South African Population Research Infrastructure Network (SAPRIN):
A National Research Infrastructure of Health and Demographic Surveillance System (HDSS) Nodes

SAMRC Ethics Reference: EC010-3/2021

Version 3.2

20th January 2024

Ethical Declaration by Principal Investigator:

I, **Abraham Jacobus ('Kobus') Herbst,** have read the Department of Health: Ethics in health research: principles, processes and structures, second edition, 2015, the Guidelines for Good Practice in the Conduct of Clinical Trials with Human Participants in South Africa, Second Edition, 2006, Department of Health, Pretoria, South Africa (where applicable), and the Declaration of Helsinki (2013) and have prepared this proposal with due cognisance of its content. Furthermore, I will adhere to the principles expressed when conducting this proposed research project.

Sponsor/Funder Monitoring

In the event of co-funding from an external entity distinct from the South African government's Department of Science and Innovation (DSI), the external funder shall have the prerogative to undertake monitoring duties to safeguard data integrity and ensure the welfare of research participants, that may include but not limited to on-site visits to HDSS nodal facilities or remotely through digital platforms or other communications such as telephone calls or written correspondence.

Protocol Team (Role and Contact details)			
South African Population Research Infrastructure Network (SAPRIN), South African Medical Research Council, 491 Peter Mokaba Ridge Road, Overport, Durban, 4091, South Africa			
Name	**Roles and Responsibilities**	**Telephone**	**Email**
Dr Abraham Jacobus Herbst	**Principal Investigator (PI)**: Director of SAPRIN oversee protocol development; data collection, data management, public engagement, financial and HR management systems; co-ordinate analysis, lead the write-up, and ensure the HDSS Nodal protocols are consistent with SAPRIN Core protocol	+27312034727	Kobus.Herbst@mrc.ac.za
Prof Mark Collinson	**Co-PI: Co-Director of SAPRIN**; co-leads protocol development; data collection, data management, ethics and public engagement, financial and HR management systems; overseeing monthly data collection activities, co-ordinate analysis, lead the write-up of report	+2712 3398572	Mark.Collinson@mrc.ac.za
Tinofa Mutevedzi	**Co-Investigator: SAPRIN Research Operations Manager;** assist with protocol development; implementation, harmonisation, quality assurance and performance, ethics, budget development and management, progress reports, analysis, write-up of report	+2712 339 8544	Tinofa.Mutevedzi@mrc.ac.za
Dr Linda Maoyi	**Co-Investigator: SAPRIN Data Scientist**; assist with protocol development; data management systems, data quality assurance, production of datasets, analysis, write-up of report	+27 31 203 4722	Linda.Maoyi@mrc.ac.za
Scientific leadership in the HDSS Nodes:			
Prof Stephen Tollman	**Nodal Director: MRC/Wits Agincourt HDSS Node**; ensure HDSS Nodal protocol is	+27-11 717 2085	stephen.tollman@wits.ac.za

	consistent with SAPRIN Core, and overall leadership and scientific implementation of the Agincourt HDSS adapted version of the protocol		
Prof Eric Maimela	**Nodal Director: DIMAMO HDSS Node**; ensure HDSS Nodal protocol is consistent with SAPRIN Core, and overall leadership and scientific implementation of the DIMAMO HDSS adapted version of the protocol	+27-15 268 4137	eric.maimela@ul.ac.za
Prof Collins Iwuji	**Nodal Director: AHRI HDSS Node**; ensure HDSS Nodal protocol is consistent with SAPRIN Core, and overall leadership and scientific implementation of the AHRI HDSS adapted version of the protocol	+27-74 233 3657	collins.iwuji@ahri.org
Prof David Everatt	**Nodal Director: GRT-INSPIRED HDSS Node**; ensure HDSS Nodal protocol is consistent with SAPRIN Core, and overall leadership and scientific implementation of the GRT-INSPIRED HDSS adapted version of the protocol	+27-11 717 3918	david.everatt@wits.ac.za
Dr Hannah Hussey	**Nodal Director: C-SHARP HDSS Node**; ensure HDSS Nodal protocol is consistent with SAPRIN Core, and overall leadership and scientific implementation of the C-SHARP HDSS adapted version of the protocol	+27-71 211 2353	hannah.hussey@westerncape.gov.za
Dr Nompumelelo Nzimande-Mbele	**Nodal Director: USINGA HDSS Node**; ensure HDSS Nodal protocol is consistent with SAPRIN Core, and overall leadership and scientific implementation of the USINGA HDSS adapted version of the protocol		nzimanden@ukzn.ac.za

Table of contents

List of abbreviations

Agincourt HDSS	Agincourt Health and socio-Demographic Surveillance System Node, University of the Witwatersrand and Medical Research Council
AHRI HDSS	Africa Health Research Institute Health and Demographic Surveillance System Node
AIDS	Acquired Immune Deficiency Syndrome
ART	Antiretroviral therapy
BMI	Body mass index
BP	Blood pressure
BREC	Biomedical Research Ethics Committee
CAB	Community Advisory Board
CAG	Community Advisory Group
CAPI	Computer-assisted Participant Interview
CASI	Computer-assisted Self Interview
cBED	BED IgG-Capture Enzyme Immunoassay
CD4	Cluster of differentiation 4
C-SHARP HDSS	Cape Town Systematic Healthcare Action Research Project, Health and Demographic Surveillance System Node
DBS	Dried blood spots
DIMAMO HDSS	DIMAMO Population Health Research Centre Node, University of Limpopo
DoH	Department of Health, South Africa
HDSA	Health and Demographic Surveillance Area
DST	Department of Science and Technology, South Africa
ELISA	Enzyme linked immunosorbent assay
EXTID	External individual identifier
FIO	Facility Information Officer
GCP	Good Clinical Practice
GRT-INSPIRED HDSS	Gauteng Research Triangle Initiative for the Study of Population, Infrastructure and Regional Economic Development, Health and Demographic Surveillance System Node
HbA_{1c}	Glycosylated haemoglobin
HCT	HIV counselling and testing
HDSS	Health and Demographic Surveillance System
HDSS-NRI	Health and Demographic Surveillance System – National Research Infrastructure

HDSS Node	An HDSS that forms a Node in the HDSS-National Research Infrastructure
HIV	Human immunodeficiency virus
HREC	Human Research Ethics Committee
ICD-10	International Classification of Diseases 10th revision
ICH	International Committee on Harmonisation
ICMJE	International Committee of Medical Journal Editors
INTID	Internal identifier
IRB	Institutional Review Board
ISAB	International Scientific Advisory Board
KZN	KwaZulu-Natal
LIMS	Laboratory Information Management System
MAB	Medical Advisory Board
MRC	South African Medical Research Council
NHLS	National Health Laboratory Service
NRI	National Research Infrastructure
PCR	Polymerase chain reaction
PHC	Primary health care
RI	Research Infrastructure
RNA	Ribonucleic acid
SAE	Serious adverse event
SAPRIN	South African Population Research Infrastructure Network
SARIR	South African Research Infrastructure Roadmap
SMS	Short Message Service
SOP	Standard Operating Procedure
TB	Tuberculosis
TYIP	Ten Year Innovation Plan (2008-2018)
UKZN	University of KwaZulu-Natal
UL	University of Limpopo
USINGA HDSS	Umlazi Surveillance Initiative to Nurture Grassroots Action, Health and Demographic Surveillance System Node
WHO	World Health Organization
Wits	University of the Witwatersrand

Key Words

Surveillance, birth, death, migration, health, HIV, service record linkage, socio-economic

Preamble

The South African Population Research Infrastructure Network (SAPRIN) of Health and Demographic Surveillance System (HDSS) Nodes will provide a data and research infrastructure to support health, social and economic development in some of the poorest South African populations. This protocol provides a standardized and harmonised summary of the network platform, with specific details for each HDSS node to be determined through contextualised protocols and local Institutional Review Board (IRB) approvals. To achieve this, this protocol document aims to establish sufficient harmony of approach and methods and a rigorous ethical framework while acknowledging the contrasting contexts in which sister HDSSs work. Detailed Standard Operating Procedures (SOPs) will be produced to standardise approaches and high-quality data production.

Protocol summary

Introduction: The South African Population Research Infrastructure Network (SAPRIN) of Health and Demographic Surveillance System (HDSS) Nodes will support major improvements in health, social and economic wellbeing in impoverished yet rapidly evolving populations. An important purpose is to provide an integral connection between the new evidence from the research platform and government ministries, both line-function ministries like the Departments of Health, Social Development, Home Affairs, and Basic Education, and cross-cutting ministries like the Presidency (Dept. of Planning Monitoring and Education) and Statistics South Africa. The overall aim is to reduce research costs for the government while making poorer South Africans healthier and socially and economically better-off. Measurable aspects of this goal will be the monitoring and evaluating of progress towards milestones in the DSI's Ten-Year Innovation Plan, the National Development Plan, the United Nation's Sustainable Development Goals Framework, and WHO's Health and Development Indicators in South Africa.

Method: A Health and Demographic Surveillance System (HDSS) is a standardised, field-based information system and research platform, with prospective data routinely collected from entire populations at both individual and household levels in impoverished and developmentally-constrained communities, both rural and urban. Individual and household indicators that will be routinely collected and assessed include vital events, i.e., births and deaths (by cause), residence status and migration, household dynamics, socio-economic status, disease risk and monitoring, employment, education status and social protection. The individual and household registration systems will be complemented by public sector records on inter alia health system utilization, school attendance and social grant receipt to enable population-level research on access to, and failure to access, publicly-provided services.

Community engagement: Communities in which HDSS nodes are embedded are crucial partners. Care is taken to interface with community structures in multiple ways, including discussion on study objectives and methods prior to commencing fieldwork, feeding back research findings, supporting the use of research data, and active engagement with community advisory boards/ groups.

Geographic Scope: The project will first integrate and standardise the three existing HDSS nodes in South Africa (namely, MRC/Wits Agincourt HDSS in Bushbuckridge, Mpumalanga, established in 1992 with a current population of 105,461; U. Limpopo DIMAMO HDSS in Capricorn, Limpopo, established in 1996, with a current population of 99,741; and the Africa Health Research Institute's Population

Intervention Platform in uMkhanyakude, KwaZulu-Natal, established in 2000, with a population of 153,410).

The network of HDSS nodes has been expanded to include new nodes in urban Gauteng (urban, Gauteng Research Triangle Initiative for the Study of Population, Infrastructure and Regional Economic Development [GRT-INSPIRED], Hillbrow, Johannesburg, and Atteridgeville and Melusi, Tshwane, Gauteng. urban Western Cape (Cape Town Systematic Healthcare Action Research Project [C-SHARP], Nomzamo and Bishop Lavis, Cape Town, Western Cape), and Umlazi Surveillance Initiative to Nurture Grassroots Action (USINGA) in eThekwini (urban)..

This expanded platform, representing over 1% of the national population (62 million), and covers a more inclusive spectrum of impoverished yet dynamically developing sub-populations and enables a new and unprecedented understanding of the dynamic bi-directional migration flows that link poor, rural communities with urban centres. The expanded scale of the full network will enable a more complete understanding of rare phenomena, such as maternal mortality.

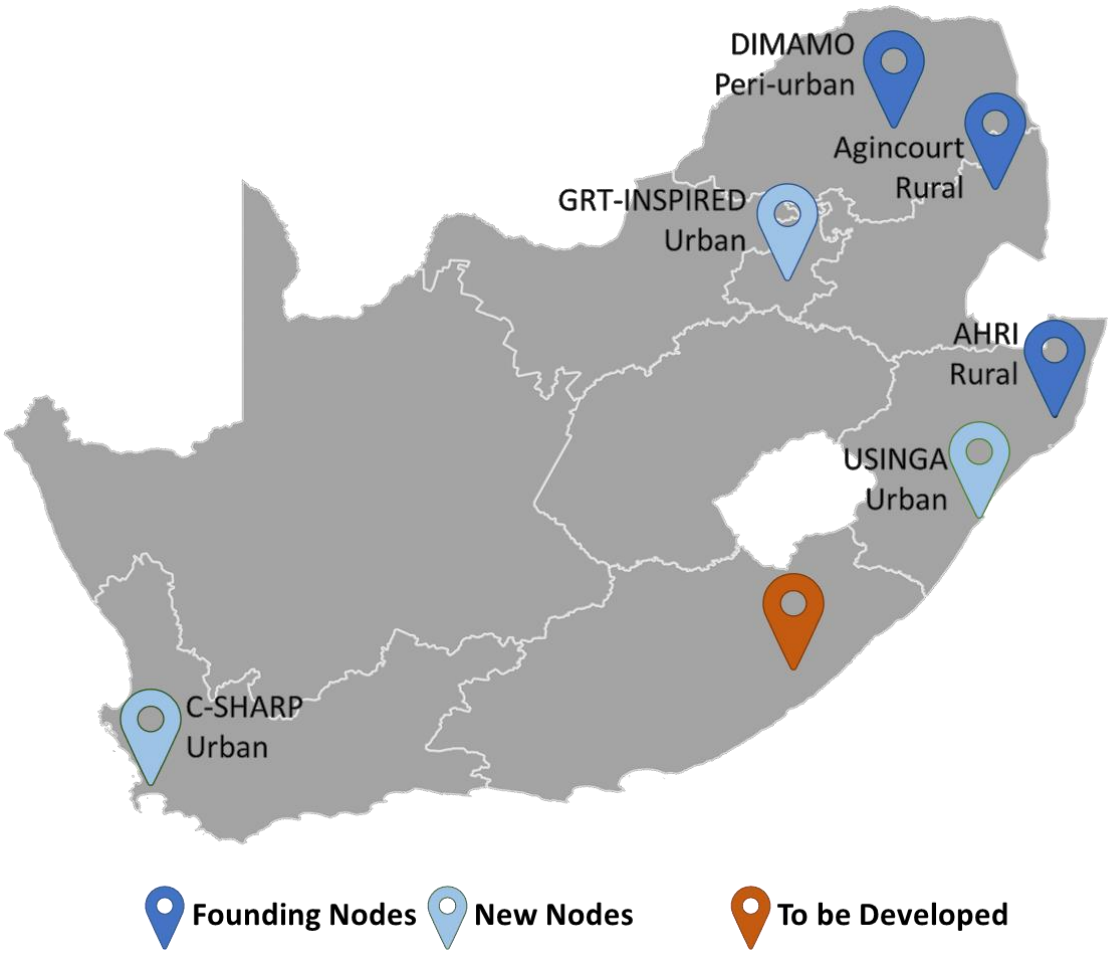

Figure: The Network has been expanded from three (3) to six (6) Health and Demographic Surveillance System (HDSS) Nodes, with a seventh rural node still to be developed either in FS, EC, NC or NW.

Expected outcomes and impacts:
The expected outcomes and impacts of the SAPRIN are:

- Provide a versatile inter-disciplinary **research platform,**
- Produce ongoing research **data** for policy development and evaluation,
- Produce routinely updated **thematic reports** on population health and socio-economic wellbeing,
- To establish clear links with the **national statistical system,**
- Effectively engage **community structures, and**
- Provide a resource for **post-graduate** research training and ongoing career development.

1. **A versatile inter-disciplinary research platform**

   Based on experience from nearly 60 combined years of running HDSS Centres, we will implement the South African Population National Research Infrastructure Network (SAPRIN), which will be an extensive, versatile and interdisciplinary platform for researchers from universities, science councils, regional and international collaborations. This platform can support the breadth of study designs and types based on observational, interventional or policy evaluation designs.

