A conversation with Jack Colford and Ben Arnold on September 16, 2013

Participants

- Jack Colford – Professor of Epidemiology, School of Public Health, UC Berkeley
- Ben Arnold – Epidemiologist, Colford Research Group, School of Public Health, UC Berkeley
- Timothy Telleen-Lawton – Research Analyst, GiveWell

Note: This set of notes was compiled by GiveWell and gives an overview of the major points made by Professor Colford and Dr. Arnold.

Summary

Professor Colford's research group, of which Dr. Arnold is a member, is advising Deworm the World and its partners in India on conducting a survey to estimate the prevalence and intensity of soil transmitted-helminths in Rajasthan and other states. The main subjects of discussion were the sampling strategy and possible sources of bias in the estimate.

Colford Research Group and Deworm the World

Professor Colford and Dr. Arnold have known Alix Zwane of Deworm the World (DtW) since she was working for the Gates Foundation. When she became the Executive Director of DtW, she asked the Colford Research Group to advise DtW and its in-country partners on designing prevalence surveys for India.

The Colford Research Group gave informal suggestions to DtW on the design of the current Rajasthan survey (this is the second survey that DtW has conducted in Rajasthan). The group is discussing with DtW a more formal advisory role for the design and analysis of future surveys sponsored by DtW.

Rajasthan 2013 prevalence survey design

The goal of the prevalence survey is to achieve an estimate of the statewide prevalence of soil transmitted helminth (STH) infections in Rajasthan with a margin of error of 3 percentage points. Intensity rates will also be estimated from the same data although they are expected to be low, since the first prevalence survey showed mostly low prevalence rates of the two species measured, and DtW has conducted statewide deworming since that time.

Sampling

Climate can have a significant impact on STH prevalence. The survey collects samples from districts that are stratified across different ecological zones of Rajasthan. The implicit stratification by ecological zone is accomplished in the first stage of Probability in Proportion to Size (PPS) sampling by ordering the districts in a list by ecological zone and using a systematic random sample. The sampling interval is determined based on the total estimated enrollment (summed across all districts) divided by the number of districts desired (N). A random start number is chosen, and N numbers are drawn by adding the interval to the random start. Districts are ordered by ecological zone, and the cumulative enrollment is calculated. A district is selected if its enrollment range includes the randomly selected numbers. Since larger districts have wider enrollment ranges, they are more likely to be

selected.

With PPS sampling, the districts are weighted in their chance of being selected as the first district based on the number of children enrolled in school in the district so that each enrolled child in Rajasthan has an equal chance of being selected for the sample.

Possible sources of bias

Below are several important possible sources of bias in the prevalence rate estimate; these are not intended to be exhaustive:

1) Since the survey is school-based, children that don't go to school may be underrepresented in the survey. The survey attempts to overcome this bias by using lists of enrolled children, and seeking out children that aren't at school on the day the sample is taking place. This wasn't done in the original 2012 Rajasthan prevalence survey. Still, children not in school will be harder to find so the final sample will likely contain a lower proportion of out-of-school children than the target population. One might imagine that kids not in school would have higher prevalence due to a correlation between absenteeism and sickness or lack of access to health care. On the other hand, such children might also have less exposure to infection so the direction of the bias cannot be assumed.

2) Among selected children that are found, participation is voluntary and if children decide not to participate on a systematic basis (for example, because they are sick), that can introduce bias to the results.

3) Estimating STH prevalence using a single stool sample from each child may underestimate STH prevalence in conditions where children have low intensity infections, as a single stool on a given day may, by chance, contain no eggs despite coming from an infected child. In the United States, doctors often use stool samples from three consecutive days to diagnose worm infections. Unlike the original 2012 Rajasthan prevalence survey, this 2013 survey attempts to measure this bias by selecting a subsample of schools in which to test on three consecutive days rather than one. If measuring samples for three consecutive days results in significantly higher infection diagnoses, that may be used to adjust the final prevalence estimate.

4) Keeping the stool samples cold until they are analyzed in a central laboratory is important for properly counting STH eggs; problems with the logistics of keeping the samples cold could bias the results.

Standard design

The survey design is a relatively standard methodology for measuring prevalence and intensity,

Prevalence mapping and geospatial modeling

Using the data generated from the above approach to build a model of local STH prevalence based on ecological data, climatic data and location within the state may allow for predictions of which districts (among those not directly sampled) are likely to have higher or lower prevalence. The advantage of the approach is that it can provide greater spatial resolution of prevalence estimates (e.g., at the district level rather than the state level), but the disadvantage is that the estimates rely on a statistical model that may or may not be correct. Depending on the sophistication and validation of such a model, it may

not be appropriate to use such a model to estimate prevalence at higher spatial resolution within the state. The validity and usefulness of the predictive models will be very context specific, and it remains unclear how well they would perform in the context of Rajasthan, India.

Research on helminths

The Colford Research Group is involved in an ongoing study that may ultimately add to the scientific understanding of the health impact of helminth infections in newborns (www.washbenefits.net).