

Wellcome Mental Health
Data Prize

**South African
exemplar datasets**



In preparation for the data prize, we have consulted with several data holders in the UK and South Africa to build an understanding of the data that we consider to be appropriate for the prize. In the process of this, we have identified several exemplar datasets that meet all of the criteria above. The data holders of each of these datasets have been consulted on the most effective way to access their data, with details of the South African datasets documented below.

Following our consultation with the South African Population Research Infrastructure Network (SAPRIN) research team in South Africa, we supported the production of a novel dataset that integrates their most recent surveillance episodes into a longitudinal dataset tailored for research into depression and/or anxiety in children and young people. This dataset will be available for all South African-led prize teams.

We welcome applications from teams that propose to conduct their research on alternative datasets, provided they meet the criteria above. Applications proposing the use of such datasets will not be disadvantaged for selection to the prize.

When making an application to a data holder for access to a dataset, please use the prefix “Wellcome MH data prize” to your project title.

SA datasets

SAPRIN

Official info:

<http://saprin.mrc.ac.za/index.html>

Profile:

SAPRIN (South African Population Research Infrastructure Network) is a network of health and demographic surveillance sites in South Africa. Between them the nodes follows more than 75 000 households (320 000 individuals) longitudinally through regular surveillance visits. Shortly after the start of the Covid-19 pandemic, SAPRIN implemented a shared Covid surveillance programme. This Covid-19 surveillance is still being conducted in these three SAPRIN nodes.

Cohort profiles:

[COLLINSON, M. A., MUDZANA, T., MUTEVEDZI, T., KAHN, K., MAIMELA, E., GOMEZ-OLIVE, F.X. et al. \(2021\) Cohort Profile: South African Population Research Infrastructure Network \(SAPRIN\). *International Journal of Epidemiology*](#)

[SIEDNER, M.J., HARLING, G., DERACHE, A., SMIT, T., KHOZA, T., GUNDA, R. et al. Protocol: Leveraging a demographic and health surveillance system for Covid-19 Surveillance in rural KwaZulu-Natal. *Wellcome Open Research*, 5:109](#)

The SAPRIN team have compiled a novel dataset for use by prize teams, incorporating the latest Covid-19 surveillance data for the children and young people followed by the programme. This allows researchers to investigate links between the many years of routinely collected demographic surveillance data for individuals and households with new data on how individuals have managed during the epidemic, including measures of depression and/or anxiety.

A full specification of the new dataset can be found here:

<https://saprindata.samrc.ac.za/index.php/catalog/80/related-materials>

Access guidelines:

The novel dataset will be made available to prize participants through the existing SAPRIN Data Repository:

<https://saprindata.samrc.ac.za/index.php/catalog/80/study-description>

It will be available for all prize teams wishing to access it for the purpose of conducting research for the data prize.

Birth to 30

Official info:

Profile:

Birth to 30 (Bt30), which began in 1989, is conducted by the University of the Witwatersrand in Johannesburg. The aim was to track a group of children from Soweto-Johannesburg, initially for 10 years, with a multidisciplinary research program encompassing their growth, health, well-being and education. As the cohort has aged, the study focus has expanded to include research into cohort behaviour in adolescence and early adulthood, employment and parenting.

Cohort profiles:

[RICHTER, L. M., NORRIS, S., PETTIFOR, J., YACH, D. and CAMERON, N. \(2009\) Cohort Profile: Mandela's children: The 1990 birth to twenty study in South Africa. *International Journal of Epidemiology*, 36\(3\), 504-511](#)

[RICHTER, L. M., VICTORA, C. G., HALLAL, P. C., ADAIR, L. S., BHARGAVE, S. K., FALL, C. H., LEE, N., MARTORELL, R., NORRIS, S., SACHDEV, H. S., STEIN, A. D. and the COHORTS group. \(2012\). Cohort profile: the consortium of health-orientated research in transitioning societies. *International Journal of Epidemiology*, 41\(3\), 621-626](#)

Online repository of study questionnaires, a research theme inventory and a list of existing publications:

https://www.dropbox.com/sh/pxe7zubyck6e6wa/AACJhYllvbPoWqLUYf_R6tiga?dl=0

Access guidelines:

Birth to 30 datasets are accessible only through direct application to the project leaders at the University of the Witwatersrand. The lead investigators welcome new collaborations from teams that expand upon current research themes and bring new skills to exploring the rich set of longitudinal data.

The Bt30 team are particularly interested in close collaboration with a potential prize team, with a preference for in-person meetings during the project.

As a result of this, Bt30 are only able to commit to collaboration with one prize team.

Applicants should consider this when planning their prize application and have an alternative data source in mind to ensure they are able to submit an application.

