Remote Pre-Analysis Plan
Tanzania Childhood Pneumonia Project, Phase II

May 29, 2020

Prepared prior to receiving data from remote data collection Round 1

Table of Contents

List of Abbreviations

ARI	Acute respiratory infection
ADDOs	Accredited Drug Dispensing Outlets
ALU	Artemether-lumefantrine
Amox-DT	Amoxicillin Dispersible Tablets
Amox-OS	Amoxicillin Oral Suspension
CTC	Care and Treatment Clinics
DMO	District Medical Office
GoT	Government of Tanzania
ILS	Integrated Logistics System
IMCI	Integrated Management of Childhood Illness
M&E	Monitoring and evaluation
ME&L	Monitoring, evaluation, and learning
MoHCDGEC	Ministry of Health, Community Development, Gender, Elderly and Children
MSD	Medical Stores Department
Peds-Amox	Pediatric Amoxicillin
R&R	Report and Request
R4D	Results for Development Institute
TFDA	Tanzania Food and Drugs Authority
URTI	Upper respiratory tract infection
WHO	World Health Organization

Introduction

This pre-analysis plan for Results for Development's (R4D) first round of remote data collection for Phase II was drafted during data collection, prior to receiving data and before any analysis.

Project Name: Monitoring and Evaluating Availability, Stocking and Dispensing of Pneumonia Treatments in Public and Private Sector Markets in Tanzania (Phase II)

Principal Investigators:

- Dr. Ntuli Kapologwe, PO-RALG
- Machumu Miyeye, MoHCDGEC
- Cammie Lee, Results for Development

Co-Investigators:

- Jean Arkedis, Results for Development
- Dr. Erin Fletcher, Results for Development
- Taylor Salisbury, Results for Development
- Yangzom Tshewang, Results for Development
- Zara Ali, Results for Development

Other partners: The partners for this project are Tanzanian Ministry of Health, Community Development, Gender, Elderly and Children (MoHCDGEC), Medical Stores Department (MSD), National Institute for Medical Research (NIMR), Tanzania Food and Drugs Authority (TFDA), and Pharmacy Council. R4D is responsible for procuring and donating amoxicillin dispersible tablets (Amox-DT) to the Government of Tanzania (GoT), with time-bound funding support, and R4D, NIMR, and MoHCDGEC are responsible for conducting the study outlined in this protocol. The survey firm, EDI Global (EDI), has been hired to conduct data collection.

Intervention Overview

I. Project Overview

R4D has partnered with Tanzania's Ministry of Health, Community Development, Gender, Elderly and Children (MoHCDGEC) to help achieve the Government of Tanzania's (GoT) goal to reduce childhood mortality. Given the disproportionate childhood mortality resulting from pneumonia, R4D and MoHCDGEC are working to increase access to the first-line pediatric pneumonia treatment – amoxicillin dispersible tablets, or Amox- DT – and to improve childhood pneumonia diagnosis and prescription.

Since May 2016, the main goal of R4D's Tanzania childhood pneumonia program has been to increase the coverage rate of Amox-DT by 15 percentage points by the end of 2021.[1] To this end, for the past three

---

[1] The program focuses on Amox-DT, as it is the recommended first-line treatment for childhood pneumonia. Amoxicillin has been shown to be 4-15% more effective in treating pneumonia than other

years, R4D has co-financed the procurement of Amox-DT for public health facilities and supported improved forecasting and planning in the public sector.

The R4D engagement is comprised of two main program areas: 1) Public sector implementation to administer catalytic funding for Amox-DT procurement and to accelerate pneumonia policy changes; and 2) A monitoring, evaluation, and learning (ME&L) component to assess changes in Amox-DT supply, understand underlying causes of misdiagnosis and mistreatment in childhood pneumonia in public health facilities, and test potential interventions to address these causes.

In Phase I of this study, data collection was conducted in public health facilities as well as private health care facilities, mainly Accredited Drug Dispensing Outlets (ADDOs). In the current phase of this study, Phase II, the public sector has been prioritized and will be the focus of data collection efforts.

II. Study Aims

This document outlines an analysis plan for the first of seven rounds of the Phase II study that will be conducted using a remote data collection methodology known as Computer-Assisted Telephone Interviewing (CATI).[2] An eighth round of data collection for the study will be conducted using an in-person data collection methodology. This data collection plan deviates from the original data collection plan of an initial in-person first round and seven subsequent remote rounds. In-person data collection began in selected health facilities on March 9, 2019, and just over a week later, on March 16, the first COVID-19 case was confirmed in Tanzania. The study investigators decided to immediately suspend the in-person survey for research ethics reasons and to comply with social distancing recommendations from the WHO.

The first remote round was originally scheduled to begin in June 2020. Due to the suspension of the in-person survey, R4D moved up the start date of the remote round to May to ensure a full round of the Phase II study was completed as soon as possible. This remote round began on May 12, 2020 and will end on or about June 8, 2020.

