IDinsight

Concept note: Impact Evaluation of CHAI's Household Contact Management for Childhood Tuberculosis Program in India

May 2024

Context

The Clinton Health Access Initiative (CHAI), in partnership with GiveWell, is implementing a community-based household contact management (HCM) intervention that aims to improve Tuberculosis (TB) Preventive Treatment (TPT) initiation rates for under five children in India. The national guidelines for Programmatic Management of TB Preventive Treatment (PMTPT) recommend TPT for all asymptomatic household contacts of index pulmonary TB patients. In an effort to scale-up this guideline, the Joint Effort for Elimination of Tuberculosis (JEET) project launched in India in 2021. Under this program, project staff are deployed at the sub-district level to screen household members of index patients and link them to health facilities for TPT initiation.

CHAI proposes to build on this existing approach in two ways: 1) CHAI will leverage strengths of the existing system by involving three cadres of community health workers: Accredited Social Health Activists (ASHAs), TB Health Volunteers (TB-HVs), and public private support agency staff (PPSAs) who will play a key role in household-level contact tracing, management, and care; 2) These community health workers will conduct symptom screening of the household members of index TB patients, will initiate TPT for asymptomatic under-fives (U5), and will refer any symptomatic household members to the health facility for further testing, evaluation, and treatment. This pilot will be conducted in 70 districts across two Indian states, Uttar Pradesh and Bihar.

CHAI and GiveWell would like to engage an independent evaluation partner to help them understand primarily the effect of the program on TPT initiation among U5s in households of active TB patients.

CHAI and GiveWell will use the evidence from this evaluation to assess the cost-effectiveness of the program and make a decision about scaling up the program in Uttar Pradesh (UP) and Bihar, as well as other states. CHAI, GiveWell, and IDinsight had an initial discussion about this opportunity in early March 2024. This note captures that discussion as well as IDinsight's proposed approach to evaluating the impact of CHAI's program on initiation of TPTs among eligible U5s. We welcome the opportunity to further discuss and refine the approach outlined in this note with CHAI. Feedback is highly welcomed.

Proposed Approach

We propose a cluster randomized evaluation, randomized at the district level, to rigorously measure the impact of CHAI's program. This approach will allow us to have the greatest confidence that observable impacts are attributable to CHAI's program.

Research questions

This evaluation aims to answer the following primary research questions:

1. What is the effect of CHAI's HCM for Childhood TB Program on TPT initiation for eligible children under 5?[1]
2. What is the effect of the program on CHW workloads and activities (both TB and non-TB related)?

The evaluation will also aim to answer the following secondary research questions:

1. What is the effect of the program on active TB treatment coverage among index patients?
2. What is the effect of the program on active TB treatment initiation among U5s?
3. What is the effect of the program on TPT completion for eligible children under 5? What is the effect on completion for eligible children under 5, conditional on initiation?
4. What percentage of U5s in households of index patients are symptom-positive for TB at the time of data collection?
5. What is the effect of the program on the percentage of households of index patients reached per national guidelines?

[1] The proposed question in the documentation that CHAI shared is, "What is the effect of the program on TPT coverage rates among children under 5 who are reached?" The implied denominator is therefore U5s in households that are reached by a CHW (or other frontline worker in the case of control districts). However, we believe a more policy-relevant denominator is U5s in households of identified index patients.

6. What is the effect of the program on the percentage of household contacts that receive counseling and referrals?
7. What is the effect of the program on the percentage of household contacts that follow-up at a health facility for additional screening and treatment?
8. What is the effect of the program on the percentage of household contacts who are positively diagnosed and screened?
9. What is the effect of the program on the average time between identification of the index patient for a frontline worker to conduct household contact tracing?
10. What is the effect of the program on the rate of active TB among under 5 children living with a household member with active TB?
11. What percentage of TB patients who come to the health facility for TB screening make it through each step of the cascade of care (have symptoms and are diagnosed with TB; initiated on active TB treatment (if positive) or TPT (if negative)?

Overview of proposed evaluation design

To estimate the impact of the HCM program, we will conduct a cluster randomized controlled trial (RCT), randomizing at the district level. In Bihar, we will randomly assign 13 districts to receive the program and 13 districts to serve as controls. In UP, we will randomly assign 22 districts to receive the program and the remaining 22 districts to serve as controls.[2]

We will randomize districts to receive CHAI's HCM program using stratified random assignment with strata defined by JEET implementation status and any other variables which the CHAI team believes are most predictive of the primary outcome variables. This could include whether the district is rural or urban, estimates of the baseline levels of the primary outcome variable from Ni-Kshay / JEET admin data, or an estimate of CHW effectiveness from NFHS (or other) data.

In addition to the main randomized controlled trial, we will conduct a qualitative process evaluation to take place midway between baseline and endline. The process evaluation will provide insights into the experiences and perceptions of the program from caregivers of U5s, CHWs, and other program implementation staff (including government stakeholders and CHAI staff) at the district and state level.

[2] Having more control districts even while maintaining the same number of treatment districts can reduce the overall sample size. However, we understand from CHAI that this is not feasible as potential evaluation districts (treatment and control) have already been identified based on TB prevalence. Further, CHAI will roll out the program to control districts after the evaluation is completed and therefore cannot add to the overall number of districts it plans to serve.

Surveys and data collection tools

We will collect both quantitative and qualitative data to answer the evaluation questions at one of three timepoints:

- **Baseline:** Data will be collected prior to the roll-out of the program. Based on CHAI and GiveWell's feedback, this will be done in the full sample of evaluation districts.
- **Process evaluation:** Rather than conducting a quantitative midline survey, we will conduct a qualitative process evaluation ~9 months after the program is launched in the evaluation areas.[3]
- **Endline:** Data will be collected ~19 months after the program is launched in the evaluation areas to measure our outcomes.

We will conduct the following types of data collection:

- **Pilot of main survey instruments** Prior to baseline data collection, we will pilot the main data collection instruments, especially the survey of households of active TB index patients. The pilot sample size for the survey of households of active TB patients will be approximately 100 households and will include patients who have been indexed at various times prior to the survey (e.g. patients who were indexed 30 days prior, 60 days prior, etc.) so that we can get a better sense of the optimal time window to ask about details of TPT initiation and completion.

- **Surveys of households of active TB index patients:** To collect data on primary outcomes, we will sample active TB patients from the Ni-Kshay registry, locate patients' households using data from Ni-Kshay, and interview household members. We will *not* use information from Ni-Kshay to attempt to select only patients living with U5s.[4] At baseline, we will sample 5,636 active TB patients indexed approximately 30-60 days prior to the survey. At the endline, we will sample 5,636 active TB patients indexed approximately 30-60 days prior to the survey and an additional 5,636 active TB patients indexed approximately 120-150 days prior to the survey for a total of 11,272 active TB patients. (We provide more details of our sampling of active TB patients in the section "Sampling of active TB index patients" below.)

[3] We have discussed with CHAI and GiveWell teams the possibility of conducting the process evaluation at the endline, simultaneously with the quantitative endline survey. We have agreed that we will take the final decision about the ideal timeline for the interviews at a later stage in the project. For the updated budget, we have increased the number of interviews that we will conduct, and can determine later how this should be split across midline and endline.

[4] Based on communication with the CHAI team, this information may be available in Ni-Kshay in program districts but using this information would create a strong risk of bias if the information is incomplete for certain types of patients.

Our survey respondents will be the index patient, head of household or other adult member of the household (other than the index patient), and the primary caregiver of the U5 child(ren) in the household (if different from either of the above). In addition to collecting outcome data, we will also collect data on various household characteristics which may help increase the precision of our estimates.

Validation of TPT initiation: During the household survey of index patients, we will supplement info on recall of TPT initiation of U5s with observation of strips of 3HP tablets to verify U5s' intake of 3HP doses. For now, we assume that caregivers are distributed tablets for a month at a time based on the last call. We will describe the operational risks of accommodating this timeline in our survey and provide ideas to mitigate the risks. The survey will also include additional probes around challenges with consumption, methods of disposal, etc. to indirectly verify recall of doses administered. Finally, we will triangulate the initiation recall with Nikshay data, acknowledging that the Nikshay data will be comparable between treatment and control districts at baseline, but may not be so at the endline if the intervention leads to data quality improvements in the treatment districts.

