POC

Andrew Mertens

Sept 28, 2020

## Descriptive statistics

Descriptive statistics of measured children by country

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Country | N | Child age | Min child age | Max child age |
| Bangladesh | 68,982 | 2.5 | 0.00 | 5 |
| PakistanPunjab | 73,384 | 2.4 | -0.03 | 5 |
| Zimbabwe | 13,428 | 2.5 | 0.01 | 5 |

Descriptive statistics of growth outcomes by country

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Country | N haz | Mean haz | Prev stunting | Prev sev stunting | N whz | Mean whz | Prev wasting | Prev sev wasting |
| Bangladesh | 22,108 | -1.3 | 28 | 8.5 | 22,065 | -0.66 | 9.9 | 2.19 |
| PakistanPunjab | 38,047 | -1.4 | 32 | 11.5 | 37,779 | -0.42 | 7.5 | 2.00 |
| Zimbabwe | 5,964 | -1.2 | 23 | 5.4 | 5,962 | -0.05 | 2.9 | 0.35 |

Descriptive statistics of diarrhea and acute respiratory infection outcomes by country

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Country | N diarrhea meas | N diarrhea | Prev diarrhea | N ari meas | N ari | Prev ari |
| Bangladesh | 23,089 | 1,589 | 6.9 | 68,982 | 530 | 0.77 |
| PakistanPunjab | 39,783 | 5,434 | 13.7 | 73,384 | 3,747 | 5.11 |
| Zimbabwe | 6,096 | 853 | 14.0 | 13,428 | 109 | 0.81 |

Binary measures of improved WASH conditions by country

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Country | N households | N imp wat | N imp san | N imp hygeine | N imp wash |
| Bangladesh | 61,242 | 1,405 | 10,160 | 11,348 | 16,626 |
| PakistanPunjab | 51,660 | 982 | 11,392 | 4,342 | 13,634 |
| Zimbabwe | 11,091 | 2,201 | 3,197 | 3,570 | 5,903 |

Binary measures of E.coli contamination by country

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Country | N households | N ec household | N ec source | N safely manh20 | N imp wash and no ec |
| Bangladesh | 61,242 | 5,031 | 2,319 | 5,035 | 4,708 |
| PakistanPunjab | 51,660 | 4,543 | 2,395 | 4,593 | 5,159 |
| Zimbabwe | 11,091 | 1,675 | 1,117 | 1,684 | 1,791 |

Drinking water conditions by country

|  |  |  |  |
| --- | --- | --- | --- |
| Country | Surface water | Unimproved | Na |
| Bangladesh | 764 | 631 | 59,847 |
| PakistanPunjab | 412 | 330 | 50,918 |
| Zimbabwe | 673 | 1,504 | 8,914 |

Sanitation conditions by country

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Country | No facility | Unimproved | Limited | Basic |
| Bangladesh | 1,186 | 8,974 | 12,981 | 38,101 |
| PakistanPunjab | 7,866 | 3,526 | 5,764 | 34,504 |
| Zimbabwe | 2,371 | 826 | 3,988 | 3,906 |

Hygeine conditions by country

|  |  |  |  |
| --- | --- | --- | --- |
| Country | None | Limited | Na |
| Bangladesh | 2,286 | 49,271 | 9,685 |
| PakistanPunjab | 1,839 | 48,698 | 1,123 |
| Zimbabwe | 705 | 9,945 | 441 |

Household E.coli contamination by country

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Country | 4 | 3 | 2 | 1 | Na |
| Bangladesh | 1,903 | 1,894 | 1,234 | 1,109 | 55,102 |
| PakistanPunjab | 951 | 1,322 | 2,270 | 3,151 | 43,966 |
| Zimbabwe | 644 | 622 | 409 | 449 | 8,967 |

Source water E.coli contamination by country

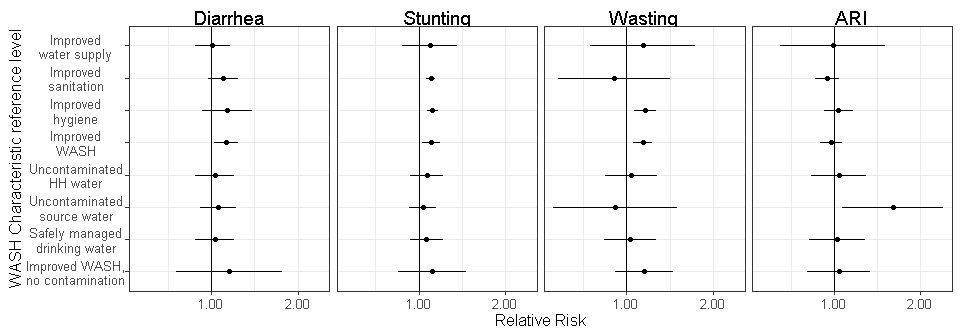
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Country | 4 | 3 | 2 | 1 | Na |
| Bangladesh | 309 | 681 | 1,329 | 3,750 | 55,173 |
| PakistanPunjab | 424 | 587 | 1,384 | 4,455 | 44,810 |
| Zimbabwe | 395 | 341 | 381 | 926 | 9,048 |

