

# Proposed RCT Evaluation Design

Prepared by IDinsight for GiveWell and New Incentives on  
June 15, 2017



Authors: Steven Brownstone, Maureen Stickel, Alison Conner, Daniel Stein, and Paul Wang

IDINSIGHT

*Table of Contents*

Evaluation Background.....	2
Goal of the Evaluation .....	2
Key Stakeholders .....	2
Existing Research on Incentives for Immunization.....	2
Research Objectives .....	3
New Incentives Program .....	5
Proposed Methodology.....	7
Key Evaluation Design Decisions .....	7
Sample Size Calculation.....	10
Evaluation Details .....	12
Discussion of Technical Risks.....	16
Ethical Considerations.....	20
Operational Risks .....	21
Timeline .....	21
Annexes .....	23
Annex 1 Relevant IDinsight Projects.....	23
Annex 2 Administrative Data Sources.....	25
Annex 3: Addition Power Calculation Information .....	26
Annex 4: Serological Techniques.....	28
Annex 5: Literature on Sampling Techniques .....	29
Annex 6: Timing of the Baseline .....	36
Bibliography.....	37

## Evaluation Background

### Goal of the Evaluation

The goal of the evaluation is to inform GiveWell's decision on whether or not to fund New Incentives' expansion in North West Nigeria. GiveWell will use the results of this evaluation, in addition to data on New Incentives' operating costs, to evaluate the cost-effectiveness of New Incentives. To this end, IDinsight's evaluation aims to precisely estimate the impact of New Incentives' program on vaccination coverage rates across a variety of clinics in Zamfara and Katsina states.

### Key Stakeholders

#### *IDinsight*

IDinsight is a client-service organization that helps social sector actors generate and use evidence to inform decisions. Our team has coordinated over 60 impact evaluations in Africa and Asia using experimental and quasi-experimental methodologies, and works with a wide range of for-profit, government and not-for-profit organizations.

Relevant projects include a clustered randomized controlled trial (RCT) in Zambia evaluating whether offering newborn and maternal HIV testing at clinics improved testing rates and / or adversely affected under-five immunization rates (Wang 2015), and a clustered RCT of non-monetary incentives to encourage facility delivery also in Zambia (Wang 2016). A more exhaustive list of relevant projects can be found in Annex 1.

#### *New Incentives*

New Incentives is an international NGO focused on leveraging the evidence around conditional cash transfers to achieve development goals. Since 2014, New Incentives has provided over 20,000 conditional cash transfers to Nigerian mothers. New Incentives began operating in Nigeria with a program designed to limit mother-to-child transmission of HIV. After re-evaluating which clinic healthcare service would be most cost-effective to incentivize, the program shifted in 2016 to focus on routine immunization. The details of New Incentives' immunization program will be discussed below.

#### *GiveWell / Good Ventures*

GiveWell is a charity research organization dedicated to finding the most cost-effective ways to improve lives globally. They are closely associated with the Good Ventures foundation which funds much of GiveWell's experimental and research work. This work includes New Incentives' immunization program and GiveWell's learning partnership with IDinsight. The evaluation of New Incentives' program falls under IDinsight's broader learning partnership goal of supporting GiveWell in their search for more top charities.

### Existing Research on Incentives for Immunization

There are a number of studies<sup>1</sup> that show incentives can have a significant impact on immunization coverage rates, especially in low baseline coverage settings. The landmark study is Banerjee and Duflo's 2010 paper on in-kind incentives to increase immunization rates in Rajasthan. The randomized study found the percent of fully immunized children in villages with the incentives and reliable immunization camps increased to 39% as compared to 6% for control villages. Villages where incentives were offered, but there was no intervention to increase the reliability of camps increased coverage to 18%. This intervention differed from New Incentives' model in that Banerjee and Duflo provided non-monetary incentives (lentils and thalis – dishware) rather than cash and the immunizations were provided at village camps rather than clinics.

---

<sup>1</sup> In addition to the research discussed below, see Loevinsohn 1986 and Chandir 2010. There is also a broader literature base on conditional cash transfers to encourage health intervention uptake summarized by Lagarde 2007.

An individually randomized RCT from Adamawa state in North East Nigeria provides evidence for the impact of incentives immunization in the Nigerian context. The study found an 800 Naira (\$5.30 USD), conditional cash transfer increased mother's tetanus vaccine take-up by 28 percentage points (Sato and Takasaki 2016, 5).

There are several ongoing evaluations studying examining the impact of incentives on immunization rates. A recent study (Gibson 2017) in Western Kenya found a modest increase in the percent of children fully immunized from 82% to 90% with SMS reminders and a 200 KES incentive per pentavalent and measles vaccine received (approximately \$2.35 at the time of the study). GiveWell is also involved in ongoing research in India and Pakistan. The Pakistan study explores incentive lotteries and the study in India looks at a primary healthcare facility-based non-monetary incentive program combined with SMS reminders. The Pakistan study uses individual randomization while the India study randomizes by clinic as our study aims to do. To determine coverage within the clinics catchment areas, the researchers randomly select seven villages out of a clinics coverage area<sup>2</sup> to do a coverage survey.

While the literature is clear that incentives can increase immunization, the extent that New Incentives will be cost-effective in the North West Nigerian context remains an open question. In the papers discussed above the magnitude of the effect ranged from eight to thirty-three percentage points. Further research is necessary to understand whether incentives for immunization are a cost-effective use of resources in the resource constrained healthcare system.

### Research Objectives

New Incentives' program was structured around the evidence discussed above that incentives can improve vaccination coverage rates. Hence, the goal of this study is to quantify New Incentives' program's impact on routine immunization coverage rates.

The primary research question is:

1. How does New Incentives' program affect the percent of infants in the community served by a clinic that complete the routine immunization schedule?

Secondary research questions include:

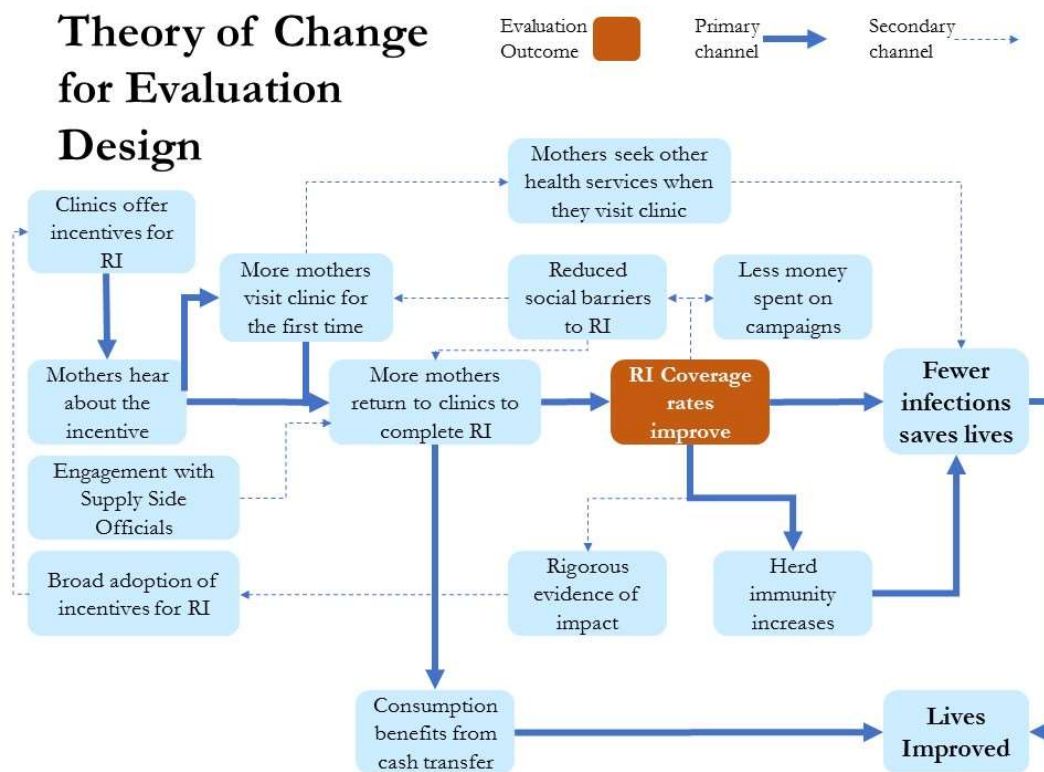
1. What is the effect for individual antigens, particularly Measles 1?
2. Does New Incentives' program improve the timeliness of vaccinations, particularly for Measles 1?
3. Does New Incentives' program result in health behavior changes beyond immunizations?

New Incentives hopes to increase coverage by drawing more mothers to visit the clinic as well as encouraging all mothers to complete their course of vaccination. The below theory of change further illustrates the many pathways through which the New Incentives' program could improve and save lives, and highlights the outcome which will be the primary focus of the evaluation: improved coverage rates of childhood immunizations. IDinsight will seek to collect indicative information on other pathways such as mothers seeking other care at clinic visits or supply side improvements as well as impact on other health behaviors, but the focus of the measurement activities will be on coverage rates.

---

<sup>2</sup> Across Haryana state, the location of the study, there were 14.8 villages on average per primary healthcare facility according to the 2014-2015 rural health statics report.

Figure 1: Theory of Change for New Incentives Routine Immunization Conditional Cash Transfer Program



#### Outcome Variables

The primary outcome variable is the program’s impact on **the percentage of fully immunized twelve to sixteen-month-old children in a clinic’s catchment area**. ‘Fully immunized’ is defined as having completed the traditional five visit routine immunization schedule when surveyed. The clinic’s catchment area is defined by the local government and available at each clinic as a list of settlements served.<sup>3</sup>

Secondary outcomes will include:

- Timeliness: Percentage of infants who receive Measles 1 when they are at least nine months and less than ten months old.
- Antigen Specific Coverage: Estimated changes in coverage rates for each vaccination. This will include all antigens listed on the child health card. Self-reporting will include Polio, BCG, PENTA/PCV (1-3), Yellow Fever, and Measles (1-2).
- Vaccination Average: Average number of antigens received per infant.
- Administrative Vaccination Levels: Number of antigens given by the clinic over four months.
- Percentage of infants who received at least one immunization.

<sup>3</sup> We will include all clinics either identified by LGA authorities or the clinic in the sample, but treat settlements where their inclusion status in the catchment is ambiguous separately in the analysis.

## New Incentives Program

### ***New Incentives in the North West Nigeria Context***

North West Nigeria has some of the highest fertility and lowest vaccination rates in the world (DHS 2013). This unfortunate combination has resulted in frequent measles outbreaks (NCDC 2016) as well as the area becoming one of the world's last locations with wild polio virus. In recent years, the donor community has invested substantially in improving supply side infrastructure for routine immunization (NRISP 2013), but coverage rates remain poor (UNICEF 2015). New Incentives is addressing this apparent shortfall in demand for immunization with the demand-side approach of cash incentives. More details on the Nigeria routine immunization system can be found in IDinsight's February [site visit report](#).

North West Nigeria has relatively low mobile phone penetration and minimal mobile money penetration. Consequently, New Incentives must provide their incentives in physical cash. While this increases operational complexity for New Incentives, using physical cash will be instrumental in changing mother's attitude towards routine immunization. It may also increase the chance of mothers directly controlling the money they receive from the program.