- **Surveys of caregivers of children who have initiated TPT for Isoscreen verification:** To validate self-reported data on TPT adherence, in treatment districts only, we will sample households of children who are recorded as having initiated TPT in Ni-Kshay approximately 10-12 weeks prior to surveying. Since TPT takes approximately 12 weeks to complete, most children in this sample should still be taking TPT. We will ask caregivers whether the child was administered a dose of TPT in the past 48 hours and, if so, administer a diagnostic test (IsoScreen) to verify adherence on the spot[5]. Exact details of how we sample these children will be decided at a later date based on evidence from the pilot and baseline. For example, prior to scheduling interviews, we may call caregivers to first ask when, i.e. which day of the week, they typically administer TPT. This would help ensure that survey visits are conducted within 48 hours of TPT being administered but may also lead to bias if it induces caregivers to administer TPT.

- **CHW surveys:** We will survey a random sample of TB-HVs and PPSAs to collect data on CHW workload and other variables. By comparing reported HCM and non-HCM workload for TB-HVs and PPSAs in the treatment and control groups between baseline and endline, we may obtain experimental estimates of the impact of the program on non-HCM workload and other variables.

  We will NOT include ASHA workers in this sample since the impact of the program on ASHA workload is likely too small to detect through experimental comparisons.[6] (See the

[5] We will discuss with CHAI the possibility of receiving training from the local CHAI team and/or other researchers on administering the IsoScreen test.

[6] There are roughly 2.5 lakh ASHA workers in Bihar and UP and, according to CHAI, there are about 4 lakh cases of active TB in the two states each year. Therefore, there are only 1.5 active TB patients per ASHA worker per year which means that it is unlikely that the program would seriously affect ASHA worker workload.

following subsection for more details on how we will estimate the impact of the program on ASHA workers' workload.)

- **Time-use surveys of CHWs through shadowing:** To estimate the impact of the program on ASHA workers' workload, we will conduct a time-use survey of ASHA workers performing household contact management. In the time-use survey, surveyors will shadow ASHA workers as they visit households of active TB patients for the purpose of household contact management. This data will allow us to estimate the time required to perform tasks related to the HCM program and thus crude estimates of the impact of the HCM program on ASHA workloads.

  We will also perform time use surveys of TB-HVs and PPSAs as they perform HCM tasks. This data will allow us to double check and validate the experimental estimates of the impact of the HCM program on these cadres' workloads. This may be useful if TB-HVs and PPSAs have an incentive to over or under report the impact of HCM tasks on their workload. (For example, CHWs may resent being assigned a new task and thus overreport the amount of time dedicated to the task.) We will conduct the time-use surveys for 6 ASHA workers, 6 TB-HVs, and 6 PPSAs across treatment and control areas at the endline.

- **Facility-based data collection:** At the endline, we will sample government health facilities involved in diagnosing and treating TB. At each sampled facility, we will interview key medical staff and inspect patient registers[7] to collect data on outcomes such as the share of children diagnosed with TB who begin DOTS.

  We will also use facility visits to validate self-reported data, such as whether household members actually followed up at a health facility. For those who report that they visited the sampled health facility for further screening or care, we could validate whether we can find those patients in the related registers.[8]

- **Semi-structured key informant interviews:** We will conduct in-depth interviews with various program stakeholders including household members, CHWs, CHW supervisors, district officials, state officials in both treatment and control areas, as well as CHAI district staff. Data from these interviews will be used to understand the experience and perception about the intervention from caregivers of U5s on TPT regimen and those involved in its

[7] We will continue to explore what administrative data sources would be feasible to use.
[8] Currently, our budget accounts for validation only in the subset of facilities that are selected for the facility survey. This means we could only validate outcomes for those who report going to the selected facility. Because registers may not be complete and name matching can be challenging in settings where people have similar names, we would not conclude that self-report is overestimated if we cannot identify them in registers. However, this exercise can provide some signal of the quality of self-report, particularly when comparing match rates across treatment and control participants. If CHAI would prefer that we do validation for a larger set of respondents, we can revise the budget. This method could only be used to cross-check outcomes for respondents who report accessing care at a facility. It cannot be used to cross-check when respondents report that they did not access care at a facility.

implementation. Based on previous experience of conducting qualitative interviews, we believe that the following sample size will allow us to generate insights on perceptions of and experience with program implementation:

1. 20 households/caregivers of U5s
2. 20 ASHA workers
3. 20 PPSA
4. 20 TB-HVs
5. 10 each of each CHW cadre's supervisors (total 30)
6. 10 each of more senior supervisors
7. 10 government officials
8. 5 CHAI staff at district level

We will work closely with CHAI to identify the appropriate themes to explore through qualitative interviews with these stakeholders. We have budgeted for some qualitative interviews at midline as well as some at endline. We will work with CHAI and GiveWell to determine the optimal stakeholders and split in interviews across the two points in the project.

Across these data collection activities, we will rely on several data collection tools and target respondents. These are summarized in Table 1.

Table 1. Intended Data Collection Tools

Data Collection Activity	Data Collection Instrument	Description	Target Respondent(s)	Timing of Use		
				Baseline	Midline	Endline
Household surveys	Index patient survey	Collects self-reported treatment outcome data on index patient and consent to speak with other HH members	Index patient	**X**		**X**
	Household member survey	Collects self-reported outcome data on non-U5 household members	Household head or other adult household member (not index patient)	**X**		**X**
	U-5 survey	Collects caregiver-reported data on U5 outcomes	Caregiver of U5	**X**		**X**
	Caregivers of children reported as having initiated TPT	Collects caregiver-reported completion data and adherence data, including verification	U5s recorded as having initiated TPT in Ni-Kshay			**X**

		for a sub-sample using IsoScreen point of care diagnostic test				
CHW surveys	CHW survey	Collects data on a sample of TB-HVs and PPSAs to understand their workload, daily activities, time allocations, etc.	TB-HVs, PPSAs (does not include ASHAs)	**X**		**X**
	CHW observation	Collects data from observing household TB visits with CHWs	ASHAs, TB-HVs, PPSAs			**X**
Facility-based data collection	Facility survey	Collects data from various TB registers to better understand the average number / percentage of patients who reach the various steps in TB cascade of care	Public health facilities (PHCs, CHCs, and DHs)			**X**
	Facility outcome verification	Uses facility registers to verify self-report about TB care utilization	Public health facilities			**X**
Qualitative process evaluation	Semi-structured interview guides	Collects qualitative data on perceptions of and experiences with the HCM for Childhood TB program	Household members, CHWs and program implementation staff, including government officials at the district and state level		**X**[9]	**X**

Sampling of active TB patients

At baseline, we will sample 5,636 active TB patients indexed approximately[10] 30-60 days prior to the survey. At endline we will sample 5,636 active TB patients indexed approximately 30-60 days

[9] We have budgeted for the option to split the qualitative interviews between midline and endline after further discussions with CHAI and GiveWell teams.
[10] We will explore the optimal timeframe during the pilot.

prior to the survey and an additional 5,636 active TB patients indexed approximately 120-150 days prior to the survey for a total of 11,272 active TB patients. The motivation for the sampling approach is as follows:

1. **Households of active patients indexed 30-60 days prior to surveying will be asked detailed questions about TPT initiation** – While we believe that household members will have longer recall of events such as a CHW visiting the household to talk about TB, seeking TB screening and treatment at a health facility, and initiating their U5 on TPT, it is unlikely that they will remember the specific details of that visit, including when exactly it happened, the specifics of what the CHW told them, etc, the longer the recall period. Therefore, we will try to interview them shortly after that visit would have occurred. On the other hand, we do not want to visit them too soon, so that the CHWs have had sufficient time to follow-up with the households after the index patient was identified, as we do not want to pick up effects on the timeliness of the CHW visit rather than whether it occurred at all.
2. **Households of active patients indexed 120-150 days prior to surveying will be asked detailed questions about TPT adherence and completion** – We have included this additional cohort at the endline so that we may gather data on TPT completion. As the TPT course takes approximately 12 weeks to complete, this cohort will allow us to gather data on the share of children who have completed, or are on track to complete, TPT. While data from this cohort can be pooled with data from the 30-60 day cohort for most outcomes, the purpose of this cohort is primarily to measure TPT completion (including testing a subsample with the IsoScreen test only in treatment districts). This cohort will only be included at the endline.
3. **We have included a full baseline to increase the precision of our endline estimates** – By sampling 5,636 active patients at baseline, we will be able to obtain rough estimates of the mean primary outcome variable by district which will allow us to increase the precision of our estimates at endline. As this is an RCT, it would also be possible to do away with the baseline and instead increase the endline sample size. After internal discussion with the CHAI and Givewell team it was decided to include the baseline sample to increase the precision of our endline estimates.

