2-2019

# Regional prevalence of health worker absenteeism in Tanzania

Tomoki FUJII
*Singapore Management University*, tfujii@smu.edu.sg



Citation
FUJII, Tomoki. Regional prevalence of health worker absenteeism in Tanzania. (2019). *Health Economics*. 28, (2), 311-316.
**Available at:** https://ink.library.smu.edu.sg/soe_research/2235

# Regional Prevalence of Health Worker Absenteeism in Tanzania*

Tomoki Fujii

*School of Economics, Singapore Management University*
Singapore Management University
90 Stamford Road, Singapore 178903
Phone: +65-6828-0279
Fax: +65-6828-0833
E-mail: tfujii@smu.edu.sg

Published in Health Economics, Feb 2019. 28 (2), pp. 311-316.
https://doi.org/10.1002/hec.3844
Pre-print


* This work grew from a consultancy work for the World Bank. I thank Gayle Martin and Waly Wane.

## Regional Prevalence of Health Worker Absenteeism in Tanzania


### ABSTRACT

Absenteeism of health workers in developing countries is common and can severely undermine the reliability of health system. Therefore, it is important to understand where the prevalence of absenteeism is high. We develop a simple imputation method that combines a Service Delivery Indicators survey and a Service Provision Assessment survey to estimate the prevalence of absenteeism of health workers at the level of regions in Tanzania. The resulting estimates allow one to identify the regions in which the prevalence of absenteeism is significantly higher or lower than the national average and help policymakers determine the priority areas for intervention.



Keywords: Sub-Saharan Africa, primary health facility, imputation, random-effects probit, service delivery indicator


### 1 Introduction

Despite the healthy economic growth witnessed in recent decades, many parts of sub-Saharan Africa continue to face various development challenges, among which is the lack of access to good and affordable healthcare services. This issue has indeed been well recognized and government spending has more than doubled and development assistance for health has more than quadrupled since the 1990s in Africa. While increased health spending may have raised the quantity and quality of health equipments, facilities and workers, it has not been matched with improvements in the delivery of healthcare services. One important issue is the absenteeism of health workers, which is prevalent in developing countries (Chaudhury et al., 2006).

Absenteeism of health workers is a serious issue because it severely undermines the reliability of the health system and potentially negatively affects health behaviors and outcomes of the public. For example, Glodstein et al. (2013) find that nurse absence on a patient's first visit significantly reduces the probability of getting a HIV test over her entire pregnancy in West Kenya.

It is important to recognize that the causes of absenteeism are diverse (Belita et al., 2013) and low income levels do not necessarily translate into a high level of absenteeism as shown in the case of Laos (Yamada et al., 2012). While the lack of adequate incentives and monitoring is often deemed an important cause of prevalent absenteeism among health workers, monitoring and incentive schemes may not work well in the long run if they can be manipulated from the inside, as demonstrated by Banerjee et al. (2008) in a field experiment in India.

The preceding discussion indicates that there is unlikely to be a panacea for the absenteeism of health workers. Nevertheless, it is important to know which regions of the country have high prevalence of absenteeism. This is because the efforts to curb absenteeism are most likely to be fruitful in geographic areas where the issue of absenteeism is most severe, even if different areas may require different kinds of efforts.

In reality, however, policymakers typically have little, if any, information on absenteeism. Even if any data on absenteeism are available, such information is often aggregated to the country level, which is at best useful for comparisons across countries or over time. This study proposes a simple method to enable within-country comparison of the prevalence of absenteeism by combining the following two data sources: Service Delivery Indicator (SDI) survey and Service Provision Assessment (SPA) survey, both of which are becoming increasingly available in sub-Saharan Africa.

## 2 Data

The first data source is the first round of SDI survey in Tanzania conducted in 2010 (World Bank, 2012). The SDI surveys, which have been conducted in a number of sub-Saharan African countries, collect information for assessing the performance of health clinics (and schools) in Africa from the perspective of citizens accessing a service. The survey data typically include the characteristics of health workers, the availability of certain essential drugs, medical equipments, infrastructure of health facility, and information on absenteeism based on unannounced visits. The SDI data used in this study include 175 primary health facilities in Mainland Tanzania. However, some of the regions are not covered in the data such that it is not possible to directly disaggregate the survey estimates of the prevalence of absenteeism to the regional level.

