Can existing improvements of water, sanitation, and hygiene (WASH) in urban slums reduce the burden of typhoid fever in these settings?

Justin Im\textsuperscript{1,*}, Md. Taufiqul Islam\textsuperscript{2,*}, Faisal Ahmmed\textsuperscript{2}, Deok Ryun Kim\textsuperscript{1}, Ashraful Islam Khan\textsuperscript{2}, Khalequ Zaman\textsuperscript{2}, Mohammad Ali\textsuperscript{3}, Florian Marks\textsuperscript{1,4}, Firdausi Qadri\textsuperscript{2}, Jerome Kim\textsuperscript{1}, John D Clemens\textsuperscript{2,5}.

1 International Vaccine Institute, Seoul, Republic of Korea
2 International Centre for Diarrheal Disease Research, Bangladesh, Dhaka, Bangladesh
3 Johns Hopkins University, Baltimore, Maryland, USA
4 Department of Medicine, University of Cambridge, Cambridge, United Kingdom
5 UCLA Fielding School of Public Health, Los Angeles, CA 90095-1772, USA

*Both authors contributed equally

Corresponding author contact information:

Justin Im

International Vaccine Institute

SNU Research Park, 1-Gwanak-ro, Gwanak-gu

Seoul 08826, Republic of Korea

justin.im@ivi.int

+82-2-881-1156

© The Author(s) 2020. Published by Oxford University Press for the Infectious Diseases Society of America.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
Main point: A household composite WASH variable developed using a machine learning algorithm, and based on naturally occurring variations in WASH in a Kolkata slum, is predictive of typhoid risk, with residence in a “good” WASH household associated with reduced typhoid risk.
Abstract

Background

Sustained investment in water, sanitation and hygiene (WASH) has lagged in resource-poor settings; incremental WASH improvements may, nonetheless, prevent diseases such as typhoid in disease-endemic populations.

Methods

Using prospective data from a large cohort in urban Kolkata, India, we evaluated whether baseline WASH variables predicted typhoid risk in a training subpopulation (n=28470). We applied a machine learning algorithm to the training subset to create a composite, dichotomous (“good”, “not good”) WASH variable based on four variables and evaluated sensitivity and specificity of this variable in a validation subset (n=28470). We evaluated in Cox regression models whether residents of “good” WASH households experienced lower typhoid risk after controlling for potential confounders. We constructed virtual clusters (radius 50m) surrounding each household to evaluate whether “good” WASH prevalence modified typhoid risk in central household members.

Results

“Good” WASH was associated with protection in analyses of all households (Hazard ratio (HR)=0.57, 95% CI: 0.37-0.90, p=0.015). This protection was evident in persons ≥5 years at baseline (HR=0.47, 95% CI: 0.34-0.93, p=0.005) and was suggestive, though not statistically significant, in younger age groups (HR=0.61, 95% CI: 0.27-1.38, p=0.235). The level of surrounding household “good” WASH coverage was also associated with protection (HR=0.988, 95% CI: 0.979-0.996, p=0.004, for each percent coverage increase). However, collinearity between household WASH and WASH coverage prevented assessment of their independent predictive contributions.
Conclusion

In this typhoid-endemic setting, natural variation in household WASH was associated with typhoid risk. If replicated elsewhere, these findings suggest that WASH improvements short of major infrastructural investments may enhance typhoid control.