## Covariate information

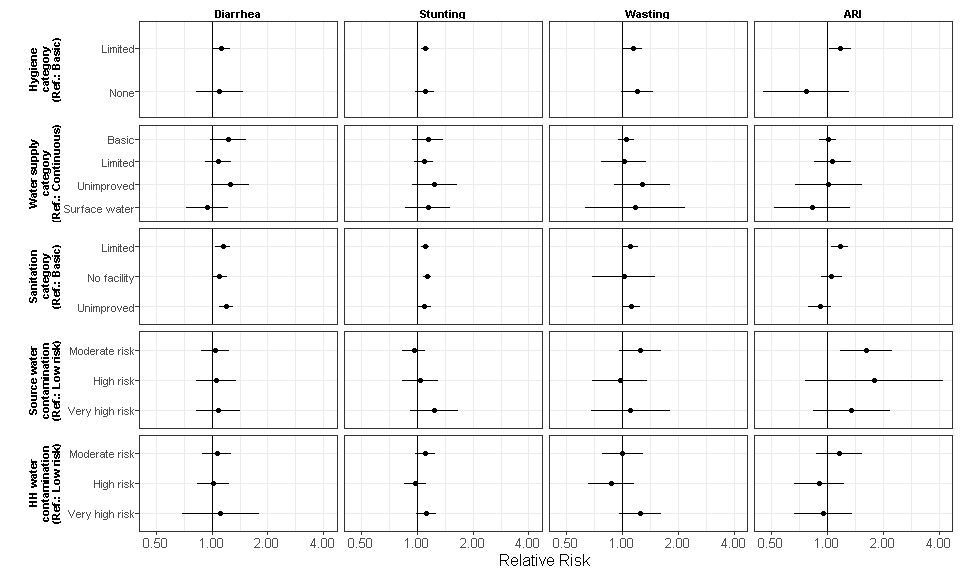
Descriptive statistics of covariate by country

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Var.1 | Bangladesh.n.23103. | Pakistanpunjab.n.39835. | Zimbabwe.n.6100. | Overall.n.69038. |
| country |  |  |  |  |
| Bangladesh | 23103 (100%) | 0 (0%) | 0 (0%) | 23103 (33.5%) |
| PakistanPunjab | 0 (0%) | 39835 (100%) | 0 (0%) | 39835 (57.7%) |
| Zimbabwe | 0 (0%) | 0 (0%) | 6100 (100%) | 6100 (8.8%) |
| Caregiver education |  |  |  |  |
| none | 6923 (30.0%) | 15583 (39.1%) | 237 (3.9%) | 22743 (32.9%) |
| primary | 6959 (30.1%) | 7433 (18.7%) | 2141 (35.1%) | 16533 (23.9%) |
| secondary | 6435 (27.9%) | 12879 (32.3%) | 3080 (50.5%) | 22394 (32.4%) |
| higher | 2771 (12.0%) | 3940 (9.9%) | 635 (10.4%) | 7346 (10.6%) |
| missing | 15 (0.1%) | 0 (0%) | 7 (0.1%) | 22 (0.0%) |
| Maternal age |  |  |  |  |
| Mean (SD) | 42.4 (14.2) | 46.5 (14.9) | 41.6 (14.4) | 44.7 (14.8) |
| Median [Min, Max] | 38.0 [17.0, 95.0] | 43.0 [15.0, 98.0] | 38.0 [15.0, 95.0] | 40.0 [15.0, 98.0] |
| Child age (days) |  |  |  |  |
| Mean (SD) | 916 (525) | 898 (526) | 927 (528) | 906 (526) |
| Median [Min, Max] | 919 [0, 1830] | 892 [-11.0, 1840] | 934 [3.00, 1830] | 905 [-11.0, 1840] |
| Child sex |  |  |  |  |
| 1 | 11952 (51.7%) | 20546 (51.6%) | 3040 (49.8%) | 35538 (51.5%) |
| 2 | 11151 (48.3%) | 19289 (48.4%) | 3060 (50.2%) | 33500 (48.5%) |
| Birth order |  |  |  |  |
| Mean (SD) | 1.77 (0.657) | 2.15 (0.851) | 2.07 (0.748) | 2.01 (0.802) |
| Median [Min, Max] | 2.00 [1.00, 4.00] | 2.00 [1.00, 4.00] | 2.00 [1.00, 4.00] | 2.00 [1.00, 4.00] |
| Urban/rural |  |  |  |  |
| Rural | 18797 (81.4%) | 29342 (73.7%) | 4241 (69.5%) | 52380 (75.9%) |
| Urban | 4306 (18.6%) | 10493 (26.3%) | 1859 (30.5%) | 16658 (24.1%) |
| Currently breastfeeding |  |  |  |  |
| 1 | 10832 (46.9%) | 11652 (29.3%) | 1769 (29.0%) | 24253 (35.1%) |
| 2 | 2213 (9.6%) | 10536 (26.4%) | 1759 (28.8%) | 14508 (21.0%) |
| missing | 10058 (43.5%) | 17647 (44.3%) | 2572 (42.2%) | 30277 (43.9%) |
| Household size |  |  |  |  |
| Mean (SD) | 5.39 (2.17) | 8.40 (4.25) | 5.49 (2.28) | 7.13 (3.82) |
| Median [Min, Max] | 5.00 [2.00, 29.0] | 7.00 [2.00, 45.0] | 5.00 [2.00, 20.0] | 6.00 [2.00, 45.0] |
| Number of children under 5 |  |  |  |  |
| Mean (SD) | 1.30 (0.553) | 2.18 (1.20) | 1.46 (0.681) | 1.82 (1.07) |
| Median [Min, Max] | 1.00 [1.00, 6.00] | 2.00 [1.00, 11.0] | 1.00 [1.00, 5.00] | 2.00 [1.00, 11.0] |
| Household floor |  |  |  |  |
| unimproved | 14804 (64.1%) | 14684 (36.9%) | 1697 (27.8%) | 31185 (45.2%) |
| improved | 8294 (35.9%) | 25128 (63.1%) | 4390 (72.0%) | 37812 (54.8%) |
| missing | 5 (0.0%) | 23 (0.1%) | 13 (0.2%) | 41 (0.1%) |
| Cookstove |  |  |  |  |
| unimproved | 19636 (85.0%) | 20592 (51.7%) | 4271 (70.0%) | 44499 (64.5%) |
| improved | 3465 (15.0%) | 19169 (48.1%) | 1818 (29.8%) | 24452 (35.4%) |
| missing | 2 (0.0%) | 74 (0.2%) | 11 (0.2%) | 87 (0.1%) |
| Chimney |  |  |  |  |
| no | 18413 (79.7%) | 20870 (52.4%) | 33 (0.5%) | 39316 (56.9%) |
| yes | 815 (3.5%) | 2030 (5.1%) | 11 (0.2%) | 2856 (4.1%) |
| missing | 3875 (16.8%) | 16935 (42.5%) | 6056 (99.3%) | 26866 (38.9%) |
| Cooking fuel |  |  |  |  |
| clean | 5 (0.0%) | 11 (0.0%) | 21 (0.3%) | 37 (0.1%) |
| solid | 19714 (85.3%) | 24322 (61.1%) | 4336 (71.1%) | 48372 (70.1%) |
| missing | 3384 (14.6%) | 15502 (38.9%) | 1743 (28.6%) | 20629 (29.9%) |
| Roof construction |  |  |  |  |
| unimproved | 255 (1.1%) | 933 (2.3%) | 1508 (24.7%) | 2696 (3.9%) |
| improved | 22840 (98.9%) | 38735 (97.2%) | 4581 (75.1%) | 66156 (95.8%) |
| missing | 8 (0.0%) | 167 (0.4%) | 11 (0.2%) | 186 (0.3%) |
| Wall material |  |  |  |  |
| natural | 2736 (11.8%) | 5944 (14.9%) | 744 (12.2%) | 9424 (13.7%) |
| rudimentary | 411 (1.8%) | 483 (1.2%) | 19 (0.3%) | 913 (1.3%) |
| finished | 19926 (86.2%) | 33274 (83.5%) | 5301 (86.9%) | 58501 (84.7%) |
| missing | 30 (0.1%) | 134 (0.3%) | 36 (0.6%) | 200 (0.3%) |
| House size (rooms) |  |  |  |  |
| Mean (SD) | 2.23 (1.26) | 2.16 (1.54) | 2.03 (0.982) | 2.17 (1.41) |
| Median [Min, Max] | 2.00 [1.00, 20.0] | 2.00 [1.00, 82.0] | 2.00 [1.00, 9.00] | 2.00 [1.00, 82.0] |

