New Incentives
Measurement Options
31 March 2017

Executive Summary

To support a GiveWell decision on New Incentives' top charity status, IDinsight is designing an impact evaluation of New Incentives' program. The evaluation will be a randomized controlled trial (RCT) and take place in North West Nigeria, a region where New Incentives believes it will have substantial opportunities to scale. We recommend randomizing at the clinic level with approximately 120 treatment clinics divided into treatment and control clinics. Because the cost of implementation and data collection at this many clinics will be substantial, IDinsight has outlined a number of activities to prepare for the RCT. These steps will allow GiveWell to make an informed funding decision about the study. To start finalizing details related to the ultimate study design and shape our preparatory activities, IDinsight has also outlined different design options for the RCT for GiveWell to provide input.

IDinsight Recommendations:

- **Matching Study:** Forgo the matching study to focus resources on data validation and piloting as well as assisting New Incentives with a small study to set the measles incentive.
- **RCT Design:** Use vaccine coverage as the outcome variable and use household coverage surveys to measure the difference in coverage between treatment and control at endline, and use administrative measures of coverage for baseline measurement.
- **Study Size:** 60 clinics will allow the RCT to detect around an 11% increase in measles coverage and a 10% increase in PENTA3 coverage at 80% power and 5% precision. (Details in Appendix 2).

Key Areas for GiveWell Input:

- **Matching Study:** Given the limitations, is a matching study prior to the RCT worthwhile?
- **RCT Design:** What role should administrative data play in the RCT? Should there be a baseline coverage survey? Should we measure mortality?
- **Study Size:** What effect size should the study be powered to detect? What level of power is necessary to inform GiveWell's cost effectiveness estimates?

Preparatory Activities for the RCT

There are four main activities we plan to undertake to prepare for the RCT and validate evaluation design choices. We will help finalize the implementation model, understand how far mothers travel for incentives, validate administrative data, pilot administrative coverage estimates, and pilot techniques for the household coverage survey.

Fieldwork for the preparatory phase are anticipated to take place in May and June. The approximate cost of the fieldwork associated with these activities is 28,000-36,000 USD as outlined in the budgets shared along with this document. Desk work associated with these activities will begin immediately.

Assist New Incentives to finalize their implementation model

New Incentives plans to use the pilot model for scale up. The overall structure and operations of the model will remain unchanged. New Incentives (NI) will make three important implementation decisions prior to the RCT:

- Decide which two states in the North West where they will operate.
    - IDinsight will help analyze administrative data and other key factors and provide an additional perspective.

- Decide the cash amount for the measles incentive.
    - IDinsight will support a small randomized evaluation to test different measles incentives[1] (e.g. 2000 Naira versus 4000 Naira). Other vaccination incentive amounts will remain the same as used during the pilot study (500 Naira).
- Decide if other components should be added to supplement NI's current model (e.g. family planning or Vitamin A).

Understand the distances mothers travel to clinics offering incentives

In order to ensure control clinics are sufficiently separated from treatment clinics, it is important to understand how far mothers are willing to travel for incentives. Furthermore, the distance mothers are willing to travel for incentives will define how we approach measuring spillovers when estimating administrative coverage. Details on all available administrative data sources can be found in Appendix 1.

We will use data from the pilot regions to estimate the distance mothers travel. These clinics are located outside the North West where the study will take place. To gain some insights into the North West, we will supplement this analysis with data from a UNICEF cash transfer program that offers incentives for mothers at clinics.

For the actual analysis of mother travel patterns, we will rely on triangulating from different data sources. There are challenges in using each type of data. A few examples are highlighted below:

1. Follow-up addresses are included on the child health register. This data is inconsistently filled out, though, and only goes to the overall neighborhood level (instead of street address). This will make the information difficult, if not impossible, to use.
2. Nurses track the number of doses of each vaccine given at the clinic on a vaccination day. Administrative data on doses of vaccine given can be used to look for declines in neighboring clinics during the period incentives were offered.[2] Because changes in clinic volumes could be due to a number of other factors, it will be difficult to identify a clear radius of affected clinics.
3. We will conduct interviews to ask mothers how they reached the clinic how much they spent.[3] This could be added as a routine question NI staff ask or involve IDinsight enumerators conducting exist interviews. Transport cost is not a perfect proxy for distance as the data is self-reported. Although mothers recalling what they spent that day to reach the clinic may not be difficult, mothers may lie if they think revealing they are from far away will affect their eligibility.

Validate administrative population data and pilot administrative coverage estimates.

Administrative coverage estimates are a potential alternative to more costly household coverage surveys in the RCT. The administrative coverage estimates will draw on micro-census and DVDMT data (see Appendix 1).[4]

We propose using micro-census data as the population denominator for administrative coverage estimates, given availability. The intensity of validation depends on how we plan to use administrative coverage in the

[1] New Incentive will randomly select mothers due for measles in the next three months to call and offer a higher incentive to return for their measles vaccination other mothers will simply receive a reminder call. There will also be a pure control group with no reminder call so the effect of the reminders can be measured.
[2] Neighboring clinics can be identified using a pre-existing geodatabase of Nigerian health facilities, and confirmed by discussions with LGA officials.
[3] If they walked or cycled we can ask how much time they spent, but self-reported durations tend to be less reliable than self-reported costs.
[4] We have not confirmed recent micro-census data is available for the final selected states.

evaluation. The goal of the validation activities is to facilitate a more informed discussion of the administrative baseline option. If the micro-census data isn't available for one of the selected states, we will discuss as a team next steps.

If discussion with GiveWell and New Incentives rule out using administrative coverage for the baseline but are still considering it for a midline, then micro-census validation activities can be skipped. Once the baseline coverage survey is conducted, we will have ample population data to validate the micro-census numbers against.