**Key words:** “water, sanitation and hygiene”; WASH; typhoid fever; protection; recursive partitioning
Introduction

Major infrastructural improvements in clean municipal water and sanitary treatment and disposal of human waste have eliminated typhoid fever, an infection caused by *Salmonella enterica* serovar Typhi (*S. Typhi*), as a significant public health problem in affluent countries. However, typhoid continues to affect populations of low- and middle-income countries (LMICs) (1-3), and worldwide, there are approximately 10.9 million cases and 116,800 deaths annually (4). Because resources in LMICs are generally too limited to enable the major infrastructural improvements achieved by more affluent countries in the foreseeable future, it may be questioned whether incremental improvements in water quality, hygiene and sanitation that are feasible and affordable in these settings will be sufficient to have a measurable impact on typhoid burden. One way of approaching this problem is to evaluate whether natural variations in the level of household water, sanitation, and hygiene (WASH) that are already present in poor, typhoid-endemic settings, and thus are affordable and feasible, impart a measurable impact on the risk of typhoid fever in those settings. In this paper, we reanalyze longitudinal data from a Kolkata slum population under comprehensive surveillance for typhoid to address this question.

Methods

*The Vi Polysaccharide trial in Kolkata*

A cluster randomized trial of the effectiveness of typhoid Vi polysaccharide (ViPS) vaccine in a poor slum area of Kolkata, where typhoid is endemic, was conducted in individuals 2 years of age or older between 2004 and 2006 (5). Two years prior to vaccine dosing for the trial, surveillance for enteric fever was initiated in the trial area and continued for two years following vaccination. All households and individuals within the study area were enumerated through a baseline census conducted at the beginning of surveillance activities. Periodic census updates were conducted throughout the study period, including immediately prior to vaccination, 1-year post-vaccination, and at the completion of the surveillance period. Census methods included the enumeration of individuals...
and collection of geolocation, demographic, socio-economic, and WASH information of the household and its residents. WASH information was collected during the baseline census only. Passive surveillance for typhoid was conducted at 5 study clinics, where residents of the study area presenting with a history of fever for at least 3 days received blood cultures which were evaluated for *S. Typhi* as described elsewhere (5). Visits for fever in which the date of onset of fever was within 14 days of the date of discharge for the previous visit were grouped together into febrile episodes. All subjects or their guardians provided written informed consent. Typhoid fever was defined as a febrile episode in which at least one blood culture yielded *S. Typhi*. The overall protective effect of the Vi vaccine was 61% (95% CI: 41-75) when measured against a hepatitis A vaccine control (5).

**Selection of WASH variables for analysis**

We analyzed whether variables characterizing the level of household WASH, ascertained in the baseline census, predicted the risk of typhoid fever in the baseline closed cohort over the ensuing four years of clinical surveillance for typhoid fever. In the census there were 5 non-binary household variables characterizing household source of drinking water, treatment of water for daily use, site of defecation, handwashing practice after defecation, and waste disposal location. We grouped the categories of each of these variables into two classifications reflecting “good” or “not good” WASH based on substantive judgements and the distribution of the population according to the different categories, but without prior knowledge of the rates of typhoid associated with each component category. We then divided the baseline population at random into two equal sized, mutually exclusive subsets, a “training” population and a “validation” population, and selected the four binary WASH variables that were associated with the hazard of typhoid fever in Cox regression models at p<0.1, after first ascertaining whether the variables met the proportional hazards assumption. These variables were household source of drinking water (private tap, private well, or private pump versus other); household daily water treatment (filtered or boiled versus other), site of defecation (private or shared flush toilet versus other), and handwashing after defecation (always with soap versus other).
Development of the WASH decision tree to predict typhoid in the training subpopulation

To create an overall binary composite variable for household WASH predicting typhoid in household members, we used a machine learning algorithm to develop a decision tree constructed with recursive partitioning (6). The decision tree was designed to predict development of typhoid fever among baseline household members over 4 years of follow-up, using 4 dichotomous household WASH variables independently associated with typhoid hazard. We initially ran the algorithm for the training population, specifying a default loss function of 1:1 for the ratio of costs of false negative and false positive classifications, and requiring that each terminal node have at least 300 observations. For cross-validation, the training population was randomly partitioned into 10 parts; 1 part was used in estimating cross-validation error. To find the optimal typhoid prediction tree in the training population, the model was pruned by the minimal complexity parameter (CP) corresponding to a minimum error with at least 2 terminal nodes in the tree. A Receiver Operating Characteristic (ROC) Curve was constructed for the selected rule, and the area under the ROC curve (AUC) was estimated (7, 8). We used maximization of the Youden index to select the cutoff probability for the ROC demarcating the overall binary WASH variable into “good” versus “not good” household WASH in relation to the risk of typhoid in inhabitants of the household.