   Knowledge will be developed in nodal and multi-nodal studies to provide evidence on rapidly changing health, social wellbeing, socio-economic profiles and their determinants. It will provide a scientific evidence-base for cost evaluation, policy-making and targeting intervention programmes, thereby improving the accuracy, efficiency and effectiveness of pro-poor health and wellbeing interventions. The overall goal is to make poorer South Africans healthier and better-off while reducing costs to the government and improving the overall return on public sector investments.

2. **Research data for policy development**

   SAPRIN will provide ongoing, up-to-date, longitudinal data that is representative of South Africa's fast-changing, poorer rural and urban communities for research and policy evaluation, interpretation and calibration of national data. The outcome will be accessible, dynamic, representative, and timely data. Key areas of data collection are population dynamics, health dynamics, social well-being and economic well-being.

3. **Routinely updated thematic reports on population health and socio-economic wellbeing.**

   SAPRIN will produce routine thematic reports to support knowledge translation using data from all the HDSS Nodes on core thematic areas, including Population Dynamics Health Dynamics, Social well-being and Economic well-being. Once per year, a full set of health and socio-economic well-being indicators will be produced. SAPRIN's web-site will provide access to back-issues of thematic reports.

4. **Links with the national statistical system**

As a component of the national statistical system, the longitudinal data from SAPRIN will calibrate and validate national datasets and serve as an early warning system for dire population issues.

5. **Engaged community structures**

   The communities within which each HDSS is embedded are crucial partners and will utilise the data developed by the HDSS nodes. There will be coordinated engagement with communities to enable two- way learning between researchers and community members, maintain community willingness to participate in research, ensure the dignity of research participants, enable the research site community and service providers to have access to and make effective use of research results, and enable effective and appropriate community participation in guiding and informing research processes at the HDSS nodes. At the national level, there will be a network-based Community Advisory structure with representatives from each nodal Community Advisory Board, with the same set of objectives that empower and dignify communities involved in the research.

6. **Resource for post-graduate research training**

   SAPRIN will host interdisciplinary post-graduate students from multiple South African universities to expand human capacity for conducting advanced research that takes advantage of longitudinal platforms and is effectively linked with national, regional and international networks. It will provide excellent opportunities to embed post-graduate training and support research career development and leadership in a broad range of interlinked disciplines.

Research Agenda:

Routine, ongoing, longitudinal data-collection will occur in key domains of health, population, and social and economic well-being in South Africa's vulnerable populations. We define vulnerable populations as a disadvantaged sub-segment of the national population requiring augmented protections from greater risk for poor health status and healthcare access, experiencing significant disparities in life expectancy, access to and use of healthcare services, morbidity, and mortality[1].

We define poverty according to the South African Multidimensional Poverty Index (SAMPI) that was developed by Statistics South Africa based on the global Multidimensional Poverty Index (MPI) to capture the severe deprivations that each person or household faces with respect to economic activity, education, health and standard of living[2].

1. <u>Population Health</u>
   By collecting information on all ages, from birth to oldest, we contribute to the vital understanding of adolescence (a key phase of development and risk period with an impact on later health outcomes as well as the wellbeing of offspring) and older populations (these are increasing rapidly with impacts on health care, social and welfare services and economic development).

a. *Births and Deaths (by cause)*

A core function of the HDSS platform is to maintain an up-to-date register of all pregnancies and their outcomes and a register of all deaths and their causes (main and contributing) for the whole population under long-term health and socio-demographic surveillance. This will enable the close monitoring of a range of key health and development indicators, including life expectancy at birth and at later ages, adult mortality, under-five, infant and neonatal mortality rates, still-birth rates, maternal mortality ratio, cause-specific mortality rates, adolescent and total fertility rates, and completeness of birth and death registration. It also supports the computation of precise individual exposure periods, which underpins the capability of the platform to support advanced studies in a range of fields.

b. *Disease and risk factor monitoring (HIV, Hypertension and Body Mass Index)*

Individual level monitoring of health events and risk factor prevalence will be undertaken for key tracer conditions, namely HIV-status through a dried blood spot (DBS) and a rapid test, hypertension status through blood pressure measurement, and body-mass index status through measuring individual heights and weights. Together with mortality data, these will provide ongoing information on the epidemiological transition and associated changes in burdens of disease over time at stages along the life course.

c. *Health care utilisation*

Individuals' use of clinics and hospitals will be recorded by observing health service utilisation patterns and linking these data confidentially with population data. This will support the ongoing monitoring of sub-groups in the population that are able to access health services and those for which barriers occur.

d. *Childhood immunisation*

Immunisation remains the most cost-effective population health intervention. Ongoing updates of coverage rates by vaccine-type will occur for each vaccine in the national schedule. Coupled with accurate knowledge of population trends and utilisation of health services, this will provide a valuable basis for monitoring and evaluating universal health coverage, a key sustainable development goal.

e. *Food Security*

Food security provides a vital backdrop for health and social and economic wellbeing. The converse also holds that interventions aimed at uplifting health social and economic well-being can be severely compromised in the context of food insecurity. Therefore, monitoring food security in poorer populations will be a focus of the research infrastructure.

2. <u>Social well-being</u>

a. *Household dynamics*

Resilience and vulnerabilities of households keenly affect their health, social and economic wellbeing, and household structures are undergoing transition. This core social unit will be carefully tracked over time through monitoring births and deaths (as described above) and residence and migration (as described below). Households are defined as people living and eating together, plus non-resident household members that have temporarily migrated away from the household but retain close ties.

b. *Residence Status and migration*

Migration and residence status are highly relevant for personal and household health, social and economic wellbeing and are frequently used as a strategy to access unequally distributed opportunities. There are age, sex and socio-economic correlates of migration, but little is known about the extent of urban and rural linkages, what these mean for the social and economic wellbeing of sending households, and the health and social consequences for migrants and their households.

c. *Education status and schooling*

Individual interviews will be the basis for ongoing monitoring of the education status of the population. The aim is to develop confidential links between school and population data to establish measures of school and individual performance and their determinants. Indicators recommended by the National Planning Commission that can be monitored by the research infrastructure include 'the percentage of children that have two years of pre-school education'; 'the percentage of schools and learners that achieve 50% and above in literacy, mathematics and science, in grades 3, 6 and 9'; and 'the percentage of schools and learners that complete all 12 years of schooling'.

3. <u>Economic well-being</u>

a. *Socio-economic status*

Household socio-economic status will be monitored using an index derived from a list of household assets, dwelling size (number of rooms), dwelling construction materials, water, sanitation and energy sources. These measures are triangulated with similar measures in STATSSA household surveys and the index is subject to ongoing research and validation{Kabudula, 2016 #15} These variables are sensitive to the socio-economic status of a household and indicate what is changing over time and what is driving the changes. Crucially, this data will help to show how socio-economic status mediates the successes and failures of health, social and economic wellbeing interventions.

b. *Labour status*

Through individual interviews, labour force participation in formal and informal sector work and/or lack of employment will be tracked over time for all people aged 15 years and over. This will enable ongoing monitoring of unemployment trends, studying the determinants of employment and keeping track of how labour markets are changing for people, including youth, in poorer sections of society. This, in turn, can be key to addressing levels of inequity in the country.

c. Social protection

Individual interviews will be the basis for ongoing monitoring of government grant uptake. The aim is to develop links between the data systems of the Department of Social Development and the longitudinal HDSS research infrastructure to establish a unique social protection monitoring and evaluation platform. This will support ongoing research in coverage and quality of social protection interventions, such as social grants (pensions, child support grants and others), as well as identifying which eligible groups are failing to access protection and evaluating the impacts of these interventions on human health, social and economic well-being.

1. Background information and rationale

1.1. Background information

On 4 Oct 2016, the Minister of Science and Technology, Minister Naledi Pandor announced that the South African Research Infrastructure Roadmap (SARIR) had been developed to facilitate a research infrastructure investment programme[3]. SARIR is intended to provide a strategic, rational, medium- to long-term framework for planning, implementing, monitoring and evaluating the provision of research infrastructures (RIs) necessary for a competitive and sustainable national system of innovation. SARIR also provides a basis for discussion concerning financing future infrastructure for research in South Africa and participating in joint international RIs. The SARIR initiative is a high-level strategic and systemic intervention to provide research infrastructure across the entire public research system, building on existing capabilities and strengths and drawing on future needs. The overall objective of SARIR is to provide a strategic, rational, medium- to long-term framework for planning, implementing, monitoring and evaluating the provision of research infrastructures necessary for a competitive and sustainable national system of innovation (NSI). In the above context, the concept of "research infrastructure" includes facilities, resources and services used by the scientific community across all disciplines for conducting cutting-edge research for the generation, exchange and preservation of knowledge. It includes major facilities, equipment or sets of instruments, collaborative networks and knowledge-containing resources such as collections, archives, databanks and biobanks.

Research in the human and social sciences is essential for social, economic and cultural development and transformation in South Africa[4, 5]. Yet a number of recent studies, such as the Consensus Study on the State of Humanities in South Africa: status, prospects and strategies [6] and the Charter for Humanities and Social Sciences [7] have highlighted the diminishing role of these disciplines in academia, and emphasised that they should be enabled to play a stronger role in the development of society, the economy and intellectual life in South Africa. The need for an elevated role of the human and social sciences in the country's development has also been recognised by the DST in its Ten-Year Innovation Plan [8], which highlights human and social dynamics as one of the grand challenges. The objective of this grand challenge is to increase and deepen research to improve scientific understanding and practice in a range of fields while contributing to the development of evidence-based public policy that improves human well-being. The science plan developed for this grand challenge [9] specifically mentions the need for research infrastructure and longitudinal studies to achieve this objective.

In 2013, a jointly funded project was launched between the European Commission (under the Trade, Development and Cooperation Agreement) and the DST for the development of a national research infrastructure roadmap. The process was led by a team of four experts (two from South Africa and two from Europe). A key deliverable of the team was a framework for a national research infrastructure roadmap for South Africa and the development of guidelines, a policy framework and the scope of the roadmap. The framework provided a comprehensive set of principles and a basis for the development of a national research infrastructure roadmap [3]. Through a bottom-up approach, including several multidisciplinary workshops, survey questionnaires and interviews with researchers and research managers from both the public and the private sectors, the expert group produced a SARIR framework

identifying 17 medium to large research infrastructures across six scientific domains: (i) humans and society; (ii) health, biological and food security; (iii) earth and environment; (iv) energy; (v) materials and manufacturing; and (vi) physical sciences and engineering. In the humans and society category, a "South African network of health and demographic surveillance sites" was identified as one of the 17 research infrastructures. The final selection of RIs took place in two steps. The first step involved the development of high-level proposals or concept documents for each of the 17 RIs identified during the initial bottom-up phase. These proposals were assessed to see whether the relevant infrastructures should progress through to the final stage of developing SARIR. Researchers were appointed to act as champions for each RI and to act as facilitators and coordinators to assist all relevant stakeholders and role players in producing the high-level proposals (meta-design reports). The champions produced 17 reports, which were subjected to an evaluation and selection process by a steering committee appointed by the DST. From the 17 meta-design reports submitted, 13 RIs were identified for incorporation into the final SARIR. The next step was to develop full proposals (including addressing the gaps identified in the meta-design reports, such as proposed governance structures and budget, consultation levels with key stakeholders, and the state of readiness and time frame required to establish the infrastructure). The 13 final, detailed proposals were reviewed and evaluated by the Steering Committee for rigour, adherence to the template for the proposals, and overall feasibility and state of readiness. The committee issued its final report to DST in this regard in 2016. Using the Steering Committee's inputs and scores, the DST next needed to assess the readiness for implementation of the 13 RIs, taking into account DST-internal financial and strategic considerations. Seven RIs, including the network of South African HDSS Nodes, now called SAPRIN, the South African Population Research Infrastructure Network, were selected to start implementation in the 2016/17 financial year.

At its commencement, the SAPRIN incorporated three existing Health and Demographic Surveillance System (HDSS) Nodes[10-12], all of which bring a track record of internationally-recognised research, effective collaborations, government support, and community engagement. Drawing on joint experience over the last two decades, we built a common surveillance platform and upgraded the existing HDSSs to this template. Three new urban nodes in Gauteng, the City of Cape Town in Western Cape, and eThekwini in KZN were developed,with a seventh node to be added in either FS, EC, NC, or NW. This would stratgegically span the seven nodes over six province in rural, urban, and peri-urban geographies.
.

1.2. Rationale

Scientifically, the uniqueness lies in having individual and population level data (vis-à-vis sampled data), longitudinal data (vis-à-vis cross-sectional data), adequate representation of poorer parts of the population (vis-à-vis data based on the middle or better-off parts of the population) and triangulated into the national system to calibrate and validate the census and national surveys[13, 14].

SAPRIN will provide a versatile research infrastructure that can support diverse study designs suited to multilevel, interdisciplinary research spanning molecular to individual and population levels. The key strengths of the HDSS nodes making up this platform are their longitudinal perspective and unique capabilities that derive from full population enrolment and comprehensive follow-up within a defined social and geographical setting[10-12]. The population base for SAPRIN will initially cover some 250 000

persons in existing contrasting rural and peri-urban settings in three provinces, thus contributing unique depth of data from the outset – some 20 years and 4.5 million person-years of observation – that is at best, only partially met by existing national datasets (where statistical power and community reach have limitations). Once fully implemented, the NRI coverage will be about 1% of the South African population.

There is no substitute for detailed microdata relating to known individuals, households and communities, updated on a regular basis over periods of time, for understanding the dynamics of health[15], welfare and social change as rapid, at times complex, development occurs[14]. SAPRIN is timely and pertinent both nationally, in the context of the goals articulated in the National Development Plan, and globally, in the priorities outlined in the 2030 Sustainable Development Agenda[14, 16].

The HDSS nodes[10-12] making up this distributed NRI across poor yet heterogeneous communities will support a broad programme of interdisciplinary research relevant to a range of government and development sectors and involving a spectrum of disciplines in the life, natural and social sciences, as well as the humanities. SAPRIN will generate knowledge about how South Africans can achieve greater wellbeing, improve skills development and enhance their socio-economic status, particularly among the poorer, often neglected communities. This will benefit the economy by enabling poorer people to participate more effectively and drive the economy through their work. Furthermore, SAPRIN will enable a highly effective evaluation of the impact of public policy across a range of sectors. Where poorer people are struggling to maximise their contributions to society due to adverse health and socio-economic conditions, then the best scientific methods should be applied to drive improvement in these conditions and bring about wider participation in the economic and social life of the country. This process of enhanced social development – with improving personal and professional productivity across all life stages – will be the key return for the investment made in this NRI.

We present a standardized proforma protocol for SAPRIN at the network level, outlining ethical standards, objectives, design, methodology, and minimum uniform data to be collected across HDSS nodal platforms. Each SAPRIN HDSS node will tailor the protocol and create consent forms for submission to its Research Ethics Committee (REC) of record.