If we do move forward with using this data for a baseline, one method for verification would be to check if the populations of settlements around the New Incentives learning sites match the micro-census data. We could potentially pay New Incentives field officers to do this and combine the census with mobilization for the program to accelerate the volume ramp-up process at learning sites.[5] We will also have population data for settlements where we pilot the coverage survey.

A lighter-touch option for validating the data is to see if the number of houses in a settlement roughly matches the reported number of houses in the micro-census data. The number of houses could be checked relatively quickly by field staff counting the number of occupied structures in a settlement or estimated remotely using satellite imagery.

To validate the numerator for administrative coverage calculations we will carefully review the WHO led data quality reports (DQS) which investigate whether tally sheet data is being aggregated properly, and also use the DVD-MT data to compare vials distributed by the LGA to a clinic against the reported number of vaccinations given. It should be noted that availability and quality of DQS and DVD-MT is not ideal.

We will select some pilot clinics and some clinics in the North West to construct administrative coverage estimates. For the pilot clinics we will use data from neighboring clinics to discount increases in the number of vaccines given by decreases in vaccines given in neighboring clinics to account for the fact that clinics will draw in mothers from outside their official catchment areas.

Pilot coverage surveys

We will select two or three clinics across the two North West states selected for scale up to run a coverage survey. This will involve a census of every child in each settlement followed by a more detailed survey determining the vaccination status of a random sample of children. We will select at least one clinic with a large catchment area including far-away settlements to ensure the feasibility of conducting a census across a wide area. To determine vaccination status, we will use child immunization cards, mothers' recall of vaccination history, and cross-referencing with clinic child health registers. Blood testing may be an option for validating PENTA vaccinations, and we will explore the cost and feasibility of incorporating this into coverage surveys as we develop the study design over the coming months.[6]

[5] The idea behind this one-time activity is to ensure volumes in learning sites more quickly simulates the expected volume during the RCT. Without mobilization, New Incentives estimates it takes about three months for volumes to level off. If house to house mobilization activities were conducted in the RCT sites new Incentives would need to commit to doing regular house to house mobilizations at scale which is likely unsustainable hence a ramp-up period has been designed into the study.

[6] Unfortunately, blood testing is only an option for PENTA because the blood test for measles cannot distinguish between vaccine derived immunity and immunity from exposure to the wild virus.

New Incentives
Measurement Options
31 March 2017

Key Decision: Should IDinsight conduct a matching study to evaluate the impact in pilot clinics?

Based on recent conversations with GiveWell, IDinsight recommends to not conduct the matching study. While the matching study could give a directional estimate of impact, it does not contribute substantially to RCT preparations. The table below outlines the advantages of conducting a matching study as well as the limitations.

Advantages	Limitations

• The study will offer a directional indication of program's impact on coverage. (Impact estimates from pilot admin data focus on retention and are pre-post)
• Synergies between matching study and other preparatory activities
  ◦ The more clinic-specific estimates of administrative coverage we create the more we will learn about the technique.
  ◦ Collecting more administrative data for clinics neighboring pilot clinics will help us learn about the distances mothers travel.
	• Data from only nine treatment clinics will make the study too low-powered to refine GiveWell's cost-effectiveness model.
	• Results in North Central and South South may not be generalizable to the North West.
	• Administrative data may be too unreliable to produce meaningful results.
	• Assumptions about mothers' travel or the absence of other programs in neighboring clinics may be wrong.

Conducting a RCT

The RCT will involve baseline measurement, selecting treatment and control clinics, midline measurement, and endline measurement. We will also register the trial with a relevant research registry. This step will involve submitting a pre-analysis plan.

Common features across most design options

- Timeline:
  - Spring 2017: state selection
  - Summer 2017: clinic selection, register trial, and IRB approval
  - Fall 2017: baseline and start of implementation.
  - March 2018: number of mothers arriving at clinics reaches steady state
  - Late Summer 2018: administrative midline with PENTA3 and limited measles results to potentially inform 2018 top charity status
  - Spring 2019: endline survey and final report
- Cohort study:
  - Rather than following individual babies from baseline to endline, we would examine coverage for two distinct cohorts of children at baseline and endline. Each cohort will be made of up children who should have finished their vaccination series, roughly aged 1 year.
- Pre-analysis plan:

- Key outcome: There are a number of valid potential outcomes. Our current thinking is to construct a weighted average of the vaccination coverage for each vaccine based on modeled impact on lives saved, but this might evolve as we learn more.
    - Other potential outcomes (*for discussion*): Standard routine immunization indicators in the public health literature such as PENTA3 and Measles coverage or the percentage of children who completed the routine immunization course.
- Secondary outcomes: attitudes towards vaccination, disease prevalence, mortality.
    - Measuring these outcomes would require additional survey modules. Adding these modules must be balanced against keeping the survey relatively short.

- Possible additional data collection to influence the cost effectiveness model
    - Collecting data to see if increased clinic visits have other health impacts.
        - Other diseases: malaria, diarrhea, and pneumonia treatment, family planning
        - Nutrition: does growth monitoring lead to improved nutritional status?
    - Collecting data to understand the prevalence and effectiveness of other supplementary immunization efforts
        - A household coverage survey will partially address this by asking mothers where their babies received a vaccine, but mothers may be unable to distinguish clinics, campaigns, and special outreaches.