Table 2: Sample size summary

	Full baseline with separate cohorts	
	BL	EL
Active patients	5,636 (30-60 days)	11,272 index patients (5,636: 30-60 days 5,636: 120-150 days)
Children reported completing TPT (IsoScreen	NA	1071 (in T only)[11]

[11] Inclusive of 300 children for IsoScreen testing

validation)		
CHWs (ASHAs, TB-HVs, PPSAs)	384	384
Facilities	140	140

Table 3: Outcomes for the evaluation

Outcome	Definition	Population	Plan for data collection
PRIMARY OUTCOMES			
TPT initiation	Number of U5s beginning any TPT regimen / Number of asymptomatic U5s in the household of an index patient	U5s who live in households of index TB patients	Self-reported by caregiver
CHW workloads and activities	Proportion of time spent on TB control and management each week or month / total time spent on CHW activities each week or month	TB-HVs and PPSAs	Self-reported by CHWs
SECONDARY OUTCOMES			
Active TB treatment coverage	Number of household members who tested positive and initiate DOTs / Number of household members who report having received a positive TB diagnosis	Household members of index TB patients	Self-report at the endline household survey
TPT completion (conditional on initiation)	Number of U5s who complete the TPT regimen / Number of eligible U5s who initiate TPT	U5s who live in households of index TB patients and initiate TPT	Self-report at the endline household survey
TPT completion (unconditional on initiation)	Number of U5s who complete the TPT regimen / Number of asymptomatic	U5s who live in households of index TB patients	Self-reported by caregiver

	U5s in the household of an index patient		
Symptom-positivity among U5s	Number of U5s with positive TB symptoms / Number of U5s in the household of an index patient	U5s who live in households of index TB patients	Self-report at the endline household survey
Households reached	Total households visited by a frontline worker for contact tracing / total households with index patients	Sampled households	Self-report at the endline household survey
Household contacts that receive counseling and referrals	Number of household contacts that report receiving counseling by a frontline worker and were referred to a health facility for further screening / all household contacts	Household members of index TB patients	Self-report at the endline household survey
Household contacts that follow-up at a health facility for additional screening and treatment	Number of household contacts that report going to the health facility for additional screening / number of household contacts that were screened Number of household contacts that report receiving TB treatment from the health facility / number of household contacts that report receiving a positive TB diagnosis	Household members of index TB patients	Self-report at the endline household survey
Household contacts who are positively diagnosed and screened	Household contacts that report receiving a positive diagnosis / household contacts that were	Household members of index TB patients	Self-report at the endline household survey

	screened		
Number of days between diagnosis of the index patient and the frontline worker visit data			Will explore if administrative data is available on frontline worker visit

Risks to the Evaluation

There are a few risks to the evaluation that we want to highlight:

- **State officials will not agree to randomization - Technical risk is high, likelihood is low:** We understand from CHAI that they have discussed this with Bihar and Uttar Pradesh officials and are confident that this will not be an issue. If it is, then we will need to find an alternate study design.
- **Ni-Kshay lists are unavailable or low quality - Technical risk is high, likelihood is low:** If the Ni-Kshay lists are unavailable or suffer from quality issues, then this could: 1) affect our ability to use the list as a sampling frame, which would significantly increase the cost of the evaluation (or reduce the sample size and power); 2) bias the external validity of our sample if some types of households are systematically excluded from the Ni-Kshay list; 3) inadvertently exclude households from the program, since we understand that the program also uses these lists to conduct the HCM. This could mean the reach of CHAI's program is not as high as the evaluation data would imply.
- **The intervention improves the availability / quality of the Ni-Kshay lists - Technical risk is high, likelihood is low (for index patients):** If the quality or completeness of the list is improved by the intervention, then our samples may not be similar across the treatment and control arms. We understand from CHAI that the intervention is unlikely to affect the quality of the list of index patients but may improve the accuracy of the data for the contacts. This is why we propose using the index patient data only for sampling rather than using the Ni-Kshay list to also sample U5s.
- **The intervention affects who is diagnosed with TB (the index patient) - Technical risk is low-medium, likelihood is low:** If households are more likely to get screened and diagnosed for TB with CHAI's program than would otherwise be the case, then our households across the treatment and control groups may no longer be comparable. Further, CHWs in the treatment arm may be less effective because they are overwhelmed by patients.
- **Spillover - Technical risk is low-medium, likelihood is low-medium:** If CHWs cross district boundaries to perform household contact tracing (which may be the case if patients seek treatment at facilities outside of their home district), there may be spillover from treatment to control or vice versa. These types of spillovers are likely present but rare and can be

mitigated by excluding patients who were initially traced by a CHW from outside their district. There may also be spillover if control districts adopt some of the practices used in treatment districts. We believe that this risk is low-medium.

Deliverables

We will generate the following deliverables for this evaluation:

- Final evaluation protocol and IRB submission
- Pre-analysis plan
- Full baseline survey instruments
- Full baseline data
- Full baseline survey technical report
- Process evaluation survey instruments
- Process evaluation data
- Process evaluation survey technical report
- Endline survey instruments
- Endline data
- Final technical report

Budget estimates for the recommended evaluation design

As recommended by CHAI and GiveWell, we share the final budget estimates for the full baseline and two cohorts option.

The overall costs fall under the following categories:

- **IDinsight staff costs**: These include the costs associated with permanent IDinsight staff members
- **Temp staff costs**: These include costs associated with salaries of surveyors and their supervisors; and training-related costs (including hall rentals, refreshments, lodging and one round trip during training). These costs include taxes
- **Non staff costs**: These include costs associated with IDinsight's team travel during field visits, fuel for surveyors during data collection, software and technology, printing, IRB, insurance and emergency. These costs include taxes

The overall project consists of five phases:

- Phase 1: Evaluation design and IRB
- Phase 2: Full baseline
- Phase 3: Interim support between different activities
- Phase 4: Process evaluation
- Phase 5: Endline

In Table 4 below we highlight the different costs based on the phases of the project and cost category.

Table 4: Data collection costs, by phases and category

Phase	IDinsight staff costs	Temp staff costs	Non-staff costs	Total
1. Evaluation design and IRB	$112,948	$0	$3,000	**$115,948**
2. Baseline	$395,500	$227,496	$230,067	**$853,063**
3. Interim support between different activities	$47,444	$0	$0	**$47,444**
4. Process evaluation	$194,623	$7,020	$22,823	**$224,466**
5. Endline	$530,734	$510,664	$369,443	**$1,410,841**
Total	**$1,281,249**	**$745,180**	**$625,333**	**$2,651,762**

Additional notes about the budget:

- We used IDinsight standard daily rates to budget IDinsight staff time
- Since the contracting will be with IDinsight Inc. taxes will only be applicable on the data collection elements of the budget, and have been included in the figures given above
- Based on our last discussion with CHAI and GiveWell we have made adjustments to the team size and data collection costs to ensure:
    - We have sufficient time and resources for testing various recall methods during the baseline

    - Adjusted process evaluation data collection costs to ensure that we can cover it in both midline and endline based on further discussions about respondents to be covered at the different points of time
    - Included an additional cohort at endline to measure TPT completion
    - Included costs associated with the IsoScreen testing
- We are assuming that CHAI and GiveWell teammates are available for co-creation at the various steps and provide timely input and feedback
- We are assuming that data required for sampling purposes - such as Ni-Kshay or the roster of CHWs - are in the necessary format and include the necessary data points
- We will require IRB approval before we conduct data collection with permission letters from health officials

Figures 1 below provide a draft, high-level timeline for this evaluation. IDinsight will develop a more detailed work plan incorporating updated information on project design and implementation timelines during the first phase. Please note that based on our discussion with CHAI and GiveWell, IDinsight can move the process evaluation (either entirely or partly) to the endline to ensure that the qualitative components are assessed during the endline as well.

