Power Considerations for Village RCT Odisha - Phase 1 - Diarrhea outcome

Alex Lehner

July2023

This document lays out power calculations for a binary outcome, measured at the individual level, for a cluster randomized trial (C-RCT) in Odisha, India. The outcome of interest is diarrhea (approximate incidence rate of 5% [latest data from NFHS]) and the treatment (inline chlorination devices for clean water) is assigned for a complete cluster (a village).

The crucial assumptions for these calculations are:

- Intra-cluster correlation (ICC) [`0.01` and `0.02`]
- (Average) size of villages (U5 children) [`30` and `50`]
- Effect size [`10%` and `20%`]

One of the main goals of the document is to illustrate the **trade-off between including more villages and increasing the number of survey rounds** to achieve power. Autocorrelation of diarrhea incidence within individuals over time is considered here as well (in the form of an AR(1) process, roughly matching the values from Pickering et. al. (2019)). For computational reasons the current benchmark version of multi-round simulations does not consider this, however, because the autocorrelation of the binary outcome is not too strong (see analysis of Pickering data below) and thus not affecting power significantly.

Gains from extra rounds of data collection. The bottom line is that there are substantial power gains from extra rounds. The intuition for this is that with every new round we get additional diarrhea cases. This is different from a scenario with a continuous outcome variable where power gains from extra periods stem mostly from averaging out noise in Y (see the 2012 paper by McKenzie).

The OLS specification that is assumed is the one from Pickering et. al. (2019). It regresses an indicator variable of diarrhea incidence on a treatment indicator at the village level and one fixed effect for every round - exploiting only within-round variation and thus controlling for unobserved variables that change over time:

$$Y_{ic} = \beta_1 \; T_c + \sum_{r=1}^{3} \gamma_r \; ROUND_{ric} + \varepsilon_{ic},$$

with standard errors clustered at the village level (or the village by individual level to account for autocorrelation if the number of rounds, r, is greater than 1).

A quick comment on cluster sizes. Please note that none of the following calculations assume variation in cluster size. In reality these are going to vary and thus harm statistical power. For details and some descriptives (i.e. the CV of sizes per district), please have a look at the dashboard that contains information on all census villages of Odisha. In practice, for the final randomization it could make sense to exclude very small and very big villages from the sample ex-ante in order to increase power (I am thinking of somewhere around the 5th and 95th percentile - depending on the state).

To inform our calculations, we analyze data from a clustered chlorination intervention in Bangladesh from Pickering et al (2019, The Lancet).

The document is structured as follows:

1) Simple benchmark calculations for only one survey round (**plug-in formulas**)
2) Consideration of multiple rounds of data collection (**simulations**, DGP targeted to match data in Pickering et. al. (2019))
3) Relevant descriptive statistics from Pickering et. al. (2019)

1) Benchmark calculations - one round only (plug-in formulas)

The following two plots show power as a function of the number of villages for two different average villages sizes: 30 and 50. The plots are grouped by ICC (`0.01` and `0.02`) and MDE (`10%` and `20%`). As noted in the introduction, the assumed diarrhea incidence is 5% across all specifications.

The plug-in formula that is being used is implemented in the R function `clusterPower::cpa.binary()`. The stata equivalent is `clustersampsi, binomial`. An example on how to use the code:

