

African Population Cohorts Consortium



Image: Thoko Chikondi, Wellcome Collection

Scoping Report



March 2021

African Population Cohorts Consortium. Scoping Report. London: Wellcome Trust; 2021

Foreword

The world is at an inflection point. The global shock caused by the coronavirus pandemic has clearly demonstrated the need for political and scientific co-operation which stretches far beyond national borders. Robust and timely data on biology, health, behaviour, socio-economics and the environment are needed to predict and combat such disasters in the future. Such data could herald a scientific revolution in Africa, driving novel causal insights with global relevance and informing African-specific interventions to improve health and social outcomes.

The African continent is experiencing an increasing burden of non-communicable diseases (NCDs) including hypertension, diabetes, cardiovascular, respiratory diseases and mental health problems. This epidemiological transition has emerged alongside existing high levels of chronic infectious diseases such as HIV/AIDS and tuberculosis and coexists with malnutrition. This emerging triple burden is unique, and the increasing numbers of people affected by multiple long-term conditions (MLTC) extremely challenging.

Changes in lifestyle, environmental exposures and poverty are to some extent driving these new patterns. We know little about this multimorbidity, how disorders cluster, the underpinning biological and social mechanisms, the long-term consequences and the health systems responses required to address them. Evidence gathered in high income countries cannot inform the specific disease burden or multimorbidity experienced in low-and middle-income country settings and interventions developed elsewhere will not necessarily translate to African contexts.

The lack of such data from Africa is a critical barrier to wider development on the Continent. There are almost no large-scale population-based studies in Africa, and African populations contribute only 3% of the world genetic data^{1a}, even though the greatest genetic diversity exists in African populations.

We need large and inter-linked longitudinal studies to understand the biological, socioeconomic and environmental causes of the multiple disease burden experienced in Africa. The scientific value of such large-scale data would be enormous and unanticipated, providing new knowledge of both local and global value. These data could drive advances in personalised medicine, be platforms to test interventions to improve health and social outcomes and be used by governments to measure progress towards the Sustainable Development Goals (SDGs). Alignment to tackle the [top scientific priorities as they are being stipulated through the African Science, Technology and Innovation Priorities \(ASP\) programme](#), implemented through the African Academy of Sciences (AAS) and the African Union Development Agency (AUDA-NEPAD) will be sought.

This report presents the result of scoping work conducted by a group of African scientists and international research funders to explore how such an ambitious undertaking could be achieved under the umbrella of an 'African Population Cohorts Consortium' (APCC). They have worked together over a period of a year including a scoping meeting in Uganda in March 2020 and subsequent remote engagement lead by Dr Nicki Tiffin (Cape Town, South Africa) to develop a vision, guiding principles, structure, and potential research themes. The drive and commitment of those involved has been outstanding and I would like to thank them for their contribution.

The vision laid out in this scoping paper is rightly ambitious. It builds on important advances in research infrastructure in Africa, such as Human Hereditary and Health in Africa (H3Africa) initiative and the network of Health & Demographic Surveillance Sites (HDSS). The task is now to work closely with national and regional policy makers in Africa, African scientists and most importantly affected communities, to refine the vision and scope of APCC. This will be used to develop an ambitious yet feasible plan which funders can use to decide whether, and how, to invest in this most pressing area for scientific and human development.

I look forward to supporting this work as it develops.



Sir Jeremy Farrar
Director, Wellcome Trust

Table of Contents

Foreword	2
List of Contributors	3
Executive Summary	6
1. Background	8
1.1 Rationale	8
1.2 Longitudinal Population Studies in Africa – building on existing infrastructure	9
1.3 Learning from international LPS	11
2. Proposed Aims and Objectives of the APCC	13
2.1 Proposed Aims	13
2.2 Proposed research objectives	13
3. Proposed Structure and activities of APCC	15
3.1 Structure	15
3.2 Proposed research activities	15
4. Potential research themes	17
4.1 Cross-cutting research themes	17
4.2 Supporting new research in high priority areas	18
Demographics	18
Health and disease	19
Non-communicable diseases	21
Infectious diseases	21
Genetics/Genomics	21
Environmental factors	23
Socioeconomic factors	23
5. Principles underpinning the APCC	26
5.1 African-led, with equitable governance structures	26
5.2 Driven by community engagement	27
5.3 Support ethical, equitable and relevant use and sharing of samples and data	28
5.4 Strengthen African capacity and leadership	29
6. The way forward	31

Phase 1: Scoping phase	31
Phase 2: Formative phase	31
Phase 3: Implementation phase	31
7. Conclusion	32
8. References	33
Appendix 1: Further contributors to the Scoping Report	37
Appendix 2: Existing Cohorts, LPS and Networks in Africa	39
Appendix 3: Possible formative and pilot projects	43
Appendix 4: Glossary and Abbreviations	46

Executive Summary

African diversity across multiple domains (human, environmental, socioeconomic, policy and health systems) can provide unparalleled research insights. These could be harnessed to provide novel causal insights with global relevance and to inform African-specific interventions to improve health and social outcomes.

Vision

The African Population Cohorts Consortium, APCC, could drive scientific discovery that enhances our understanding of the interacting biological, genetic, socioeconomic and environmental factors underlying health and wellbeing. This promises to accelerate evidence-based improvement of health and social outcomes on the Continent and to monitor progress towards the SDGs.

Aims and objectives

APCC has two proposed overarching and synergistic aims:

- *To strengthen and enhance research infrastructure for population research in Africa*
- *To harness this robust infrastructure to enable high-quality scientific research in high-priority areas*

Proposed research objective comprises:

1. *Enable world-leading discovery science to answer the most pressing health issues on the Continent*
2. *Provide quality population data for surveillance and monitoring progression towards the SDGs*
3. *Assess the impact of policy interventions to support national and regional priorities*

Principles

The following principles will guide the development and implementation of APCC:

1. *African-led, with equitable governance of the initiative*
2. *Driven by community engagement*
3. *Support ethical, equitable and relevant use and sharing of samples and data*
4. *Strengthen African capacity and leadership*

Structure

APCC is proposed as an African-led, African-governed collaborative platform for large longitudinal population studies

(LPS). It would build on existing research infrastructure to collect, collate and analyse multi-dimensional data and samples from diverse populations, and be a platform for add-on studies.

A 'hub and spoke' model aims to ensure Africa-wide geographical representation and inclusion of underrepresented regions and populations. APCC is proposed to consist of a network of core and affiliated sites representing diverse countries across the Continent.

Core sites with data collection from population samples 'typical' of the country would build on existing research infrastructure including large cohorts, HDSS, biorepositories, and established linkage to routine health, social and environmental data.

Affiliate sites would collect a minimum dataset and may not initially have data linkage or biorepositories. Affiliate sites will be supported to participate through capacity strengthening and can progress to become Core sites.

APCC seeks to support participation from countries that are at very different stages of research capacity and would enable progression to Core site status through clear and transparent criteria.

APCC is proposed to be governed by a **Managing Committee** comprising representatives from the participating Core and Affiliate sites. They would be supported by a **Coordinating Centre** which establish standardised protocols and core data standards, ensure data harmonization/inter-operability and support cross-country analyses. An **Independent Scientific Advisory Board (ISAB)** would provide independent scientific advice to the Managing Committee.

By strengthening relationships with national ministries and agencies, APCC aims to leverage data linkage to routine and existing data resources, in turn supporting national efforts with a reciprocal flow of new data. These ongoing intersectoral collaborations aim to ensure relevance of the research outputs, bridge the transition from research output to translation into policy, and mitigate risks to long-term sustainability through effective buy-in from national policy-makers.

Research Vision

This scoping report has identified a wide range of research domains which APCC could contribute to. These include surveillance, discovery research and intervention platforms with direct impact on policy. Work is now required to focus

this vision on a smaller number of areas which have the potential to deliver significant advances in knowledge as well as health and social impact.

Potential research domains include:

- *The changing epidemiological transition of African populations including the causes and impacts of demographic shifts due to migration, morbidity and mortality.*
- *How the genetic diversity of humans, pathogens and vectors can contribute to population-level and individual health.*
- *The changing burden and determinants of both infectious and non-communicable diseases, including understanding and predicting emerging diseases.*
- *Multi-morbidity including interactions between infectious and non-communicable diseases as well as mental health in different environments and across the life course.*
- *Socioeconomic and environmental drivers of health and wellbeing for example the impact of a changing climate on health and social outcomes including the changing distribution of disease vectors.*
- *How health and wellbeing can impact economies.*

These domains are inter-related, with cross-cutting themes including a life course approach, precision medicine and precision public health, migration and mobility, planetary health, and health systems research including universal health coverage.

The pan-African diversity of APCC would enable comparison of relationships between determinants and outcomes from diverse social, cultural, economic, environmental, geographical and genetic backgrounds over time. This diversity in exposures and outcomes would be a core strength of APCC. Coupled with the ability to track the impact of the rapid pace of change in African populations over time, and the consequence of this change on health and social outcomes.

As key enablers to the success of the Consortium, APCC will engage with policymakers from the outset to ensure research is designed in collaboration with and meets the needs of key stakeholders. Ensuring early buy in through co-development will help to shape the research effectively, and build understanding of the research process and outputs amongst policymakers, supporting APCC to achieve its objectives and bridge the transition from research output to translation into policy.”

Phases

We propose that APCC is further developed in two phases:

The **Formative Phase** would be led by a consortium of African scientists and policy makers. They would refine the scope and scientific objectives of APCC, finalise the structure of APCC including the governance and management structure and establish best practice for ethics and data governance. The primary outcome of this phase would be a White Paper outlining these. Decisions on investment in APCC and whether the initiative will go forward will be taken after this phase.

The **Implementation Phase** would involve an open selection of Core and Affiliate sites and the creation of the Coordinating Centre. Each site would develop country-specific research priorities in partnership with local policy makers and communities. Pilots would be conducted in each site to finalise study design, research protocols and standards for data collection and harmonisation protocols before data collection commences.

Summary

It is time for a step-change in ambition for population-based science in Africa. Recent developments on the Continent including large scale genomics research (such as H3Africa), capacity building programmes (such as the Developing Excellence in Leadership, Training and Science (DELTAS) programme and existing research infrastructure (including a network of HDSS sites, bioinformatics hubs and biorepositories), mean that a more co-ordinated and ambitious vision is within reach.

1. Background

1.1 Rationale

A substantial and reliable evidence base is needed for scientific approaches to improve the lives of African populations and realise the SDGs (Figure 1) across the Continent.

Large-scale detailed, reliable and comparable data are needed to monitor health, socio-economic and environmental metrics, to track changes over time, to identify emerging challenges and to measure ongoing and changing

relationships between key determinants and outcomes for health and wellbeing throughout the life course and over time^{1b}. Achieving SDG-3, *Good health and wellbeing*, depends on achieving other environmental and socio-economic SDGs; and good health in turn contributes to achieving those SDGs. This requires an integrated approach to identify and harness the synergies and relationships between health, social, economic and environmental determinants².

Figure 1: The Sustainable Development Goals (from <https://www.un.org/sustainabledevelopment/>)



Successful African-led initiatives are growing, with expanding skills and infrastructural capacity to generate, manage and analyse large and complex data, and to store collected samples³. Technological innovation and a growing research landscape are bringing precision public health within reach for Africans, addressing historical underrepresentation of African populations in global research and identifying how clinical, environmental, socio-economic and genetic drivers of health work in concert to affect health and wellbeing⁴⁻⁶. Increasing depth and granularity of a wide range of data types can inform the African-led design of interventions tailored to African individual and population needs, providing a reliable evidence-base to ensure maximum efficacy of health care through the design of appropriate individual- and population-level preventative, diagnostic, prognostic and therapeutic tools and interventions. The [AAS](#) provides a unifying framework for pan-African research coordination and oversight, and Africans are well-positioned to find African solutions to African problems whilst simultaneously making significant contributions to global scientific discovery. Africa Centre for Disease Control, [Africa CDC](#), addresses disease threats and outbreaks and supports related public health programmes on the Continent. Partnerships between funders, researchers and policy makers combined with growing experience in appropriate and effective community engagement, research ethics and data governance underpin ethical, participant-centric data use, as well as effective and ethical governance of onward data use in support of Open Science initiatives.

1.2 Longitudinal Population Studies in Africa – building on existing infrastructure

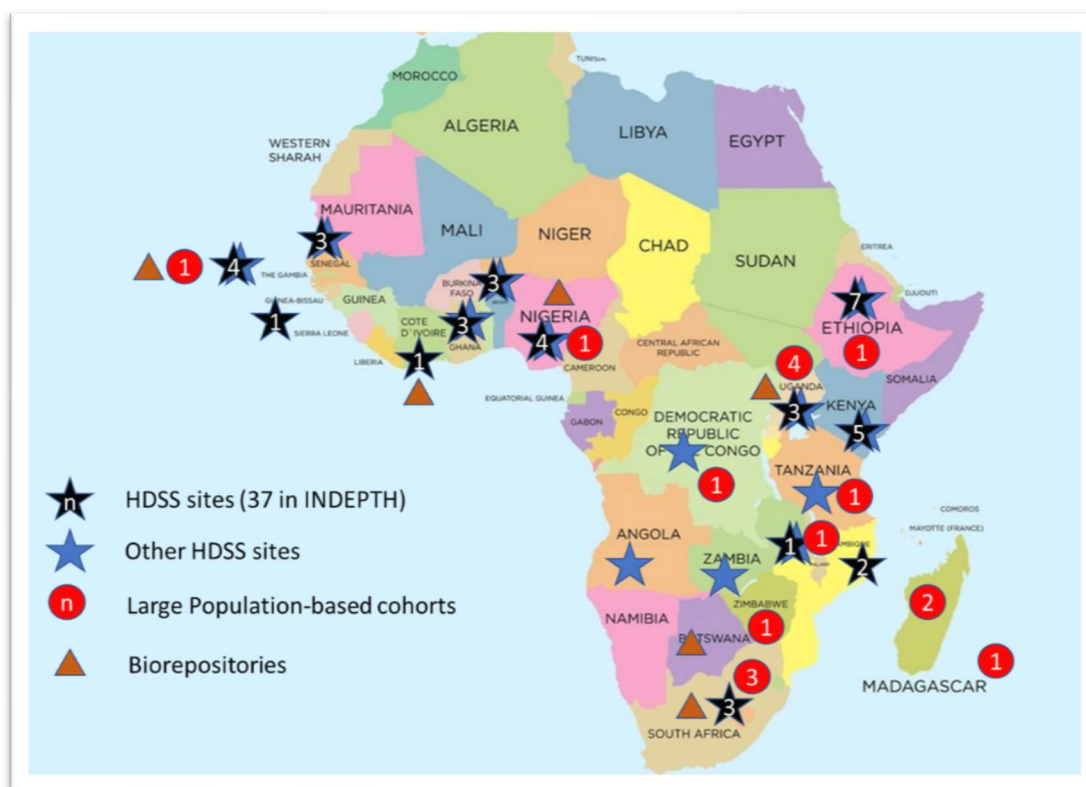
LPS include cohorts, panel studies and biobanks. They offer high-quality multi-dimensional data from across the life-course, providing an evidence base for interventions and health care provision. With appropriate data linkage, interoperability and multi-purposing of data, resource use can be maximised to ensure both research and service delivery impact, enabling improved outcomes for cohort participants as well as the wider population. With appropriate consents and data governance in place, cohort-related resources can

be well-positioned to be discoverable, accessible and reusable to ensure maximal learning and benefit from collected data.

Longitudinal cohorts provide an opportunity for relationship-building with participants over time, leading to meaningful and engaged informed consent processes, community involvement and participation, and stakeholder involvement and engagement. This long-term investment into building of capacity, engagement and resources over time can support improved translation of findings into direct positive impacts on health and social outcomes.

APCC seeks to leverage established LPS resources across the Continent (Figure 2) by expanding existing sites as well as linking LPS together in the Consortium. Existing resources can provide a baseline from which APCC can develop a framework for the detailed clinical, biological, socioeconomic and environmental data collection envisaged for the Consortium.