Selecting Clinics

While New Incentives plans to eventually scale-up to groups of neighboring clinics for operational reasons, selecting well-spaced individual clinics for the RCT will maximize power relative to cost. IDinsight will map clinics in the states where the RCT is conducted using an existing geodatabase of Nigerian health facilities, and develop a list of clinics sufficiently far apart so as to avoid mothers traveling between them (spillover). The distance required will be determined by the analysis conducted during the preparatory phase as described above. Guided by IDinsight's mapping activities, New Incentives will screen clinics to ensure the program can be feasibly run at identified well-spaced sites. IDinsight will randomly select treatment clinics from the resulting list of clinics. Some areas may need to be excluded altogether to avoid overlapping the study with similar programs such as the UNICEF cash transfer program.

Key Decisions: What role should administrative data play in the RCT? Should we measure mortality? Is individual randomization still an option?

The tables below presents several design options for the RCT. IDinsight's current recommendation is to use administrative data for the baseline and conduct a coverage survey for the endline. Note that the recommendation for the administrative baseline is conditional on the data being validated during the preparatory phase.

Unless otherwise noted, all options would require a similar number of treatment and control clinics. For options 1-4 differences from the previous option are highlighted in bold. For options 5-6 bolding simply indicates key aspects of the option.

A breakdown of the cost of the measurement activities in options 1-4 are outlined in the budgets shared along with this document.

Option 1: Coverage Survey and Census Baseline and Endline

Summary	• At baseline and endline conduct a household census to find children in an age cohort expected to be influenced by the program. • Sample 40 babies to determine vaccine status by triangulating data from child health cards, self-reported histories, and clinic records.[7][8] • Stratify random selection into treatment and control using baseline coverage. • Compare coverage in treatment and control clinics at endline using baseline coverage as a control variable in the final analysis.
Assumptions	• GiveWell's modeled relationship between coverage and mortality is accurate. • Mothers from control clinics do not travel to treatment clinics for vaccinations. • Treatment and control groups balanced on any variables beyond the few used for stratification.
Budget	• Approximately $350,000 each for baseline and endline coverage surveys.
Operational Risks	• A large discrepancy between the number of self-reported vaccinations and the number of vaccinations that can be verified using the cards and clinic records may make the results hard to interpret. • The logistics of managing large survey teams may be challenging. • A change in the security situation may make field team management difficult at endline • If an immunization campaign occurs immediately prior to endline will report one set of results including campaign and another adjusting for it by discounting vaccinations given during the time period of the campaign.[9]

Option 2: Coverage Survey Baseline with Census and Endline Sample Survey

Summary	• At baseline conduct a household census to find children in an age cohort expected to be influenced by the program, **and babies who will be in that age cohort by endline.** • Sample 40 babies to determine vaccine status by triangulating data from child health cards, self-reported histories, and clinic records. • Stratify random selection into treatment and control using baseline coverage. • Compare coverage **among babies sampled at baseline who can be found at endline** in treatment and control clinics using baseline coverage as a control variable.
Assumptions	• GiveWell's modeled relationship between coverage and mortality is accurate. • Mothers from control clinics do not travel to treatment clinics for vaccinations. • Treatment and control groups balanced on any variables beyond the few used for stratification. • **Babies identified at baseline that can't be found at endline do not differ meaningfully from babies included in the study**

[7] See Appendix 2 for details on power calculations used to determine 40 babies per clinic optimizes power.
[8] As is common in the literature, we will report results based on coverage estimates derived from self-reported data alone and self-reported data verified by administrative records.
[9] Mother's self-reported dates of vaccination can be made more accurate by using major community events to anchor responses

Budget	• Approximately $350,000 for baseline. • **Skipping the census at endline would save about $150,000 dollars at endline.**
Operational Risks	• A large discrepancy between self-reported vaccinations and verified vaccinations. • The logistics of managing large survey teams may be challenging. • A change in the security situation may make field team management difficult. • An immunization campaign occurring immediately prior to endline. • **Endline delays could result in babies identified during the baseline being too old for their vaccination status to be easily determined (mothers forget details of vaccinations and administrative records are archived or destroyed).**

Option 3: Administrative Baseline and Coverage Survey Endline

Summary	• **Generate administrative coverage estimates to stratify the random selection of clinics into treatment and control groups.** • At endline conduct a household census to find children in an age cohort expected to be influenced by the program. • Sample 40 babies to determine vaccine status by triangulating data from child health cards, self-reported histories, and clinic records. • Compare coverage in treatment and control clinics at endline using **baseline administrative coverage estimates** as a control variable.
Assumptions	• GiveWell's modeled relationship between coverage and mortality is accurate. • Mothers from control clinics do not travel to treatment clinics for vaccinations. • **Randomization stratified using administrative coverage estimates led to a balanced sample i.e. average coverage across the 60 control clinics without the program would have been approximately the same as average coverage across the 60 treatment clinics at endline (we will be verifying administrative data to add confidence to this assumption).** • **Administrative data is accurate enough to group the clinics (e.g. low, medium, and high baseline coverage) for sub-group analysis.**
Budget	• Approximately $350,000 for the endline. • **Approximately $30,000 for baseline, and additional costs associated with more intensive verification of administrative data sources.**

Operational Risks	• A large discrepancy between self-reported vaccinations and verified vaccinations.
	• The logistics of managing large survey teams may be challenging.
	• A change in the security situation may make field team management difficult at endline.
	• An immunization campaign occurring immediately prior to endline.
	• **Administrative data concerns:**
	  ○ DVD-MT data could be inaccurate due to irregularities in recording vaccines or addition errors aggregating data.
	  ○ We have received reports from CDC consultants and the Gates team that micro-census data can be skewed.
	• **Piloting is sufficient to verify the feasibility of the endline coverage survey.**
	• **If the validation timeline extends beyond June, the RCT might have to be delayed if the admin data turns out to be extremely unreliable and the baseline**

	survey will have to be carried out.

