Appendix 1: Mental health data landscaping in India and South Africa

This appendix provides a very high-level overview of the mental health data landscape in India and South Africa. Please note that Wellcome has not conducted a comprehensive scoping exercise of either country, therefore, the supplier would need to determine feasibility of running a data challenge on either of the two proposed geographies (see the go/no go criteria included below). Alternatively, if the supplier is aware of promising and relevant datasets elsewhere in the global south, we would be open to suggestions. To be clear, we do not expect suppliers to scope the entire LMIC landscape for mental health related data, which is why we are providing some high-level information on existing datasets of potential relevance in India and South Africa. At the end of this Appendix, we have also listed a couple of potentially interesting datasets from other LMICs.

When assessing the potential of these (or other) datasets, we suggest that you consider how they fit with the go/no go criteria for a mental health data challenge outside the UK, included here:

- **Is the necessary technical and regulatory data infrastructure in place?** In other words, do necessary datasets exist and are they accessible according to legal regulations and governance? This is particularly important, since the data should be accessible as soon as the data challenge is launched.
- **Is the mental health problem relevant within the country being considered?**
- **Is the necessary data capacity in place?** Ideally there should be sufficient data science capacity, such that a data challenge prize like this would enable capacity building in the selected geography.
- **Is there enough mental health research expertise within the country?** As above, suppliers would need to ensure that the necessary mental health research expertise is already available locally.
- **Are there networks and/or champions to ensure participant recruitment?** Because participation in a data challenge relies on raising enough interest and motivation, it would be important to ensure that there are influential mental health networks/champions in the country under consideration, which would be able to secure participation in the challenge.
- **Is there any potential for leveraging attention from other funders/stakeholder?** Challenges are a good way of leveraging attention from other funders, and so it might be worth considering if the country in question would offer such an opportunity.

Below we have included some possible datasets which may be worth exploring.

**India**
Aside from considering the datasets included below, we would also encourage you to review Appendix 2, which provides a brief overview of other leads the suppliers could potentially try and follow up on (e.g., organisations who may be owners of mental health data and/or are potential partner organisations for a mental health data challenge) and a brief insight into the regulatory landscape in India.

1. The Strengthening the Evidence base on school-based interventions for promoting adolescent health (SEHER) project

Data collected in 2015, 2016 and 2017

- The SEHER project was designed to develop and evaluate a whole-school multicomponent health promotion intervention delivered in government-run schools in Bihar, India. In 2015, a three-arm cluster randomised controlled trial was set-up, involving 75 schools who were randomised to receive (i) the SEHER intervention delivered by a lay counsellor, (ii) the SEHER intervention delivered by a teacher, or (iii) a control group in which only the standard government-run classroom-based life-skills Adolescent Education Programme was implemented. Students enrolled in grade 9 (aged 13-15 years) in the 2015-16 academic year. Students were exposed to the intervention for two years, and the outcome assessment was conducted at three timepoints, including: (i) at baseline, in June 2015; (ii) at 8-months follow-up, in March 2016; and (iii) at 17-months follow-up, in December 2016, when the students were in grade 10. During the first year of implementation, the trial was extended for a second academic year, in 2016-17, to assess whether the effects were sustained in the longer term.
- Data was collected from 15,232 students, and the following outcome measures were used:
  - The primary outcome was school climate, which was measured with the Beyond Blue School Climate Questionnaire (BBSCQ).
  - The secondary outcomes included
    - Depressive symptoms, which were measured with the Patient Health Questionnaire-9 (PHQ-9);
    - Experience of bullying, which was measured with the Bullying Victimisation Questionnaire;
    - Violence, where participants were classified as a perpetrator of violence or not, based on a series of closed-ended responses;
    - Attitude towards gender equity, measured with the Gender Equitable Men Survey;
    - Knowledge of reproductive and sexual health was measured based on WHO’s Illustrative Questionnaire for Interview-Surveys with Young People.
  - Researchers also assessed a range of exploratory behavioural outcomes: current smoking and/or chewing of tobacco, drinking alcohol, consumption of other substances, and sexual behaviour using a set of questions based on WHO’s Illustrative Questionnaire for Interview-Surveys with Young People.
  - Student participants were also asked to report the number of suicide attempts they had made.
• To request data access, please complete this online application form, and send a data request email to bhargav.bhat@sangath.in
• For more information, see: https://www.sangath.in/seher/

2. Consortium on Vulnerability to Externalizing Disorders and Addictions (c-VEDA)

Data collected between 2017 and 2021 (ongoing)

• The c-VEDA project, which is jointly funded by the Indian Council for Medical Research (ICMR) and the Medical Research Council (MRC), was set-up to investigate if environmental and genetic risk factors in industrialised countries and emerging societies shape brain function and behaviour in distinct ways, thus leading to different risk constellations and neurobehavioural trajectories for substance misuse and externalising disorders.

