Agenda

- WSP’s Rural Sanitation and Handwashing programs
- Outcome Measures
- Challenges (1): underreporting of symptoms
- Challenges (2): attrition
- In-focus: Appropriate recall periods for diarrhea and other caregiver-reported symptoms
WSP’s Rural Sanitation Program

- East Java, Indonesia (29 districts)
- Madhya Pradesh, India
- Tanzania (10 districts)
WSP’s Hand Washing Program

- Perú
- Senegal (4 regions)
- Vietnam (3 provinces)
- Tanzania (10 districts)
## Health and Welfare Impacts

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Diarrhea prevalence</td>
<td>Caregiver reported symptoms collected in a 14-day health calendar</td>
<td>Household questionnaire</td>
</tr>
<tr>
<td>Child growth</td>
<td>Anthropometric measures: - Weight - Height - Arm and head circumference</td>
<td>In-household collection of anthropometric measures</td>
</tr>
<tr>
<td>Child developmental and social-emotional screening</td>
<td>Composite measures</td>
<td>Ages and stages questionnaires adapted</td>
</tr>
<tr>
<td>Anemia</td>
<td>Hemoglobin concentration (&lt; 110g/L per international standards)</td>
<td>In household collection and analysis of capillary blood using the HemoCue photometer</td>
</tr>
</tbody>
</table>
## Outcome Measures (II)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Parasite prevalence</td>
<td>Egg counting of stool preserved samples with modified Kato Katz method; ELISA for protozoan</td>
<td>In-house collection of stool from children (0-24 mo) at baseline</td>
</tr>
<tr>
<td>Environmental contamination (selected countries &amp; measures)</td>
<td>• Membrane filtration for prevalence of E.coli; total coliforms</td>
<td>In-house collection of water samples, rinse of sentinel objects and hands, water source samples</td>
</tr>
</tbody>
</table>
| Economic                          | • productivity and time savings—caretakers  
• facilities and implements (e.g. soap) costing - capital cost and maintenance  
• opportunity costs             | Household questionnaire |
| Education                         | School enrollment and attendance | Household questionnaire |
### Handwashing Behavior and Determinants

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Handwashing with soap behavior</td>
<td>Direct observation of place for handwashing stocked with soap and water</td>
<td>Household questionnaire</td>
</tr>
<tr>
<td></td>
<td>Self-reported handwashing with soap behavior</td>
<td>Household questionnaire</td>
</tr>
<tr>
<td></td>
<td>By event: time, juncture, by who, handwashing materials, both hands</td>
<td>Structured Observations</td>
</tr>
<tr>
<td>Determinants of handwashing with soap behavior</td>
<td>Opportunity, ability, and motivation determinants</td>
<td>Household questionnaire</td>
</tr>
<tr>
<td>Determinants of sanitation behavior</td>
<td>Opportunity, ability, and motivation determinants</td>
<td>Household questionnaire</td>
</tr>
</tbody>
</table>
Issues in Outcome Measurement

- Behavior change
  - Hand washing (structured observations)
  - Latrine/toilet use
- Measures of “intangibles”: privacy, security, prestige
- CLTS and collective responses: reputation, shame (trust games?)
- Open-defecation free communities: 100%? (transect walks?)
- Self-report: not truth but socially acceptable answers? i.e. knowledge vs action effects (list randomization, non-verbal response methods?) social desirability bias
- Subjective appreciation: different answers for same type of facilities? (vignettes?)
### WSP’s Impact Evaluations – Sample Summary

<table>
<thead>
<tr>
<th>ARMS (T,C)</th>
<th>clusters/arm</th>
<th>HH sample size</th>
<th>Stool sample size</th>
<th>Water (hh)</th>
<th>Community Survey</th>
<th>HW structural obs.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>BL</td>
<td>EL</td>
<td>BL</td>
<td>EL</td>
<td>BL</td>
</tr>
<tr>
<td><strong>INDONESIA</strong></td>
<td>2</td>
<td>80</td>
<td>2,090</td>
<td>2,500</td>
<td>0</td>
<td>1,120</td>
</tr>
<tr>
<td><strong>MP, INDIA</strong></td>
<td>2</td>
<td>40</td>
<td>2,000</td>
<td>3,000</td>
<td>260</td>
<td>1,200</td>
</tr>
<tr>
<td><strong>PERU</strong>*</td>
<td>3 (T), 2 (C)</td>
<td>40</td>
<td>3,500</td>
<td>3,500</td>
<td>160</td>
<td>1000</td>
</tr>
<tr>
<td><strong>VIETNAM</strong></td>
<td>2</td>
<td>140 (T), 70</td>
<td>3,150</td>
<td>3,150</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>SENEGAL</strong></td>
<td>2</td>
<td>54, 56</td>
<td>1,550</td>
<td>1,550</td>
<td>110</td>
<td>0</td>
</tr>
<tr>
<td><strong>TANZANIA</strong>*</td>
<td>3 (T), 1 (C)</td>
<td>47(3), 48</td>
<td>1,700</td>
<td>4,500</td>
<td>NA</td>
<td>2,000</td>
</tr>
</tbody>
</table>

**TANZANIA***:  
T1 = handwashing  
T2 = handwashing and rural sanitation  
T3 = rural sanitation

**PERU***:  
T1 = mass media treatment at the provincial level  
T2 = social mobilization treatment at the district level  
T2-schools = promotion of handwashing behavior in primary schools  
C-schools= children who attend similar primary school where HW promotion not offered
Challenges (1): Symptom underreporting

- Large declines in both diarrhea and non intervention or season related symptoms:
- Underreporting due to health cadres collection effect?