This pre-analysis plan was drafted prior to analyzing any Phase II data and is intended to cover all seven remote rounds. However, we intend to revisit the content of the survey as the COVID-19 situation progresses. If items are added to address companion research questions in subsequent rounds, we will submit amendments to this pre-analysis plan to reflect those changes. No subtractions to the plan are anticipated. The pre-analysis plan for a final in-person round will be described in a companion document.

While the remote and in-person rounds of this study are similar, the survey instruments for the remote rounds have fewer questions. The remote round instruments have been shortened to reduce the time it takes to complete the survey and thereby limit respondent fatigue, which sets in more quickly over the phone than in-person.

---

antibiotics. As a result, WHO revised its guidelines in 2011 to recommend amoxicillin over co-trimoxazole. In 2013, UNICEF began supporting dispersible tablets over oral suspension as the recommended formulation of amoxicillin for children under five since they are easier to dose and administer to children, have a longer shelf life, are cheaper, and offer supply chain advantages due to lower volume and weight (UNICEF, 2013).

[2] In a CATI survey, an enumerator collects information from respondents by calling them on the phone and recording their answers on a tablet.

The primary aim of Phase II[3] is to:

> **Estimate availability[4] and stocking[5]** levels of pediatric amoxicillin[6] (peds-Amox) in the public health facilities

The secondary aim[7] is to:

> **Estimate and identify availability and stocking levels of other childhood illness treatments.** In particular, we will focus on the availability and stocking of cotrimoxazole oral suspension (cotrim OS) – a broad spectrum antibiotic that is often used as an alternative treatment for non-severe pneumonia – artemether lumefantrine (ALU)/Coartem, Zinc/ORS co-pack, Gentamycin, ferrous sulphate/folic acid (FeFol), Benzylpenicillin injection (penicillin G/Chrystapen), and paracetamol

Research Methodology

This round of surveying is intended to capture changes in Amox-DT availability and stocking as a result of the first three years of R4D's partnership with GoT to facilitate Amox-DT importation and distribution across Tanzania. However, since some R4D-funded Amox-DT had already been distributed to health facilities during the initial round of data collection (Phase I, Round 1), a true baseline could not be established. Consequently, any differences measured between baseline and this round of data collection may underestimate R4D's influence on Amox-DT availability and stocking.

---

[3] The in-person round includes one additional primary aim: Assess provider knowledge of pneumonia diagnosis and prescription protocols. This primary aim is excluded from the remote round primary aims as no interview is conducted with a health care provider in the remote rounds. This is because, as referenced above, the remote survey instrument is shorter; it is also focused on one primary respondent: the stock room attendant.

[4] Availability is a binary variable reflecting the presence of one or more units of a medicine in a facility versus zero units.

[5] Stocking is defined as whether there has been a stockout in the 90 days prior to the facility visit by the survey team.

[6] Pediatric amoxicillin refers to the two primary pediatric formulations of amoxicillin that are intended for use with children under five, Amox-DT and amoxicillin oral suspension/syrup (Amox-OS)

[7] The in-person round includes on additional secondary aim: Assess how providers alter prescription behavior when faced with a peds-Amox stock out. As described in the footnote above, no interview is conducted with health care providers in the remote rounds.

I. Research Questions & Indicators

The goal is to monitor whether Amox-DT, and peds-Amox more broadly, are reaching health facilities and being dispensed to children appropriately. The four main research questions that guide Phase II are[8]:

1. What is the availability of peds-Amox in public health facilities?
2. Has the availability of peds-Amox in public health facilities changed over the past three years? If so, by how much?
3. What are the stocking levels of peds-Amox in public health facilities?
4. Has the stocking of peds-Amox in public health facilities changed over the past three years? If so, by how much?

The primary and secondary outcomes from the survey used to explore each of these questions are presented in the Outcomes section, in *Table 1*.

II. Sampling Method

Phase I

Prior to Phase I, R4D worked with the MoHCDGEC to obtain a list of all public health facilities in Tanzania at three levels: district hospitals, health centers and dispensaries. Using this list, we implemented a stratified two-stage sampling method to select a nationally representative sample of public health facilities across the 26 mainland regions of Tanzania to be included in the baseline and subsequent rounds of data collection. The sampling process is outlined below.

1. Randomly selected three districts from each of the 26 regions of mainland Tanzania[9,10,11]
2. Randomly selected facilities based on the category of facility from each council:
   - One district hospital was selected per district (if available)[12]
   - One health center was selected per district (if available)
   - Seven dispensaries were selected per district[13]

---

[8] In line with the explanation above, the in-person round includes an additional research question related to interviews with health care providers: Has provider knowledge of proper pneumonia diagnosis and treatment changed over the past three years? In the remote rounds, no interview is conducted with health care providers.