Figure 1: Overview of activities and timelines: Full baseline

Phase	Activity	2024								2025												2026												2027		
		May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar
Evaluation design and IRB	Theory of Change																																			
	Survey instruments development																																			
	IRB application																																			
	Pre-analysis plan																																			
Full Baseline	Pilot																																			
	Revision of instruments																																			
	Training																																			
	Data collection																																			
	Analysis and baseline report																																			
Interim support and consultations																																				
Process evaluation	Survey instruments development																																			
	Pilot																																			
	Revision of instruments																																			
	Training																																			
	Data collection																																			
	Analysis and PE report																																			
Interim support and consultations																																				
Endline	Review evaluation design																																			
	Survey instruments development																																			
	Pilot																																			
	Revision of instruments																																			
	Training																																			
	Data collection																																			
	Analysis and baseline report																																			

Team Structure

We have a strong India-based team with a deep understanding of India's evolving health landscape. This team is supported by a strong network of globally distributed in-house talent with diverse experience in process and impact evaluation.

Our team structure incorporates experience across the health sector in India across a variety of service areas such as evaluation design, strategy and advisory. The team is organized as follows:

Table 7: Team structure and utilization

Role	Utilization	Key Responsibilities
Director	20-30%	Strategic oversight and final approval of all project deliverables
Economist	25%	Advisory services and technical support to ensure project activities and deliverables meet highest technical standards
Manager / Senior Manager	50-100%	Overall project and stakeholder management
Associate / Senior Associate	100-200%	Responsible for conducting day-to-day activities, including preparing for and overseeing implementation of data collection
Field Manager	100-250%	Responsible for leading and implementing all data collection activities

Specific teammates will be selected based on availability and needs of that specific activity. Given the complexity of the study, IDinsight will involve additional staff members who specialize in large-scale data collection during the peak data collection periods.

To complement this team, we will leverage the expertise of an external, local TB researcher who can provide valuable input on our study protocol, survey instruments, data collection methods as well as the interpretation of the data. We recommend budgeting additional $15,000 for local researchers which would include an average of 20 hours per month for 15 months for a cost of $50/hour (inclusive of taxes).

These are sample profiles for the type of individuals who might guide and work on this project:

Dr. Neha Raykar – Neha is a Director at IDinsight, based in New Delhi. She works on building and managing projects in the areas of public health, nutrition, and adolescent wellbeing. Prior to

IDinsight, Neha was a Senior Consultant at Oxford Policy Management, where she led evaluations, operations research, and surveys studying health and nutrition in India, Ethiopia, and Nigeria. She was also the Lead Economist at the Public Health Foundation of India. She is a Senior Fellow of the Atlantic Fellows for Health Equity program based at George Washington University in Washington DC. Neha holds a PhD in Economics from the University of California, Riverside, an MA in Economics from Mumbai University, and a BA in Economics from St. Xavier's College, Mumbai.

Dr. Alison Connor – Alison Connor is the Director of Health at IDinsight, based in Nairobi, Kenya. Alison leads IDinsight's health strategy, identifying new opportunities in the health sector and providing expertise to health projects across all IDinsight offices. Previously, Alison was based in Zambia with IDinsight for 2.5 years. Her past projects include a community-based cluster RCT on childhood immunization financial incentives in rural Nigeria, a learning partnership with the government of Ethiopia, and several demand-focused evaluations with the government of Zambia in both health and social protection. Alison was also a founding member of IDinsight's technical team and has led internal training and professional development initiatives for the organization. Prior to joining IDinsight, Alison completed a community-based cohort study in rural Nepal for her dissertation, measuring the prevalence of misuse of uterotonics during labor and delivery and the adverse health consequences for mothers and newborns. Alison holds a BA in Government from Harvard University and a PhD in Global Disease Epidemiology and Control from Johns Hopkins Bloomberg School of Public Health.

Dr. Will Thompson – Will Thompson is an Economist at IDinsight based in New Delhi, India. Will oversees the technical design and analysis of impact evaluations, process evaluations, and design of complex sample surveys. He currently supports several projects, including the measurement of key performance indicators across India's Aspirational Districts, and IDinsight's agriculture learning partnership with the government of India. Will's sectoral focus is primarily in agriculture, and his research interests include capacity building and human/social capital formation, economic resilience, and behavioral economics. Prior to joining IDinsight, Will was a PhD student in the Department of Agricultural and Applied Economics at the University of Georgia, where his dissertation research focused on productive asset transfers to poor and marginalized women, as well as household resilience to shocks. While completing his PhD, Will led a large-scale randomized control trial of Heifer International's productive asset transfer programs in Nepal, evaluating a broad range of welfare outcomes. He also led or supported impact evaluations in Haiti and West Africa. Will holds a B.A. in History from Davidson College, and a PhD in Applied Economics from the University of Georgia.

Dr. Liza von Grafenstein - Liza von Grafenstein is an Economist at IDinsight, based in Delhi, India. Prior to IDinsight, Liza used machine learning to predict real-time child malnutrition outcomes at the Indian district level as an Innovative Metrics and Measures for Agriculture and Nutrition Actions (IMMANA) Career Development Fellow at Georg-August-University Göttingen, Germany, the Center of Development Studies (ZEF), Germany, and the South Asia office of the International Food Policy Research Institute (IFPRI) in Delhi, India. Liza completed her PhD in Economics at the Georg-August-University Göttingen. In addition, she holds an MPA with a focus on International

Development from Cornell University, US, and a BA in Political Science and South Asian Studies from Ruprecht Karl University of Heidelberg, Germany. She speaks English, Hindi, and German.

Abhishek Sharma – Abhishek Sharma is a Senior Manager at IDinsight, based in New Delhi, India. Abhishek is a development professional who has worked with Kudumbashree National Resource Organization (NRO), UNDP, World Bank, and Harvard's Evidence for Policy Design (EPoD). As a Senior Research and Capacity Building Manager at EPoD, he managed multiple research projects, including a large-scale evaluation in Chhattisgarh, India. Previously in Kudumbashree NRO, under the National Rural Livelihood Mission mandate, he helped share the best practices from Kerala regarding local governments, decentralized governance, and poverty eradication with other state governments. Abhishek holds a bachelor's degree in Electrical Engineering from the National Institute of Technology Hamirpur, Post Graduate Diploma in Rural Management (PGDRM) from Institute of Rural Management Anand (IRMA), and an MPA with a focus on International Development from Cornell University.

Shreya More - Shreya More is a Senior Associate on the DataDelta team at IDinsight, based in New Delhi, India. Shreya works with the Data on Demand team, on Niti Aayog's Aspirational Districts Program project that aims to improve administrative data quality. She has experience in indicator selection, measurement tool design, data collection, data quality processes, and analysis. She is also currently supporting the impact evaluation of an intervention that aims to embed socio-emotional learning in public education.Prior to IDinsight, Shreya worked as an analyst at PwC US Advisory, with clients in the technology and telecom sector. Shreya holds a Bachelor's degree in Economics from Shri Ram College of Commerce, Delhi University; and a Masters Degree in Economics from Delhi School of Economics.

