change and Public Health

Sabine Gabrysch (Potsdam Institute for Climate Impact Research)
Head of Research Department on Climate Resilience

Appendix B
Emerging research priorities

In preparing this report, we consulted a selection of people working at the climate-health nexus and asked them about their priorities for research and the challenges they face using health and climate data. The participant list was compiled from our own contacts, those of the team at the Wellcome Trust and those recommended by participants. These conversations were not intended as a rigorous research gap analysis, but the responses are listed below and we include here a brief summary of emerging themes. Overall, comments from participants were consistent with, but not limited to, those reported in two recent reviews of climate and health research (Berrang-Ford et al., 2021; Scheelbeek et al., 2021) which found major gaps in evidence from LMICs and in research on mental health, undernutrition, maternal and child health and adaptation or mitigation options and their consequences for health.

Several categories of research activity were identified: detecting and understanding associations between climate and health; supporting practical interventions and public health programmes; detecting the impacts of climate change on health that have already occurred and predicting future impacts; supporting and evaluating adaptation to climate change; how health-risk behaviours contribute to greenhouse gas emissions and how climate change mitigation policies affect health; methodologies needed to address climate-health research gaps.

Some common priorities emerged that cut across these categories of research. Understanding the role of mediating factors in climate-health relationships was repeatedly identified, in particular economic factors, how behaviour can modify associations, and how climate can affect health system functioning. A need to focus on LMIC countries was very clearly expressed, as was the dearth of knowledge on urban health and the sub-urban scale data needed for research and operational risk surveillance. Heat was the most frequently mentioned climate exposure and several people raised concerns about compound events (e.g. drought followed by flooding, or heat in conjunction with wildfire and poor air quality). Nutrition and communicable diseases (mainly malaria and dengue) were the health outcomes that were mentioned most often. Children were the priority population group for several participants and pregnant women were also mentioned, but it was also clear that we do not yet understand which groups are vulnerable to which aspects of climate, particularly in LMICs. Research to identify thresholds for health impacts was explicitly mentioned by some people, especially for early warning systems and operational decision-making, but is implicit in other topics that arose, such as detection and attribution of the health impacts of extreme events to climate change, which requires an event definition. Thresholds are needed for a range of climate exposures, at different levels of exposure (not just the most rare/extreme events), for different health outcomes (not just mortality) and for different population groups including those with pre-existing conditions. Few people mentioned mitigation, but adaptation was a clear priority: what are the options for adaptation? are they feasible and affordable? how can we evaluate them in practice? and how do we ensure that we are adequately prepared for an uncertain future?

Research priorities cut across the full range of climate timescales (from weather to long-term trends), with several people raising predictability of health outcomes across lead-times from days to seasons. Timescales were not always mentioned explicitly by participants, but on closer inspection it becomes apparent that the majority of research questions require analyses of weather at intra-annual and inter-annual timescales. For example, heat thresholds associated with health impacts would involve analysis using daily health and meteorological data, whereas understanding how Dzud (a very severe Mongolian winter) is connected with infectious diseases in children would require analysis on interannual timescales to sample a sufficient number of winter seasons, with particular attention given to the timing during the year of health surveillance in children (given the seasonal nature of this particular pathway). Although long-term in their perspective, research gaps related to climate change also span the full range of climate timescales, from detection and attribution studies focused on extreme events, to predicting the timing of crossing critical health thresholds in future decades, an impossible task to accomplish with any accuracy (Nissan and Conway, 2018; Nissan et al., 2021). Where long-term trend analysis was mentioned, it was generally in the context of understanding how background trends are affecting the impact of weather events, seasonality and interannual variability through shifts in baseline climate conditions.