1.3. Study objectives

SAPRIN's population platform has the following objectives that will be pursued by hosting observational, intervention and policy-evaluation research studies:

In population health:

a. To obtain accurate data on population dynamics (births and deaths by cause) to enable the close monitoring of a range of health and development indicators and to compute precise individual exposure periods that support advanced studies in a range of fields; these speak to critical life course stages including childhood, adolescence, adulthood and later life phases.
b. To obtain accurate measures of disease burden and risk factor prevalence in key tracer domains, namely, HIV, hypertension and body-mass index.

c. To obtain accurate measures of individual and population-level access to health services, both directly through interviews and indirectly through record linkage.
d. To monitor vaccine coverage rates through individual interviews and recording information from Road-to-Health Cards of all children aged 6 years and younger.
e. To monitor food security through household interviews.

In social wellbeing:

f. To track household dynamics over time on the whole population of households through careful monitoring of births, deaths, residence status, and in- and out-migrations.
g. To keep track of all individuals' residence status and migration events in the study populations.
h. To monitor education status and education outcomes at the level of individuals, schools, and the links between them.

In economic wellbeing:

i. To monitor socio-economic status, measured by modern assets, fuel used for cooking, heating and lighting, access to water and sanitation, ownership of transport, livestock and quality of housing.
j. To keep track of labour status at the individual level for all people aged 18 years and older.
k. To keep track of the uptake of social protection measures through individual and household interviews and through links with the Department of Social Development databases.

With links to other RIs:

l. To link with other DST-funded research infrastructures in environmental observation, digital languages and genetics/genomics research.

2. Study design

The SAPRIN is based on a network of partnerships between the government, universities, research councils and HDSS Nodes that will each pursue their own research agenda, while accommodating the SAPRIN objectives, which are to:

- Undertake population-based research to provide interdisciplinary evidence on rapidly changing health, wellbeing, and socio-economic profiles and their determinants.
- Determine how to target policies and interventions to address population health issues, poverty, inequality and unemployment.
- Monitor and evaluate the effectiveness of such policies.

HDSS Nodes will provide access to geographically well-defined research populations and facilities suitable to execute the surveillance protocol. Nodes will implement standardised longitudinal data collection protocols comprising of individual, household and service utilisation data consistent with the SAPRIN protocol as a minimum requirement. Where aspects are new to a Node and its study communities, it may be necessary to pilot prior to full introduction.

The eligible population are members of households that are resident within the health and demographic surveillance area (HDSA)[10-12, 17]. These study site boundaries are predefined for each of the HDSS Nodes[10-12]. See Figure 2 for the geographic location of the HDSS nodes.

The longitudinal population platform will have the following components:

a. **Household component** consisting of one in-person interview and two telephonic interviews per year, with a household informant.
b. **Individual component** conducted during one visit (coinciding with the visit to the household they are a resident member of) per year.
c. A **verbal autopsy** interview with the care giver or close relative of everyone who has died;
d. **Linkage of** individuals in the study population to individual-level **service delivery data** obtained from public health, welfare, home affairs and education authorities;

These components will be described in more detail in section 5.

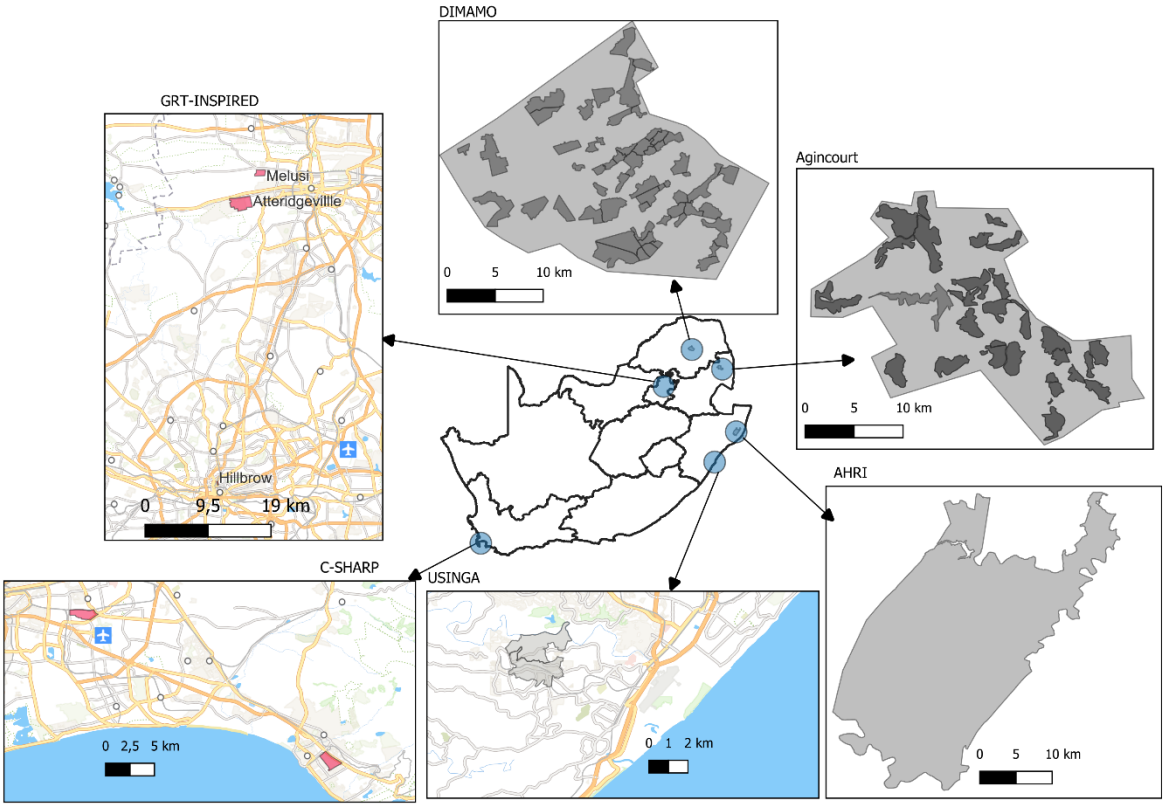

Figure 1: Location of the current 5 nodes of the SAPRIN

3. Selection and withdrawal of participants

3.1. Household component

Households resident in locations within the study area will be eligible for inclusion in the household component of SAPRIN. All individuals identified during an interview by the household proxy informant as a member of the household will be enumerated.

A resident household member is an individual that intends to sleep the majority of the time at the location occupied by the household over a four-month period. Households will include resident and non-resident members. An individual is a non-resident member if they have close ties to the household but do not physically reside with the household most of the time. They can also be called temporary migrants, and they are enumerated within the household list. Because household membership is not tied to physical residency, an individual may be a member of more than one household.

Households will be free to refuse to participate in the surveillance and to withdraw from the study at any time, without any impact on routine health care or other services to which they are entitled.

3.2. Individual component

All individuals aged 15 years and older who are resident members of households in the study population will be eligible for inclusion into the individual component of SAPRIN (see section 5.1.2).

Individuals are free to refuse to participate in and to withdraw from the individual component without any impact on routine health care or other services to which they are entitled. If a household refuses to participate in the household component, its members will not be invited to participate in the individual component during that same household visit.

3.3. Verbal autopsy

All deaths among resident or non-resident members of households will be eligible for verbal autopsy, irrespective of where the death occurs, as explained in section 4.3 below.

3.4. Routine service delivery data

Information on service delivery to households and individuals in the study population will be collected from service delivery departments (including the Departments of Health, Social Welfare, Home Affairs and Basic Education). While specific principles for linking to service records are outlined in section 4.5 below, the following enumerates some of the routine service delivery data intended for linkage with population surveillance data:

- Existing registers and databases will be linked to the HDSS, including public health registers like the TIER.NET system (an electronic patient records management system contains information on all treatment visits for people on ART), ETR.Net and EDR-Web (tuberculosis registers).
- National Health Laboratory Services (NHLS). Links to the routine laboratory results for all patients in the respective study populations.
- Department of Social Development - Social Security Information System.

- Department of Basic Education. Access to the school records of children in the respective study population.
- Home Affairs. Access to the civil registration and vital statistics register.

The SAPRIN Management Unit will negotiate with the respective departments to facilitate access to this information.

3.5. Research studies conducted in the surveillance populations

These are historical, current or future studies conducted with enrolled participants who are resident within the SAPRIN node surveillance areas. To obtain the maximum research benefit from such studies, it will be beneficial to link the records of study participants in such studies to their SAPRIN surveillance data. The following benefits are anticipated:

a. It will reduce the research burden on individuals who live within the study-area by reducing the collection of duplicate data by different studies.
b. It will increase the longitudinal reach of study measures, by, for example, the ability to look at distal outcomes for study interventions that took place at an earlier time, by linking historical study data with outcome data (e.g. survival) measured by SAPRIN.

4. Study procedures and evaluations

4.1. Measurements

4.1.1. Household component

Informed consent will be obtained from the household informant. (See Appendix A for list of types of informed consent forms to be developed by each HDSS node).

The following information will be collected as part of the household component (Appendix B contains full list of core data elements):

a. **Locations**. A location is a building, or a group of buildings, on land belonging to a single person or organisation, and used for one main purpose. locations are primarily used for residential purposes by one or more households. Each location is geo-located using global positioning system (GPS) measurements and allocated a unique identifier number that will be optionally displayed on a plaque attached at the entrance to the location. The type and status (occupied, under construction, broken down) will be recorded at each surveillance visit. The purpose of collecting this information is to ensure that the total eligible population is covered in the surveillance and to enable the spatial analysis of research results based on the location of an individual's residence and its proximity to service delivery infrastructure.
b. **Households**. A household is defined as a social group of one or more individual members who may or may not be resident at the time of the interview. Household members are usually, but not always, related. They share in the joint household resources and know each other well enough to provide information about each other. In each household, one of the members is the household **head**. The head of the household is the household member considered by other household members to be their head. It is usually, but not always, a senior male member of the

household. The household informant ideally should be a senior resident member of the household and not necessarily the head. The informant should be knowledgeable about the different members, their presence at the location and their relationship to the head of the household. Information is collected about household migration, that is, when a household moves as a unit to another location. Information about household assets, energy and water sources, and sanitation are also collected. The purpose of collecting this information is to ensure that the total eligible population (all household members) is covered in the surveillance, to keep track of households as they move within the study site, to contextualise information collected on individuals within the household and to control for socio-economic status of households during data analysis.

c. **Individuals**. Household members, where a **resident** is a member of a household who normally lives (i.e. intends to sleep the majority of the time over a four-month period) at the same location as the household; and a **non-resident** is a member of a household who does not normally live at the same location as the household but is nevertheless considered a member of the household. Information is collected from the household informant about the union status of each household member and their relationship to each other and the household head. Any known pregnancies and pregnancy outcomes (stillbirths, live births) of household members are recorded, as well as deaths and in- and out- migrations. Migrations include both movements internally within the HDSA as well as movements into and out of the HDSA. The purpose of collecting this information is to accurately measure the residential and household membership exposure of all individuals in the population and to determine population dynamics (fertility, mortality, migration rates, population age composition and life expectancy). On registration, everyone will be allocated two unique identifiers:
    i. **An external identifier (EXTID).** This unique identifier will appear on systems or documentation where the individual is identified (e.g. household member listings or referral slips) but will not appear on any datasets used for analysis. The EXTID will be kept in a single table in the research database together with other identifiable personal information. Access to this table will be restricted to the research data management staff.
    ii. **An internal identifier (INTID)**. This identifier will not appear on any documentation or system where the individual is identified and will be used to identify the individual in datasets used for analysis. This identifier will be used internally by the research database to tie together all research data related to the individual. The INTID and EXTID will appear together only on the single table mentioned in (i) above.

4.1.2. Individual component

In addition to the individual-level data on household members obtained from household informants in the household component, information will be collected from resident individuals aged 15 years and older who agree to participate in the **individual component** of the study. Written consent to participate will be obtained (see Appendix A for the type of individual informed consent forms to be developed by each HDSS node for). Written consent will also be obtained to communicate clinically relevant biomarker results back to participants and, where necessary, to facilitate linkage to care. Cell phone type and number will be collected to enable contact with the study participant. The following information and specimens will be collected (Appendix B lists the core data elements):

a. **Individual demographics**. Basic demographic information, including union, educational and employment status. The purpose of collecting this information is to communicate clinically relevant screening results and risk information to participants when appropriate and facilitate linkage to care where required (see Section 5.3), to enable follow-up for nested studies, and to control for individual characteristics during data analysis.
b. **Individual health**. The purpose of collecting this information is to characterise the health care utilisation, the prevalence of hypertension as a tracer condition for other non-communicable conditions; the prevalence of self-reported TB treatment, diabetes or stroke, and the HIV treatment cascade in the study population. It is complemented by the cause of death data obtained from the verbal autopsies to provide population-level mortality profiles.
c. A **finger-prick blood specimen** either in the form of dried blood spots on filter paper or a micro-capillary tube. All specimens will be tested for HIV using an enzyme-linked immunosorbent assay (ELISA) as part of the routine suite of tests for research purposes; participants will be offered a separate HIV testing service for their own health, as described in section 5.3. Sufficient blood will be collected for biomarkers of non-communicable diseases, such as HbA_{1c} for diabetes and CRP. Where these measures are indicative of a need for further clinical assessment and management of the participant, there is an obligation to feed results back to the participant and to establish appropriate protocols for linkage to care (See section 4.3.5). The purpose of collecting this information is to characterise the continuing evolution of the HIV epidemic and its response to intervention measures and to track the emerging non-communicable disease burden and risk.
   At the time of the individual interview, broad-based consent will be requested from participants for such use of the specimen (Section 10.7). This will include consent to store any specimens for future research testing. Separate ethical approval will be sought for such use of the collected specimens. For example, the following additional tests may be done on some or all specimens:
   i. Sensitive detection methods for infectious agents (including but not limited to HIV) will be applied, including those to quantify the number of blood-borne pathogens. These tests will be the most sensitive available at the time for the form of a specimen being analysed. Where present, infectious agents may be further characterised by pathogen gene sequencing. Such data will be used, for example, for detecting markers of drug resistance or transmission potential.
   ii. From time to time, based on the research objectives of SAPRIN or requests from investigators, other measures for other infectious diseases or non-communicable diseases may be conducted either retrospectively on stored specimens or prospectively on collected specimens. Separate ethical approval will be sought ffrom the relevant ethical committees for such use of the collected specimens.
d. **Height, weight, and blood pressure.** The purpose of collecting this information is to characterise the study population with regards to the prevalence of obesity and hypertension and to track the evolution of non-communicable disease risks. Blood pressure measurements will be performed in accordance with the World Health Organization (WHO) STEPwise approach to Surveillance (STEPS) method[18, 19].

4.1.3. Verbal autopsy

A verbal autopsy interview based on the latest World Health Organisation (WHO) standard questionnaire will be conducted with the last caregiver of all deceased household members. The

purpose of collecting this information is to determine cause-specific mortality rates. Deaths will be identified through the household component. Written informed consent will be obtained from the respondent. Verbal autopsy interviews will take place at least a month but no later than one year after the death occurred to allow for the immediate mourning period to conclude.

4.1.4. Routine service delivery data

Information on service delivery to households and individuals in the study population will be collected from service delivery departments (including the Departments of Health, Social Welfare, Home Affairs and Education), and linked to the population platform data. The purpose of collecting this information is to obtain accurate measures of individual and population-level access to services. Access to the service utilisation data will be obtained in terms of a memorandum of agreement between the Institution and the service delivery authority. We will request a waiver of individual written consent for the linkage of routine service delivery data with the platform data because (1) the research involves no more than minimal risks to the participants (2). After linking, data is anonymised before creating the combined dataset. The final dataset is strictly limited to anonymized information, ensuring minimal risk to patient confidentiality through the linkage process , and (3) analytical results will be reported at population or group level and not at the level of the individual participant. Prospective consent will be obtained from individuals to use the linked data as a basis to contact individuals for enrolment into research studies.

