STATISTICAL ANALYSIS PLAN

"Marklate don cam: Scaling bundled health services in rural Sierra Leone"

Authors and institutional affiliation (alphabetical):

- Sara María Gómez Mesa, Yale University
- Niccolò Francesco Meriggi, PHD, University of Oxford and Wageningen University
- Ahmed Mushfiq Mobarak, PHD, Yale University
- Maarten Jan Voors, PHD, Wageningen University
- Uday Wadehra, Wageningen University

Implementation Partners: Ministry of Health of the Government of Sierra Leone, Concern Worldwide

Country: Sierra Leone

Date of latest draft: 10 November 2024

Contents

Regression specifications and hypotheses

1. Analysis of primary outcomes

Our primary estimand is the Intend To Treat effect.

For hypothesis 1: To estimate the impact on vaccine coverage we use an OLS regression at the village level:

$$y_{c,t} = \alpha_k + \beta\, Treat_c + \delta_t + X_c'\gamma + \varepsilon_{c,t} \quad (1)$$

Where $y_{c,t}$ is an indicator of the count of children under 5 and girls 10-17 years old in community c at time t has received verified government-recommended vaccines and health products (Vitamin A/Deworming/IPTi) in the routine childhood immunization schedule divided by the eligible population; $Treat_c$ is an indicator whether community c was assigned to receive the bundle; α_k is randomization stratum fixed effect; and δ are survey wave fixed effects, $X_c'\gamma$ is a vector of (average) baseline covariates including gender, age, education level, baseline coverage and $\varepsilon_{c,t}$ is the robust error term.

As descriptive analysis we will also assess impacts on (i) coverage by verbal attestation, (ii) uptake, including additional people that show up (from other villages, those missed during our village census and temporary migrants, etc), (iii) the number of zero-dose children, the share of doses missed, the timeliness of vaccination and the number of partly immunized children.

For hypothesis 2: To estimate the impact on the availability and use of chlorine we use an OLS regression at the household level:

$$y_{j,c,t} = \alpha_k + \beta\, Treat_c + \delta_t + X_j'\gamma + \varepsilon_{j,c,t} \quad (2)$$

Where $y_{j,c,t}$ is an indicator that household j in community c, at time t, has chlorine available or verified use of chlorine in drinking water; $Treat_c$ is an indicator whether community c was assigned to receive the bundle; α_k is randomization stratum fixed effect; and δ are survey wave fixed effects, $X_j'\gamma$ is a vector of baseline household covariates including (average) gender, age, education level, baseline vaccination status and $\varepsilon_{j,c,t}$ is the robust error term clustered at the community level.

Since we are testing multiple hypothesis, we will p-values correcting for the false discovery rate (FDR). For all tests, we will also report the "naïve" or "per comparison" p-value.

If there is large non-compliance we will also estimate the Local Average Treatment effect.

2. Analysis of secondary outcomes

For SH1: To estimate the impact on use of ORS/Zinc we use an OLS regression at the individual level in project households:

$$y_{i,j,c,t} = \alpha_k + \beta Treat_c + \delta_t + X_{i,j}'\gamma + \varepsilon_{i,j,c,t} \quad (3)$$

Where $y_{v,i,j,c,t}$ is an indicator of whether a child 0-5 years old i, in household j, in community c at time t has used ORS/Zinc in the past 6 months; $Treat_c$ is an indicator whether community c was assigned to receive the bundle; α_k is randomization stratum fixed effect; and δ are survey wave fixed effects, $X_{i,j}'\gamma$ is a vector of baseline household covariates including gender, age, education level, baseline vaccination status and $\varepsilon_{v,i,j,c,t}$ is the robust error term clustered at the community level.

Impacts on knowledge and attitudes

For SH2: To estimate the impact on knowledge of vaccines and health products, we use equation (2) for adult caregivers in project households

For SH3: To estimate the impact on attitudes toward vaccines and health products, we use equation (2) for adult caregivers in project households

Health outcomes

For SH4: To estimate the impact on the incidence of malaria for children under 5, we use equation (3) for the eligible population

For SH5: To estimate the impact on the incidence of diarrhea for children under 5, we use equation (3) for project households

Indirect effects

For SH6: To estimate the indirect impact on diarrhea for household members of targeted children, we use equation (3) to estimate the incidence of diarrhea amongst children above 5 and adults in project households.

For SH7: To estimate the indirect impact on use of ORS/Zinc to treat diarrhea for household members of targeted children, we use equation (3) to estimate the use of ORS/Zinc for children above 5 and adults in project households

For SH8: To estimate the indirect impact on malaria for household members of targeted children, we use equation (3) to estimate the incidence of malaria amongst children above 5 and adults in project households.

SH9: To estimate the indirect impact on diarrhea for non-project households, we use equation (3) to estimate the incidence of diarrhea amongst children in non-project households.

Health Expenditures

For SH10: To estimate the impact on curative health expenditures, we use equation (2) to for project households

3. Treatment effects by distance to health facilities

For SH11: Treatment effects (across all primary and secondary outcomes) are larger for villages further away from health facilities.

We also use equation (1) – (3), to estimate heterogeneous treatment effects adding distance from a community the closest health facility and its interaction with the treatment indicator. As distance measures we use both:

- Travel distance in minutes from the community to the closest health facility in minutes.
- Travel distance cost from the community to the closest health facility.

We will also analyse impacts on (i) coverage by verbal attestation, (ii) uptake, including additional people that show up (from other villages, those missed during our village census and temporary migrants, etc), (iii) the number of zero-dose children, the share of doses missed, the timeliness of vaccination and the number of partly immunized children.

4. Heterogeneity in treatment effects

We propose a further exploratory analysis of background conditions that may moderate treatment effects. We incorporate (average) respondent characteristics listed below as well as interactions with the treatment indicator into equations (1) - (3):

- Respondent gender
- Respondent age group: 18-40, 41-60, 61 or older.
- Respondent education level: none, primary or secondary, university or higher.
- Respondent baseline vaccine status
- Respondent baseline income
- Respondent baseline vaccine trust and hesitancy indices.
- Baseline attitudes towards traditional authorities.

5. Descriptive Analyses of Process and Context

We will prepare simple descriptive summaries of the process data from households and Health Outreach Team and contextual data from the health facility surveys. These summaries will be used to understand more precisely how the intervention functioned and also what contextual factors may help to explain both patterns in the control group and patterns in effects.

6. Cost effectiveness

The costs for all the resources needed for the programme's delivery will be recorded and organized into relevant categories. These categories include i) the cost of recruiting, training, and deploying (eg. transport and wages) Health Outreach Teams, ii) the beneficiary costs both for a) opportunity cost of time spent attending sessions and b) people's travel time to health facilities, c) possible treatment and check-up costs incurred by beneficiaries, e) opportunity costs for beneficiaries visiting clinics for check-ups or treatments, iii) treatment costs incurred by the health sector for services falling under the "free healthcare" policy.

This information will be analyzed in combination with study outcomes to calculate the cost-effectiveness/utility of the proposed intervention and compare it to routine care. The analysis will pay particular attention to the different cost categories, and it will be used to calculate DALY and QALY under different scenario simulations. The simulation will allow to e.g. estimate changes in cost-effectiveness associated with scaling up of the intervention, knowing that, fixed costs will

likely decrease as the initiative scales, whereas the variable costs will increase proportionally and therefore, capture the incremental cost per DALY. The cost-effectiveness analysis will be complemented by a budget impact analysis to determine the affordability of the intervention for GoSL, MoH, and their partners.