Figure 2: Current population-based resources in Africa



Within Africa, the INDEPTH network harmonises data from many HDSS, across Africa, while the Department of Science Innovation/South African Medical Research Council South

African Population Research Infrastructure Network (SAPRIN) links sites across South Africa (Box 1).

Box 1: Examples of HDSS networks in Africa

The INDEPTH Network consists of HDSSs that investigate disease- and pathogen-specific morbidity, cause-specific and total mortality. The network primarily undertakes population-level surveillance of key indicators, but sub-cohorts may also be followed up by phone or home visits to collect clinical data or laboratory specimens. Innovations such as unique electronic or biometric indicators facilitate effective data linkage from routine data ecosystems such as health systems, population registers and community resources. The use of the Comprehensive Health and Epidemiological Surveillance System^{7a} (CHES) provides a framework for key metrics. Collaboration with the extensive INDEPTH network can provide synergies with established data collection programmes ongoing in many parts of Africa.

The South African Population Research Infrastructure Network (SAPRIN) is funded by the Government of South Africa. SAPRIN integrates standardised population, health and socio-economic data for geographically defined rural and urban populations from three HDSS nodes, with a fourth node preparing for its baseline census in urban Gauteng, and with several more nodes under development. There is a common protocol for household visits and linkage of public sector records to enable the analysis of indicators for births and deaths, residence and migration, socioeconomic indicators (labour status, education, social protection), and measures of health and wellbeing.

The current population is 355 000 consented individuals in rural and peri-urban areas. Clinic attendance records are linked to the population data and analysed as evidence for policymaking. The population platform triangulates with national census data in the overlapping census enumerator areas, to enable comparison and validation of findings through which the national data are better understood. Commitments are in place to produce evidence for policy makers.

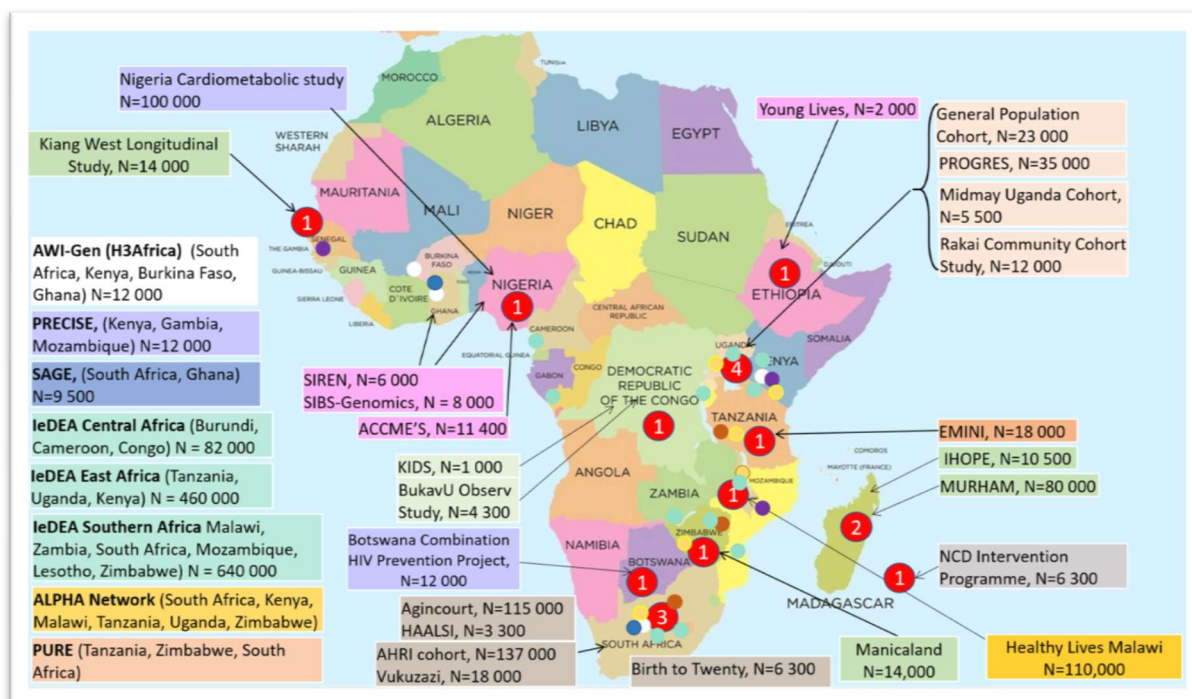
Existing infrastructure and expertise from other large disease-specific studies can be leveraged by APCC, for example: the International Epidemiology Databases to Evaluate AIDS

(IeDEA) network has multiple nodes in Central, Southern, East and West Africa and collates longitudinal HIV/AIDS data for large cohorts of People Living with HIV (PLHIV) on the

Continent; the Botswana Harvard AIDS Institute Partnership (BHP) has biorepository capacity for 1.5 million specimens; the Mali International Centres for Excellence in Research (ICER) has extensive infectious disease research infrastructure; and the [H3Africa](#) disease-specific case-control studies which explore genetic risk factors for a variety of infectious and non-communicable diseases has three biorepositories in South Africa, Nigeria and Uganda,

genomics expertise^{7b}, and bioinformatics and computational training and infrastructure across the Continent ³ (Figure 2). The [DELTA Africa Initiative](#) comprising 11 programmes has been developing skills development, scientific research and leadership capacity on the Continent, providing a well-established framework for similar activities by APCC. Both H3Africa and DELTA Africa are managed by the AAS.

Figure 3: Large Population Studies in Africa



1.3 Learning from international LPS

APCC could engage with existing international large-scale LPS as well as other international Consortia, leveraging existing skills, tools and experience whilst tailoring and repurposing these to African contexts as appropriate. In turn, APCC outputs could contribute to expanding global perspectives, improving representation of African populations in global efforts thereby increasing global equity in international initiatives, and interface with the International HundredK+ Cohorts Consortium (IHCC) (see Box 2) in order to synergise with this global collaboration of cohort studies.

Box 2: International 100K+ Cohorts Consortium

The [IHCC](#) is a global network of large cohorts with multi-dimensional data from diverse populations collaborating to enhance scientific understanding of the biological, environmental, and genetic basis of disease and to improve population health, facilitating meta-analyses to answer global health questions that no single one can answer alone, thus enhancing the value of each and leveraging the investments already made in them. IHCC member cohorts are unselected for disease and aim to recruit 100,000 participants or more with available biospecimens (or the possibility of collecting them) with longitudinal follow-up. Cohorts with less than 100,000 participants are accepted as members if they include low- and middle-income country/countries, and/or cohorts of disadvantaged populations in high income countries, and/or collect data from exceptional, unique, or hard to accrue groups. The IHCC currently consists of 103 participating cohorts located in 43 countries, with total current sample sizes across all cohorts of > 50 million participants. The IHCC Cohort Atlas, launched in 2020, is a platform which enables researchers to search across IHCC cohort metadata (<https://atlas.ihccglobal.org/>). Underpinning the atlas is a harmonized metadata model that provides mappings into IHCC cohort data dictionaries. The Atlas is the first step towards creating a federated interoperable network for the discovery, access, and cross cohort analysis.

Internationally there is growing recognition of the enormous potential in linking genetic and genomic data to routine health data and other exposures in population cohorts, so that the relationship between genetic factors in concert with other exposures can be assessed in relationship to a variety of

outcomes experienced by an individual over their lifetime. Examples include [DeCode](#) (Iceland), [Biobank Japan](#) (Japan), [100 000 Genomes Project](#) (UK), [UK Biobank](#) (UK) and [All of Us](#) (USA) (see Box 3) and those described by Stark et al (2019)⁸.

Box 3: Health-related longitudinal cohorts linking genetic data to digital health records⁹

[deCODE Genetics](#) is a genotyped population cohort comprising of over half of the adult population in Iceland. Electronic medical records have been linked to genetic data as well as a comprehensive genealogy database.

[UK Biobank](#) is a large, detailed prospective study with over 500 000 participants in the UK. It collects extensive socio-economic, environmental and lifestyle data from participant interviews, questionnaires and wearable tech¹⁰. Follow up for health-related outcomes occurs mainly through linkage to existing routine health data.

[All of Us Research Program](#), run by the National Institutes of Health in the USA, recruits participants who are willing to share electronic health records data linked to genomic, sociodemographic, lifestyle and other collected metrics.

[The BioBank Japan Project](#)¹¹ is a large, patient-based biobank of more than 200 000 participants, that links genomic data from bio-banked samples to electronic health records, with a focus on 47 target diseases.

2. Proposed Aims and Objectives of the APCC

There is remarkable diversity across Africa in terms of demography, disease burdens, health systems, environments, geographies, economies, cultures and the extensive genomic diversity of African populations.

Within individual LPS, it can be difficult to disentangle the impacts of such diversity. APCC would be uniquely positioned to harness the advantages of this extensive African diversity. Through core data harmonisation and standardisation of rich and varied data capture across highly diverse contributing sites, it would be possible to tease out the true impacts of these diverse determinants on priority outcomes. APCC could provide a rich and unique source of data for exploring and better understanding the biological, social and environmental determinants of wellbeing, health and social outcomes, and facilitating large-scale, high-powered studies to conduct meaningful meta-analyses using harmonised datasets across all stages of the life course. Using these data, APCC could evaluate changing dynamics of determinants and outcomes, and their relationships to each other.

2.1 Proposed Aims

APCC is proposed to have two overarching and synergistic aims:

1. *To strengthen and enhance research infrastructure for population research in Africa.*

The first aim is to build a robust and sustainable pan-African core research infrastructure for LPS, ensuring an enduring platform for long term scientific achievements driving improved population health, wellbeing and social outcomes. This platform would leverage existing infrastructure and skills on the Continent, ensuring that knowledge- and skills-sharing from existing programmes have pan-African reach and impact through APCC collective.

2. *To harness this infrastructure to enable high-quality scientific research in high-priority areas*

The second aim is to use APCC research platform to enable high quality, focused research programmes addressing high priority research questions informed by national and regional needs, to generate transformational new knowledge to accelerate improvements in population health and wellbeing and to facilitate the achievement of the SDGs on the Continent. The large scale of aggregated data would ensure

valid, significant and representative research outputs that are relevant to local, national and regional stakeholders, providing insights that can lead to meaningful change and improvements for the general population.

2.2 Proposed research objectives

Three research objectives are proposed spanning discovery research, surveillance and policy evaluations. Whether it is desirable and feasible to include all three objectives should be explored in the formative phase.

Enable world-leading discovery science to answer the most pressing health issues on the Continent

The proposed design of APCC allows for open-ended research goals that can evolve over the lifetime of the programme. APCC could inform the design of evidence-based interventions in the health, social services, education, labour and environmental sectors. It could also facilitate precision public health approaches for African populations for better preventative, diagnostic, prognostic and therapeutic approaches to health and wellbeing.

Provide quality population data for surveillance and monitoring progression towards the SDGs

APCC could provide a unique opportunity to provide direct measurements of burden of disease, accurately measuring the true prevalence (proportion of patients with that condition) and incidence (rate of new cases occurring) of any health condition in the LPS population. Demographic trends including birth and mortality rates would provide accurate data for Offices of National Statistics. As new health challenges emerge, APCC sites would be positioned to rapidly generate pan-African data that can be aggregated and compared to support national efforts and assist surveillance and tracking of emerging challenges.

Assess the impact of policy interventions to support national and regional priorities

This could include monitoring the impact of national, regional or local or policy interventions, engaging with policy makers where appropriate. Through pro-active engagement with

service providers and policy-makers, APCC could support and enhance existing service provision; through ongoing relationship building and collaboration at the research-policy nexus, APCC researchers could learn to effectively contribute to the policy-making progress, whilst policy-makers could simultaneously learn from APCC research enterprise.

3. Proposed Structure and activities of APCC

3.1 Structure

APCC aims to establish a large network of collaborating LPS from across the Continent, leveraging existing infrastructure in Africa such as existing LPS, biobanks and data infrastructure.

APCC could have the following structure, with clear and transparent criteria for sites to be defined during the formative phase of the programme:

Core sites: will be mature LPS that include populations 'typical' of a country or region, collecting granular and multi-dimensional data including core metrics defined during the formative phase, collecting biological samples and with capacity for, or aiming to develop, data linkage to routine clinical, social and environmental data resources.

Affiliate sites: will have existing data collection, for example an existing HDSS site, or less well-phenotyped LPS which may collect some but not all the core data, may lack biological sample collection or the ability to link to routine data. Affiliate sites may work to become Core sites, for example by building longitudinal capacity or increasing the type of data or samples collected.

A Managing Committee supported by a **Coordinating Centre** is proposed to support a governance structure that could include members from core and affiliate sites, providing oversight and logistical support, and facilitating the sharing of best practices, data harmonisation, inter-operability and Consortium-wide meta-analyses.

An **ISAB** would provide independent scientific advice to the Managing Committee.

The final APCC structure, inclusion criteria, management and governance structure will be discussed and finalised during the Formative phase through wide stakeholder engagement.

3.2 Proposed research activities

Leveraging existing resources and infrastructure

APCC proposes to leverage existing infrastructure, experience and knowledge already present in Africa to establish a collaborative network of LPS. This would provide a backbone of collaborating programmes which work together to establish standardised protocols, sites inclusion criteria, data standards to generate harmonised metrics and harmonised research frameworks. From this strong baseline,

additional nascent and ongoing LPS that do not yet fulfil core membership criteria could be brought into the Consortium with support to develop infrastructure and skills capacity to achieve these standards. This could be done for example by increasing the breadth and depth of data collected, extending geographical representation, and implementing agreed data standards.

Enhancing the Consortium to ensure diversity of representation

The Consortium will seek to identify underrepresented regions and populations and could support affiliated population cohort programmes to ensure diverse and equitable representation of African populations. Africa-wide participation will ensure a wide range of diversity both within and between African populations, thus increasing external validity and generalisability of findings. The long-term nature of APCC should facilitate relationship-building with participants with meaningful and engaged informed consent processes, community participation and stakeholder engagement, thus enhancing scientific citizenship whilst simultaneously improving the quality and impact of resources generated. This long-term investment and building of capacity, engagement and resources will provide an environment that is well-positioned for improving translation of findings into direct positive impacts on health and wellbeing.

Standardising and harmonising data collection

APCC proposes to define, agree and pilot a study design that will ensure that core data from the participating cohort studies will be comparable across the Consortium and suitable for combined analyses or meta-analyses. Whilst enabling comparison across sites by ensuring standardisation of common data elements, this approach will not preclude individual sites to pursue additional independent objectives, recognising that not all metrics collected can be replicated at all sites. This flexibility will also accommodate developing subsequent site-specific nested and subset analyses that will be developed over time, providing a strong standardised baseline that will enhance and enable - rather than stifle - specialised onward analyses.

APCC would seek define a core set of LPS meta-data that describe each of the participating LPS in a standardised way, and store essential documents and information to ensure

transparency; for example including descriptors such as location, type of location (urban/rural), number of participants, gender representation (% female), age distribution in standardised categories, age range of total cohort, year of initiation, funders, ethics review boards with oversight, ethics approval references, protocols, participant information documents and informed consent templates.

The LPS will provide population references - core sets of metrics that describe overall health and wellbeing, providing reliable metrics for comparison to data generated elsewhere, contributing to enhancing and validating existing national metrics. Returning enriched data and analyses to service providers and data generators such as national ministries and departments will also be explored. Optimising appropriate data linkage, interoperability and multi-purposing could ensure both research and service delivery impact, ensuring improved outcomes for cohort participants as well as the wider population. In addition, a set of core indicators common and accessible in all cohorts could be explicitly defined, to maintain a dashboard of a selected subset of key indicators relating to SDGs at each participating site.