4.2. Data collection

4.2.1. Household and individual component

The desired frequency of data collection for the household component is a function of the resolution required to accurately detect key vital events, specifically births & still births, deaths (especially neonatal deaths)[20] and migration.

Reporting of still births and neonatal deaths can be improved by recording pregnancies and then subsequently enquiring about the pregnancy outcome. This implies that the household informant needs to be aware of the pregnancy status of each female household member. This is more likely to be the case during the second two trimesters of the pregnancy. Thus, an interview frequency that guarantees at least one interview during the last two trimesters of pregnancy will have an increased chance of recording the pregnancy – this implies an interview frequency of at least three times per year at the household level.

The accurate recording of residency episodes (the period between birth or in-migration and out-migration or death) requires at least one interview during the period of residency to record the presence of a resident household member. The frequency of household interviews will, therefore, determine the shortest residency episode that can be reliably detected; in other words, three household interviews per year will reliably detect residency episodes with a duration of more than four months.

However, frequent home visits increase the burden on household informants and raise data collection costs. Given the penetration of mobile phone ownership (>96%) in the study populations, it is possible to reduce the frequency of household visits but retain the benefit of accurately recording pregnancy outcomes and migration at a lower cost both to households and the Institution by augmenting

household visits with telephone interviews. Therefore, SAPRIN-nodes will conduct **at most a single household visit each year** and 2 telephone interviews with the household informants where the household has given consent and is reachable by phone.. Such telephonic interviews will review the household membership roster with the informant noting the current residential status of each member and the pregnancy status of female household members between the ages of 12 and 50. Where a pregnancy, birth, death or migration event is detected, the relevant questionnaire will then be telephonically administered. Telephone interviews and contacts will be conducted by trained interviewers based at a call centre at the nodal office. Household proxy respondents will provide written consent during the home visit for telephonic interviews. Only households granting consent will qualify for telephonic interviews. During the actual telephonic interview, digital audio consent will be recorded from the household proxy respondent.

The household and individual components will be combined into a single location visit during which the household informant will be invited to be interviewed (household component) as well as all eligible resident household members aged > 15 years (individual component). Interviews will be done by a team of fieldworkers who will visit locations according to a predefined visit schedule. Due to youth being in school at the time of some interviews, visits can be planned for afternoons and weekends.

Field-based data collection visits in an area will be preceded by community entry activities conducted by a community engagement team (see section 4.6).

4.2.2. Verbal autopsy

Verbal autopsy interviews will be conducted independently from the household visits, a minimum of one month after the date of death of the deceased, to allow for an appropriate mourning period. Interviews will be conducted by specially trained fieldworkers either telephonically or in person, depending on the availability of the care giver to be interviewed.

4.2.3. Routine service delivery data

Access to the service utilisation data will be obtained in terms of a memorandum of agreement between the Node and the service delivery authority, facilitated by the SAPRIN Management Hub, and complaint to the Protection of Personal Information Act (POPIA or the Act), No. 4 of 2013, and any associated regulations, including the Code of Conduct of Research. Currently, the AHRI Node has such agreements with the KwaZulu-Natal Provincial Department of Health and the National Health Laboratory Service and similar agreements will be entered into with other service delivery departments, such as Social Welfare, Home Affairs and Education. Data will be transferred to the mutually agreed trusted information system directly from the service delivery authority, and the transfer procedure will be specified in a node-specific standard operating procedure.

4.3. Screening, counselling, testing, and referral

4.3.1. Overview

The current protocol covers the communication of HIV test results,facilitation of linkage to HIV care, and communication of blood pressure and body mass index (BMI). In the future, other diagnostic and risk information may be collected during the individual component, where communication of the

information is warranted. A protocol amendment will be submitted at that point to describe how such communication will be handled and, if appropriate, the procedures to facilitate linkage to care.

4.3.2. HIV counselling & testing (home)

In addition to the blood specimen that is obtained for research purposes, household members will be offered rapid HIV testing during the household visit. All resident members who are aged ≥15 years will be eligible to be offered HCT. For participants who give consent for this component (and assent in the case of those aged 15 to 18 years), the test will be performed by field workers trained in HIV counselling and testing (HCT) in accordance with national guidelines [21]. Individuals may choose to have HCT even if they do not consent to participate in the other elements of the individual component if they are on the eligibility list for the individual component.

The test will be done in a private space in the home after written informed consent. The rapid test methodology will meet or exceed the requirements of the South African national protocol for rapid HIV testing and will consist of a minimum of two tests from different manufacturers. The rapid test result will be communicated at the time and place of testing, except in case of a discrepant result, where the ELISA result will be communicated during a scheduled household visit.

Other vulnerable individuals (e.g. people with mental health problems) identified by the fieldworker team will also be visited, assessed and tested (if consented) by a suitably qualified individual (e.g. project nurse). If a Node is unable to provide a suitable qualified individual for this purpose, vulnerable individuals will be excluded from testing.

Results from HCT at the household or the mobile clinics will be entered into the HDSS database using the same procedures as for the household and individual components (see Section 7) to monitor and evaluate linkage to HIV care.

4.3.3. Linkage to care for HIV and non-communicable conditions

Clinically relevant findings during research data collection require linkage to care. Clinically relevant findings should be determined to a level where care can be readily initiated if required by the care service to which the participant is linked. Nodal clinical governance structures to determine what the requisite diagnostic certainty is by considering the capabilities of local care services.

An HIV-positive participant (newly diagnosed and participants previously diagnosed but not yet on ART) will receive a referral slip or card for HIV care to facilitate the linkage process at the clinic. Participants will be encouraged to attend a clinic of their choice within the next 7 to 10 days after testing positive. HDSS nodes will establish referral systems based on either having study staff presence at clinics or collaborative agreement with the local government health department. These referral systems will include follow-up procedures to verify that referred individuals were initiated on ART.

Where measurements (e.g. blood pressure) are taken from research participants and have clinical implications (e.g. participant may require treatment for hypertension), the national department of health standard of care protocol will be used to determine the need to refer the individual for further clinical management at a nearby health facility. To minimise over-referral, and so overwhelming of

government health services, HDSS nodes will take steps to a) confirm diagnoses before referral, where possible; b) offer lifestyle advice as an alternative to referral, where appropriate and c) ensure that referral health services have appropriate equipment for further diagnostic evaluation.

4.4. Laboratory procedures

Section 5 and 9 provide detailed and comprehensive guidelines for managing personally identifiable information, while the focus here is on the principles of laboratory procedures.

4.4.1. Specimen collection

All eligible individuals will be asked to consent, in writing, to give a finger-prick blood specimen. Collection will be done either in the form of dried blood spots (DBS) on filter paper or a micro-capillary tube. For safety, retractable, single-use lancet needles will be used for blood collection. The blood specimen will be only labelled with a barcode number, which uniquely links the specimen to the participant's study ID number (INTID) through scanning the specimen barcode into a computer tablet at point of collection.

4.4.2. Specimen handling and transport

After DBS collection, filter papers will be kept in a rack placed in a closed container labelled as a biosafety hazard and transported to the laboratory where the DBS cards will be kept overnight at room temperature. The following day, DBS cards will be prepared for transportation to the analysis laboratory, where they will be stored in a freezer at -20 degrees Celsius.

At the end of each working day, fieldworkers will submit all specimens to their fieldwork supervisor, who in turn will submit the specimens to the laboratory technician, who will control the drying, packing and shipping of the samples. Specimens will be transported on a regular basis from the site to the laboratory for analysis and final storage.

4.4.3. Laboratory procedures

Upon receipt in the laboratory, the barcode of each specimen will be scanned into the Laboratory Information Management System (LIMS). Specimens not suitable for analysis will be reported to the study coordinator, along with a daily LIMS report on specimen submission. The specific assays and methods that will be used for each test will be described in a study-specific laboratory procedures manual.

4.4.4. Specimen storage

It is the aim of the laboratories to keep all collected samples in good condition for analysis. Therefore, all specimens and aliquots stored in the nodal laboratories will use appropriate storage containers with required labels which include the laboratory specimen ID or specimen ID and date of processing/receipt. The original specimen (e.g. in case of dried blood spots) or aliquots will be stored under appropriate temperature conditions. Where derivative specimens are generated after testing, e.g. nucleic acid extracts or polymerase chain reaction (PCR) amplicons, excess specimens will also be stored under appropriate temperature conditions. The storage location of all specimens, aliquots and derivatives will be maintained within the LIMS.

As a default practice, specimens are retained for the operational lifespan of a HDSS node, essentially archived indefinitely, according to Good Laboratory Practices (GLP). However, in instances where a node is unable to sustain storage due to cost or space constraints, SAPRIN should be promptly notified of the decision. A review will be conducted to explore options such as providing funding to support continued storage or permitting the node to proceed with the destruction of specimens in accordance with its established procedures.

4.5. Record linkage

4.5.1. Routine Service delivery Records

Routine service delivery records (Section 4.2.3) will be linked to data from the study population. To protect the confidentiality of the individuals involved, the linkage will be conducted in two stages:

i. Personal identifiable information (names, identity numbers, birth date, sex, local area of residence, phone number, service location) will be obtained without any individual-specific service delivery information, e.g. the result of a laboratory test or the allocation of a social grant to the individual. Each of these records will contain an identifier that the service delivery organisation can associate with the service delivery records belonging to that individual but which contains no identifiable information about the individual when viewed in isolation. We will generate a similar file of individually identifiable characteristics from our research database without any other information we may hold on the individual. This data will also contain an identifier that will anonymously link the individual to the rest of our research data on that individual. These two files containing only personally identifiable information will be subjected to a matching procedure based on the individual characteristics to produce an output that contains only two identifiers (one originating from the service provider and one from our information on the individual). The matching procedure is done in isolation from the service delivery data to protect the disclosure of this data to the staff involved in the matching. The matching will be done using a secure information system environment trusted by both parties. The specific matching methodology, acceptable secure environment and tools used in the linkage will be specified in a standard operating procedure approved by both parties.
ii. This file will then be used to retrieve the service delivery data (excluding any personally identifiable information) belonging to the matched individual for linkage to the research data about that individual. This procedure will take place on a dedicated secure computer located on the Node's secure network. Only a restricted number of data managers will have access to this computer and the matching datasets. Scientists and other data users will only receive an anonymous identifier with the linked service delivery and research data and will not have access to the underlying personally identifiable information.

4.5.2. Research Studies in the SAPRIN Node surveillance areas

Study records from these studies will be linked to the SAPRIN records by using either a shared unique individual identifier, in which case the data will be anonymously linked (e.g. without reference to personally identifiable data), or using personally identifiable information, in which case the same method as for service utilisation data will be used.

We will not seek individual consent to do this linkage because:

a. The process does not involve the collection of new data from an individual.
b. The process involves no more than minimal risks to the participants,
c. Only the study data manager and principal investigator of the study being linked will have access to personally identifiable information if required by the linkage process,
d. The analytical datasets resulting from the linkage will be de-identified to prevent the disclosure of personally identifiable data and
e. The results of analyses based on the linked data will only be reported or published at an aggregate level, further reducing the risk of identity disclosure.
f. All future studies that are collecting potentially sensitive information will still request permission from the applicable research Ethics Committee to link such data to the SAPRIN database.

4.6. Community engagement

HDSS Nodes will ensure appropriate community engagement with the research population, which is essential owing to the longitudinal nature of HDSSs and their location within vulnerable populations. Working with a Community Advisory Board (CAB) or similar stakeholder body, a comprehensive programme of community-based information sharing, collaborative and multidirectional consultation, and discussion will assist in ensuring ongoing relationships of mutual trust and respect within each Node.

4.6.1. Overview

During the period of health and socio-demographic surveillance preceding this protocol, a strong and comprehensive relationship has been established with the local communities in the existing Nodes. An ongoing interactive and integrated community engagement strategy will build on these existing strong partnerships and thus contribute to the sustainability of each Node to achieve the following aims:

a. enable local communities to contribute to setting research priorities and the research agenda
b. comprehensively prepare the community to participate in research studies, considering cultural and social sensibilities
c. consult with community members regarding strategies to deal with issues that might arise from research activities and emerging ethical issues
d. communicate to and discuss key scientific results with the local population
e. encourage data requests for use in community development from the local population
f. encourage evidence-based individual behaviour change based on scientific results
g. collaborate with community members to devise strategies to facilitate the uptake of scientific results by local service providers and
h. maximise the social value of research activities.

4.6.2. Community engagement strategy

To achieve these aims, tried–and–tested programmes of community engagement activities will continue in the existing nodes. A diverse range of engagement activities will likely be used in the different settings. Activities will generally fall into the following three areas: consultation and opinion-seeking; interaction and information-sharing with entire communities, including knowledge dissemination; and

targeted interaction and information-sharing. Nodes will ensure that the following activities are included in their community engagement programmes at a minimum.

Consultation and opinion-seeking

Each Node will:

a. Work with a CAB, whose main role is to ensure that the rights of research participants are respected, and to act as a bridge between the community and the Nodal staff and activities. The CAB is comprised of elected or nominated community representatives, who also receive training in this role.
b. Conduct community dialogues to assess community concerns, perceptions, misconceptions, attitudes and knowledge; respond to specific issues raised during research activities; and ensure appropriate knowledge dissemination activities are planned and implemented.
c. Meet with village leadership, local traditional authorities, municipal authorities and other community structures, as well as service providers – especially from the Departments of Health, Social Development and Education – to explain and discuss new studies being planned and to provide feedback results from previous studies. These meetings are also used to discuss the relevance of the research as well as any concerns from HDSS participants or service providers.
d. Participate in municipal and other public forums, either by invitation or when the HDSS requests time to present.

Interaction and information sharing with entire communities, including knowledge dissemination

a. Knowledge dissemination and discussion activities occur where village leaders call the community to gather at a school or other meeting place to hear presentations from Node staff about the research and discuss the relevance of findings. Government service providers, such as nurses and social workers, accompany the Node staff and can also answer questions. These activities inform people about forthcoming research, including when and how data collection will be done, explain study procedures, communicate key scientific results often specific to that community, and provide time to listen to community issues and answer questions.
b. Social media can be used to communicate with and engage communities. For example, a reverse-charge SMS portal for communities to communicate concerns and issues can be part of the call centre for telephonic data collection interviews. Such an SMS portal can also be used to raise study awareness and communicate scientific results to community members who have consented to telephonic contact.

Targeted interaction and information sharing

a. Engage the youth through sports and music tournaments, tailored knowledge products for inclusion in the school curriculum or youth group activities, or arrange interactive educational and general health awareness sessions.
b. Participate in and host annual national and international events of celebration, such as World AIDS Day.

c. Invite targeted populations such as youth or village leadership into the research centre for educational purposes as well as to increase understanding of research activities.