[9] One and two councils were sampled from Mbeya and Songwe regions, respectively. In 2016, the Songwe region was created from the western part of Mbeya region. Two councils in Kagera region were excluded because the survey was piloted in those areas. Two councils from Arusha region were replaced at the request of the survey firm, EDI, due to the councils being located inside national parks, which would have presented significant logistical difficulties.

[10] Councils were weighted according to the proportion of public health facilities in each council so that councils with a higher number of public health facilities would have a greater probability of selection.

[11] The list of active public health facilities at the dispensary, health center, and district hospital level was downloaded from MoHCDGEC's Health Facility Registry in January 2017.

[12] District hospitals were over-sampled since one of the survey objectives focused on the availability of electricity and oxygen at district hospitals.

[13] Districts with fewer than seven dispensaries included all dispensaries.

This resulted in a total sample of 624 health facilities. The sample was comprised of 53 district hospitals, 50 health centers, and 521 dispensaries, out of a total of 61 district hospitals, 477 health centers, and 4,471 dispensaries in Tanzania (according to Tanzanian Government Open Data information from July 2016). When possible, the same facilities were surveyed in each round of data collection to allow outcomes within each facility to be compared over time.

Phase II

For the eight rounds of data collection in Phase II, we selected a sub-set of approximately 250 health facilities from the 2017 Phase I sample of public health facilities using a geographically stratified randomization approach. The sampling process was as follows:

1. Ran power calculations and determined the need to select 10 regions.
2. Randomly selected one region from each of the eight geopolitical zone in Tanzania,[14] for a total of eight regions selected.
3. Purposively selected the Dar es Salaam and Mwanza regions to ensure population and urban representation, resulting a total of 10 regions selected.

In the 2017 study, there were an average of 25 facilities sampled per region, across three districts. All facilities from the 2017 Phase I sample in these ten regions were included in our Phase II sample, resulting in a sample size of 251 health facilities. This approach will allow outcomes to be compared over time within each facility from 2017 to 2021 and allows to detect a change of 8.51 percentage points.

If a facility is permanently closed over the course of Phase II data collection, a facility of the same category (district hospital, health center, dispensary) will be randomly selected from the same district to serve as a replacement facility in the sample.

III. Data Collection

Round 1 of Phase I was conducted in March 2017,[15] Round 2 was conducted in July 2017, and Round 3 was conducted November 2017.

Data collection for the first remote round of Phase II will be conducted over 3 weeks from May 12, 2020 to June 8, 2020.

Data collection methods

The remote public health facility survey employs a CATI data collection method in which enumerators speak with respondents over the phone and record responses in a tablet as the responses are given. The

---

[14] Geopolitical zones were defined as those used in the 2015/2016 Demographic and Health Survey (DHS) and 2017 Malaria Indicator Survey.

[15] The first round of data collection in March 2017 was intended to serve as a baseline measure of Amox-DT availability, stocking, and dispensing before R4D-funded Amox-DT reached health facilities. However, since peds-Amox had already reached emergency stocking levels nation-wide, R4D and the MoHCDGEC agreed to start Amox-DT distribution upon its arrival in country in late 2016. The first shipment of R4D-administered Amox-DT reached MSD zonal stores on January 15, 2017 and began distribution to health facilities shortly thereafter.

survey is a structured interview that includes requests for respondents to conduct stock counts and carry out record reviews.

The survey is comprised of three questionnaires:

1. **Facility questionnaire:** Conducted with the facility in-charge to obtain consent to survey health facility staff and collect data remotely; identify potential respondents for the remote data collection rounds; and gather background information about the facility and information on the availability of oxygen treatment, pulse oximeters, and electricity.

2. **Respondent introduction questionnaire:** Conducted with the staff member identified by the facility-in-charge as someone who is familiar with the availability and stocking of pediatric medications at the facility, typically the stock room in-charge or stock room attendant. The goal of this questionnaire is to introduce the respondent to the remote data collection methodology and provide an instructions sheet on how to collect the medicine availability and stock levels of peds-Amox and other comparator medicines from facility dispensing outlets and stock rooms.

3. **Dispensing outlet and stock room questionnaire:** Conducted with the same staff member who participated in the respondent introduction questionnaire (questionnaire 2). The goal of this questionnaire is to asses availability of peds-Amox and other comparator medications, as well as to record the stocking levels of peds-Amox in all dispensing outlets (excluding Care and Treatment Clinics, or CTCs) and all stock rooms at the facility.

Table A1 in the *Appendix* summarizes the three questionnaires.

Survey Firm

For Phase II data collection, R4D is partnering with the same survey firm, EDI Global, that conducted the three initial rounds of data collection in Phase I.

EDI Global (EDI) is responsible for:

- Recruiting and hiring the enumerator team
- Programming the data collection tools in proprietary software, Surveybe©
- Translating tools and instructions into Swahili
- Conducting enumerator training
- Managing the logistics of data collection
- Administering surveys
- Supervising enumerators to ensure high quality data collection
- Collating and cleaning the data

The EDI data collection team is led by a Team Leader and Project Coordinator with the assistance of a Data Processing Officer. The survey is being conducted by four enumerators who are overseen by two supervisors.

