Institute for Health Metrics and Evaluation

Data Release Information Sheet

Data summary

Project name: Access, Bottlenecks, Costs, and Equity (ABCE) project in Tamil Nadu

Date of release: March 27, 2018

Summary:
This dataset is from the Access, Bottlenecks, Costs, and Equity (ABCE) project in Tamil Nadu, jointly conducted by the Public Health Foundation of India (PHFI) and the Institute for Health Metrics and Evaluation (IHME). The dataset provides information at the facility-year level. Data include information on services offered, expenditure, revenue, personnel by category, and other variables related to facility operations. In total, a nationally representative sample of 168 facilities were surveyed. Data were collected through interviews of health providers, direct observation of facility areas, and assisted observation of facility resources.

Relevant publications:

Institute for Health Metrics and Evaluation (IHME), Public Health Foundation of India (PHFI). *Assessing Facility Capacity, Costs of Care, and Patient Perspectives:* Tamil Nadu. Seattle, WA: IHME and New Delhi: PHFI, 2018.

Acknowledgments

Collaborating organizations:

- Institute for Health Metrics and Evaluation (IHME)
- Public Health Foundation of India (PHFI)
- Tamil Nadu state government and district officials

Funding institution:
Bill & Melinda Gates Foundation

Suggested Citation: Institute for Health Metrics and Evaluation (IHME), Public Health Foundation of India. Access, Bottlenecks, Costs, and Equity (ABCE) project in Tamil Nadu, India: ABCE Health Facility Survey 2012-2013. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2018.

File inventory

File name	Description	Date Finalized
IHME_ABCE_IND_TN_2012_2013_FACILITY_DATA_Y2018M03D27.CSV	Public use survey dataset – CSV format	March 27, 2018
IHME_ABCE_IND_TN_2012_2013_FACILITY_DATA_Y2018M03D27.DTA	Public use survey dataset – DTA format	March 27, 2018
IHME_ABCE_IND_TN_2012_2013_FACILITY_CODEBOOK_Y2018M03D27.CSV	Codebook, with descriptions of all variables included in the dataset	March 27, 2018
IHME_ABCE_IND_TN_2012_2013_FACILITY_QUESTIONNAIRE_Y2018M03D27.PDF	Health Facility Questionnaire	March 27, 2018
IHME_ABCE_IND_TN_2012_2013_FACILITY_INFO_SHEET_Y2018M03D27.PDF	Data release information sheet	March 27, 2018

Data structure

The data are structured in such a way that each row represents a facility-year (i.e., the variable "year" in each row corresponds with a given fiscal year).

Methodological statement

Data collection

Data from health facilities in Tamil Nadu were collected with the ABCE Facility Survey from October 2012 to January 2013 by research associates (RAs) hired by PHFI. All data were collected with DatStat, which allows for survey adaptation and customization based on respondent answers. All collected data went through a thorough verification process between IHME, PHFI, and the ABCE field team. Following data collection, the data were methodically cleaned and re-verified, and securely stored in databases hosted at IHME and PHFI.

All instruments used for the ABCE project in Tamil Nadu are available through IHME's Global Health Data Exchange (GHDx): http://ghdx.healthdata.org/record/india-tamil-nadu-access-bottlenecks-costs-and-equity-project-2012-2013.

Sampling/population

To construct a nationally representative sample of health facilities in Tamil Nadu, a two-step stratified random sampling process was used. Districts, from which facilities would be drawn, were selected using three strata to maximize heterogeneity: proportion of full immunization in children aged 12-23 months as an indicator of preventive health services; proportion of safe delivery (institutional delivery or home delivery assisted by skilled person) as an indicator of acute health services; and proportion of urban population as an indicator of overall development. The districts were grouped as *high* and *low* for urbanization based on median value, and into three equal groups as *high*, *medium* and *low* for the safe

delivery and full immunization indicators. Twelve districts were selected randomly from each of the various combinations of indicators, and in addition the capital district was selected purposively.

Twelve districts were selected randomly from each of the various combinations of indicators, and in addition the capital district was selected purposively. Facilities were sampled from each selected district across the range of platforms identified in Tamil Nadu. In India, sampled health facilities included district hospitals, sub-district hospitals, community health centres, primary health centres, and sub health centres.

A total of 13 districts were selected through the county sampling frame, and 168 facilities from those districts participated through the facility sampling frame:

- One district hospital
- One sub-district hospital for each sampled district hospital
- Two community health centres for each sampled sub-district hospital
- Two primary health centres for each sampled community health centre
- One sub health centre for each sampled primary health centre

Facility type	Final sample
District hospital	13
sub-district hospital	26
Community health centre	24
Primary health centre	54
Sub health centre	51
Total health facilities	**168**

Data processing and cleaning

Data were processed and appended together. All feedback gathered through the collection and verification process was incorporated into the dataset. Clear data entry-errors and skip patterns were removed. Generally, facilities that answered N/A to questions were treated as zeros while facilities that answered “don’t know” or “refuse to answer” were treated as missing data. Several variables were identified to clean more extensively (Annex 1). These variables were reviewed by comparing the difference between totals and subcomponents collected in the data as well as ratio outlier detection, among other methods. For a list of assumptions applied to these variables, please contact IHME.

The other variables in this dataset have not been analyzed or examined in detail beyond removing skip patterns and apparent data-entry issues.