4.7. Staff responsibilities

Here, we outline the key HDSS staff categories and their respective responsibilities crucial for the successful implementation of this protocol. The focus is on leveraging innovative, efficient, and cost-effective systems, underpinned by rigorous quality assurance and quality control practices.

a) Operational manager

Responsible for overall protocol implementation activities and quality assurance, strategic organisational management of progress and quality assurance of implementation activities in an ethical, innovative, cost-efficient, and cost-effective manner.

b) Section coordinators

Responsible for coordinating operational activities of a specific workstream namely, community engagement, data collection and data management. Focus on ensuring there are clear standard operating procedures, supplies, quality control activities, quality assurance measures, effective training programmes and sufficient progress.

c) Team supervisors

Responsible for direct day-to-day logistical support, performance, and quality standards supervision of a team of operational staff, like team of data collectors.

d) Data collectors

Responsible for collecting data and biomarkers from participants through either computer-assisted personal interviews (CAPI) or computer-assisted telephonic interviews (CATI), this includes specialised data collectors such as those with skills to track hard-to-reach participants.

e) Data managers

Responsible for development and programming of electronic data collection platforms, uploading, chain of custody and synchronisation of electronic data records, quality control and quality assurance of data, data linkage and integration, extraction, documentation, and archiving.

f) Communitry engagement officers

Responsible for establishing and implementation of platforms and activities for engagement with local communities through relationship building; consultation and opinion-seeking; interaction and information-sharing with entire communities, education and knowledge dissemination; and preparing communities for data collection.

5. Data management

5.1. Data management responsibilities

5.1.1. Data custodian

Data will be utilised for analysis and publication as determined by the scientific strategy of the SAPRIN Steering Committee. The data custodian for SAPRIN will be the SAPRIN Director. SAPRIN will make datasets available, according to levels of access and set criteria, via an appropriate data repository and complaint to the Protection of Personal Information Act (POPIA or the Act), No. 4 of 2013, and any associated regulations, including the Code of Conduct of Research. It is the responsibility of the data custodian to ensure that data access is consistent with the terms of the ethical approval obtained for this protocol. The custodian for HDSS data at each node will be the nodal PI.

5.1.2. Data security

The HDSS node will put in place robust data security policies compliant with national statutory requirements, such as POPIA, that clearly delineate roles and access levels for individuals with authorized access to research participant information. Additionally, the HDSS node will implement procedures for controlled access to data. All electronic devices containing data in electronic format, whether desktop or portable systems, external storage media, or servers, will undergo regular backups in accordance with the HDSS node's documented procedures and data recovery plan. This process will incorporate strong authentication measures to safeguard data access. Furthermore, backup databases and files will be encrypted in compliance with established encryption standards, and they will be securely stored in a designated location.

5.1.3. Data collection

Household and individual components, verbal autopsy data collection and clinic linkage management is the responsibility of the nodal research operations head or his/her delegate, who is responsible for timely and accurate data collection as specified in section 5.2.

5.1.4. Research data management

Once data have been stored in the longitudinal population and clinical databases, they become the responsibility of the research data management at each Node. Research data management will regularly assess data quality and implement measures, such as data constraints and data integrity rules, to ensure data quality. Research data management will provide on an annual basis; within two months of the conclusion of the annual household data collection round, the core data elements listed in Appendix B. SAPRIN will maintain a central data repository that will consolidate the core data elements from all SAPRIN nodes. SAPRIN will develop and implement network-wide data quality metrics. SAPRIN nodes will comply with agreed data quality standards set by SAPRIN in consultation with nodal research data management.

5.2. Data capture methods

5.2.1. Data collection

The node-specific electronic data collection systems will be used to create and administer electronic questionnaires. These questionnaires will be displayed on a tablet computer, and all collected data will be encrypted on the device. The electronic forms will implement skip patterns and data validation checks. Two data collection modes will be used:

i. **Computer-assisted Participant Interviews (CAPI)**. The data collector reads out the question and records the response from the participant on the electronic form. This mode will be the default for data collected in SAPRIN.
ii. **Computer-assisted Self Interview (CASI)**. This mode will be used for sensitive questions. The interviewer will hand the tablet to the participant, who will answer the questions without disclosing the answers to the interviewer. Where needed, auditory assistance will be provided (ACASI)

5.2.2. Call centre

Telephonic interviews will be initiated and conducted by trained operators based at a call centre at the node from a list of household informants produced from the study population database. Operators will capture the result of the telephonic interviews directly into a computerised questionnaire. The call audio is stored and archived in the call centre system, linked to the interview record in the database, and can be retrieved as needed through access-controlled protocols.

5.3. Types of data

5.3.1. Personally identifiable information

All personally identifiable information will be collected and handled in compliance with the Protection of Personal Information Act (POPIA or the Act), No. 4 of 2013, and any associated regulations, including the Code of Conduct of Research. All personally identifiable information (including names, identity numbers, addresses, geolocation data, and telephone numbers) will be stored in a separate table in the database with restricted access and identified by a separate identifier (see section 5.1.1). This information will not be included in datasets on the data repository. If access to this data is required for research purposes, it must be authorised in each case by the data custodian. The data will only be made available in a data enclave. The data enclave is a virtual machine launched from the researcher's own desktop but operating on a remote server under the control of the Institution, like remotely logging into another physical computer. The virtual machine is isolated from the user's physical desktop computer, restricting the user from downloading files or parts of files to their physical computer. The virtual machine is also restricted in its external access, preventing users from emailing, copying, or otherwise moving files outside of the secure environment, either accidentally or intentionally. Once a data user has produced the output from the analysis, the output will be vetted by a member of the research data management section and if the analytical results contain no personally identifiable information, the results will be made available for download from the data repository.

5.4. Data Sharing and Dissemination

Data sharing and dissemination will be done according to SAPRIN's data use policy. SAPRIN will maintain a publicly accessible repository where research datasets will made available compliant with FAIR data principles and national protection of private information legislation (e.g. POPIA). Datasets will be documented using a generally accepted data documentation standard (currently Data Documentation Initiative v2.5) and assigned a Digital Object Identifier (DOI). Dataset metadata will be open access, with the data content accessible based on a click-through data use agreement or by submitting a request through the data repository, providing information on the affiliation of the data user and a description of the purpose for which the data will be used. Data requests are approved by the SAPRIN Director or delegate. Ownership of collected HDSS data rests with the nodal institution; core data elements submitted to SAPRIN are co-owned by SAPRIN and the nodal institution.

5.5. Data processing and storage

Data will be processed and stored on servers under the physical control of the Node in industry-standard relational databases with data integrity and user authentication for access control. Data will be replicated on at least a daily basis to an alternative site to provide secure offsite storage of data. Transactional logs will be backed up every 30 minutes to enable recovery of data in the event of equipment failure.

All users of the system will be authenticated through individual passwords with minimum complexity and regular change rules. The node will use industry-standard malware and intrusion detection with at least annual penetration tests by a reputable outside security audit company.

At the Node, a client-server architecture will be implemented where data is not stored on laptops or local workstations but only on a central server with restricted physical access.

Because this is a longitudinal HDSS with no end date, data are permanently stored. Each HDSS node will detail how it will dispose of data if it ceases to be part of the SAPRIN network, when it runs out of storage space, or its host institution ceases to exist.

5.6. Data collection computer tablets

To ensure secure data collection, computer tablets will be subject to password access control, with a remote wiping feature activated in the event of loss or theft to prevent unauthorized access. Each data user will be assigned a unique account for accountability. Overnight, computer tablets will be stored in access-controlled cabinets. Additionally, regular synchronization is mandatory, with data transferred to the server at least twice daily—once at the commencement and once at the conclusion of the working day.

6. Quality control and quality assurance

6.1. Standard Operating Procedures

All study procedures and activities will be standardised through the development of a suite of standard operating procedures (SOPs).

6.2. Staff training

All study staff will be trained on SOPs that are relevant to their role in the study. Training programmes, based on SOPs, will be delivered using learning outcomes approaches that focus on developing practical skills about procedures study staff should know or be able to do.

All study staff will be trained in map reading, household entry, research ethics (including obtaining informed consent), Good Clinical Practice (GCP), face-to-face and telephonic interviewing skills, electronic data collection techniques, handling of social problems including child protection issues, and in linking participants into care at DoH clinics.

Fieldworkers will be required to successfully complete DoH accredited HIV counselling course to conduct HIV testing. The training includes skills like pre and post-HIV counselling, linkage to care, and working with vulnerable groups, including adolescents. Additional training in finger-prick collection and handling of dried blood spots (DBS) and micro-capillary blood specimens, with further specific training added as needed.

At the start of each year, at least two weeks will be allocated for HDSS-specific staff training. Regular operational staff meetings focused on discussing and training on issues identified by fieldworkers and/or through supervision and quality control programmes (see below) will be held.

6.3. Supervision and data collection quality control

Fieldworkers will be organised in teams, each led by a supervisor. These teams enable close supervision of each fieldworker. Each fieldworker will have at least two supervised data collection interviews in a month. During supervised interviews, the supervisor observes the fieldworker practically going through all data collection procedures and recording the data. The supervisor will give feedback of their observation to the fieldworker, including on-the-job training, where necessary, in addition to compiling supervision reports, which will be analysed to inform and develop occasional refresher training programmes. An additional data collection quality control measure involves randomly selecting at least two interviews conducted by each fieldworker per month for a repeat visit by a supervisor. This aims to verify adherence to procedures, measure how accurately questions were asked, and responses captured.

6.4. Data quality control and assurance

All data recorded on computer tablets will be subjected to rigorous checking. First, the electronic data collection platform will be programmed with validations, constraints, and skip rules to guide the data collector in the interview. Data quality controllers, who will employ automated electronic quality control system to l check for accuracy, data logic, internal consistency and completeness of data recording of every interview. All cases with errors will be rejected for correction by the fieldworker and

the supervisor who originally did the data collection. Further, errors are recorded and analysed to identify trends and causes. Resultant quality control reports will be discussed in monthly field team meetings. Every HDSS node will generate an electronic quality control and validation document detailing validation rules, constraints, and skip rules for each core data element. Additionally, HDSS nodes will compile a list of errors identified during electronic quality checks, regularly produce quality metrics, and conduct data profiling.

A common harmonised longitudinal database will be designed by SAPRIN and deployed at each HDSS Node. Equally, each HDSS node, unless given special exception, will extract data from the HDSS nodal data collection platform into the SAPRIN database using a centrally developed data extraction utility tool, designed with robust temporal data validation rules, to ensure all data is of the same quality standard and follow same variable labels and response codes.

6.5. Annual operational Audit of HDSS Nodes

To ensure adherence to quality standards, adherence to procedures and protection of research participants, the SAPRIN Management Hub will conduct an annual operational audit at each HDSS node. Additionally, data submitted to SAPRIN by the HDSS nodes will undergo further plausibility data quality assurance.

7. Protocol oversight

7.1. Nodal Directorate

The nodal director and principal investigator of the HDSS will be responsible for the scientific oversight of the protocol as implemented at the nodal level.

7.2. SAPRIN Steering Committee

This is the decision-making body that upholds the SAPRIN vision of open and excellent science while balancing the resources available, overseeing spending, and holding responsibility for the actions of the SAPRIN Management Hub. The Steering Committee will ensure the scientific work focusses on issues relevant to improving health and socio-economic status and that the full potential of the infrastructure is realised.

7.3. Requirements of the Nodes and the SAPRIN Steering Committee

a. Multi-nodal studies will be solicited and approved at the NRI level and facilitated by the SAPRIN Management Hub.
b. b. Nodes must maintain the capability to host properly budgeted multi-nodal studies initiated by SAPRIN, subject to approval by the SAPRIN Steering Committee or as provided in the data access and sharing policy.
c. As autonomous Research Units, HDSS Nodes will solicit and approve the research studies that they host independently of SAPRIN.

d. Nodes will maintain registers of proposed and active studies to be shared with the Management Hub to keep track of the research portfolio generated by the National Research Infrastructure and to facilitate knowledge translation from the research outcomes.
e. Nodes will acknowledge SAPRIN in their research outputs related to the HDSS infrastructure.

8. Safety considerations

8.1. Risks of harm

8.1.1. Response to positive disease screening test

The response to a positive disease screening test varies from one person to another and may include shock, crying, agitation, stress, guilt, withdrawal, anger and outrage. The fieldworker or research nurse conducting the test will allow the participants to deal with the news in their own way and give them the opportunity to express their feelings. The fieldworker or research nurse will determine the needs of the individual participant and will deal with them accordingly as part of the post-test counselling process. Any participant in severe distress will not be left alone; the research team will ensure that the participant has appropriate support. In some instances, crisis intervention will be required, especially if there is any suggestion of suicidal ideation or intent. In these cases, urgent referral to appropriate care at the local district hospital will be made, or emergency admission to hospital will be facilitated.

8.1.2. Inadvertent disclosure of screening status

Through the mechanisms to support linkage to care, including text messages and phone calls (section 5.3.4) there is the potential for the inadvertent implicit disclosure of disease status if community members become aware that such messages and calls are only used for positive participant status. However, text messages and phone calls will be utilised for other purposes unrelated to a participant's disease status. All cell phone numbers will be verified by research personnel during the household visit. No text message will explicitly mention a disease, and participants will select a coded message from a range of options that can be sent to them to support their linkage to care.

8.2. Reporting procedures

Any concern regarding potential risks and harm to participants will be reported to the hosting institution's management through the appropriate lines of responsibility. An initial assessment of the risk of harm will be made and any serious adverse event (SAE) will be reported to the Steering Group. All SAEs will be forwarded to the Human Research Ethics Committee within seven days.

8.3. Safety oversight

8.3.1. Nodal Directorate

The Nodal Directorate will be the senior scientific management group of the hosting institute. This will include the Principal Investigator of the HDSS Node. This group will meet on a regular basis in accordance with the scientific strategy of the hosting institute. The group will ensure high standards of clinically related work in SAPRIN Nodes, and the relevant protocols will be scrutinised to avoid any potentially adverse impact of our work.

8.3.2. Clinical Governance

The Node Directorate ensures that clinical governance needs are identified and addressed within all research activities and makes an initial assessment of the risks and harm to participants during the conduct of the study. This allows for continuous assessment of policies and procedures to identify the need for changes, training, and staff development.

8.3.3. SAPRIN Steering Committee

. The Steering Committee has oversight of the NRI budgets, spending, and the focus of the scientific work, as well as ensuring quality and safety. The SAPRIN Steering Committee and centrally-based Management Hub will work with the Nodal Directorates to ensure that ethical reviews are of the highest quality and safety considerations are strictly adhered to.

9. Ethical considerations

9.1. Background

The research outlined in this protocol will be conducted in accordance with the Declaration of Helsinki[22]; the principles of Good Clinical Practice as laid down in the ICH Harmonised Tripartite Guideline for Good Clinical Practice[23]; and the ethics guidelines of the Department of Health of the Republic of South Africa[24].

The introduction of individual rapid HIV testing (with results) to participants aged 15 years and over, in addition to the dried blood spot specimen for HIV sero-surveillance, has been prompted by ethical concerns locally and internationally to communicate positive results back to individuals. In the context of i) a locally high incidence of HIV, ii) highly effective treatments for HIV, and iii) a need for substantially higher testing coverage in the area, there is an ethical imperative to directly offer participants the opportunity to learn their HIV status. Similarly, given high rates of hypertension and escalating risks for non-communicable diseases such as diabetes, it is important to refer participants with a high blood pressure reading for care.

