Detailed Descriptions of HDPSA Datasets

This document profiles 12 aggregated datasets compiled for the HDPSA project. Each dataset summarises key indicators from nationally representative surveys conducted in South Africa. The profiling includes variable definitions, data types, ranges, distributions, structural characteristics, and methodological notes.

1. Access to Health Care (national, ZAF)

Focus: Accessibility of primary healthcare services, including facility distance, travel time, barriers to access, and antenatal/postnatal coverage.

- Variables: Distance to facility (categorical, <30 mins, 30–59 mins, ≥1 hour); medical aid coverage (binary, yes/no); ANC visits (numeric, 0–10).
- Data types: Categorical, binary, and numeric.
- Value ranges: Proportions 0–100%.
- Distributions: Typically, right skewed for distance/time, normalized percentages for service coverage.
- Structure: Aggregated proportions, stratified by sex, age, and urban/rural.

2. Anthropometry (national, ZAF)

Focus: Nutritional status of children under five years.

- Variables: Stunting (height-for-age z-score < −2), wasting (weight-for-height z-score < −2), underweight (weight-for-age z-score < −2), overweight.
- Data types: Numeric (z-scores) and categorical (prevalence categories).
- Value ranges: Z-scores typically between −6 and +6; prevalence reported as %.
- Distributions: Bell-shaped for z-scores, prevalence clustered around WHO thresholds.
- Structure: National proportions disaggregated by age group and sex.

3. Child Mortality Rates (national, ZAF)

Focus: Neonatal, infant, and under-five mortality.

- Variables: Neonatal mortality (deaths <28 days), infant mortality (<1 year), under-five mortality (<5 years).
- Data types: Numeric (per 1,000 live births).

- Value ranges: Typically, 10–60 per 1,000.
- Distributions: Declining trend over years; higher in rural/poorer quintiles.
- Structure: Time-series rates derived from retrospective birth histories.

4. COVID-19 Prevention (national, ZAF)

Focus: Behavioural and preventive practices during COVID-19.

- Variables: Mask usage (binary), handwashing frequency (ordinal), vaccination awareness (binary).
- Data types: Binary, categorical.
- Value ranges: 0–100% prevalence.
- Distributions: Skewed towards high uptake of basic preventive measures.
- Structure: Cross-sectional, age/sex disaggregation.

5. HIV Behaviour (national, ZAF)

Focus: Risk behaviours and HIV-related knowledge.

- Variables: Condom use at last sex (binary), multiple partners (numeric), HIV testing history (binary), knowledge of PMTCT (binary).
- Data types: Binary, numeric.
- Value ranges: Binary (0/1), proportions 0–100%.
- Distributions: Urban/rural differences; higher condom use among youth.
- Structure: Individual-level survey items aggregated nationally.

6. Immunisation (national, ZAF)

Focus: Child vaccination coverage.

- Variables: BCG, DPT, Polio, Measles, PCV, Rotavirus, fully immunised (binary/coverage %).
- Data types: Binary and percentage.
- Value ranges: 0–100%.
- Distributions: Skewed towards high coverage for BCG, lower for measles.
- Structure: Aggregated for children 12–23 months.

7. Infant and Young Child Feeding (IYCF) (national, ZAF)

Focus: Breastfeeding and complementary feeding.

- Variables: Early initiation, exclusive breastfeeding (0–5 months), minimum dietary diversity (6–23 months).
- Data types: Binary, categorical.
- Value ranges: 0–100% prevalence.
- Distributions: Exclusive breastfeeding low, dietary diversity uneven across wealth quintiles.
- Structure: Aggregated by child age bands.

8. Literacy (national, ZAF)

Focus: Household literacy and education proxies.

- Variables: Adult literacy (binary: can read a sentence), education attainment (categorical levels).
- Data types: Binary, categorical.
- Value ranges: 0–100%.
- Distributions: Urban–rural disparities; female literacy lower in certain regions.
- Structure: Aggregated national proportions.

9. Maternal Mortality (national, ZAF)

Focus: Maternal health outcomes.

- Variables: Maternal mortality ratio (numeric, per 100,000 live births), skilled birth attendance (binary), facility delivery (binary).
- Data types: Numeric, binary.
- Value ranges: Ratios typically 100–500.
- Distributions: Higher ratios in rural provinces.
- Structure: Modelled estimates with survey indicators.

10. Symptoms of Acute Respiratory Infection (ARI) (national, ZAF)

Focus: Prevalence of ARI symptoms among under-fives.

- Variables: Cough, rapid breathing, care-seeking for ARI.
- Data types: Binary, categorical.

- Value ranges: Prevalence typically 5–20%.
- Distributions: Higher among poorer households.
- Structure: Two-week recall aggregated nationally.

11. Toilet Facilities (national, ZAF)

Focus: Sanitation access.

- Variables: Toilet type (improved/unimproved), shared facility (binary).
- Data types: Categorical, binary.
- Value ranges: 0–100% proportions.
- Distributions: Skewed by urban–rural divide.
- Structure: Aggregated at national level.

12. Water (national, ZAF)

Focus: Drinking water access.

- Variables: Source type (improved/unimproved), on-premises availability, time to fetch.
- Data types: Categorical, binary.
- Value ranges: 0–100% prevalence.
- Distributions: Improved sources dominant in urban areas, longer collection times in rural.
- Structure: Aggregated proportions nationally.

Data Sources and Methodology

All datasets are based on the South Africa Demographic and Health Survey (SADHS) 2016 and related national health surveys. The SADHS employed a two-stage stratified cluster sampling design: in the first stage, 750 enumeration areas (EAs) were selected from the national sampling frame, stratified by province and urban/rural status. In the second stage, households were systematically sampled within EAs. A total of 11,083 households were selected, with interviews conducted for 8,514 women (15–49 years) and 3,618 men (15–59 years) (Statistics South Africa et al., 2017).

Indicators such as anthropometry and immunisation were collected through biomarker measurements and vaccination card/recall, while literacy and access to services were captured via household and individual questionnaires. Mortality estimates were derived from retrospective birth histories.

References

- Statistics South Africa, South African Medical Research Council (SAMRC) & ICF. (2017). *South Africa Demographic and Health Survey 2016: Key Indicators Report*. Pretoria, South Africa, and Rockville, Maryland, USA: Stats SA, SAMRC, and ICF.
- The DHS Program. (2015). *Demographic and Health Survey Sampling and Household Listing Manual*. Calverton, Maryland: ICF International.
- The DHS Program. (2018). *Guide to DHS Statistics, DHS-7*. Rockville, Maryland: ICF.
- United Nations Children's Fund (UNICEF). (2021). *Global databases on child health, nutrition, and WASH*. New York: UNICEF.
- World Health Organization (WHO). (2019). *World Health Statistics 2019: Monitoring health for the SDGs*. Geneva: WHO.