Further analyses

Because analysis of predictive rules developed in a training population may overestimate the performance of the rule if analyzed in the same subpopulation, we evaluated the rule in a separate validation population. Procedures for analysis of the validation subpopulation were the same as for the training subpopulation. After confirming that the predictive rule for “good” WASH performed similarly in the training and validation populations with respect to sensitivity (the proportion individuals developing typhoid who lived in households with “not good” WASH) and specificity (the proportion of individuals not developing typhoid who lived in households with “good” WASH) in predicting typhoid, we undertook further analyses in the total population. First, because the first 2
years of surveillance occurred in a population that had not been vaccinated with ViPS, whereas the second 2 years occurred during the ViPS trial, we evaluated whether the predictive rule performed similarly with respect to sensitivity and specificity before and during the ViPS trial. Second, we evaluated in the total population the protection against typhoid associated with “good” WASH in the household. To measure the protective association between “good” WASH in the household and the subsequent risk of typhoid fever in household inhabitants, we used Cox proportional hazard regression models after evaluating proportionality assumptions for independent variables. To evaluate independent variables not fulfilling proportionality assumptions, we included the variables as interaction variables with time of follow-up. Models were adjusted for age in years with time interaction, number of household members with time interaction, and monthly household expenditure (in Indian rupees), all of which were associated at p<0.05 with the hazard of typhoid in bivariate Cox models. The hazard ratio (HR) for culture positive typhoid fever was estimated by exponentiation of the coefficient of living in a household with “good” WASH in the model, and the 95% confidence interval (CI) for the HR was estimated using a robust sandwich method where each household was assumed as a cluster in the model. We also used Cox models to evaluate the protective association between the percentage of “good” WASH households ascertained by GIS in the baseline census as located within a 50m radius of the central household under analysis, on the one hand, and the rate of typhoid in the total population during the ensuing four years, on the other. The analysis was performed using “rpart” package for decision tree modeling, “rpart.plot” package for tree plotting, “pROC” packages for ROC curve, “survival” for Cox model and “dplyr” package for data management under R-Studio analytical software (9-11). In all statistical analyses, p<0.05 (2-tailed) was taken as the margin of statistical significance.
Results

Characteristics of the baseline population

During the first census, 10944 households and 56940 residents were enumerated and included in the study closed cohort (Figure 1). During the follow-up period, 276 incident cases of typhoid were identified (132 cases per 100000 person-years), from which 4 were recurrent secondary cases and excluded from analysis. Of 272 typhoid episodes analyzed, 113 occurred before the trial and 159 during the trial. The total population, matched on individuals, was randomly partitioned into equal sized training and validation subsets with 28470 individuals in each and 5415, and 5529 households respectively. The mean age was 27.8 years in both populations; and the proportion of males in each population was 53.8% and 54.2%, respectively. Reported typhoid cases were also distributed evenly in the two populations, 136 cases each.

Development and evaluation of the WASH decision tree to predict typhoid

Bivariate analysis of individual WASH variables and the risk of typhoid in the training population showed an association for private or shared flush toilet (HR=0.30, 95% CI: 0.1-0.95, p=0.041), source of drinking water from private tap, well, or pump (HR=0.31, 95% CI: 0.14-0.71, p=0.005), and water treatment by filter or boiling for daily use (HR=0.45, 95% CI: 0.18-1.1, p=0.081) (Table 1). There was some evidence for association with always hand washing with soap after defecation (HR=0.74, 95% CI: 0.52-1.05, p=0.097) and no evidence for association with waste disposal practice (p=0.958).

