Health and Demographic Surveillance Systems and the Post-2015 Agenda

Samuel Clark
University of Washington

Strengthening the Demographic Evidence Base for the Post-2015 Development Agenda

New York, NY
October 5-6, 2015

Health and Demographic Surveillance Systems - HDSS

HDSS as a Method I

HDSS is intensive LONGITUDINAL data collection - linked through time

- Most HDSS sites motivated by a need to comprehensively account for participants in *trials*
- Population – usually everyone living within a geographical boundary
- Aften an initial census, all ins/outs monitored
  - 'ins' = births and in-migrations
  - 'outs' = deaths and out-migrations
- Very few HDSS follow/track people outside surveillance area – reidentifying people when they move (back) in is a big challenge

HDSS as a Method II

- All households visited 1-4 times per year, at each visit status of all household members queried
  - vital and migration events updated
  - many other status variables updated, e.g. household assets (SES), employment/education status, biomarkers, etc.
- Frequency of visits often dictated by *pregnancy monitoring* to accurately describe all birth outcomes

HDSS as a Method III

- Beyond demographic structure and dynamics, key items recorded at almost all HDSS sites
  - Time-evolving *links* between people and households, people and places (residence) and households and places
  - Household socio-economic status (SES) through assets
  - Cause of death through verbal autopsy (VA)
  - Individual & household level status indicators – many!
  - Evolving array of biomarkers, usually connected ongoing trials

HDSS as a Method IV

- Result is data that are *prospective, densely linked and very detailed*
- Very few events missed, and if they are, they are recorded on next visit $\rightarrow$ all data improve over time
- 'Typical' HDSS site
  - Contiguous *demographic surveillance area* of several hundred square kilometers
  - $\sim$ 80,000 people under surveillance
  - $\sim$ 12,000 households
  - 2-3 visit 'rounds' per year
  - Operating for 10-20 years, some much older

Strengths and Weaknesses of HDSS I

- Strengths
  - Very high quality data from populations with comparatively few health and population data
    - Longitudinal
    - Detail, including biomarkers
    - Dense links between entities
  - Highly functional platforms to conduct randomized, controlled trials
  - Accumulated linked, detailed data allow wide variety of retrospective, population-based studies
  - Through both trials and observational studies, address questions of **cause & effect**

Strengths and Weaknesses of HDSS II

- Weaknesses
  - **No statistical framework for generalization**
    - HDSS study design is effectively 100% sample of population in demographic surveillance area
    - *Not* a traditional sample of a larger population; does not 'represent' anything larger
    - *Cannot* generalize to larger populations in the manner of a sample survey
  - Variety of reasons to expect HDSS study populations to differ from similar surrounding populations
    - 'Hawthorne Effect' - HDSS study populations intensively observed over long periods of time
    - HDSS study populations participate in trials whose aims are to directly change health and behavior

Networks of HDSS Sites I

There are two formal networks of HDSS sites

- The **INDEPTH Network** based in Accra, Ghana
  - 52 HDSS sites in 20 Countries, mainly Africa and Asia, $\sim$ 3M people under observation
  - Coordinates multi-site projects
  - Organizes annual scientific meeting
  - 'Professional organization' for HDSS sites and HDSS scientists
  - **Does not own or directly control any data**
  - Operates two public-access data repositories, more below
  - Clearinghouse for HDSS information, methods, etc.
  - `www.indepth-network.org`

Networks of HDSS Sites II

- The **ALPHA Network** based at the London School of Hygiene and Tropical Medicine, London
  - 10 member sites in East and Southern Africa, *overlaps with INDEPTH*
  - Specifically concerned with HIV; member sites must operate HIV sero-surveillance
  - Focuses on specific HIV-related scientific investigations
  - Conducts standardized analysis on pooled data
  - Maintains large collection of clean, harmonized data; not publicly available
  - `alpha.lshtm.ac.uk`

Insights from HDSS

Too many to summarize neatly ...

- Numerous consequential results from biomedical trails: mosquito bednets, nutritional supplementation, contraceptive program effectiveness, HIV prevention & treatment, etc.
- Basic demography: structure & dynamics
- Relationships between household SES and health and demography
- Family structure and risk associated with various things: death, movement, education, health
- Google Scholar searches using HDSS site names will reveal thousands of publications

Cause & effect results most useful; biomedical results potentially generalizable, others less so

Availability of HDSS Data

- Contact site directly and negotiate access to data, usually in context of a project and grant
- Work through either of the two prominent networks of HDSS sites – INDEPTH or ALPHA
- Access HDSS data on the INDEPTH Network's data repositories
  - **iShare** `www.indepth-ishare.org`
    anonymous individual-level data
  - **INDEPTHStats** `indepth-ishare.org/indepthstats`
    aggregated data

Potential Contribution of HDSS to Post-2015 Agenda

Potential Contribution of HDSS to Post-2015 Agenda

- HDSS data have unusually strong strengths and weaknesses
- Worth thinking carefully how they might contribute to post-2015 agenda
  - Cause & Effect
  - Triangulation & Data Amalgamation
  - Calibration of 'Big Data'
  - Training

Cause and Effect

HDSS should continue doing what they are best at: conducting trials

- LMICs need locally-conducted trials
- HDSS sites are designed to conduct such trials, and net of the Hawthorne Effect, they are good at it
- Continue ...

Triangulation & Data Amalgamation

Triangulating or *amalgamating* data from several sources is the most obvious and likely least expensive way of rapidly making more and better data available where they are needed most

- HDSS sites can contribute unusually detailed data to amalgamation exercises
- HDSS data can be used to create models that represent finely-disaggregated indicators in relationship to each other
- Such models can be used to fill-in, extrapolate, etc. sparser information from other sources
- Models of this type can also be used to design *highly efficient, targeted* sampling strategies that can be used to conduct very inexpensive sample surveys in much bigger populations

Calibration of 'Big Data' I

The *data exhaust* version of big data are informative artifacts of ongoing processes that have an information component: cell phone calls, social media, web searches, etc.

- Many of these may yield valuable information about the populations that produce them
- *Big data of this type do not have a statistical design that dictates how they are related to the population*, therefore no way to systematically
  - Understand bias
  - Characterize uncertainty
  - **Generalize**

Calibration of 'Big Data' II

- **Likely most valuable way that HDSS can contribute is to help solve these problems with big data**
  - HDSS can be used to characterize and understand the relationship between big data and the population, then this can be used to correct/calibrate big data-derived indicators in general
- Example: *cell phone call metadata* - 'call detail records'
- HDSS can include study of cell phone ownership/usage
- Results of the HDSS cell phone study can be compared to cell phone call metadata for HDSS study population and used to understand the biases and omissions in population counts, movements, etc. inferred from cellphone call metadata

Training

The biggest challenge to the 'Data Revolution' is the lack of people to produce and work with the data, especially in lower- and middle-income countries (LMICS) where it is needed most

- This will require quickly scaling up training in data-oriented fields in LMICS: data science, analytical methods, related IT
- Many HDSS sites are already linked to universities and run training programs at graduate level that involve integrate data, analysis and substantive learning
- HDSS sites offer a unique 'hands-on' opportunity to train people in *all* of the variety of related tasks involved in producing good data and useful results from those data
- Training and internship programs situated at HDSS sites should be rapidly built up

Discussion

Discussion I

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 very frequent updates
- Long time series of very accurate population and health indicators for the HDSS study population

The primary weakness of HDSS

- The data describe only the HDSS study population and cannot be generalized beyond that

Discussion II

With respect to the post-2015 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
4. Contributing more to the rapid training of data-oriented professionals, especially in the population and health fields