Several quality assurance measures are being implemented by EDI to ensure high-quality data, along with oversight measures by R4D. Quality assurance measures include:

- **Training oversight:** During the five-day enumerator training, an R4D Senior Project Officer and Senior Program Associate were present to answer questions and provide explanations about the protocols.

- **Electronic data collection:** Electronic data collection allows for automated skip patterns and built-in consistency checks to identify errors and missing values during the interview, allowing enumerators to correct their submissions while still at the facility.

- **Field review:** Before data is transmitted to the EDI headquarters, a supervisor checks each survey completed by enumerators for possible errors and areas needing clarification with the facility.

- **EDI consistency checks:** The EDI Data Processing Officer and Team Leader perform a range of checks on incoming data throughout data collection to check for inconsistencies and errors across field teams.

- **Call observations (spot checks):** EDI supervisors dial in to listen to a subsample of interviews – one per enumerator per questionnaire – and complete observation checklists. Observations serve as an opportunity to provide feedback on data collection, enforce best practices in data collection, and send generalizable notes to the full survey team.

- **Survey back-checks:** EDI created an abbreviated version of the original questionnaires for re-interviews in a subset of public health facilities. The responses from re-interviews are compared to the original responses. Re-interviews occur in a representative sample in terms of region, facility type, and questionnaire, and are completed for 20% of interviews per enumerator.

- **Facility follow-up during data cleaning:** When the EDI data processing team finds errors or inconsistencies during data cleaning, the team contacts the corresponding enumerator or respondent on the phone to clarify the response and update the interview file.

Analysis

I. Summary Statistics

First, we will present summary statistics of key metrics for outcomes related to availability and stocking of peds-Amox. This will allow us to compare changes in our metrics of interest over time. Summary statistics for health facilities in Phase II will be visualized in comparison with Phase I, Rounds 1-3 values of the same outcomes.

Summary statistics will further be disaggregated as applicable by:

- Amoxicillin type (Amox-DT, Amox-OS, and peds-Amox)
- Health facility type
- Integrated Logistics System (ILS) subgroup
- Region
- MSD zone

This study focuses on the availability and stocking of peds-Amox, but we also want to understand the context of how the system is operating overall. Therefore, our survey also tracks the availability of other comparator medications: cotrim-OS, ALU, Zinc/ORS co-packs, Gentamycin, FeFol, Benzylpenicillin injection, and paracetamol. Where analysis is disaggregated by amoxicillin type, we will also compare the availability and stocking of peds-Amox to these comparator medications.

II. Outcomes

In addition to visualizing summary statistics, we will also statistically analyze changes in outcomes between Phase I, Round 1 and each round of Phase II data collection. This component of the analysis will assess how R4D's catalytic funding has influenced availability and stocking of peds-Amox in the three years since Phase I, Round 1.

Table 1 below presents the primary research questions (left column) and maps each research question to specific outcome metrics from the survey (right column). Outcome metrics are categorized as primary or secondary outcomes; additional metrics and questions from the surveys will be measured and analyzed to provide context to the primary and secondary outcomes we measure.

Table 1: Research questions and outcomes

Research Question	Outcome Metrics
Availability **1. What is the availability of peds-Amox in public health facilities?** **2. Has the availability of ped-Amox in public health facilities changed over the past 3 years? If so, by how much?**	**Primary:** % of facilities with peds-Amox *anywhere in the facility* at the time of the survey **Secondary:** Amount of peds-Amox in the facility at the time of surveying; calculation of mean and median number of treatments available % facilities with Amox-DT *anywhere in the facility*
Stocking: **3. What are the stocking levels of peds-Amox in public health facilities?** **4. Has the stocking of peds-Amox in public health facilities changed over the past 3 years? If so, by how much?**	**Primary:** % of facilities that experienced stock outs (one day or more without stock) of peds-Amox in the *last 90 days* **Secondary:** Mean number of days of current peds-Amox stock out % of facilities that have resolved a stock out of Amox-DT in the last 90 days Mean number of days of the most recent stock-out that has been resolved Mean number of days since facility received most recent Amox-DT shipment **Additional metrics that provide context:** How the most recent stock out was resolved % of facilities whose MSD Amox-DT order has not been filled

III. Analysis Description

Levels: Availability and Stocking

The primary research question we aim to answer with Phase II of this study is: What is the availability of peds-Amox in public health facilities? To answer this question, we will provide a mean peds-Amox availability estimate across all facilities in our sample, as described below. The mean stocking levels of peds-Amox in public health facilities will be reported the same way.