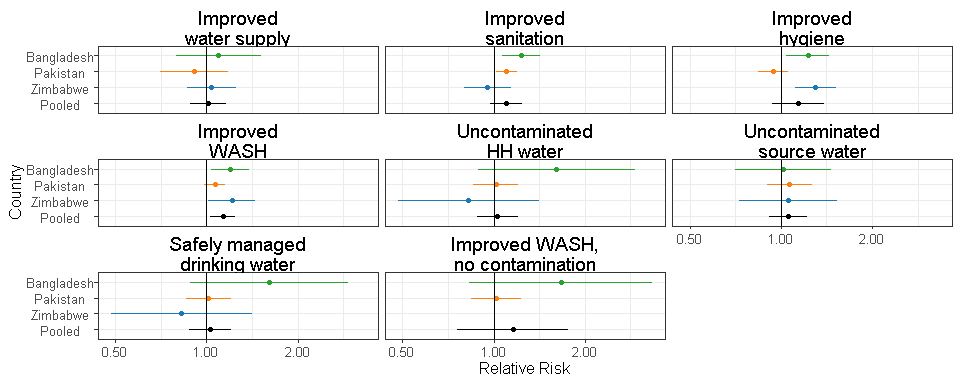
## Primary figures



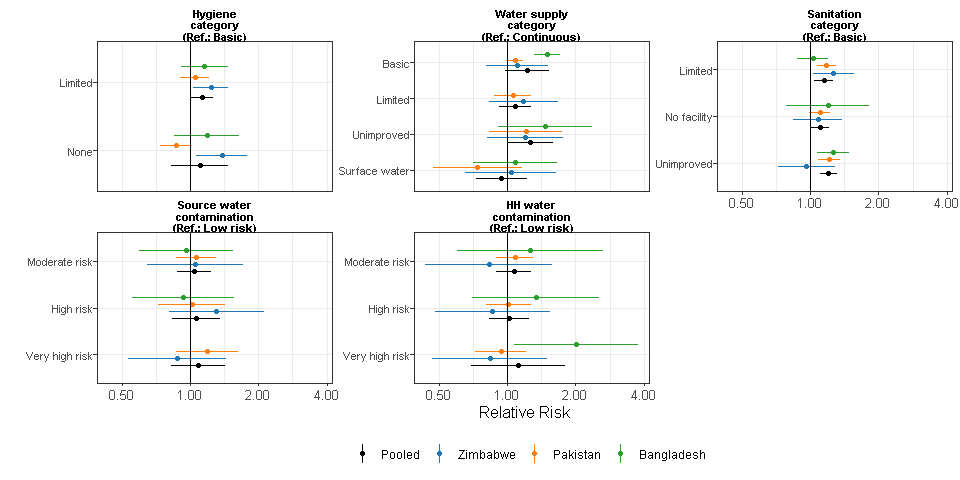
**Figure 1.** Random-effects pooled relative risks, pooled over the Bangladesh, Zimbabwe, and Pakistan MICS-6 surveys, for the 4 outcomes. Panels are divided by outcome, with ARI standing for acute respiratory infection. The Y axis lists the reference (improved) level of contrast, with the relative risk estimates between the listed reference level and the higher-risk (unimproved) level. For example, households with any level of E. coli contamination were compares to houses with uncontaminated HH water. Among other significant estimates, unimproved hygiene (no water and soap at handwashing station), unimproved sanitation, and unimproved combined WASH was associated with a higher risk of stunting, and unimproved combined WASH was associated with a higher risk of diarrhea. Larger confidence intervals may represent associations estimated less precisely (e.g. for a rarer outcome like ARI) or represent heterogeneity between the country-specific estimates.