In the training population, the 4 WASH variables associated with risk of typhoid (p<0.10) were included in a decision tree model for binary classification of household-level WASH. The optimal cutoff probability maximized the Youden Index (0.004343) applied to an ROC curve that had an AUC=58% (95% CI: 54-61) (Figure 2). The resulting decision tree delineated 2 rules for “not
good” WASH and 4 rules for “good” WASH (Figure 3). The dominant bifurcation in the model was primary source of drinking water for the household, where private ownership of a tap, well or pump, regardless of other WASH characteristics indicated “good” household WASH (Figure 3). The sensitivity and specificity of the classification criterion to predict typhoid given by the model was 91.2% (95% CI: 85.1-95.4) and 19.5% (95% CI: 19.0-19.9), respectively, in the training set, and evaluation of the rule in the validation population yielded similar results, 90.4% (95% CI: 84.2-94.8) and 19.8% (95.5 CI: 19.3-20.3). We therefore applied the rule to the total population. Evaluation of the composite variable in the total population before the trial and during the trail resulted in a sensitivity and specificity of 92.0% (95% CI: 85.4-96.3) and 19.6% (95% CI: 19.3-19.9), respectively, before and 89.9% (95% CI: 84.2-94.1) and 19.6% (95% CI: 19.3-19.9) during the trial.

Evaluation of the protective association between WASH and the rate of typhoid

In the total population, in comparison with residence in a “not good” WASH household, residence in a “good” WASH household was associated with 59% (HR=0.41, 95% CI: 0.27-0.63, p<0.001) crude reduction in hazard for culture confirmed typhoid overall and a 43% (HR=0.57 (95% CI: 0.37-0.90, p=0.015) hazard reduction in the adjusted model (Table 2). When stratified by age at baseline, both the younger and older age groups residing in “good” WASH households experienced a reduction in typhoid hazard, 39% (HR=0.61, 95% CI: 0.27-1.38, p=0.235) and 44% (HR=0.56 95% CI: 0.34-0.93, p=0.005), respectively; the association was statistically significant only in the older age group, likely due to the lower number of typhoid cases in the younger age group. We next stratified the population into high versus low levels of “good” WASH in households surrounding the central household within 50m virtual clusters, taking the cut point (34%) as the prevalence of “good” WASH at which there was a noticeable decrease in the incidence of typhoid in the central households (Figure 4). Our data failed to show that the level of “good” WASH coverage in the community surrounding a central household under analysis modified the protective association between the level of household WASH and the risk of typhoid among members in the central household (Table 3). We next measured
the association between the level of surrounding households with “good” WASH and the risk of typhoid in members of the central household in a model that controlled for potentially confounding variables (age, household size, and household expenditure). In this analysis, we found that the association was significant (HR=0.988, 95% CI: 0.979-0.996, p=0.004, for each percent increase of coverage). However, it was not possible to evaluate the independent protective associations with typhoid fever for household WASH on the one hand and level of neighborhood WASH on the other hand due to collinearity between these two variables.

Discussion

We used recursive partitioning to develop a dichotomous classification rule for household-level “good” WASH based on 4 independent non-binary WASH variables collected at baseline. When applied to the study population, residence in a “good” WASH household was associated with a moderate decrease in risk for contracting typhoid during the study period. When analyzed separately by age group, the protective association was significant for the older age group and suggestive for the younger age group. The neighborhood level of “good” WASH coverage appeared to be associated with the risk of typhoid among members of central households, however it did not modify the association between the level of WASH in the central household and the risk of typhoid in household members.