Option 4: Administrative Baseline and Endline

Summary	• Generate administrative coverage estimates to stratify the random selection of clinics into treatment and control groups. • **Compare administrative coverage estimates in treatment and control clinics at endline** using baseline administrative coverage estimates as a control variable.
Methodological Assumptions	• GiveWell's modeled relationship between coverage and mortality is accurate. • Mothers from control clinics do not travel to treatment clinics for vaccinations. • Randomization stratified using administrative coverage estimates led to a balanced sample • Administrative data is accurate enough to group the clinics for sub-group analysis. • **Errors in the administrative data are unbiased towards treatment and control clinics.** • **Coverage estimates cannot be systematically biased (uniformly high or low) if they are used to determine coverage rates without the program in the CEA.**
Budget	• **$30,000 for each administrative data collection** at baseline and endline. • Additional costs associated with more intensive verification of administrative data sources.

Operational Risks	• Administrative data concerns:
	  ○ DVD-MT data could be inaccurate due to irregularity in recording vaccines or addition errors aggregating data.
	  ○ We have received reports from CDC consultants and the Gates team that micro-census data can be skewed.
	• **Data from neighboring clinics is sufficient to correct for the effect of incentives drawing in babies from outside their catchment area inflating the number of vaccinations given relative to population.**
	• **Administrative data systems may change during the course of the study. In particular, polio micro-censuses may end.**

Option 5: Mortality Study

Summary	• Compare estimated **under-5 mortality rates** in treatment and control clinic catchment areas. • Could **extrapolate mortality in a smaller age cohort** affected by the program to an estimate of under-5 mortality so that they study is less than 5 years.
Methodological Assumptions	• For the CEA case fatality rates during the study period must be the same as when New Incentives scales. • **Immunizations will not be added or improved** in the routine immunization schedule.
Budget	• To detect a 2% change, the maximum possible given reasonable assumptions in the CEA, **we would need 500 clinics in the study**. (see Appendix 2 Table 2 for details) • A mortality study this large would cost millions.
Operational Risks	• New Incentives has the **operational capacity to scale to 250 clinics**.

Option 6: Individual Randomization

Summary	• **Randomize mothers arriving at the clinic** for BCG to incentive and non-incentive groups. • Compare the fraction of **babies who complete the vaccination schedule** between groups.
Methodological Assumptions	• **Retention is a sufficient outcome variable** (understanding that it will likely underestimate cost-effectiveness of the program by not capturing new mothers drawn to the clinic by incentives). • Control mothers knowing other mothers who received the incentive will not affect their behavior. • The difference in operations relative to unrandomized at scale operations will not significantly affect impact. • The few clinics involved in the study are broadly representative of scale-up sites.
Budget	• **Only 1-2 clinics** would be necessary to detect the expected effect. • As **New Incentives administrative data would be used**, additional IDinsight field costs would be minimal.
Operational Risks	• **Another program** may severely effect operations at one of the few study clinics.[10] • **Local leaders may be angered** by the fact not all mothers receive incentives. • New Incentives field staff would have **strong incentive to accept bribes** from mothers if there is anyway field staff could influence treatment assignment.[11]

[10] This could be mitigated by increasing the number of study clinics beyond the few required to achieve statistical power, but the individual design can only be justified from a cost perspective if it has fewer clinics than a clinic clustered design.
[11] One way to prevent field staff from influencing treatment assignment would be to use biometrics, but that caries its own operational risks.

New Incentives
Measurement Options
31 March 2017

Summary

Option	IDinsight's perspective	New Incentive's perspective
1) *Coverage Survey and Census Baseline and Endline*	Most rigorous option, but information gained from a baseline coverage survey does not justify its cost.	New Incentives favors this option because it does not risk reliance on skewed population estimates or potentially incomplete administrative data for stratification.
2) *Coverage Survey Baseline with Census and Endline Sample Survey*	Potential bias from excluding migrant and other populations lost to follow-up is not worth cost savings	New Incentives believes that having accurate population estimates for baseline estimates of coverage is critical. We are more skeptical that micro-census or other currently available population data is not reliable.
3) ***Recommended by IDinsight:*** *Coverage Survey Endline and Administrative Baseline*	With reasonable assumptions will provide similar results to option 1, but at a far lower cost.	New Incentives is highly concerned about using skewed data to estimate coverage at baseline since this data will be used both to stratify clinics and also to control for baseline coverage in endline analysis.
4) *Administrative Baseline and Endline*	There are too many risks with the administrative data to rely on it for the primary impact estimate at endline.	Similar to IDinsight's perspective
5) *Mortality Study*	The number of clinics required is unfeasibly large.	Similar to IDinsight's perspective
6) Individual Randomization	Too many assumptions are required to link study results to New Incentives cost-effectiveness at scale	Similar to IDinsight's perspective. Particularly, concerned about issues regarding spillover and not measuring actual coverage.

Details on RCT Measurement Techniques

Below are details on how IDinsight plans to conduct coverage surveys, estimate coverage using administrative data, and measure mortality impacts. These details are meant to make the options in the table above more tangible, but the details of any of these data collection strategies could change as we pilot and learn more about the context.

Coverage Survey

A coverage survey is a household survey which determines a population's coverage rate by directly interviewing a representative sample of the population. The following steps would be taken to conduct a coverage survey for the RCT:

- Determine the relevant age of the cohort under analysis. Children should be old enough to have completed their routine immunizations and young enough that the program is likely to affect their vaccination status (approximately 1 year-old children)
- Conduct a census of all children in all settlements within the official catchment area of a clinic. Catchment area will be defined by the official settlement list and local immunization officers in the event of any ambiguity.
- If more than 40 eligible babies randomly select 40 to assess vaccination status in detail. We will triangulate information from child health cards, mother's recall of medical details of past vaccinations, and clinic records to determine what vaccinations a child has received.

