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%) and the treatment (inline chlorination devices for clean water) is assigned at the household level for a complete cluster (a village).

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

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 sice [15% and 20%]

Autocorrelation of diarrhea incidence within individuals over time is considered here as well (in the form of an AR(1) process), but seems to be less of a concern for power.

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.

The document is structured as follows:

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

1) Descriptives - diarrhea counts

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

round	vitamins (control)	chlorine
1	0.0686275	0.0623853
2	0.0990826	0.0731225
3	0.1111111	0.0770791
4	0.0862745	0.0518672
5	0.0517598	0.0617021
6	0.0478360	0.0436893
7	0.0526316	0.0193548

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.
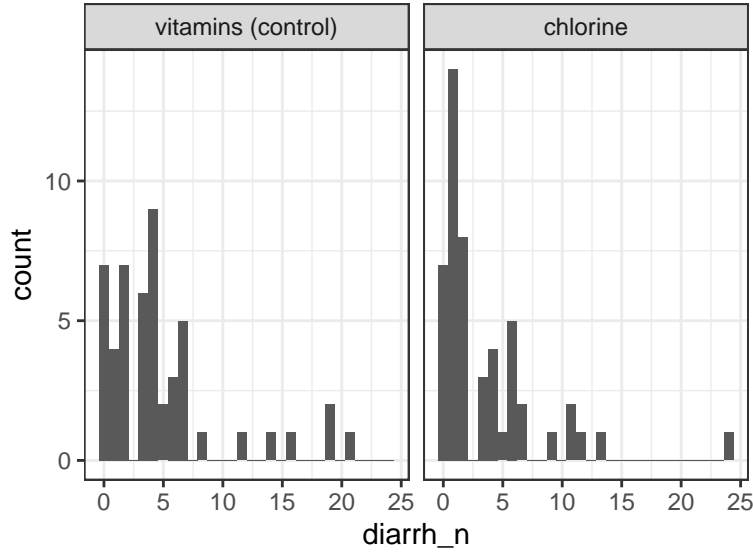
Checking counts per cluster

Figure 1: Diarrhea counts per cluster (pump), summed over all 7 rounds - approximately Poisson.

2) Benchmark calculations (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.

On the backend, the code uses the function `clusterPower::cpa.binary()` many times to compute each point in the plots.