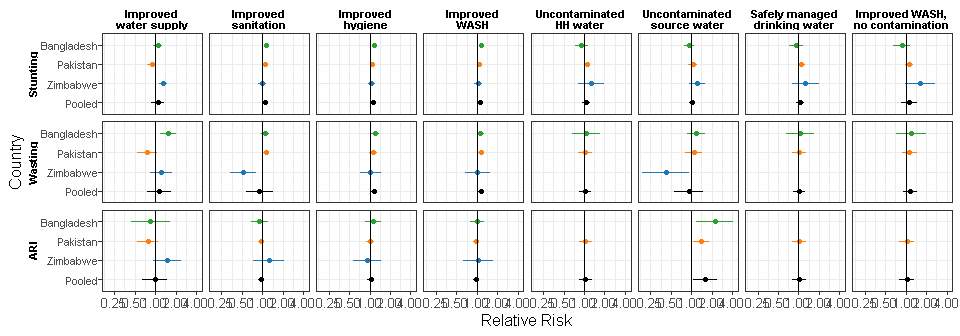
**Figure 2.** Random-effects pooled relative risks, pooled over the Bangladesh, Zimbabwe, and Pakistan MICS-6 surveys, across the 4 outcomes for WASH conditions stratified into levels of the Joint Medical Program WASH ladder or by level of E.coli concentration. The row labels list the WASH characteristic and the reference (most improved) level of contrast, and the Y-axis labels are the contrasting higher level of risk.



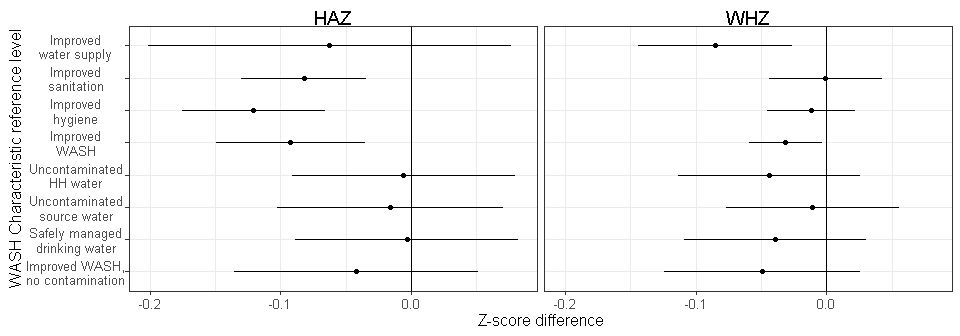
**Figure 3.** Country-specific and random-effects pooled relative risks for the primary outcome of diarrheal disease. Panels are divided by WASH condition, with the relative risk estimates between the reference level listed at the top of the panel and the higher-risk (unimproved) level. Missing estimates denote an outcome with too much sparsity in one strata of the contrast. For example, there are less than 10 cases of diarrhea in houses with improved WASH without E.coli contamination, so a relative risk was not estimated.



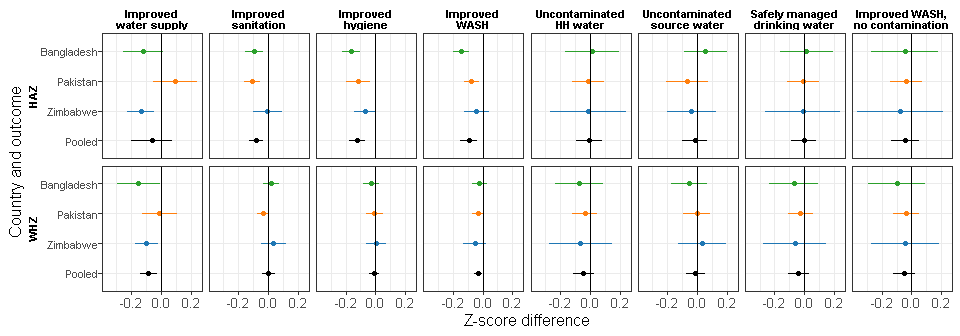
**Figure 4.** Country-specific and random-effects pooled relative risks, pooled over the Bangladesh, Zimbabwe, and Pakistan MICS-6 surveys, for the primary outcome of diarrhea across WASH conditions stratified into levels of the Joint Medical Program WASH ladder or by level of E.coli concentration. The row labels list the WASH characteristic and the reference (most improved) level of contrast, and the Y-axis labels are the contrasting higher level of risk, and the color of points denotes the country of the survey.