Our second data source is the Tanzania Service Provision Assessment Survey 2006 (National Bureau of Statistics and Macro International, 2007). The SPA surveys, which are conducted as a part of the Demographic and Health Survey, are nationally representative sample surveys of formal sector health facilities drawn from all regions by two-stage random sampling and cover about 10 percent of all health facilities in Mainland Tanzania.[1] The SPA data contain an array of observations on health facility and health workers, some of the questions covered, such as the availability of certain drugs and infrastructure and demographic characteristics of health workers, appear to be comparable to those in the SDI surveys. As such, we can combine these two surveys and make comparisons of the outcome of interest—which is the prevalence of absenteeism in this paper—at a disaggregated level, because the standard sample size of a SPA survey is larger than that of a SDI survey. While SPA surveys cover various types of health facilities, we only use the information of health workers in primary health facilities to match the coverage of the SDI survey.

[1] It also covers 36 percent of all health facilities in Zanzibar, but since SDI survey does not cover the health facilities in Zanzibar, we will not use the observations from Zanzibar.

## 3 Methodology

We impute the indicator of absenteeism from the SDI data into the SPA data. Our method is built on the small-area estimation by Elbers et al. (2002, 2003), in which the individual welfare indicator of interest is repeatedly imputed at the household level and then aggregated up to small areas to obtain point estimates and standard errors. We use a similar method, but because we are only interested in the prevalence of absenteeism, we use a binary regression model with random effects instead of a linear regression. It should be also noted that, even when we do not need to disaggregate the estimates of the prevalence of absenteeism, combining the SDI data with the SPA data (or any data that are larger than the SDI data) potentially has an advantage from the perspective of efficiency, because the means estimated with imputed values can be more accurate (i.e., have lower standard errors) than the mean directly computed from the sample (Matloff (1981) and Fujii and van der Weide (2013)).

Specifically, we consider the following random-effects probit model:

$$y_{ci} = \mathbf{1}(x_{ci}^T\beta + \eta_c + \varepsilon_{ci} > 0), \tag{1}$$

where $y_{ci}$ is the binary outcome variable of interest for individual $i$ in cluster $c$ and $x_{ci}$ the vector of covariates, where an individual and a cluster respectively represent a health worker and a health facility in our application. The cluster- and individual-specific random effects terms, $\eta_c$ and $\varepsilon_{ci}$, are assumed to be orthogonal to each other and normally distributed with the variances of $\sigma_\eta^2$ and unity (by normalization), respectively. We use maximum-likelihood estimation to obtain the point estimates of $\beta$ and $\ln \sigma_\eta^2$ and their variance-covariance matrix.

We then repeatedly impute the outcome variable into the SPA records using a simulation technique similar to Elbers et al. (2002, 2003), but we bootstrap the SPA sample in two stages in each round of the simulation to take into account the sampling design. Following the approach taken by Elbers et al. (2014), we replicate each observation in the bootstrapped SPA sample by the facility weight in each round of simulation to represent the whole mainland Tanzania. As a result, we have an imputation sample for each round of simulation.