9.2. Human Research Ethics Committee

The protocol will be submitted for review by the Biomedical Research Committee (BREC) of the South African Medical Research Council (SAMRC). In addition, each Node will submit a protocol to their local Human Research Ethics Committee (HREC) based on this core protocol but adapted to their local needs for ethical approval in accordance with the HREC requirements. Further, the HDSS node protocol might be subject to approval by other ethics, including those of collaborating scientists and external funders.

9.3. Elements of Informed consent

Every informed consent form contemplated in Appendix A will be developed to include all the basic elements of as required by South African (SA) ethics guidelines and other international base practices, namely;

a) **A statement that the study is research**; provide a clear and understandable explanation of the research and its purpose.

b) Voluntary Participation: emphasize that participation is entirely voluntary, and participants can withdraw at any time without consequence.

c) **Study Procedures**: Clearly outline the details of the research procedures, including the duration, nature, and frequency of any interventions or interactions.

d) **Risks and Benefits**: Disclose potential risks and benefits associated with participation in the study, ensuring a balanced and comprehensive overview.

e) **Confidentiality**: Describe the measures taken to safeguard participant confidentiality and how data will be handled.

f) **Contact Information**: Provide contact information for the researcher or a designated representative who can address participant questions or concerns, as well as details and contacts of the research ethics committee that provided ethical clearance for the study.

g) **Alternatives**: Explain any reasonable alternatives to participation, where applicable, and the implications of choosing not to participate.

h) **Language readability level**: all informed consent and all participant facing documents will be written to comply with the South African (SA) ethics guidelines standards of English language readability level of Grade 8 or lower on the MS Word Flesch-Kincaid Reading Grade score and will be translated and administered in the local language(s) set out in the application for ethics clearance.

i) **Understanding**: Ensure that participants comprehend the information provided by employing clear language and assessing their understanding.

j) **Opportunity to ask questions**: Allow participants the opportunity to ask questions and seek clarification regarding any aspect of the study.

k) **Written informed consent**; although some components of this study, like household surveillance, is minimal risk, we have taken the decision to have all data collection activities done after obtaining written consent as a way to ensure compliance with POPIA Act.

l) **Age of consent**: In compliance with South African laws that set age of majority at 18 years, we will consider individuals aged 18 and above are to have the legal capacity to provide independent consent. For the individual health surveillance segment of our study, which includes individuals aged 15 and above, we will initially obtain parental/guardian consent for minors aged 15-17 before obtaining the independent assent of the minor.

9.4. Informed consent process

The informed consent process will be conducted by trained fieldworkers. Node specific consent forms in the language spoken locally will be developed to comply with South African (SA) ethics guidelines standards of a readability level of Grade 8 or lower on the MS Word Flesch-Kincaid Reading Grade score. An approach to the assessment of capacity to consent will be implemented at each node.

Consent to approach adolescents aged 15-17[1] years will be sought from the parent or guardian. If they agree to such an approach, adolescents will sign an assent form after being provided with information about the study in an understandable form using age-appropriate language. Study procedures will be done privately without direct observation by a parent or guardian. Adolescents of all ages have a right for their HIV and other test results to remain private and confidential, although disclosure to the parent, guardian or another appropriate adult will be encouraged and supported. HDSS Nodes will put in place detailed procedures of how to handle follow-up visits and linkage to care where an adolescent does not want to disclose HIV status to parents.

We will treat adolescents aged 15-17 years as emancipated minors if they: a) live in a child-headed household where there are no adults; b) are married, and/or c) are a parent, unless an HDSS Node's Ethics Committee of Record indicates a different framework. We will document the specific circumstances of child emancipation, and we will request waiver of parental consent for all documentable cases of emancipated children.

The informed consent process for pre-literate participants will be conducted in the presence of an independent witness nominated by the participant but cannot be a staff member of the HDSS node. Verbal consent will be obtained from the participant, and their consent will be indicated by a mark on the electronic consent form. The independent witness will be asked to sign the informed consent form on behalf of the participant.

The official consent record maintained by the HDSS node will always be in electronic format. Electronic written informed consent, including signatures, will be obtained through touchscreen computer tablets in the data collection application. In cases where e-signatures are not supported on tablet devices, consent signatures will be recorded on a hard copy, which will then be captured into the data collection application as a photograph. In all instances, participants will always receive a hard copy of the consent, signed by both the participant and the data collector.

In Section 5.4, data processing and storage, we outlined the security features of data collection devices and servers, emphasizing password-controlled access. Data collection devices will undergo regular synchronization, transferring data to the server at least twice daily, at the start and end of the working day. After passing quality control and assurance, consent personally identifiable information (including signatures and/or photographs) will be moved to a protected table accessible only to a designated data manager and the HDSS Nodal PI. Notably, consent information is excluded from analytical datasets or when sharing data.

[1] Children from their 15th birthday up to and including the day prior to their 18th birthday.

9.5. Inclusion of vulnerable groups

The inclusion of adolescents is important in these surveillance areas where hypertension and diabetes are not well studied and where the incidence of HIV infection in young women is high. There is, therefore, a need to better understand risks in this group. Clearly offering tests for non-communicable conditions and HIV testing to this age group brings more complex ethical issues, and thus, they will be treated differently within the study-specific protocols with additional professional expertise and support. Each HDSS node will outline, in its node specific protocol and standard operating procedures, details of how it will appropriately manage adolescents and deliver expert professional support services.

As a minimum standard, an arrangement with the local Dept. of Social Development (DSD) will be entered into to allow the referral of cases requiring social support directly to DSD social workers stationed in each of the local municipal wards.

9.6. Guidance on Approaches to Child Protection

Legal responsibilities to protect the well-being of children are clear. Conducting interviews in homes leads to several potential child protection challenges:

i) Field staff may witness or encounter a possible or actual child protection concern with a participant, and this may or may not relate to node specific protocol questions on, for example, sexual health.
ii) Through their presence in homes and communities, field staff may witness or encounter a possible or actual child protection concern with non-participant children.
iii) There is a small risk of research organisation staff being impersonated or abusing their position to gain access to children (or property).

Research organisations are bound to respond to these concerns.

i) In the process of completing individual questionnaires, information may be obtained that explicitly or implicitly provides evidence of a sexual offence (e.g. statutory rape). As researchers, there is an obligation to report the commission of sexual offences against children according to the Criminal Law (Sexual Offences and Related Matters) Amendment Act 32 of 2007 and Amendment Act 5 of 2015[22, 23]. Interviews are conducted in people's homes in the locality, as fieldworkers systematically move from one location to the next. **This core protocol does not contain any questions on sexual relationships or behaviour**. However, should a node specific protocol based on this protocol contain questions on sexual relationships or behaviour, we recommend that participants give information on sexual relationships directly onto computer tablets and, as such, not disclose sexual behaviour directly to the fieldworker. Fieldworkers should be unable to access the answers given. These approaches are consistent with those outlined in a key memo on the topic[25]. These approaches should be discussed and agreed upon with the local office of the Department of Social Development and a written undertaking of their ability to handle social issues (including child protection) referred to them should be obtained. However, whether data is collected electronically or through interview, participants can be encouraged to seek help

through the interviewer or field worker. The fieldworker should, in turn, be trained to give an appropriate response; responses might include a locally tailored list of self-referral helplines and services or an ability to obtain an urgent call-back for advice. Finally, self-referral options can be embedded in interviews, triggering a call back from a research nurse within a specified number of days.

ii) Field staff should be trained in appropriate responses, with relevant supporting SOPs, should they encounter a child protection concern with a non-participant.

iii) The appearance of fieldwork staff should be standardised by nodes, with uniform and ID badges, to help community members ensure as far as possible that staff are bona fide.

1. **Where 15 to 17-year olds enter on tablet computer, or disclose in an interview, that a sexual partner is 2 or more years older.**

   If these individuals are reported to the police, as suggested by one interpretation of the law, major social harm will result. Criminalisation at scale will occur as young people (who have been systematically contacted) are entered into the criminal justice system - and some will be entered on the sexual offences register. This process itself may put young people at higher risk from partners, family members or others.

   Nodes will need to develop responses to protect the well-being of young people whilst avoiding referral into the criminal justice system.

2. **Where participants enter on tablet computer that a sexual partner is 2 or more years younger and aged under 18 years old.**

   Similarly to 1 above, we do not recommend the collection of any details of sexual partners other than age. Participants may or may not know the true age of their partners and may or may not know their contact details. We believe, given these uncertainties, that screening of collected data for such reports, followed by mandatory reporting of the participant (as 'perpetrator') to the police, would cause social harm and is likely to put some young people at higher risk.

9.7. Privacy and confidentiality

Lack of timely linkage to care is currently one of the major obstacles to success in HIV treatment programmes and the management of hypertension, diabetes and other non-communicable conditions. We believe that we have an obligation to make referrals to high quality care and to offer some simple interventions to support participants to obtain life-saving treatment. Given that the majority (>90%) of individuals eligible (>15 years) for individual health surveillance possess cell phones, the default and preferred method for follow-up contact for linkage to care will be through confidential telephone calls or text messages. However, a text message sent by Short Message Service (SMS) and a phone call could be considered intrusive, and there is a potential risk of inadvertent and indirect disclosure of HIV status and other health information through these mechanisms. We will mitigate this risk by verifying all phone numbers provided by participants and by offering a range of coded message options that will protect confidentiality. In instances where cellphone numbers are unavailable, or telephone contact has not been consented to, follow-up for linkage to care will involve in-person visits to the homestead by trained

staff, ensuring the confidential handling of HIV status and treatment information. To uphold confidentiality standards, every HDSS node will establish and execute an employee confidentiality or non-disclosure agreement with personnel who will come into contact with participant confidential, personally identifiable information.

9.8. Future use of stored specimens and data

Nodes participate in major multicentre and multi-country studies, such as Human Heredity and Health in Africa (H3Africa) – and, in the process, ensure full compliance with ethical precepts and the principles of community engagement. Broad consent from the participants will be sought for storage of blood and other specimens, such as saliva and urine, where applicable, and for future use by researchers according to the Node's scientific strategy. Any request for access to and use of the stored specimens from the Node's biobank will need to be reviewed and approved by the Principal Investigator and Director of the Node. Ethics approval from the Human Research Ethics Committee (HREC) will be required for use of stored specimens for purposes other than those described in this protocol; plans to feed back the results of specific tests must be clearly stated in the ethics application. Unless specifically indicated by the HREC, participants will not be re-contacted to consent to the use of their specimens for this research. Such tests may result in findings which require health intervention in the opinion of the local clinical governance body. Should this be the case, participants will be contacted and referred to appropriate care, pending HREC approval. Participants will be able to withdraw their consent to storage of their specimens at any time.

9.9. Compliance with POPI Act

All personally identifiable information will be collected and managed in compliance with the Protection of Personal Information Act (POPIA or the Act), No. 4 of 2013, and any associated regulations, including the Code of Conduct of Research.

10. Publication & data sharing policy

10.1. Publication policy

Each HDSS Node is an independent Research Unit linked to national and international collaborators and funding partners. The Node Director oversees the quality of research underway and ensures fairness of collaborations, including aspects like co-authorship.

Multi-nodal projects will, in addition, be monitored by the SAPRIN Management Hub. Publication plans for projects using data from multiple Nodes should be reviewed for fair participation and co-authorship. The SAPRIN Director retains the discretion (which shall not be unreasonably exercised) to require an amendment to any draft abstract/paper, to amend authorship, to refuse permission for publication, or to require a delay in submission for publication providing that it is within the best interests of the SAPRIN.

We would expect full acknowledgement of the SAPRIN in any publications that draw on the HDSS platform, and if the collaboration was substantial, we would expect co-authorship. Authorship guidelines, as recognised by the International Committee of Medical Journal Editors (ICMJE) (http://www.icmje.org/) will be followed. Authors must acknowledge the funder(s), the study site, and the participants. A standardised wording of acknowledgement will be provided.

10.2. Data sharing policy

When a Node joins SAPRIN, it will contribute its historical data on the core protocol variables (Appendix B) for archiving in the first SAPRIN data repository available for use by nodal and the broader scientific community.

The HDSS Node will contribute the annual update to the SAPRIN data repository as specified by the minimum protocol. The data should be provided within 3 months after the close of the data collection round for the preceding calendar year.

a. The SAPRIN-identified data repository will share the contributed data (including information derived from the service utilisation linkage) with appropriate controls to limit the disclosure risk of private and confidential data and complaint to the Protection of Personal Information Act (POPIA or the Act), No. 4 of 2013, and any associated regulations, including the Code of Conduct of Research.
b. The SAPRIN will encourage Nodes to utilise the data repository to produce research outputs supporting the SAPRIN research agenda.

11. Capacity development

a. HDSS Nodes will each be part of or in partnership with a South African tertiary educational institution and will embed postgraduate research training and early career mentoring in their research agenda, thus contributing to scientific career development and research leadership.
b. Nodes will be required to employ entry-level staff, i.e. field staff, and some grades of data and admin staff, from their research populations and have a programme of in-house skills development.
c. SAPRIN and the HDSS Nodes will promote science awareness, appreciation and learning in local schools as part of an active public engagement programme.

12. Budget

d. SAPRIN budget is provided by DSI in 3-year circles through a contract with SAMRC, with an annual transfer
e. In turn, SAMRC has a contract with each of the HDSS Nodes within SAPRIN
f. Each node is contracted for funding of a maximum population of 100,000 individuals with an indication of the SAPRIN core protocol activities to be implemented in each year. The budget is

proportionally adjusted if the nodal population is less than 100,000 or some activities are not implemented.

g. The budget costs are activity-driven.
h. The budget is calculated based on 45 consecutive weeks per calendar year data collection schedule
i. The current 3-year Medium Term Framework (MTEF) budget (2024-2026) is shown in Table 2 below.

Table 2: SAPRIN Budget (2021-2024)

Component	Apr 21- Mar 22	Apr 22-Mar 23	Apr 23-Mar 24	Total
	YR7	YR8	YR9	YR7-9
Agincourt Node	R10 100 000	R10 050 000	R10 595 754	R30 745 754
AHRI Node	R10 100 000	R10 050 000	R10 595 754	R30 745 754
Dikgale Node	R10 100 000	R10 050 000	R10 595 754	R30 745 754
GRT-INSPIRED	R11 280 000	R11 100 000	R11 702 773	R34 082 773
C-SHARP	R11 280 000	R11 100 000	R11 702 773	R34 082 773
USINGA	R11 280 000	R11 100 000	R11 702 773	R34 082 773
Node 7	R5 500 000	R10 050 000	R10 595 754	R26 145 754
SAPRIN Management Hub	R5 392 580	R5 584 339	R5 863 556	R16 840 475
MRC (5% overheads)	R3 751 629	R3 954 217	R4 167 745	R11 873 591
Total	R78 784 209	R83 038 556	R87 522 638	R249 395 403
VAT Inclusive	**R90 601 840**	**R95 494 339**	**R100 651 034**	**R286 747 213**

13. References

1. Shivayogi P. Vulnerable population and methods for their safeguard. Perspectives in Clinical Research. 2013;4(1):53-7.

2. Statistics South Africa. The South African MPI: Creating a multidimensional poverty index using census data. Pretoria: Statistics South Africa; 2014.

3. Department of Science and Technology. A South African Research Infrastructure Roadmap Framework. 2013.

4. Kabudula CW, Houle B, Collinson MA, Kahn K, Gómez-Olivé FX, Tollman S, 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. 2017;5:e924-35.