### ***The Incentive System***

New Incentives provides cash incentives to caregivers who bring their children for immunizations. The incentives follow the below schedule:

*Table 1: Schedule of Immunizations Incentivized by New Incentives*

Visit	Birth	6 weeks	10 weeks	14 weeks	9 months
<b>Immunizations</b>	BCG	Penta1, PCV1	Penta2, PCV2	Penta3, PCV3	Measles
<b>Incentive Amount<sup>4</sup></b>	₦500	₦500	₦500	₦500	₦TBD <sup>5</sup>

To be eligible for an incentive, the infants. In general, infants will be vaccinated if they are less than the maximum age for receiving any given vaccine<sup>6</sup>. New Incentives field staff work with nurses to ensure clinics follow maximum age guidelines so that infants are not turned away due to nurses enforcing lower maximum ages than state guidelines recommend.

Mothers should also be from the catchment area of a clinic to be eligible, but New Incentives recognizes this eligibility criteria cannot be enforced 100%. To prevent flooding a clinic with mothers from neighboring communities, New Incentives conducts limited advertising. Currently, the clinics have small posters and mothers who have received their first incentive get SMS reminders<sup>7</sup> for the remaining four visits. All other marketing is via word of mouth. Town criers are occasionally used to advertise the program.

New Incentives has a team of field officers responsible for disbursing incentives to mothers. On each vaccination day, the field officers check vaccine quality and stock and then prepare to disburse incentives. Incentives are paid in cash by a New Incentives' staff member who also assesses the validity of the infant for

<sup>4</sup> At current exchange rates ₦500 is approximately \$1.40, but exchange rates are currently unstable. As recently as February, ₦500 was approximately \$1.

<sup>5</sup> New Incentives is in the process of finalizing the measles incentive amount. IDinsight is providing evidence to support that decision by advising New Incentives on a phone RCT offering mothers in pilot clinics reminders or reminders and surprise bonuses in the incentive they will receive for measles.

<sup>6</sup> New Incentives will not be enforcing a strict maximum age

<sup>7</sup> In New Incentives' pilot clinics 75% of mothers listed at least one phone number New Incentives could use to reach them. However, Zamfara and Katsina states are poorer and have worse mobile phone penetration than the states where the pilot took place.

vaccination. The general principle is the incentive is given with respect to the infant, not the caregiver. This means the incentive is paid to whoever brings the infant to the clinic as long as that person also has the child health card. In practice, mothers tend to bring their infants. Mothers with twins get double the incentive amount for each visit.

### ***Fraud Prevention***

Fraud is prevented by ensuring photos are taken of all immunization records as well as photos of each woman holding their infant and the incentive they received. According to New Incentives' staff, the cash incentive seems to overcome social taboos about photos. Before and after photos of field officers' cash on hand are also required to compare to the recorded valid disbursements.

To prevent mothers from bringing infants multiple times to repeatedly receive the incentivized vaccines and thus the incentive award, all mothers are required to show their child health card with the infant's All Babies ID (See Annex 2 for a discussion of administrative data sources). The presence of a BCG scar would prevent mothers from bringing in infants to receive multiple child health cards under different names. However, there is some risk of mothers colluding with nurses to not record vaccinations in the child immunization register and then pretending to lose their child health cards to get a replacement showing incentive eligibility. While field staff would not be able to immediately address this threat, New Incentives' staff reconciling the electronic record would notice the double payout, triggering an investigation. More details on fraud prevention can be found in the IDinsight's February site visit report. New Incentives has since introduced a variety of anti-fraud measures including: enrolling infants only at the BCG stage (with fresh BCG mark), crossing out ABAE IDs at the last visit (Measles 1), and stamping ABAE IDs to prevent them from being used by other caretakers.

### ***Expansion Plans***

New Incentives plans to scale their program to states across North West Nigeria. They will first focus on Zamfara and Katsina states which are the two states in which the RCT will take place. These states were chosen after a careful scoping process. Zamfara and Katsina were chosen for a variety of factors including the lack of other planned programs which would have biased the results of this study as well as the receptiveness of local state governments to research.

During the study, New Incentives plans to operate at well-spaced clinics. Consequently, there will likely be many non-program clinics between each program clinic. However, at scale, New Incentives plans to cluster program clinics since spillovers between treatment and control sites will no longer be a concern. This change may reduce crowding at program clinics which may result in even larger effects on catchment area coverage. Additionally, New Incentives may begin radio or other untargeted advertising when operating at scale. This kind of widespread advertising will not be possible during the RCT due to spillover risk.

The exact clinic-level scale-up criteria is yet to be defined, but will likely include some combination of clinic size, to ensure sufficient volume for cost effectiveness, and a supply-side readiness assessment. The results from the RCT will include a range of clinic types and could provide information to inform the final scale-up criteria.

## Proposed Methodology

### ***Overall Design***

The evaluation will be structured as a two-arm cluster RCT. One arm will serve as control and the other arm will receive New Incentives full program. Since the goal is to evaluate New Incentives' program rather than incentives in isolation, there is no need to disaggregate the effects of New Incentives field staff on clinic operations from the incentives' overall effect.

There will be baseline, midline and endline measurement rounds. The baseline and endline will use a coverage survey and administrative data while the midline will use only administrative data. The midline will take place 12 months after baseline and the endline will take place approximately 22 months after baseline. The justification for the baseline will be discussed in depth below. The proposed midline is to provide an early indicator of program impact to guide GiveWell funding decisions prior to endline results.

### ***Outline of Study Design Phases***

The main steps involved in conducting the RCT are outlined below. Each step will be discussed in more detail later in the document.

1. Choose clinics to include in study.
  - a. Determine how far apart treatment and control clinics need to be to avoid spillover.
  - b. Use clinic maps to identify sets of well-spaced clinics for New Incentives to investigate.
  - c. Confirm that clinics meet the supply side criteria and other criteria New Incentives would use at scale.
2. Conduct a baseline coverage survey of the identified clinics.
  - a. Map the boundaries of the clinic's catchment area.
  - b. Use compact segment sampling to sample the catchment area.
  - c. Determine vaccination status for sampled infants.
3. Randomize clinics into treatment and control groups, stratifying on baseline coverage rates.
4. Use administrative data on vaccinations administered to conduct midline survey.
5. Conduct endline coverage survey.
6. Compare percentage of fully immunized infants (and other key outcomes) between treatment and control clinics.

## Key Evaluation Design Decisions

### ***Randomizing at the Clinic Level***

Clinic-level randomization is proposed over the alternatives of individual level and LGA level randomization. Randomizing at the clinic level allows the study to precisely measure the program's true impact on coverage using a feasible number of clinics. Clinic level randomization's primary disadvantage is that mothers may travel from other clinics to take advantage of the incentive. While there is a limit to how far a mother could practically travel with a newborn, many mothers may still travel to treatment clinics from surrounding clinics' catchment areas despite New Incentives' eligibility criteria. However, there are ways to limit or account for spillover of this kind, which are mention in further detail later in the document.

Individual randomization is not recommended due to the difficulty of randomizing mothers into the program when they have not yet visited a clinic. This would require a large effort to identify new births and communicate incentives to mothers on an individual level. This was deemed not feasible operationally. Additionally, it is possible that if mothers in the control group know that their neighbors have received incentives, this could



dissuade them from vaccinating their infants and bias the study results. Conversely, Sato (2016) found the individual incentives had significant positive peer effects.

One feasible alternative to individual randomization would be to randomize among mothers who come to the clinic to get BCG, providing incentives for further vaccinations. However, we determined that this has the risk of drastically underestimating the true cost effectiveness of New Incentives' model. An important part of New Incentives' theory of change is incentivizing mothers who would not otherwise go to a health facility to come to the health facility and vaccinate their children. While in many contexts most mothers come to the clinic for BCG after birth and fail to come back for other vaccinations, in North West Nigeria, all available data sources suggest BCG coverage rates are low. To capture the program's effect on bringing mothers into the clinics it is necessary that mothers who have not yet brought in their children for BCG know they will receive the incentive if they go to the clinic. For these reasons, we deemed individual-level randomization to be infeasible.

Randomizing larger geographic units such as local government areas (LGAs) does not fully address the spillover issue since any geographic grouping of clinics will border areas with clinics without incentives. Furthermore, if LGA level randomization is pursued, it would be cost prohibitive to use a reasonable number of clusters<sup>8</sup> to achieve the statistical power required to precisely determine New Incentives' impact.

### ***Spacing Between Treatment and Control Clinics***

As mentioned above, treatment and control clinics must be spaced so that mothers from control clinics do not travel to treatment clinics for vaccinations. The distance will be determined by triangulating different data sources from clinics where New Incentives operates. Information from the North West will be given additional weight due to its contextual relevance. These information sources include:

- Exit interviews with mothers to understand the distance they traveled to the clinic.
- Analysis of follow-up addresses recorded in clinic child health registers.<sup>9</sup>
- Analysis of trends in the number of vaccinations administered for clinics surrounding clinics with incentives.
- Information collected by New Incentives field staff on cost of travel.
- Cost of public transportation in the clinic area.

Zamfara and Katsina states are large enough that the 150 clinics can be spaced up to approximately 20km apart. When choosing the final spacing of clinics, we will error on the side of spacing them further apart than necessary since New Incentive's feels that reducing spillover risk is worth the increased operational complexity of the program clinics being spaced farther apart. If we find that women frequently travel further than 20km to receive incentivized vaccinations, this may require reducing the number of clinics in the study to ensure proper spacing. We do believe this scenario to be unlikely, however.

### ***Conducting a Baseline Coverage Survey***

A baseline coverage survey is proposed over the alternatives of not conducting any baseline or using administrative data to establish baseline coverage rates.

Baseline measurement will provide modest benefits to the study rigor. Primarily these benefits will come from increasing the likelihood of balance between treatment and control arms. The baseline will also provide

---

<sup>8</sup> The number of clusters required is largely insensitive to cluster size. Thus to achieve similar precision the study would need about as many LGAs randomizing at the LGA level as it needs clinics randomizing at the clinic level.

<sup>9</sup> The data will be collected from before and after the incentives are offered to analyze the changes

operational insights into data collection, and more accurate baseline coverage estimates to inform GiveWell interim cost effectiveness model.

The argument for using administrative data or forgoing baseline measurement all together is avoiding the substantial cost and capacity required to achieve the modest improvements to study rigor outlined below. The table below addresses the advantages and limitations of a baseline in detail. While IDinsight believes the benefits to the overall evaluation of conducting a baseline are relatively small, we understand that given GiveWell’s preference for rigor, the added expense of a baseline is justified.