Now, let us consider a specific round $t(\in [1, \cdots, T])$ of the simulation, where $T$ is the total rounds of simulation. We draw the simulated parameter $\tilde{\beta}^{(t)}$ and $\ln(\tilde{\sigma}_\eta^{(t)})^2$ from a joint normal distribution with the estimated mean and variance-covariance matrix. The simulated error terms $\tilde{\eta}_c^{(t)}$ [$\tilde{\varepsilon}_{ci}^{(t)}$] are drawn from a normal distribution with the variance of $\tilde{\sigma}_\eta^{(t)}$ [one] for each cluster [individual] in each round. With these draws, we obtain the imputed outcome $\tilde{y}_{ci}^{(t)}$ for each individual in the imputation sample for round $t$ by replacing $y_{ci}$, $\beta$, $\eta_c$, and $\varepsilon_{ci}$ with $\tilde{y}_{ci}^{(t)}$, $\tilde{\beta}^{(t)}$, $\tilde{\eta}_c^{(t)}$, and $\tilde{\varepsilon}_{ci}^{(t)}$ in eq. (1). By taking an average of $\tilde{y}_{ci}^{(t)}$ within region $R$, we obtain a region-level mean outcome $\tilde{Y}_R^{(t)}$ for region $R$ in round $t$. The point estimate $\hat{Y}_R$ for region $R$ is given by the mean of $\tilde{Y}_R^{(t)}$ over $t$ and its standard error $s.e.(\hat{Y}_R)$ by the standard deviation of $\tilde{Y}_R^{(t)}$ over $t$. The lower bound ($p_5$) and upper bound ($p_{95}$) of 90-percent confident interval as well as the estimate of the median ($p_{50}$) are also produced from the corresponding percentiles of $\tilde{Y}_R^{(t)}$ with respect to $t$.

## 4 Results

We run a random-effects probit regression of the indicator variable for the absence from health facility where the random effects are included at the level of health facilities. The regression results are reported in Table 1. It should be noted that our measure of absenteeism only takes into account whether the health worker was present at the health facility. Therefore, those health workers who are absent for a "legitimate" reason such as training will be still counted as being absent. On the other hand, those health workers who are at the health facility will be counted as being present whether or not they are seriously working.

As with Elbers et al. (2002, 2003), our estimates are used for imputation and not for causal inferences. Thus, we are not intrinsically interested in the regression results. Nevertheless, a few points are worth mentioning about them. First, somewhat surprisingly, the prevalence of absenteeism tends to be higher in facilities that have electricity, toilet, and sphygmomanometer. This may be a reflection of better training opportunities in better equipped facilities. Second, the table also shows that older male workers tend to be more likely to be absent, which is in line with the findings of existing literature. Third, the table also shows that there is a significant spatial heterogeneity in the prevalence of absenteeism, which motivates our study.

Finally, Table 1 also shows that the variance $\sigma_\eta^2$ of $\eta$ is small. In fact, the null hypothesis that the intracluster correlation is equal to zero cannot be rejected by a likelihood-ratio test. Therefore, we will subsequently discuss the consequence of using a standard probit model, which corresponds to the case where $\sigma_\eta^2$ is dropped from eq. (1) such that the error terms are independent across individuals. Nevertheless, because the cluster-specific random effects do not cancel out as quickly as the individual-specific random effects through aggregation, we allow $\sigma_\eta^2$ to be strictly positive in our main analysis to produce conservative standard errors.

Based on the regression estimates reported in Table 1, we randomly draw the relevant parameters and error terms for 2,000 rounds of simulation and aggregate up to the level of 21 regions in Mainland Tanzania as reported in Table 2. This table shows that the following regions have relatively low prevalence of absenteeism: Dodoma, Mtwadra, Ruvuma, Iringa, Mbeya, Singida, Tabora, and Rukwa. In particular, the Iringa, Mbeya, Singida, and Tabora regions have significantly lower prevalence of absenteeism than the national average at a 5 percent significance level using the imputed estimates (see also Table 3). On the other hand, the following regions have a relatively high absence rate: Kigoma, Shinyanga, Kagera, Mwanza, and Mara. In fact, all of these regions have a significantly higher absence rate than the national average at a 5 percent level.

Table 1: Random-effects probit regression of the indicator for the absence of health workers.

| Variable | Mean | Coef | | (s.e.) |
|---|---|---|---|---|
| Health facility has electricity | 0.705 | 0.237 | ** | (0.118) |
| Health facility has toilet | 0.927 | 0.321 | | (0.217) |
| Health facility has a sphygmomanometer | 0.896 | 0.754 | *** | (0.196) |
| Age of the health worker | 42.22 | -0.012 | ** | (0.006) |
| Female health worker | 0.734 | -0.089 | | (0.113) |
| Coastal Zone | 0.322 | 0.283 | * | (0.146) |
| Northern Zone | 0.173 | 0.306 | * | (0.171) |
| Lake Zone | 0.257 | 0.738 | *** | (0.148) |
| Constant | | -1.594 | *** | (0.397) |

Source: Tanzania SDI Survey.