The findings of this study must be considered in light of several important limitations. Firstly, our construction of a composite WASH variable to classify households relied on very simply ascertained WASH variables that were collected at baseline only as potential confounders for analysis of the vaccine trial. More in depth characterization of WASH variables might have resulted in a more powerfully predictive WASH composite variable and created the opportunity for further assessment of independent contributions of WASH component variables. For example, although it is likely that drinking water from a household’s private tap, well, or pump was microbiologically superior to that
from other sources, we do not have direct data on this issue. Secondly, WASH variables were ascertained at baseline and the population was followed for 4 years. Although the population was quite stable, it is possible that the WASH status of households changed over time. However, such misclassification bias, if random, would have made our analyses conservative. Thirdly, we used a machine learning algorithm that created a dichotomous composite WASH variable. Although useful, particularly in classifying the level of WASH coverage in surrounding households, this dichotomization may have resulted in some loss of information about the variables, again making our analyses conservative. Fourthly, our analysis of the protective association of household WASH and the risk of typhoid by age was limited by sparse data in the younger age stratum. Fifthly, we partitioned the training and validation sets by individuals rather than households, with the result that there were some shared households in the training and validation sets. However, sensitivity and specificity values for the composite variable were nearly identical in the training and validation subsets when we excluded households that were shared by the two subpopulations. Finally, the period of follow-up for our study included periods before and during a ViPS vaccine trial. However, as indicated in the results, separate analyses of the predictive rule during these periods revealed consistent results.

Despite these potential limitations, several factors strengthened our findings. First, the analyses utilized data prospectively collected in the context of a clinical trial for which population information was collected systematically and typhoid ascertainment was done in a comprehensive manner and was confirmed by blood culture positivity. Secondly, the population was relatively stable over time, and thus patterns of migration were unlikely to affect our assessments. And lastly, our dataset was large enough to both train and validate our model in mutually exclusive populations to minimize overfitting.
These results indicate that improvements to WASH short of major municipal infrastructure developments might have an appreciable effect on the risk of typhoid in urban settings of South Asia. The results reinforce the notion that household transmission was an important contributor to disease transmission in this urban setting, and moreover, that household source of drinking water appears to be a dominant determinant of typhoid risk, although the effect of drinking water source can be modified by a variety of other WASH variables. This evaluation also suggests that the effect of WASH in the household was not modified by the level of WASH in the surrounding neighborhood.

The ability of our analysis to demonstrate that household WASH may influence the risk of typhoid in household residents encourages future studies of this topic. We recommend that future studies characterize WASH in a more comprehensive fashion, including WASH behaviors, and that cohort studies of this topic monitor and analyze changes in household WASH that occur during follow-up. As well, because of the theoretical likelihood that protection conferred by newer generation typhoid vaccines may be modified by the intensity of the typhoid inoculum to which vaccinees are exposed, it will be of great interest to examine the interactions between levels of household WASH and typhoid vaccine protection in future vaccine evaluations.
Funding

This work was supported by the Bill & Melinda Gates Foundation (OPP1171432). The research was supported by the Diseases of the Most Impoverished Program of the Bill & Melinda Gates Foundation.

Conflict of interest

The authors state no commercial or other association that might pose a conflict of interest with regards to the findings presented in this manuscript.
References

1. Tack B, Phoba MF, Van Puyvelde S, Kalonji LM, Hardy L, Barbe B, et al. Salmonella Typhi From Blood Cultures in the Democratic Republic of the Congo: A 10-Year Surveillance. Clinical Infectious Diseases. 2019;68(Supplement_2):S130-s7.

2. Mogasale V, Maskery B, Ochiai RL, Lee JS, Mogasale VV, Ramani E, et al. Burden of typhoid fever in low-income and middle-income countries: a systematic, literature-based update with risk-factor adjustment. The Lancet Global Health. 2014;2(10):e570-80.

3. Antillón M, Warren JL, Crawford FW, Weinberger DM, Kürüm E, Pak GD, et al. The burden of typhoid fever in low- and middle-income countries: A meta-regression approach. PLoS Neglected Tropical Diseases. 2017;11(2):e0005376.

4. GBD 2017 Typhoid and Paratyphoid Collaborators. The global burden of typhoid and paratyphoid fevers: a systematic analysis for the Global Burden of Disease Study 2017. Lancet Infect Dis. 2019;19(4):369-381. doi:10.1016/S1473-3099(18)30685-6.