*Table 2: Advantages and Associated Limitations of a Baseline Coverage Survey*

Advantages	Limitations
<p>Increased assurance that the treatment and control groups will be balanced on baseline coverage</p> <ul style="list-style-type: none"> <li>• A variable will be balanced by design at baseline if is used for stratification.</li> <li>• If no baseline is conducted it is impossible to know whether any given randomization led to balance on key variables at baseline.</li> </ul>	<p>Balance on baseline characteristics doesn’t guarantee the treatment and control groups would have remained balance at endline in a world without the intervention, and in general it is always possible for the study to be unbalanced on unobservables.</p>
<p>Ability to analyze outcomes across clinics with different baseline characteristics</p> <ul style="list-style-type: none"> <li>• To model New Incentives impact as they expand into areas with lower and higher baseline coverage GiveWell can analyze the program’s differential impact on lower and higher baseline coverage clinics within the RCT.<sup>10</sup></li> </ul>	<p>For the heterogeneous treatment effects to have significance, there would need to be large differences in outcomes in areas with different levels of baseline coverage (see Annex 3 for details on power)</p>
<p>Marginal increase in statistical power.</p> <ul style="list-style-type: none"> <li>• If the correlation between baseline and endline coverage rates is high, incorporating baseline data can improve the studies power.</li> </ul>	<p>Baseline data’s effect on power is small. Even if a third of the variation in vaccine coverage was explained by the baseline, the study’s power would increase by only 5%. (Annex 3 for details on power)</p>
<p>Operational learnings about the data collection plan that can be used to improve endline measurement</p>	<p>The operational context can change in the years between baseline and endline limiting the relevance of some operational learning about data collection.</p>
<p>Accurate baseline coverage values to include in GiveWell’s cost effectiveness model for the period before endline results are available.</p>	<p>The extent to which GiveWell will need to make funding decisions about New Incentives prior to endline results is unclear.</p>

Conducting a baseline coverage survey will enhance the rigor of the study, and reduce the risk of the results being biased. The baseline will also allow for an analysis of New Incentives’ effect across different types of clinics. This analysis may be important as GiveWell adapts the New Incentives cost effectiveness model to account for New Incentives’ expansion. Furthermore, baseline will provide a perspective on changes in other health behaviors and attitude towards vaccination.

<sup>10</sup> Depending on the effect size the study may not be powered to detect different characteristics impact on coverage precisely. However, these sub-sample analysis can still provide important information to guide scale-up.

### ***Vaccination Coverage as the Outcome***

Vaccination coverage (as measured by surveys and clinic records) is recommended over mortality and biological immunity. A mortality study is infeasible due to the hundreds of clinics required to detect the expected change in mortality, and a serological study with current technology does not justify the added operational complexity at baseline. We are still considering incorporating biomarkers as a robustness check at endline. More details on the promises and limitations of different serological techniques can be found in Annex 4.

### **Sample Size Calculation**

GiveWell has determined that New Incentives' would need to achieve a 7% increase in coverage to be considered for top charity status, while a 25% coverage increase would make New Incentives GiveWell's highest ranked top charity by being twice as cost-effective as AMF. Due to the infeasibility of detecting a 7% increase which would require at least 300 clinics in the study, the study will be powered to detect a 10% increase in the percent of fully immunized children with a p-value less than 5% we have proposed a sample of 150 clinics with 75 treatment and control clinics. Within each clinic, only 40 infants will be surveyed since surveying additional infants doesn't increase power materially.

We caution below calculations follow conventional study design norms, and thus may not fully address GiveWell's decision criteria. The alpha in the calculations below is the likelihood the reported effect is in truth zero. However, since a 7% effect is the threshold for top charity status GiveWell will also need to consider the probability the true effect is less than seven percent. IDinsight is currently working on a statistical simulation to better estimate this probability in the context of the study. A very conservative estimate is that to ensure there is less than a 5% chance the true program effect is less than 7%, the mean estimated effect would need to be 17%. However, IDinsight is still working on refining these estimates, and would appreciate feedback from GiveWell on how they approach the confidence intervals around effectiveness when comparing existing top charities.

*Table 3: Treatment Clinics Required to Detect Different Effect Sizes at Different Levels of Power. (See Annex 1 for more graphs)*

Detected Effect		Measles Vaccine		PENTA Vaccine	
		80% Power	90% Power	80% Power	90% Power
7% increase in coverage	5% alpha	151	203	145	193
	10% alpha	120	166	114	158
8% increase in coverage	5% alpha	116	156	111	148
	10% alpha	92	127	88	121
10% increase in coverage	5% alpha	<b>75</b>	100	72	95
	10% alpha	60	82	57	78
11% increase in coverage	5% alpha	62	83	60	79
	10% alpha	50	68	47	65

Figure 2: Relationship Between Study Size and Power to Detect Different Effects Relative to a Null Hypothesis of no Effect

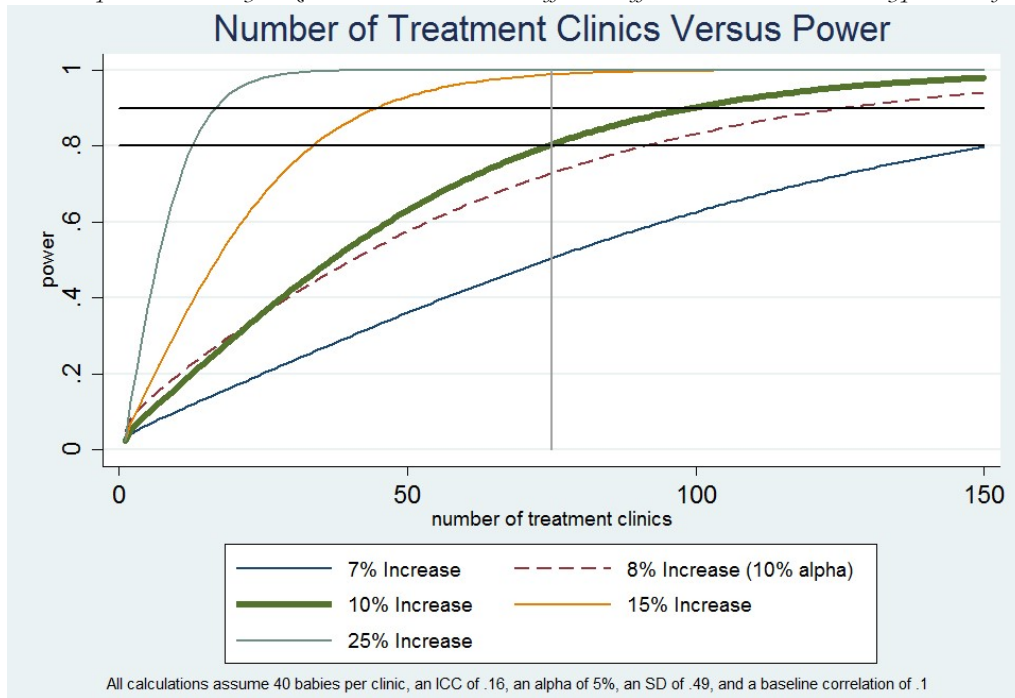
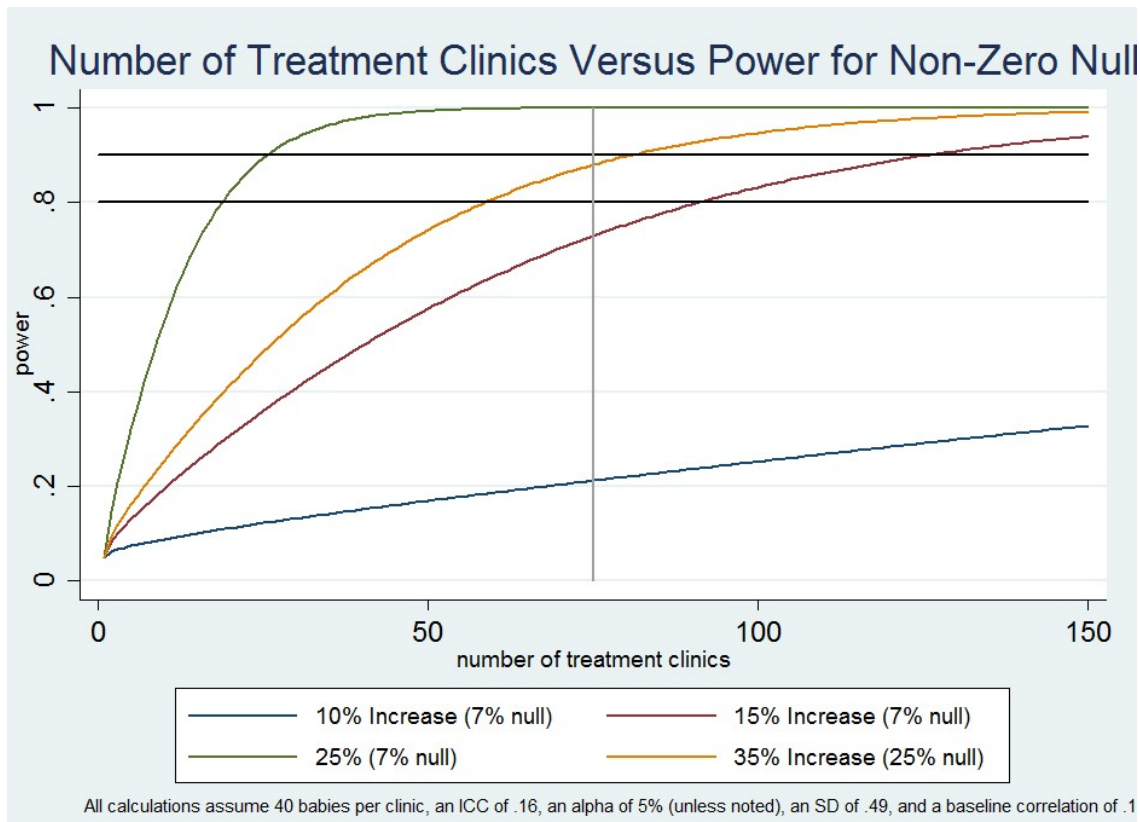


Figure 3: Relationship between Study Size and Power to Detect Different Effects Relative to Different Decision Thresholds



We recognize that GiveWell and New Incentives both would ideally want the study to be powered at greater than 80%, and to detect even a 7% increase in coverage. However, in order to ensure we can space treatment and control clinics sufficiently, we feel it is important to keep the number of study clinics limited. Also, having some areas of the state where other experimental interventions can be rolled out without effecting the study may be beneficial. A final practical consideration is that a lower number of treatment and control clinics ensures strong oversight of data collection.

GiveWell noted in recent communication that the amount of funding New Incentives is likely to receive increases as the estimated effectiveness approaches 25%. Thus, not detecting a seven or eight percent impact would not impose nearly as high of an opportunity cost as not detecting a 15% increase in coverage. The study is well powered to detect these larger potential impacts.

## Evaluation Details

### ***Steps 1: Selecting Clinics***

Using clinic maps available online<sup>11</sup> and obtained from eHealth Africa, IDinsight will identify groups of clinics spaced so that it is very unlikely mothers would be willing to travel between them to access incentives.

Once a list of potential study clinics has been produced, New Incentives will visit these clinics to ensure operational feasibility. New Incentives plans to screen for basic supply side readiness, only periodic stock outs, and a sufficient number of immunization days to make the program practical. Clinics with extremely low numbers of women attending vaccination will not necessarily be screened out if they appear to have a large catchment population.

In the event two nearby clinics are both considered eligible by New Incentives, IDinsight will randomly select one to be included in the study. In cases, where an identified clinic is very close to a clinic also offering large numbers of immunizations, it may make sense for the program to operate in both clinics to reduce spillover risk, but with only one clinic included in the study.

The final operational screening criteria will be developed in coordination with New Incentives and shared with GiveWell once New Incentives gains operational experience in the North West.