Research questions by topic

Detecting and understanding associations between climate and health

Exploratory research

- Climate effects on waterborne disease
- How the health effects of climate impact the functioning of society
- Climate impacts on ecosystems in the Amazon and consequent effects on human health
- Environmental conditions as determinants of long-term health

Understanding pathways of impact

- Impacts of climate on nutrition via pathways other than food security
- Effects of heating in urban areas on mosquito survival and the dynamics of dengue
- Climate's indirect effects on household income and behaviour and consequences for vector-borne disease
- How is climate linked to seasonal mobility and seasonal disease dynamics?
- Climate's role in driving exposure to endocrine-disrupting chemicals via many hypothesized pathways including fetal exposure.
- How does Dzud affect children (specifically height and infectious diseases) in Mongolia?
- How does heat affect waterborne diseases via pathogens in water supplies
- How does behaviour modify climate-health associations?
- Climate and infectious disease pathways in young children
- ENSO effects on cholera (much debate around hypothesized pathways)
- How does birthday affect health outcomes?
- How are humidity and temperature related in different contexts?

Practical interventions and health programming

- Predictability of extremes for early warnings on interannual, seasonal, sub-seasonal and individual extreme event timescales
- Associations between ENSO and health
- Impacts of climate on delivery and success of interventions
- Thresholds for health impacts
- Vulnerability to climate of people with pre-existing conditions (e.g. HIV & drought; kidney disease & heat)
- Understanding people's perception of climate risk
- Risk communication strategies for heat risk in already-hot regions
- Sub-urban scale climate surveillance to support operational decision-making for malaria
- Impact-based forecasting of infectious diseases
- Seasonal heat preparedness
- What heat interventions are appropriate in the African context, for different vulnerable groups?
- What barriers exist that prevent people seeking treatment or help during heat waves?
- Sub-urban scale heat risk mapping

Thresholds

- Thresholds for serious and less severe health impacts (morbidity, wellness, occupational health)
- Thresholds for impacts on upstream drivers of health (e.g. employment prospects and education)
- Thresholds for different exposures including both climate and downstream environmental exposures (e.g. heat and flooding)
-

Climate change

- Detection and attribution to climate change of the health impacts of extreme events, for loss & damage and as evidence for mitigation policies and action
- Emerging health risks, such as new infectious diseases
- Changing disease patterns e.g. interplay of dengue and malaria
- Effect of shifting climate baselines on the health impacts of climate variability (weather extremes or ENSO)
- Expansion/contraction of vector-borne disease regions in peripheral zones
- Prediction:
  - When will critical health thresholds be crossed?
  - What is the feasibility of eradicating malaria by 2030/2050?
  - What are the emerging health risks on a 10-30 year lead-time?
- How is heat risk changing in terms of seasonality, frequency, severity and variability of heat extremes

Adaptation

- What are the limits of adaptability? Could compound events serve as analogues?
- Evaluations of adaptation measures
- How to prepare for new, emerging infectious diseases
- Research to support contingency planning for health programming under climate change
- Identifying adaptation options
- Robust planning and preparedness for climate change, such as by running pilots to stress-test current programs to (for example) investigate the sustainability of more frequent early warnings being triggered
- Can we learn from implementation science to close the gap between evidence and implementation?
- How can urban planning measures reduce UHI?
- Which critical infrastructures that are important for health are vulnerable to heat risk?
- What adaptation options are available for heat in LMICs?
- What legal frameworks are needed to protect people during heat waves
- Costs of extreme events on healthcare systems
- Costs of adaptation options

Mitigation

- What are the health impacts of climate change mitigation policies?
- How do different lifestyles (particularly in LMICs) contribute to greenhouse gas emissions?
- How do health risk behaviours cluster with greenhouse gas emissions?

Methodologies

- How can climate adaptation interventions be evaluated?
- How can success of heat interventions be evaluated?
- How can climate be accounted for routinely in evaluations of health programs?
- What methods are available to account for the dynamic nature of climate-health associations over time, including the role of adaptation as a confounder?
- How can behavioural responses to climate events be accounted for in research on climate-health associations?
- Methods to deal with poor/incomplete data series
- New statistical methods to combine different datasets
- How can monitoring, evaluation and learning be improved, in particular peer-to-peer learning?
- Can life-course epidemiology incorporate environmental conditions as a new determinant of long-term health?
- What analytical methods can be used to explore how climate change interacts with other planetary boundaries that drive health outcomes?
- What quality control is needed for research on complex topics like migration?

