Request for Expressions of Interest (REOI) in the Establishment of Additional Health and Demographic Surveillance Nodes in the South African Population Research Infrastructure Network (SAPRIN)

***Issue Date:** 27 May 2019*
***Submission Deadline:** 18:00 30 June 2019*

TABLE OF CONTENTS

1 INTRODUCTION

The **South African Population Research Infrastructure Network** (SAPRIN)[1] is issuing this Request for Expression of Interest to gauge the interest of stakeholders in the expansion of its current network of three health and demographic surveillance system (HDSS) nodes, with urban nodes in Gauteng province, Cape Town and eThekwini metropolitan areas, and a rural node in the Eastern Cape.

SAPRIN is funded by the Department of Science and Technology (DST) as part of the South African Research Infrastructure Road Map (SARIR) and hosted by the South African Medical Research Council (SAMRC).

HDSS nodes implement a standard protocol for the prospective and longitudinal collection of population, health and socio-economic data in geographically-defined sections of impoverished and developmentally-constrained communities, both rural and urban. Since its inception in 2016, SAPRIN has integrated the three existing HDSS nodes in South Africa, namely, MRC/Wits University **Agincourt HDSS** in Bushbuckridge District, Mpumalanga, established in 1992 with a current population of 115,000; University of Limpopo **DIMAMO HDSS** in the Capricorn District of Limpopo, established in 1996, with a population of 35,000, currently expanding to 100,000; and the **Africa Health Research Institute** (AHRI) HDSS in uMkhanyakude District, KwaZulu-Natal, established in 2000, with a population of 125,000, linked to the University of KwaZulu-Natal.

SAPRIN is issuing this Request for Expressions of Interest (REOI) to gauge interest in and readiness of South African research organisations to establish and operate HDSS nodes consistent with the SAPRIN research protocol (Appendix A). Eligible research organisations or consortia must be associated with a South African university or national research council, have experience in population-based research and established links to the community where the node is to be located.

The **deadline** for submissions is **30 June 2019.** All submissions must be electronic and emailed to Nosicelo Mchiza - Nosicelo.Mchiza@mrc.ac.za.

An information webinar for interested parties will be held on **11 June 2019**. **Please RSVP by 12h00 on Friday, 31 May 2019 to** Nosicelo.Mchiza@mrc.ac.za.

2 THE OPPORTUNITY

South Africa is striving to emerge from **a legacy of gross social injustice** and consequent health and socio-economic inequality, to becoming a country where all citizens can access ever-broadening opportunities to build productive lives for themselves. But, stewardship of the country faces several challenges, including high to very high levels of inequality, with a Gini coefficient of 0.66; an unemployment rate of 27%; a poverty headcount ratio of 57%; as well as colliding epidemics of HIV, TB and non-communicable diseases. This underlines **the need for improved, science-based information and advice to direct development-oriented decision-making, investments and interventions**.

[1] http://saprin.mrc.ac.za/

Through the **South African Research Infrastructure Roadmap** (SARIR)[2], the Department of Science and Technology is establishing world-class scientific research infrastructure in a range of key sectors, including humans and society, where SAPRIN, the **South African Population Research Infrastructure Network**, was established. SARIR has been developed to facilitate a research infrastructure investment programme. 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 as outlined in the White Paper on Science, Technology, and Innovation[3].

The research infrastructure established by SAPRIN consist of a network of health and demographic surveillance system, each of which is a standardised, field-based information system engaged in the prospective and longitudinal collection of population, health and socio-economic data for geographically-defined sections of impoverished and developmentally-constrained communities, both rural and urban. Data and findings apply widely and are easily scalable. Individual and household indicators that are routinely collected and assessed include: **vital events, i.e. births and deaths (by cause), residence and migration, socio-economic status, disease monitoring, and measures of wellbeing represented by labour status, education and social protection.** The HDSS registration systems will be complemented by **linking to public sector records** of health system utilization, school performance and access to social grants, to enable research on the factors associated with access to services or lack thereof. The anticipated outcomes and impact of SAPRIN are:

1. **Regular Data Releases**: up-to-date, longitudinal data, representative of South Africa's fast-changing poorer communities for research, interpretation and calibration of national datasets conforming to the SAPRIN Data Access Policy (Appendix B).
2. **National Statistics Triangulation**: longitudinal SAPRIN data triangulated with National Census data for calibration of national statistics and studying mechanisms driving the national statistics.
3. **Interdisciplinary Research Platform:** an infrastructure for conducting observational and interventional scientific research, at population level, for researchers from universities, science councils, and other organisations, including regional and international collaborations.
4. **Policy Engagement:** providing evidence to underpin policy-making for cost evaluation and targeting intervention programmes, thereby improving the accuracy and efficiency of pro-poor, health and wellbeing interventions.
5. **Scientific Education**: training associated with SAPRIN research projects at related universities, providing expanded capability in science that is effectively linked with national, regional and international networks.
6. **Community Engagement**: coordinated engagement with communities to enable two-way learning between researchers and community members, maintain willingness to participate in research, ensure the dignity of research participants, enable research site communities and service providers

[2] https://www.dst.gov.za/images/Attachments/Department_of_Science_and_Technology_SARIR_2016.pdf
[3] https://www.dst.gov.za/images/2019/FINAL-White-Paper-to-Cabinet_11-March-2019.pdf

to have access to and make effective use of research results, and enable effective and appropriate community participation in guiding and informing research processes.

SAPRIN's impartiality and independence from vested interests makes it well positioned as a credible knowledge partner for collaborating with key government departments to strengthen policy and better inform decision-making on health and socio-economic matters. The scientific evidence generated by SAPRIN will potentially contribute to objective, informed evaluations of the implications of policy options from a science perspective. To fully realize the potential benefits of public investments in Research Infrastructures (such as SAPRIN), the aim of SAPRIN is to be seen as a source of trusted science advice (based on robust, peer-reviewed scientific evidence) to government and to civil society.

For SAPRIN to be an optimal national asset it needs to be part of a nexus of stakeholders: national public research infrastructure, core service ministries, and knowledge-building institutions. These three types of stakeholder have different roles to play in partnering with SAPRIN.

