HEALTH AND DEMOGRAPHIC SURVE

Sam

Health and Demographic Surveillance Systems and the 2030 Agenda

Samuel J. Clark

Abstract

HDSS is an old method for intensively monitoring a population that is hosting a trial of some sort. The strengths of HDSS include very detailed descriptions of whole populations with frequent updates. This often provides long time series of accurate population and health indicators for the HDSS study population. The primary weakness of HDSS is that the data describe *only* the HDSS study population and cannot be generalized beyond that.

With respect to the 2030 agenda, HDSS can contribute by: 1) continuing to conduct cause-and-effect studies; 2) contributing to data triangulation or amalgamation initiatives; 3) characterizing the bias in and calibrating *big data*, and 4) contributing more to the rapid training of data-oriented professionals, especially in the population and health fields.

Contents

1. Inroduction

Over the past decades health and demographic surveillance systems (HDSS) have proliferated in low- and middle-income countries (LMIC) as population-based research platforms designed to develop and test public health-oriented interventions. Many survive their founding project and grow into general purpose population and health research ins that contr-o54bute new e ano54d ndings in a wide var-o54ey of elds.

This note brie y describes the HDSS 'method' with a summary of its strengths and weaknesses and then speculates on how HDSS might work toge with other traditional and new sources of data and analytical approaches to strengthen the demographic, health and development evidence base in the context of the SDGs.

2. Health and Demographic Surveillance Systems - HDSS

2.1. HDSS as a Metho

The fundamental motivation for HDSS is the need to accurately account for the full population *at risk* of X or participating in a tr-al of some sort. With that in mind, a population is clearly de ned and regularly mon-o54rored and all the in/out ows of people to/from tho54e popo54ulation are fully documented. The population isypically de ned as *everyone* living within a geographical region, and the ows are *birtho]TJ/F15 0 1 09 9nsT a2d.8d8 0 Ta h](and)T]u/ F19 10r90he Tf 21.384 0 Td [(in-mig* 'outs'.

After an initial census, each household within the demographic surveillance area (DSA) is visited at least annually and a detailed s of information on eh household member is updated. Most HDSS sites conduct so-called 'update' rounds on either a quartely, 3-monthly or semi-annual basis. This provides frequent updates on individual people, the households they live in and their community as a whole. The obvious balance tho54at must be struck is between round frequency and cost, and the key facrors that drive that balance are how well the site wants to characrerize pregnancy-related outcomesy child especially miscarr-ages, still births mortality. To do this well, the round frequency needs to be short enough totify all pregnancies and their outcomes.

This basic design creates an observ platform capable of extremely intensive mon-roring with respect to space and a wide var-ety of social/meic d-mensions. HDSS data are *prective, densely linke and very detaile* . This provides the opportunity for simultaneous disaggregation along many d-mensions, and more unusual and useful, potential study of *cct* because the same entities (people, households, communities) are followed through time.

Although the details v widely from site to site, HDSSypically data on basic demog-raphy[1], socio-economic status (SES) through household asset information [2], cause of (COD)

---

[1] Vital rates, migration, sex-age structure.
[2] In most cases adapte from World Bank/DHS household asset surveys.

through verbal autopsy (VA) [3], various biomarkers and a wide variety of other social and biomedical data.

The typical population size for an HDSS site is 80,000 individuals in 12,000 households. Long-running HDSS sites, such as Matlab in Bangladesh and several in West Africa, have been operating since the 1960{70s, and there are a number in East and Southern Africa that are 20 years old.

2.2. Strengths and Weaknesses of HDSS

HDSS data are likely the most timely and detailed of all population/health data regularly generated in LMICs, and the HDSS platform is often used to conduct rigorous randomized, controlled trials (RCT) and less rigorous observational studies of cause-and-e ect. The accumulated population and health data generated by HDSS sites are often used to conduct detailed retrospective population-based studies that include both sexes, all ages, all SES levels, etc. Data like these are usually extremely rare in LMICs [4].

This very level of intensity and detail is also the challenge associated with HDSS. The HDSS 'study design' is a 100% census of a geographically-de ned population and therefore does not *represent* a any larger population in the sense of a sample survey. Consequently results produced from HDSS data *cannot* be generalized to larger populations, although it is very tempting to do so. This lack of a statistical design that guarantees generalizability is one of two very signi cant challenges for HDSS data.

The second is the *Hawthorne E ect*[5]. HDSS study communities are observed comprehensively for long periods of time and participate in multiple, often overlapping, trials explicitly aimed at changing their health or behavior. Even if adequate control groups are maintained for *a* study, those people are involved in other studies, and over the course of decades of being studied, it is certain that there is no una ected control group, if that were even ethically feasible.

The big strengths of HDSS data are their temporal coverage, detail, dense linkage and the fact that they exist at all for the chronically under-documented populations in LMICs where HDSS sites operate.

The important weaknesses of HDSS data are their lack of generalizability, brought about by the lack of a statistical framework that relates HDSS study populations to a larger population and by the Hawthorne E ect, the fact that they are intensely observed, and this in and of itself, modi es the population.