### ***Step 2: Conduct a Baseline Coverage Survey***

#### *Sampling Procedure*

The study's population of interest is the birth cohort who would be 12 to 16 months old at the time of endline living within a clinic's catchment area. While many studies examine 12 to 24-month-old infants, the large number of clusters means only 40 infants per catchment need to be identified. The age-group for the New Incentives study is focused on 12-16 month olds to ensure the program will be fully operational even for the oldest sampled babies at the endline. Moreover, data quality and mother recall is likely to be best amongst younger infants<sup>12</sup>.

To ensure that the baseline survey can be completed on time, with reasonable budget, and with high quality supervision, we propose using compact segment sampling to sample a clinic's catchment area. An overview of the literature comparing compact segment sampling to conducting a full listing can be found in Annex 5<sup>13</sup>.

---

<sup>11</sup> <https://africaopendata.org/dataset/health-facilities-in-nigeria>

<sup>12</sup> Based on population estimates derived from the polio campaign data there should be on average 107 12 to 16 month olds in a given catchment area. If there are instances where at least 40 infants are not identified, the expected solution will be to census more households.

<sup>13</sup> Note that given the relative importance of the baseline in terms of the overall evaluation, marginally changing the theoretical rigor of the baseline has even more marginal effect on the studies overall rigor.

To implement compact segment sampling, we propose the procedure outlined in table 3 below.

*Table 4: Compact Segment Sampling Strategy*

High Level Procedure	Practical Considerations
1. All settlements within a clinic's catchment are divided into segments of roughly equal size	<ul style="list-style-type: none"> <li>The number of segments per settlement will be determined by the catchment areas' approximate population as estimated from polio immunization campaign data.</li> <li>The process will be done remotely using satellite maps of the settlements and drawing borders along roads or clearings. Since the process is quick and geotagging settlements remotely is difficult, all settlements within xkm of a clinic will be divided.</li> </ul>
2. One segment from each settlement is randomly selected	<ul style="list-style-type: none"> <li>Maps illustrating these segments will be distributed in print and electronically to field teams responsible for a clinic catchment.</li> </ul>
3. Field teams will census the selected segment of each settlement within a clinic's catchment. <sup>14</sup>	<ul style="list-style-type: none"> <li>Advance teams will confirm which settlements are in the catchment area of a clinic.</li> <li>The census team will ask about living and deceased infants born into the cohort.</li> <li>Households with multiple eligible infants will be treated with the principle one observation per mother<sup>15</sup> since the mothers receive the incentive and bring infants to the clinic.</li> <li>Community events and holidays will be used to facilitate birth date recall if paper record not available.</li> </ul>
4. Around 40 eligible infants, living or deceased, will be randomly selected from those listed for an in-depth survey	<ul style="list-style-type: none"> <li>If not enough eligible infants are censused within the initial segments, an additional segment to census will be selected at random until all segments are exhausted.</li> </ul>

#### *Coverage Survey Data Collection*

The main unit of analysis for the study is a mother-infant pair. Household, mother, and infant data will be collected using a household survey. Clinic data will either be geographic or derived from administrative sources.

The household survey will consist of four modules:

1. Self-reported vaccination history, Child Health card check and a BCG scar check for living infants

<sup>14</sup> The team is still considering the best solution for clinic catchments with large numbers of settlements (20-30).

<sup>15</sup> For the purposes of listing we will count twins as one infant. If a twin is selected one of the twins will be randomly selected for further surveying. Infants with different mothers, but the same father will be counted as individual observations.

- Globally standardized questions which use details such as location of the immunization to enhance accuracy.
  - Leverage community events to increase accuracy of reported dates of immunization.
2. Demographics and socioeconomic status including other health behaviors
  3. Attitudes towards vaccination
  4. Exposure to incentives

While at each household, interviewers will ask for infants or deceased infant's child health cards. To increase the likelihood mothers will have child health cards available, community leaders will be enlisted to announce the coverage survey team so that mothers will have time to find their cards. In cases where a card is not available a member of the survey team will look for the infant in clinic records so that the mother's self-reported vaccination history can be verified. Cases that can't be further verified will be treated as missing in the main analysis and used to bound estimates in robustness checks.

If there are a large number of cases where a child's vaccination status can't be determined during baseline measurement, we will consider strategies to strengthen administrative record keeping at treatment and control sites and card retention from the period the endline cohort is born through the endline.

Coverage estimates using clinic administrative data described in the section on midline data collection will be in a robustness analysis.

### ***Step 3: Clinic Randomization***

After the baseline coverage survey, the clinics will be randomized. The proposed randomization scheme will have two levels of stratification. First, we will stratify by state since state-level contextual factors may have a major influence on the program's impact. Next, we will stratify by baseline coverage, dividing clinics in each state into around 4 strata of similar coverage rates.<sup>16</sup> Thus the 150 clinic sample will be divided into around 8 groups of 16 or 17. Other important variables such as catchment area size and population density will be controlled for explicitly in the analysis.

### ***Step 4: Midline Data Collection***

In early September 2018, we will use administrative data to derive preliminary estimates of the program's impact. The infants initially enrolled in November and December should be due for their Measles vaccination by the end of August, as well as infants who only got BCG and enrolled when they were slightly older. More importantly, the main sample cohort of infants born in April and May should have largely received PENTA 3 vaccinations by the end of August.

We will collect August DVD-MT data in the form of the clinic-level tally sheets, the official record of vaccinations given, (see Annex 2 for a discussion of administrative data sources) from study clinics and the clinics surrounding treatment clinics. We plan to construct coverage rates by dividing the adjusted number of doses given in July and August by the estimated size of a two-month birth cohort in a clinic's catchment. This estimate will be based on using the partial census conducted at baseline to adjust the polio population data. After June fieldwork and baseline data collection, we will have a better sense whether administrative data sources are sufficiently accurate to create these estimates.

Adjusting the dose statistics is necessary for treatment clinics because many of the doses given may be for children from neighboring clinics who traveled to a treatment clinic to be eligible for an incentive. The adjustment will be made by discounting the number of vaccinations in treatment clinics by the change in vaccinations in neighboring clinics since baseline. This adjustment will only work if no new interventions began

---

<sup>16</sup> The exact amount of strata will be determined by looking at the distribution of coverage rates.

at neighboring clinics during the study. However, we will be working with the relevant authorities to target other interventions in areas away from study clinics.

The coverage results for all vaccines aside from measles combined with retention data from learning and pilot sites could form the basis for a 2018 top charity recommendation, or support additional funding to help New Incentives prepare to scale once the endline results are available. Other rounds of low cost administrative data collection and analysis could take place as relevant to support GiveWell’s decision-making.

Given the relatively short time period between the study’s endline results and 2019 top charity recommendations, it may be important for New Incentives to receive additional funding prior to the study results so that New Incentives will not be ineligible for top charity status on account of inability to absorb funds, and to ensure there are no implementation funding gaps between the study and a top charity decision.

***Step 5: Endline Data Collection***

Endline data collection will proceed similarly to baseline data collection, but integrate improvements based on learning from the baseline. For instance, it may be necessary to increase the number of households censused in order to obtain the correct number of infants, or to improve the procedure for verifying vaccination status. One key planned difference is that the compact segment selection process will not be repeated as the same compact segments from baseline will be used to census at endline.

The endline data collection will take place 12 months after New Incentives clinics began operating at full volume. Based on previous experience, New Incentives thinks this will take place in April 2018, but there is flexibility to push back the endline if New Incentives program scales into the treatment sites slower than expected.

The 12-month interval is necessary for disaggregating the program’s impact on vaccination timeliness and overall vaccination status. Mothers in control clinics in particular may come later than nine months to receive their measles vaccinations. If the endline took place nine months exactly from birth, these infants would be considered unvaccinated. Since the endline will take place when infants are at least 12 months, we will be able to correctly classify these infants as untimely but vaccinated.

***Step 8: Compare Percent Fully Immunized Between Treatment and Control Clinics***

While the full range of analyses will be detailed in the study’s pre-analysis plan, the primary analysis will be a comparison of coverage rates for infants in the coverage areas of treatment and control clinics. This specification will be an analogue of the familiar ANCOVA model frequently used in impact studies, but modified to take into account the fact that this is a repeated cross-section:

$$V_{E,ij} = \beta_0 + \beta_1 * T_j + \beta_2 * C_j + \beta' * \gamma_{ij} + \alpha_p + \epsilon_{ij}$$

- $V_{E,ij}$  is the applicable endline vaccination status for infant  $i$  in clinic  $j$ .
- $T_j$  is the treatment status of clinic  $j$ , which contains infant  $i$
- $C_j$  is the baseline coverage rate<sup>17</sup> at clinic  $j$ , which contains infant  $i$
- $\gamma_{ij}$  is a vector of infant-level covariates
- $\alpha_p$  is a vector of categorical factors corresponding to the clinic, as well as stratification dummies used in the randomization
- $\epsilon_{ij}$  is the error term for infant  $i$  in clinic  $j$  clustered at the clinic-level

---

<sup>17</sup> Will adjust standard errors to reflect that baseline coverage is measured with error.



- Infant  $z$ 's outcomes are weighted inversely proportional to the probability of being selected into the sample

A full pre-analysis plan will be created after the baseline and the study will be registered. Waiting to produce the full pre-analysis plan until after baseline measurement will allow us to include any new hypothesis that emerge from the data collection exercise, and incorporate insights from New Incentives operations at the learning sites.

## Discussion of Technical Risks

### ***Ambiguous Results***

There is a chance that the impact evaluation will not unambiguously show the program was either effective or ineffective at increasing vaccination coverage in the selected clinics. In particular, it may be unclear whether the program increased coverage greater than the 7% threshold for top charity status.

The most likely scenario, especially for baseline measurement, would be a large number of infants where only self-reported vaccination status is available. In the 2013 DHS survey vaccination coverage rates were almost double when self-reported vaccination unsupported by cards health cards were included (DHS 2013, Table X).

We intend to gain more confidence on self-reported vaccination by cross referencing to the child immunization register, but when New Incentives compared child health cards to child immunizations registers they found that while 80% of PENTA 3 vaccinations could be cross-referenced in the immunization registers only 53% of measles vaccinations could be similarly verified.

If this issue is pronounced at baseline, we will explore different options to improve card retention and clinic record keeping for the endline birth cohort.

If data issues remain unaddressed it is possible different data sources may imply different results. For example, the program appears to increase coverage by 15% ( $p > 95\%$ ) if coverage is calculated using mother's recall alone, but when the only vaccinations that can be verified by administrative records are considered the program increases coverage by only 5% ( $p > 60\%$ ).

Possible additional scenarios include:

- Cost-effective results for only a subsample of clinics: the overall increase in coverage is estimated at 5% with a p-value of 80%, but for clinics with low baseline coverage there was a 15% increase in coverage with a p-value of 90%
- The program doesn't increase coverage, but increases timeliness: The program doesn't appear to increase measles coverage, but mothers in treatment clinics are 50% more likely to vaccinate their infants at precisely 9 months rather than returning sometime between 9 and 12 months.

### ***Partial Compliance***

New Incentives may fail to implement their program in selected treatment clinics or implementation stops in some clinics during the course of the study. An intent to treat (ITT) analysis would need to include all clinics, even those where implementation didn't take place, thus reducing the chance of the study detecting an effect as the estimate would underestimate the true treatment effect.

Strategies to mitigate this threat:

- Created a learning clinic phase to ensure operational readiness for scale.