Data linkage using existing routine data resources

Linkage to existing data resources would be pursued where possible and aim to enhance data completeness and granularity. This includes national and regional administrative and civil registries such as records for social services, birth and death registries, health-related data and disease specific registries (including for notifiable diseases). Whilst full electronic medical records may not yet be established in partnering sites, administrative data from health services can still be effectively harnessed and linked, and APCC could engage with existing data collation and linkage efforts seeking to strengthen both core health service delivery by national agencies as well as the LPS research enterprise.

Unified digitalisation of health care data is becoming more common, with many examples of digital implementation across Africa, e.g. [SmartCare in Zambia](#), [OpenMRS in Rwanda](#), or the Provincial Health Data Centre (PHDC) health information exchange in the Western Cape, South Africa⁸³. Other digitalised records include national population registries (births and deaths), social services data, education service data, municipal household mapping data, and climate data. The formative and implementation phases of APCC could explore potential sources of administrative data and pilot data linkage approaches in collaboration with the departments that manage those data, building collaborative relationships with those stakeholders. This could include piloting the linkage of retrospective data in addition to prospective linkage activities, greatly increasing the time span for which data are available for cohort participants. APCC could strive to facilitate linked data from birth onwards for younger sectors of the population.

In addition, APCC members could access the experience, infrastructure schema and open software from existing initiatives including the PHDC¹² and SAPRIN (Box 1) in South Africa.

4. Potential research themes

4.1 Cross-cutting research themes

Intersectoral themes speak to common drivers of health, wellbeing and social outcomes. The SDGs (Figure 1) reflect fundamental issues impacting health and wellbeing that are relevant to all populations globally. The scoping phase has identified the following research themes and frameworks that could be explored or guide the Consortium research efforts.

A Life Course Approach to health, wellbeing and social outcomes recognises temporal and social perspectives across the life of individuals and generations, studying the changing physical and social determinants of health and wellbeing across all ages and stages of life including pregnancy, perinatal and maternal health, infancy, adolescence, adulthood including the ageing population¹³. A life course approach will identify opportunities for early interventions at appropriate stages of life in order to improve health and wellbeing of African populations. Through this lens, APCC could bring novel insights that reflect, for example, the generally younger population structure in Africa as well as the significant infectious disease challenges faced by neonatal and paediatric sectors of the population. The life course lens and longitudinal approach will provide insights into the interplay between infectious diseases and non-communicable diseases in individuals as they age and will help to understand the drivers of these transitions.

Precision Medicine and Precision Public Health harness granular data about environmental, clinical, socioeconomic and genomic data to better tailor health interventions to individuals and also to whole populations¹⁴. APCC would follow precision medicine principles by ensuring that detailed genetic, health, socioeconomic and environmental data are collected for every individual, and that with appropriate statistical approaches these data can be harnessed to tailor health interventions to best fit individuals and populations represented in APCC cohorts. Population-level pharmacogenomic data, for example, can inform prescribing practices across the Continent, while understanding the impact of unique gene-environment interactions experienced by participants in African can help to elucidate disease aetiology. This research could be supported through an

'omics' platform, including genomics, that maximises expertise and infrastructure existing in the Continent.

Migration and Mobility of population groups and workforces as well as rural-urban migration all contribute to complexity of analysis and diversity in African populations. APCC could provide invaluable insights into real, ongoing population dynamics and describe determinants and outcomes of migration, including socioeconomic factors, health and wellbeing. These data could also describe the impact of the environment and climate change on migration and could provide insights about vulnerable migratory populations that include undocumented individuals and refugees. Migration in Africa is driven by many different factors including the need to find work, urbanisation, political unrest or war, and food insecurity driven by climate change and extreme weather. Collating these data at a large scale will provide new and invaluable insights into their impacts on populations and how negative impacts might be mitigated.

One Health articulates the interconnectedness of many different sectors with the health and wellbeing of human populations.¹⁵ APCC would collect context-specific data from both urban and rural sites that describe the health of the environment including climate change, agricultural practices, livestock, living conditions, sanitation and antimicrobial resistance (AMR)¹⁶⁻¹⁹. This approach will underpin the analysis of geographical, environmental and socio-economic determinants of health and wellbeing. Using a One Health lens could, for example, equip APCC to rapidly understand the interrelated drivers of emerging zoonotic pathogens (for example the Ebola virus) and treatment efficacy for pathogens, including the spread of AMR, at a pan-African scale.

Universal Health Coverage is the provision of high quality health services, including prevention, promotion, treatment, rehabilitation and palliation, to all people regardless of their socioeconomic situation or resources^{20,21}. According to the World Health Organisation ([WHO](https://www.who.int/)), universal health coverage embodies equitable access to health care of sufficient quality without incurring financial risk. APCC could collect health systems and healthcare access data, and identify gaps in health care access together with health and wellbeing outcomes for vulnerable groups (for example, migrants, adolescents, the elderly), as well as assessing the impacts of

policy-level interventions for universal health coverage²²⁻²⁴. Such research across the diverse health systems on the Continent could, for example, compare and contrast the impact of different modes of health care provision on key health and well-being outcomes, facilitating shared learning of most effective healthcare practices from across the Continent, informing best practices based on an extremely wide evidence base, and seeding new ideas for effective healthcare provision.

Multimorbidity/Multiple long-term conditions (MLT). The burden of diseases across the globe is changing as populations age and countries develop. On the African continent, an increasing burden of non-communicable diseases, such as hypertension, diabetes, cardiovascular, respiratory diseases and mental health problems is developing alongside existing high levels of chronic infectious diseases, new emerging infectious diseases and malnutrition. We know very little about the causes and long-term consequences of this unique disease burden. Changes in lifestyle, environmental exposures and poverty are to some extent driving these new patterns but we know little about this multimorbidity, how disorders cluster, the underpinning mechanisms, and the health systems responses required to address them. The multidimensional data captured in LPS that APCC seeks to link would be well positioned to bring the evidence needed to address the multimorbidity patterns that are relevant in the Continent.

4.2 Supporting new research in high priority areas

APCC could provide an opportunity to address a large variety of scientific questions that can range from investigations of immediate policy relevance relating to Sustainable Development Goals (SDGs, Figure 1) to advancing discovery science by answering novel scientific questions. APCC would enable collaborations that support basic science research, with stored samples and data used for discovery research into aetiological and biological mechanisms and processes for non-communicable diseases and microbiomes, or infectious disease research involving human hosts, vectors and pathogens. The ASP programme Alignment, implemented through the AAS and the AUDA-NEPAD, has set out to identify the top scientific priorities across the continent and could inform the scientific direction of APCC. High priority research questions can be explored and finalised in the formative phase of the programme. These research questions will harness the large scale of data collected as well as the diversity of exposures and outcomes for statistically powered comparisons and meta-analyses across APCC.

Potential research domains include:

- *The changing epidemiological transition of African populations including the causes and impacts of demographic shifts due to migration, morbidity and mortality.*
- *How the genetic diversity of humans, pathogens and vectors can contribute to population-level and individual health.*
- *The changing burden and determinants of both infectious and non-communicable diseases, including understanding and predicting emerging diseases.*
- *Multi-morbidity including interactions between infectious and non-communicable diseases as well as mental health in different environments and across the life course.*
- *Socioeconomic and environmental drivers of health and wellbeing for example the impact of a changing climate on health and social outcomes including the changing distribution of disease vectors.*
- *How health and wellbeing can impact economies.*

Demographics

Demographic measurements undertaken across APCC could provide immediate data about current population structure, family structures, mobility, migration and demographic dividend. In the short and intermediate term, changes in these metrics will describe fluctuations and dynamic demographics which may be seasonal or influenced by urbanisation, environmental, socioeconomic or policy-level changes. The use of harmonised metrics will allow for agile demography with rapid detection of changes in the population. In the longer term, changing demographic metrics will reveal fundamental shifts in population indicators, including births, life expectancy, and mortality. Ongoing research will seek to identify drivers of these changes, such as policy or economic changes, environmental changes including overcrowding, urbanisation, impact of agricultural activities, pollution or natural resources depletion, and extreme weather events such as extreme heat, and climate change.

APCC could explore appropriate and accurate identification methods for participants, including possible biometric identifiers, capitalising on individual sites experiences to ensure acceptability to the participants and communities where they will be used. Core metrics will be agreed for defining gender, socioeconomic status, lifestyle, health care access and wellbeing among others. Consideration of civil status will be similarly inclusive beyond traditional concepts of marriage, divorce and partnerships - whilst retaining sensitivity to social and cultural norms and participant safety.

Box 4: Priority research questions – Demographics

- *What is the structure of the population (age categories, gender, civil status, education, occupation) and how does this change over time?*
- *Is the demographic dividend reaching its potential, and are the expanding youth and young adult population receiving adequate service provision?*
- *What family structures predominate, and what are the causes and impacts of changing demographics? How can vulnerable households be supported, and do targeted health and social policies reach these vulnerable households?*
- *What are the migratory and mobility patterns of the population, including urbanisation, and what are the causes and impacts of migration and changes in migration?*
- *What rates of fertility, births and birth outcomes exist in the population, and how do these relate to changes in other determinants of health and wellbeing?*
- *What is the true mortality rate, what are causes of mortality, and how do they relate to disease burden and mortality? What are the determinants and impacts of changes in mortality?*

Changes in demographic data measuring attrition from, and addition to, the population over time will provide information about how migration impacts key exposures and outcomes and the wellbeing of migrants and their households. This will facilitate comparisons between permanent migration and temporary work-related migration where an individual remains a household member.

APCC could establish and deploy appropriate methods for collecting reliable, detailed data for cause of death during the formative phase of the programme. For example, verbal autopsy through interviews with close relatives or caregivers of deceased individuals could be piloted to provide granular cause of death data where medical certification of death is not feasible or civil registration and national collection of vital statistics is not sufficient (see examples of work done in South Africa²⁵, Uganda²⁶, Ghana²⁷). Pragmatic and practical approaches building on existing tools, for example the [WHO Verbal Autopsy Standards](#)²⁸ could be piloted and validated²⁹. Similarly, the use of minimally invasive autopsy could be explored (see example of validation study in Mozambique³⁰). Using medical records to infer cause of death is another approach that could be assessed (see examples of retrospective work done in Ethiopia³¹, Nigeria³²⁻³³, Cameroon³⁴) as well as looking at triangulation of methods. Acceptability of verbal autopsy, minimally invasive autopsy and other approaches to ascertain cause of death would

need to be determined through sensitive, carefully conducted and culturally appropriate community engagement, led by local facilitators using appropriate sensitivity and language use to avoid offence or culturally insensitive or taboo probing.

Tracking demographic metrics and how they change will also identify possible bias in individual LPS: selection bias can be determined by identifying which sectors of the total population are agreeing to participate, who is declining, and identifying sectors that are not being reached by recruitment efforts. In addition, demographic metrics could be used to assess attrition bias and which sectors of the population are preferentially lost to follow-up through opting out of participation, changing patterns of health-care or other interactions, or through leaving the cohort catchment area. Demographic metrics could be used to ascertain data integrity and internal validity and reliability of cohort data, for example by using in-depth analysis and curation of a subset of participants, and by checking how representative the cohort is of the wider population's demographics.

Health and disease

The formative phase of APCC would seek to identify priority diseases within Africa across sites based on national and regional disease burden, knowledge about disease determinants, and the potential for intervention/prevention. These include the increasing role of common NCDs such as diabetes, cardiovascular disease (including hypertension), cancer and mental health conditions; and infectious diseases, both chronic such as HIV, TB, malaria, and current and new emerging infections. This would provide accurate baseline measurements of the burden of disease across the Continent and describing the prevalence of diseases in different environments (rural/urban), regions, geographical domains, climates and life stages. From this baseline, trends in key determinants and outcomes can be measured. This holistic approach to tracking changes in population-level burden of disease over time as well as changing health profiles in individuals through their life course will provide new insights into the interplay between infectious and non-communicable diseases in diverse African populations.

By design, APCC could also facilitate a holistic view of individuals, without restricting data collection to specific diseases or exposures, and provide insights into the determinants, burden and nature of multi-morbidity in African populations. The large scale and scope of the Consortium could facilitate studies of less common diseases and neglected tropical diseases, providing large enough case numbers even for rare conditions. In addition, the harmonised metrics and large coverage facilitate agile and rapid monitoring of emerging health conditions and pathogens across the Continent.

Data generated by APCC could track changes in the burden of disease over time, highlighting vulnerabilities, neglected diseases and neglected sectors of the population susceptible

to particular diseases and providing evidence to inform appropriate interventions. Access to health care, health systems and quality of care can be measured, evaluating the match/mismatch between existing and legacy health systems designed around episodic disease and primary health care needs and the current requirements of populations facing chronic health challenges. Determinants and drivers of key diseases will be identified, and changes in burden and impacts of disease over time will be related to changes in key determinants alongside the impact of changing policy and health care provision and access. Existing standards can be

used or adapted, such as the [SDG for Disease](#) standards and the [WHO Quality of Care](#) metrics³⁵.

Biomedical interventions are most often developed and tested in non-African populations. APCC data could facilitate characterising adverse drug reactions in individuals and efficacy of prescribed drugs in participants by capturing failures to respond to treatment and treatment switching. Giving the significant impact on African patients of adverse drug reactions or failure to respond to primary lines of treatment, reliable evidence on these issues will provide great value to policy makers and governments.

Box 5: Priority research questions – Health and disease

For high priority diseases identified during the formative phase:

- *What is the true prevalence and incidence of common diseases across Africa? What are the accurate baseline measurements for priority diseases across the Continent, and how do these metrics change over time and through the life course?*
- *What is the incubation period (for infectious diseases), duration and severity of illness for common diseases? What is the impact of demographic, environmental, socioeconomic and genomic exposures on patient outcomes, how do these determinants change over time and through the life course, and what are the causes and impacts of those changes?*
- *What are common comorbidity profiles, and what are their impacts on patient outcomes? How do comorbidity profiles change over the life course and in response to changes in external determinants of health?*
- *What is the impact of interplay between infectious and non-communicable diseases, and how does this relate to the stage and rate of epidemiological transition?*
- *What current preventative measures are in place (including vaccination programmes, screening programmes and prophylactic treatments), how effective are they, and how is their efficacy impacted by changes in other determinants of health and wellbeing?*
- *How are diseases diagnosed, and are there good prognostic markers for patient outcomes? Are diagnostic tests safe and appropriate for Africa, in terms of costs, efficacy and environmental conditions for storage and use?*
- *How do health systems and access to care as well as quality of care impact burden of disease and health outcomes, and what are the impacts over time of changes in policy and access to health care? To what extent are resources available to realise policy and interventions?*
- *What treatments are in use, and how safe and effective are they for African populations? What factors, including pharmacovariants, impact efficacy of medications?*
- *What are the causes and impacts of behavioural and societal determinants of health (social epidemiology)?*
- *What is the prevalence of mental health conditions? What are their causes and impacts on general health and wellbeing and how do they interface with socioeconomic, environmental and genetic determinants across the life course? How is mental health impacted by changing determinants including life course transitions, access to quality care, and interventions or policy changes?*

For infectious diseases, additional research questions include:

- *What is the long-term impact of COVID-19 on individuals' health and on health systems and access to health care?*
- *What is the Antimicrobial Resistance (AMR) profile for populations, how is it impacted by environmental issues and agriculture, social and economic factors, genomic diversity, health systems and access to health care? How is AMR changing in response to changing determinants of health and wellbeing, environmental and climate change?*
- *What are host, vector and pathogen dynamics that drive disease outcomes? How do these change in response to changes in environment, climate, health systems, policy, social and economic determinants? What are the dynamics of transmission networks, and how do these impact the rates of transmission?*
- *What are host, vector and pathogen genetic factors that impact susceptibility and severity of disease and treatment efficacy? How can these be harnessed in surveillance for existing and emerging pathogen outbreaks?*

Non-communicable diseases

Common chronic NCDs such as cardiovascular disease (including hypertension), mental health conditions, diabetes, cancer and kidney disease can be explored to predict risk over time and to inform the design of early interventions such as screening, treatment initiation and promotion of lifestyle changes. Genetic/genomic contributions to developing NCDs will also be modelled in combination with clinical (multimorbidity), observational, demographic and socioeconomic determinants and drivers of NCDs to identify the contributions of all these factors to the overall disease state. Samples collected from participants would be biobanked and used to research the changing disease profile in individuals over time. Key interactions and the dynamic interplay between NCDs and infectious diseases would be explored.