Note: Number of observations is 773. Number of clusters is 173. The point estimate and standard error of $\sigma_\eta$ are 0.116 and 0.166, respectively. $^{*}$, $^{**}$, and $^{***}$ denote a statistical significance at 10, 5, and 1 percent levels, respectively. *P*-value for the likelihood ratio test of zero intracluster correlation ($H_0 : \sigma_\eta^2/(1+\sigma_\eta^2) = 0$) is 0.357. Southern Highland and Central Zones are the base category for geographic zones.

Table 2: Estimates of the prevalence of absenteeism of health workers.

| Region | Mean | (s.e.) | $p_5$ | $p_{50}$ | $p_{95}$ |
|---|---|---|---|---|---|
| Dodoma | 17.24 | (4.77) | 11.54 | 16.43 | 25.61 |
| Arusha | 24.73 | (4.94) | 17.71 | 24.16 | 33.76 |
| Kilimanjaro | 25.77 | (5.02) | 18.48 | 25.33 | 34.88 |
| Tanga | 20.84 | (4.99) | 14.03 | 20.19 | 29.96 |
| Morogoro | 26.18 | (4.29) | 20.38 | 25.79 | 33.67 |
| Pwani | 23.61 | (4.55) | 17.57 | 23.04 | 31.70 |
| Dar Es Salaam | 23.16 | (4.46) | 17.34 | 22.52 | 31.08 |
| Lindi | 21.97 | (5.00) | 15.05 | 21.35 | 30.89 |
| Mtwara | 17.77 | (4.85) | 11.99 | 16.95 | 26.55 |
| Ruvuma | 18.62 | (4.64) | 12.71 | 17.96 | 26.46 |
| Iringa | 16.51 | (4.64) | 10.86 | 15.73 | 24.41 |
| Mbeya | 16.83 | (4.53) | 11.45 | 16.06 | 24.53 |
| Singida | 14.76 | (4.87) | 9.10 | 14.03 | 23.20 |
| Tabora | 15.33 | (4.70) | 9.88 | 14.46 | 23.61 |
| Rukwa | 18.88 | (4.66) | 12.91 | 18.24 | 27.56 |
| Kigoma | 40.11 | (4.27) | 33.14 | 40.12 | 47.05 |
| Shinyanga | 32.60 | (4.04) | 26.37 | 32.35 | 39.46 |
| Kagera | 36.78 | (3.89) | 30.44 | 36.79 | 43.14 |
| Mwanza | 33.21 | (4.02) | 27.04 | 32.95 | 40.06 |
| Mara | 37.66 | (4.00) | 31.24 | 37.52 | 44.26 |
| Manyara | 27.10 | (5.23) | 19.15 | 26.73 | 36.43 |

Source: Tanzania SDI Survey and TSPA Survey, 2006.

Table 3: Estimates of the prevalence of absenteeism of health workers by zones.

| Estimation | (1) Survey only | | (2) Imputation | |
|---|---|---|---|---|
| Zone | Mean | (s.e.) | Mean | (s.e.) |
| Coastal Zone | 26.95 | (2.88) | 23.20 | (4.23) |
| Southern Highland Zone | 16.84 | (3.74) | 17.54 | (4.39) |
| Northern Zone | 23.56 | (4.52) | 24.75 | (4.77) |
| Lake Zone | 36.96 | (3.39) | 35.31 | (3.62) |
| Central Zone | 19.31 | (5.16) | 15.86 | (4.59) |
| Mainland Tanzania | 26.68 | (1.77) | 24.74 | (3.27) |

Source: Tanzania SDI Survey and TSPA Survey, 2006.

While the estimates reported in Table 2 are useful for identifying the regions in which the prevalence of absenteeism is high, there may be some concerns about the reliability of such estimates. In particular, our method rests on the assumption that the model parameters estimated with the SDI survey is applicable to the SPA survey. This assumption may be questionable, because there is a gap of four years between the two surveys.