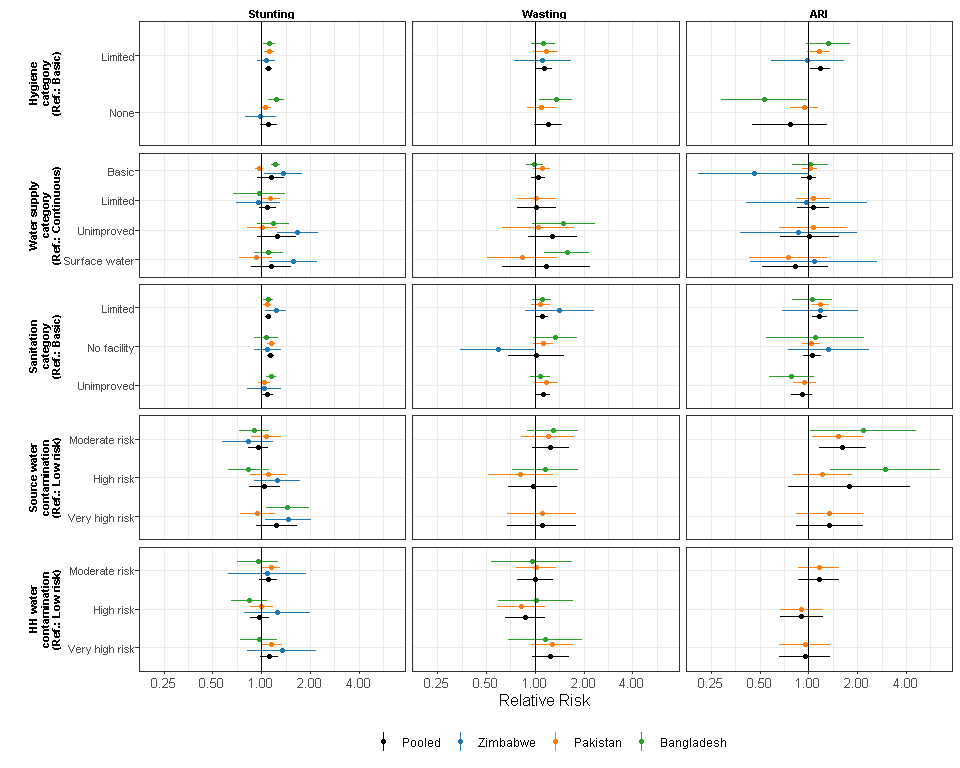
**Figure 5.** Country-specific and random-effects pooled relative risks for the secondary outcomes. Panels are divided by WASH condition, with the relative risk estimates between the listed reference level and the higher-risk (unimproved) level.



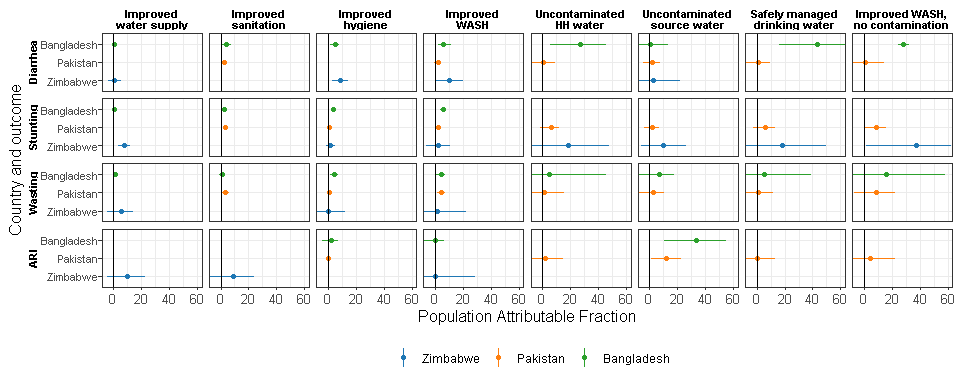
**Figure 6.** Random-effects pooled Z-score differences, pooled over the Bangladesh, Zimbabwe, and Pakistan MICS-6 surveys, for the height-for-age Z-score (HAZ) and weight-for-height Z-scores (WHZ) outcomes. The Y axis lists the reference (improved) level of contrast, with the relative risk estimates between the listed reference level and the higher-risk (unimproved) level.



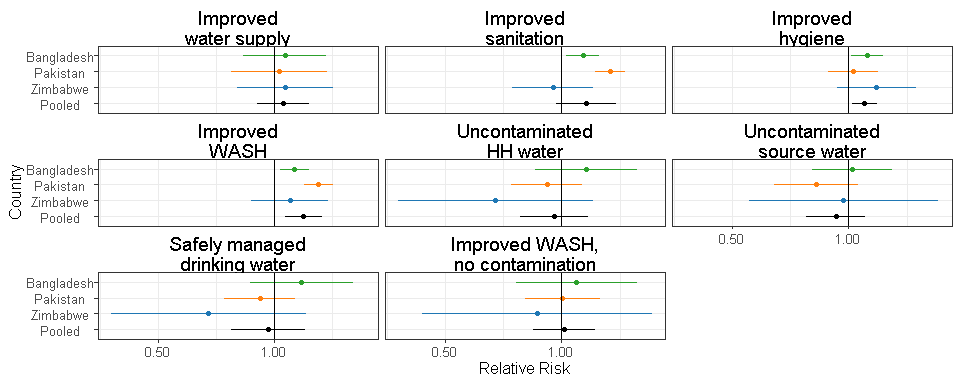
**Figure 7.** Country-specific and random-effects pooled Z-score differences, pooled over the Bangladesh, Zimbabwe, and Pakistan MICS-6 surveys, for the HAZ and WHZ outcomes. The Y axis lists the reference (improved) level of contrast, with the relative risk estimates between the listed reference level and the higher-risk (unimproved) level.