Akshita Sharma - Akshita Sharma is a Senior Associate at IDinsight, based in India. Currently, she supports IDinsight's partnership with Project Sampoorna, a social-emotional learning (SEL) initiative led by the Government of Jharkhand, where she has led efforts to incorporate a participatory lens into research design and dissemination. Prior to IDinsight, Akshita worked with the Social and Political Research Foundation as a Research Associate. She also worked as a Research Trainee with Outline India and a Research Intern with Koan Advisory Group. Akshita holds a Master's in Economics from Dr. B.R Ambedkar University, Delhi, and a Bachelor's in Economics from Delhi University.

Syed Maqbool - Syed is a Field Manager Lead at IDinsight with extensive experience in conducting qualitative and quantitative surveys across various states. Syed holds 20+ years of experience in the socio-economic research platform. Prior to joining IDinsight, Syed worked as a Project Associate and Field Monitor at CMF, the Institute for Financial Management and Research, and MIT's Abdul Latif Jameel Poverty Action Lab (J-PAL). Syed holds a Master's degree in Sociology from Osmania University.

About IDinsight

IDinsight is a global advisory, data analytics, and research organization (501c3 non-profit) that helps development leaders generate and use evidence to make effective decisions and maximize the impact of their programs. We tailor a wide range of data and evidence tools, including randomized evaluations (RCTs) and qualitative research tools, to help decision-makers explore the questions they want answered, within the timeline and budget they have available. We work with governments, multilaterals, foundations, and innovative non-profit organizations in Asia and Africa across a range of sectors, including agriculture, health, nutrition, WASH, education, and social protection. We aim to improve millions of lives by transforming how the social sector innovates, learns, and improves.

IDinsight has well over 300 staff members in offices in seven countries: India, the Philippines, Kenya, Senegal, South Africa, the United States, and Zambia. In India we have worked across several states such as Uttar Pradesh, Bihar, Rajasthan, Punjab, Delhi, Telangana and Meghalaya.

Our services include large-scale primary data collection and analysis, impact evaluations (including RCTs and quasi-experimental evaluations), monitoring, process evaluations, program design, and embedded evaluation teams. We have conducted over 200 monitoring and evaluation engagements throughout Africa and Asia; we tailor each one to our clients' decision-making needs and deliver rapid, iterative, and high-quality insights to amplify their organizational impact. Past clients include UNICEF, UNDP, the World Bank, MCC, DFID, GIZ, the Bill & Melinda Gates Foundation, Big Win Philanthropies, GiveWell, GiveDirectly, Educate Girls, Indus Action and governments across Africa and Asia.

Relevant Experience

Below we provide a non-exhaustive list of projects with geographic or topical overlap with this opportunity:

Southern New Hampshire University (SNHU)-Kepler | Rwanda | Quasi Experimental Evaluation | 2013 - 2016

The SNHU Kepler quasi-experimental evaluation project aimed to assess the impact of the SNHU-Kepler program on students' academic and employment outcomes post-graduation. The study involved SNHU-Kepler graduates and a matched comparison group, with the evaluation focusing on various academic assessments and employment prospects. The graduates were assessed on computer literacy, cognitive skills (English, Math, and logic), and English reading and writing. The comparison group was selected based on previous evaluations of the SNHU-Kepler program, matching students on age, gender, previous academic achievement, family background, and baseline test scores for the 2014 cohort. The study reported results disaggregated by cohort and also by gender and baseline performance on academic assessments.

Educate Girls Development Impact Bond | India | Impact Evaluation and Machine Learning Support | 2015 - ongoing

IDinsight conducted an RCT of a remedial tutoring program in government schools in India. IDinsight designed and conducted a three-year RCT of Educate Girls' program in 500 schools in rural Rajasthan, surveying 10,000 students. The study used modified ASER assessments to test for foundational literacy (Hindi & English) and numeracy. The results of IDinsight's evaluation determined the outcome payments for the world's first Development Impact Bond. Since the completion of this RCT, IDinsight has worked with Educate Girls to inform where it should scale in a $100M-funded expansion of its program in India. IDinsight built and tested machine learning models to determine an appropriate method to predict which areas had concentrations of out-of-school girls. IDinsight's final prediction model allows Educate Girls to locate between 50 and 200% (depending on geographic scope) more out-of-school children for approximately the same operational cost per village.

RCT with Give Directly and Movva | Malawi | Impact Evaluation | 2022 - 2023

We partnered with GiveDirectly and Movva to assess the impact of SMS nudges on child school attendance and parental engagement. GiveDirectly's project funder wanted to have the intervention scaled to other parts of rural areas if it had an impact. We conducted a randomized controlled trial among 2333 GiveDirectly cash transfer recipients across households in 32 of 209 villages in Lilongwe. While we found no impact across all pre-specified indicators, 97% of recipients found messages to be either very useful or extremely useful, and significant information was shared between households. If GiveDirectly chose to scale up its program, we recommended (1) specific adjustments they should make to the intervention and (2) continued monitoring and evaluation of its implementation.

AgResults, a consortium governed by the American, Australian, British, and Canadian governments, BMGF, and the World Bank, is a $152mn initiative that uses pay-for-results (PfR) prize competitions to incentivize the private sector to fill agricultural market gaps and improve the lives of smallholder farmers. **IDinsight is evaluating AgResults projects in East Africa, Senegal, and Ethiopia.** IDinsight has worked to find the most rigorous evaluation design that meets the feasibility constraints of the various stakeholders. Each project involves a **mixed-method impact evaluation, a process evaluation, and a cost-effectiveness analysis**.

Evidence Action IFA | India | Process evaluation | 2023 - Ongoing

IDinsight is supporting Evidence Action in estimating coverage of India's Iron and Folic Acid (IFA) Supplementation strategy and estimating the effect of Evidence Action (EA)'s Technical Assistance (TA) on IFA coverage. The technical assistance aims to support program delivery and improve program coverage to reduce the prevalence of anemia for the target populations of i) preschool-age children aged 6-59 months, ii) government school-going children 5-19 years, and iii) out-of-school adolescent girls 10-19 years. IDinsight will conduct a deep coverage analysis using data from annual survey rounds collected by Evidence Action and a process evaluation across five states, assessing the improvement in programme delivery through EA's technical assistance across two project phases.

NITI Aayog's Aspirational Districts Program - Health and Nutrition | India | Learning Partnership | 2017-2020

IDinsight functioned as NITI Aayog's knowledge partner for the Prime Minister's flagship initiative, the Aspirational Districts Programme (ADP). The program reaches 250 million people living in 100+ of India's poorest districts. With support from the Bill and Melinda Gates Foundation (BMGF), IDinsight set up the Action-focused Measurement and Learning (AMAL) Unit at NITI Aayog, which comprised sector-specific learning teams that used behavioral and experimental methods to help the government design policies and evaluate what works to inform scale-up. To address poor maternal and child nutrition outcomes in India, the Government launched the National Nutrition Mission, POSHAN Abhiyan. This program promotes SBCC across various platforms to encourage nutrition-seeking behaviors, e.g., exclusive breastfeeding and complementary feeding. In collaboration with 8+ development partners, the Ministry of Women and Child Development (MWCD) and NITI Aayog, IDinsight designed and conducted a large-scale SBCC survey that measured the implementation of POSHAN Abhiyaan's SBCC interventions and addressed critical evidence gaps on the relative efficacy of nutrition-related SBCC content. We collected data from over 10,000 pregnant women across 27 Aspirational Districts, studying the reach of different platforms, the recall of different messages, knowledge levels, self-efficacy, and best practices. Our findings informed several decisions of MWCD to increase the uptake of nutrition-seeking behaviors of mothers and their young children across India. In particular, the MWCD increased its focus on interpersonal communication platforms such as home visits and community-based events, which had higher reach and recall, and de-invested in other platforms, increasing the cost-effectiveness of the national program.

New Incentives Impact Evaluation | Nigeria | Impact Evaluation | 2017 - 2022

IDinsight conducted a randomized evaluation of the New Incentives All Babies Are Equal Initiative's (NI-ABAE) conditional cash transfer program for routine immunizations in 2017 in northwest Nigeria. In 2021, New Incentives engaged IDinsight to design a measurement strategy to collect this information through periodic coverage monitoring surveys. Through its field staff, New Incentives now routinely collects information from households to monitor vaccination coverage in cohorts of areas it operates. In July 2022, New Incentives requested additional support from IDinsight to analyze this data according to the pre-specified analysis plan, conduct quality checks, assess the fidelity of the sampling protocol, and prepare the results from the first cohort of follow-up surveys.