Administrative Coverage

We believe rough coverage estimates can also be derived from administrative data. The process we would use to derive coverage estimates is described below:

- Determine which clinics located in close proximity to treatment clinics and are likely to be impacted by the program (via drawing in mothers to the treatment clinics and away from nontreatment clinics).
- Obtain administrative data on vaccine doses given by the treatment clinics and all surrounding clinics.
- Determine the change in vaccinations given across the entire area to discount additional vaccinations in the treatment clinic that are due to diverting mothers from other clinics.
- Convert this impact in terms of vaccinations given to coverage estimates[12] using micro-census data by dividing the number of vaccinations given by the population of the treatment clinic's catchment area.

Measuring Mortality

Mortality can be examined as part of a household survey. An outline of the process is described below:

- Determine which age cohort's mortality would be analyzed. If measles mortality is the focus, all children ages one to five would be the relevant cohort medically, but data from children ages 12-24 months could also be extrapolated to this broader cohort where we expect New Incentives will have mortality impact.
- Determine the settlements in treatment clinics' catchment areas.

[12] Strictly speaking, we could simply report results in terms of additional vaccinations caused by the program, but a coverage calculation would have to be made eventually to feed the results into GiveWell's CEA.

- Create a timeline of key community events in each settlement to anchor birth and death dates provided by mothers.
- Create a census of the catchment area focusing on the age cohort of analysis.
- Through household sample surveys determine how many children in the relevant age cohort died in treatment and control catchment areas since babies eligible for the program began receiving the relevant vaccines.

Individual Randomization

New Incentives has indicated that they believe it is operationally feasible to randomize select mothers within a clinic. The general process is described below:

- Immediately before receiving the BCG shot the babies will be randomized into treatment and control.
- Treatment mother receives a New Incentives ID sticker, the control mothers do not. All children with BCG scar and no New Incentives ID sticker are clearly control group. [13]
- Control mothers will have no further interaction with New Incentives field staff.
- The number and kinds of vaccines treatment and control babies receive after BCG will be compared.

The described method cannot measure New Incentives' program drawing mothers into the clinic that normally would not have ever attended. This could result in a significant underestimate of cost-effectiveness.

Key Decision: What increase in immunization coverage should the study be designed to detect? How important is knowing the programs impact across different kinds of clinics?

Our review of the GiveWell cost effectiveness model suggests that roughly a 10% increase in coverage is the minimum effect New Incentives would need to demonstrate to potentially achieve a top charity rating. [14] The power calculations outlined in Appendix 2 show that 60 clinics should be sufficient to detect an increase of approximately that magnitude with 5% significance (Appendix 2 Figure 1). However, as shown in Appendix 2 Figure 1, powering the study to detect any effect size less than 10% under standard conservative assumptions[15] would require exponentially increasing numbers of clinics since 10% is near the tail of the power curve (Appendix 2 Figure 1).

If New Incentives achieves a 20% increase in coverage, then 20 clinics will be sufficient to detect an effect. [16] Thus with 60 clinics we will be able estimate whether the program was effective in low, medium, and high baseline coverage clinics. If the program is impactful across these three kinds of clinics, GiveWell can have

[13] To deal with the issue of mothers losing their child health cards or New Incentive's staff influencing eligibility in exchange for bribes mother's fingerprints could also be scanned.
[14] This estimate was based on examining scenarios where New Incentives would be three times more cost effective than GiveDirectly and 25% less cost effective than AMF based on GiveWell's current estimates. An 8% increase in coverage under some assumptions could still lead to top charity status. With 60 clinics this we calculate a power of 64% to detect an 8% increase with 10% precision.
[15] 80% power and 5% significance with baseline and midline outcomes not meaningful correlated with variations in effect size. Graphs in Appendix 2 explore the implications of other assumptions.
[16] Sophie, Chelsea, Isabel, and Natalie all assumed increases around this magnitude in the cost effectiveness models. Studies in both Nigeria and India found at least 20% increases in vaccination with similar sized cash transfers. The retention increases compared to historical retention observed during the pilot are also greater than 20%.

more confidence that the program's impact is generalizable and not dependent on the average baseline conditions in the study sites.

The mortality study would require an infeasible number of clinics because even the most optimistic estimates of program's impact would only result in a 1-2% change in mortality. Detecting such a small change in an RCT is very difficult. A conservative power calculation suggests 500 clinics would be the minimum necessary (Appendix 2 Table 2). Once New Incentive's scales up, a better approach may be to compare mortality rates in the next DHS survey between states where New Incentives operates and a synthetic control made up of other Nigerian states.[17]

Next Steps

Once GiveWell provides guidance on the key decisions outlined in this document, IDinsight will work with New Incentives to take the following steps:

- Prepare a more detailed proposal for the activities listed in the preparatory phase.
    - Prepare a funding request for necessary field activities during the preparatory phase.
    - Fieldwork for the preparatory phase will take place in May/June 2017.
- Draft a proposal and pre-analysis plan for the RCT & apply for IRB approval.
    - Drafting will be heavily influenced by learnings from the preparatory phase.
    - Decision and funding for baseline in summer 2017.

[17] Synthetic controls would involve weighting other states in Nigeria based on their similarity to the states where New Incentives operates. This weighting process results in a theoretical state which is comparable on observables to the states where New Incentives operates.