**Figure 8.** Country-specific and random-effects pooled relative risks, pooled over the Bangladesh, Zimbabwe, and Pakistan MICS-6 surveys, for the secondary binary outcomes. The row labels list the WASH characteristic and the reference (most improved) level of contrast, and the Y-axis labels are the contrasting higher level of risk, and the color of points denotes the country of the survey.



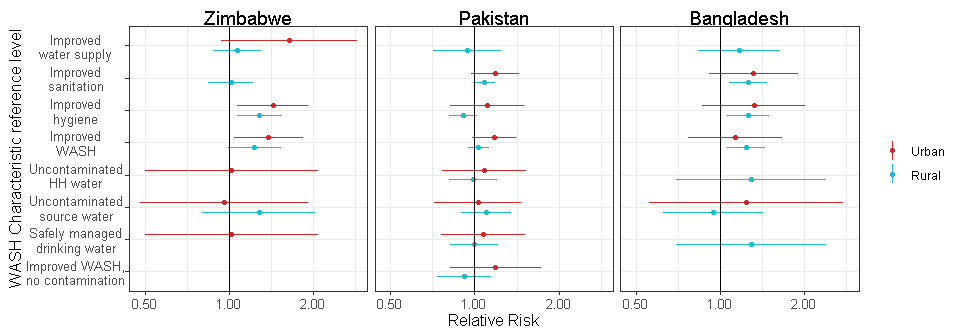
**Figure 9.** Country-specific population attributable fractions (PAFs) estimated from adjusted relative risks for the binary outcomes and WASH characterists. PAF’s were only estimated for estimated relative risks greater than one. The confidence intervals are bootstrapped, with 100 bootstrap iteractions for the proof-of-concept analysis (the full analysis will use 1000 iterations).



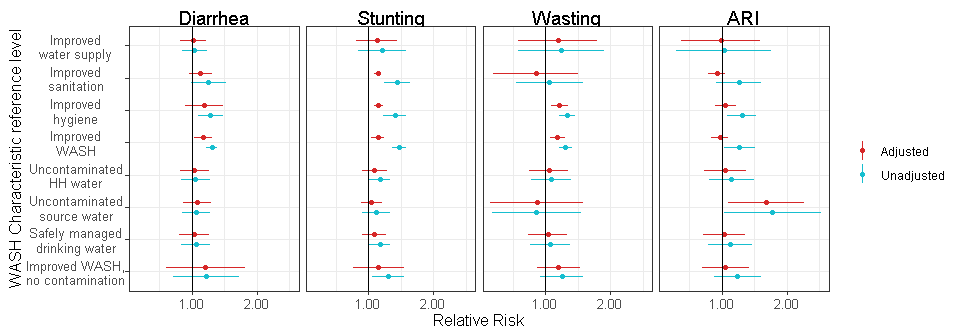
**Figure 10.** Country-specific and random-effects pooled relative risks for all-cause child mortality. Panels are divided by WASH condition, with the relative risk estimates between the reference level listed at the top of the panel and the higher-risk (unimproved) level. Missing estimates denote an outcome with too much sparsity in one strata of the contrast.

## Sensitivity analysis figures

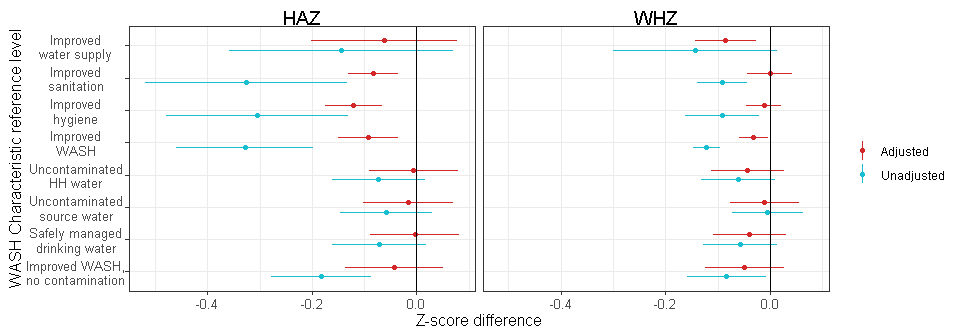
Note: the sensitivity analyses are based on pooled estimated from binary codings of WASH characteristics for easier comparison between the primary and sensitivity results across the WASH conditions.



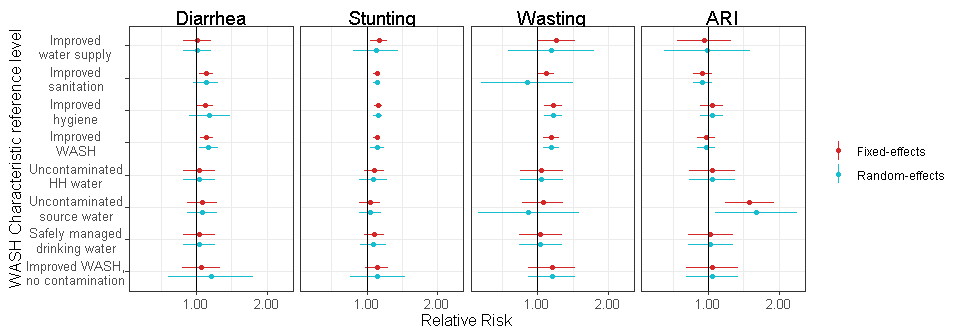
**Figure S1.** Comparison of adjusted relative risks between urban and rural households across country-specific estimates of the relative risk of diarrhea. Eighty-one percent of Bangladeshi households are rural, 72% of Pakistani households are rural, and 65% of Zimbabwean households are rural. Missing estimates are from WASH-condition and outcome combinations with rarity in one strata. Relative risks are generally higher in urban populations, though the difference is small or imprecise.