- Extended the study timeline to give New Incentives time to ramp up operations so that the likelihood of treatment clinics failing to deliver reliable RI services or New Incentives failing to deliver reliable cash transfers is reduced.

Control clinics receiving the program would also be considered partially compliant. While it is highly unlikely New Incentives would accidentally implement their program in control clinics, other NGOs could roll-out similar cash transfer programs in the control clinics during the study resulting in the study underestimating impact. (Immunization campaigns, discussed in the external validity section, are another type of program we have to monitor)

Strategies to mitigate this threat:

- Working closely with state primary healthcare development agencies to ensure they direct other programs away from control clinics.
- Coordinating with other organizations working to improve routine immunization coverage in Nigeria to avoid overlap. (New Incentives is doing this currently, and avoiding overlapping programs was part of state selection)

### ***Missing Data***

It may be impossible to verify the vaccination status of some infants from the clinic records, which may decrease power. Furthermore, if these infants are affected by the program differently than infants for which we can determine vaccination status the results may be biased.

Strategies to mitigate this threat:

- Provide capacity building for record keeping in treatment and control clinics<sup>18</sup>.
- Ensure trends in administrative data match coverage survey results.
- Biomarker testing at endline to establish vaccination status for ambiguous cases.

Some people may refuse to be interviewed or to provide samples for biomarker testing if it is used at endline. These infants may be affected by the program differently than infants in families who consent to be interviewed making the results non-representative.

Strategies to mitigate this threat:

- Work with community leaders to encourage people to participate in the survey
- Ensure enumerators are friendly and culturally sensitive.
- Ensure household enumerator team is female so that they can speak with the mothers directly.

A general approach to dealing with data which couldn't be verified from the clinic records is to assume the missing observations are either all vaccinated or unvaccinated in order to bound the coverage estimates as part of the study's robustness analysis.

### ***Spillovers***

New Incentives has observed in pilot clinics that their program pulls in mothers from beyond the official catchment areas of the clinics in which they work. While many aspects of the methodology proposed above are meant to mitigate this risk, spillovers remain a threat to the evaluation.

---

<sup>18</sup> This would be an IDinsight led activity with technical input from New Incentives. IDinsight would likely sub-contract a third-party service provider for implementation

Mothers from other areas may come to treatment clinics for vaccinations to access incentives. If mothers in the catchment area of control clinics come to the treatment clinics and receive vaccinations this could bias the impact estimate downward.

Strategies to mitigate this threat:

- Ensure treatment and control clinics are placed far apart.
- Clinics near transport hubs can be avoided.

The program could affect a state's overall vaccine supply and the supply of Child Health Cards. Officials may prioritize facilities with incentives during supply shortages causing the study to overestimate the effect of the program.

Strategies to mitigate this threat:

- Work closely with state officials to encourage them to treat all clinics equally.
- Carefully monitor any supply side issues so that they can be explicitly controlled for in the final analysis.<sup>19</sup>

General excitement or information around vaccination could spread from treatment areas to control area potentially increasing immunization rates in control clinics as well leading the study to underestimate program impact.

Strategies to mitigate this threat:

- Measure attitudes towards vaccination at baseline and endline in both treatment and control sites to monitor this phenomenon

General solution to spillovers:

- Include distance to nearest treatment clinic in the analysis of control clinics so that any spillover not addressed by the design can be quantitatively modeled using spatial models.
- New Incentives will collect information on proxies for the distance mother's travel as part of their routine monitoring which will help quantify the extent of the spillover issues raised above.
- Monitor spillovers by collecting information on where mothers sought ante-natal care if applicable, and how they heard about the incentives.

### ***Evaluation Driven Effects***

Since self-reported data will play a role in the data collection process, there is an enhanced risk of psychological factors biasing the results. News about the program may also spread as far as control clinics with unpredictable implications for mother behavior.

Mothers in the treatment group may say they vaccinated their child because they know there is a vaccination program even if they really haven't vaccinated their child resulting in the study overestimating impact.

Strategies to mitigate this threat:

- Verifying mother reported vaccination against administrative records.

---

<sup>19</sup> The exact work plan for supply-side monitoring will be determined once the nature of administrative data support is decided.

Mothers in the control group may not vaccinate their children because they resent the treatment group or plan to wait to vaccinate until their own clinics offer incentives resulting in the study overestimating impact.

Strategies to mitigate this threat:

- Ensure control clinics are physically and socially distant from treatment clinics.
- One potential strategy could be making mothers eligible for incentives if they enroll in the program at BCG, the first vaccine, thus mothers would have to start anticipating future roll-out almost two years before it would be plausible.
- New Incentives will not make any explicit promises of scale-up to control sites to state officials and make clear that if such scale-up occurs it will take place months after the study completes.
- Ensure that any capacity building or supply-side initiatives at control sites are conducted under a different name that is not related to the New Incentives program (e.g. N-STOP).

A general strategy for reducing evaluation effects is to try and separate the survey team from the research to the greatest extent possible. Consequently, the research team will frame its work as coverage surveying, a routine activity in Northern Nigeria, rather than being associated with the New Incentives program. Baseline measurement will occur prior to any program activities and thus not be subject to any of these evaluation-driven threats.

### ***External Validity***

In general, external validity concerns are mitigated by the fact that the study will be used to justify scale-up funding of the program within the same broad cultural and political context of the study by the same implementer. An analysis which reveals consistent program impact across a wide variety of clinics would provide further evidence that the program's effectiveness should remain consistent as New Incentives scale.

The main external validity concern is whether the study's temporal context will be generalizable to the future. Major vaccination events in Nigeria may make it difficult to generalize results from the study to New Incentives' future implementation context.

A measles supplementary immunization campaign would result in much higher measles vaccination rates in control clinics than might be expected generally resulting in the study underestimating impact. The next measles campaign is scheduled for October/November 2017

Strategies to mitigate this threat:

- Complete the baseline prior to the next campaign without changing the implementation schedule. The endline age cohort would not be born at the time of the Oct/Nov campaign. There will likely not be another campaign until after endline if Nigeria maintains its campaign schedule.
- If the campaign can't be avoided, when interviewing mothers carefully ask about the timing of vaccination using community events to anchor dates and then report results including and excluding campaign vaccinations.

The roll-out of a new vaccine such as the rota-virus vaccine could have supply side or demand creation effects. During the roll-out of new vaccines there is more publicity around vaccination in general which may artificially increase coverage rates in control clinics. Authorities focus on the new vaccine may result in supply side bottlenecks for the old vaccines which may limit the programs impact.

Strategies to mitigate this threat:

- Measure attitudes towards vaccination at baseline and endline in both treatment and control sites to monitor this phenomenon.
- Carefully monitor any supply side issues so that they can be explicitly controlled for in the final analysis.

Targeted immunization campaigns or health camps could also skew the results. Mother's whose infants receive a vaccination through a campaign may be less likely to bring their children in for vaccination. An immunization campaign disproportionately targeting control clinics would result in the study overestimating the "status quo" coverage rate.

Strategies to mitigate this threat:

- Randomizing at the clinic level will ensure treatment and control clinics are not overly clustered in any LGA or geographic area which would be the likely targets of campaigns.

## Ethical Considerations

### ***Data Handling and Security***

All raw data will be directly uploaded from enumerator data collection devices to secure encrypted services. The research team will follow strict data management protocols to limit access to the raw data to those who require the data for survey management or initial analysis. Data will be anonymized prior to data dissemination or sharing.

### ***Informed Consent***

With the advice of our data collection partners, we will use locally appropriate informed consent forms which will be administered orally prior to the survey and census. Proper administration of an informed consent protocol will be an important topic in enumerator training. The mothers will provide informed consent on behalf of their infants as is common practice in pediatric studies.

### ***Ethical Clearance***

The survey team will obtain ethical clearance from federal and state officials in Nigeria. The team will also inform local and traditional leaders about the study and seek their approval for undertaking research. At the state-level New Incentives has already informed leaders that survey research will take place alongside the roll-out of their program.

### ***Ethical Risks***

As with any randomized controlled trial, the control clinics will be the subject of research, but not receive the treatment. Since the funder is not willing to commit funds to scale the program across one hundred and fifty clinics based on the existing evidence, limited resources necessitate that some clinics do not receive the treatment. If the program does find positive effects, control clinics will likely be some of the first clinics to receive the treatment. In general, more accurate information on vaccination coverage rates in North West Nigeria will be broadly useful to the government and public health community in their efforts to improve the routine immunization system and thus indirectly benefiting the control group.

The program itself also poses some ethical risks. First, financial incentives may reduce mother's intrinsic motivation to vaccinate their children by monetizing vaccination. The study will also measure attitudes towards vaccination at baseline and endline to determine whether the program seems to be reducing mother's intrinsic motivation to vaccinate. Questions particularly to address the benefits of vaccination or asking mothers directly if they would vaccinate if they lived in a settlement ineligible for an incentive will be added. There is also a

possibility that the program will improve mother's perceptions of vaccination by reducing social taboos or overcoming fears of vaccination through experience.

Another ethical risk imposed by the program is that giving cash to mothers could spark social conflict. We believe this risk is low since other research in Northern Nigeria showed a friend receiving an incentive to vaccinate increased an individual's propensity to vaccinate rather than sparking conflict (Sato and Takasaki 2016). At a community level, we will further mitigate this risk by explaining to local leaders that there were only sufficient resources for some clinics to receive incentives and the team needed to spread out these clinics to better understand the impact of the incentives across different environments.

### Operational Risks

Due to Nigeria's complex operating and regulatory environment, IDinsight determined it would be necessary to partner with a local organization to facilitate data collection. However, the organization IDinsight partners with may fail to produce high quality work within the contracted deadlines. To mitigate this risk IDinsight will draw on its experience working with survey firms in other contexts. IDinsight will create performance-based contracts to ensure organizations are incentivized to deliver high quality work. IDinsight staff will also carefully supervise all aspects of data collection from hiring staff to back-checks. Most of this supervision will take place in the field, and we anticipate IDinsight staff to be located in Nigeria throughout major data collection activities.

Some aspects of the broader Nigerian operating context that would be most likely to impact operations and contributing to the decision to contract a survey firm include the unstable security situation, challenging financial environment, and poor transportation infrastructure.

A change in the security situation could make close supervision of data collection by international staff infeasible. While based on IDinsight's current assessment of the security environment, we are comfortable locating staff in Northern Nigeria to oversee data collection the situation may change, especially by endline.

Exchange rate fluctuations and exchange control policies could make transferring necessary funds into Nigeria much costlier or time consuming than expected. Paying data collection staff through a survey firm will reduce the number of necessary money transfers required, and thus mitigate this risk partially.

Transportation infrastructure in Nigeria is weak. Recently the Abuja airport was closed for six weeks and gasoline shortages are not uncommon. It is likely many treatment and control settlements will be located far from tarred roads. Working with local data collection organizations used to managing field logistics in this environment will reduce the risk of fieldwork being delayed due to transportation delays.

### Timeline

IDinsight recommended baseline measurement begins in August in order to ensure the majority of baseline data collection is completed prior to a potential measles campaign currently slated for October/ November this year. However, the baseline may still be delayed by field events or challenges working with a survey firm. See Annex 6 for more details on this decision to target an August baseline.