Power for total number of villages, avg size: 50

Power for total number of villages, avg size: 30
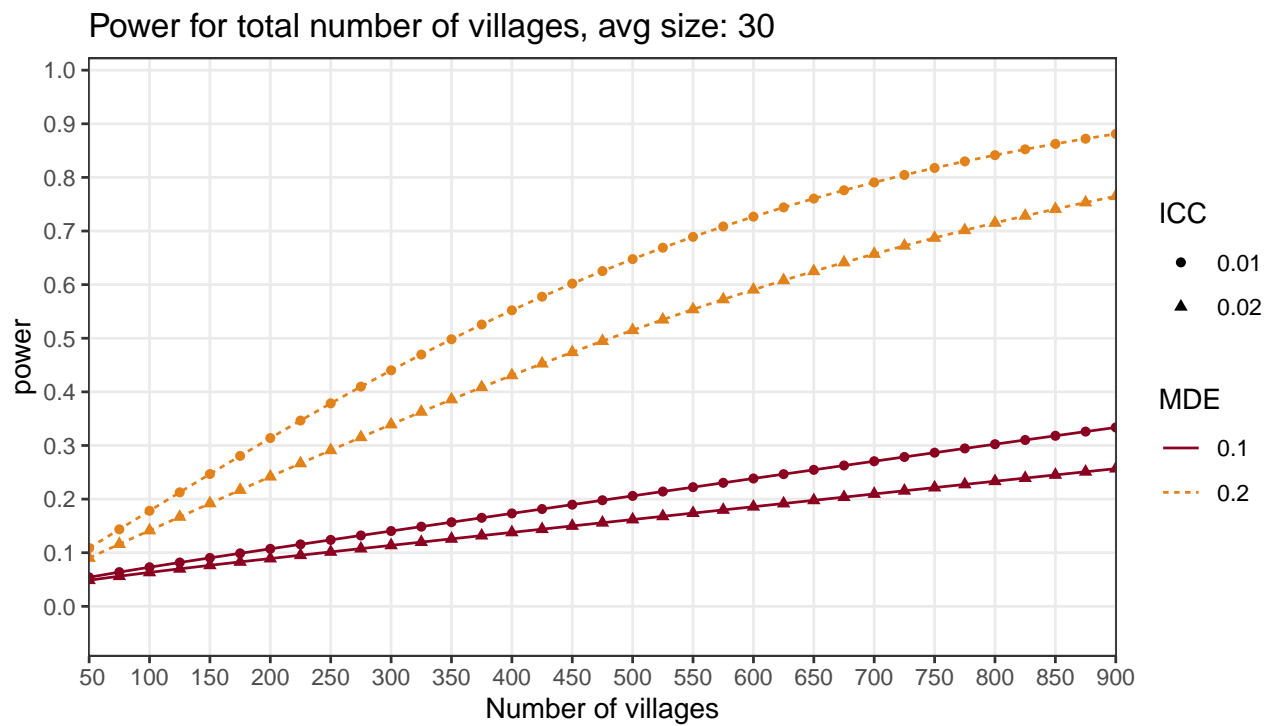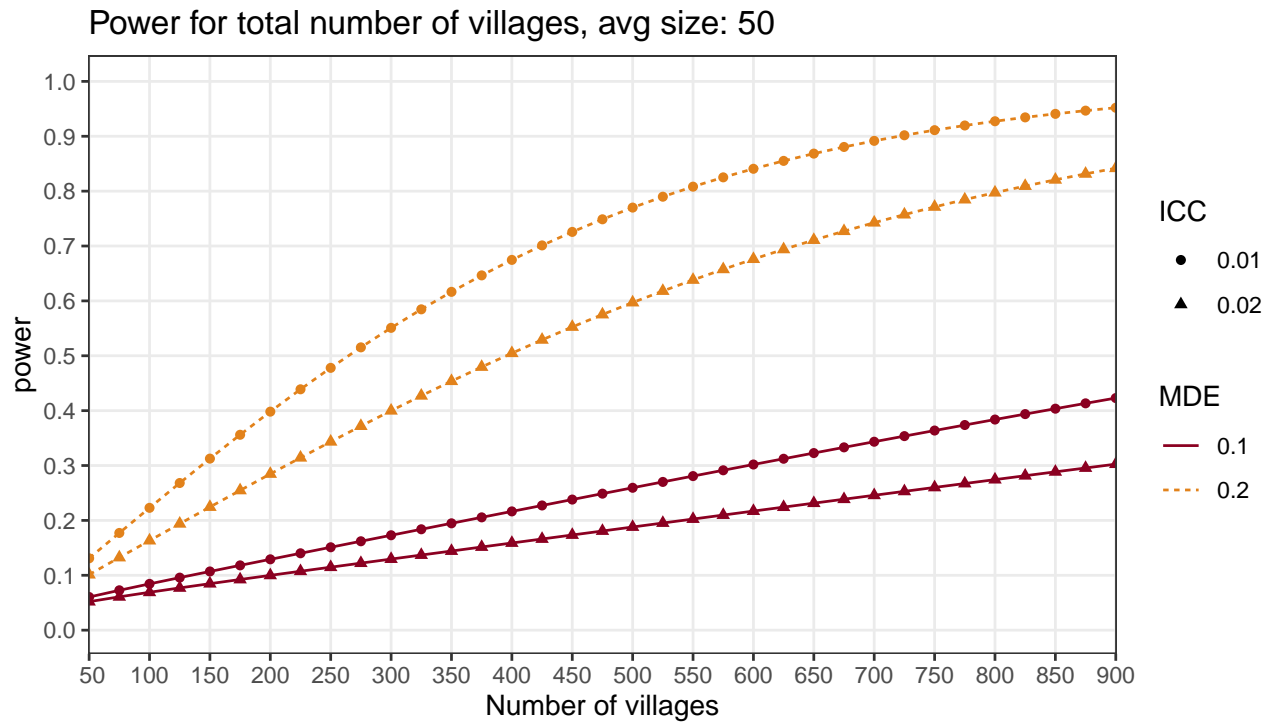
3) Simulation: multiple rounds of data collection