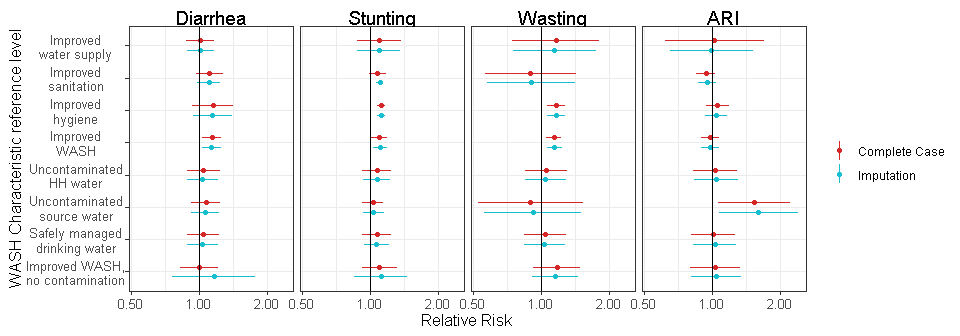


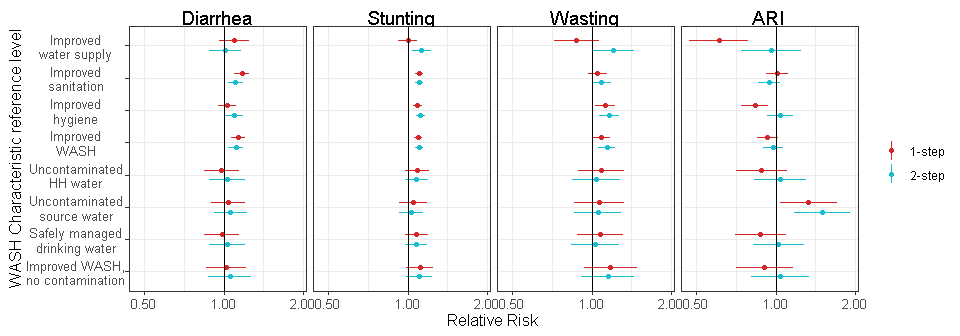
**Figure S2.** Comparison of unadjusted and adjusted relative risks across pooled estimates and binary measures of improved WASH conditions. Unadjusted estimates are generally further from the null, indicating that covariate adjustment is correcting for some confounding between WASH conditions and child health outcomes (likely from socio-economic status as well as other factors).



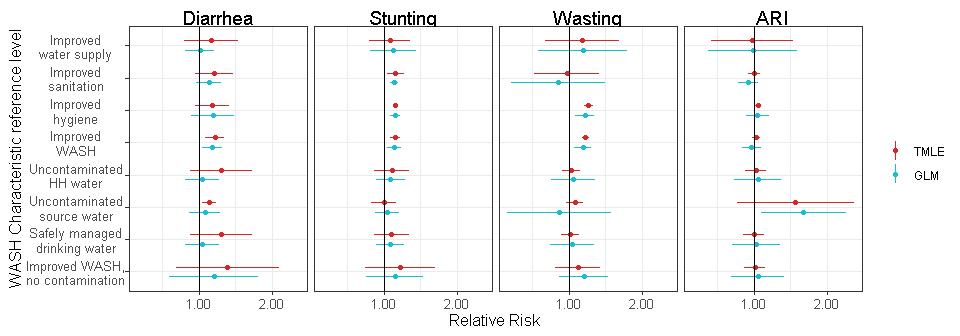
**Figure S3.** Comparison of unadjusted and adjusted Z-score differences across pooled estimates and binary measures of improved WASH conditions. Unadjusted estimates are generally further from the null, indicating that covariate adjustment is correcting for some confounding between WASH conditions and child growth.

 **Figure S4.** Comparison of random- and fixed-effects pooled relative risks across outcomes and binary measures of improved WASH conditions. Point estimates are very similiar, though random-effects estimates have larger confidence intervals when study-specific estimates are heterogeneous.

 **Figure S5.** Comparison of pooled relative risks across outcomes and binary measures of improved WASH conditions from the primary analysis, where missing covariate information is imputed, and from a complete-case sensitivity analysis. Estimates are very similar, though the association between improved sanitation and stunting is larger and more precise after imputation. For categorical variables, a “missing” category was added, and for continuous variables missingness was imputed using the median. Multiple-imputation will be used in the full analysis.

 **Figure S7.** Comparison of 2-step (primary) and 1-step (sensitivity) pooled relative risks across outcomes and binary measures of improved WASH conditions. In 2-step meta-analyses, country-specific estimates are estimated first, and then pooled using traditional meta-analysis methods. In 1-step meta-analyses, the estimate of interest is estimated from a single model estimated using a dataset of all country-specific survey results. Here, the 2-step estimates are pooled using fixed effects to be more comparable to the 1-step estimate, as the random-effects for survey have not been implemented yet for the 1-step analysis. One and 2-step estimates are generally similiar here, with no discernible pattern in differences.