Infectious diseases

The current COVID-19 pandemic has highlighted the increasing risk of future infectious disease-related disasters that connected and global societies will face. Understanding the long-term impact of COVID-19 in African populations and population subgroups is likely to become a scientific and health priority. APCC could have a valuable contribution thanks to the long-term follow up of participants, particularly if COVID-19 becomes an endemic or seasonal concern similar to cholera, influenza and other recurring pathogens. Similarly, APCC could be well placed to respond rapidly to new emerging infections in the Continent. This is in addition to the existing infectious disease priorities across Africa such as HIV, TB, malaria, hepatitis, sexually transmitted infections and tropical diseases – including neglected tropical diseases. Region-specific sentinel programmes could monitor key infectious diseases and accurately measure incidence, prevalence and determinants of susceptibility, severity, and patient outcomes including quality of life for those infectious diseases, providing an ideal context for piloting, validating and evaluating novel testing strategies and novel diagnostics. Pathogen samples could be biobanked and used to understand evolution of pathogens, transmission networks for specific pathogen strains and antimicrobial resistance. This high-quality data can be used in infectious disease modelling approaches to predict disease burden, spread and the efficacy of proposed interventions. High quality data for temporal dynamics of infectious disease transmission and patient outcomes can provide an important benchmark for model building and validity testing.

Focus areas within the infectious disease research programme include in-depth analysis of burden and determinants of maternal and neonatal morbidity and mortality due to infectious diseases, particularly in the first seven days after birth, and the interplay between co-infections and NCDs could be explored in detail to

understand how multi-morbidities impact patient outcomes and onward transmission of infectious diseases. Analysis of determinants of transmission include geographical and environmental determinants, social factors such as interpersonal or community/population-level behaviours, violence, stigma, alcohol or drug abuse, household structures and social practices, as well as socioeconomic drivers such as education level, financial security, employment status and working conditions. Detailed dispensing and treatment data and an understanding of networks of transmission and household and social dynamics can contribute to better analysis of AMR, especially when combined with local information about agricultural antibiotic use.

Geographical data about households, topography and the movement of people in the population cohort can also inform the identification of transmission hotspots and networks. This work can be reinforced within the population cohort research through phylogenetic analysis of pathogens, phylodynamics and the use of Geographical Information Systems (GIS) to identify the physical spread of pathogens. These methodologies and mapping approaches can also be used in concert with meteorological data to identify the impact of climate change on emerging diseases and new epidemics.

Genetics/Genomics

Health genomics is increasingly recognised as a key component of improving population and individual health, and health genomics research within Africa is growing⁵. Until recently, most medical research has been undertaken in northern hemisphere populations, and diagnostic and therapeutic tools are designed for and tested in those populations. African populations, however, have much greater genetic depth and diversity than rest-of-world populations because modern humans originated in Africa, expanding and diversifying on the Continent before only a small genetic subset left Africa and subsequently populated the rest of the world³⁶. Over time, African-specific environmental factors including disease challenges have also driven natural selection of certain genetic traits (for example the impact of endemic malaria on genetics³⁷). Despite this, African populations remain underrepresented in global health genomics research⁶ and diagnostic tools and therapeutics are still not optimised for Africans.

APCC could provide a unique opportunity to conduct sufficiently powered analyses of genetic factors in concert with complete clinical profiles, environmental and socioeconomic factors, and changes over time in all of these metrics. This can elucidate the aetiology of both common and rare diseases in African populations to inform better diagnostic, prognostic and therapeutic approaches. APCC could aim to establish a genomics platform by leveraging existing infrastructure on the Continent, such as the bioinformatics, genomics, biobanking and skills capacity

developed during the H3Africa programme³ (Box 6). The genomic data generated by the H3Africa programme provides valuable genomic reference panels for many African populations³⁸, which could be augmented by APCC through whole genome sequencing (WGS) of representative participant panels. With complex, granular longitudinal phenotypes described as they evolve over time, APCC could generate new knowledge about complex genetic interactions and their impact on patient outcomes over time. Better understanding of African genetics underlying disease can also help to elucidate biological mechanisms of disease to the benefit of all global populations.

Box 6: The Human Heredity and Health in Africa Consortium^{38,39} (www.h3africa.org)

The H3Africa Consortium has created an African network of research collaborations for undertaking health genomics research in Africa, supported by the Wellcome Trust (UK), the National Institutes of Health (USA) and the African Academy of Sciences (AAS). H3Africa enables competitive genomic research on the Continent, undertaken by African scientists. The initiative consists of 51 disease-specific projects, including case-control genomic studies of common non-communicable disorders such as heart, renal and metabolic diseases, as well as infectious diseases, using genetic, clinical and epidemiological methods to identify hereditary and environmental contributors to disease. Infrastructure developed includes biorepositories for secure sample storage and laboratories with capacity for processing samples for -omics research. The H3Africa bioinformatics network (H3ABioNet, www.h3abionet.org) has developed African-specific genomics tools and pipelines, for example the H3Africa genotyping chip³ and also undertaken extensive training and development of bioinformatics skills and capacity across the Continent⁴⁰. Data generated by H3Africa projects can provide important insights into background genomic diversity and African-specific aetiological variants on the Continent.

Human genomic data would be generated for participants in a disease- or hypothesis-agnostic approach. WGS of a subset of participants can provide population-specific reference panels. Genotype data – a genome-wide summary of genetic diversity using key marker variants across the genome - can be used in genome-wide association studies (GWAS) to identify markers that are associated with outcomes. The greater genetic diversity of African populations means that the sections of DNA represented by each marker - called

haploblocks - are shorter than in other populations, so that fine-mapping to identify true causal genetic variants can be more easily accomplished. Measuring this African diversity would also assist global efforts to identify aetiological variants by confirming findings from studies in other world populations and fine-mapping candidate variants, and also by excluding candidate variants that appear to be rare in other populations suggesting they are suitable as disease candidates but are in fact common in the background of greater African diversity and unlikely to drive disease. These important developments would contribute to global efforts to define the clinical impact and relevance of all human genetic variation and especially genetic determinants of disease³⁸ (for example the ClinVar database, USA).

Box 7: Priority research questions - Genomics

- *What are the genetic determinants of susceptibility and severity for common diseases in African populations, and how can these be identified through African-tailored approaches such as whole genome sequencing, GWAS, admixture mapping, polygenic risk scores, complex statistical modelling of combinations of genetic and other determinants working in combination, and measurements of complex longitudinal phenotypes?*
- *How do genetic factors work together, and in concert with environmental, socio-economic and demographic factors to drive disease phenotypes?*
- *What genetic factors can provide African-specific preventative, diagnostic and/or prognostic information?*
- *How do pharmacovariants affect efficacy and tolerance of treatment regimens for common diseases in African populations?*
- *How can pathogen genomics be used for surveillance and transmission mapping?*

Precision medicine approaches can help to improve patient outcomes through designing solutions tailored specifically to their needs, and precision public health approaches will use aggregated data to inform public health decisions about whole population groups even where genetic data are not available for all individuals. Data about genes involved in processing medications – pharmacogenes⁴¹ - can guide dispensing practice at the population-level, reducing adverse events and improve patient outcomes⁴². These data may also inform the adoption of new medications based on population-level pharmacogenomic profiles. An illustrative example is the evaluation of Efavirenz dosing in sub-Saharan Africa based on pharmacogenomic profiling of African patients⁴³⁻⁴⁵. In addition, microbiome data can provide holistic insights, and might include environmental^{46,47}, agricultural^{48,49} and human^{50,51} microbiome analysis; and pathogen genomic data can be used for surveillance of existing and emerging pathogens and their routes of transmission.

Environmental factors

The extensive impact of environmental factors is well recognised in studies of health, wellbeing and social outcomes. Environmental factors assessed by APCC could include: water quality, availability and waste water disposal; zoonotic pathogens; quality of air and associated air-borne industrial and biological pollutants through measurement of particulate density⁵²; food security and availability, nutritional value, food-related behaviours and social norms; human-animal cohabitation that can be agricultural or companionable (pets); seasonal changes in weather and weather extremes such as drought, floods and extreme heat; and long-term trends such as climate change, rising temperatures, extreme heat, changing agricultural capacity and water security. In addition, APCC would be uniquely positioned to explore human behaviours that in turn impact the environment and sustainability of resources.

SDG-6 aims to ensure availability and sustainable management of water and sanitation for all, reflecting the impacts of water scarcity, unsafe drinking water supplies and poor waste water disposal on human health. APCC could facilitate the analysis of Water, Sanitation and Hygiene (WASH) interventions in both urban⁵³ and rural^{54,55} settings, particularly relating to waterborne parasites⁵⁶⁻⁵⁸, WASH in healthcare facilities⁵⁹⁻⁶¹, and impacts on vulnerable populations such as children under 5 years^{62,63}. APCC analyses could monitor progress towards achieving SDG-6 by assessing the impact of changes in WASH metrics over time or after wider interventions and/or enhancement of clean water and sanitation resources, and monitor progress towards achieving SDG 6.

Box 8: Priority research questions – Environmental determinants of health and wellbeing

Research Questions:

- *What is the impact of environmental determinants on health and wellbeing - such as industrial pollution, the built environment, weather and climate, water and sanitation, environmental microbes, agriculture and food security - on health and wellbeing?*
- *How can we effectively measure the rapid pace of change in these metrics, the causes and impacts of ongoing changes, and inform agile interventions that can match the rate of change?*
- *What are the impacts of WASH determinants on health and wellbeing in rural and urban areas? How do changes in WASH impact health and wellbeing over time?*
- *How do climate and weather impact health and wellbeing? What are the immediate impacts of extreme weather events, and what are the long-term impacts of climate change and utilisation patterns for natural resources?*

Other environmental determinants which could be explored by APCC include food security and sustainable agriculture⁶⁴, impact on food resources – including protein-rich nutrition - of marine environments⁶⁵, inland lakes and dams in Africa⁶⁶. Climate change and significant shifts in weather patterns over time also impact human health and well-being⁶⁷, and APCC research would capture these temporal trends and changes, and their impact on health and wellbeing in alignment with **SDG-13** on climate action. In particular, the Consortium could be well-positioned to employ a [water-energy-food-ecosystem nexus](#) approach to explore these dynamics. Impact of extreme weather can similarly be assessed and temporal data will provide before and after comparison data to identify challenges in food security, secure WASH, and infectious disease challenges^{68,69} arising from events such as droughts⁷⁰, heatwaves⁷¹, floods⁷², storms and cyclones^{73,74}. The impact of livestock or other animal-related exposures to zoonotic infections in conjunction with other environmental, climate change and socioeconomic determinants will also be investigated.

With the increasing frequency of meteorological extreme events such as floods, droughts and heatwaves, the association between projected climate risks and emerging diseases will be measured in both urban and rural areas in a context of low-resilience health systems, facilitating the analysis of health system responses and adaptation to emerging diseases and climate change. Data linkage to existing neighbourhood and regional environmental data will be leveraged wherever possible, building on the learnings of similar programmes elsewhere^{75,76}. APCC could address these issues through One-Health, Eco-Health, Global Health and Planetary Health approaches. GPS and geolocation metrics will support these investigations into the importance of the immediate environment in determining health and wellbeing.

Socioeconomic factors

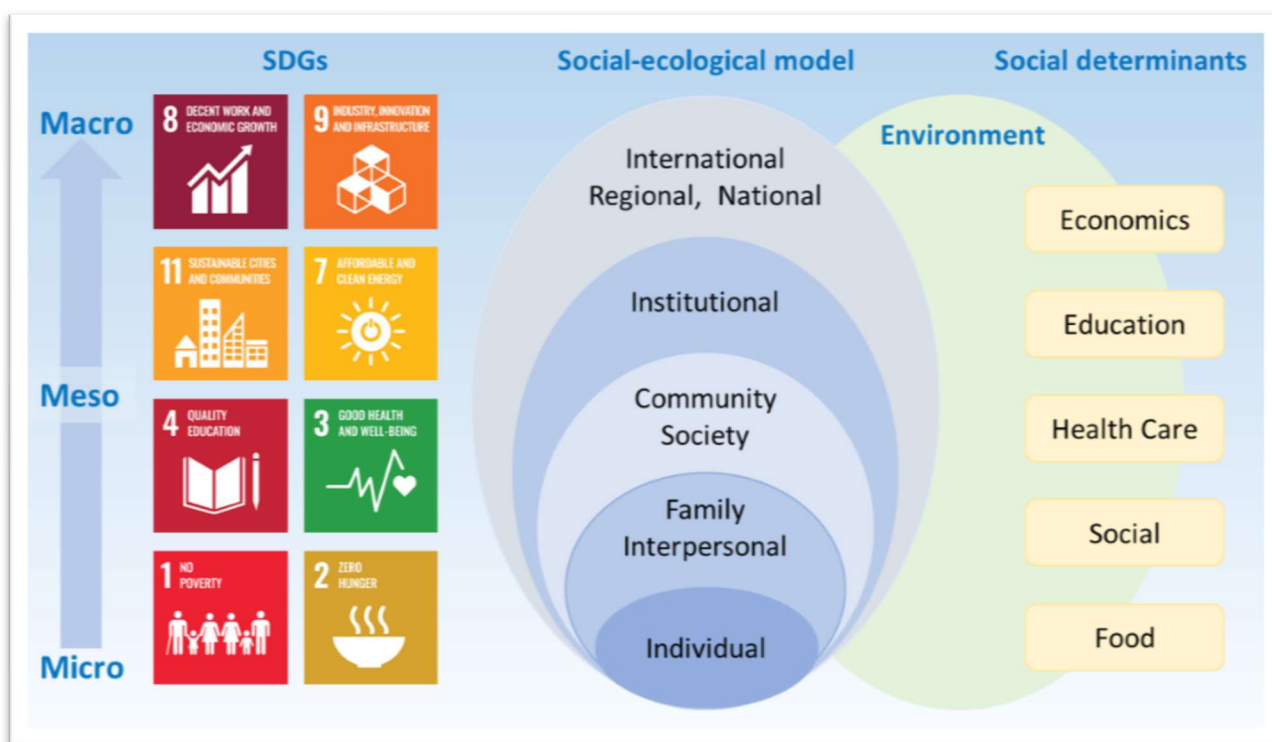
APCC could have the ability to monitor progress towards the related SDGs (Figure 1). Granular socioeconomic data, with complete and holistic data coverage at an individual level as well as community-level metrics would allow a deeper understanding of how health can impact social/human capital and wellbeing, as well as the converse relationship. Social determinants of health are well described⁷⁷, as well as how health and well-being can in turn drive socioeconomic outcomes in individuals, populations, and impact economies at a national level⁷⁸⁻⁸⁰. APCC could investigate the causes, impacts, pathways and transitions in changing socioeconomic determinants through the life course, over time, and in response to policy change. In particular, the Consortium would be well-positioned to explore the dynamics within the interface of social, economic, environmental, biological and clinical drivers of disease, health and wellbeing. APCC could explore causes and impacts of the rate of change of

socioeconomic drivers and outcomes that underlie health and wellbeing.

To provide meaningful insights that can be compared across African countries, core metrics for assessing socio-economic status, social inclusion/exclusion, living standards,

consumption and assets measurements would need to be compiled, standardised and validated. Specific metrics that define each socioeconomic stratum need to be adjusted to local norms, customs and lifestyles, and an optimal and feasible universal metric will need to be developed and validated.