• To address the project’s main aim, a longitudinal dataset involving >14,000 individuals aged 0-25 years and based in rural and urban Indian settings (Chandigarh, Kolkata, Imphal in Manipur, Rishi Valley in Andhra Pradesh, Bangalore and Mysore), was established. The dataset includes a comprehensive assessment of mental health, externalising behaviour and substance use disorders, as well as environmental, biological, and neuroimaging data. Assessment instruments have been selected to allow for comparison with the IMAGEN and ALSPAC UK longitudinal datasets, which have been highlighted in ODI's scoping report (Appendix 2).
  o Please note that the neuroimaging, genetic and epigenetic data is available for a random selection of 1,000 individuals aged 10-23 years.
  o Aside from collating data on physical health (e.g., height, weight, nutrition), the clinical assessment asks recruited participants about their:
    – Circumstances at birth (e.g., mother’s health, birth weight, etc.)
    – Family medical history
    – Exposure to (i) psychological stress (e.g., adverse childhood circumstances) and (ii) environmental toxins
    – Current mental functioning (including thoughts, moods and behaviours), and in the case of adolescents and adults, use of tobacco, alcohol and other drugs, other behaviours like internet use, cell phone use and gambling
    – Temperament/personality
    – Psychological tests for mental abilities
  o The main measure of relevance to anxiety and depression is the MINI International Neuropsychiatric Interview, and we have access to more details on the c-VEDA measures, should that be helpful.

• The c-VEDA project has completed baseline data collection, and data is accessible to the wider scientific community via an anonymised databank web-based retrieval system. Details on how to request data access can be found at: https://cveda.org/access-dataset/

• For more information, see: https://cveda.org/. Alternatively, for details on the protocol and methods, see: https://bmcpsychiatry.biomedcentral.com/articles/10.1186/s12888-019-2373-3 and https://www.nature.com/articles/s41380-020-0656-1
3. Understanding the lives of adolescents and young adults (UDAYA)

Data collected in 2015-16 and 2018-19 (ongoing?)

- Since 2015, the Population Council in India has been leading a longitudinal study of 10-19-year-olds in two large states of India – Bihar and Uttar Pradesh, where one fourth of India’s adolescent population resides. The focus of the study has been on both unmarried girls and boys aged 10-14 years and 15-19 years, as well as married girls aged 15-19 years.
- To date, UDAYA has conducted two waves of data collection, the first in 2015-16 and the follow-up in 2018-19.
  - In the first wave, the Population Council interviewed a total of 10,433 adolescents in Bihar and 10,141 adolescents in Uttar Pradesh (i.e., >20,000).
  - In the second follow-up wave, the Council re-interviewed >16,000 young people.
- The UDAYA study has collected a range of data from the adolescents involved, including information on their education, employability and economic inclusion; media and technology use; agency, community and citizenship; health and nutrition; entry into marriage and motherhood; violence; parental engagement; and entitlement.
  - Mental health data has been collated using the Patient Health Questionnaire-9 (PHQ-9), which assesses depressive symptoms, as well as some additional questions about suicidal ideation, among others.
- Data from the first wave of the UDAYA study can be accessed through the Harvard Dataverse, at: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/RRXQNT, and the Population Council will be making the data from the second wave available to the public soon.
- For more information about the UDAYA study, see: https://www.projectudaya.in/.

4. The National Mental Health Survey (NMHS) of India

Data collected from 2015 to 2016

- The Indian Government Ministry of Health and Family Welfare commissioned the National Institute of Mental Health and Neurosciences (NIMHANS) to undertake a NMHS in India, to develop data on prevalence, patterns and outcomes for mental disorders. The NMHS aimed to estimate the prevalence and burden of mental health disorders in India and to identify the magnitude of treatment gaps, existing patterns of healthcare service utilization, challenges and barriers in mental health delivery, as well as provide an understanding of the impact and disability associated with these disorders. Ultimately, the objective of the NMHS was to provide a rationale for enhancing investment in mental health care in India, as well as evidence to strengthen and implement nationwide mental health policies and programmes.
The survey was cross-sectional and collected data from 2015 to 2016, covering 39,532 individuals (including 1,200 individuals aged 13-17 years and 36,000 aged 18+) from 12 states in India.