2.3. Networks of HDSS Sites

There are two prominent networks that coordinate some of the activities of HDSS sites. The INDEPTH Network is an NGO based in Accra, Ghana that provides a clearinghouse function

---

[3] Often using either the INDEPTH Network standard VA instrument or the WHO standard VA instrument.

[4] For example, prospectively collected data on all-age mortality.

[5] The 'observer' e ect in which the act of observing people changes their behavior above and beyond active intervention.

for HDSS activities, coordinates funding and activities of some large multi-site projects, conducts regular scientific meetings for its member sites and provides a wide range of resources to HDSS sites. The INDEPTH web site www.indepth-network.org provides information on the network and its members sites and is updated continuously.

The ALPHA Network is a group of HIV surveillance sites in East and Southern Africa that includes a number of HDSS sites that collect HIV biomarkers. The ALPHA Network is organized by members of the Department of Population Health at the London School of Hygiene and Tropical Medicine in London and maintains a strict focus on HIV-related health and population investigations. The ALPHA Network maintains standardized data sets contributed by each member site and conducts analytical/training workshops to produce publications on HIV-related questions. The ALPHA Network web site alpha.lshtm.ac.uk is updated frequently and provides additional information.

2.4. Insights from HDSS

Over decades HDSS sites have contributed significant insights in many areas ranging from basic demographic indicators to specific trial results related to vaccines, dietary supplements, mosquito bed nets, contraceptive effectiveness, HIV treatment and prevention effectiveness, effects of different approaches to health care system design, etc. Because the data are so rich, they can be used in very many new combinations and consequently speak to a very wide range of health and social questions and the often the intersection of those questions - such as the effects of different approaches to health care systems. The INDEPTH Network web site has a publications page that lists many recent published articles, but there are many more that can be found with more general searches using the names of individual HDSS sites.

2.5. Availability of HDSS Data

Because they are densely-linked [6], prospective data updated on a frequent basis, HDSS data are comparatively complex. This requires HDSS-specific data models, often implemented using relational database management systems. Consequently, HDSS data are difficult to present in an easily understandable form, and it is difficult or not possible to adequately anonymize HDSS data for open access. These factors have prevented or slowed down the movement to public access that is becoming more common with other types of population-based data.

For most serious studies, data must be requested directly from the HDSS sites. This process requires understanding what data are available and their general structure and then making a request to an expert data manager at the site who writes a customized structured query language (SQL) script to extract and manipulate the raw data into the form necessary for the study. This is an *ad hoc*, time consuming process that makes taking full advantage of HDSS data difficult. The INDEPTH Network data repositories are working to make this process less tedious and much quicker.

Finally, there are thousands of published articles using HDSS data that provide additional results, including a growing series of 'cohort profile' articles published in the *International Journal of Epidemiology* that provide detailed descriptions of HDSS sites, their data and their key results [7].

3. Potential Contribution of HDSS to 2030 Agenda

The 2030 Agenda broadly calls for disaggregated population and health indicators describing national-scale populations at sub-national levels with frequent updates. In its current form the HDSS method does all of this except describe large populations, and consequently it is worth identifying what can be learned and scaled-up from HDSS and how existing HDSS sites might contribute to the production of national-scale data.

3.1. Cause-and-Effect

HDSS is designed to study cause-and-effect relationships, and there will continue to be an ongoing need to conduct trials of all sorts. HDSS sites should continue doing this, and perhaps they should be expanded and replicated to provide LMICs with additional capacity to test pharmaceuticals, vaccines and behavioral interventions locally.

3.2. Triangulation & Data Amalgamation

Although many LMICs do not have traditional vital statistics and economic monitoring systems, they do have a variety of sources of data that can be combined to provide a reasonable description (arieam(addiO4(praO5(cott.54(ofdtional)-3 [(ad-337(cm.h)-aieam1-305(win96(to)- n96(to)8(e)-3traditi)-a(ork)-310(

Amalgamation of this type is possible through several general approaches: 1) 'merging' or linking data sets from di erent sources that describe the same or similar units of analysis and then imputing data that are missing from each source, 2) calculating indicator values and uncertainty de ned in the same way from each data source and merging those into a single data set that has wider temporal and spatial extent, and 3) applying a statistical model to a dataset of indicator values from several sources to identify and remove bias in each source and take advantage of spatial and temporal structure to 'smooth' the data and thereby produce nal estimates with smaller uncertainty and both spatial and temporal continuity.

HDSS data o er a fourth possibility. Because HDSS data describe the population so comprehensively, it is possible to understand relationships between indicator values that are widely collected in other sources of data, such as SES, and particular outcomes of interest. This information can be leveraged to design *very e cient, targeted* sampling designs for sample surveys that produce representative indicator values in large populations that are similar to the HDSS from which the relationship is understood. Potentially this makes existing HDSS sites much more valuable without requiring them to change at all.

3.3. Calibration of `Big Data'