Appendix A - Sample Size Calculations

Sample of active TB patients

Sample calculations for estimating impact of intervention on TPT initiation using 30-60 day cohort only

We estimate that we will require 78 patients per district in Bihar and 82 patients per district in UP for a total sample size of 5,636 at both baseline and endline to achieve the MDEs specified by CHAI and Givewell. Note that:

1. These figures assume that we will only use households of active patients indexed 30-60 days prior to surveying to estimate the impact of the program on TPT initiation. (For more information on what our MDE would be if we also used the 120-150 day cohort to measure TPT initiation see section below.)
2. These figures represent the total number of active patients to be sampled and not our final analytical sample (which will be significantly smaller as many patients won't live with children under 5 or respond to the survey).
3. If the share of patients who do not respond or who do not live with children under 5 is higher than anticipated we would increase our sample size to account for this.

The table below provides more information on the inputs used in these sample size calculations.

Parameter	Value	Explanation
Number of treatment and control districts	13 treatment and 13 control (Bihar); 22 treatment and 22 control (UP)	See main text
Intra-class correlation (prior to accounting for the baseline)	.09	To estimate ICC, we calculated the district-level ICC for whether a woman had met with a CHW worker in the previous 6 months using the most recent NFHS data. Note that the value of ICC that we use in the final power calculation command is $ICC * (1 - R_{dist}^{2})$. See below for more details
R_{dist}^{2}	.2	R_{dist}^{2} is the r squared from a regression of endline district means for the outcome variable on baseline district means for the

		outcome variable. With a repeated cross section design, the primary way a baseline improves precision is by reducing the cluster level variance at endline. (See equation 6 of Jacob, Zho, and Bloom.) We estimate R_{dist}^2 using DHS data from Pakistan. Pakistan had a DHS round in 2017-18 and a limited round in 2019. Districts in Pakistan are also similar in size to districts in India. After adjusting the sample size to match our final analytical sample size per district, R_{dist}^2 was .25 (having >= 4 ANC visits), .11 (total ANC visits) and .14 (been visited by CHW in previous 12 months). As this is a key input and we have little prior information on which to base the input estimate on, we have included sensitivity analysis which looks at how our MDEs would change if this input was higher or lower than we anticipate in appendix D.
Minimum Detectable Effect	15 ppts (Bihar); 14 ppts (UP)	Values are based on the memo shared by CHAI. Based on our understanding, using these values for MDE ensures that if the true impact of the program is roughly what CHAI anticipates, we will probably (>80%) find the effect in each state statistically significant and the point estimate will be larger than the threshold value in order for the program to be cost effective.
Baseline levels of primary outcome	13% (Bihar); 36% (UP)	Values are based on memo shared by CHAI.
Non-response	30%	Value based on IDinsight experience locating pregnant and lactating women using ASHA registries.
Share of index patients living with at least one child under 5	40%	According to NFHS data, 38% and 40% of households in UP and Bihar respectively have under 5 kids. Note that some

		households may have more than one kid under 5 but outcomes among kids in the same household are likely highly correlated and thus we consider them 1 data point for the purpose of estimating power.
Among hhs with children under 5, total children under 5	1	
Power	.8	Standard value used.
Alpha	.05	Standard value used.

We provide the exact Stata commands below for reference. (Note that after calculating the sample size using the code below we must adjust for non-response and for the proportion of patients living with children under 5.)

```
# BIHAR power calcs
power twoproportions .13 .28, k1(13) k2(13) rho(.072)
# UP power calcs
power twoproportions .36 .50, k1(22) k2(22) rho(.072)
```

Note that while the MDE for Bihar and UP is 14 and 15 percentage points respectively, the MDE for both states combined is approximately 9.74 percentage points. To estimate the combined sample size we divided the MDE by 2^0.5 to account for the effective doubling of sample size and multiplied by 95% to account for the increased degrees of freedom due to the increase in the number of clusters.

Sample calculations for estimating impact of intervention on TPT completion using 120-150 day cohort only

If we sample an additional 5636 active TB patients at endline, we estimate that our MDE for overall completion (i.e. the overall share of eligible children under 5 who complete TPT) would be 13.5 ppts in Bihar and 13.1 ppts in UP and our MDE for conditional completion (i.e. the share of children who complete TPT among those who initiate TPT) for both states combined would be 13.8 ppts. We note that the estimate of the impact of the program on conditional completion is not a true experimental estimate since we are conditioning on TPT initiation which is affected by treatment. If we use this additional sample to also estimate the impact of this program on TPT initiation then our revised MDEs for TPT initiation would be 13.3 ppts in Bihar and 12.6 ppts in UP.

The table below provides more information on the inputs (beyond those noted above) use for these sample size calculations.

Parameter	Value	Explanation
Conditional completion in control districts	70%	This is the share of children who complete TPT among those who initiate TPT in control districts. Value based on CHAI inputs.

We provide the exact Stata commands below for reference.

```
# UP - overall completion
power twoproportions .252 , k1(22) k2(22) m1(22) m2(22)
rho(.072) power(.8)
# Bihar - overall completion
power twoproportions .091 , k1(13) k2(13) m1(23) m2(23)
rho(.072) power(.8)
# Both states combined - conditional completion
power twoproportions .7, k1(35) k2(35) m1(6) m2(6) rho(.09)
power(.8)
# UP - revised MDE for TPT initiation
power twoproportions .36 , k1(22) k2(22) m1(46) m2(46)
rho(.072) power(.8)
# Bihar - revised MDE for TPT initiation
power twoproportions .13 , k1(13) k2(13) m1(44) m2(44)
rho(.072) power(.8)
```

Surveys of caregivers of children who have initiated TPT for Isoscreen verification

We will sample 1071 children who are reported as having initiated TPT in Ni-Kshay in treatment districts only, survey caregivers of these children approximately 10-12 weeks after TPT initiation, and administer Isoscreen to children whose caregivers report that they administered TPT to the children in the previous 48 hours. This will allow us to estimate TPT adherence among children whose caregivers claim adherence to plus or minus 6.4 percentage points (95% CI).

We note that, depending on the exact sampling strategy, there may be some overlap between children in this sample and children in the main sample of households of active TB patients.

According to CHAI, there are approximately 400,000 active TB patients across Bihar and UP. If we assume 40% of TB patients live with children under 5, 1 child under 5 per household, 42% of

children living with an indexed patient initiate TPT (the same assumptions as used above) and a one month sampling window, we should have 400,000*(35/(38+75))*.4*.42/12=1734 children in our treatment districts who have initiated TPT over a one month window. Thus, if we sample 1071 children in a one month window, we will effectively sample roughly 62% of all children who have initiated TPT. If we find that a one month window is too narrow, we can increase the window by using a longer survey period.

The table below provides information the additional assumptions (beyond those already made above) used in this calculation:

Parameter	Value	Explanation
Reported conditional completion	80%	We assume 70% conditional in control districts (see above) and a 10 ppt increase in treatment districts
Among children whose caregivers report administering TPT, share whose caregivers report having administered TPT in previous 48 hours	50%	If this were completely random we would expect only 2/7 of children to have taken TPT in the previous 48 hours since the medication is administered weekly. We think that we can increase the chances of a caregiver having recently administered TPT by conducting surveys just after the weekend when caregivers are more likely to administer TPT. We may also call caregivers ahead of time.
Actual TPT adherence among those with claimed adherence	80%	
Isoscreen sensitivity / specificity	93% / 99%	Taken from here https://erj.ersjournals.com/content/erj/early/2014/01/16/09031936.00132613.full.pdf

To calculate the sample size for the validation sample, we use the following formula which is the basic formula for a clustered sample but with an additional correction for test accuracy adopted from Rogan and Gladen (1978).

$$SE = \sqrt{\frac{r*(1-r)*DE}{N}};\ DE = \frac{1+\rho(m-1)}{(se+sp-1)^2}$$

Where se is sensitivity and sp is specificity.