The survey used a combination of quantitative and qualitative methods to assess the burden of mental health problems and the status of mental health systems in India. The study instruments included:

- a sociodemographic questionnaire developed to collect household (e.g. household number, cluster type, family composition, income) and individual details (e.g. age, gender, education, occupation, marital status);
- the MINI 6.0 (& MINI kid for 13-17-year-olds) for assessment of mental morbidity, including suicidal risk;
- the Fagerstrom questionnaires for tobacco use disorders;
- the WHO-SEARO screening instrument for epilepsy;
- a brief screening instrument developed by NIMHANS was used for assessing intellectual disability and autism spectrum disorder;
- the Sheehan Disability Scale was used to measure levels of mental-health related functional impairment in primary care settings, developed to assess functional impairment in three inter-related domains: work/school, social and family life;
- an instrument asking about healthcare seeking and utilisation patterns
- the WHO-DAS 2.0 was conducted, based on a set of 7 questions looking at subjective reporting of overall difficulties, duration of these difficulties in the past 30 days, its impact on routine activities, expenditure due to illness, respondent missing on family, social or leisure activities due to illness.

For a summary of the NHMS: [http://indianmhs.nimhans.ac.in/Docs/Summary.pdf](http://indianmhs.nimhans.ac.in/Docs/Summary.pdf)

The NMHS data can be made available for individual researchers.

**Caveat:** Please note that this is a cross-sectional dataset, which will likely limit how you are able to analyse the data. The reason we have included this dataset is because it is the largest reported survey of mental health in India and may therefore hold potential for some interesting analysis to be done.

---

**South Africa**

1. **Birth to Twenty (Bt20)**

*Data collected from 1989 on (ongoing)*

- The Bt20 is a large and long running longitudinal birth cohort study, which intends on being multidisciplinary, tracking the growth, health, well-being and educational progress of urban children in Soweto-Johannesburg.
- The project was initially called Birth to Ten (Bt10) but has recently changed its name to Birth to 20+ (Bt20+), focusing on transitions to adulthood (e.g., education, livelihoods, metabolic disease risk).
• The initial sample size was of 3,273 live births.
• Data has been collected on at least a yearly basis, beginning in the antenatal period, with data on pregnancy and birth. Because the researchers adopted a life-course approach, the project collected data on many of the major issues confronting the individuals at the developmental periods under investigation, including for e.g.:
  o Early years: environmental influences (e.g., poverty, migration and political violence), access to health services, nutrition, childcare, growth and development.
  o Primary school years: cognitive ability, school performance, social adjustment.
  o Later rounds of data collection explored: early lifestyle risk factors (e.g., diet and weight gain, parental monitoring and supervision, educational failure, and sexual experimentation) and a wide range of physical (e.g., body composition and bone mineral density scans) and physiological measures (e.g., pubertal development, biochemical markers of insulin resistance)
  o Adolescent years: prediction and measurement of risk, including sexual and reproductive behaviour (e.g., sexual debut, unplanned pregnancy and sexually transmitted infections), early expression of the metabolic syndrome (e.g., obesity, hypertension, and insulin resistance) and (iii) social marginalization (e.g., school dropout, substance abuse, and conflict with the law).
• Mental health data has been assessed at several timepoints, including for e.g.:
  o Maternal mood, which was measured at 6 months with the Pitt depression inventory and at 10 years with the Centre for Epidemiologic Studies Depression scale (CES-D).
  o Child psychological functioning (including internalising and externalising behaviour), which was assessed at 10 years with the South African Child Assessment Schedule (SACAS).
• A study synopsis is available at: https://journals.lww.com/epidem/Fulltext/2005/09000/_Birth_to_Twenty__A_Study_of_Health_and.390.aspx
• Collaborations can be established through formal agreements with the principal investigators.

2. Cape Area Panel Study (CAPS)


• The CAPS is a longitudinal study, that follows the lives of a large and representative sample of adolescents in Cape Town. The initial sample included 5,256 households, including 4,752 young people aged 14-22 years at the time of recruitment in 2002.
• Although the main objective of the study was to investigate how poverty and inequality are reproduced across generations, the study collected data on a wide range of outcomes, including schooling, employment, health, family formation, and intergenerational support systems. Data on the parents and other older household members of the individuals enrolled in the study was also collected.
To assess participants' mental health/psychological distress, the Kessler-6 (K6) screening scale, which asks respondents about whether they felt nervous, hopeless, restless, sad and worthless, was used.