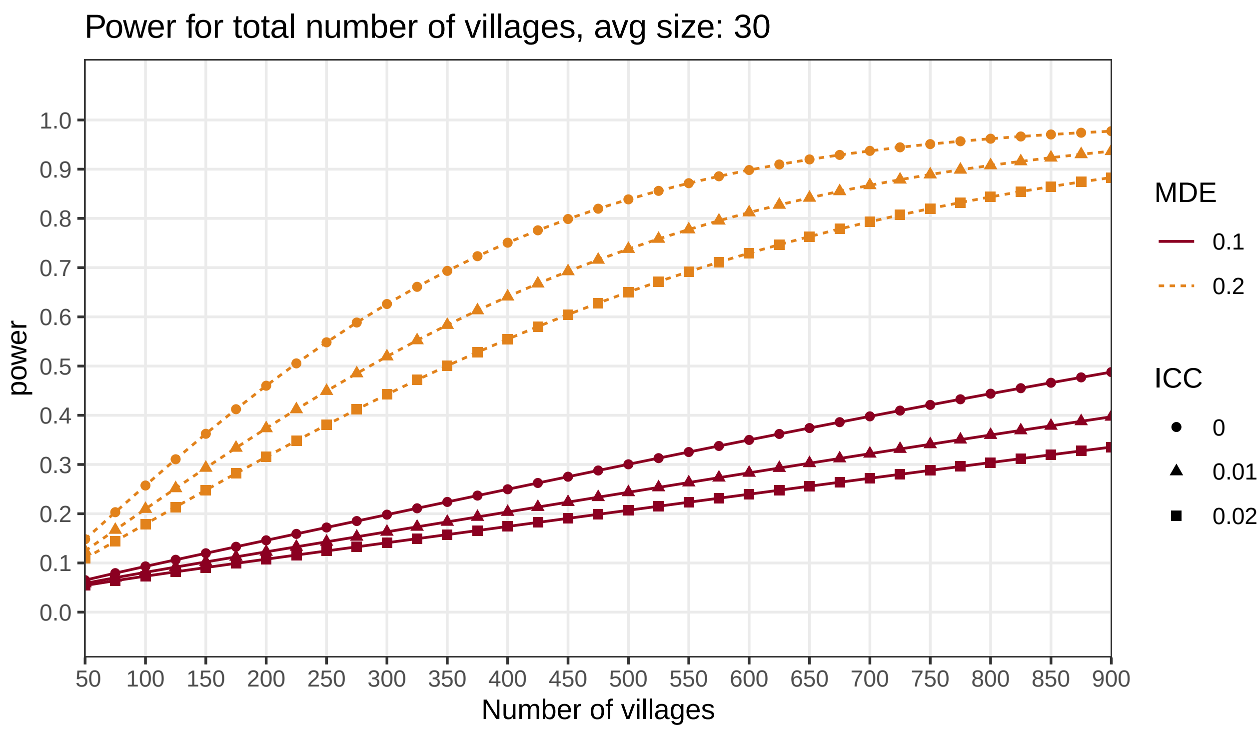
```
cpa.binary(nsubjects = 50, # average clustersize
           CV = 0, # no variation in clustersize assumed
           power = .8, # desired power
           ICC = 0.01,
           p1 = 0.05, # diarrhea incidence of 5%
           p2 = 0.05 * (1 - 0.20)) # MDE of 20%
# Function returns the number of clusters needed PER condition, i.e. multiply by two.
```

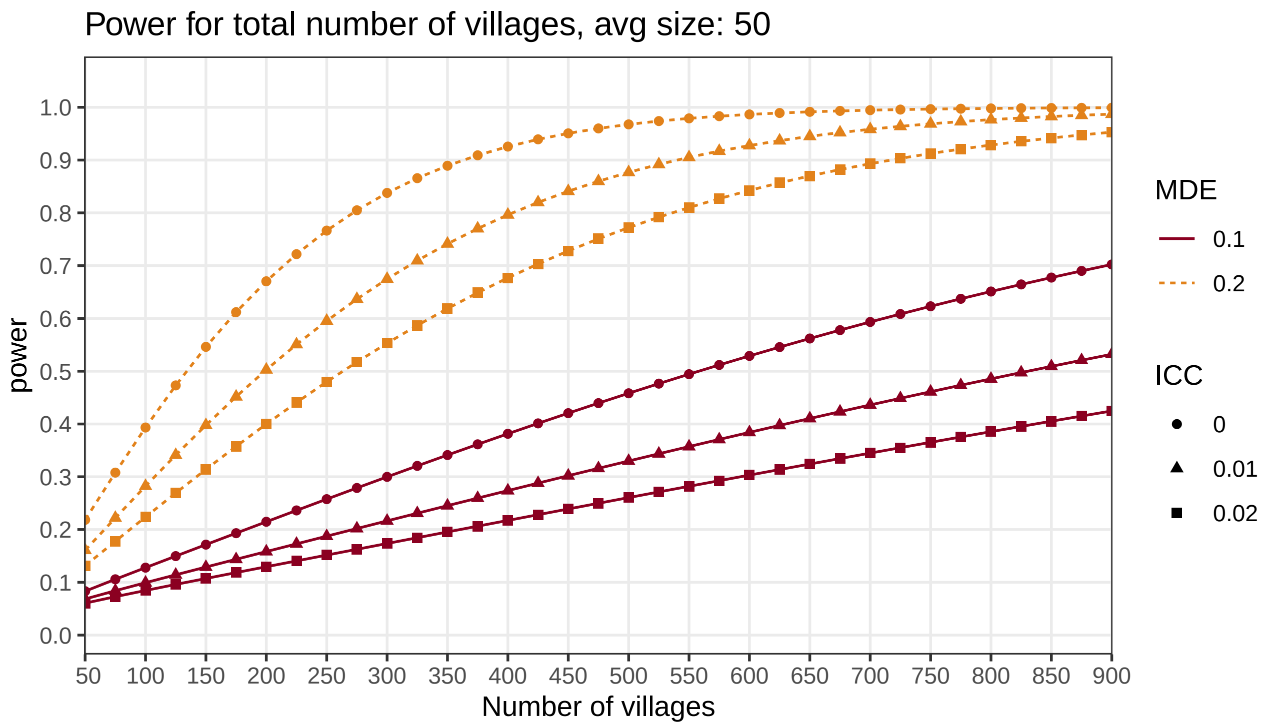
The plots below **trace out the full power curve** for many potential combinations of inputs. The bottom line of these calculations is that for an MDE of 10% we are not going to be sufficiently powered. For the MDE of 20%, the following table summarizes the different scenarios where power is approximately 80% (extracted from the data that can be seen in the plots).