Priority variables and contexts for research

Exposures:

- Heat (mentioned most)
- Flooding
- Saltwater intrusion
- Drought
- Compound events: heat+wildfires/air pollution/disasters

Health outcomes:

- Nutrition
- Water-borne diseases
- Vector-borne diseases – dengue, malaria deaths and cases
- Endocrine system (via chemical exposure)
- Asthma
- Mental health
- Obesity
- Child development (height)

Mediating factors:

- Health systems
- Household income
- Behavioural responses to climate
- Risk perception
- Migration
- Infrastructure

Populations:

- Children
- Pregnant women
- Which population groups are vulnerable (e.g. people vulnerable to heat including migrants, refugees, elderly infants, pregnant women, outdoor workers etc.)

Geographical settings:

- LMIC
- Urban: intra-urban and urban vs peri-urban vs rural
- Conflict settings
- Heat in different climatic zones in Africa

Appendix B References

Berrang-Ford, L. et al. (2021). Systematic mapping of global research on climate and health: a machine learning review. *The Lancet Planetary Health*. Doi: 10.1016/S2542-5196(21)00179-0.

Scheelbeek, P. F. D. et al. (2021). The effects on public health of climate change adaptation responses: a systematic review of evidence from low- and middle-income countries. *Environmental Research Letters*, **16**, p. 073001. Doi: 10.1088/1748-9326/AC092C.

Nissan H. and Conway, D. (2018). From advocacy to action: projecting the health impacts of climate change. PLoS Medicine, **15**, e1002624-e1002624.

Nissan, H., Ukawuba, I. and Thomson, M. (2021). Climate-proofing a malaria eradication strategy. Malaria Journal, **20**, no. 1, pp. 1–16. Doi: 10.1186/s12936-021-03718-x.

Appendix C
Behavioural Science and the Climate Emergency

Stephen Reicher (University of St Andrews)

One of the core lessons from the response to the Covid-19 pandemic has been the need to integrate physical and social science – particularly behavioural science – perspectives. It is crucial to understand how the virus transmits and what mitigates the process of transmission using insights from virology, epidemiology, modelling, engineering and other disciplines. But this is of little use unless we understand how to change people's behaviours accordingly in order to reduce virus spread – whether that be a matter of spatially distancing, wearing masks, self-isolating or getting vaccinated.

This lesson is of particular importance in responding to the climate emergency. Of course, we need to understand the antecedents of warming, the consequences of warming and the actions necessary to contain warming. But all this is of little use unless it can be used to generate effective action at multiple distinct levels: individual consumption; institutional practices; government policies; coordinated global action.

This gives rise to six sets of behavioural questions:

First, how can individuals be influenced to change their patterns of consumption, recognising that this is a matter of information and messaging, of opportunity (i.e. having the necessary resources) as well as of motivation? In terms of motivation, what is the role individual level approaches (use of incentives) and group level approaches (developing social norms) and how do the two inter-relate?

Second, what is the role of leadership – at community, institutional and governmental levels – in promoting climate change behaviours? How is trust, as a basis of influence, generated and how is it undermined?

Third, what is the basis of climate activism – of people acting collectively in order to influence institutional and governmental climate policies/practices? What determines involvement and what sustains involvement in such movements?

Fourth, how can an internationalist perspective be promoted in which there is concern for the impact of climate change in other countries and support for action to contain it? Conversely, how can 'climate nationalism' be avoided?

Fifth, what is the basis of opposition to climate change action, both at an individual and a collective level? What factors give traction to disinformation on the climate emergency and what is the most effective way of countering such disinformation?

Sixth, what is the role of social inequalities both in terms of the impact of the climate emergency and the response to it? Are different groups more or less concerned and likely to act at an individual level? Are members of these different groups more or less likely to participate in climate emergency activities – and, if so, why?

This list is not meant to be exhaustive. Moreover, each individual question and set of questions is open to further elaboration. Nonetheless, a successful outcome of the climate emergency depends upon generating positive attitudes and active support from the public and hence paying as much attention to the dynamics of the public response as to the dynamics of climate change itself.


wellcome