Baseline data collection will be phased by state for logistical simplicity and so that each state strata in the study is surveyed at the same time. Finishing any given state or area with an equal number of treatment and control clinics faster will help in reducing the risk that treatment and control clinics' 12 to 16 month olds were born at substantively different times. This will be especially important if the measles campaign begins during data collection. New Incentives will begin operations in treatment clinics in the first state after data collection finishes for that state.

We will use August 2018 administrative data to conduct a midline. The results will be available in late September to inform GiveWell 2018 funding decisions.

Since the study design does not involve tracking individual babies from baseline, the timing of the endline is somewhat flexible. We will schedule the endline based on the date from which New Incentives program clinics achieve normal operations. At pilot sites, New Incentives found that it took a few months for information about the program to spread among the mothers in the community. In the learning sites, New Incentives noticed there is an initial spike in volume caused by mothers bringing older infants who never received BCG to the clinics. Currently, New Incentives estimates this “ramp-up” period should take three to four months. Based on this estimate, the endline data collection should take place in the late spring of 2019 with results available by summer.

### ***Immediate Next Steps***

The immediate next steps leading up to baseline data collection are outlined in the Gantt chart in Annex 6. The most time sensitive next step is survey firm contracting. In order to sign a survey firm contract we will need a funding decision from GiveWell by mid to late June. We will submit a final financial proposal for baseline fieldwork after we have had time to analyze the survey firms’ financial proposals. However, we estimate the cost will be between **\$250,000-350,000**.

IDinsight anticipates it will have staff in Nigeria throughout the next few months overseeing June data collection from learning sites, preparations for the RCT, and RCT data collection. July preparatory activities will include:

- Piloting the full survey
- Translating the full survey and testing translations during piloting
- Hiring key staff
- Developing enumerator training curriculum.
- Hiring enumerators

We expect enumerator training and the start of data collection to take place in August.

## Annexes

### Annex 1 Relevant IDinsight Projects

The project list below includes projects that meet two of these three criteria: took place in Africa, focused on health, and was a randomized controlled trial

Client	Project Description	Africa	Health	RCT
UNICEF Zambia / Zambia MoH (Mama Kits) <sup>20</sup>	IDinsight conducted a cluster randomized controlled trial (RCT) that revealed small, non-monetary mama kit incentives can cost-effectively increase rural facility delivery rates.	✓	✓	✓
Zambia Ministry of Health, Ministry of Community Development, Mother and Child Health <sup>21</sup>	IDinsight conducted a clustered randomized controlled trial to assess two interventions to increase HIV testing on the number of HIV exposed infants identified, number of HIV exposed infants tested for HIV, and percentage of infants immunized.	✓	✓	✓
Zambia's National Malaria Control Centre <sup>22</sup>	Designed and conducted RCT measuring community-level point distribution dynamics for insecticide-treated nets distribution and the impact of CHW hang-up on net usage in rural Zambia.	✓	✓	✓
UNICEF VMMC <sup>23</sup>	Designed and conducted a 3 arm RCT evaluating the impact of two SMS-based campaign interventions to promote uptake of voluntary medical male circumcision on the U-Report platform.	✓	✓	✓
Kangaroo Mother Care	Designed and conducted a detailed situation assessment and baseline study involving clinic administrative data collection and created a detailed evaluation design.	✓	✓	

<sup>20</sup>Wang, P., Connor, A. L., Guo, E., Nambao, M., Chanda-Kapata, P., Lambo, N., & Phiri, C. (2016). "Measuring the impact of non-monetary incentives on facility delivery in rural Zambia: a clustered randomized controlled trial." *Tropical Medicine and International Health*, 21(4), 515-524.

<sup>21</sup> Wang, P. C., Mwango, A., Moberley, S., Brockman, B. J., Connor, A. L., et al. (2015). "A cluster randomised trial on the impact of integrating early infant HIV diagnosis with the expanded programme on immunization and HIV testing rates in rural health facilities in southern Zambia." *PLoS ONE*, 10(10), e0141455.

<sup>22</sup> Wang, P. and Connor, A. L et al. (2016). "Community point distribution of insecticide-treated bed nets and community health worker hang-up visits in rural Zambia: a decision-focused evaluation." *Malaria Journal*, 15, 140.

<sup>23</sup> Leiby, K., Connor, A., Tsague, L., Sapele, C., Kaonga, A., Kakaire, J., and Wang, P. (2016). "The Impact of SMS-Based Interventions on VMMC Uptake in Lusaka Province, Zambia: A Randomized Controlled Trial." *Journal of Acquired Immune Deficiency Syndromes (1999)*, 72(Suppl 4), S269-S277



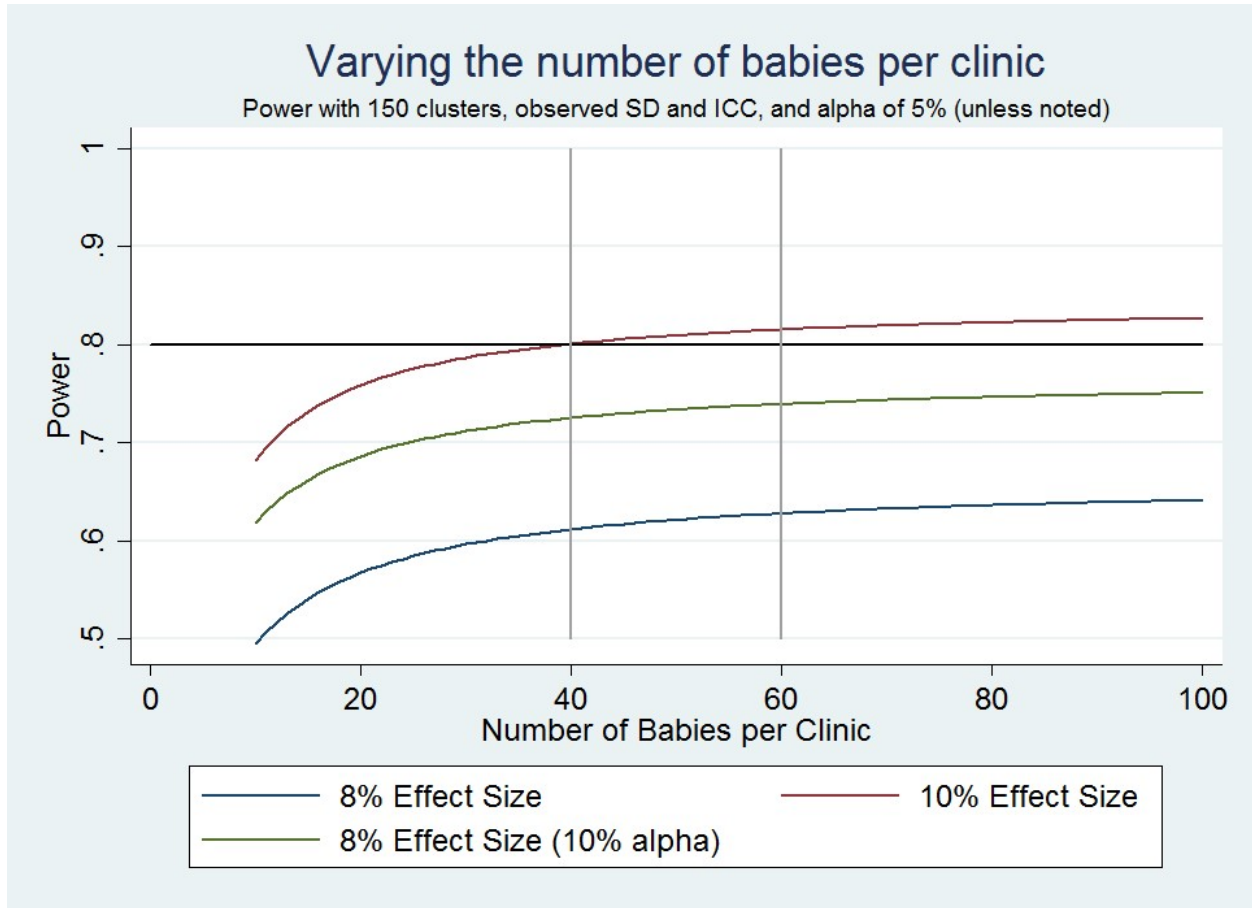
Client	Project Description	Africa	Health	RCT
TechnoServe R&D Coalition	Designing and conducting 7-8 evaluations of innovations that stand to benefit agribusinesses and smallholder farmers. Working with three companies in Mozambique and Uganda.	✓		✓
d.Light Solar Home System	Designed and conducted an RCT to evaluate for d.light, a manufacturer of solar lighting products, in conjunction with USAID DIV.	✓		✓
Cola Life	Designed and conducted a price elasticity assessment for ORS kits using nth-price auctions in rural and urban Zambia.	✓	✓	
UNICEF WASH	Conducting tailored decision-focused evaluations and M&E training for UNICEF WASH teams in Kenya and Philippines.	✓	✓	✓
Esoko	Designed and conducted RCT evaluating whether Esoko's mobile based farmer extension and price information services improved farmer output and loan repayment.	✓		✓
Bill and Melinda Gates Foundation Root Crops Project	Designed and currently conducting an RCT evaluating an innovative program for distributing improved sweetpotato varieties to farmers in Uganda and Tanzania.	✓		✓
Results for Development Measurement and Operational Research Project	Designed and currently conducting a national baseline to measure availability, stocking, and dispensing of amoxicillin dispersible tablets to treat childhood pneumonia in public and private health facilities in Tanzania.	✓	✓	

## Annex 2 Administrative Data Sources

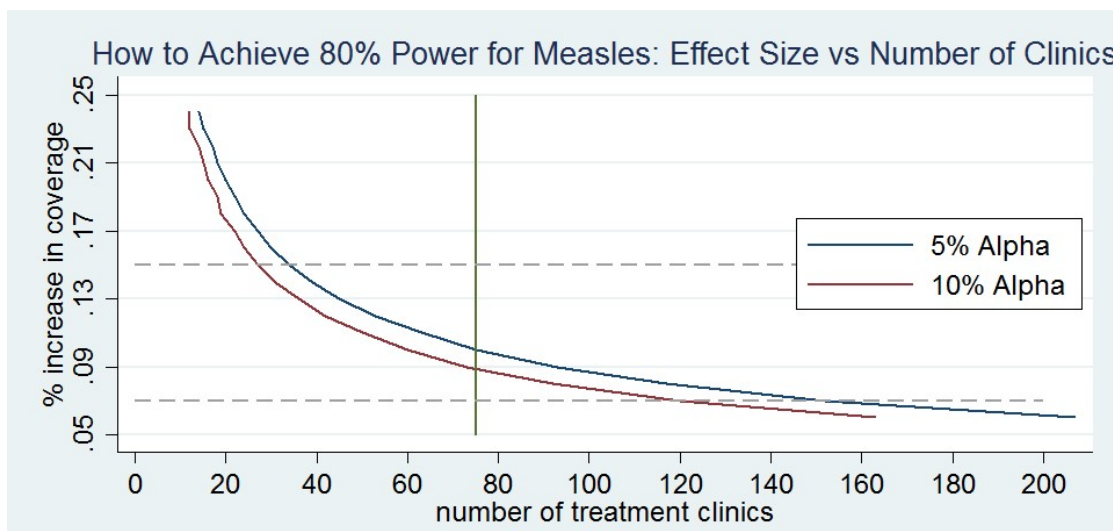
Data Source	Description	Planned Use	Known Issues
Micro-Census	As part of the polio eradication campaign, health workers periodically go house to house to count the number of under-5 children in order to set campaign targets.	We plan to use the micro-census as the denominator in administrative coverage 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 (tally sheet data)	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 infants 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 infant 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 if it's lost. New Incentives adds an All Babies ID sticker to these cards	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 lose the child health cards. Data on cards that have been replaced and transcribed from the child immunization register may be inaccurate.