- For more information, see: https://www.saldru.uct.ac.za/surveys/cape-area-panel-study-caps/
- For more details on the study and access to the publicly available data, please see: https://datafirst.uct.ac.za/dataportal/index.php/catalog/266

3. National Income Dynamics Study (NIDS)


- The NIDS is the first nationally representative longitudinal study in South Africa, involving 28,000 individuals in 7,300 households. It is an initiative of the Department of Planning, Monitoring and Evaluation (DPME), and the first five waves of data collection were implemented by the Southern Africa Labour and Development Research Unit (SALDRU) based at the University of Cape Town.
- NIDS examines the livelihoods of individuals and households over time. A particularly unique feature of the NIDS is that respondents that have moved from their original households are still tracked, so long as they remain within South Africa. Additionally, because the NIDS collects data on the individuals initially recruited and those in their households, data has been collected from more than the initial 28,000. In 2017, for example, 40,000 individuals were interviewed.
- The NIDS dataset includes over 70 million data points, covering a wide range of themes, including: education, income, wealth, health (including emotional health, wellbeing and social cohesion), fertility, vulnerability, migration, labour market participation and economic activity, work life, household composition and structure, consumption, and more. 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.
  - To assess depressive symptoms, the Centre for Epidemiologic Studies Depression scale (CES-D) was used.
- In 2020, a special follow-up with a subsample of the adults from households interviewed in 2017 is being conducted. This survey is called the National Income Dynamics Study: Coronavirus rapid Mobile Survey (NIDS-CRAM).
- For more information, see: http://www.nids.uct.ac.za/
- The data is freely available at: http://www.nids.uct.ac.za/nids-data/data-access

Select examples from the wider mental health data landscape

1. Pelotas Birth Cohort Study: Brazil

Data collected from four birth cohorts (born in 1982, 1993, 2004, and 2015), with subjects from all cohorts followed-up every few years (ongoing)
The Pelotas Birth Cohort Study first began in 1982 when the research team began to record data from infants born to mothers living in the urban area of Pelotas, in Southern Brazil.

The study has collected extensive information on health and socioeconomic variables and has had different focuses at distinct developmental periods. It has, for example, investigated the role of adolescent development in influencing high-risk behaviours and, when the cohort members become young adults, the main emphasis is on examining the risk factors for chronic disease, reproductive history and mental health.

The sample size varies depending on the birth cohort: 1982 (n=5,914), 1993 (n=5,249), 2004 (n = 4,231) and 2015 (n= 4,426), but the study has managed to maintain high retention rates.

- In terms of mental health data, the 2004 cohort is likely to be the best bet. Children’s mental health was assessed by clinical psychologists through maternal report using the Development and Well-Being Assessment (DAWBA). Maternal depressive symptoms were assessed using the Edinburgh Postnatal Depression Scale (EPDS).

Collaborations can be established through formal agreements with the principal investigators. The Wellcome Trust has connections with the principal investigator(s) and may therefore be able to assist in setting up these collaborations.

2. StrongMinds Dataset: Uganda & Zambia

Data collected from 2016 to 2019

- StrongMinds is a social enterprise founded in 2013 that provides mental health services to impoverished African women.

- StrongMinds maintains anonymized data from around 80,000 former clients in both Uganda and Zambia, who were treated between 2016 and 2019. Client ages range from 12-95 years, and the average client is 35 years old. Adolescents aged 12-24 years make up approximately 20% of the total sample.

- The following demographic and descriptive statistics have been collected from all individuals: marital status, head of household status, monthly income and source, food consumption, broad health status, shelter status, and social network access.

- The primary mental health measure used by StrongMinds is the Patient Health Questionnaire (PHQ)-9, which assesses depressive symptoms. Baseline (i.e., before therapy), midpoint (i.e., midway through therapy), and endpoint (i.e., after the last session of therapy) symptoms scores are available for all clients who completed therapy. Additionally, representative samples are resurveyed by external firms at two additional timepoints after the therapy has concluded, including at 2-weeks follow-up and 6-months follow-up.

- The data are divided into a series of cohorts, by country, year and treatment cycle. StrongMinds would therefore need some minimal staff time support to aggregate them before sharing more widely, should that be of interest. The Mental Health
Priority Area at the Wellcome Trust has connections with the StrongMinds team and would therefore be able to help coordinate this process.

- For more information on StrongMinds, see: https://strongminds.org/.

3. **Global Early Adolescent Study (GEAS): Global**

Data collection started in 2015

- The GEAS is a longitudinal study that follows the experiences of over 15,000 adolescents (aged 10-14 years) on five continents, annually, for up to five years (until they are 15-19 years old). Participants were predominantly sampled from poor urban areas in the following countries: USA, Ecuador, Chile, Brazil, Belgium, the Democratic Republic of the Congo, Kenya, Malawi, South Africa, China and Indonesia.
- The study seeks to better understand how gender socialization in early adolescence occurs around the world, and how it shapes health and wellness for individuals and their communities, focusing specifically on body comfort and pubertal satisfaction, gender-based and interpersonal violence, mental health, school retention, and sexual and reproductive health.
- The GEAS collects both observational and interventional data.
- For more information, see: https://www.geastudy.org/
- It is unclear whether the data is easily accessible.