Halted repeated rounds
Challenges (2): Attrition

- Some countries showed a decline in sample size in first and subsequent rounds
- Long questionnaires effect? (country-particular)
- Child not present
- Supervision problems?

![Graph showing large attrition in Peru and Senegal](image-url)
Some strategies: Repeated measurements and attrition

- Supplement samples (original replacement rules)

Eligible households:
- Living in the cluster at time of baseline (not moved in)
- Age eligibility projected forward

- Supervision problems?
  - Cash prize for enumerator team with lowest rate of attrition
  - Adherence to replacement rules- if hh not found, workload is not reduced
In Focus:
What is the appropriate recall period to ask about diarrhea symptoms?

- Numerous studies have looked into recall bias in caregiver-reported symptoms and diarrhea
- It is presumed that reporting closer to the interview is more accurate (Ross 1986)
- Recall up to 2-3 days is the least biased
- Recall beyond 7 days is very poor
- Recommendation has been to ask about diarrhea in the past 2-3 days to minimize bias.
- Our recommendation differs from this – we will explain why.
In the WSP baseline, diarrhea reporting declines after day 3 and falls precipitously after day 7.
When used to estimate treatment effects (RR or RD), poor symptom recall creates a bias-variance tradeoff.

- Unbiased and highly precise estimators are desired but often unavailable.
- It is often better to use a slightly biased/precise estimator than an unbiased/noisy estimator.

With symptom recall, we face a tradeoff:
- Short recall (2 days) = minimal bias
- Longer recall = more information (lower variance)
- Where is the “optimal” recall window?
Diarrhea Recall
The MSE helps us chart the bias-variance tradeoff

- The Mean Squared Error (MSE) is a function that combines bias and variance:

\[ \text{MSE} (\hat{\theta}) = \text{Var} (\hat{\theta}) + \text{Bias} (\hat{\theta}, \theta)^2 \]

- Estimators with lowest MSE optimize the bias-variance tradeoff (Geman 1992)

- In simulations from WSP data, bias dominates the MSE.

- But, bias is small (0.1 – 0.2 %) with up to 7 days of recall
A 7 day recall window minimizes the MSE for the risk difference in diarrhea in 4 of the 5 WSP cohorts.
Diarrhea Recall:
Use of 2 day versus 7 day recall requires larger samples

### Relative Sample Size Required
2 day vs. 7 day Recall

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Period Prevalence</th>
<th>Point Prevalence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Senegal</td>
<td>1.52</td>
<td>1.20</td>
</tr>
<tr>
<td>Indonesia</td>
<td>1.77</td>
<td>1.27</td>
</tr>
<tr>
<td>India MP</td>
<td>1.73</td>
<td>1.40</td>
</tr>
<tr>
<td>India HP</td>
<td>1.89</td>
<td>1.51</td>
</tr>
<tr>
<td>Peru</td>
<td>1.92</td>
<td>1.52</td>
</tr>
</tbody>
</table>

Note:
Period prevalence = any diarrhea in the last X days?
Point prevalence = number of days ill / days of observation
Diarrhea Recall: Summary

- Use a 7 day recall with daily symptom records
- Recall windows beyond 3 days are biased, but the bias is very small in RR and RD estimators
- Recall beyond 7 days is very unreliable
- 7 day recall optimizes the bias-variance tradeoff
- Reducing the recall window from 7 days to 2 days would require an increase in sample size in the WSP cohorts of 20 – 92%
Contacts and references

Bertha Briceno  bbriceno@worldbank.org
Ben Arnold  benarnold@berkeley.edu

WSP Impact Evaluation Baseline Reports:


http://www.wsp.org/
Additional References


Ross, D. A. & Vaughan, J. P. Health interview surveys in developing countries: a methodological review.. *Stud Fam Plann*, 1986, 17, 78-94


Schmidt, W.-P.; Genser, B.; Barreto, M.; Clasen, T.; Luby, S.; Cairncross, S. & Chalabi, Z. Sampling strategies to measure the prevalence of common recurrent infections in longitudinal studies. *Emerging Themes in Epidemiology*, 2010, 7, 5-


Simulated cohorts with and without recall bias using a validated model for recurrent infections (Schmidt 2007, 2009, 2010) with parameters from WSP

Relative risk estimators with 7 day recall have little bias (< 0.1%)

Risk difference estimators with 7 day recall also have very little bias: 0.1 – 0.2 percentage points.

A typical profile from the India MP cohort:

True Risk Difference (RD)  -1.18%
RD at 7 days with error  -1.07%
Bias in the RD  -0.11%