1. **National public research infrastructure**:
   This group of stakeholders provide research standards and priorities, and ensure high levels of public embeddedness and value for public policies.

    a. Department of Science and Technology (DST); South African Medical Research Council (SAMRC); Human Sciences Research Council (HSRC); Statistics South Africa (Stats SA); Department of Planning, Monitoring and Evaluation (DPME)

2. **Core service ministries and donors**
   These stakeholders will have two-way relationships with SAPRIN. Data on utilisation of public services needs to be linked to the SAPRIN databases in population nodes. Ministries providing public services will benefit from SAPRIN data to target new policies and evaluate the outcomes of policies aimed at improving the efficiency and impact of services and programmes.

    a. Health; Social Development; Home Affairs; Basic Education; Labour; Human Settlements
    b. International programme funders (UNAIDS, PEPFAR, etc)

3. **Knowledge-building institutions**:
   These stakeholders will use SAPRIN data to support research projects, with single- or multi-centre longitudinal population data. The platform can be extended by embedding research projects, using external study-specific sources of funding to conduct observational or interventional research.

    a. Pre-eminent scientists in Universities and Centres of Excellence
    b. Their doctoral students and research officers
    c. International science funders (Wellcome Trust, Gates Foundation, NIH, etc)

2.1 Funding

SARPIN will fund participating nodes to implement the standard SAPRIN surveillance protocol, but nodes are encouraged to seek co-funding to extend the protocol and source external funding for research studies hosted on the population platform so established (see figure 1). The SARIR vision is the long-term funding (10-15 years) of research infrastructure, but typically SAPRIN nodes receive funding contracts for three years of operation at a time.

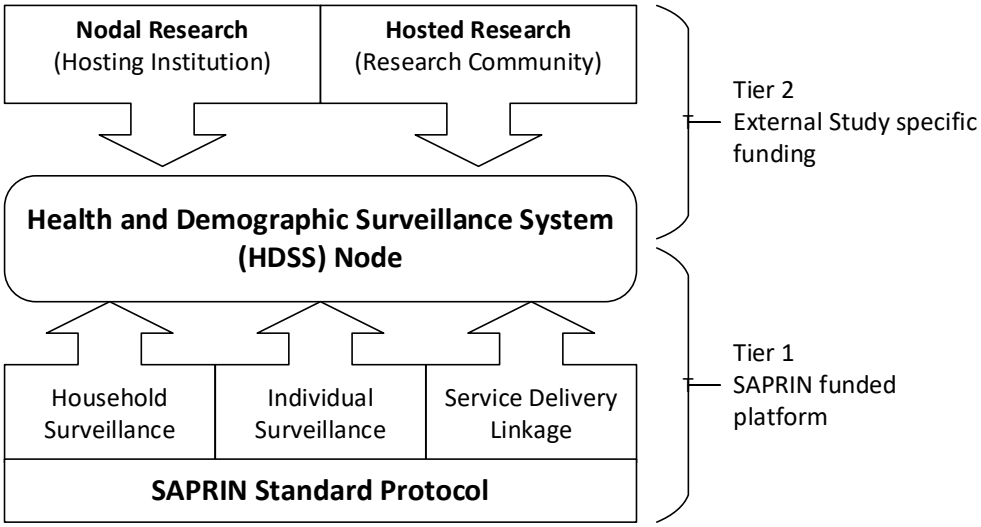

Figure 1 : SAPRIN Nodal Funding Model

2.2 Eligibility Criteria

SAPRIN seeks expression of interest from eligible research organisations or consortia to implement and operate SAPRIN nodes to be established in the Gauteng province, the cities of Cape Town and eThekwini metropolitan areas and in a rural or peri-urban area in the Eastern Cape province. There is no precondition of exactly where in these broad areas the node should be situated, except that it should comprise of geographically well-defined communities with a surveillance population of approximately 100,000. For a precise definition of the surveillance population see the attached SAPRIN protocol.

To be eligible to respond, a research organisation or consortium must fulfil the following criteria:

1. Be a legal entity (or part of a legal entity, or in the case of a consortium, a designated institution that will act as the legal entity on behalf of the consortium) in associated with or part of a South African university or national research council,
2. Possess a track record of conducting community-based research,
3. Possess an existing relationship with and access to the communities where they intend to base the node,
4. Express a vision and strategy for the population-based research that will be hosted by the node.
5. Have the experience and capability to manage the complex data management and field operations associated with longitudinal population surveillance

2.3 Implementation Schedule

The establishment of the additional SAPRIN nodes will commence in 2020 and conclude in 2023, with approximately one node established each year. The sequence in which the nodes will be established will depend on the information gathered by this REOI.

2.4 REOI Process

An **information webinar** will be arranged on **11 June 2019** after the release of this REOI where interested parties can interact with SAPRIN staff to address any additional information required to aid in their response to the REOI.

The information gathered by this **REOI will not be used in selecting candidates for SAPRIN funding**. After assessment of the REOI submissions, SAPRIN will announce the process by which the entities responsible for implementing a particular SAPRIN node will be selected. This will involve further detail submissions by interested entities and an independent review process, that will be communicated in future requests for proposal. However, it is important for interested parties to submit the requested information, because it will guide us in structuring the requests for proposals and selection criteria.

Submissions should be lodged electronically on the SAMRC's electronic submission platform, no later than 18:00 on 30 June 2019.

3 SUBMISSION REQUIREMENTS

The submission should be structured as follows:

1. **Organisation** (1 page maximum)
   a. Name and contact information
   b. Affiliation with South African Universities or National Research Councils
   c. A brief history of the organisation, including years of operation
   d. The organisation's leadership

2. **Statement of interest and experience (2 pages max)**
   a. Why is the organisation interested in hosting a SAPRIN node?
   b. What is the organisation's experience in working on a similar project?
   c. What is the organisation's vision for the research to be hosted on the nodal infra-structure?
   d. What in kind contributions and special capabilities can the organization bring to SAPRIN and/or the node?