Appendix 1 Administrative Data Sources

Data Source	Description	Planned Use	Known Issues
Micro-Census	As part of the polio eradication campaign, health workers periodically go house to house to count the number of under-5 children in order to set campaign targets.	We plan to use the micro-census as the denominator in administrative coverage estimates. We will use population distributions provided by other surveys such as the DHS to convert the under-5 population estimate to an under-1 population estimate.	Figures are sometimes inflated so that vaccination workers can receive more days of pay. Estimates are also sometimes an undercount with polio vaccination teams reporting greater than 100% coverage for some settlements.
DVD-MT	The DVD-MT system is the primary source of administrative data on immunization in Nigeria. Clinics tally each vaccination they give and these tallies are aggregated and digitized. The DVD-MT system also contains information on vials of vaccine distributed each LGA to clinics	We plan to use the DVD-MT data to derive the numerator for our administrative coverage calculations. We will also check the accuracy of the vaccination numbers by comparing against the vial distribution information.	There are sometimes errors in aggregation or counting with the tally sheets. Vials distributed are only a rough proxy for vials used because the cold chain for some vials can be broken during the distribution process resulting in wastage that is rarely recorded.
WHO DQS	To improve the accuracy of the DVD-MT system WHO consultants regularly check that the tally sheets are accurately aggregated and inputted into the DVD-MT system for a sub-sample of clinics	Carefully reviewing the DQS report will be an important part of the administrative data validation process.	Only a few clinics are selected for a DQS review each quarter and it is unclear what sampling procedure is used to identify them. DQS reviews might not have taken place recently in all states.
Child Immunization Register	Each clinic keeps a child immunization register where basic information on each child served and the date of each vaccination for that child is recorded	During the coverage surveys we will cross-reference self-reported vaccinations for babies whose mothers have lost their child health cards against the child immunization register.	Many potential issues. For example, the same infant is sometimes recorded multiple times meaning his/her immunization history is scattered throughout the register. Other times certain vaccinations will not be recorded at all.
Child Health Cards	A mother is given a child health card for her baby at her first visit. Each vaccination is recorded on the child health card. Mothers are reissued new cards based on data in the child immunization register	In the coverage survey one of the key sources of data on a child's vaccination status will be the child health cards provided by the mother. The cards are the most definitive indicator of vaccination.	Mothers frequently loose the child health cards. Data on cards that have been replaced and transcribed from the child immunization register may be inaccurate.

	if they lose the card.		

Appendix 2 Power Calculations

The graphs and tables below illustrate the tradeoffs inherent in determining the size of the study. We have focused on graphs for measles as they are slightly more conservative than those for PENTA3. In the study, our primary desired outcomes variable will rely on knowing the full vaccination history of individual babies. However, measles and PENTA3 coverage rates are good proxies for overall immunization.

The baseline coverage rates, standard deviations, and inter-cluster correlations were all calculated using variations in ward coverage rates in Kebbi state. The ward coverage rates were calculated by dividing the total number of vaccinations given as recorded in administrative data by ward population estimates derived from the micro-census.

Figure 1 illustrates why 60 clinics was selected. The orange line represents 60 clinics, assuming 40 babies per clinic catchment. The choice of 40 babies is discussed later in the document. The line intersects the power curves at effect sizes needed to obtain results with a 5% and 10% probability of type-1 error. Type-1 error is the probability an observed result is different from the null hypothesis only due to random chance. It is often referred to as alpha. The green line indicates the effect size necessary to obtain results for 20 clinic subgroups with a probability of type 1 error of 5% and 10% respectively. Table 1 lists the effect sizes where these lines intersect the graph.