Table 1: Number of villages needed for different scenarios (total, both arms - numbers extracted from the power-curves in the figures below).

avg_clustersize	diarrhea incidence	MDE	ICC	villages_needed	power
30	0.05	0.2	0	600	0.9
30	0.05	0.2	0.01	775	0.9

avg_clustersize	diarrhea incidence	MDE	ICC	villages_needed	power
30	0.05	0.2	0.02	900	0.9
50	0.05	0.2	0	375	0.9
50	0.05	0.2	0.01	550	0.9
50	0.05	0.2	0.02	725	0.9




Power for total number of villages, avg size: 50
power
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
50
100
150
200
250
300
350
400
450
500
550
600
650
700
750
800
850
900
Number of villages
MDE
0.1
0.2
ICC
0
0.01
0.02

2) Simulation: multiple rounds of data collection

The simulation code that is used here models the data generating process (DGP) as draws from a Bernoulli distribution with a cluster level random effect that induces ICC. The events (diarrhea incidence) within individuals across rounds are modeled with an **i.i.d. error to first give a benchmark scenario**. In the subsequent section we assume a correlated error term that induces an AR(1) process to approximately match the data in Pickering et. al. (2019). The simulations are carried out 1,000 times for each dot that is visualized.

To showcase that the code delivers meaningful power estimates, the plots at the end of this section showcase the scenario where the number of rounds in the simulations is 1. It can be seen that the resulting power curves are very similar to the results from above, verifying that they can approximate the results from the plug-in formula well. Note that the curve for MDEs of 10% is a bit wobbly due to the small effect size, a larger amount of iterations would straighten them out further.

2.1) Simulations with 3 rounds, iid errors

As indicated in the introduction, there are substantial power gains from additional rounds - mostly because we add additional diarrhea incidences to the data. The following table summarizes the scenarios where power is approximately 80%, extracted from the values that you can see in the plot - as in the plug-in formula case. **To get a better picture for the power numbers it is ideal to check the interpolated values in the plots**. The plots trace out the full power curve (note that there are less dots as compared to the above because for every value we need a substantial amount of simulations, which is computationally very costly given the many different scenarios).

Table 2: Effectsize 20%, 3 rounds. Number of villages needed for different scenarios (total, both arms - numbers extracted from the powercurves in the figures below).

avg_clustersize	diarrhea incidence	MDE	ICC	villages_needed	power
30	0.05	0.2	0.01	300	0.90
30	0.05	0.2	0.02	400	0.92
50	0.05	0.2	0.01	200	0.88
50	0.05	0.2	0.02	325	0.91

Table 3: Effectsize 10%, 3 rounds. Number of villages needed for different scenarios (total, both arms - numbers extracted from the powercurves in the figures below).