3. **Vision for the project specified in the SAPRIN protocol (3 pages max)**
   a. Which specific geographic location do you have in mind?
   b. What is your prior experience in conducting research in this location?
   c. Have all homesteads in this area been mapped and have geospatial coordinates?
   d. What is the extend of your community engagement in this area and what additional engagement will be required to establish nodal operations?

e. What is the nature of your relationship with local authorities in this area, including service delivery authorities, such as local health services?
f. What local physical infrastructure is available as a base for field operations and data collection?
g. How do you intend to manage electronic data collection and telephonic survey components of the protocol?
h. How do you intend to manage the biomeasures and do you have access to laboratory facilities for laboratory assays on the dried blood spots?
i. Highlight any aspect of the SAPRIN standard protocol that you envisage having difficulties to implement.

4 LIMITATIONS

4.1 Confidentiality

SAPRIN will not disclose any content or concept information regarding submissions to any third party other than SAPRIN officials and employers, members of the SAPRIN Steering Committee, and persons specifically engaged by SAPRIN to assist with assessing the responses. Any documented assessment of the EOI Responses will reveal neither the identity nor the affiliation of EOI Respondents.

4.2 Communication

Any queries relating to the EOI Process should be directed to the SAPRIN Director, Kobus Herbst, at the following e-mail address: **kobus.herbst@mrc.ac.za**

4.3 Reserved Rights

SAPRIN and the MRC reserve the following rights in respect of this EOI:

1. The right not to proceed with the implementation if additional SAPRIN nodes, or a subset thereof.
2. The right to change the schedule for node rollout,
3. The right to expand or restrict organisational participation in any subsequent requests for applications or proposals.
4. The right to request additional information from the REOI responders.

Appendix A : Study Protocol

South African Population Research Infrastructure Network (SAPRIN)

A National Research Infrastructure of Health and Demographic Surveillance System (HDSS) Nodes

Co-Principal Investigators:

Associate Prof Mark Collinson

Dr Abraham Jacobus Herbst

Scientific leadership in the HDSS Nodes:

Professors Stephen Tollman, Kathleen Kahn, Deenan Pillay, Marianne Alberts

Version 2.4

8 August 2017

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
DBS	Dried blood spots
Dikgale HDSS	Dikgale Health, and Demographic Surveillance System 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
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
WHO	World Health Organization
Wits	University of the Witwatersrand

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. The purpose of this annexure is to ensure the effectiveness and high value of the platform from its start and into the future. To achieve this, the 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 costs for 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 DST'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 120,000; U. Limpopo Dikgale HDSS in Dikgale, Limpopo, established in 1996, with a current population of 38,000; and the Africa Health Research Institute's Population Intervention Platform in uMkhanyakude, KwaZulu-Natal, established in 2000, with a population of 168,000).

Then, the network of HDSS nodes will be expanded to include new nodes in Gauteng (urban), eThekwini (urban), Eastern Cape (rural) and Western Cape (urban). The existing HDSS nodes, which have decades of experience, will substantially reduce the risks of developing new nodes.

This expanded platform, representing over 1% of the national population (55,6 million), will cover a more inclusive spectrum of impoverished yet dynamically developing sub-populations; and enable 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.

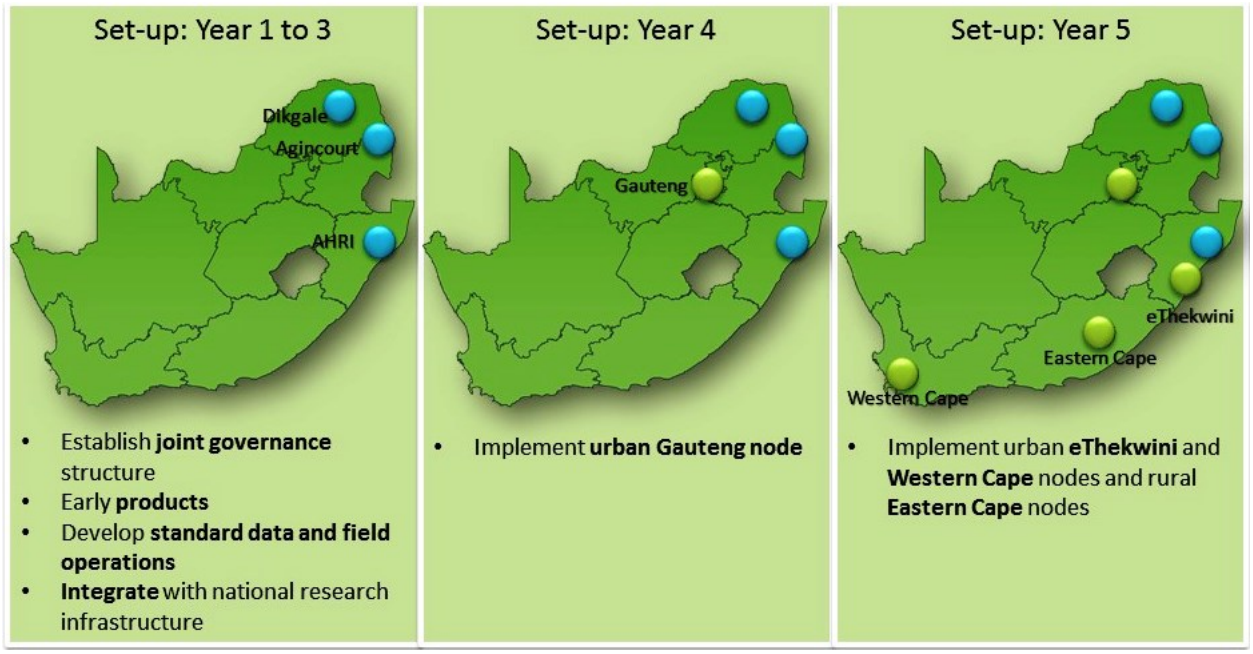

Figure: Plans to expand the Network from three (3) to seven (7) Health and Demographic Surveillance System (HDSS) Nodes, three urban and four rural, in five years.