This suggests that we need a final sample of 300 children whose caregivers a) respond to the survey, b) claim TPT adherence, and c) report administering TPT in the previous 48 hours. This leads to a total sample of 300/(.5*.8*.7)=1071.

CHW sample

We estimate that we will require 8 CHWs per district in Bihar and 4 per district in UP for a total of 384 CHWs across both states.

To estimate the sample size of CHWs required for the CHW survey we assumed that the primary outcome would be the number of hours per day CHWs spend on non-program tasks. To perform the power calculations, we used the same parameters as used for the main sample calculations for number of treatment and control districts, ICC, power, and alpha. We explain the values used for other parameters below.

Parameter	Value	Explanation
Minimum Detectable Effect	.5 hours	Based on our experience, ASHA workers typically work about 5 hours a day. If TB-HVs and PPSAs work similar hours this would represent an approximately 10% change in workload.
Standard deviation of	1 hour	Based on IDinsight staff experience.
Non-response	0%	In our experience, response rates for CHWs is very high

We provide the exact Stata commands below for reference:

```
* power calculation for Bihar
power twomeans 5 4.5, sd(1) k1(13) k2(13) rho(.09)
* power calculation for UP
power twomeans 5 4.5, sd(1) k1(22) k2(22) rho(.09)
```

Facility survey

We estimate that we will require 2 public health facilities per district for a total of 70 facilities. These power calcs assume that our goal is to detect an overall (rather than state specific) difference in the share of children under 5 who start TB treatment among those who are diagnosed with TB. We note that a) the HCM program may affect denominator (i.e. the total number of children under 5 diagnosed with TB) rendering interpretation of this effect difficult and b) inputs to the power calculations are somewhat speculative due to the lack of facility-level patient data to base these calculations on.

Parameter	Value	Explanation
Minimum Detectable Effect	5 ppts	Based on intuition
Current share of children under 5 who start TB treatment among those who are diagnosed with TB	90%	Based on figures from Subbaraman et al (2016)
Standard deviation	10 ppts	Based on intuition.
ICC	.09	

```
power twomeans .9 .95, sd(.1) k1(35) k2(35) rho(.09)
```

Appendix B - Statistical note on a full vs partial baseline

In earlier discussions, we discussed the possibility of conducting a partial, rather than a full baseline. In this appendix, we discuss the advantages and disadvantages of a full vs partial baseline. In earlier versions of the concept note we provided sample size calculations for the partial baseline option.

If we perform a full baseline our final regression specification will probably look something like this:

$$y_{d,i} = \alpha + \beta_1 T_d + \beta_2 \bar{ypre}_d + \varepsilon_d + \varepsilon_i$$

Where $y_{d,i}$ is the outcome for child i in district d at endline, T_d is a binary variable equal to 1 if district d was in the treatment and 0 otherwise, and $\bar{ypre}_d$ is the mean primary outcome for district d at baseline. The primary benefit of a repeated cross section baseline like the one we are considering is that it reduces the variation in ε_d and thus marginally increases the precision of our estimates. More specifically, if the R squared from a regression of district level endline average outcomes on district level baseline average outcomes is X then we effectively reduce the ICC by 1-X. (See equation 6 of Jacob, Zho, and Bloom.) With two rounds of panel data (i.e. the baseline and endline include the same households), a baseline can provide a valuable safeguard against attrition / different selection (since we have baseline data on both attriters and those who remain in the sample) but with a repeated cross section a baseline doesn't help in this regard.

While a repeated cross section baseline would increase the precision of our estimates, we estimate that we could increase precision more cost effectively by simply increasing the sample. We should caveat that these calculations are highly dependent on our estimate of ICC and the R squared of endline district means on baseline district means, two parameters we have high uncertainty about. In addition, a small additional benefit of conducting a baseline is that it would provide us with a better estimate of our ICC. If we find that our ICC is significantly higher than we anticipated, we could increase our endline sample size to compensate. (Though, even if we conduct a full baseline, we would only be able to estimate ICC to within about plus or minus .04 with a 95% confidence interval and if ICC is above .11 or so the sample size becomes too large to be practical.)

Appendix C - Sample options shared with CHAI and GiveWell in March 2024

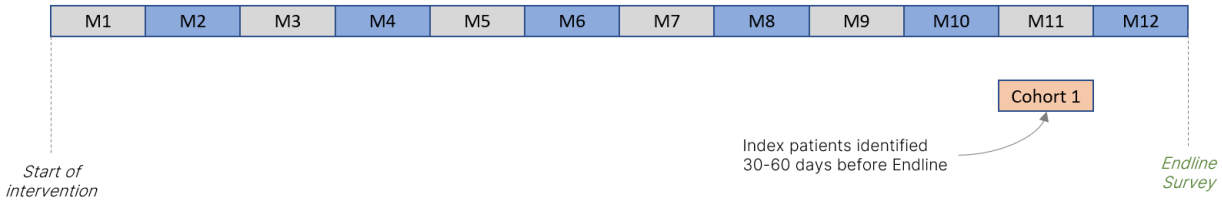
Context

On 28 March 2024, CHAI, GiveWell, and IDinsight discussed different options for constructing an evaluation sample. Based on the options shared below, CHAI And GiveWell recommended that we go ahead with Options 4 and 5. Below we share the original options for reference.

Note: Across all options, we assume that:

1. 30 days is the minimum amount of time that should pass between when one was identified as an index patient and the timing of the endline survey. This would allow time for the frontline worker to conduct the initial household visit, initiate TPT, refer HH members to a health facility, and for them to go to that health facility. *It is possible that this process would take more time in the control arm than in the treatment arm.*
2. A cohort would include index patients identified over a 1 month / 30 day timeframe.
3. A single cohort (i.e. a month of patients) in a single district would include approximately 300 patients, significantly larger than the ~110 patients we estimate that we need per district.[12]

OPTION 1: Conduct Endline on 1 Cohort of Index Patients Identified 30-60 Days Prior



Description

- IDinsight would conduct one endline using a cohort of index patients identified 30-60 days prior.

Advantages

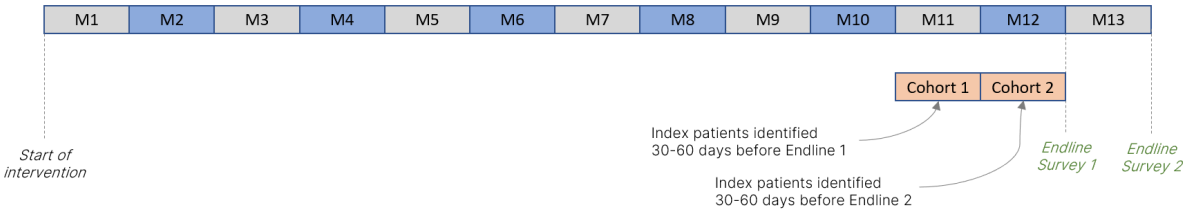
[12] According to previous communication, CHAI estimates that there are approximately 411,000 active TB patients each year in UP and Bihar combined. Dividing this figure by the number of districts in UP and Bihar combined (113) and 12 months yields a little over 300 patients per month.

- This option would minimize the recall period between the endline and a) frontline worker visits; b) initiation of TPT; and c) HH member follow-up at the health facility

Limitations

- This option would not allow us to estimate treatment completion, since it does not allow for enough time to pass for U5s to complete the 90 day TPT regimen before the endline.
- If control households are likely to take longer to a) be visited by a frontline worker and/or b) follow-up at a health facility and/or c) initiate an U5 on TPT, then this option may overestimate the increase in desired outcomes, since the apparent increase may simply be improved *timeliness*. (I.e., in an extreme case, if the intervention does not change the percentage of household members who go to a health facility for screen or the percentage of U5s who initiate on TPT but does shorten the length of time it takes for these outcomes to occur, the endline may suggest that the program is very impactful, when it actually has not increased the percentage of U5s who initiate TPT.)