Even though it is not possible for us to prove or disprove this assumption, it is possible to establish the plausibility of our results by comparing at an aggregated level the estimates directly estimated from the SDI survey and the ones based on the imputation described in Section 3. In Table 3, we report the point estimates and standard errors for the former [latter] estimates in column (1) [column (2)] at the level of five zones. As the comparison between the two columns indicates, the differences in the point estimates can be attributed to statistical errors. Therefore, the impuation-based estimates in Table 2 are consistent with the aggregate estimates derived only from the SDI survey.

## 5 Discussion

In this paper, we developed an imputation method and applied to the estimation of the prevalence of absenteeism for health workers in Tanzania by combining SDI and SPA surveys. We chose to present the results based on the cluster errors to be conservative. However, it should be reiterated that we could not reject the null hypothesis of zero intracluster correlation. Therefore, it is not unreasonable to drop the cluster-specific random-effects term. When we do so, the standard errors for regional estimates generally become smaller by around 1-2 percentage points, which makes the accuracy of the regional estimates comparable to the survey-only estimates at the zone level.[2] The point estimates in this case remain similar to those reported in Table 2.

While our conservative standard errors for regional-level estimates in Table 2 are somewhat large, we are able to identify regions that are well above and below the national average.

[2] Detailed results are available upon request.

Therefore, our results provide useful estimates of the prevalence of absenteeism of health workers in Tanzania, which policymakers can use to prioritize the geographic areas for policy intervention.

Since our method does not rely on any Tanzania-specific context, it is readily applicable to the absenteeism of health workers in other countries with SDI and SPA data. Further, since our method is sufficiently general, it can also be readily used for other binary outcome indicators of interest.

## References


Banerjee, A. V., R. Glennerster and E. Duflo (2008) "Putting a Band-Aid on A Corpse: Incentives for Nurses in the Indian Public Health Care System" *Journal of the European Economic Association*. 6(2-3): 487—500.

Belita, A., P. Mbindyo and M. English (2013) "Absenteeism amongst Health Workers—Developing a Typology to Support Empiric Work in Low-Income Counties and Characterizing Reported Associations." *Human Resources for Health* 11:34.

Chaudhury, N., J. Hammer, M. Kremer, K. Muralidharan, and F.H. Rogers (2006) "Missing in Action: Teacher and Health Worker Absence in Developing Countries" *Journal of Economic Perspectives*. 20(1): 91–116.

Elbers, C., J.O. Lanjouw, and P. Lanjouw (2002) "Micro-level Estimation of Welfare." Policy Research Working Paper 2911, World Bank.

Elbers, C., J.O. Lanjouw, and P. Lanjouw (2003) "Micro-level Estimation of Poverty and Inequality." *Econometrica* 71(1), 355–364.

Elbers, C., J.O. Lanjouw, P. Lanjouw, and P.G. Leite (2014) "Poverty and Inequality in Brazil: New Estimates from Combined PPV-PNAD Data." In *Inequality and Economic Development in Brazil,* ed. C.E. Velez, R.P. Barros, and F.G. Ferreira (World Bank Publications) chapter 5, pp. 81–104.

Fujii, T., and R. van der Weide (2013) "Cost-effective Estimation of the Population Mean Using Prediction Estimators." Policy Research Working Paper 6509, World Bank.

Goldstein, M., J.G. Zivin, J. Habyarimana, C. Pop-Eleches, and H. Thirumurthy (2013) "The Effect of Absenteeism and Clinic Protocol on Health Outcomes: The Case of Mother-to-Child Transmission of HIV in Kenya" *American Economic Journal: Applied Economics*. 5(2): 58—85.

Matloff, N. (1981) "Use of Regression Functions for Improved Estimation of Means." *Biometrika* 68, 685–689.

National Bureau of Statistics and Macro International (2007) *Tanzania Service Provision Assessment Survey*. National Bureau of Statistics, United Republic of Tanzania and Macro International Inc.

World Bank (2012) *Service Delivery Indicators: Tanzania*. World Bank.

Yamada, H., Y. Sawada, and X. Luo (2013) "Why is Absenteeism Low among Public Health Workers in Lao PDR?" *Journal of Development Studies*. 49(1): 125—133.