Applicants should complete a Birth to 30 Memorandum of Agreement, detailing their proposed research question and the members of the Bt30 research team with whom they would like to work closely.

Over the 28 years of data collection, a very large range of data have been collected. Applicants should be specific about the data required to answer their research question.

Please submit your Memorandum of Agreement by 6th May to have a reasonable expectation of receiving a decision before the prize application deadline.

Cape Area Panel Study

Official info:

<https://www.saldru.uct.ac.za/surveys/cape-area-panel-study-caps/>

Profile:

The Cape Area Panel Study (CAPS) is a longitudinal study of the lives of youths and young adults in metropolitan Cape Town, South Africa. The project was a collaboration between the University of Cape Town, the University of Michigan and Princeton University. With five waves between 2002 and 2009, the study covers a wide range of outcomes, including schooling, employment, health, family formation, and intergenerational support systems.

Cohort profile:

[LAM, D., ARDINGTON, C., BRANSON, N., CASE, A., LEIBBRANDT, M., MAUGHAN-BROWN, B., et al. \(2012\) The cape area panel study: overview and technical documentation waves 1–2–3-4-5 \(2002–2009\). Cape Town: University of Cape Town](#)

Link to existing studies:

<https://www.datafirst.uct.ac.za/dataportal/index.php/catalog/266/related-publications>

Access guidelines:

The vast majority of the data collected as part of the Cape Area Panel Study is publicly available through the DataFirst portal:

<https://www.datafirst.uct.ac.za/dataportal/index.php/catalog/266/>

Please see the section below on [accessing data through the DataFirst portal](#).

Access process	Data	Access timelines
DataFirst portal	All except below	Instant, following registration
Secure Research Data Centre	Identifiers, such as DOB, occupation and school names and locations	5 weeks for decision Access to be confirmed upon centre re-opening

National Income Dynamics Study

Official info:

<https://www.saldru.uct.ac.za/project/national-income-dynamics-study-nids/>

Profile:

The NIDS is a nationally representative longitudinal study in South Africa, involving 28,000 individuals in 7,300 households. The study ran between 2008 and 2017 with a nationally representative sample of over 28,000 individuals in 7,300 households across the country. NIDS examines the livelihoods of individuals and households over time. It also provides information about how households cope with positive or negative shocks, such as a death in the family or an unemployed relative obtaining a job. Other themes include changes in poverty and well-being; household composition and structure; fertility and mortality; migration; labour market participation and economic activity; human capital formation, health, and education; vulnerability, and social capital.

Cohort profile:

Technical papers for all waves of NIDS can be accessed here:

<http://www.nids.uct.ac.za/publications/technical-papers>

A video exploring some of the findings after five waves of NIDS can be viewed here:

<https://www.saldru.uct.ac.za/2019/02/28/south-africa-at-a-glance-looking-back-at-10-years-and-5-waves-of-nids/>

Link to existing studies:

<https://www.datafirst.uct.ac.za/dataportal/index.php/citations/?collection=NIDS>

Access guidelines:

The vast majority of the data collected as part of the National Income Dynamics Study is publicly available through the DataFirst portal:

<https://www.datafirst.uct.ac.za/dataportal/index.php/catalog/NIDS>

Please see the section below on [accessing data through the DataFirst portal](#).

Access process	Data	Access timelines
DataFirst portal	All except below	Instant, following registration
Secure Research Data Centre	Identifiers, such as DOB; locations and names of occupations, schools and households	5 weeks for decision Access to be confirmed upon centre re-opening

DataFirst

Most data on the DataFirst portal is accessible following registration.

Secure variables are only accessible via on-site access to the Secure Research Data Centre at the University of Cape Town.

To apply for access, researchers should return a completed application form to support@data1st.org, detailing the nature of the research, the duration of time access is required for, any outputs required to leave the centre and any new variables that will be created in the course of the research.

Research teams should ensure that they require access to the secure variables to achieve their research objectives before applying for access.

DataFirst aim to respond to requests for access within five weeks, so applications to access secure variables should be submitted by April 29th to allow enough time for approval to be received before the end of the data prize application window. Researchers are allocated space at the Data Centre on a first-come first-served basis.

Further information about applying for access, along with all application forms, can be found at:

<https://www.datafirst.uct.ac.za/services/secure-data-services>

Applicants should note that the Secure Research Data Centre is currently closed, with no fixed re-opening data scheduled. Accordingly, we recommend that teams intending to use the Cape Area Panel Study or National Income Dynamics Study plan their research such that they can rely solely on the publicly available data for the first six months of the prize.

Linking new data:

It is possible to apply to link new data to the secure access datasets held at the Secure Research Data Centre, providing researchers can confirm that:

- they have permission from the producers of the new data to be linked
- linking the data will not compromise data confidentiality