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
- establish clear links with the **national statistical system**
- effectively engage **community structures**
- 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 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 government, and improving the overall return on public sector investments.

2. Research data for policy development

The 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, 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

The 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 Wellbeing and Economic Wellbeing. Once per year a full set of health and socio-economic wellbeing indicators will be produced. The SAPRIN 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 the 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**

   The 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, social and economic well-being in South Africa's vulnerable populations.

1. <u>Population Health</u>
   By collecting information on all ages, we contribute to the vital understanding of adolescence (a key phase of development and risk period with 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 dry blood spot 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 hospital 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, social and economic wellbeing. The converse also holds that interventions aimed at uplifting health, social and economic wellbeing can be severely compromised in a 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 is highly relevant for personal and household health, social and economic wellbeing and is 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 health and social consequences for migrants and their households.

c. Education status and schooling

Individual interviews will be the basis for ongoing monitoring of education status in 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

Through household interviews assets will be monitored, ranging from modern assets such as ownership of a fridge and stove; fuels used for cooking, heating, and lighting; access to water and sanitation; transport availability and keeping livestock. Size and quality of housing will also be monitored. 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 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 grant 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) has been developed to facilitate a research infrastructure investment programme. 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 for 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. Yet a number of recent studies, such as the Consensus Study on the State of Humanities in South Africa: Status, prospects and strategies (1) and the Charter for Humanities and Social Sciences (2) 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 (3), 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 (4) 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 (5). 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 commencement, the SAPRIN will incorporate three existing Health and Demographic Surveillance System (HDSS) Nodes, 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 will build a common surveillance platform and upgrade the existing HDSSs to this template; and then develop an additional four HDSS Nodes to make a total of seven Nodes, spanning a strategic geography of seven provinces, which will bring national representation spanning rural and urban settings.

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.

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. The population base for the 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 increase to 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, welfare and social change as rapid, at times complex, development occurs. The 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.

The HDSS nodes 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 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.

1.3. Study objectives

The SAPRIN 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 asset ownership and socio-economic status, including 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 15 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 undertaking piloting prior to full introduction.

The eligible population are members of households that are resident within the health and demographic surveillance area (HDSA). These study site boundaries are predefined for each 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.

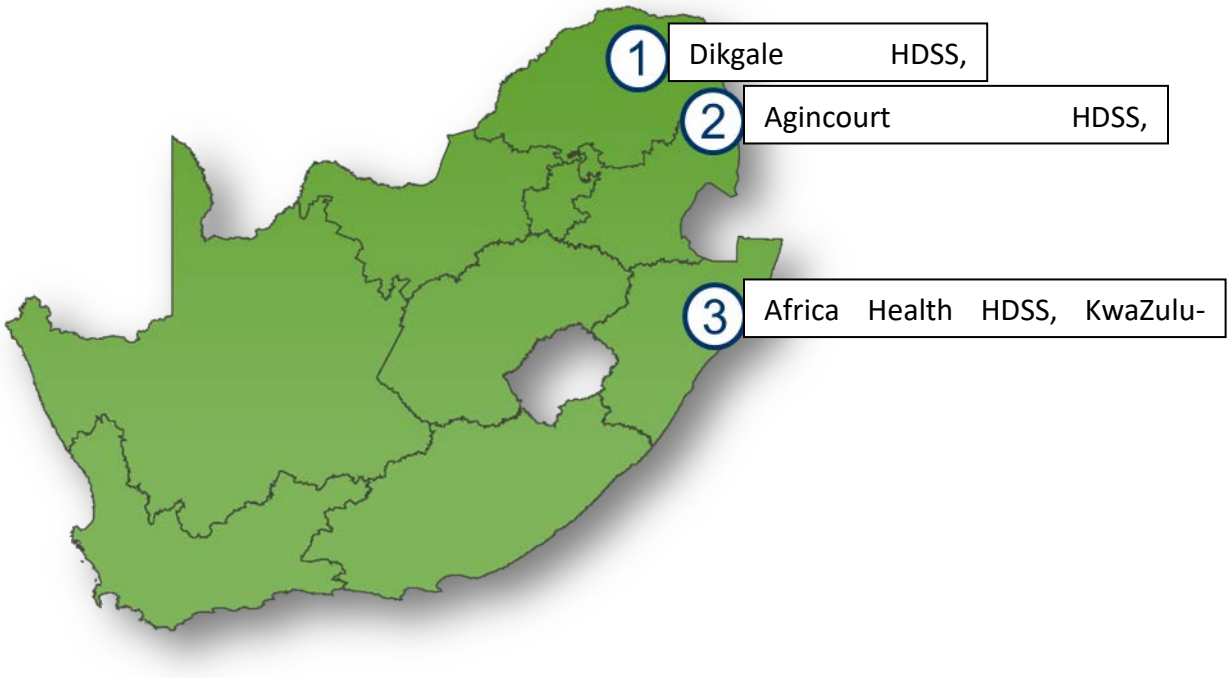

Figure 2: Location of the initial 3 nodes of the SAPRIN

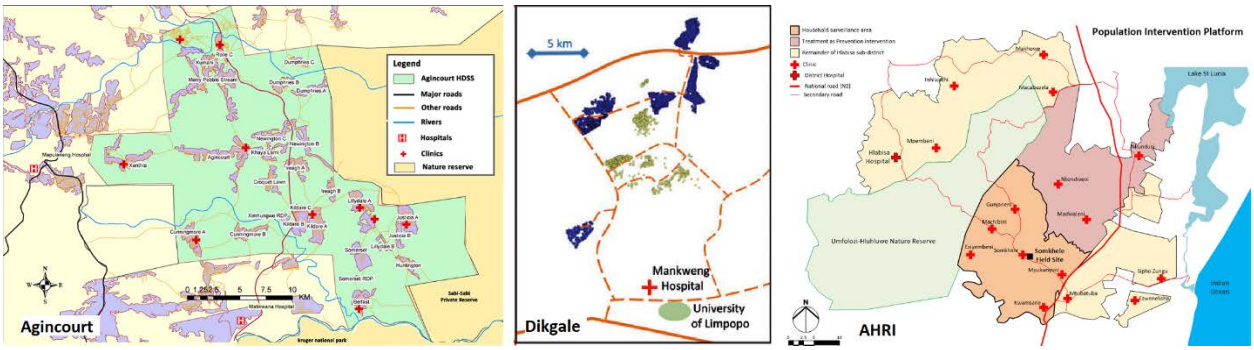

Figure 3: Detail location Maps, from left to right: Agincourt, Dikgale, AHRI

3. Selection and withdrawal of participants

3.1. Household component

Households resident in dwellings within the study area will be eligible for inclusion in the household component of SAPRIN. All individuals identified 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 time at the dwelling 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 {Leitao, 2013 #43}, 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).