Figure 4: The social-ecological model, environmental and social determinants of health, and related SDG



In addition, APCC could establish high quality qualitative studies to better understand social epidemiology through socioeconomic determinants and outcomes in African populations, using qualitative research approaches through interviews and focus groups to carefully explore the impacts of sensitive issues such as sexual practices, domestic violence, religious practices – at all times ensuring context- and socially appropriate approaches managed by local researchers who can ensure adherence to social and cultural norms and sensitivities.

Given the unique and diverse socioeconomic environments across the Continent, APCC would need to develop methods and tools, reviewing, assessing, adapting and validating existing standards and developing new tools where

appropriate, for measuring the following socioeconomic determinants:

- *Measurement of food security, including cultural variations in nutritional resources and definitions of food units and meals,*
- *Evaluation of numbers of livestock, rodents, monkeys, pets and other proximal animals, with standardised methods to define location and vector density,*
- *Data derived from meteorological stations, land-satellite images for measuring climate variability and specific weather events such as heatwaves, cold snaps, floods, and other extreme weather events,*
- *Digital data derived from wearable tech.*

Box 9: Priority research questions – Socioeconomic determinants of health and wellbeing

Health and wellbeing, and socioeconomic status may be modelled as both determinants and outcomes in recognition of the interrelatedness of these domains.

- *How can socioeconomic status be measured in an effective, accurate and comparable way for African populations, using metrics such as: life expectancy, morbidity and mortality, personal and household income, education status, workplace safety, employment status, family structure, access to social grants and services, access to health care, discrimination, food security, housing, early childhood resources and development, crime, transportation, built environment and neighbourhood, WASH, recreation and leisure, nutrition and food availability.*
- *What are the effects of socioeconomic factors on health and wellbeing, and how do health and wellbeing in turn affect socioeconomic outcomes at individual, community and national levels? How do the type and rate of change of causes and determinants impact outcomes?*
- *What are the impacts and outcomes of urbanisation through migration and also insertion of urbanisation into rural environments? How does the pace of change impact health and wellbeing, and how do health, environmental and social systems adapt to these rapid changes?*
- *How do socioeconomic determinants and outcome change during the life course and over time? What is the impact of the rate of change?*
- *What are the socioeconomic impacts of intersectoral and national policies and interventions? How is local health and wellbeing affected by changing global policies?*

Data from APCC could then be used to assess the impact of intersectoral policies^{81,82} and changes in intersectoral policies over time, exploring the tension between economy and health, as illustrated recently in the extensive debate about the impact of COVID-19 lockdown, or with examples such as smoking restrictions overriding the tobacco economy, and restrictions on advertising alcohol products having economic effects. Existing models and frameworks can be explored and adopted where appropriate, for example the socio-ecological model⁸³, social determinants of health⁷⁷.

5. Principles underpinning the APCC

The following principles will guide the development and implementation of APCC:

1. *African-led, with equitable governance of the initiative*
2. *Driven by community engagement*
3. *Support ethical, equitable and relevant use and sharing of samples and data*
4. *Strengthen African capacity and leadership*

5.1 African-led, with equitable governance structures

Proposed governance structure

APCC would be led and governed by African scientists, promoting equitable partnerships and working to the highest participant-centric ethical standards. Decision-making would be by consensus with equitable representation of all partners, with transparency about governance, processes and operations. Activities of APCC could be aligned with regional, national and Continental health and wellbeing priorities, and be guided by the SDGs. A collaborative approach will ensure that partnerships with service delivery agencies are mutually beneficial, supporting and enhancing data ecosystems and ongoing service delivery activities in each country, and maximising ongoing participant and other stakeholder benefits from data generated.

APCC would be committed to ensure fair practices, assessing and addressing potential conflicts of interest that may arise through differing priorities of diverse stakeholders that include participants, community leaders, researchers, funders and national agencies. By openly and transparently anticipating, recognising and addressing potential conflicts of interest, APCC could mitigate their negative impacts.

APCC would aspire to representativeness, inclusiveness and equity across the network and participating sites, and aim for equitable practices and benefit-sharing with the participants and communities who hold the centre of APCC. The Consortium would establish an additional layer of governance through a scientific advisory board which will have high African representation to provide contextualised and relevant oversight.

Proposed research management structure model

APCC leadership structure could include an overarching committee which will provide oversight of the activities of

APCC. A representative from each site would be included in the steering committee and working groups. A **Managing Committee** supported by a **Coordinating Centre** would be established to ensure the operational success of APCC. An **ISAB** would provide independent scientific advice to the Managing Committee. Specific domain and working groups would be established to ensure the effective progress, management, scientific integrity, research skills and infrastructural capacity strengthening, and sustainability of APCC. These working groups could include:

A **Community Engagement group** that will establish principles, facilitate shared learning and oversee and coordinate community engagement activities across the Consortium.

An **Ethics Working Group** to provide ethical oversight, advice and guidance to ensure ethical conduct and research activities across the Consortium.

A **Stakeholder Engagement Group** to undertake a stakeholder mapping exercise as well as developing and implementing strategies to engage effectively and meaningfully with the wider stakeholder community, including the public, government ministries, policy makers and the media.

A **Data Collection and Analysis group** to oversee the development of data collection tools and data standards, exploring interoperability, for example Fast Healthcare Interoperability Resources (FHIR) or other data specifications. This group would build data standards for cross-Consortium core metrics, as well as study-specific data standards. The group would be responsible for data harmonisation, data quality assurance and data analysis models; and would assist with coordinating data meta-analyses. Workstreams would be established for data linkage, biospecimens, epidemiology, statistics and computational analyses.

A **Data Governance Group** would be responsible for protocols to implement security, data access controls (structural and procedural), data anonymisation and/or perturbation, data storage, managing a federated data storage model, establishing a data access standard procedure, establishing Memorandum of Understanding (MOU) templates, collaborative agreements, sub-study approvals, publication and authorship protocols.

Code of conduct

At the start of the programme, APCC would draw up, by consensus, a code of conduct providing clear measures for network participants to report or address any experiences of bullying, discrimination, exclusionary behaviours, as well as providing mechanisms for general within-consortium feedback. Structures could be put in place to ensure inclusive practices and devolved power, ensuring meaningful participation and engagement by all members of the Consortium. Structural approaches could be developed to help to devolve excluding hierarchical structures, ensuring that opportunities are created and supported for all members of the Consortium through from senior to junior researchers and fieldworkers.

5.2 Driven by community engagement

Community engagement

APCC recognises many stakeholders in the activities of APCC who can be identified at every level of the social-ecological model shown in Figure 4. APCC could aim to develop principles of meaningful community engagement, recognising that individual LPS sites are best placed to implement engagement, and that participants carry the burden of research risks from participating in long-term LPS, including risk of breach of privacy and confidentiality as well as the possibility of disruption and inconvenience. APCC could identify community engagement partners to hear and understand the interests, needs and concerns communicated by individuals. Qualitative research should be undertaken to assist in identifying priorities, opinions and concerns within the general population about the impact and processes of LPS, and to assist APCC research programmes to align with community needs and priorities. To ensure engagement with community members is meaningful, APCC could adhere to the following principles in community engagement activities:

- *Activities will be ethical, mutually respectful, genuine, inclusive of diverse communities, and meaningful through genuinely seeking views and opinions that build the work of APCC.*
- *Local teams who speak the same language as community members and are sensitive to social norms and practices in the community will be employed.*
- *Community engagement will be central to all research activities, involving communities in the design phase of new research activities.*
- *APCC sites will work through local community leaders where possible, whilst seeking appropriate ways to ensure that there is equitable representation in community discussions from all members and sectors of the population, including women, adolescents and minority groups. Where appropriate, additional engagement*

mechanisms will be used, such as approaching administrative and traditional leaders.

- *Community engagement programmes will be designed to generate valid and reliable evidence to inform participant information and informed consent process, recruitment methods, data sharing, and the general conduct of researchers and fieldworkers within the community.*
- *Community advisory boards will be supported or new boards established at all research sites, and ongoing community engagement activities will be actively funded, remaining integral to all research programmes.*

Stakeholder engagement

APCC could seek to engage with stakeholders from all sectors, including at the macro level – national agencies and departments, regional and pan-African organisations involved in policy making and provision of services to African populations; at the meso level – organisations, health care facilities and local agencies that provide services to specific communities and populations; and at the micro level – individuals including healthcare providers, social workers, researchers and members of the general public, all who have an interest in the undertakings and outcomes of the work of APCC.

As key enablers to the success of the Consortium, APCC would need to engage with policymakers from the outset to ensure research is designed in collaboration with and meets the needs of key stakeholders. Ensuring early buy in through co-development would help to shape the research effectively, and build understanding of the research process and outputs amongst policymakers, supporting APCC to achieve its objectives and bridge the transition from research output to translation into policy. This is a key aim to mitigate risks to long term sustainability of the Consortium.

APCC could, in the formative phase, conduct a stakeholder mapping exercise to identify key stakeholders, recognising that these may differ according to region, country and context of proposed LPS sites. Key questions could include: who are the stakeholders and what are their lenses on LPS research; which APCC outputs are they invested in supporting; what is their level of influence in helping to achieve these outputs; and how can APCC effectively engage these stakeholders in complementary and collaborative relationships that are mutually beneficial to all parties?

Representatives of stakeholder sectors could be engaged through structured interviews to understand priorities, needs, and relevant outputs from the work of APCC, and to begin building collaborative relationships, guiding study design and practices to ensure outputs are relevant and useful to key stakeholders. Through active community and stakeholder engagement, and meaningful qualitative research APCC aims to reconcile and align the agendas of all stakeholders with the aims and outcomes of APCC programmes.

Box 10: Examples of stakeholders in longitudinal population studies

Micro-level stakeholders are individuals and individual-level participants, reflecting individuals’ intentions, feelings, and beliefs. **Meso-level** stakeholders represent groups of people such as units and organisations. **Macro-level** stakeholders are those in political and administrative environments, such as national, regional and global policy makers and organisations.

Micro-level stakeholders	Meso-level stakeholders	Macro-level stakeholders
Participants	Communities	Inter/National agencies, civil registration
Family members	Research Institutions	National Departments of Health
Households	Local health care service providers	Inter/National Public health networks
Traditional Healers	Local social service providers	Inter/National Social Services
Healthcare providers	Non-Governmental Organisations	Environmental Services
Researchers	Schools	Parliamentarians
Field workers	Institutes of higher learning	Policy makers
Community Health Workers	Information & Communications Technology (ICT) firms, digital services providers	Regional organisations
	Ethics and regulatory boards	
	Media	
	Inter/National Science Funding Bodies (academic, industry or philanthropy)	
	Other cohort studies	

5.3 Support ethical, equitable and relevant use and sharing of samples and data

Ethical consent framework

As part of the formative phase of the Consortium, APCC could prioritise developing a consensus ethical framework for its research activities. This framework would build on guidelines and recommendations from the [AAS Data and Biospecimen Governance Committee](#)⁸⁴ and the experiences of Consortium members in other research on the Continent, promoting appropriate use of data and biospecimens to maximise benefits to participants, populations and scientific discovery, whilst recognising that not all data may be appropriate to share in areas of specific sensitivities, and ensuring respect for participant autonomy, choice and right to privacy.

APCC could ensure that the consent process is transparent and fit for purpose with a focus on providing meaningful and accessible information to participants, ensuring autonomy in their decision to participate, and ensuring that the consent process provides a clear, informed and unequivocal mandate for sharing of anonymised data for meta-analysis and comparative purposes.

Clear guidelines could be established for participant referral to appropriate health care and social services, compulsory disclosure for notifiable diseases, returning actionable versus non-actionable incidental findings in the context of available

health care, and the return of incidental findings directly to health departments through data linkage.

Box 11: Consent models in common use Open research values

Specific Consent is consent by the participant for just the current study and no other onward or secondary use of samples or data.

Tiered Consent⁸⁵ provides detailed information about the intended specimen/data use and storage, and provides a variety of types of data/sample sharing options that the participant can opt into according to their preferences, thus enabling participants to individually select a level of specimen and/or data sharing that they are comfortable with through recorded responses to specific questions

Dynamic consent^{86,87} creates a two-way, ongoing dialogue between the researcher and participant through which the participant can decide whether to contribute their samples and/or data to new research studies, as the information about those studies becomes available.

Broad consent⁸⁸ works on the principal that once the sample and/or data are retained by the researcher, then that researcher may decide, with ethical review board oversight, where and how to continue using those resources without further input from the person who donated them.

APCC could support Open Science values, ensuring visibility and discoverability of research output, and will facilitate data access aligned with APCC principles, ethics and informed consent protocols.

APCC could address specific consent for sharing of data internationally outside the Consortium, as this may be legally required in some countries, and special consideration would be given to the onward use of genomic data because they cannot be anonymised. During the formative phase, a scoping exercise could explore the legal requirements for data sharing to ensure legal compliance by APCC partners, and participant information and consent processes will be explicit and transparent about these special considerations.

In defining these principles, consideration would need to be given to the potential dual nature of APCC as a research platform as well as its support of public health services through data linkage and data return. All activities would remain within the remit of national Health Acts, Children's Acts and Privacy laws. All research activities by APCC would be subject to ethical review, and during the implementation phase recommendations could be drawn up regarding which committees will be approached for ethics review, respecting the different ethics review requirements at different institutions and in different countries.

Data and biospecimen governance

APCC aims to collect, analyse and distribute data and biospecimens under a clear data governance framework that upholds key principles of governance to facilitate appropriate, secure and appropriately managed data and biospecimen sharing and use, whilst simultaneously preventing data breach, misuse, or unethical or illegal repurposing of data or samples. The consortium will uphold the principles of responsible sharing by taking on the role of stewards, recognising and respecting limits to sharing, and acknowledging that ownership of samples and data is highly contested and is held by the participants. In order to achieve this aim, APCC could build a data and biospecimen governance framework that will uphold these principles, drawing on the experiences of existing consortia and research programmes and building on existing governance frameworks⁸⁶. Elements of data governance could include:

- *Ensuring ethical use of resources (data and biospecimens) in line with informed consent given.*
- *Exploring National/regional legislation about permitted data and biospecimen use, including cross-border data and sample sharing, and providing guidelines for ensuring compliance.*
- *Developing guidelines for structural and technological data protection for data access control, including firewalls, encryption, database architecture, passwords and system administration to manage access to data. Data backup*

protocols and disaster management plans will also be developed.

- *Providing procedural protocols for transparent, standardised and well-documented processes for requesting access to resources, data and biospecimen access committee constitution and functioning, and other procedures that will govern the appropriate onward use of APCC resources.*
- *Implementing a sustainability plan to ensure that data are well documented, standardised and backed up, and meta-data are up to date and standardised, supporting Open Data⁸⁹, FAIR principles⁹⁰, TRUST principles⁹¹ and SHARE principles where appropriate and ethical.*

5.4 Strengthen African capacity and leadership

Skills and research leadership capacity could be strengthened through supporting Principle Investigators, mid- and early career scientists. APCC could provide training and mentorship to develop new research leadership on the Continent and provide emerging scientists with research and career opportunities to retain them within Africa.

Capacity strengthening objectives

Capacity strengthening will enhance the immediate APCC research infrastructure, as well as more widely providing a substantial contribution to robust and sustainable scientific research infrastructure, knowledge and skills on the Continent. Through interactions and collaborations with existing programmes, leveraging existing infrastructure and interfacing with existing activities such as population surveys, APCC offers high value for money. In addition, APCC activities would be able to enhance capacity for stakeholder engagement, developing ongoing relationships between research and service delivery enterprises, and create durable avenues to engagement with policy makers and to contributing to policy decisions.