Survey Weights

For Phase I, we will calculate mean peds-Amox availability using facility-level observations weighted by the probability that a facility was selected from the overall full sample of facilities in Tanzania. Results will be reported with 95% confidence intervals. The probability will be calculated as: the inverse of the probability of a district being selected multiplied by the inverse of the probability of a facility being selected within a district. We construct these sample weights under the assumption that selection into the sample is endogenously related to the number of facilities in each region and district, and that distribution and usage are correlated to regional characteristics, some of which we cannot observe.

$$Weight_{Phase_I} = \frac{1}{\Pr(district\ being\ selected) * \Pr(facility\ being\ selected\ within\ a\ district)} \quad [Eq\ 1]$$

For Phase II outcomes, we will present outcomes weighted by Phase I weights multiplied by the inverse of being selected for this round:

$$Weight_{Phase_{II}} = Weight_{Phase_I} * \frac{1}{\Pr(facility\ being\ selected\ from\ Phase\ I\ sample)} \quad [Eq\ 2]$$

For Phase II, as mentioned in *Section II: Sampling Method*, a subsample was drawn from the Phase I sample. Therefore, the results of Phase II analysis will not be nationally representative. Eight regions were randomly selected using a geographically stratified randomization approach. To ensure population and urban representation in the Phase II subsample, Dar es Salaam and Mwanza were purposively selected. Including Dar es Salaam and Mwanza, the two regions with the country's largest population concentrations and urban centers, makes the Phase II subsample more reflective of Tanzania as a whole.

For Phase II, we will calculate mean availability both including and excluding Dar es Salaam and Mwanza to explore how the "purely randomly-selected" subsample differs from our "purposively-selected" subsample.

Our preferred measure of availability will be mean availability as calculated across *all 10 regions* in our subsample using the survey weights. Results will be reported with 95% confidence intervals.

Changes over Time: Availability and Stocking

The primary aim of this study is to understand overall availability of peds-Amox in Tanzania. However, we are also interested in whether the change in availability levels is associated with R4D's three years of catalytic funding. Although we cannot definitively state a causal relationship due to the absence of a true baseline and a randomized treatment, we present specifications that control for other factors that may have influenced availability levels – both positively and negatively. The simplest measure of change in availability and stocking of peds-Amox would be to subtract mean levels between Phase II, Round 1 and Phase I, Round 1. However, a simple difference may not account for other factors – aside from R4D's three years of catalytic funding – that might influence changes in those outcomes. Including control covariates, fixed effects, clustering of standard errors, as well as the robustness checks outlined below, allow us to estimate the effect of R4D's activities, controlling for other factors that have certainly changed.

To measure the effect of three years of R4D's catalytic funding on our outcomes of interest (reported in *Table 1*), we will use two specifications: paired outcomes and unpaired outcomes.

First Specification: Paired Outcomes Analysis

Our aim to is to explore how outcomes have changed within facilities, on average, as a result of three years of R4D's catalytic funding. We will use the bivariate fixed effects regression reported in [*Eq 3a*]. Standard errors will be clustered at the facility level. This model measures the change in an outcome within each facility over time and the takes the mean across all facilities in the sample. In this specification, each facility has outcomes measured at two points in time – in other words, all of the outcomes are "paired."

This model will be used to analyze outcomes for facilities which had data in both Phase I, Round 1 and Phase II, Round *T,* where *T* is the round of interest. The analysis will be repeated on every outcome metric listed in *Table 1*.

$$Y_{jt} = \beta_0 + \beta_1 R_j + \sum_{j=1}^{N} \beta_{2,j} \eta_j + \varepsilon_{jt} \qquad [Eq\ 3a]$$

Y_{jt} is the value of the outcome variable for facility *j* in round *t=T*, where *T* is the round of interest. Y_{jt} will be one of the outcome metrics listed in the right column of *Table 1*; the regression will be run for all outcome metrics in the table. R_j is an indicator variable equal to 1 if the observation for facility *j* is from Phase II, Round *T* of data collection or 0 if the observation is from Phase I, Round 1 (the "baseline"). The coefficient on R_j, β_1 is our coefficient of interest. β_1 will report the mean effect of three or more years of R4D's catalytic funding on our outcomes of interest. $\sum_{j=1}^{N} \eta_j$ represents facility category (hospital, health center, or dispensary) fixed effects. This is a set of dummy variables for each facility category; one category will be omitted to prevent collinearity. ε_{jt} is the error term for facility *j* in round *t*. Subscripts specify the observation: *j* identifies the individual facility, and *t* denotes the data collection round. We elected to use facility type rather than individual facility characteristics as with the limited data collection, many facility characteristics are highly correlated with facility type.

Second Specification: Unpaired Outcomes Analysis

Some facilities will have data on select outcomes in Phase I, Round 1 but lack the same data in Phase II, or vice versa. This may be due to, for example, a facility closing (in which case a replacement facility would be selected) or an inability to contact a facility. In the above specification, these facilities would drop out of the analysis since there is not two time periods for comparison – in other words, the observations are "unpaired." To avoid losing these observations, we will run a second specification that includes these unpaired observations. In these instances, we will use a bivariate regression model with standard errors clustered at the facility level.