5. Sur D, Ochiai RL, Bhattacharya SK, Ganguly NK, Ali M, Manna B, et al. A cluster-randomized effectiveness trial of Vi typhoid vaccine in India. The New England Journal of Medicine. 2009;361(4):335-44.

6. Strobl C, Malley J, Tutz G. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. Psychological Methods. 2009;14(4):323-48.

7. Hand DJ. Measuring classifier performance: a coherent alternative to the area under the ROC curve. Machine Learning. 2009;77(1):103-23.

8. Faraggi D. RB. Estimation of the area under the ROC curve. Statistics in Medicine. 2002;21(20):3093-106.
9. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12:77.

10. Terry M. Therneau PMG. Modeling Survival Data: Extending the Cox Model. New York: Springer; 2000.

11. Hadley Wickham RF, Lionel Henry and Kirill Müller dplyr: A Grammar of Data Manipulation. R package version 0.7.6.; 2018.

12. Ali M, Sur D, Kanungo S, Qadri F, Kim DR, Islam T, et al. Re-evaluating herd protection by Vi typhoid vaccine in a cluster randomized trial. International Health. 2020;12(1):36-42.

13. Ali M, Qadri F, Kim DR, Islam T, Im J, Ahmmed F, et al. Unmasking herd protection by an oral cholera vaccine in a cluster-randomized trial. International Journal of Epidemiology. 2019;48(4):1252-61.
Tables

Table 1: Bivariate analysis of WASH variable and typhoid risk in the training population

| WASH variable                        | Yes       |     |     |     | No       |     |     |     | Hazards Ratios [HRs (95% CI)] | Crude HRs | p-value |
|--------------------------------------|-----------|-----|-----|-----|----------|-----|-----|-----|--------------------------------|-----------|---------|
|                                      | n         | Typhoid | PY  | IR per | n         | Typhoid | PY  | IR per | Crude HRs | p-value |
| Site of defecation:                  |           |         |     |        |           |         |     |        |              |           |         |
| Private or shared flush toilet       | 1956      | 3      | 7090| 42    | 26514     | 133    | 96063| 138   | 0.30 (0.1,0.95) | 0.041     |         |
| Main source of drinking water:       |           |         |     |        |           |         |     |        |              |           |         |
| Private tap, well, or pump:          | 3655      | 6      | 13288| 45    | 24815     | 130    | 89865| 145   | 0.31 (0.14,0.71) | 0.005     |         |
| Treatment of water for:              | 2189      | 5      | 8008 | 62    | 26281     | 131    | 95146| 138   | 0.45 (0.18,1.1)  | 0.081     |         |
### daily use:

| Filtered or boiled |  |  |  |  |  |  |  |  |  |
|--------------------|---|---|---|---|---|---|---|---|---|
| Hand wash practice after defecation: |  |  |  |  |  |  |  |  |  |
| Always with soap | 19286 | 83 | 69926 | 119 | 9184 | 53 | 33227 | 160 | 0.74 (0.52, 1.05) | 0.097 |
| Waste disposal location: |  |  |  |  |  |  |  |  |  |
| Specific | 28067 | 134 | 101693 | 132 | 403 | 2 | 1461 | 137 | 0.96 (0.25, 3.79) | 0.958 |

**Note:** p-values are adjusted using robust standard error; each household is treated as a cluster

PY = person years, IR = incidence rate
Table 2: Household-level protection of Good WASH in the total population

| Age   | Good WASH | Not Good WASH | Hazard Ratios [HRs (95% CI)] |
|-------|-----------|---------------|-------------------------------|
|       | N Typhoid PY | IR per 10^5 PY | N Typhoid PY | IR per 10^5 PY | Crude HRs | p-value | Adj. HRs | p-value |
| All   | 11154     | 25           | 40598        | 62           | 45786     | 247      | 165842  | 149     | 0.41 (0.27,0.63) | <0.001 | 0.57 (0.37,0.90)* | 0.015 |
| 0 to <5 | 708     | 7            | 2599         | 269          | 3693      | 58       | 13407   | 433     | 0.62 (0.28,1.35) | 0.224 | 0.61 (0.27,1.38)† | 0.235 |
| ≥5    | 10446     | 18           | 37998        | 47           | 42093     | 189      | 152435  | 124     | 0.38 (0.24,0.62) | <0.001 | 0.56 (0.34,0.93)‡ | 0.024 |