Skills strengthening and retaining skills on the Continent

APCC could adopt different avenues for skills strengthening to ensure scientific quality and sustainability over time, such as research training activities. A range of different training modalities will ensure that early career scientists receive the necessary high-quality training and experience to equip them as future African scientific leaders with both scientific and leadership skills. Careful succession planning through mentoring and development of a clear career path towards more senior roles should be included by participating cohorts to equip early career scientists. APCC could also be instrumental in ensuring that African scientists are retained on

the Continent, providing employment for local researchers and fieldworkers, and providing exciting opportunities to undertake high quality research and to be competitive in applications for international funding. APCC could draw on the experiences of ongoing skills development programmes such as the [DELTA Africa Initiative](#), training undertaken by [H3ABioNet](#) and [Fogarty training programmes](#), to inform the skills strengthening plan. Training focus areas will include data science and analytical skills, data linkage, epidemiology, biostatistics, health informatics, qualitative research, community engagement, bioethics and health systems policy and research. Training should also be developed for financial and project management.

Infrastructure capacity strengthening

Infrastructural strengthening by APCC could maximise and enhance existing systems whilst adding new infrastructure where needed. Data collection and analysis will require stable, secure and large data storage and processing servers, as well as sufficient and secure networks for the transfer of large datasets. Data linkage to existing electronic service delivery platforms will require both effective and secure, mutually acceptable modes of data access and data return, as well as appropriate skills development for data analysts able to develop and manage data linkage algorithms. Growing availability of within-Africa cloud capabilities should be explored for data management, whilst remaining mindful of data sensitivity, security and governance requirements. In addition, growing mobile phone use, increasing spread of smart phones and internet/mobile network availability provide evolving opportunities over the medium and long term for expanding digital health innovations. Examples include: extending mHealth solutions for field workers, providing interactive platforms for participants, access to call centres, messaging services and Q&A forums (e.g. WhatsApp, SMS, social media platforms) and the use of wearable tech for data collection. Increasing digital access and literacy will also provide opportunities for improved community engagement and dynamic consent processes. Where samples will be collected, biorepositories such as the established H3Africa biobanks³ will provide a strong basis from which to develop APCC biobanking infrastructure and laboratory capacity across the Continent.

6. The way forward

Phase 1: Scoping phase

This report presents the findings from the scoping phase. This phase included a meeting in Uganda (March 2020) where the need, vision and ambition for a research population data platform to address the Continent's most pressing health and socioeconomic needs was agreed. This was followed by a consultative process to write this scoping paper, involving the Steering Group, research funders, and led by Dr Nicki Tiffin. The result of the scoping phase is agreement on the need for, vision and broad structure of APCC, and clear areas identified for further exploration in the formative phase.

The scoping phase is exploratory and to date funders have not committed any funds to the implementation of APCC.

Phase 2: Formative phase

The focus activities of the formative phase will provide a consensus blueprint for the activities of APCC core and affiliate sites, as well as establishing ongoing community and stakeholder relationships that will endure throughout the programme. The formative phase will include activities to build out plans for the proposed structure and key principles for APCC. This will include developing detailed specifications, protocols and research plans to determine a consensus blueprint that can be used to establish APCC successfully. This will include:

1. Scientific aims: articulate a research vision for the surveillance, discovery and policy evaluation aims of the Consortium and prioritise the specific scientific aims that will guide the implementation phase.
2. Establish the requirements for the platforms that will enable delivering the scientific vision:
 - Population Data Platform: the LPS (e.g. sites inclusion criteria, core data to be collected, data collection methods, etc.)
 - Biobank & 'omics' platform
 - Data platform enabling discoverability, accessibility and sharing of APCC sites data: establish requirements, existing models and on-going initiatives in the Continent.
 - External data linkages: what current models of data linkages exist in the Continent, what's possible in this space

3. Agreed model of governance and management structure of APCC
4. Agreed ethics and stakeholder engagement principles
5. Measures of success
6. Conduct a risk assessment
7. Provide full costings for the implementation phase and 5 to 10-year costs

Funders would aim to deliver this phase over a 12-month period in partnership with an African-led Agency or Consortium awarded through an open tendering process. The product of the formative phase will be a White Paper outlining the blue print and costs for APCC. Funders will use this to decide whether they contribute funds to the full programme of work.

Phase 3: Implementation phase

Once funders have agreed to commit funds, the **implementation phase** can begin. Funds would be awarded through a competitive process to Core and Affiliate sites, and to fund the secretariat/co-ordinating hub. Pilot studies would be conducted for example to finalise engagement plans and core data collection methods, before data collection would start.

7. Conclusion

The vision laid out in this scoping paper is rightly ambitious. But it is time for a step-change in ambition for population-based science in Africa. Recent developments on the Continent including large scale genomics research (such as H3Africa), capacity building programmes (such as DELTAS) and existing research infrastructure (including a network of HDSS sites, bioinformatics hubs and biorepositories), mean that a more co-ordinated and ambitious vision is within reach.

The task is now to work closely with national and regional policy makers in Africa, African scientists and most importantly affected communities, to refine the vision and scope of APCC, and to make this ambitious vision a reality.

8. References

- 1a Popejoy B.A., Fullerton S.M. Genomics is failing on diversity, *Nature* 538:161-164
- Nabukalu, J. B., Asamani, J. A. & Nabyonga-Orem, J. Monitoring Sustainable Development Goals 3:
1b Assessing the Readiness of Low- and Middle-Income Countries. *Int J Health Policy Manag* 9, 297–308
(2020).
- 2 Aftab, W. *et al.* Implementation of health and health-related sustainable development goals: progress,
challenges and opportunities - a systematic literature review. *BMJ Glob Health* 5, (2020).
- 3 Mulder, N. *et al.* H3Africa: current perspectives. *Pharmgenomics Pers Med* 11, 59–66 (2018).
- 4 Swart, Y., van Eeden, G., Sparks, A., Uren, C. & Möller, M. Prospective avenues for human population
genomics and disease mapping in southern Africa. *Mol Genet Genomics* 295, 1079–1089 (2020).
- 5 Rotimi, C. N. *et al.* The genomic landscape of African populations in health and disease. *Hum Mol Genet*
26, R225–R236 (2017).
- 6 Popejoy, A. B. & Fullerton, S. M. Genomics is failing on diversity. *Nature* 538, 161–164 (2016).
- 7a Sankoh, O. CHESS: an innovative concept for a new generation of population surveillance. *The Lancet*
Global Health 3, e742 (2015).
- 7b Choudhury, A., Aron, S., Botigué, L.R. *et al.* High-depth African genomes inform human migration and
health. *Nature* 586, 741–748 (2020)
- 8 Stark, Z. *et al.* Integrating Genomics into Healthcare: A Global Responsibility. *Am J Hum Genet* 104, 13–
20 (2019).
- 9 Marx, V. The DNA of a nation. *Nature* 524, 503–505 (2015).
- 10 Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of
complex diseases of middle and old age. *PLoS Med.* 12, e1001779 (2015).
- 11 Nagai, A. *et al.* Overview of the BioBank Japan Project: Study design and profile. *J Epidemiol* 27, S2–S8
(2017).
- 12 Boulle, A. *et al.* Data Centre Profile: The Provincial Health Data Centre of the Western Cape Province,
South Africa. *IJPDS* 4, (2019).
- 13 World Health Organisation & International Longevity Centre UK. A Life Course Approach To Health.
(2000).
- 14 Bilkey, G. A. *et al.* Optimizing Precision Medicine for Public Health. *Front Public Health* 7, 42 (2019).
- 15 Nuttall, I., Miyagishima, K., Roth, C. & de La Rocque, S. The United Nations and One Health: the
International Health Regulations (2005) and global health security. *Rev. - Off. Int. Epizoot.* 33, 659–668
(2014).
- 16 Domenech, J., Lubroth, J., Eddi, C., Martin, V. & Roger, F. Regional and international approaches on
prevention and control of animal transboundary and emerging diseases. *Ann. N. Y. Acad. Sci.* 1081, 90–
107 (2006).
- 17 Pike, J., Bogich, T., Elwood, S., Finnoff, D. C. & Daszak, P. Economic optimization of a global strategy to
address the pandemic threat. *Proc. Natl. Acad. Sci. U.S.A.* 111, 18519–18523 (2014).
- 18 Cleaveland, S. *et al.* One Health contributions towards more effective and equitable approaches to health
in low- and middle-income countries. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 372, (2017).
- 19 Wernli, D. *et al.* Evidence for action: a One Health learning platform on interventions to tackle antimicrobial
resistance. *Lancet Infect Dis* (2020) doi:10.1016/S1473-3099(20)30392-3.
- 20 Rodin, J. & Ferranti, D. de. Universal health coverage: the third global health transition? *The Lancet* 380,
861–862 (2012).
- 21 Watkins, D. A. *et al.* Universal Health Coverage and Essential Packages of Care. in *Disease Control*
Priorities: Improving Health and Reducing Poverty (eds. Jamison, D. T. *et al.*) (The International Bank for
Reconstruction and Development / The World Bank, 2017).
- 22 Moreno-Serra, R. & Smith, P. C. Does progress towards universal health coverage improve population
health? *The Lancet* 380, 917–923 (2012).
- 23 Organisation mondiale de la santé & Groupe de la Banque mondiale. *Tracking universal health coverage:*
first global monitoring report. (World Health Organization, 2015).
- 24 Osakunor, D. N. M., Sengeh, D. M. & Mutapi, F. Universal Health Coverage in Africa: Coinfections and
Comorbidities. *Trends Parasitol.* 34, 813–817 (2018).
- 25 Fraser, A. *et al.* Time-critical conditions: assessment of burden and access to care using verbal autopsy in
Agincourt, South Africa. *BMJ Glob Health* 5, e002289 (2020).
- 26 Mpimbaza, A. *et al.* Verbal Autopsy: Evaluation of Methods to Certify Causes of Death in Uganda. *PLoS*
ONE 10, e0128801 (2015).

- 27 Herrera, S. *et al.* A systematic review and synthesis of the strengths and limitations of measuring malaria mortality through verbal autopsy. *Malar. J.* **16**, 421 (2017).
- 28 Nichols, E. K. *et al.* The WHO 2016 verbal autopsy instrument: An international standard suitable for automated analysis by InterVA, InSilicoVA, and Tariff 2.0. *PLoS Med.* **15**, e1002486 (2018).
- 29 Uneke, C. J., Uro-Chukwu, H. C. & Chukwu, O. E. Validation of verbal autopsy methods for assessment of child mortality in sub-Saharan Africa and the policy implication: a rapid review. *Pan Afr Med J* **33**, 318 (2019).
- 30 Castillo, P. *et al.* Validity of a Minimally Invasive Autopsy for Cause of Death Determination in Adults in Mozambique: An Observational Study. *PLoS Med.* **13**, e1002171 (2016).
- 31 Misganaw, A., Mariam, D. H., Araya, T. & Ayele, K. Patterns of mortality in public and private hospitals of Addis Ababa, Ethiopia. *BMC Public Health* **12**, 1007 (2012).
- 32 Adeolu, A. *et al.* Pattern of death in a Nigerian teaching hospital; 3-decade analysis. *Afr Health Sci* **10**, 266–272 (2010).
- 33 Afolabi, B. M., Clement, C. O., Ekundayo, A. & Dolapo, D. A hospital-based estimate of major causes of death among under-five children from a health facility in Lagos, Southwest Nigeria: possible indicators of health inequality. *Int J Equity Health* **11**, 39 (2012).
- 34 Einterz, E. M. & Bates, M. Causes and circumstances of death in a district hospital in northern Cameroon, 1993–2009. *Rural Remote Health* **11**, 1623 (2011).
- 35 Hanefeld, J., Powell-Jackson, T. & Balabanova, D. Understanding and measuring quality of care: dealing with complexity. *Bull. World Health Organ.* **95**, 368–374 (2017).
- 36 Shriner, D., Tekola-Ayele, F., Adeyemo, A. & Rotimi, C. N. Genome-wide genotype and sequence-based reconstruction of the 140,000 year history of modern human ancestry. *Sci Rep* **4**, 6055 (2014).
- 37 Kariuki, S. N. & Williams, T. N. Human genetics and malaria resistance. *Hum. Genet.* **139**, 801–811 (2020).
- 38 Choudhury, A. *et al.* High-depth African genomes inform human migration and health. *Nature* **586**, 741–748 (2020).
- 39 H3Africa Consortium *et al.* Research capacity. Enabling the genomic revolution in Africa. *Science* **344**, 1346–1348 (2014).
- 40 Mulder, N. J. *et al.* H3ABioNet, a sustainable pan-African bioinformatics network for human heredity and health in Africa. *Genome Res.* **26**, 271–277 (2016).
- 41 Mpye, K. L. *et al.* Disease burden and the role of pharmacogenomics in African populations. *Glob Health Epidemiol Genom* **2**, e1 (2017).
- 42 Alessandrini, M., Chaudhry, M., Dodgen, T. M. & Pepper, M. S. Pharmacogenomics and Global Precision Medicine in the Context of Adverse Drug Reactions: Top 10 Opportunities and Challenges for the Next Decade. *OMICS* **20**, 593–603 (2016).
- 43 Nyakutira, C. *et al.* High prevalence of the CYP2B6 516G-->T(*6) variant and effect on the population pharmacokinetics of efavirenz in HIV/AIDS outpatients in Zimbabwe. *Eur. J. Clin. Pharmacol.* **64**, 357–365 (2008).
- 44 Masimirembwa, C., Dandara, C. & Leutscher, P. D. C. Rolling out Efavirenz for HIV Precision Medicine in Africa: Are We Ready for Pharmacovigilance and Tackling Neuropsychiatric Adverse Effects? *OMICS* **20**, 575–580 (2016).
- 45 Ngaimisi, E. *et al.* Importance of Ethnicity, CYP2B6 and ABCB1 Genotype for Efavirenz Pharmacokinetics and Treatment Outcomes: A Parallel-Group Prospective Cohort Study in Two Sub-Saharan Africa Populations. *PLOS ONE* **8**, e67946 (2013).
- 46 Miller, E. T., Svanbäck, R. & Bohannan, B. J. M. Microbiomes as Metacommunities: Understanding Host-Associated Microbes through Metacommunity Ecology. *Trends Ecol. Evol. (Amst.)* **33**, 926–935 (2018).
- 47 Kvich, L., Burmølle, M., Bjarnsholt, T. & Lichtenberg, M. Do Mixed-Species Biofilms Dominate in Chronic Infections?-Need for in situ Visualization of Bacterial Organization. *Front Cell Infect Microbiol* **10**, 396 (2020).
- 48 Sarkar, A. *et al.* Microbial transmission in animal social networks and the social microbiome. *Nat Ecol Evol* **4**, 1020–1035 (2020).
- 49 Checcucci, A. *et al.* Exploring the Animal Waste Resistome: The Spread of Antimicrobial Resistance Genes Through the Use of Livestock Manure. *Front Microbiol* **11**, 1416 (2020).
- 50 Celiker, C. & Kalkan, R. Genetic and epigenetic perspective of microbiota. *Appl. Microbiol. Biotechnol.* **104**, 8221–8229 (2020).
- 51 Miro-Blanch, J. & Yanes, O. Epigenetic Regulation at the Interplay Between Gut Microbiota and Host Metabolism. *Front Genet* **10**, 638 (2019).
- 52 Bulto, T. W. Influence of particulate matter on human health in selected African provinces: mini-review. *Rev Environ Health* (2020) doi:10.1515/reveh-2020-0013.
- 53 Mackinnon, E. *et al.* 21st century research in urban WASH and health in sub-Saharan Africa: methods and outcomes in transition. *Int J Environ Health Res* **29**, 457–478 (2019).
- 54 O'Donovan, J. *et al.* Participatory approaches, local stakeholders and cultural relevance facilitate an impactful community-based project in Uganda. *Health Promot Int* (2020) doi:10.1093/heapro/daz127.
- 55 Whitley, L. *et al.* A framework for targeting water, sanitation and hygiene interventions in pastoralist populations in the Afar region of Ethiopia. *Int J Hyg Environ Health* **222**, 1133–1144 (2019).