# {r, echo=F, warning=F, message=F, fig.width = 10, fig.height = 3.5} # # p\_tmle\_glm\_comp\_RR # # \_\_Figure S8.\_\_ Comparison of pooled relative risks across outcomes and binary measures of improved WASH conditions, estimated either with GLM and targeted maximum likelihood estimation (TMLE) models (fit using glm models). While the targeted estimation of the parameter of interest (here, the relative risk) can # #



**Figure S8.** Comparison of pooled relative risks across outcomes and binary measures of improved WASH conditions, estimated either with GLM and targeted maximum likelihood estimation (TMLE) models, which use ensemble machine learning models in the estimation of relative risks. The TMLE models were estimated using library with simple means, generalized linear models, generalized additive models, penalized regressions, and gradient boosting machines. TMLE estimates are more precise and with different point estimates in certain cases, but were very slow to run, taking 7 days across all exposure-outcome combinations.

## MICS surveys to use in full analysis

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Region | Country | Year | Round | Status | Datasets |
| East Asia and the Pacific | Kiribati | 2018-2019 | MICS6 | Completed | Available |
| East Asia and the Pacific | Korea, Democratic People’s Republic of | 2017 | MICS6 | Completed | Restricted |
| East Asia and the Pacific | Lao People’s Democratic Republic | 2017 | MICS6 | Completed | Available |
| East Asia and the Pacific | Mongolia | 2018 | MICS6 | Completed | Available |
| East Asia and the Pacific | Mongolia (Khuvsgul Aimag) | 2016 | MICS5 | Completed | Available |
| East Asia and the Pacific | Mongolia (Nalaikh District) | 2016 | MICS5 | Completed | Available |
| East Asia and the Pacific | Samoa | 2019-2020 | MICS6 | Data processing / analysis |  |
| East Asia and the Pacific | Tonga | 2019 | MICS6 | Completed | Available |
| Eastern and Southern Africa | Lesotho | 2018 | MICS6 | Completed | Available |
| Eastern and Southern Africa | Madagascar | 2018 | MICS6 | Completed | Available |
| Eastern and Southern Africa | Zimbabwe | 2019 | MICS6 | Completed | Available |
| Europe and Central Asia | Georgia | 2018 | MICS6 | Completed | Available |
| Latin America and Caribbean | Paraguay | 2016 | MICS5 | Completed | Available |
| Latin America and Caribbean | Suriname | 2018 | MICS6 | Completed | Available |
| Middle East and North Africa | Iraq | 2018 | MICS6 | Completed | Available |
| Middle East and North Africa | Tunisia | 2018 | MICS6 | Completed | Available |
| South Asia | Bangladesh | 2012-2013 | MICS5 | Completed | Available |
| South Asia | Bangladesh | 2019 | MICS6 | Completed | Available |
| South Asia | Nepal | 2014 | MICS5 | Completed | Available |
| South Asia | Pakistan (Punjab) | 2017-2018 | MICS6 | Completed | Available |
| South Asia | Pakistan (Sindh) | 2014 | MICS5 | Completed | Available |
| West and Central Africa | CÃƒÂ´te d’Ivoire | 2016 | MICS5 | Completed | Available |
| West and Central Africa | Congo | 2014-2015 | MICS5 | Completed | Available |
| West and Central Africa | Congo, Democratic Republic of the | 2017-2018 | MICS6 | Completed | Available |
| West and Central Africa | Gambia | 2018 | MICS6 | Completed | Available |
| West and Central Africa | Ghana | 2017-2018 | MICS6 | Completed | Available |
| West and Central Africa | Guinea-Bissau | 2018-2019 | MICS6 | Completed | Available |
| West and Central Africa | Nigeria | 2016-2017 | MICS5 | Completed | Available |
| West and Central Africa | Sierra Leone | 2017 | MICS6 | Completed | Available |
| West and Central Africa | Togo | 2017 | MICS6 | Completed | Available |

Questions \* include severe growth outcomes? \* change set of adjustment covariates \*change TMLE to be just diarrhea outcome and E-Coli contamination

Things to do \* Make plot with adjusted TMLE \* Compare Bangladesh estimates to WASH Benefits (focus on rural estimates). Zimbabwe to shine.

Things to note \* Modified poisson, clustered SE’s on the household level with with glm() function \* Convergence error when trying to use GEE for the modified poison \* Sandwich SE \* No weights implemented in the TMLE fits \* no mortality outcome \* Median, though Multiple-imputation will be used in the full analysis. Note CC analysis not that much different. \* TMLE seems overly precise… increased statistical efficiency is one of the advantages, but checking into it

* Make note on average number of covariates chosen > summary(str\_count(dfW$W, pattern = “,”) + 1) Min. 1st Qu. Median Mean 3rd Qu. Max. 1.000 3.000 11.000 9.135 14.000 16.000
* Adjusted TMLE very slow - took 4 days to run for 3 countries, so choose just primary diarrhea outcome for full analysis
* Note cutoff for MICS inclusion for full analysis
* PAF
* Check that the right contrasts were included.

Things to note \* proposed changes: \* TMLE sensitivity analysis just for primary diarrhea outcome and measures of HH water contamination \* Same with time-intensive stochastic intervention \* Clustered SE on unique HH instead of inclusing a multi-level random effect for cluster and household. \* Add sensitivity analysis for clustering on the cluster ID

* Add details on how covariates with many levels are coded -coded from levels of given in the MICS documentation