- 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.

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).

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

a. **Dwellings**. A dwelling is a building, or a group of buildings, on land belonging to a single person or organisation, and used for one main purpose. Dwellings are primarily used for residential purposes by one or more households. Each dwelling 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 dwelling. 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 dwelling 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 dwelling. Information about household assets, energy and water sources, and sanitation is 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 dwelling as the household; and a **non-resident** is a member of a household who does not normally live at the same dwelling 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, each individual 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 participant information sheets and informed consent forms). 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; prevalence of hypertension as a tracer condition for other non-communicable conditions; 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 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 for such use of the collected specimens from the relevant ethical committees.

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(6, 7).

4.1.3. Verbal autopsy

A verbal autopsy interview based on the latest World Health Organisation (WHO) standard questionnaire will be conducted with last care giver 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 linkage of routine service delivery data with the platform data because (1) the research involves no more than minimal risks to the participants, (2) this combined dataset will be restricted to anonymised data hence there is no risk of the confidentiality of patients being compromised through the linkage, 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)(8) 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 a 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 duration of more than four months.

Frequent household visits, however, have a higher burden of participation for household informants and cost more. 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 **a single household visit each year**, and a short, scheduled telephone interview with the household informants twice a year. Such interviews will quickly review the household membership roster with the informant noting the current residential status of each member, and 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. Consent to conduct a telephonic interview will be obtained from household respondents in the household visit prior to the start of telephonic interview round.

The household and individual components will be combined into a single dwelling 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 that will visit dwellings 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.

Dwelling visits in an area will be preceded by community entry activities conducted by a community engagement team (see section 5.6.2).

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. 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 institute directly from the service delivery authority.

4.3. Screening, counselling, testing, and referral

4.3.1. Overview

The current protocol covers the communication of HIV test results and 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 that 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 & testing (HCT) in accordance with national guidelines (9). 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 requires 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 taking into account 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.

Where measurements (e.g. blood pressure) are taken from research participants where such measurements have clinical implications (e.g. participant may require treatment for hypertension) an appropriate diagnostic protocol must be developed that will determine the need to refer the individual for further clinical management at a nearby health facility. Such a protocol must be consistent with the current local standard of care and specific enough not to put an undue management load on the referral health facilities but maximise the benefit targeted to those who need clinical management. It is recognised that appropriate referral for continuously variable measures (hypertension and diabetes) is much more challenging than for binary conditions (eg HIV). It is notoriously difficult to arrive at a confirmed diagnosis of hypertension or type 2 diabetes, and in any case lifestyle advice is often all that is indicated at the milder end of the spectrum. Research organisations have a responsibility to minimise over-referral, and so overwhelming of government health services. Steps will be taken 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

4.4.1. Specimen collection

All eligible individuals will be asked to consent, in writing, when asked 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 be also stored under appropriate temperature conditions. The storage location of all specimens, aliquots and derivatives will be maintained within the LIMS.

4.5. Record linkage

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 personal 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.

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 whole 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.

4.6. Community engagement

HDSS Nodes will ensure appropriate community engagement with the research population, 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 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 {Participants in the Community, 2013 #15}. 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, comprises 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 and Social Development, and Education – to explain and discuss new studies being planned, and to 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 together the community 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 and

explaining 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 or 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.

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. It is the responsibility of the data custodian to ensure that data access is consistent with terms of the ethical approval obtained for this protocol.

5.1.2. 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.

5.1.3. 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 record 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.

5.3. Types of data

5.3.1. Personally identifiable information

All personally identifiable information (including names, identity numbers, address, geolocation data, telephone numbers) will be stored in a separate table in the database with restricted access and identified by 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 processing and storage

Data will be processed and stored on servers under the physical control of the Node until datasets are made available on the data repository. This data will be de-identified and can then be downloaded for processing on the data user's computer.

Data will be stored on 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 workstation, but only on a central server with restricted physical access.

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 on 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 the DoH accredited HIV counselling certificate course to conduct HIV counselling and testing, as well as receive training in finger-prick collection and handling of dry blood spots (DBS) and micro-capillary blood specimens. Additional specific training will be added when and if required.

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

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 of 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.

6.4. Data quality control

All data recorded on computer tablets will be subjected to rigorous checking by data quality controllers; who will check for accuracy, data logic, internal consistency and completeness of data recording. 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.

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 nodal level.

7.2. SAPRIN Steering Committee

This is the decision-making body that holds the NRI 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. Nodes will need to retain some capacity to host NRI-initiated multi-nodal studies (properly costed), when approved by the Steering Committee.

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 participant 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 message and phone call (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 through the appropriate lines of responsibility to the hosting institution's management. 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

This group has overall responsibility for the national research infrastructure network of HDSS Nodes. The Steering Committee has oversight of the NRI budgets, spending, and the focus of the scientific work, as well as ensuring the 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(10); the principles of Good Clinical Practice as laid down in the ICH Harmonised Tripartite Guideline for Good Clinical Practice(11); and the ethics guidelines of the Department of Health of the Republic of South Africa(12).

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.

9.3. 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. An approach to the assessment of capacity to consent will be implemented at each node.

Consent to approach adolescents aged 15-17[4] 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 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.

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. We will document the specific circumstance 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. 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.

9.4. 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.

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

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.