*Note: p-values are adjusted using robust standard error; each household is treated as a cluster

PY = person years, IR = incidence rate

* HR is adjusted by age in years, age in years with time interaction, number of household members, number of household members with time interaction, average household expenditure

† HR is adjusted by age in years, number of household members, number of household members with time interaction, average household expenditure

‡ HR is adjusted by age in years, number of household members, average household expenditure
Table 3: Household-level protection of WASH stratified by WASH coverage in the community

| Coverage (%) | Good Wash | | Not Good Wash | | Hazard Ratios [HRs (95% CI)] |
|--------------|-----------|-----------|----------------|-----------|---------------------------|
|              | n         | Typhoid   | IR per 10^5 PY | n          | Typhoid | IR per 10^5 PY | Crude HRs | p-value | Adj. HRs | p-value |
| 0 to ≤34     | 5589      | 15        | 20367          | 74         | 41070   | 226         | 148752     | 152 | 0.48 (0.29, 0.81) | 0.006 | 0.63 (0.37, 1.07)* | 0.085 |
| >34 to 100   | 5565      | 10        | 20231          | 49         | 4716    | 21          | 17090      | 123 | 0.40 (0.18, 0.91) | 0.029 | 0.52 (0.21, 1.28)† | 0.154 |

Note: p-values are adjusted using robust standard error; each household is treated as a cluster

PY = person years, IR = incidence rate

* HR is adjusted by age in years, age in years with time interaction, number of household members, average household expenditure

† HR is adjusted by age in years, number of household members, average household expenditure
Figure legend

Figure 1: Flow diagram showing population at first census and incident typhoid cases before and after vaccination classified by residence in “Good” or “Not Good” WASH household

Figure 2: Receiver Operator Characteristic (ROC) curve for defining cutoff probability for defining dichotomous composite WASH variable

Note: AUC = 58% (95% CI: 54-61); best cutoff using the Youden Index is 0.004343

Figure 3: Decision tree for categorization of “Good” and “Not Good” WASH based on 4 dichotomous WASH variables

*(Typhoid episodes/household population)

Figure 4: Relationship between neighborhood WASH coverage and risk of typhoid

Note: Typhoid incidences were calculated for 40 equal subgroups of WASH coverage; a smooth line was drawn using a locally estimated scatterplot smoothing (LOESS) function
Figure 1

Population at first census
n=56940

Outmigration
n=4241

Typhoid episodes
n=276

Recurrent episodes
n=4

Before vaccination
n=113

After vaccination
n=159

Good WASH
n=9

Not Good WASH
n=104

Good WASH
n=16

Not Good WASH
n=143
Figure 3

Drinking water
  +-----------------+-----------------+
  | Yes             | No              |
  +-----------------+-----------------+
  | Hand wash with  | Flushing toilet |
  | soap            |                 |
  +-----------------+-----------------+
  | Yes             | No              |
  +-----------------+-----------------+
  | Flush toilet    |                 |
  +-----------------+-----------------+
  | Yes             | No              |
  +-----------------+-----------------+
  | Daily water use |                 |
  +-----------------+-----------------+
  | Good (13/7353)* | Good (1/908)*  |
  | Good (5/1793)*  |                 |
  +-----------------+-----------------+
  | Not Good (15/630566)* | Good (6/1100)* |
  | Not Good (91/15220)*         |                 |
Figure 4

Note: Typhoid incidences are calculated for 40 equal subgroups of each coverage.