

# Lit Review

Susanna Makela

January 24, 2017

## Kumar and Vollmer (2013) “Does access to improved sanitation reduce childhood diarrhea in rural India?”

- **Goal:** to quantify the effect of improved sanitation on diarrhea incidence in young children in rural India
- **Data:** DLHS-3 (2007-2008), restricted to rural areas
- **Treatment variable:** whether household has access to improved sanitation (flush toilet, ventilated pit, composite toilet)
- **Outcome variable:** 0/1 dummy for whether child had diarrhea in past 2 weeks
- **Methodology:** nearest-neighbor matching with replacement based on propensity scores (tolerance of 0.01 on propensity score scale); some robustness checks using propensity-weighted regression (but with a LPM; not clear why they didn’t do propensity-weighted logistic regression, if that’s a thing)
- **Results:**
  - ATT: improved sanitation averts 0.8 diarrhea episodes per household-year (large in the context of an average of 3.1 diarrhea cases per child-year, or 3.9 diarrhea cases per household-year)
  - Heterogeneous treatment effects, particularly by whether HH treats water, SES, and gender – but, none of these were specified a priori and none of the differences across categories are statistically significant, except for water treatment.

## Jalal and Ravallion (2003) “Does piped water reduce diarrhea for children in rural India?”

- **Goal:** is a child less vulnerable to diarrheal disease if s/he lives in a household with access to piped water? Do children in poor, or poorly educated, households realize the same health gains? Does income matter independently of parental education? *They actually specified their interactions a priori!*
- **Data:** household survey done by the National Council of Applied Economic Research in 1993-1994; nationally representative survey of rural households (only covers 16 states)
- **Treatment variable:** whether household has access to piped water (a tap either inside or outside the house)
- **Outcome variable:** 0/1 dummy for whether child had diarrhea (not clear what the time period is!); they also look at duration of diarrhea (in days)
- **Methodology:** propensity score matching with 5 nearest neighbors to minimize the squared difference in estimated *odds* of treatment. Not clear if they did the matching with or without replacement. The estimate of impact is this bizarre estimator that I can’t figure out what on earth it’s doing:

$$\Delta \bar{H} = \sum_{j=1}^T \omega_j \left( h_{j1} - \sum_{i=1}^C W_{ij} h_{i0} \right).$$

I think  $j$  indexes the treated children (i.e. children in treated households; they don’t account for the fact that households can have multiple children and just treat each child-household pair as independent),  $i$  indexes the untreated children matched to the  $j$ th treated one. Then  $C$  is 5, the number of untreated children matched to each treated one and  $T$  is the total number of treated children in the sample. The

health status of the  $j$ th treated child is  $h_{j1}$ , and  $h_{ij0}$  is the health status of the  $i$ th untreated child matched to the  $j$ th treated child. The  $W_{ij}$ 's are "the weights applied in calculating the average income of the matched non-participants" – no idea what this means. Does it mean they didn't weight the five untreated neighbors equally in terms of income? Why are they being weighted by income? Maybe  $W_{ij}$  is the income of the household of the  $i$ th untreated child matched to the  $j$ th treated child, normalized by the total income of the households of all  $C$  matched untreated children? They don't say! Finally, the  $\omega_j$ 's are "the sampling weights used to construct the mean impact estimator". What does this mean? Are they just the sampling weights for each treated child? Meaning the sampling weights for the households of the treated children? Or are the sampling weights adjusted somehow, so they sum to 1? Again, we don't know because the authors don't specify!

- **Results:**

- Significant treatment effects on both prevalence and duration of diarrhea, but when stratifying by income, beneficial effects only appear for households in the highest 60% of income. Similarly, when stratifying by highest education of a female household member, benefits only accrue to women with more than primary education. Stratifying by both income and education reveals that education matters much more in poorer income quintiles than in richer ones. (Weird that economists use "stratify" to mean that they run the regression separately in particular subgroups.)

## Hammer and Spears (2016) "Village sanitation and child health: effects and external validity in a randomized field experiment in rural India"

- **Goal:** 1) estimate effects of rural sanitation on child height; 2) document evidence suggestive of externalities; 3) external validity by analyzing districts that were selected for but did not receive intervention.
- **Data:** three survey rounds collected as part of a sanitation promotion intervention run by the government of Maharashtra with the World Bank Water and Sanitation Program. Baseline collected in Feb 2004, intervention started shortly thereafter, midline survey in August 2004, and endline in August 2005 (about 18 months from baseline to endpoint).
- **Treatment variable:**

$$\begin{aligned} y_i &\sim N(\theta_{j[i]}, \sigma_y) \\ \theta_j &\sim N(\mu, \tau) \\ \mu, \tau &\sim p(\mu, \tau) \end{aligned}$$

$$\begin{aligned} y_i &\sim N(\theta_{j[i]}, \sigma_y) \\ \theta_j &= \mu + \tau\eta \\ \eta &\sim N(0, 1) \\ \mu, \tau &\sim p(\mu, \tau) \end{aligned}$$