9.5. Guidance on Approaches to Child Protection

Legal responsibilities to protect the wellbeing 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(10, 11). Interviews are conducted in peoples' homes in the locality, as fieldworkers systematically move from one dwelling 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(13). These approaches should be discussed and agreed 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 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 help lines 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 field work 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 in to 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 at the same time 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 to the police of the participant (as 'perpetrator'), would cause social harm and is likely to put some young people at higher risk.

9.6. 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. 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.

9.7. 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 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.

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 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) and the study site and 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 as specified by the minimum protocol to the SAPRIN data repository. 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 disclosure risk of private and confidential data.
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. References

1. Academy of Science of South Africa (ASSAf). Consensus study on the state of the Humanities in South Africa: Status, prospects and strategies.; 2011.
2. Department of Higher Education and Training. Charter for Humanities and Social Sciences Report. 2011.
3. Department of Science and Technology. Innovation Towards a Knowledge-based Economy, Ten-Year Innovation Plan for South Africa (2008-2018). 2008.
4. Department of Science and Technology. Human and Social Dynamic in Development Grand Challenge Science Plan. 2010.
5. Department of Science and Technology. A South African Research Infrastructure Roadmap Framework. 2013.
6. 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.
7. World Health Organization. The WHO STEPwise approach to Surveillance of noncommunicable diseases (STEPS). Geneva: World Health Organization; 2003.
8. World Health Organization. Neonatal and perinatal mortality: country, regional and global estimates. Geneva2006.
9. Department of Health RoSA. National HIV counselling and testing policy guidelines2015.
10. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. Jama. 2013;310(20):2191-4.
11. ICH Harmonised Tripartite Guideline. Guideline for good clinical practice1996.
12. Department of Health RoSA. Ethics in health research: principles, processes and structures. Second edition ed. Pretoria: Department of Health; 2015.
13. Strode A, Slack C. Selected ethical-legal norms in child and adolescent HIV prevention research: Consent, confidentiality and mandatory reporting. Pietermaritzburg: HAVEG; 2014.

Appendix A: Participant Information Sheets, Informed Consent Forms and Informed Assent Forms

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.

Type	Purpose of Form

Appendix B: Core Data Elements

These data elements will be incorporated into the node specific questionnaire, which 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. Dwelling. A place of residence within the boundaries of the demographic surveillance area ii. Clinic. A primary health care facility, without overnight facilities iii. Health Centre. A primary health care facility that is larger than a clinic with some overnight care facilities iv. Hospital v. Primary School. Teaching up Grade 7 vi. Secondary School. Teaching from Grade 8 to 12 vii. Creche. A pre-school child care facility viii. Police station ix. Tribal authority x. Social service office xi. Home affairs office
b. Area	Village or other higher level spatial grouping of locations
c. 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 dwelling as the household, non-resident household members are individuals who do not share the same dwelling as the household, but who eat from the same pot when they are present in the household's dwelling.
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 dwelling as the household; and a **non-resident** is a member of a household who does not normally live at the same dwelling 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. Sex	Male (1), female(2) or unknown(0)
k. Birth	The Birth event of this individual or the DeliveryEvent resulting in the birth of this individual if the individual enters surveillance through birth
l. EndEvent	The death event of this individual or the Observation at which this individual was last observed
m. MotherId	The unique identifier of the mother of the individual, if the mother is enumerated in the surveillance
n. FatherId	The unique identifier of the father of the individual, if the father is enumerated in the surveillance
o. MotherSurname	The surname of the mother of the individual
p. MotherFirstName	The first name/s of the mother
q. FatherSurname	The surname of the father of the individual
r. 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. Old age ii. Disability iii. War veterans iv. Care dependency v. Foster child vi. Child support vii. Grant-in-aid
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 dwelling, 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 1-4 vi. Undergraduate degree vii. Post-graduate degree viii. Not attending
i. Currently employed	i. Yes ii. Part-time iii. No
j. Not employed	Type/reason for unemployment i. Caring for others/household duties ii. Looking for work iii. Student or in training iv. Unable to work due to illness/disability v. Other reason

k. 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
l. EmploymentType	i. Works as employee ii. Work for themselves iii. Do odd jobs/piece jobs
m. Employer	i. Central government ii. Provincial administration iii. Local / regional authority iv. Public corporation v. Private sector employer vi. Non-profit institution vii. Self-employment viii. Another household member
n. FinancialStatus	Self-reported financial status i. Very Comfortable ii. Comfortable iii. Just Getting By iv. Poor v. Extremely Poor
o. MaritalStatus	i. Never married ii. Married iii. Polygamous Marriage iv. Divorced/Separated v. Widowed
p. PartnershipStatus	i. Marital Partnership ii. Regular Partnership iii. Casual Partnership(s) iv. No Partnership
q. HealthStatus	Self-reported health status i. Excellent, Very Good or Good ii. Fair iii. Poor
r. Tuberculosis	i. Ever treated - Yes/No ii. Treatment started in last 12 months – Yes/No iii. Currently on TB treatment – Yes/No

s. HIV	i. Ever received a test result for HIV - Yes/No ii. Ever had a positive HIV result – Yes/No
t. 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)
u. If HIV-	i. When last HIV- result (>1yr, <1yr ago)
v. HIVResult	HIV serostatus from dried blood spot
w. Hypertension	i. Ever treated - Yes/No ii. Treatment started in last 12mos iii. Currently on treatment
x. Diabetes	i. Ever treated - Yes/No ii. Treatment started in last 12mos - Yes/No iii. Currently on treatment - Yes/No
y. 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
z. 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
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 – to stand/house ii. Piped – Public tap/kiosk iii. Borehole/well iv. Rainwater v. Flowing river/stream vi. Dam/standing water vii. Protected spring viii. Water carrier or tanker
b. Toilet	What kind of toilet does the household use i. Flush ii. Ventilated improved pit iii. Other pit iv. Bucket v. Chemical
c. Electricity supply	Is the household connected to the electricity grid? i. Yes ii. No