avg_clustersize	diarrhea incidence	MDE	ICC	villages_needed	power
30	0.05	0.1	0.01	1000	0.78
30	0.05	0.1	0.02	1000	0.58
50	0.05	0.1	0.01	900	0.89
50	0.05	0.1	0.02	1000	0.71

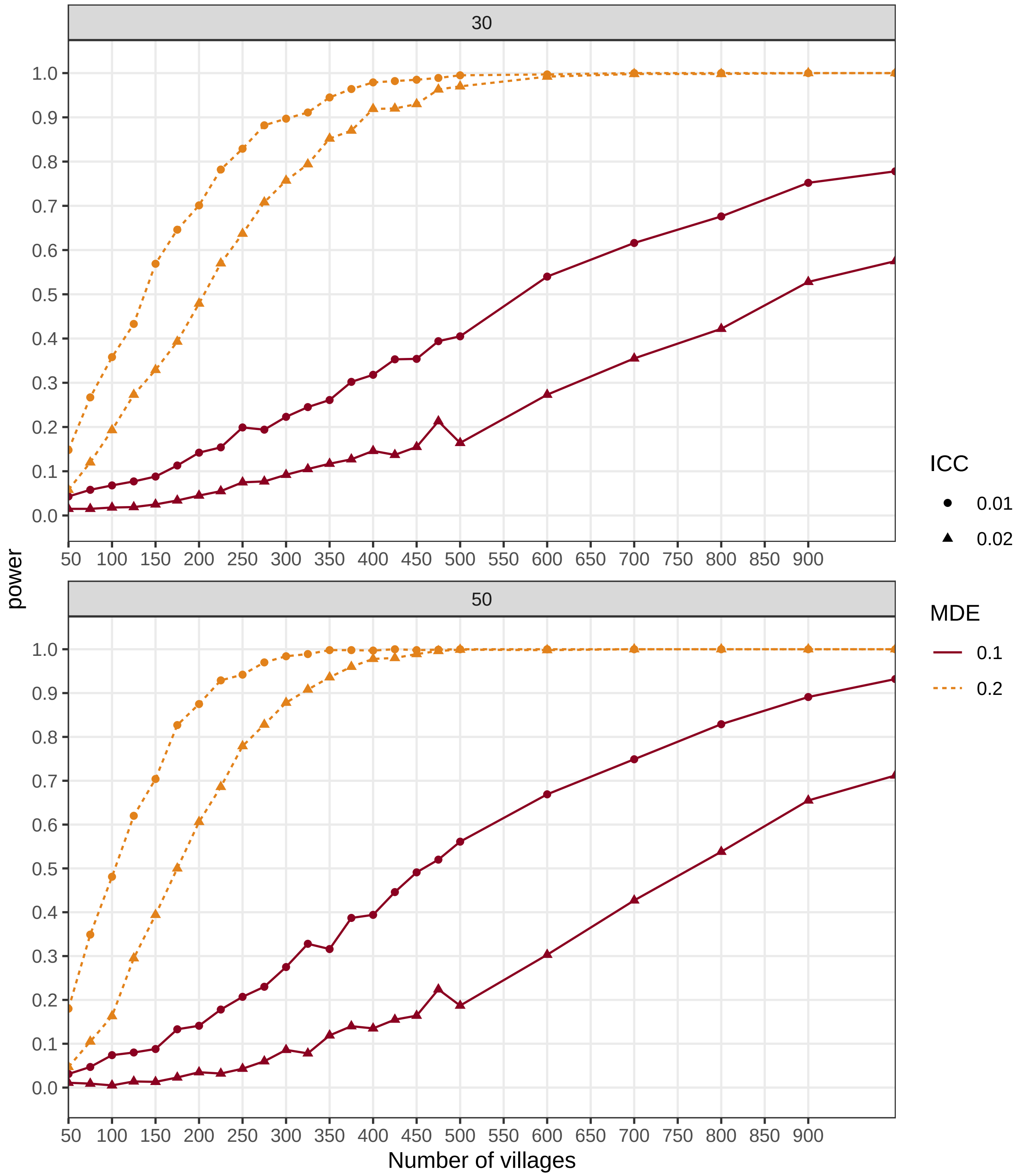


Figure 1: Power with three rounds of data collection. Upper panel: 30 U5 children per village. Lower panel: 50 U5 children per village.

2.2) Simulations with 3 rounds, AR(1) errors

to be added. For the case of 3 rounds, results won't change substantially, however.