5. Kabudula CW, Houle B, Collinson MA, Kahn K, Tollman S, Clark S. Assessing Changes in Household Socioeconomic Status in Rural South Africa, 2001–2013: A Distributional Analysis Using Household Asset Indicators. Social Indicators Research. 2016;127(3).

6. Academy of Science of South Africa (ASSAf). Consensus study on the state of the Humanities in South Africa: Status, prospects and strategies.; 2011.

7. Department of Higher Education and Training. Charter for Humanities and Social Sciences Report. 2011.

8. Department of Science and Technology. Innovation Towards a Knowledge-based Economy, Ten-Year Innovation Plan for South Africa (2008-2018). 2008.

9. Department of Science and Technology. Human and Social Dynamic in Development Grand Challenge Science Plan. 2010.

10. Tanser F, Hosegood V, Bärnighausen T, Herbst K, Nyirenda M, Muhwava W, et al. Cohort Profile: Africa Centre Demographic Information System (ACDIS) and population-based HIV survey. International Journal of Epidemiology. 2008;37(5):956-62.

11. Kahn K, Collinson MA, Gómez-Olivé FX, Mokoena O, Twine R, Mee P, et al. Profile: Agincourt Health and Socio-demographic Surveillance System International journal of epidemiology. 2012;41(4):988-1001.

12. Marianne Alberts SAD, Solomon Choma, Perpetua Modjadji, Felistas Mashinya, Sandra Burger, Ian Cook, Sanette Brits, Peter Byass. Health & Demographic Surveillance System Profile: The Dikgale Health and Demographic Surveillance System. International journal of epidemiology. 2015;44(3):1-7.

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

14. Sankoh O. Why population-based data are crucial to achieving the Sustainable Development Goals. International journal of epidemiology. 2017;46(1).

15. Sankoh O, Arthur S, Nyide B, Weston M. The history and impact of HIV&AIDS. A decade of INDEPTH research. HIV & AIDS Review. 2014;13(3):78-84.

16. Sankoh O, AbouZahr C, Adami H, Tollman S, Byass P, Tanner M. Universal health coverage and reliable global health estimates. Lancet. 2013;382(9886).

17. F Xavier Gómez-Olivé LM, Ryan G Wagner, Chodziwadziwa W Kabudula, Julia K Rohr, Kathleen Kahn, Till Bärnighausen, Mark Collinson, David Canning, Thomas Gaziano, Joshua A Salomon, Collin F Payne, Alisha Wade, Stephen M Tollman, Lisa Berkman. Cohort Profile: Health and Ageing in Africa: A Longitudinal Study of an INDEPTH Community in South Africa (HAALSI). International journal of epidemiology. 2018;47(3):689-90j.

18. Malaza A, Mossong J, Barnighausen T, Newell ML. Hypertension and obesity in adults living in a high HIV prevalence rural area in South Africa. PloS one. 2012;7(10):e47761.

19. World Health Organization. The WHO STEPwise approach to Surveillance of noncommunicable diseases (STEPS). Geneva: World Health Organization; 2003.

20. World Health Organization. Neonatal and perinatal mortality: country, regional and global estimates. Geneva2006.

21. Department of Health RoSA. National HIV counselling and testing policy guidelines2015.

22. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. Jama. 2013;310(20):2191-4.

23. ICH Harmonised Tripartite Guideline. Guideline for good clinical practice1996.

24. Department of Health RoSA. Ethics in health research: principles, processes and structures. Second edition ed. Pretoria: Department of Health; 2015.

25. Strode A, Slack C. Selected ethical-legal norms in child and adolescent HIV prevention research: Consent, confidentiality and mandatory reporting. Pietermaritzburg: HAVEG; 2014.

14. Appendices

14.1. Appendix A: List of the types of informed consent forms to be developed by each HDSS node.

Type	Purpose of Form
1. Household Component Information and Consent form	▪ **To provide information about the household data collection component of the study and obtain consent of household proxy respondent** to complete household questionnaires, conduct telephonic data collection
2. Individual component Information and Consent form	▪ **To provide information about the population-based data collection activities and obtain consent of all eligible individuals, including parental consent for all participants aged 15-17 years and assent from the adolescents themselves. Individuals will be asked to consent to:** answer general heath questions, give a research blood specimen, future research using specimens (when approved by research ethics committee), link records (including of own children) from Departments of Health, Social Development (social grants), Basic Education and Home Affairs to population heath research records, be contacted in future, blood pressure, height and weight measurements, a rapid HIV test and to be linked into care.
3. Clinic Attendee component information and consent form	▪ **To provide information about the clinic-based data collection and consent for everyone attending a department of health clinic within the study area. Individuals will be asked to consent to** give personal identifying information, reason for attendance, and to be linked to population-based study information given in the past.
4. Verbal autopsy component information and consent form	▪ **To provide information about data collection and consent for recording of information on causes of death. For each death, an identified person most knowledgeable about the deceased's terminal illness or accident will be asked to consent to** give details on the signs, symptoms and circumstances of the death.

14.2. Appendix B: Core Data Elements

These data elements have been jointly developed into standard question wording and coding, are agreed to by all SAPRIN HDSS Nodes, and will be submitted to the node specific ethical approval process.

Item	Notes
A. Base Entities	The core entities on which longitudinal data will be collected. Each entity has a dated start and end event
1. Location	A place to which geo-coordinates can be assigned

a. Type	i. Residential . A place of residence within the boundaries of the demographic surveillance area a. Flats b. Informal dwelling c. formal dwelling d. cluster housing e. hostel ii. Health Service- facility offering health services a. Clinic b. Health Centre c. Hospital d. Drop-in centre e. Place of safety f. Alternative therapy g. Traditional healers h. Hospice iii. Personal Services a. Salon b. Nail bar c. Health and beauty d. Spa e. Reflexology iv. Retail a. Shops i. Shopping mall ii. Retail outlets for clothing iii. Housewares iv. Showrooms v. Electronic and domestic appliances vi. Communications equipment vii. Spaza/Tuckshop viii. Hardware b. Food store i. Butchery ii. Bakery iii. Grocery iv. Supermarket v. Convenience Store c. Pharmacy v. Entertainment a. Social i. Restaurants ii. Bars/Shebeens/Taverns iii. Takeaways iv. Clubs v. Cinemas b. Sports i. Gymnasia ii. Sports clubs iii. Swimming pools iv. Stadiums vi. Industrial a. Factory (manufacturing of all kinds)

	b. Workshop i. Repair ii. Maintenance vii. Offices a. Commercial space b. Professional services (e.g. engineers, solicitors, etc) viii. Education a. Creche b. Primary school c. Secondary school d. Tertiary College e. University ix. Travel a. Transport i. Petrol station ii. Bus and taxi stand iii. Parking garage b. Short stay residential i. Bed and Breakfast ii. Hotel iii. Campsite iv. Lodge/Guesthouse x. Public Services a. Government Services i. Police ii. Social Services iii. Licensing iv. Crematoria v. Post Office b. Parastatal i. Water ii. Electricity iii. Refuse disposal iv. Correctional services c. Community services i. Place of worship ii. Library iii. Community hall d. Stadium (non-specific public gathering) xi. Outbuilding a. Garage b. Carport c. External toilet d. Shed xii. Inaccessible a. Claimed b. Captured c. Derelict xiii. Open Space a. Park b. Public open space c. Building site d. Playground

	xiv. Other (specify)
b. State	i. Under construction ii. Usable iii. Broken down
c. Functional	i. Yes ii. No
d. Area	Village or other higher level spatial grouping of locations
e. Coordinate	Latitude and longitude on the WGS84 coordinate system
2. Household	A household resident at a location within the boundary of the surveillance area. A household is defined as a social group of one or more individual members. Resident household members share the same location as the household, non-resident household members are individuals who do not share the same location as the household, but who eat from the same pot when they are present in the household's location.
a. Identifier	A unique household identifier which the household retains irrespective of its current residence
3. Individual	Household members, where a **resident** is a member of a household who normally lives (i.e. intends to sleep the majority of the time) at the same location as the household; and a **non-resident** is a member of a household who does not normally live at the same location as the household but is nevertheless considered a member of the household.
a. Identifier	A globally unique identifier used internally to uniquely identify the individual
b. External identifier	An identifier used on data collection instruments to uniquely identify the individual, but is removed together with other personally identifiable information when data is made available for analysis
c. Surname	Family or last name
d. Alternative surname	An alternative surname (formerly) used by the individual, e.g. maiden name
e. First name 1	The first given name of the individual
f. First name 2	The second/alternative given name of the individual
g. Civilian ID number	The 13-digit South African civilian identity number
h. Citizenship	The country of which the individual is a citizen
i. Nationality	The country of origin of the individual (the country of which the person is/was a citizen by birth)
j. Contact number	Cellphone or landline
k. Sex	Male (1), female (2) or unknown (0)
l. Birth	The Birth event of this individual or the DeliveryEvent resulting in the birth of this individual if the individual enters surveillance through birth
m. EndEvent	The death event of this individual or the Observation at which this individual was last observed
n. MotherId	The unique identifier of the mother of the individual, if the mother is enumerated in the surveillance
o. FatherId	The unique identifier of the father of the individual, if the father is enumerated in the surveillance
p. MotherSurname	The surname of the mother of the individual
q. MotherFirstName	The first name/s of the mother
r. FatherSurname	The surname of the father of the individual

s. FatherFirstName	The first name/s of the father
B. Events	Demographic events that delineate episodes of observation during longitudinal surveillance
1. Migration	A change in location of usual residence
a. Date	Date of migration
b. Type	i. Internal (change of location within surveillance area) ii. External (change of location into/out of the surveillance area)
c. Direction	i. In (migration into this location) ii. Out (migration out of this location)
d. Origin	Origin of the migration (location)
e. Destination	Destination of the migration (location)
f. Unit	Individual or household
2. Birth	The birth event of an individual
a. Date	Date of birth
b. Birth weight	Birth weight as obtained from the Road to Health card in grams
c. BirthCertificate	Whether the birth certificate was observed
d. DateRegistered	The date on which the birth was registered, or null if the birth hasn't been registered yet
3. PregnancyOutcome	The end event of a pregnancy
a. Date	Date of birth
b. Type	i. Spontaneous abortion ii. Assisted abortion iii. Caesarean iv. Assisted v. Normal vaginal Types i-ii is only applicable if the pregnancy end prior to the 28th week of pregnancy, types iii-v is applicable only on or after the 28th week of pregnancy
c. Place	i. In a health care facility ii. Outside a health care facility, but not at home iii. At home
a. BirthAttendant	i. Doctor ii. Midwife iii. Traditional birth attendant iv. Lay person
b. LiveBirths	The number of babies born alive resulting from this delivery
c. Stillborn	The number of still born babies resulting from this delivery
4. Death	The death of an individual
a. Date	The date of the death
b. Place	i. In a health care facility ii. Outside a health care facility, but not at home (road, etc) iii. At home
c. DeathCertificate	Observed
d. Verbal autopsy	WHO standard verbal autopsy questionnaire items. SAPRIN will adopt the worldwide standardised World Health Organisation (WHO) Verbal Autopsy questionnaires to determine the cause of deaths reported in the surveillance population. The latest version of the WHO Verbal Autopsy Questionnaire is 2016, edited and cognitively tested to facilitate the use of publicly available analytical software for assigning the cause of death.

5. Enumeration	The initial event at enumeration of an entity. Can only be used during the base census of the demographic surveillance, or when a new area is added to the surveillance
a. Date	Date of enumeration
6. Membership	A change in household membership
a. Date	The date of the change
b. Type	i. Start ii. End iii. Household dissolution
7. Household Headship	A change in the head of the household
a. Date	The date of the change
b. Type	i. Start ii. End
8. UnionEvent	A change in the union (conjugal relationship) between two individuals
a. Date	The date of the event
b. Type	i. Start ii. Marriage iii. Separation iv. Divorce v. Partner died
9. Observation	A surveillance visit at a location
a. Date	Visit date
b. Location	Location where the visit took place
c. DataCollector	The person responsible for data collection during the visit
d. Respondent	The primary respondent at the visit
C. Episodes	Used to record associations longitudinally between base entities. Episodes are always started and ended through events.
1. Residence	An episode during which an individual is resident (sleeping most of the time there) at a particular location that falls within the surveillance area. An individual can only be resident at one location at a time, i.e. residency episodes cannot overlap
a. Individual	The individual identifier
b. Location	The location identifier
c. StartEvent	The event that started the episode, can only be Enumeration, Birth, Migration (Direction: In)
d. EndEvent	The event that terminated the episode, can only be Death, Migration (Direction: Out) and Observation (in which the implication is that the individual is last known to be resident)
2. HouseholdResidence	An episode during which a household is resident at a particular location that falls within the surveillance area. A household can only be resident at one location at a time, i.e. residency episodes cannot overlap
a. Household	The household identifier
b. Location	The location identifier
c. External identifier	The identifier associated with the household during this residence
d. StartEvent	The event that started the episode, can only be Enumeration, Household formation, Migration (Direction: In)
e. EndEvent	The event that terminated the episode, can only be Household dissolution, Migration (Direction: Out) and Observation (in which the implication is that the household is last known to be resident)