d. Cooking fuel	What is the main fuel used for cooking? i. Wood ii. Gas (LPG) iii. Coal iv. Electricity
e. Dwelling construction	What is the construction materials of the walls? i. Brick ii. Cement iii. Other modern building material iv. Stabilised mud v. Traditional mud vi. Wood vii. Other informal structures What is the construction material of the floor? i. Tiles ii. Cement iii. Modern carpet iv. Wood v. Other modern material vi. Dirt vii. Mat viii. Other traditional How many bedrooms does your household occupy at this dwelling?
f. Assets	Does the household have any of the following items in good working order? i. Telephone ii. Cellphone iii. Primus Cooker, Sikeni iv. Electric hot plate v. Electric stove with oven vi. Gas Cooker vii. Fridge/Freezer viii. Electric kettle ix. Television set x. Video cassette recorder/DVD xi. Radio/Stereo xii. Sewing machine xiii. Block maker xiv. Car or Bakkie xv. Motorcycle/Scooter xvi. Bicycle xvii. Kombi/Lorry/Tractor xviii. Bed xix. Table and chairs xx. Sofa/Sofa set xxi. Kitchen sink xxii. Car battery for Electricty xxiii. Wheelbarrow xxiv. Hoe,Spade or Garden Fork xxv. Bed Nets xxvi. Cattle xxvii. Other Livestock (chickens etc)
g. 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
h. 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
i. Food security	In the last 12 months did you or any other member of your household ever cut the size of your meals or skip meals because there was not enough money for food? i. Yes – how often (Almost every month, some months but not all, only once or twice) ii. No

Appendix B : SAPRIN Data Access Policy

Purpose

The purpose of the SAPRIN Data Access Policy is to promote access to the longitudinal population data collected by SAPRIN Nodes in terms of the memoranda of agreement between the nodal institutions and the South African Medical Research Council (SAMRC).

Scope

This policy shall apply to the data defined as 'Core Data Elements' in Appendix B of the SAPRIN Protocol (attached to this policy as Appendix B). The policy also applies to data collected by the SAPRIN nodes prior to their membership of SAPRIN in so far as such data falls within the definition of the 'Core Data Elements'.

Policy Framework

This policy should be read in conjunction with the following agreements and policies:

1. *Project Funding Agreement* (DST/CON 0236/2016) between the Department of Science and Technology (DST) and the South African Medical Research Council (SAMRC). Clause 9.2 "… a recipient that undertakes Research and Development using funding from DST shall own or co-own the Intellectual Property."
2. *Management and Commercialisation of Intellectual Property Policy* of the SAMRC. Clause 3.2.1.6 "Where the research … was sponsored or contracted on a Full Cost basis, ownership of the IP resulting from the Sponsored Research or Contract Research shall vest in the MRC, unless otherwise negotiated between the MRC and the sponsor or contracting party." And clause 3.2.1.11 "The MRC will claim ownership of original databases, electronic data, websites, training materials, games, and other works with commercial value that have been developed by MRC Employees in the course and scope of employment." This clause applies to SAPRIN staff as MRC employees.
3. *Memoranda of Agreement* between the South African Medical Research Council and each of the institutions representing the SAPRIN Nodes. Clause 7.4: "Foreground Intellectual Property from all work on the Project will be retained by the Party generating the Intellectual Property, provided that where Intellectual Property is jointly created, it will be shared proportionate to the contributions." Clause 4.3.2 "The NRI data repository will share the contributed data (including information derived from the service utilisation linkage) with all scientists affiliated with participating institutions as well as the national research community and government departments, with appropriate controls to limit disclosure risk of private and confidential data."

Principles

1. As a public-funded research infrastructure, the research data collected by SAPRIN Nodes are a public good, produced in the public interest, which should be made openly available with as few restrictions as possible in a timely and responsible manner.
2. SAPRIN and nodal data management policies and plans should be in accordance with relevant standards and community best practice. Data with acknowledged long-term value should be preserved and remain accessible and usable for future research. Specifically, SAPRIN will adhere to the FAIR data principles[1] and ensure that SAPRIN data is **f**indable, **a**ccessible, **i**nteroperable and **r**eusable.
3. To enable SAPRIN data to be discoverable and effectively re-used by others, sufficient metadata should be recorded and made openly available to enable data users to understand the research and re-use potential of the data. Published results should always include information on how to access the supporting data.
4. SAPRIN recognizes that there are ethical and legal constraints on the release of research data. To ensure that the research process is not damaged by inappropriate release of data, SAPRIN and SAPRIN Nodes should ensure that these are considered at all stages of the research process.
5. In order to recognize the effort and intellectual contributions of SAPRIN Nodes in producing the data, all users of SAPRIN data should acknowledge the source of the data and abide by the terms and conditions under which they are accessed.
6. SAPRIN will make all data products based on the core data elements available under a Creative Commons BY-NC license (attribution required, no commercial use).

Definition of Terms

1. *SAPRIN*. The South African Population Research Infrastructure Network, funded by the Department of Science and Technology (DST) as part of the South African Research Infrastructure Roadmap (SARIR) and hosted by the South African Medical Research Council (SAMRC).
2. *SAPRIN Node*. A health and demographic surveillance system collecting longitudinal data as a basis for node-based and multi-centre observational and interventional policy-relevant research in accordance with the SAPRIN Protocol and funded by SAPRIN. The host institution of the SAPRIN Node has a memorandum of agreement with the SAMRC governing the funding arrangements and the obligations of the SAPRIN Node.
3. *SAPRIN Protocol*. The common protocol governing the collection of longitudinal population data by SAPRIN nodes, agreed to in and attached as an appendix to the memorandum of agreement entered into by the nodal institution with the SAMRC.

---

[1] Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data. 2016;3:160018.