This model will compare the mean of the outcome for all Phase I facilities with the mean outcome for all Phase II facilities.

$$Y_{jt} = \beta_0 + \beta_1 R_j + \varepsilon_{jt} \qquad [Eq\ 4a]$$

Y_{jt} is the value of the outcome variable for facility *j* in round *t=T*, where *T* is the round of interest. Y_{jt} will be one of the outcome metrics listed in the right column of *Table 1*; the regression will be run for all outcome metrics in the table. R_j is an indicator variable equal to 1 if the observation for facility *j* is from

Phase II, Round *T* of data collection or 0 if the observation is from Phase I, Round 1 (the "baseline"). The coefficient on R_j, β_1 is our coefficient of interest. β_1 will report the mean effect of three or more years of R4D's catalytic funding on our outcomes of interest. ε_{jt} is the error term for facility *j* in round *t*. Subscripts specify the observation: *j* identifies the individual facility, and *t* denotes the data collection round.

The two methods provide two different measurements and the effect reported by the two will differ. It is important to note that there will be selection bias in paired outcomes analysis because it excludes facilities that have closed since the initial data collection in Phase I, Round 1 or facilities that we were unable to contact. However, both models are useful as they describe the effects of R4D's program in different ways. The paired outcomes analysis examines changes in outcomes within facility over time, while the unpaired outcomes examines changes in average performance on outcomes for the whole sample over time.

All regressions will be weighted using the weights described in the *Analysis, Levels: Availability and Stocking* section above. Regressions will be run both including and excluding Dar es Salaam and Mwanza to explore how the "purely randomly-selected" subsample differs from our "purposively-selected" subsample.

Our preferred measure of change in peds-Amox availability will be the mean effect as calculated across *all 10 regions* in our subsample using the survey weights.

IV. Robustness Checks

Purely Random Selection vs. Purposive Selection

As a further robustness check on selection, we will run the specifications above including an indicator variable S_j equal to 1 if facility was in one of the eight districts from the purely random sample and equal to 0 if the facility was in the purposively selected districts of Dar es Salaam and Mwanza. The coefficient of interest in this regression, coefficient β_3, will measure if being in the purely randomly-selected sample has any influence on outcome measures as compared to being in the purposively-selected sample.

$$Y_{jt} = \beta_0 + \beta_1 R_j + \sum_{j=1}^{N} \beta_{2,j} \eta_j + \beta_3 S_j + \varepsilon_{jt} \qquad [Eq\ 3b: Paired\ outcomes\ analysis]$$

$$Y_{jt} = \beta_0 + \beta_1 R_j + \beta_3 S_j + \varepsilon_{jt} \qquad [Eq\ 4b: Unpaired\ outcomes\ analysis]$$

Health Facility Quality and Size

We hypothesize that the quality and size of a health facility would influence our outcome measures. However, the direction of the effect is ambiguous. For example, a larger health facility may have capacity to stock greater quantities of medications, but because it serves a larger number of referral patients it may also face a greater demand for those medications. Similarly, better quality health facilities may be better at ordering medicines from MSD and therefore receive more medicines; however, there could be a greater demand for services and medications at better quality facilities.

We will run the same models specified in the *Analysis* section but include facility-level variables collected in the survey instead of facility category fixed effects.

$$Y_{jt} = \beta_0 + \beta_1 R_j + \delta_{i,j} X_{i,j} + \varepsilon_{jt} \qquad [Eq\ 3c\text{:}\ Paired\ outcomes\ analysis]$$

$$Y_{jt} = \beta_0 + \beta_1 R_j + \delta_{i,j} X_{i,j} + \varepsilon_{jt} \qquad [Eq\ 4c\text{:}\ Unpaired\ outcomes\ analysis]$$

In the equations above [*Eq 3c*] and [*Eq 4c*], X_i are facility-level variables intended to proxy for the quality and size of each facility. We will include the following variables as control variables in our model:

- Mean availability of all medication surveyed (taking the mean across binary availability indicators for each medication surveyed)
- Number of health care providers for children under 5
- If healthy facility has a functional stethoscope
- If health facility is connected to the electricity grid
- If healthy facility provides oxygen therapy

Fixed Effects

The primary paired outcomes analysis model [*Eq 3a*] includes facility category fixed effects. We will also report the results of running the model with alternate fixed effects:

1. Individual facility fixed effects
2. Region fixed effects – this specification will be done removing the weighting specified in the *Analysis* section.
3. Sampling strata fixed effects