3. Membership	An episode during which an individual is a member of a household. An individual must be a member of at least one household to be under surveillance. An individual can be a member of more than one household at a time (membership episodes may overlap). In the case of multiple memberships, the designated household of an individual will be the household the individual is co-resident with, ranked according to the closeness of the members relationship to the household head (self, spouse, child, grandchild, parent, sibling, other relationship)
a. Individual	The individual identifier
b. Household	The household identifier
c. StartEvent	The event that started the episode, can only be Enumeration, Birth or Membership start
d. EndEvent	The event that ended the episode, can only be Death, Membership end or Observation (implying that the membership is current)
4. Household Head Relationship	Household head relationships are linked to a household membership and record the relationship between the individual and the head of the associated household. If the head of household change, all current household members start a new household head relationship episode
a. Membership	The household membership episode associated with this household head relationship
b. Relationship	The relationship between the individual member (to whom the membership belongs) and the current household head of the household i. Self (the individual who is the subject of the membership is the household head) ii. Spouse (incl partner in stable relationship) iii. Child (incl adopted/foster child) iv. Son/daughter-in-law (incl individuals in stable relationship with any child of the household head) v. Grandchild vi. Parent vii. Parent-in-law (incl parent of partner in stable relationship) viii. Grandparent ix. Other relative x. Domestic worker or tenant xi. Unrelated household member
c. StartEvent	The event that started the episode. Can only be Enumeration, Birth, Membership start, or Household Headship start
d. EndEvent	The event that ended the episode. Can only be Death, Membership end, Household Headship end or Observation
5. Union	The episode during which two persons are in an informal or formal conjugal relationship. For a conjugal relationship to exist the following factors should be considered: i. Shelter. Do the partners live under the same roof? ii. Sexual and personal conduct. Do the partners have sexual relations; do they maintain an attitude of fidelity to each other; do they eat their meals together? iii. Services. Do they share household responsibilities? iv. Social. Do they participate together in social activities; does their society recognise them as a couple? v. Support. Do they support each other financially? vi. Children. Do they have children together?
a. Individual1	The individual identifier of one of the parties to the union. By convention this will be the female in a heterosexual union.
b. Individual2	The individual identifier of the second party to the union.
c. StartEvent	The start of the union. Can only be Enumeration or Union start

d. MarriageDate	The date on which the union has been formalised as a marriage
e. EndEvent	The end of the union. Can only be Union – partner died, Union – separation (if there is no marriage date), Union – divorce (if there is a marriage date), or Observation if it a current union
6. Pregnancy	The period of being pregnant. Also used to record maternity histories retrospectively
a. Woman	The individual identifier of the woman who experienced the pregnancy
b. ANCVisits	The number of antenatal care visits during this pregnancy
c. Duration	Duration in weeks of the pregnancy
d. Outcome	A delivery event or Observation if the pregnancy is still current at the time of observation
7. Social support	The period during which an individual receives a government social support grant
a. Grantholder	The individual identifier of the person holding the grant
b. Beneficiary	The individual identifier of the intended beneficiary of the grant. This may be the same as the identifier of the grant holder in the case where the holder is the beneficiary, e.g. old age pension, or different as in the case of care dependency grants, where the grant holder receives the grant on behalf of someone else, e.g. a child
c. Type	The type of grant i. Care dependency ii. Child support iii. Covid Relief Grant iv. Disability v. Foster care vi. Grant-in-aid vii. Old age Pension viii. War veterans ix. Social relief of distress x. Tertiary Education Grant xi. None xii. Don't know xiii.
d. StartDate	The start date of the social support
e. EndDate	The end date of the social support
D. StatusObservation	Information collected at a particular observation, valid only at the time of the observation. State may be imputed between consecutive status observations but is not known to be valid. This is in contrast with episode where the assertion is that the state represented by the episode is valid for the duration of the episode.
1. Individual StatusObservation	A set of data elements collected about an individual either during a face-to-face visit or telephonic interview with the individual or from a proxy informant
a. Individual	The unique individual identifier of the subject of the status observation
b. Observation	The observation at which the status observation was made
c. ResidentStatus	Physical presence in the location, recorded as the number of months since the previous observation visit

d. MotherStatus	i. Same household ii. Same area (village/isigodi) iii. Elsewhere in surveillance area iv. In the immediate surroundings outside surveillance area v. Elsewhere vi. Died vii. Unknown status
e. FatherStatus	i. Same household ii. Same area (village/isigodi) iii. Elsewhere in surveillance area iv. In the immediate surroundings outside surveillance area v. Elsewhere vi. Died vii. Unknown status
f. HighestSchoolLevel Completed	Grade 1 – 12
g. HighestNonSchool Education	i. Undergraduate degree ii. Postgraduate degree iii. ABET 1-4 iv. NQF 1-4
h. CurrentEducation	If the individual is currently attending an educational institution, at what level: i. Creche ii. Pre-school iii. Grade 1-12 iv. ABET 1-4 v. NQF 1-6 /N1-N6/NTC 1-6 vi. Undergraduate degree vii. Post-graduate degree viii. Not attending

i. If 5 yrs and below, childcarer	Name of childcarer Child attend: i. Grade R ii. Preschool/nursery school/Grade 00/Grade 000 iii. Creche/ educare centre iv. Child is cared at home [Go to question 5.4 below] v. Home/community play group [Go to question 4 below] vi. None [Go to questions 5.3 and 5.4 below] Not attending, where child spend most time in day i. At home with parent, foster parent or guardian. ii. At home with another adult iii. At home with someone younger than 18 years iv. At someone else's dwelling v. Other (specify) Main reason child cared at home i. Prefer that the child stays at home/with someone else. ii. These facilities do not exist in our area. iii. Too expensive. iv. Other
j. Currently employed	i. Yes ii. Part-time iii. No
k. Not employed	Type/reason for unemployment i. Awaiting the season for work ii. Caring for others/household duties iii. Looking for work iv. Student or in training v. Health reasons vi. Disabled or unable to work (handicapped) vii. Pregnancy viii. Retired or too old ix. Don't want to work x. Other reason (specify_____) xi.

l. EmploymentSector	i. Agriculture/Fishing/Forestry ii. Mining iii. Manufacturing iv. Electricity and water v. Construction vi. Wholesale/retail vii. Restaurant/Hotels/Sport/Tourism viii. Transport and communication ix. Finance x. Educational services xi. Health services xii. Legal services xiii. Research xiv. Domestic services xv. Armed forces xvi. Informal sector, e.g. street vendor xvii.
m. EmploymentType	i. Self-employed ii. Employee iii.
n. Employer	i. Central government ii. Provincial government iii. Local / regional authority iv. Public corporation v. Private sector employer vi. Non-profit institution vii.
o. FinancialStatus	Self-reported financial status i. Very Comfortable ii. Comfortable iii. Just Getting By iv. Poor v. Extremely Poor
p. MaritalStatus	i. Not married ii. Monogamous Marriage iii. Polygamous Marriage iv. Divorced/Separated v. Widowed
q. PartnershipStatus	i. Marital Partnership ii. Regular Partnership iii. Casual Partnership(s) iv. No Partnership
r. HealthStatus	Self-reported health status i. Very Good ii. Good iii. Moderate iv. Bad v. Very Bad

s. Little interest or pleasure in doing things in last 2 weeks	i. Not at all ii. Several days iii. More than half the days iv. Nearly every day
t. Feeling down, depressed, or hopeless in last 2 weeks	i. Not at all ii. Several days iii. More than half the days iv. Nearly every day
u. Ever been admited	Government or private hospital (Yes/No) Visted out-patient dept at govt or private hospital (Yes/No) Visted govt health clinic or private doctor (Yes/No) Visited pharmacy/chemist (Yes/No) Visted traditional healer (Yes/No)
v. Tuberculosis	i. Ever treated - Yes/No ii. Treatment started in last 12 months – Yes/No iii. Currently on TB treatment – Yes/No
w. HIV	i. Ever received a test result for HIV - Yes/No ii. Ever had a positive HIV result – Yes/No
x. If HIV+	i. When first HIV+ result (>1yr, <1yr ago) ii. When last HIV- result (>1yr, <1yr ago) iii. When first started ART (Never, <1yr, >1yr) iv. Currently on ART (Yes/No)
y. If HIV-	i. When last HIV- result (>1yr, <1yr ago)
z. HIVResult	HIV serostatus from dried blood spot
aa. Hypertension	i. Ever treated - Yes/No ii. Treatment started in last 12mos iii. Currently on treatment
bb. Diabetes	i. Ever treated - Yes/No ii. Treatment started in last 12mos - Yes/No iii. Currently on treatment - Yes/No
cc. Health care utilisation	i. Admitted to hospital past month - Yes/No ii. Visited a clinic past month – Yes/No iii. Visited private doctor past month – Yes/No iv. Used pharmacy/chemist past month – Yes/No v. Visited traditional healer past month – Yes/No
dd. Vaccination history (child <=6yr)	For each vaccination record: date received and source of information (Road to Health Card or recall) i. At birth – BCG & Polio 0 ii. At 6w – Polio1, DTab+IPV+HiB1, HepB1, Rota1, PCV1 iii. At 10w – DtaP+IPV+Hib2, HepB2 iv. At 14w – DtaP+IPV+Hib3, HepB3, Rota2, PCV2 v. At 9mos – Measles1, PCV3 vi. At 18mos – DtaP+IPV+Hib4, Measles2 vii. At 6yr - DT

ee. Adult vaccination	Ever vaccinated i. Tetanus and diphtheria (and pertussis) (often around age 12 or during pregnancy) ii. Human Papilloma virus (often around age 9). If queried: "the HPV vaccine is an injection given in the thigh or the upper arm when you are in Grade 4 at school as a protection against cervical cancer" iii. Hepatitis B (sometimes given during pregnancy) iv. Influenza v. Covid-19 vi. Other, specify vii. None Vaccinated past 12 months i. Tetanus and diphtheria (and pertussis) ii. Human Papilloma virus (HPV) iii. Hepatitis B iv. Influenza v. Covid-19 vi. Other, specify Agree to be vaccinated (or get vaccinated again) for Covid-19 i. Definitely would ii. Probably would iii. Probably would not iv. Definitely would not Reasons to choose to get [another] vaccine for Covid-19 i. To protect others in my family ii. To protect my community from COVID-19 iii. To protect myself from getting sick with COVID-19 iv. I have a chronic health condition v. If my doctor, nurse, or health worker recommends it vi. It would be the best way to avoid getting seriously ill from COVID-19 vii. It would allow me to feel safe around other people viii. Life won't go back to normal until most people are vaccinated ix. Other, specify x. None Reasons to choose NOT to get [another] vaccine for Covid-19 i. I am allergic to vaccines ii. I don't like needles iii. I don't get vaccines in general iv. People in my community do not get vaccines in general v. I'm not concerned about getting ill from the coronavirus vi. I would be concerned about getting infected with the coronavirus from the vaccine vii. I am concerned about side effects from the vaccine viii. I don't think that the vaccine will work ix. The coronavirus outbreak is not as serious as some people say it is x. It might be expensive xi. It will be too late because we will have already been infected xii. I expect vaccination site will require long/expensive travel

	xiii. I expect vaccination site will be open inconvenient hours/require long wait times. xiv. I expect that the vaccination site would not have vaccines. xv. Other, specify. xvi. None Reasons that would make you more likely to get [another] vaccine for Covid-19 i. If my employer required me to get one ii. If I needed to have a vaccine to access health services iii. If I needed to have a vaccine to get into shops and restaurants iv. If I could get paid time off to get vaccinated v. If the government made a vaccine mandatory vi. If I didn't have to go too far to get the vaccine vii. If the vaccine was proven to be safe viii. If it was recommended to me by a religious or community leader ix. To stop following COVID-19 guidelines x. If it was available through a regular vaccination campaign xi. If childcare was available whilst I get my vaccine xii. If there was an incentive (e.g. food, phone credit) xiii. If I could choose the brand of vaccine xiv. Nothing would make me more likely to get a COVID-19 vaccine xv. Other, specify Where would you prefer to get [another] vaccine for Covid-19 i. Hospital ii. Health centres iii. Community centre/meeting hall/local shop/hall iv. Workplace v. Pharmacy vi. Mobile vaccination centre vii. Dedicated vaccination centre viii. Place of worship ix. None, I would not get a COVID-19 vaccine anywhere x. Other, specify
2. Household Status Observation	Set of data elements collected from a household informant, during a face to face or telephonic interview

a. Water source	The most commonly used (during last year) source of drinking water i. Piped (tap) water in the dwelling/house ii. Piped (tap) water in yard iii. Borehole in yard iv. Rain-water tank in yard v. Neighbour's tap vi. Public/communal tap vii. Water-carrier/tanker viii. Water vendor ix. Borehole outside yard x. Flowing water/stream/river xi. Stagnant water/dam/pool xii. Well xiii. Spring xiv. Other (specify) xv. xvi. Distance of water source from dwelling/house if not in yard i. Less than 200 metres ii. 201-500 metres iii. 502 metres- 1kilometre iv. More than 1 kilometre v. Don't know
b. Toilet	What kind of toilet does the household use i. Flush toilet connected to a public sewerage system ii. Flush toilet connected to a septic or conservancy tank iii. Pour flush toilet connected to a septic tank (or septage pit) iv. Chemical toilet v. Pit latrine/toilet with ventilation pipe (VIP) vi. Pit latrine/toilet without ventilation pipe vii. Bucket toilet viii. Ecological sanitation systems (e.g. urine diversion) ix. Open defecation (e.g. no facilities, field, bush) x. Other (specify) xi. xii.
c. Electricity supply	Is the household connected to the electricity grid? i. Yes, interconnected grid electricity ii. Yes, off grid (generator, solar, other electricity sources) iii. No iv.

d. Cooking fuel	What is the main fuel used for cooking? i. Electricity from grid ii. Solar energy iii. Generator energy iv. Gas (LPG) v. Paraffin vi. Wood vii. Coal viii. Animal dung ix. None x. Other, specify. xi. Alternative cooking energy during loadsheding i. Solar energy ii. Generator energy iii. Gas (LPG) iv. Paraffin v. Wood vi. Coal vii. Animal dung viii. None ix. Other, specify. Alternative lighting energy during loadsheding i. Solar energy ii. Generator energy iii. Gas (LPG) iv. Paraffin v. Candles vi. Battery operated light vii. None viii. Other, specify. Main alternative source of energy during loadshedding if grid electricity is main source i. Generator ii. Inverter operated from batteries charged when there is grid electricity. iii. Inverter operated from batteries charged from solar panels and/or wind. iv. Battery operated equipment or power bank v. None vi. Other, specify.

e. Location construction	What is the construction materials of the walls? i. Brick ii. Cement block/concrete iii. Corrugated iron/zinc iv. Wood v. Plastic vi. Cardboard vii. Mud and cement mix viii. Wattle and daub ix. Tile x. Mud xi. Thatching/ grass xii. Asbestos xiii. Other (specify What is the construction materials of the roof? i. Cement block/concrete ii. Corrugated iron/zinc iii. Wood iv. Plastic v. Cardboard vi. Mud and cement mix vii. Wattle and daub viii. Tile ix. Mud x. Thatching/ grass xi. Asbestos xii. Other (specify What is the construction material of the floor? i. Earth/Sand ii. Dung iii. Wood/ Planks iv. Parquet/ polished wood v. Vinyl or asphalt strips vi. Ceramic tiles vii. Cement viii. Carpet Other (specify How many bedrooms does your household occupy at this location?
f. Assets	Does the household have any of the following items in good working order?

	i. Television set ii. Swimming pool iii. DVD player/ Blu ray player iv. Pay TV (M-Net/DSTV/ Top TV) Subscription v. Air ci\conditioner (excluding fans) vi. Computer/Desktop/ laptop vii. Vacuum cleaner/ floor polisher viii. Dish washing machine ix. Tumble dryer x. Fridge/Freezer xi. xii. Electric Stove/ Gas Stove xiii. Microwave Oven xiv. Built-in Kitchen sink xv. Home security service xvi. Home theatre system/ Hifi/Stereo xvii. Geyser (providing hot water) xviii. Landline telephone xix. Cellphone smartphone xx. Livestock xxi. Car or Bakkie xxii.
g. Household access to internet	i. From a network at home ii. From household members' cellphone(s) iii. From a network at work iv. From elsewhere (e.g. internet café) v. No access to internet
h. Household agricultural activity	i. Livestock production (cattle goats, sheep, pigs, etc) ii. Poultry production (chickens, ducks, geese, guinea fowl, ostrich, etc) iii. Vegetable production iv. Production of crops (grains, fruits, sugarcane, etc) v. Fodder grazing /pasture/grass for animals vi. Other vii. None
i. Crime	Has any resident member of the household been a victim of any of these crimes in the past 12 months? i. None ii. Theft iii. Assault iv. Murder v. Other crime
j. Financial situation	How would the household classify its financial situation these days? i. Very comfortable ii. Comfortable iii. Just getting by iv. Poor v. Extremely poor

k. Food security	Last 12 months: Worry about not enough i. Yes ii. No iii. Don't know Unable to eat healthy and nutritious i. Yes ii. No iii. Don't know Ate only a few kinds of foods i. Yes ii. No iii. Don't know Skipped a meal i. Yes ii. No iii. Don't know Ate less than thought i. Yes ii. No iii. Don't know Ran out of food i. Yes ii. No iii. Don't know Hungry but not eat i. Yes ii. No iii. Don't know Didn't eat all day i. Yes ii. No iii. Don't know Children skipped or cut meal size i. Yes ii. No iii. Don't know