OPTION 2: Conduct Rolling Endline on 2 Cohorts of Index Patients Identified 30-60 Days Prior[13]



Description

- IDinsight would sample patients at the end of month 11 and survey these patients at the end of month 12. IDinsight would sample patients again at the end of month 12 and survey these patients at the end of month 13.[14]

[13] This option could also be used for Option 4, if Option 4 will equally not yield enough index patients to meet our sample size requirements with 1 cohort. (It could also be used for Option 3 but this would be even more costly.)

[14] If we obtain Nikshay lists, sample patients, and immediately share the list of sampled patients with field teams for surveying we would only have to sample patients once. In practice, we anticipate that it will take at least a couple of weeks to obtain lists, clean them, and sample patients.

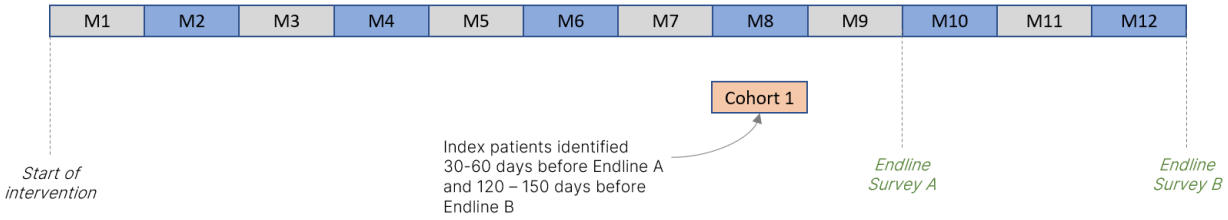
Advantages

- This option would also minimize the recall period between endline and a) frontline worker visits; b) initiation of TPT; and c) HH member follow-up at the health facility
- If we do not think we would get sufficient sample with only 1 cohort, this option would allow us to use a 2nd cohort without increasing the recall period. **Therefore, this option only makes sense if we do not think that we will have sufficient sample with 1 cohort. Based on the numbers shared by CHAI, we *do* think that we will have sufficient sample with 1 cohort, therefore we do not recommend this option at this time. This can be confirmed at baseline.**

Disadvantages

- This option will add significant cost.
- The limitations from Option 1 still apply.

OPTION 3: Conduct 2 Endlines at Different Time Points on 1 Cohort of Index Patients



Description

- IDinsight would conduct one endline 30-60 days after index patients are identified and then another endline on the same cohort 90 days after the first endline.
- There are two variations to this option: A) Conduct 2nd endline on full sample; B) Conduct 2nd endline only with those who had initiated at the 1st endline.

Advantages

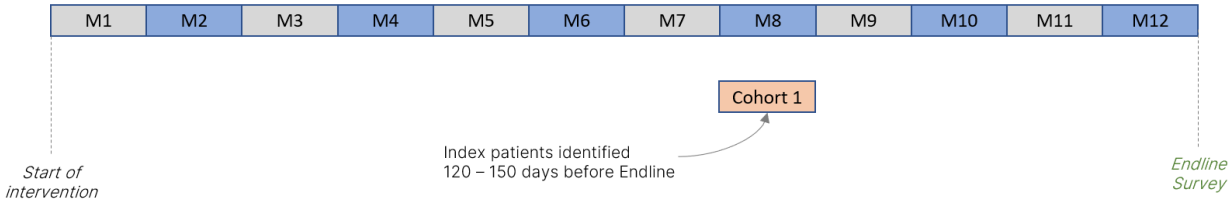
- This would allow us to estimate TPT initiation while minimizing the recall period, but then would also allow us to estimate coverage with the second endline.
- While both variations are more expensive than Option 1, Variation B would be less expensive than Variation A.
- Variation A would allow for more time to measure initiation of TPT among control households, mitigating the limitation for Options 1 and 2.

Disadvantages

- Asking about TPT initiation in the first survey may influence people's behavior if they think people are checking up on their compliance.
- For ethical reasons, surveyors will share information at endline 1 about the importance of following up, initiating TPT for U5s, etc, which could influence the outcome at the 2nd endline.
- Variation B would have similar limitations to estimating initiation as Option 1.

Because this option can itself affect the outcome, we do not recommend this option.

OPTION 4: Conduct Endline on 1 Cohort of Index Patients Identified 120-150 Days Prior



Description

- IDinsight would conduct one endline using a cohort of index patients identified 120-150 days prior.

Advantages

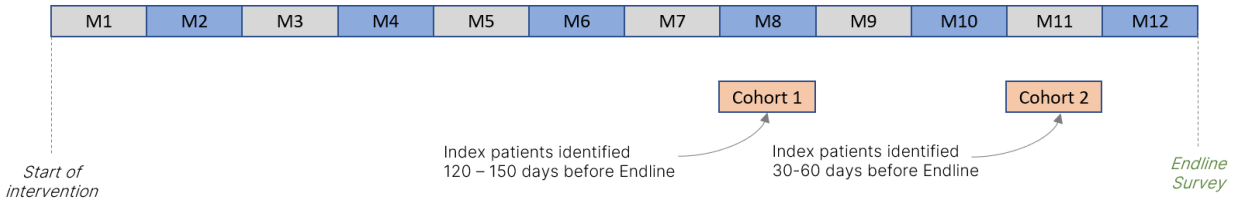
- This option would allow us to estimate treatment coverage at minimal cost.

Disadvantages

- Recall for earlier outcomes such as whether a frontline worker visited, what they did / said during that visit, etc may be poor that far out. Given the stigma associated with TB in this context, **we believe that the recall of a frontline worker visiting to speak about TB, diagnosing someone with TB, and initiating TPT would be memorable, and therefore, recall would still be reliable at the suggested timeframe.** However, it would be infeasible to get accurate estimates of *when* the visit happened.
- As with the concerns in Option 1, we may not be able to measure full completion in the control arm if they initiate treatment later than the treatment group on average.

- If the control group is still using the 9 month TPT regimen, we cannot measure completion (nor would it make sense to compare this with completion of a 3 month regimen).

OPTION 5: Conduct Endline on 2 Cohorts: 1 Cohort of Index Patients Identified 120-150 Days Prior and 1 Cohort of Index Patients Identified 30-60 Days Prior



Description

- IDinsight would conduct one endline on separate cohorts: one cohort of index patients identified 120-150 days prior and one 30-60 days prior. The cohort of index patients identified 30-60 days prior would be the main cohort, since U5 initiation of TPT is the primary outcome.
- There are a few variations within this Option: A) Cohort 1 could be a smaller number of treatment and control patients (i.e., this option need not double the sample size); B) Cohort 1 could be done among the treatment group only if one only needs an estimate of coverage among the treatment group, rather than a comparison of coverage between the treatment and control groups.

Advantages

- This option would allow us to estimate both initiation and coverage.
- The advantage of this option over Option 4 is that it shortens the recall period for measuring initiation (and other outcomes related to the frontline worker visit) while still allowing us to measure coverage.

Disadvantages

- If we wish to power the study to detect effects on coverage we would require a larger sample size (and thus larger budget) than option 4 so we would only recommend this option if the longer recall is likely to meaningfully affect the quality of our primary outcome.

- This option has the same limitations as Option 1 (in terms of timeliness of initiation) and Option 4 (in terms of whether it will allow for enough time for completion in the control arm).

Appendix D - Full baseline sensitivity analysis

Our power calculations for the option in which we include a full baseline rely on estimates of R_{dist}^2, the r squared from a regression of endline district means for the outcome variable on baseline district means for the outcome variable. In the power calculations above, we assume that $R_{dist}^2 = .2$ but this estimate is based solely on data from Pakistan and thus we have low confidence in this estimate. We show the MDE for Bihar and UP respectively if we use a full baseline with the sample sizes given above but R_{dist}^2 is lower than .2.

Overall, we think that the state-specific MDEs are still reasonable even with much lower values for R_{dist}^2.

R_{dist}^2	Bihar MDE (ppts)	UP MDE (ppts)
0	16.14	15.02
.05	15.84	14.77
.1	15.53	14.51
.15	15.22	14.24
.2 (current assumption)	15	14