Figure 1

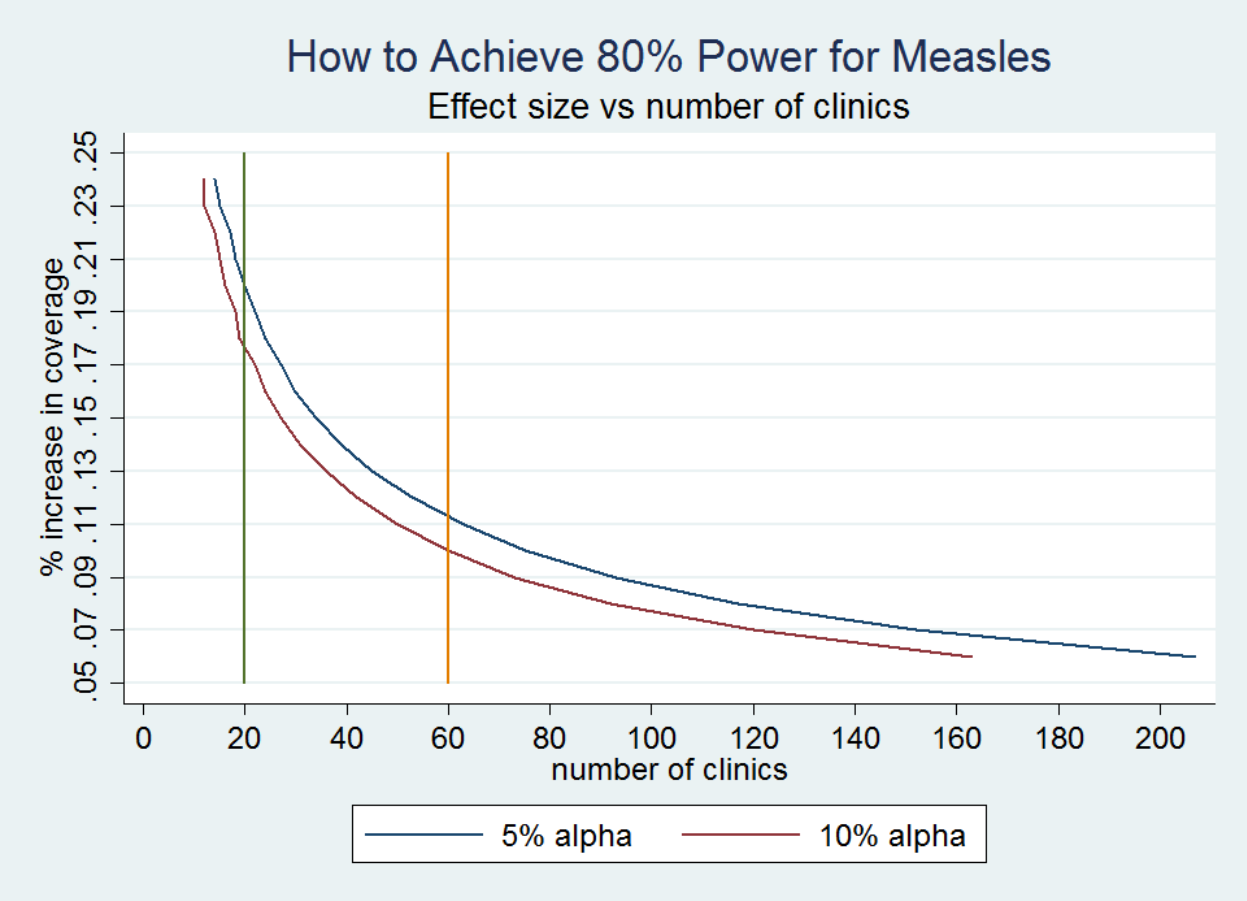

Table 1: Detectable Effect Sizes

	Measles		PENTA3	
	60 Clinics	**Group of 20**	**60 Clinics**	**Group of 20**
5% Probability of Type 1 Error	11%	20%	11%	19%
10% Probability of Type 1 Error	10%	17%	10%	17%

Figure 2 illustrates the robustness of choosing 60 clinics. The graph relates power to the effect size necessary to achieve that power. Power curves for different inter-cluster correlations and correlations from baseline and midline survey rounds are included. The second ICC used for the power curves, .25, is taken from the WHO handbook on cluster coverage surveys. The ICC used in all other power calculations is .16 which is derived from the administrative coverage estimates for Kebbi state discussed above. A vaccine incentive study in Western Kenya observed a village ICC of .14 which suggests using an ICC of .16 may be reasonable. The power curve accounting for endline results being highly correlated with baseline and midline statistics is meant to illustrate the potential impact on power of additional rounds of surveying. A correlation of .31 is illustrated since using the baseline budget to add treatment clinics, seven based on current budget estimates, would result in a similar power curve.

Figure 2

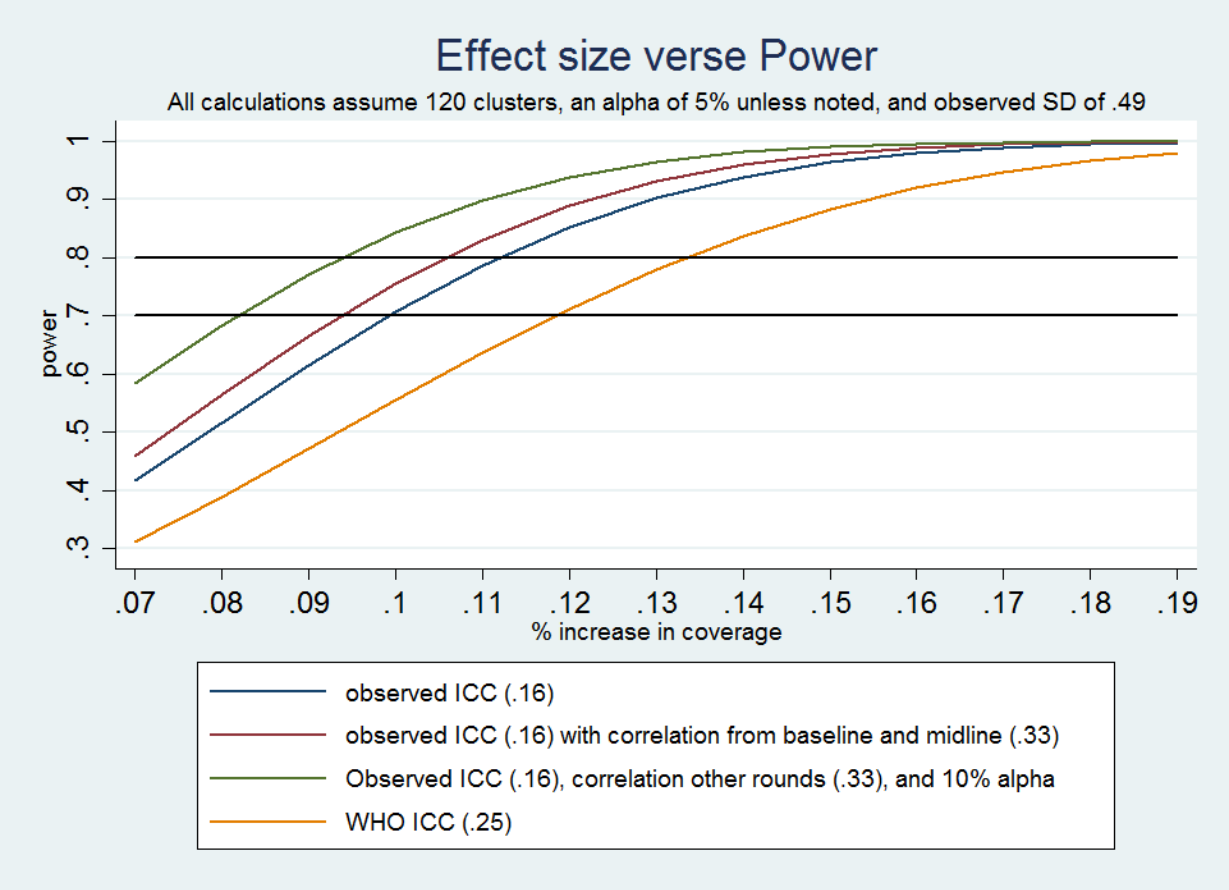

Figure 3 illustrates why 40 babies per clinic maximizes power while minimizing costs. The grey lines illustrate the power provided by 40 and 60 babies per clinic respectively. Note the effect sizes which allow 40 and 60 baby designs to achieve 80% power are only .2% different.