Technical Challenges

I. Multiple Hypothesis Testing

To account for the large number of hypotheses being tested in the analysis, we will adjust our results for multiple hypothesis testing by using the Holm-Bonferroni correction. This correction imposes a more conservative threshold for statistical significance since standard cutoffs would likely result in a statistically significant outcome by chance alone (i.e., false positive). This ensures that statistically significant findings are more likely to reflect actual change rather than a false positive.[16]

The Holm-Bonferroni correction is an extension of the popular Bonferroni correction, which very simply adjusts the threshold for significance by a factor equal to the number of tests being conducted. Instead of adjusting by a constant factor for all tests, the Holm-Bonferroni correction makes a sequential step-wise adjustment by increasing the threshold of the most significant result by a factor equal to the total number

---

[16] If we do multiple comparisons, the probability of a false positive is: $1-(1-a)m$, where m is the number of tests and a is the significance level (set at 0.05 by convention). For example, if we were to ask seven questions from the same data at the 5% level of significance, we would increase the chance of finding at least one false positive from 5% (with one question) to about 30%.

of tests, the next most significant results with a factor equal to the total number of tests – 1, etc.[17] The benefit of the Holm-Bonferroni test is that it accounts for outcomes that are dependent in a less conservative way than Bonferroni, thereby improving the statistical power of the test.[18] For this correction, our "families" of hypotheses are as outlined in *Table 1*.

II. Missing Data

Missing values can take the form of non-response (e.g. uncompleted surveys), partial response (e.g. "Don't know" responses), or errors in the data. Minor and random missing values can be omitted from the analysis. If we find that missing values are unexpectedly high, we may check the sensitivity of the complete case estimate using imputation methods to replace missing values. However, we expect minimal missing data since the digital survey form includes constraints that require data to be entered before the enumerator can proceed with data collection.

There are three main reasons a facility surveyed at baseline may not be surveyed in Phase II data collection rounds:

1. Temporary closures – In Phase I, Round 2, two facilities from the original sample were not surveyed due to temporary closures. These two facilities were reopened for Phase I, Round 3 surveying, but we may see other facilities drop from the Phase II sample for similar reasons.
2. Permanent closure – If a public health facility surveyed in Phase I has since closed, we will randomly select another public health facility of the same level from the same district not previously included in the study.
3. Inability to contact facility – Before the in-person round of Phase II was suspended, the survey teams visited 97 facilities of the 251 Phase II facilities and collected new contact information. The 154 remaining facilities will need to be contacted using phone numbers provided by the District Medical Officer (DMO) and phone numbers collected in Phase I. If these numbers are incorrect or outdated, we may not be able to contact the health facility. In this instance, we will randomly select another public health facility of the same level from the same district not previously included in our survey.

III. Evaluation Effects

There is some risk of various evaluation effects (also known as observer, or Hawthorne, effects), the process by which a participant of a study changes their behavior due to the knowledge that they are under study and for what specific purpose. For example, healthcare workers may try to request (unnecessarily)

---

[17] In practice, this is performed by sorting outcomes into groups called families whereby outcomes or tests within a family are correlated (dependent). Following outcome sorting, the change in the outcome across survey rounds is estimated using either model. Unadjusted p-values from each test within a family are then sorted in ascending order and corrected by multiplying the p-value of a given rank by (m – n), where m is the total number of tests within a given family of related outcomes and n is the rank order of the p-value from 0 (smallest) to m (largest). If the adjusted p-value is less than 0.05, then that outcome is statistically different between rounds.

[18] Abdi, H (2007). Bonferroni and Šidák corrections for multiple comparisons. *Encyclopedia of measurement and statistics*, 3, 103-107.

additional Amox-DT stock if they have knowledge that this is one of our primary outcomes – and especially if they discern that we think having Amox-DT stock is desirable. For these reasons, the survey tool contains additional questions about other treatments for childhood illnesses to lessen the likelihood that a healthcare worker will identify pneumonia and its associated treatments as being our major outcome of interest. By doing so, we also aim to decrease the risk of the observer effect influencing the quality of our data.

IV. Other Study Limitations

Additional limitations of this study are due to data sources, time constraints, and potential respondent biases.

These limitations include:

- The quality of the data differs based on the data source. The research team may be able to collect observable and verifiable data in some cases but will have to rely on respondent recall for others. For example, the research team may be able to verify if the medications were available in the stock rooms and dispensing outlets on the day of the visit. However, data on stock outs relies on a combination of records and recall, dependent on whether the respondent has records available for reference as well as the quality of those records.

- The most knowledgeable respondent is not always available. Enumerators are given instructions on who within the facility should respond to each questionnaire. However, the most appropriate respondent is not always available (due to illness, leave, other obligations, etc.) and the enumerator will have to interview the next most appropriate respondent who may not have as much knowledge of the topic. For example, if the person who manages the stock room is not available, their assistant may have to respond to the questionnaire.