4. *SAPRIN Core Data Elements*. A standard set of data elements collected by a SAPRIN Node in terms of the SAPRIN Protocol and specified in Appendix B to the SAPRIN Protocol.
5. *Legacy Data*. Data collected by a SAPRIN Node prior to entering into the nodal agreement with the SA MRC that is equivalent to the SAPRIN Core Data Elements.
6. *Node Specific or 'Own' Data*. Data collected in addition to the core data elements as part of the longitudinal population-data or collected by the node during the conduct of node-initiated research.
7. *SAPRIN Data Repository*. A web-based system to manage access to SAPRIN data and data documentation.
8. *SAPRIN Data Products*. Curated and documented datasets produced by SAPRIN from the SAPRIN Core Data Elements and legacy data and shared on the SAPRIN Data Repository.
9. *SAPRIN Data User*. An individual or institution registered on the SAPRIN Data Repository.
10. *SAPRIN Data Use Agreement*. An agreement that a data user must accede to prior to gaining access to a SAPRIN Data Product. See Appendix A.
11. *Direct access*. A data user can directly download a data product from the SAPRIN Data Repository once the user has documented the intended use of the data and has read ('clicked through') the SAPRIN Data Use Agreement.
12. *Restricted access*. A data user must first submit a request to access the data product and such access must be approved in terms of this policy before access is granted to download the dataset from the data repository.
13. *SAPRIN Steering Committee*. The Steering Committee established in terms of the project funding agreement between DST and the SAMRC.

Policy

1. SAPRIN Nodes will on an annual basis allow the extraction of the SAPRIN Core Data Elements from their research databases. This extraction should take place as soon as the data collection for the calendar year has been completed and the collected data verified as being complete, but no later than the start of the second quarter of the following year. Access should be provided to *legacy data* to ensure that retrospective changes to the longitudinal data are included.
2. SAPRIN will retain the extracted data in an anonymised form and produce SAPRIN data products from these extracted data with full acknowledgment to their origins.
3. SAPRIN will document each data product using an internationally recognized meta-data standard prior to making the data available on the SAPRIN Data Repository.
4. Data users will be required to register with a verifiable email address on the SAPRIN Data Repository prior to gaining access to any SAPRIN data product.
5. The access to SAPRIN data products by users will be restricted based on the risk of identity disclosure associated with the dataset. Datasets with minimal disclosure risk will be made available for direct access.
6. Upon accessing a SAPRIN dataset, a data user must indicate the intended use of the data and agree to the SAPRIN Data Use Agreement (Appendix A).

7. Requests for access to restricted data sets will be approved by the SAPRIN Director or designate. Requests for access to data sets must be responded to within 5 working days.
8. If a request is denied, a data user may appeal to the SAPRIN Steering Committee. The decision by the SAPRIN Steering Committee to agree to or deny access will be final.
9. If SAPRIN data are intended to be used for commercial purposes, permission for such use must to be obtained from the SAPRIN Steering Committee.
10. Data users may request access to data forming part of the SAPRIN core data elements, but which are not yet available on the SAPRIN data repository. Such 'ad hoc' data requests will be treated as restricted data sets and once produced, will be made available on and accessed via the SAPRIN data repository by the data user requesting the 'ad hoc' data set. No data will be made available to data users other than through the SAPRIN data repository.
11. Although co-authorship is not a requirement for the use of a SAPRIN dataset, where assistance was given in the development of an ad hoc dataset by SAPRIN staff or SAPRIN nodal staff or affiliates to the extent that such assistance qualifies for the generally accepted norms of co-authorship, such co-authorship must be granted by the data user.
12. SAPRIN nodes can use and share the core data elements they themselves are collecting as part of their own data products without prior permission from SAPRIN. This is consistent with the intellectual property clause in the SAMRC-Node agreement. The use of data containing any of the core data elements or data that depends on the surveillance infrastructure funded through SAPRIN should contain a standard acknowledgement of the DST funding[2] and participation in SAPRIN.
13. Data users requesting access to data elements that are not part of the SAPRIN core but are collected by one or more SAPRIN nodes will be referred to the relevant SAPRIN node/s and will be subject to the node's own data access mechanism.
14. Any dispute concerning the distinction between core data elements and node specific or 'own' data will be resolved by the SAPRIN Steering Committee.
15. SAPRIN may share links to SAPRIN data products on the data repository with other reputable data repositories to improve the visibility and access to SAPRIN data products.
16. The SAPRIN Management Hub will analyse data requests and produce an annual report to the SAPRIN Steering Committee on data set downloads and outputs produced based on the SAPRIN data.

[2] The exact wording of the acknowledgement is current being considered by the DST and this policy will be updated as soon as this has been confirmed.

Appendix A. SAPRIN Data Use Agreement

1. The data and other materials provided by SAPRIN will not be redistributed or sold to other individuals, institutions, or organizations without the written agreement of SAPRIN.
2. The data will be used for statistical and scientific research purposes only. They will be used solely for reporting of aggregated information, and not for investigation of specific individuals or organisations. The Data User will neither use nor permit others to use the data in any way other than listed in the original application (Analysis Plan) for access to the dataset.
3. No attempt will be made to re-identify respondents, and no use will be made of the identity of any person or establishment discovered inadvertently. Any such discovery should immediately be reported to SAPRIN.
4. No attempt will be made to produce links among datasets provided by SAPRIN, or among data from SAPRIN and other datasets that could identify individuals or organizations.
5. The Data User will ensure that the data are kept in a secured environment and that only authorized users have access to the data.
6. Any books, articles, conference papers, theses, dissertations, reports, or other publications that employ data obtained from SAPRIN will cite the source of data in accordance with the Citation Requirement provided with each dataset.
7. An electronic copy of all reports and publications based on the requested data will be sent to SAPRIN.
8. The original collector of the data, SAPRIN, and relevant funding agencies bear no responsibility for use of the data or for interpretations or inferences based upon such uses.
9. Once the data set has served its indicated purpose it must be destroyed. If the dataset needs to be lodged for publication purposes, a reference (a digital object identifier will be maintained by SAPRIN for this purpose) to the original dataset on the SAPRIN data repository should be used. Derived or aggregated datasets produced from the original dataset do not fall within this provision and may be lodged as publication datasets. If the same dataset is needed for a different purpose, the dataset should be re-requested and the new purposes indicated.

By continuing past this point to the data retrieval process, you signify your agreement to comply with the above-stated terms and conditions and give your assurance that the use of statistical data obtained from SAPRIN will conform to widely-accepted standards of practice and legal restrictions that are intended to protect the confidentiality of respondents.