**Annex 3: Addition Power Calculation Information**

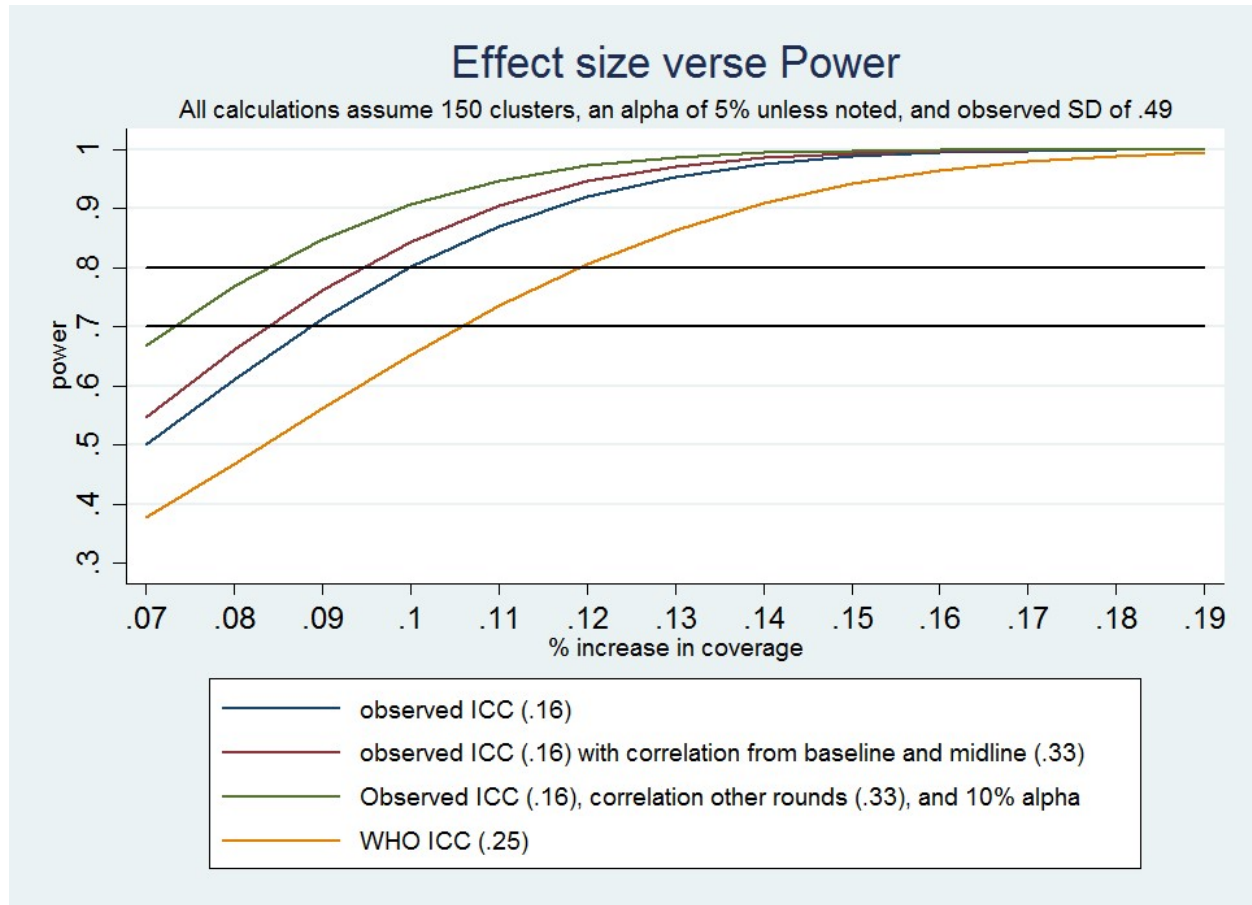
The choice of 40 babies per clinic is illustrated by the following graph. Note that the number of babies per clinic maters even less for higher values of alpha.



The following graph illustrates how many clinics are needed to achieve 80% power at different effect sizes. The dashed grey lines are for a 15% effect and a 7% effect.



The following graph illustrates the minimal impact of correlation from baseline on power. The red line represents a .33 correlation coefficient from baseline covariates which is far higher than we will likely observe.



#### Annex 4: Serological Techniques

IDinsight investigated different bio-marker techniques, and had conversations with a number of key researchers. A core limitation of all serological techniques is vaccine attribution. For measles, current techniques can't differentiate between virus induced immunity and vaccine induced immunity. For the pentavalent vaccination, current techniques can't determine conclusively whether an infant received one or three doses of the vaccine.

The current gold standard technique which is serum blood collection is logistically challenging to collect in low resource settings (Travassos 2015). Oral fluid and dried blood spots techniques involve less invasive sample collection, but still have major challenges. Oral fluid samples are collected through a process analogous to vigorous tooth brushing while the blood for the dried blood spot techniques is collected using capillary blood from finger pricks.

Current oral fluid technology requires a strong cold chain between infant and lab (Emelda Okiro, Personal Communication, April 26, 2017). Other researchers have not found oral fluid produces accurate results in the context of their studies. (Hayford 2013).

While dried blood spot technology doesn't have the same cold chain requirements it still requires relatively sophisticated laboratory procedures and technology to properly analyze the samples. Sometime even experienced labs face difficulties (James Nokes, Personal Communication, April 24, 2017) The research team would likely need to bring international experts and reference sample to Nigeria in order to facilitate a Nigerian lab undertaking a dried blood spot analysis.

There is some emergent technology which may make incorporating bio-marker testing more feasible at endline. In particular, a research consortium is developing an oral fluid and capillary blood point of care tests for measles. Similar, to a pregnancy test enumerators can apply the capillary blood or oral fluid to the test strips directly which will visually indicate the presence of measles antibodies. Field testing of these tests will take place throughout this year (Lenesha Warrenner, Personal Communication, April 27, 2017). Hopefully, by the time of endline data collection there will be sufficient evidence around the accuracy of these tests for IDinsight to consider incorporating them into endline data collection.

The primary advantage of incorporating biomarkers into the endline would be to be an additional source of triangulation when attempting to assess a infant's vaccination status. For example, for infants where only mother reported vaccination status was available biomarker testing could be performed in conjunction with clinic records checks to verify the vaccination status.

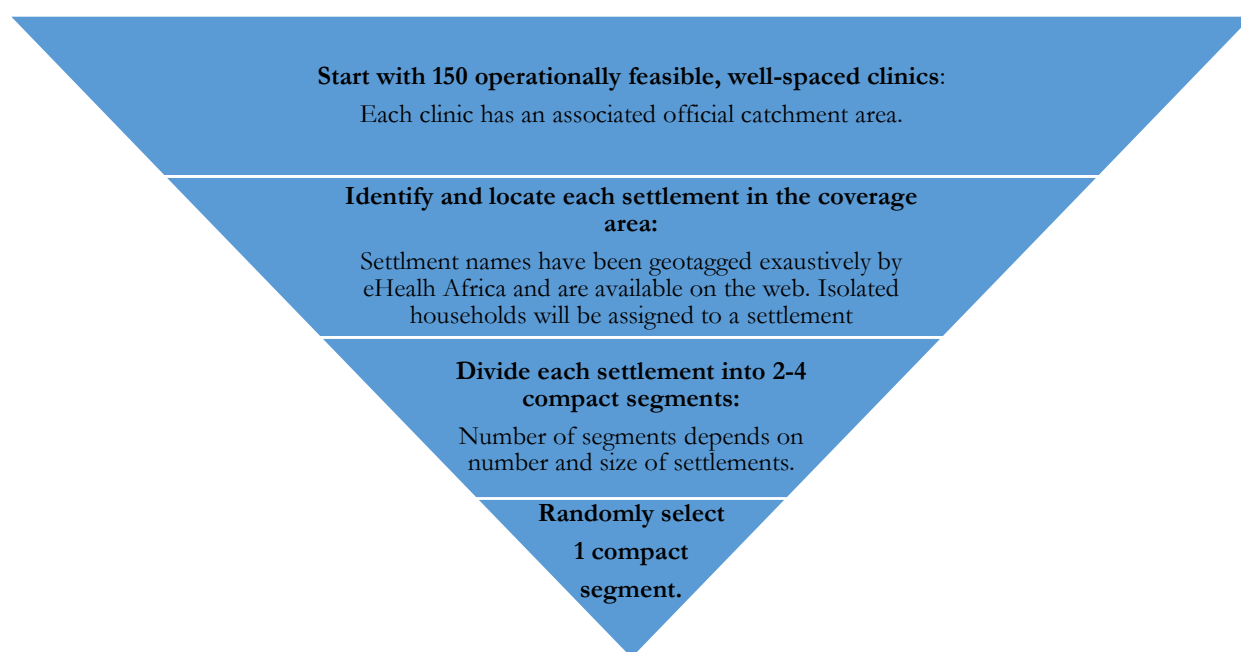
One possible use of biomarker testing at endline would be fraud identification. Biomarker testing would allow for definitive identification of any cases where records showed a vaccination was received in order to receive an incentive, but the infant has no biological indication of vaccination. However, fraud involving infants receiving multiples vaccinations could not be detected in this manner.

## Annex 5: Literature on Sampling Techniques

### **Introduction**

In an effort to reduce the risk of an unbalanced RCT, GiveWell and IDinsight have decided to undertake a baseline coverage survey. A crucial part of any coverage survey is the sampling methodology for a population of infants. Nigeria, like many developing countries, lacks a reliable birth register which could be used for sampling. While an exhaustive census of infants is the most rigorous method for establishing a sampling frame, censuses are generally costly and time consuming. The discussion below examines different options for avoiding this exhaustive census, and the likelihood they may result in an unrepresentative sample. It is worthwhile noting that having both baseline and endline survey rounds will allow the analysis to largely control for any unrepresentative by taking into account baseline characteristics. Figure 4 below illustrates the application of the sampling technique to New Incentives study.

*Figure 4: Compact Segment Sampling in the New Incentives Evaluation*



### **Background**

Given the impracticality of a full census prior to a coverage survey, the World Health Organization came up with a no-census sampling technique as part of its Expanded Program on Immunization (EPI).<sup>24</sup> As the sampling technique for the coverage surveys associated with the program grew in popularity, the term 'EPI' came to refer to the sampling technique as well as the program. Roughly, EPI sampling involves enumerators walking to the center of a village, spinning a pen, and interviewing all households along the imaginary line created by the pen, thus a random walk technique.

As researchers became concerned with an overreliance on EPI random walk sampling, Bennet (1994) used a computer simulation to examine bias at the household level for different sampling schemes. This study was brought about due to the authors noting that in many developing countries, there is no list or map of households to be randomly selected for studies and resources to completely enumerate and map all households in a

<sup>24</sup> For more detail on the traditional coverage survey methodology associated with the EPI program, click [here](#).

community is limited. The main outcome of the study was that modification of the standard EPI that enabled enumerators to visit slightly more households, performed similar to having a full list for surveying.