Figure 3

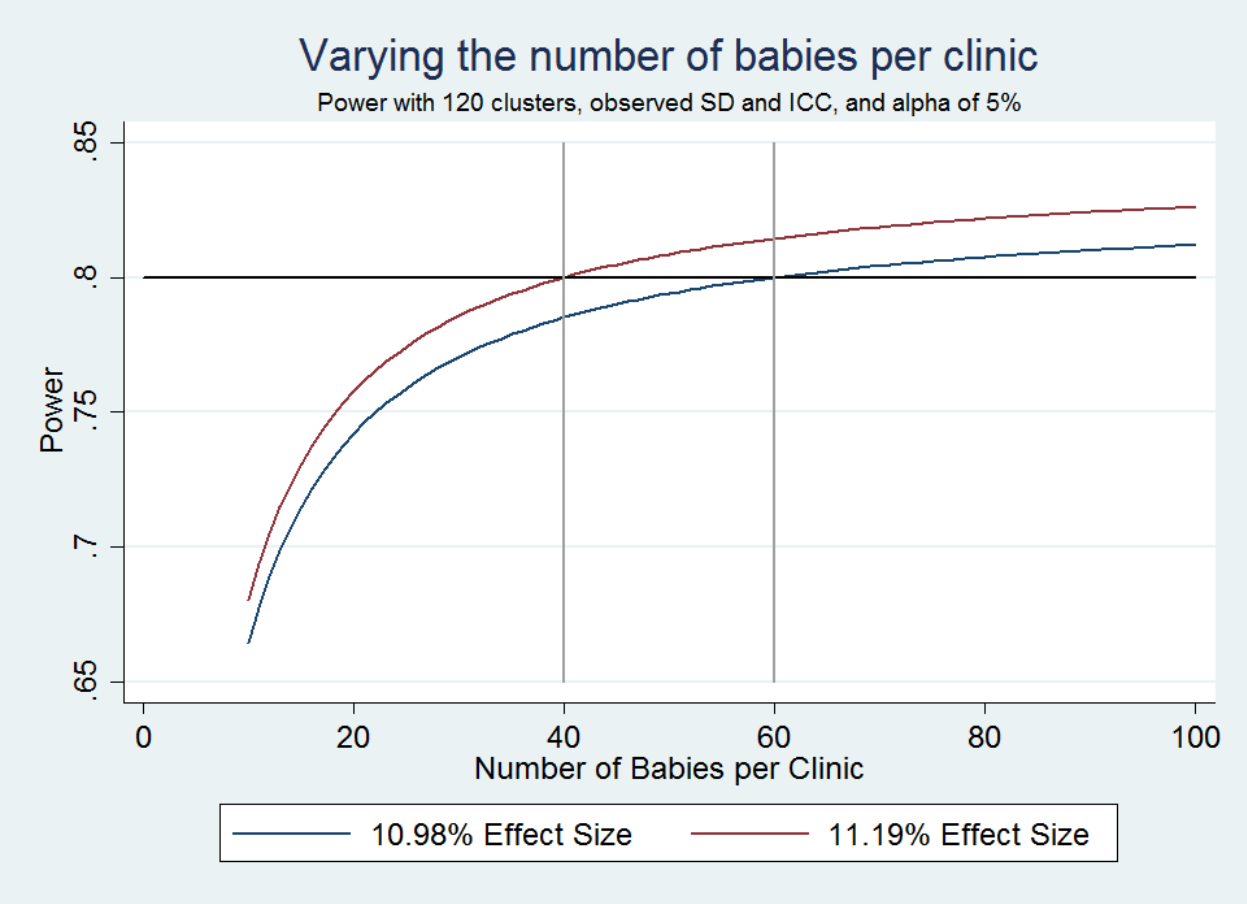

Table 2 shows the hypothetical number of clinics needed to achieve 80% power for a variety of situations not well illustrated by the graphs. Difference in baseline rate and standard deviation are not explored since baseline rates ranging from 30% to 70% all yield standard deviations of approximately .5 due to the nature of the binomial distribution. The assumptions behind the mortality power calculation are also illustrated.

Table 2: Number of Clinics Required for 80% Power

Scenario	Baseline	More babies per clinic	Higher ICC	Baseline	More babies per clinic	Higher ICC	Conservative Power
Outcome	Measles Coverage	Measles Coverage	Measles Coverage	PENTA3 Coverage	PENTA3 Coverage	PENTA3 Coverage	Mortality
Effect	11%	11%	11%	11%	11%	11%	2%
Rate at Baseline	46%	46%	46%	47%	47%	47%	7%
Assumed SD	0.50	0.50	0.50	0.50	0.50	0.50	0.25
Assumed ICC	0.17	0.17	0.25	0.16	0.16	0.25	0.10
Assumed Precision	5%	5%	10%	5%	5%	10%	5%
Observations per cluster	40	60	40	40	60	40	2000
Number of Clinics	**63**	**60**	**70**	**60**	**57**	**70**	**248**