- Some comparisons across time are not possible because of changes in the data collection tools. Data collection tools were refined after Phase I, Round 1 data collection to facilitate better respondent understanding, to add more color to the questions being asked, to streamline the data collection process, or to probe further into questions that came up in Phase I, Round 1. In some cases, these changes might mean that variables are measured differently between rounds, which may prohibit direct comparisons across rounds.

- Differences across rounds may reflect different respondents rather than true change. In some cases, respondents in Phase II may be different than those in previous rounds. Thus, differences in responses may reflect different levels of knowledge across respondents rather than change over time.

Ethical Considerations

I. Risks

This study poses limited risk to study participants. There is no risk involved in surveying public healthcare facilities or other study components for either the enumerators or the facility staff. Facility staff's regular workload and compensation is unlikely to be affected. Some survey activities may briefly stop healthcare

workers from performing their duties, but it will be stated clearly that workers are free to stop or pause at any time (such as in a medical emergency).

II. Benefits

This study contributes to the understanding of the availability and stocking of pneumonia treatments in Tanzania. Findings will benefit public and private health facilities, the Tanzanian government, and medical distributors by providing information that can improve treatment of pneumonia in Tanzania. Ultimately, the study will benefit Tanzanian children with pneumonia by improving available treatment.

III. Informed Consent

All healthcare workers will be read an informed consent form prior to the interview. The informed consent form will emphasize that participation is completely voluntary. Participants will not be provided any compensation.

An important point to note with regards to consent is that we are planning on including a more generalized study title in written information and consent forms to blind study participants and mitigate Hawthorne effects. The study title listed on the consent and written information forms will be "Monitoring and Evaluating the Availability, Stocking and Dispensing of Childhood Illness Treatments in Public and Private Sector Markets in Tanzania".

IV. Data Management and Confidentiality

Risks involving data management and data security are minimal. Trained enumerators will use smart phones and tablets to collect all data. Electronic questionnaires will be encrypted with a password. We will use Stata (version 14) to analyze the data. Primary data analysis will include quality control checks, such as checks for missing data and data entry.

Results Sharing

Any publication of study results will be decided among investigators before submitting results for publication. Results from each round of data collection will be disseminated annually, alongside any salient updates on research plans, with the relevant MoHCDGEC departments, e.g. RCHS, prevention and policy, etc. and other bodies that could benefit from the data, such as the Medical Stores Department and the Pharmacy Council. The final list of stakeholders to be invited for dissemination will be decided upon in collaboration with the MoHCDGEC. The presentation may take the form of in-person presentations to key stakeholders at the leadership level of the MoHCDGEC, MSD, and TFDA.

These annual disseminations at the national government level could be complemented with disseminations at the local level, where results could be shared among key decision makers, such as district medical officers, and regional medical officers. This may be done through sharing a brief report on the results accompanied by any salient updates on research plans as well as a list of best practices and key takeaways from the study in regard to childhood pneumonia case management. Similar to the national

level dissemination, the final list of local level stakeholders would be decided upon in collaboration with the MoHCDGEC.

The policy implications of this study are potentially twofold:

1. Short-term policy implications: These policies can result from the first few rounds of data collection and results sharing. One potential outcome could be a new policy that targets the strengthening of public supply chains for facilities with frequent pneumonia treatment shortages.
2. Long-term policy implications: After the dissemination of results following the final round of data collection in December 2021, R4D hopes to help with shaping new policies and approaches to pneumonia case management in children in Tanzania. By shedding light on a childhood illness that does not attract sufficient funding from donors and NGOs as compared to other illnesses, this case study could assist the MoHCDGEC in securing funding and programs that continue to target childhood pneumonia.

In addition to disseminating the findings of this research with key stakeholders in Tanzania, R4D will also share the findings with GiveWell. R4D will consult with its GoT partners prior to the findings being made public on GiveWell's website.

APPENDIX I: Data collection tools

Table A1: Public Facility Questionnaires Overview

Questionnaire	Purpose	Respondent	Contents
1. Facility questionnaire	Attain consent to talk to healthy facility staff, to collect data remotely, and identify respondents for the remote data collection rounds. Gather descriptive information about the facility (e.g. number of staff, availability of oxygen therapy, electricity supply).	Facility in-charge	• Structured interview (approximately 15 minutes)
2. Respondent introduction questionnaire	Introduce the remote data collection methodology and provide the instructions sheet on how to collect the medicine availability and stock levels of peds-Amox and other comparator medicines from facility dispensing outlets and stock rooms.	Staff member identified by the facility-in-charge as someone who is familiar with the availability and stocking of pediatric medications at the facility, typically the stock room in-charge or stock room attendant.	• Structured interview (approximately 10 minutes)
3. Dispensing outlet and stock room questionnaire	Understand availability of peds-Amox and other comparator medicines and stock levels of peds-Amox in the facility dispensing outlets and stock rooms.	Same staff member who participated in the respondent introduction questionnaire (questionnaire 2)	• Structured interview (approximately 25 minutes) • Stock count (dispensing outlet and stock room)