2.3) Showcase that simulations with 1 round match formulas

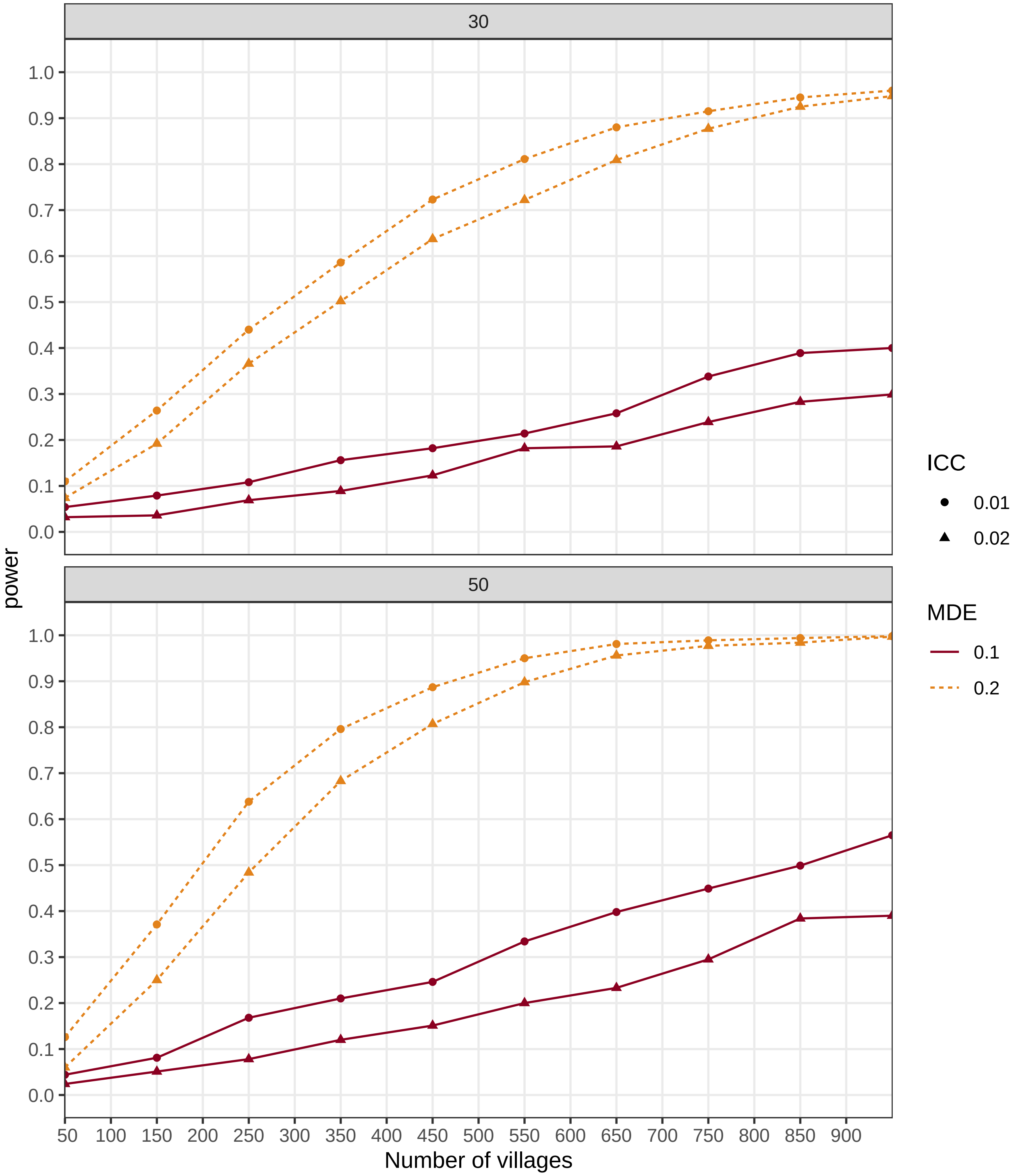


Figure 2: Showcase the working of the simulation code by setting the number of rounds to 1 and matching the results from the plug-in formulas above.