- 56 Cissé, G. Food-borne and water-borne diseases under climate change in low- and middle-income countries: Further efforts needed for reducing environmental health exposure risks. *Acta Trop.* **194**, 181–188 (2019).
- 57 Ahmed, S. A., Guerrero Flórez, M. & Karanis, P. The impact of water crises and climate changes on the transmission of protozoan parasites in Africa. *Pathog Glob Health* **112**, 281–293 (2018).
- 58 Archer, J. *et al.* Intestinal Schistosomiasis and Giardiasis Co-Infection in Sub-Saharan Africa: Can a One Health Approach Improve Control of Each Waterborne Parasite Simultaneously? *Trop Med Infect Dis* **5**, (2020).
- 59 Weber, N. *et al.* Strengthening Healthcare Facilities Through Water, Sanitation, and Hygiene (WASH) Improvements: A Pilot Evaluation of ‘WASH FIT’ in Togo. *Health Secur* **16**, S54–S65 (2018).
- 60 Huttinger, A. *et al.* Water, sanitation and hygiene infrastructure and quality in rural healthcare facilities in Rwanda. *BMC Health Serv Res* **17**, 517 (2017).
- 61 Mulogo, E. M. *et al.* Water, Sanitation, and Hygiene Service Availability at Rural Health Care Facilities in Southwestern Uganda. *J Environ Public Health* **2018**, 5403795 (2018).
- 62 Momberg, D. J. *et al.* Water, sanitation and hygiene (WASH) in sub-Saharan Africa and associations with undernutrition, and governance in children under five years of age: a systematic review. *J Dev Orig Health Dis* 1–28 (2020) doi:10.1017/S2040174419000898.
- 63 Momberg, D. J. *et al.* Intersectoral (in)activity: towards an understanding of public sector department links between water, sanitation and hygiene (WASH) and childhood undernutrition in South Africa. *Health Policy Plan* (2020) doi:10.1093/heapol/czaa028.
- 64 Herrero, M. *et al.* Farming and the geography of nutrient production for human use: a transdisciplinary analysis. *Lancet Planet Health* **1**, e33–e42 (2017).
- 65 Costello, C. *et al.* The future of food from the sea. *Nature* (2020) doi:10.1038/s41586-020-2616-y.
- 66 Beveridge, M. C. M. *et al.* Meeting the food and nutrition needs of the poor: the role of fish and the opportunities and challenges emerging from the rise of aquaculture. *J. Fish Biol.* **83**, 1067–1084 (2013).
- 67 Mpandeli, S. *et al.* Climate Change Adaptation through the Water-Energy-Food Nexus in Southern Africa. *Int J Environ Res Public Health* **15**, (2018).
- 68 Tonnang, H. E. Z., Kangalawe, R. Y. M. & Yanda, P. Z. Predicting and mapping malaria under climate change scenarios: the potential redistribution of malaria vectors in Africa. *Malar. J.* **9**, 111 (2010).
- 69 Bryson, J. M. *et al.* Neglected Tropical Diseases in the Context of Climate Change in East Africa: A Systematic Scoping Review. *Am. J. Trop. Med. Hyg.* **102**, 1443–1454 (2020).
- 70 Ngcamu, B. S. & Chari, F. Drought Influences on Food Insecurity in Africa: A Systematic Literature Review. *Int J Environ Res Public Health* **17**, (2020).
- 71 Pasquini, L., van Aardenne, L., Godsmark, C. N., Lee, J. & Jack, C. Emerging climate change-related public health challenges in Africa: A case study of the heat-health vulnerability of informal settlement residents in Dar es Salaam, Tanzania. *Sci. Total Environ.* **747**, 141355 (2020).
- 72 Codjoe, S. N. A. *et al.* Impact of extreme weather conditions on healthcare provision in urban Ghana. *Soc Sci Med* **258**, 113072 (2020).
- 73 Pozniak, A., Atzori, A., Marotta, C., Di Gennaro, F. & Putoto, G. HIV continuity of care after Cyclone Idai in Mozambique. *Lancet HIV* **7**, e159–e160 (2020).
- 74 Devi, S. Cyclone Idai: 1 month later, devastation persists. *The Lancet* **393**, 1585 (2019).
- 75 Thomson, D. R. *et al.* Extending Data for Urban Health Decision-Making: a Menu of New and Potential Neighborhood-Level Health Determinants Datasets in LMICs. *J Urban Health* **96**, 514–536 (2019).
- 76 the SALURBAL Group *et al.* Building a Data Platform for Cross-Country Urban Health Studies: the SALURBAL Study. *J Urban Health* **96**, 311–337 (2019).
- 77 Catalyst, N. Social Determinants of Health (SDOH). *NEJM Catalyst* (2017).
- 78 Ngwenya, N. *et al.* Behavioural and socio-ecological factors that influence access and utilisation of health services by young people living in rural KwaZulu-Natal, South Africa: Implications for intervention. *PLOS ONE* **15**, e0231080 (2020).
- 79 Kabudula, C. W. *et al.* Socioeconomic differences in mortality in the antiretroviral therapy era in Agincourt, rural South Africa, 2001–13: a population surveillance analysis. *Lancet Glob Health* **5**, e924–e935 (2017).
- 80 Pillay-van Wyk, V. & Bradshaw, D. Mortality and socioeconomic status: the vicious cycle between poverty and ill health. *Lancet Glob Health* **5**, e851–e852 (2017).
- 81 Lencucha, R. & Thow, A. M. Intersectoral policy on industries that produce unhealthy commodities: governing in a new era of the global economy? *BMJ Glob Health* **5**, (2020).
- 82 Watkins, D. A. *et al.* Intersectoral Policy Priorities for Health. in *Disease Control Priorities: Improving Health and Reducing Poverty* (eds. Jamison, D. T. *et al.*) (The International Bank for Reconstruction and Development / The World Bank, 2017).
- 83 Holt-Lunstad, J. Why Social Relationships Are Important for Physical Health: A Systems Approach to Understanding and Modifying Risk and Protection. *Annu Rev Psychol* **69**, 437–458 (2018).
- 84 Nordling, L. Africa’s science academy leads push for ethical data use. *Nature* **570**, 284–285 (2019).
- 85 Nembaware, V. *et al.* A framework for tiered informed consent for health genomic research in Africa. *Nat Genet* (2019) doi:10.1038/s41588-019-0520-x.
- 86 Dankar, F. K. *et al.* Dynamic-informed consent: A potential solution for ethical dilemmas in population sequencing initiatives. *Comput Struct Biotechnol J* **18**, 913–921 (2020).

- 87 Kaye, J. *et al.* Dynamic consent: a patient interface for twenty-first century research networks. *European Journal of Human Genetics* **23**, 141–146 (2015).
- 88 Hofmann, B. Broadening consent—and diluting ethics? *Journal of Medical Ethics* **35**, 125–129 (2009).
- 89 Elsevier. The evolution of open science – how digitization is transforming research. *Elsevier Connect* <https://www.elsevier.com/connect/the-evolution-of-open-science-how-digitization-is-transforming-research>.
- 90 The FAIR Data Principles. The FAIR Data Principles | FORCE11. *The FAIR Data Principles* <https://www.force11.org/group/fairgroup/fairprinciples>.
- 91 Lin, D. The TRUST Principles for digital repositories. 5.

Appendix 1: Further contributors to the Scoping Report

Reviewers & Interviewees

University of Oxford, UK	Rory Collins
London School of Hygiene and Tropical Medicine, UK & Malawi Epidemiology and Intervention Research Unit, Malawi	Mia Crampin
NCDE, London School of Hygiene & Tropical Medicine, The African Computational Genomics Research Group & Uganda Medical Informatics Centre, MRC/UVRI & LSHTM, Uganda	Segun Fatumo
Duke Centre for Applied Genomics and Precision Medicine, US	Geoffrey Ginsburg
Bristol University, UK	Deborah Lawlor
Department of Science and Innovation, South Africa	Sagren Moodley
University of Nigeria	Obinna Onwujekwe

Attendees of scoping meeting in Uganda

Elhadji Ba	Institut de Recherche pour le Développement, Senegal
Bassirou Bonfoh	Afrique One-ASPIRE programme at Centre Suisse de Recherches Scientifiques en Côte d'Ivoire (CSRS)
Andrew Boule	University of Cape Town, South Africa
Elaine Boylan	MRC
Mark Collinson	South African Population Research Infrastructure Network
Mia Crampin	Malawi Epidemiology and Intervention Research Unit
	University of Glasgow
	London School of Hygiene and Tropical Medicine
Mary De Silva	Wellcome
Kirsty Le Doare	MRC/UVRI at LSHTM Uganda Research Unit
	St. George's University of London
Audrey Duncanson	Wellcome
Jacques Emina	Population and Health Research Institute
	Department of Population and Development Studies, University of Kinshasa
	Democratic Republic of Congo
Bruna Galobardes	Wellcome
Geoffrey Ginsburg	Duke University
	Global Genomic Medicine Collaborative
	International 100K Cohort Consortium
Branwen J Hennig	Wellcome
Kobus Herbst	DSI-MRC South African Population Research Infrastructure Network (SAPRIN)
	Africa Health Research Institute, South Africa
Momodou Jasseh	MRC Unit The Gambia at London School of Hygiene and Tropical Medicine
Jill Jones	MRC
Dan Kajungu	Makerere University Centre for Health and Population Research, Uganda
Pontiano Kaleebu	MRC/UVRI & LSHTM Uganda Research Unit
	Uganda Virus Research Institute
	LSHTM
Andre Pascal Kengne	Non-Communicable Diseases Research Unit, South African Medical Research Council
Catherine Kyobutungi	African Population and Health Research Center, Kenya
Eusebio Macete	National Directorate of Public Health, Ministry of Health, Mozambique

Nyovani Madise	Manhiça Health Research Centre, Manhiça Foundation, Mozambique
Jennifer Maroa	African Institute for Development Policy, Kenya
Pamela Mason	AAS
Rachel Miles	ESRC
Jeffrey Mphahlele	Wellcome
Nicola Mulder	South African Medical Research Council
Joseph Mwelwa	University of Cape Town, South Africa, H3ABioNet
Moffat Nyirenda	Joint Minds Consult, Botswana
Mayowa Ojo Owolabi	MRC/UVRI & LSHTM Uganda Research Unit
	World Federation for NeuroRehabilitation
	Faculty of Clinical Sciences & Centre for Genomic and Precision Medicine, University of Ibadan, Nigeria
Michèle Ramsay	University of the Witwatersrand, South Africa
Rindra Randremanana	Institut Pasteur de Madagascar
Janet Seeley	LSHTM
	MRC/UVRI and LSHTM Uganda Research Unit, Uganda
	Africa Health Research Institute, South Africa
Anjali Sharma	Centre for Infectious Disease Research in Zambia (CIDRZ)
Izukanji Sikazwe	Centre for Infectious Disease Research in Zambia
Emma Slaymaker	ALPHA Network/London School of Hygiene and Tropical Medicine
Abdrmane Soura	Institut Supérieur des Sciences de la Population, Burkina Faso
Neo Tapela	Botswana-Harvard AIDS Institute Partnership, Botswana
	Nuffield Department of Population Health, University of Oxford
Ricardo Thompson	Chokwe Health Research and Training Centre, Mozambique
Stephen Tollman	MRC/Wits University Rural Public Health and Health Transitions Research Unit (Agincourt), South Africa
	University of the Witwatersrand
	Medical Research Council South Africa
Emily Wong	Africa Health Research Institute, South Africa
Marcel Yotebieng	Albert Einstein College of Medicine, New York, USA

Appendix 2: Existing Cohorts, LPS and Networks in Africa

A2.1. Table of existing cohort and population studies in Africa

This table details many of the existing studies and networks ongoing. The landscape of studies ongoing is complex and changing and this list may not be complete.

Corrections and additions to this table are very welcome.

Study name	HDSS site	Region	Location (Africa)	Approximate sample size
ACCME'S (African Collaborative Centre for Microbiome and Genomics Research's HPV and Cervical Cancer Study)	N	West	Nigeria	11 400 (final evaluation 5 400)
Agincourt	Y	South	South Africa	60,000 (1992), 117000 (2019)
AHRI Cohort	N	South	South Africa	137 000
ALPHA network (10 HIV-focused, 9 HDSS sites)	N	Multi	Kenya, Malawi, Tanzania, South Africa, Uganda Zimbabwe	
ANDLA Network	Y	Multi	Kenya	
ANDLA Network	Y	Multi	Malawi	
ANDLA Network	Y	Multi	South Africa	
ANDLA Network	Y	Multi	Tanzania	
ANDLA Network	Y	Multi	Uganda	
Arba Minch HDSS	Y	East	Ethiopia	
AWI-Gen (H3Africa)	N	Multi	Burkina Faso	12 000
AWI-Gen (H3Africa)	N	Multi	Ghana	12 000
AWI-Gen (H3Africa)	N	Multi	Kenya	12 000
AWI-Gen (H3Africa)	N	Multi	South Africa	12 000
Bandafassi HDSS	Y	West	Senegal	6 500 (1984), 14 500 (2017)
Bandim HDSS	Y	West	Guinea-Bissau	200 000
Birth to Twenty Cohort Study	N	South	South Africa	6 300 (3 300 at enrolment)
Botswana Combination HIV Prevention Project	N	South	Botswana	12 000
Bukavu Observ Study	N	Central	DRC	4 300
Butajira HDSS	Y	East	Ethiopia	
Centre de Recherche en Sante de Kimpese	Y	Central	DRC	63068
Chokwe HDSS	Y	East	Mozambique	
Cross River HDSS	Y	West	Nigeria	

Study name	HDSS site	Region	Location (Africa)	Approximate sample size
Dabat HDSS	Y	East	Ethiopia	
Dande HDSS	Y	Central	Angola	59 635
Dikgale HDSS	Y	South	South Africa	
Dodowa HDSS	Y	West	Ghana	
EMINI	N	East	Tanzania	18 000
Farafenni HDSS	Y	West	The Gambia	16 000 (1981), 58 000
General Population Cohort	N	East	Uganda	10 000 (1990), 22 000 (2018)
Gilgel Gibe HDSS	Y	East	Ethiopia	
HAALSI - Health and Aging in Africa (nested in Agincourt HDSS)	Y	South	South Africa	5 000
Harar HDSS	Y	East	Ethiopia	
Healthy Lives Malawi, Rural & Urban	Y	South	Malawi	110 000
leDEA - Central Africa	N	Central	Burundi, Cameroon, Congo, DRC, Rwanda	82 000
leDEA – East Africa	N	East	Tanzania, Uganda, Kenya	460 056
leDEA – Southern Africa	N	South	Malawi, Zambia, South Africa, Mozambique, Lesotho, Zimbabwe	640 000
Ifakara	Y	East	Tanzania	170 000
Iganga/Mayuge HDSS	Y	East	Uganda	
IHOPE	N	East	Madagascar	8 310 enrolment; 10 508
Karonga DHSS -is now converted into Healthy Lives	Y	East	Malawi	35 000
Kaya HDSS	Y	West	Burkina Faso	
Kiang West Longitudinal Population Study	Y	West	The Gambia	15 000
Kilifi HDSS - waiting form PI	Y	East	Kenya	
Kilite Awlaelo HDSS	Y	East	Ethiopia	
Kinshasa Integrated data Systems (KIDS)	N	Central	DRC	1 000
Kintampo HDSS	Y	West	Ghana	
Kisesa (Magu HDSS)	Y	East	Tanzania	40 000
Kisumu	Y	East	Kenya	220 000
Kombewa HDSS	Y	East	Kenya	
Kyamulibwa HDSS	Y	East	Uganda	
Manhica HDSS	Y	East	Mozambique	190 000
Manicaland	N	South	Zimbabwe	14 000
Masaka	Y	East	Uganda	20 000
Mbita HDSS	Y	East	Kenya	