Inclusion of administrative data

Public health facilities in Tamil Nadu rarely provide information about expenditure toward medical supplies and pharmaceuticals. Instead, medical supplies and pharmaceuticals are often paid for through central sources, such as the national or state government.

Pharmaceutical expenditures

State ministry data from 2006 to 2012 have been added to supplement the pharmaceutical expenditure reported by the facility, to capture expenditure exchanged above the facility level. Since vaccines were distributed from central agencies, the expenditure on vaccines was calculated based on the number of doses stored or administered.

Imputation and constructed variables

Missing values in the following variables have been filled in using the Amelia II package in R (Honaker et al. 2011) using lags and leads of two years. The model was run 50 times, and the final dataset includes the median values over these 50 runs. All variables were imputed at the same time.

The model used to fill in missing values was chosen through out-of-sample validation to compare model fit with other viable models; ultimately, the model that minimized the root mean square error (RMSE) across all variables was chosen. This strategy was consistent across all ABCE countries and states. For each variable, there is a corresponding variable with the prefix "**i_**", which indicates an imputed value.

The complexity of imputation for such a large dataset prevented researchers from successfully imputing missing values for the over 1,500 variables in the ABCE dataset. As a result, we imputed all major outputs, expenditures, and personnel, but excluded many subcategories of outputs and expenditures.

Known data-quality issues

To date, the variables pertaining to inpatient visits, outpatient visits, births, immunization doses, personnel totals, and expenditure totals have received the most scrutiny and review by ABCE research team members. Outliers involving other variables have yet to be thoroughly examined in such detail.

The ABCE Facility Survey structure varies slightly across countries and Indian states. Therefore, some variables used for the ABCE project in Tamil Nadu may not be directly comparable to those used in other ABCE countries and Indian states.

The expenditures reported with these data include any expenditure at the facility level as well estimates of expenditure through central government agencies. Expenditure data do not include the cost of donated pharmaceuticals or medical supplies and equipment.

Facility outputs were collected using forms that facilities use to record patient volumes and services provided. This method of record keeping can introduce some inconsistencies between outputs. For instance, for a small number of facilities, the number of family planning visits was higher than the total

number of inpatients recorded. For certain outputs such as inpatient visits and expenditure subcomponents, the totals were more reliable then the subcomponents, so the subcomponents may not add up to the total. For other outputs, the subcomponents were more reliable than the total reported.

Some facilities included in the sample had inadequate data for a subset of variables. There were instances where all five years of a certain output or expenditure category had to be imputed. These values are less reliable than the imputation results from facilities where at least one year of data has been reported. You can identify these cases if the variable has "**all_imp_"** preceding its name and the variable you are interested in is 1. If a value has been imputed, note that the total will not be the sum of the subcomponents.

Codebooks

Variable names and labels can be found in the Excel file "**IHME_ABCE_IND_TN_2012_2013_FACILITY_CODEBOOK_Y2018M03D27.CSV**." Each codebook contains variable names pertaining to a given module of the survey. You will also find indicators of which variables have been imputed, supplemented with any additional information and variables generated by IHME. Any variable with "**fy**" in the name varies over the five-year panel of the data, whereas variables without "**fy**" in the name do not vary and were measured at only a single point in time.

Public Use Dataset Notes

This is a public use dataset. The data have been de-identified. Variables determined to contain identifiable private information, or potentially identifiable private information, for health facilities, health workers, and/or other individuals have been removed in accordance with IHME's microdata release protocol. The protocol's determination for variables that constitute identifiable private information is based primarily on HIPAA'S De-identification Standard.

Citations

Honaker J, King G, Blackwell M. Amelia II: A program for missing data. *Journal of Statistical Software*. 2011: 1-47.

Additional Information

Terms and Conditions

http://www.healthdata.org/about/terms-and-conditions

Contact information

To request further information about the ABCE project in India or in other countries, please contact IHME:

Institute for Health Metrics and Evaluation
2301 Fifth Ave., Suite 600
Seattle, WA 98121
USA
Telephone: +1-206-897-2800
Fax: +1-206-897-2899
Email: data@healthdata.org
www.healthdata.org

These files may be updated periodically, so we appreciate hearing feedback or additional information about how these data are being used.

Annex 1. Variables that have been cleaned, imputed, or both for the ABCE project in Tamil Nadu

Variable name	Variable description	Cleaned	Imputed
Expenditure			
te_fy_tot	Total expenditure	X	
atq_fy_1	Total administrative and training expenditure	X	X
nm_fy_1	Total non-medical expenditure	X	X
iu_fy_1	Total infrastructure expenditure	X	X
mse_fy_1	Total medical supplies and equipment, pharmaceuticals, and vaccine expenditure	X	X
ps1_fy_1	Total expenditure on personnel	X	X
ps2_sal_fy_1	Total expenditure on salaries	X	
Personnel			
pers_fte_fy_1	Total personnel	X	
Doctors_fy	Total # of doctors	X	X
Nurses_fy	Total # of nurses	X	X
Paramed_fy	Total # of other medical personnel	X	X
Nonmed_fy	Total # of non-medical personnel	X	X
Outputs			
opc_fy	# of outpatient visits	X	X
ipc_fy	# of inpatient visits	X	X
Births_fy	# of deliveries	X	X
Opcv_fy	# of immunization doses	X	X
Inputs			