3) Descriptives from Pickering et. al. (2019, The Lancet)

Pickering et. al. (2019): 100 shared water points (clusters) in two low-income urban communities in Bangladesh were randomly assigned (1:1) to have their drinking water automatically chlorinated at the point of collection by a solid tablet chlorine doser (intervention group) or to be treated by a visually identical doser that supplied vitamin C (active control group). The trial followed an open cohort design; all children younger than 5 years residing in households accessing enrolled water points were measured every 2–3 months during a 14-month follow-up period (children could migrate into or out of the cluster). The primary outcome was caregiver-reported child diarrhoea (more than 2 loose or watery stools in a 24-h period [WHO criteria]) with a 1-week recall, including all available childhood observations in the analyses. Children in the treatment group had less WHO-defined diarrhoea than did children in the control group (control 216 [10.0%] of 2154; treatment 156 [7.5%] of 2073; prevalence ratio 0.77, 95% CI 0.65–0.91).

They collected 7 rounds in total and found an effect of roughly 25% (diarrhea incidence reduction). The study was block-randomized by matching pairs, i.e. there are 50 randomization blocks with one cluster each assigned to treatment and control.

Table 4: Diarrhea incidence for every round of collection, broken down by treatment and control group.

round	vitamins (control)	chlorine
1	0.069	0.062
2	0.099	0.073
3	0.111	0.077
4	0.086	0.052
5	0.052	0.062
6	0.048	0.044
7	0.053	0.019

Intra-cluster correlation (ICC) We calculate the raw ICC to be 0.0224 and 0.0167 after taking into account the blocking structure of the experiment. The sample size is not very large and we thus have reasons to believe that these estimates are not very precise. From earlier data work on child mortality with substantially larger sample sizes we obtained estimates of around 0.01. We therefore want to use 0.01 and 0.02 as benchmark values in our power calculations.

Autocorrelation Using the residual of the main specification of interest - a regression of diarrhea on a treatment indicator with block and round fixed effects - from Pickering et. al. (2019), we can estimate an autocorrelation parameter. Specifically, we are regressing the error from that regression on its lagged values, forcing the intercept to be 0:

$$e_{it} = \rho \; e_{i,t-1} + \epsilon_{it}.$$

Alternatively, we can also check the raw correlation between the diarrhea incidence in every round with its lagged value (omitting round 1 by force because we to not observe $t = 0$.):

Table 5: Raw correlation of diarrhea incidence and its lagged value, broken down per round

round	corr
1	NA
2	0.149
3	0.149
4	0.034
5	0.077
6	0.151
7	0.016

The following table shows the counts of the total number of diarrhea cases over the seven rounds per child. The maximum number of cases per child was 5, while most children that had diarrhea only had it in one out of seven rounds.

```
##
##    0    1    2    3    4    5
## 4601 1244  267   74   12    7
```

The percentage of children in the control group that ever had diarrhea is 23.1 and in the control group 19.6 percent of children had diarrhea recorded at least once.

Checking counts per cluster

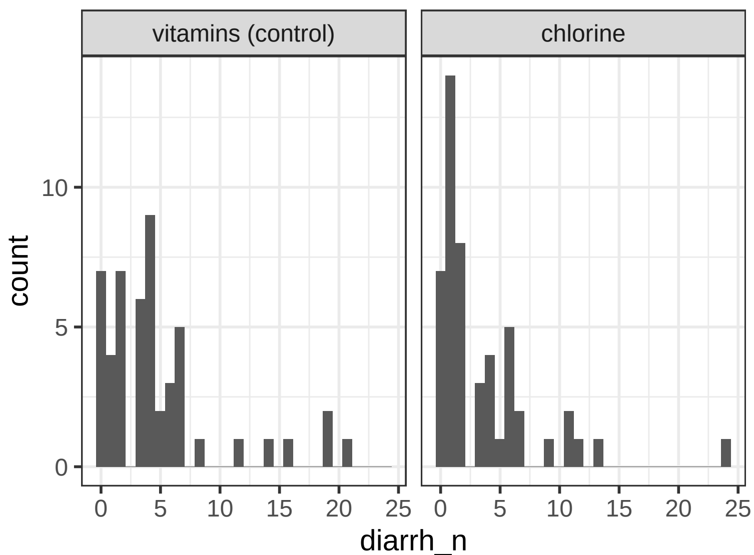


Figure 3: Diarrhea counts per cluster (pump), summed over all 7 rounds - approximately Poisson.