Study name	HDSS site	Region	Location (Africa)	Approximate sample size
MHURAM	N	East	Madagascar	78 000
Mildmay Uganda Cohort	N	East	Uganda	5 500
Mlomp HDSS	Y	West	Senegal	6 000 (1985), 8 400 (2018)
Nahuche HDSS	Y	West	Nigeria	
Nairobi	Y	East	Kenya	71 000
Nanoro HDSS	Y	West	Burkina Faso	
Navrongo HDSS	Y	West	Ghana	
NCD-Intervention Programme	N	East	Mauritius	6 300
Niakhar HDSS	Y	West	Senegal	24 000 (1984), 49 000 (2018)
Nigeria Cardiometabolic Study	N	West	Nigeria	100 000
Nouna HDSS	Y	West	Burkina Faso	
Ouagadougou HDSS	Y	West	Burkina Faso	
PRECISE Network	N	Multi	Kenya	12 000
PRECISE Network	N	Multi	Mozambique	12 000
PRECISE Network	N	Multi	The Gambia	12 000
PROGRESS/PREPARE	N	East	Uganda	aim 35 000
PURE-Prospective Urban and Rural Epidemiological Study	N	Multi	South Africa	
PURE-Prospective Urban and Rural Epidemiological Study	N	Multi	Tanzania	
PURE-Prospective Urban and Rural Epidemiological Study	N	Multi	Zimbabwe	
Rakai	Y	East	Uganda	12 000
RCCS - Rakai Community Cohort Study	N	East	Uganda	12 000
Sapone HDSS	Y	West	Burkina Faso	
SAPRIN	Y	South	South Africa	38,000-350,000
Seychelles Child Development Study	N	West	Seychelles	549
SIBS-Genomics	N	West	Nigeria, Ghana	8 000
SIREN	N	West	Nigeria, Ghana	6 000
Study on Global Ageing and Adult Health (SAGE)	N	Multi	Ghana	4500 and 5000
Study on Global Ageing and Adult Health (SAGE)	N	Multi	South Africa	4 500 and 5 000
Taabo HDSS	Y	West	Cote d'Ivoire	
uMkhanyakude/AHRI	Y	South	South Africa	168 000
Vukuzazi cohort	N	South	South Africa	18 000
Young Lives	N	East	Ethiopia	2 000

A2.2. Existing Multinational Cohort Networks in Africa

Network name	Region	Description
ALPHA Network	Eastern and Southern Africa	Population-based HIV surveillance in eastern and southern Africa, coordinated by the London School of Hygiene & Tropical Medicine
ANDLA Network	Eastern and Southern Africa	African Non-Communicable Disease Longitudinal data Alliance. Understanding non-communicable diseases (NCD) and the role of infection in Africa, Partnership of African institutions, NMIR
leDEA Network	Pan-African	Epidemiology Databases to Evaluate AIDS, International Consortium. (1) Central Africa programme led by Albert Einstein College of Medicine/City University of New York, USA; (2) East Africa region led by Indiana University School of Medicine, USA; (3) West Africa region led by Institute of Public Health, Epidemiology & Development (ISPED), France (4) Southern Africa region led by University of Berne, Switzerland and University of Cape Town, South Africa.
INDEPTH Network	Multiple partners in the global South, pan-African	Longitudinal research based on Health and Demographic Surveillance System (HDSS) field sites in low- and middle-income countries (LMICs), coordinating/resource centre in Ghana
PRECISE Network	East, West and Southern Africa	Pregnancy cohort of 12,000 women across East, West and Southern Africa, led by King's College London (UK)
PURE Network	Southern Africa represented	Prospective Urban Rural Epidemiology (PURE) study is a large-scale epidemiological study across 17 low-, middle-, and high-income countries around the world; investigating societal influences on human lifestyle behaviours, cardiovascular risk factors, and incidence of chronic noncommunicable diseases. Led by Population Health Research Institute, Hamilton Health Sciences and McMaster University, Hamilton, Canada

Appendix 3: Possible formative and pilot projects

Theme	Ideas	Outputs
Community engagement	Build a generic framework for community engagement. Pilot using a well-defined qualitative evaluation protocol to assess efficacy, acceptability and relevance of the framework.	<ol style="list-style-type: none"> 1. Generic framework (guidelines) for community engagement for LPS. 2. Scientifically sound evaluation results assessing efficacy, acceptability and relevance of community engagement protocol.
Informed consent	Build a generic participant information document for LPS using appropriate and accessible language and concepts, and including key elements for different aspects of the research. Build an informed consent template with key questions pertaining to the different consents being requested. Build a structure for capturing consent details for each participant including mechanisms for when consent is revoked i.e. putting in underlying structures to be able to move to dynamic consent. Pilot using a well-defined qualitative evaluation protocol to assess acceptability, accessibility (understanding) and relevance of the processes.	<ol style="list-style-type: none"> 1. Generic participant information and informed consent templates for LPS, that can be used to guide the participant information and informed consent processes in new studies going forward. 2. Data structure for ensuring appropriate capture of details of individual consent 2. Scientifically sound evaluation results showing acceptability, accessibility and effectiveness of participant information as well as informed consent protocol.
Equity	Build a framework for implementing equity in LPS, defining stakeholders and guidelines to ensure their representation. Define metrics and processes for ongoing monitoring and evaluation of equity in LPS over time. Pilot the framework and validate the M&E metrics	<ol style="list-style-type: none"> 1. A generic framework for ensuring equitable practices in LPS 2. A validated M&E framework for ongoing assessment of equity during LPS
Data Governance	Scoping exercise to identify and document relevant National legislation for data and biospecimen sharing; including Privacy acts, Health Acts, Children's Acts etc	Guidelines for ensuring legal compliance in ethics and data governance practices
Data Governance	Build a generic data governance workflow with definitions, principles of data access, monitoring framework with key metrics/checklists to assess transparent, equitable and inclusive data sharing. Pilot monitoring framework.	Data governance framework for LPS, Monitoring framework for ongoing assessment of data governance.
Stakeholder engagement	Stakeholder mapping process, and structured interviews with key decision-makers from represented groups (micro-individual members of general public, meso-community leaders, local government representatives e.g. health depts, macro – national, regional policy makers) to understand priorities, needs, what is not wanted	Thematic analysis to describe priorities for stakeholder groups and understand their preferences and concerns.
Data Collation and Linkage	Scoping study to document the implementation of routine data platforms in African countries, and to develop recommendations for supporting and/or developing routine data ecosystems in participating countries	Documenting current data systems in use/being implemented in African countries, including stakeholder engagement to understand and recommend how APCC can synergise with and support ongoing efforts.

Data Collation and Linkage	Programmes to build core data linkage hubs to support cohort studies in collaboration with local Health Services, for ongoing upload and collation of routine records	Established and/or nascent health information exchanges with linkage of social, economic and other relevant routine data resources, with virtuous two-way data exchange between health services and the research enterprise. Would need an established governance structure as part of the outcomes. Would need collaborators from service delivery partners, and would need to demonstrate alignment with governmental systems and processes.
Data Collation and Linkage	A series of small, short-term (1yr) grants (e.g. GBP 50 000 – 70 000) to design and pilot automated return data flows from research databases to service delivery platforms (e.g. in Govt Depts of Health). Encourage collaboration between researchers within-country to ensure a single consolidated workflow per government platform without duplication of efforts.	Established mechanism up and running for virtuous circle that returns new data. Could be encouraged as a pre-requisite for all future applications to fund data collection (helped by having within-country collaboration so one mechanism can be used by all future researchers to return data to Govt services).
Core metrics	Review and describe current health priorities for African countries and regions (including what is know about burden of disease – NCDs and infectious diseases, and where are the gaps)	Combination of systematic review and consultations to identify priorities for individual countries, with recommendations for research agendas to address those priorities
Core metrics	Review and describe different WASH metrics in African countries, using an ecological study design to describe relationships with current knowledge about burden of disease. Develop core metrics for WASH for onward use in APCC	Identify current WASH metrics as baseline measurements, Systematic information about gaps and missing data. Data standards for WASH metrics for onward use.
Core metrics	Review and describe existing platforms for meteorological data, pilot linkage to meteorological data, encouraging collaboration with key government and regional meteorological agencies. Identify gaps and develop core metrics for a data standard	Identify existing meteorological data resources, pilot data linkage and develop data standards.
Core metrics	Define data standards for core metrics for collection of longitudinal demographic data, health data - generic metrics for all NCDs and IDs, and specific metrics for each common disease, pathogen and vector data, socio-economic data, environmental data; population meta-data; use a harmonisation approach through comparison of metrics collected by existing African cohorts studies, existing standards, and augment/fill in gaps where necessary. Also build a data 'code book' for all the core metrics. Design instruments for collecting these data where this will be done in the field directly with the participant	Data standards for core metrics for LPS. Data structure/schema for core metrics Data coding for core metrics Data collection instrument (can be split out into multiple small pilots each taking on one domain – may result in better products)
Core metrics	Develop and pilot methods and tools for collecting accurate cause of death data, develop a data standard for cause of death data. Measure acceptability of methods through community engagement.	A culturally acceptable tool for collecting standardised cause of death data.
Capacity development -skills	Ensuring opportunities exist for appropriate qualitative and data analysis/management skills in cohort researchers, through scoping exercises (cataloguing what is out there already), identifying gaps, and building	Packaged, reviewed and approved training modules that can be rolled out, re-used, possibly provided online (review, piloting and robust external evaluation should all be part of the funded activities)

	workshops, degree modules, online training to fill those gaps (especially health informatics, coding/data science and statistics)	
Study Design	Undertake a scoping/review process to inform the design of a cohort meta-data standard, and to build a specification for core indicators to be displayed for each cohort on a dashboard. Pilot the standard and the dashboard specification.	A cohort meta-data standard that is harmonised to other existing standards but also brings in additional and/or African-specific metrics as needed. A validated specification for an APCC online dashboard showing key indicators for all participating sites in a real-time, online display.
Study Design	Propose, pilot and test an appropriate study design for APCC cohorts, including a Consortium-wide population cohort selection framework; a sampling framework for the cohort studies to ensure diversity; guidelines for selection criteria that include clear and detailed guidelines about participant selection criteria, sub-cohort criteria, standardised checks for selection bias, core health metric standards, genomic heterogeneity, cohort size (with an instrument for calculating minimum cohort size/statistical power).	A definite and unequivocal specification of study design and criteria for cohorts belonging to APCC
Sample collection	Establish standards for sample collection, describing what samples should be collected routinely and procedures for collecting, storing and processing those samples. Pilot the procedures at a diverse selection of sites.	A set of well-defined, detailed standard operating procedures for standardised sample collection across cohorts, with piloting/validation at multiple sites.

Appendix 4: Glossary and Abbreviations

Abbreviations

AAS	African Academy of Sciences
AESA	Alliance for Accelerating Excellence in Science in Africa
Africa CDC	Africa Centre for Disease Control
AMR	Antimicrobial resistance
APCC	African Population Cohorts Consortium
ASP	African Science, Technology and Innovation Priorities
AUDA-NEPAD	African Union Development Programme
BHP	Botswana Harvard AIDS Institute Partnership
BMGF	Bill & Melinda Gates Foundation
CHESS	Comprehensive Health and Epidemiological Surveillance System
DELTAS	Developing Excellence in Leadership, Training and Science
ESRC	Economic & Social Research Council
FHIR	Fast Healthcare Interoperability Resources
GIS	Geographical Information Systems
GWAS	Genome-wide association study
H3Africa	Human Heredity and Health in Africa Consortium
HDSS	Health and Demographics Surveillance System/Site
ICER	International Centres for Excellence in Research
ICT	Information and Communications Technology
leDEA	The International Epidemiology Databases to Evaluate AIDS (leDEA) network
IHCC	International 100K+ Cohorts Consortium
ISAB	Independent Scientific Advisory Board
LPS	Longitudinal population study
LSHTM	London School of Hygiene and Tropical Medicine
MoU	Memorandum of Understanding
MLTCs	Multiple Long-term Conditions
MRC	Medical Research Council
NCD	Noncommunicable disease
NIH	National Institutes of Health
PHDC	(Western Cape Government Health) Provincial Health Data Centre
PLHIV	People living with HIV
SAPRIN	South African Population Research Infrastructure Network
SDG	Sustainable Development Goal
TB	Tuberculosis
WASH	Water, sanitation and hygiene

WGS	Whole genome sequencing
WHO	World Health Organisation

Glossary

Data anonymisation	Anonymisation entails removing any means of identifying the individual from whom the data arose, and also removing any data points or sequence of data points that could potentially facilitate relinking data to an individual. Anonymised data cannot be relinked to the originator of those data. Genomic data cannot be anonymised. Because of the extensive unique combinations of genetic variation in each individual, even given less than 100 variants it is possible to re-identify the individual from whom those genetic data were generated.
Cohort study	A population-based cohort study is a specific category of epidemiological study in which a defined population is followed up over time (longitudinally) to observe exposures and risk factors, and how they are associated with different outcomes. The incidence of different outcomes of interest is measured by identifying new cases occurring in a specified period of time. A relative risk ratio, or incidence rate ratio, is used to compare incidence in those who are exposed compared to those who are not exposed to exposures of interest.
Data de-identification	De-identification entails removing data variables that link data directly to an individual, such as include names, identity or hospital numbers. Granular de-identified data might still be re-linked to an individual, for example a series of hospital visits on particular dates could be used to re-identify an individual by referring to hospital registers. A birthweight given to four decimal places, with a date and facility of birth could be used to re-identify an individual from a maternity ward register.
Data perturbation	Perturbation refers to undisclosed modifications to data to prevent re-identification of individuals from whom the data originate. Whilst algorithms may be developed for more complex data perturbation approaches, examples of simple methods include adding a constant undisclosed integer to all dates to retain epidemiological interpretation whilst preventing re-identification; providing age in weeks (perinatal), months (infants) or years (children onwards) to avoid disclosing date of birth; round up numerical values with many decimal places.
Demographic dividend	The growth in an economy that is the result of a change in the age structure of a country's population
Epidemiological transition	The theory of epidemiologic transition focuses on the complex change in patterns of health and disease <i>and</i> on the interactions between these patterns and their demographic, economic and sociologic determinants and consequences ⁹⁹ . Epidemiological transition often reflects the intermediate state of change from a major burden of infectious disease to a major burden of NCDs, presenting as a dual burden of infectious and NCDs.
Haploblock	A section of DNA on a chromosome in which several neighbouring, tightly linked genetic variants lie on the same section of DNA and are inherited together.
Longitudinal population study	A longitudinal population study (LPS) is an observation study of a population that measures a diverse set of determinants and tracks health, wellbeing and social outcomes with repeated observations of the same metrics over time. LPS include cohorts, panel surveys and biobank.
Microbiomes	The combined genetic material of all the microorganisms found in a particular context – the environment, parts of the body etc
Minimally invasive autopsy	A postmortem examination using non-invasive techniques, causing minimal disruption to major body parts
Multimorbidity/MLTC	The presence of two or more health conditions in the same individual and is an increasingly common occurrence which places patients at higher risk of adverse drug interactions and related poor adherence to medication, creates a greater need for integrated and holistic health care, and places individuals under much greater personal, social and financial strain due to multiple competing and compounding health requirements. The WHO recognises the burden of multimorbidity / Multiple Long-Term Conditions (MLTC) and the requirement for health systems-based approaches to address this burden ¹⁰⁰ .

Pharmacogenes	A gene that encodes proteins that interact with and process pharmaceutical compounds
Precision medicine	The application of medical care that is specifically tailored to best suit the individual, especially based on their genetic profile
Precision public health	The application of medical care and health policy that is tailored to a specific population based on their characteristics and environment, and including the aggregated genetic profile of that whole population
Social epidemiology	A branch of epidemiology that focuses particularly on the effects of social-structural factors on states of health and wellbeing
Verbal autopsy	A method of collecting information of the cause of death through discussions and interviews with caregivers and relatives of the deceased, and their medical practitioners.

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

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

The Wellcome Trust is a charity registered in England and Wales, no. 210183.
Its sole trustee is The Wellcome Trust Limited, a company registered in England and Wales, no. 2711000
(whose registered office is at 215 Euston Road, London NW1 2BE, UK).