Despite, EPI being the traditional sampling technique for analyzing immunization coverage in developing countries, other cluster sampling methods and modification have come up and been tested by various researchers. According to Brogan (1994), EPI popularized the use of cluster sample survey design in developing countries for rapid assessment of vaccination coverage. However, the EPI sampling technique has a wider confidence interval and lower accuracy that desired for accurately measuring small changes in coverage.<sup>25</sup> Turner (1996), went a step further to modify the clustering method by introducing compact segmentation. In order to put the standard EPI design on a more solid probability footing, Turner (1996) proposed improving on the EPI random walk technique by creating sub-clusters or 'segments' of approximately equal size, selecting one segment at random, and interviewing eligible members of the various survey defined target groups within all households in selected segments. This is the technique referred to as compact segment sampling (CSS).

Despite critiques from Brogan, Turner, and others, EPI random walk sampling remains popular. Marmamula (2012), reiterated earlier researches concerns about validity of EPI random walk as compared to CSS. For example, the selection of the first household in the cluster when an enumerator starts their random walk may be subject to interviewer bias. The household selection is, therefore, not truly random in a statistical sense and may introduce an unquantifiable degree of bias. EPI favors centrally located households which may be different from those in the periphery of the cluster. Lastly, ignoring households where individuals are not available and excluding non-responders can result in biased estimates, as those who have not responded may be at a higher risk or systematically different from those who have responded to the interview or were examined. These issues are addressed by CSS which removes subjectivity and minimizes a possible bias due to the household selection process followed in the EPI random walk method. CSS also facilitates a re-visit to households when there is no response.

USAID also included segment sampling in the Demographic and Health Surveys (DHS) Sampling Manual of 2012. Segment sampling is noted as an essential extension of probability sampling for household surveys. When faced with large clusters, as is the case with Nigeria, DHS Sampling Manual 2012 recommends segmenting the clusters into several smaller segments and only one of which will be included in the survey and listed. The manual however states that caution needs to be taken when creating the segments as sample segments too small relative to the sample will not ensure that variation in the population is captured. Our proposed segments, are approximately 35% the size of the overall sampling frame mitigating this risk.

### ***Empirical Evidence***

Since introduction of the CSS by Turner (1996), several authors have carried out research to prove its validity.

Chao (2012) used an exhaustive census of small business in a South African township to measure biasness and variability from the mean stemming from different sampling techniques. He found that CSS is relatively unbiased and in some respects produces better results than a true probability sample from the census.

Chao used simulations to calculate the probability any given sampling technique will produce 95% confidence intervals containing the population mean. Chao found the 'standard EPI' method i.e. no revisits (denoted as ESMnre) had an average error rate of 40%, 'segmented sampling method with first and revisits and first level weight' (denoted SSMre in the paper but analogous to CSS) had a 12% error rate, and 'probability sampling method with revisits and probability weight' (PSMrp) had an error rate of 5%. Although a seven percentage

---

<sup>25</sup> Brogan (1994) defined accuracy as 'how a sample estimator is, on the average, to the population parameter being estimated'.

point increase in error is not large, the size of the compact segments relative to the universe may have exacerbated the error rate Chao observed. The segments consisted of only 220 stalls out of a sampling frame of 3,117 stalls.

Both ESMre and SSMre were actually more likely to correctly match the sign and significance of regressions using the full universe of data than PSMrp. PSMrp's under identifies significant relationships because of the greater variance, and thus wider of confidence intervals, of samples obtained using this method. Following from these results, if we assume that including baseline characteristics in the analysis will largely account for sample bias, the compact segment sampling by reducing sample variance may actually produce more precise estimates of New Incentive's effect than a full probability sample of the entire catchment area.

While Chao (2012) focused on small business, other studies have empirically compared sampling techniques for estimating immunization coverage.

Milligan (2004) validated the use of compact segment sampling by comparing it with EPI random walk method. The study aimed at estimating vaccination coverage in western region of Gambia within 3 months of each other in 2000–2001. The study found that point estimates for vaccination coverage from the two surveys rarely differed by more than 2%. The slight difference was attributed to household selection. The study concluded that compact segment sampling is generally preferred as it ensures objectivity in household selection and permits the estimation of population totals (such as those unvaccinated). The authors also note that while the EPI survey team in their study was highly trained and disciplined this likely not the case in all EPI studies, and that CSS had the advantage that enumerator compliance to the sampling plan can be easily monitored.

Luman (2007) compared EPI random walks to systematic random sampling (SystRS) in both rural and peri-urban Ethiopia. SystRS appeared to perform better resulting in significantly different estimated coverage rates<sup>26</sup>. The fact that Luman found disagreement between SystRS and EPI while Milligan (2004) found agreement between EPI and compact segment sampling does not mean compact segment sampling would have necessarily disagreed with SystRS if implemented in the context of the Luman study. Living near the center of a cluster probably has a greater effect on whether a child will be vaccinated by a supplementary immunization campaign, the context of Luman (2007), than whether a child will be routinely immunized, the context of Milligan (2004). However, the study adds empirical evidence to the theoretical arguments that EPI sampling is not sufficiently accurate, and a new standard should be developed. IDinsight considered SystRS as a sampling strategy for the evaluation, but was concerned about ensuring enumerator compliance to the strategy and decided against it. CSS as strategy works well to tackle both these concerns and thus was a preferred methodology.

### ***Application***

There is has been an increase in the use of compact segment sampling in the field. The table below contain examples of research papers which apply versions of compact segment sampling in varying contexts.<sup>27</sup>

---

<sup>26</sup> Coverage point estimates generated by the two techniques differed by up to 5 percentage points, but the 95% confidence intervals generally overlapped

<sup>27</sup> However, it is worthwhile noting that although researchers are applying compact segment sampling, they do not explicitly refer to it as such and thus making it difficult to trace and refer to the paper.



Study	Method	Sample Characteristics
<b>Replication RCT Evaluating Immunization Incentives and SMS Reminders Program at Scale</b>	The study covers a total of 140 Primary Health Centres (PHCs) across 6-7 low performing districts of the Indian state of Haryana. From the villages covered by each of the 140 PHCs, 7 villages per PHC are randomly sampled <sup>28</sup> . Within each of the 980 villages, 15 eligible households i.e. households with children between 0-36 months are sampled from a complete census of the village.	Ongoing study funded partially by GiveWell
<b>Technical Description of the Health and Retirement Survey (HRS) Sample Design</b>	The HRS sample is selected under a multi-stage area probability sample design. The primary stage of sampling involves probability proportionate to size (PPS) selection of U.S. Metropolitan Statistical Areas (MSAs) and non-MSA counties. The next stage is a second stage sampling of area segments (SSUs) within sampled primary stage units (PSUs). The third stage of sample selection is preceded by a complete listing (enumeration) of all housing units (HUs) that are physically located within the bounds of the selected SSU. The third sampling stage is a systematic selection of housing units from the HU listings for the sample SSUs. The final stage is the selection of the household financial unit within a sample HU.	N/A
<b>A National Survey of Musculoskeletal Impairment in Rwanda: Prevalence, Causes and Service Implications</b>	A nationally representative sample of the population was selected through cluster sampling with probability proportionate to size. A list was produced of all the enumeration areas and their respective populations, and a column was created with the cumulative population across the settlements. The total population (i.e. 8,441,000) was divided by the number of clusters required (i.e. 105) to derive the sampling interval (i.e. 80,390). The first cluster was selected by multiplying the sampling interval with a random number between 0 and 1. The resulting number was traced in the cumulative population column and the first cluster was taken from the corresponding enumeration area. Households within clusters were selected through compact segment sampling. Maps of each selected cluster (i.e. enumeration area) were obtained. These maps included the locations of the head of ten-household communities, thus showing approximate population distribution. The enumeration area was visited 2-3 days before the survey and the village leaders were asked to update the map. IDinsight plans to pilot the accuracy of satellite maps in the June study. The enumeration area was then divided into segments, so that each segment included approximately 80 people. One of the segments was chosen at random by drawing lots and all households in the segment were included in the sample sequentially until 80 people were identified.	A total of 8368 individuals were enumerated and 6757 were screened (Response rate = 80.8%), 1596 (19.1%) were absent, 10 (0.1%) refused and 5(0.1%) were unable to communicate. The response rate was higher in women (84.8%) than in men (76.3%). The age- and gender-distribution of the sample was near largely similar to that of the population.

<sup>28</sup> Across Haryana state, the location of the study, there were 14.8 villages on average per primary healthcare facility according to the 2014-2015 rural health statics report.

Study	Method	Sample Characteristics
<b>Rapid Assessment of Avoidable Blindness in Nakuru District, Kenya</b>	<p>Electoral role data was used as the sampling frame for this survey. A list was produced of polling stations and their respective population sizes. The approximate population size of those age <math>\geq 50</math> (481,051) was divided by the 76 clusters required to derive a sampling interval of 6330. The first cluster was selected by multiplying the sample interval by a random number between 0 and 1. The resulting number was traced in the cumulative population column and the first cluster was taken from the corresponding polling station. The other clusters were obtained by adding the sampling interval to the previous number. This systematic sampling procedure is random and selects clusters with a probability proportional to the size of the population. The second stage was through compact segment sampling. The polling stations were visited 2-3 days before the survey and village elders produced sketch maps of polling areas showing landmarks with approximate distribution of villages and households. The polling area was then divided into segments so that each segment included approximately 50 people aged <math>\geq 50</math> years. One of the segments was selected at random by drawing lots and all households in the segment were included in the sample sequentially until 50 people <math>\geq 50</math> years of age were identified.</p>	<p>The study population consisted of 3784 people but 3503 (92.6%) were responsive. There was no difference in mean ages of those who were unresponsive and those who were responsive. There were however more non-responsive people were females. The sample of 3503 examined included 1669 men (47.6) and 1834 women (52.4%). There was a slight overrepresentation of elderly people (<math>\geq 80</math>) in the sample, particularly elderly women.</p>
<b>Rapid assessment of avoidable blindness and needs assessment of cataract surgical services in Satkhira District, Bangladesh</b>	<p>64 clusters of 50 adults aged <math>\geq 50</math> years were required for this survey. This survey was part of a larger research project, and 106 clusters were selected. The clusters were selected through probability-proportionate to size sampling using updated data from the 1991 national census as the sampling frame. Households in clusters were selected by a modification of compact segment sampling. The enumeration area was visited 2–3 days before the survey, and the village leaders were asked whether they could produce a sketch map of the enumeration area showing major landmarks and the approximate distribution of households. On the day of the survey, the enumeration area was divided into segments, so that each segment included about 50 people aged <math>\geq 50</math> years. One of the segments was chosen at random by drawing lots and all households in the segment were included sequentially until 50 people aged <math>\geq 50</math> years were identified.</p>	<p>4,868 people (91.9%) were included in the survey. The sampled population was relatively representative of the district population in terms of age and sex distribution, although women were slightly over-represented in the sample.</p>

### ***Conclusion***

Given the large sampling frame of the New Incentives evaluation, conducting a census of every household in every clinics catchment area would contribute only marginally to the studies overall rigor and increase logistical

and survey management risks substantially. Such a census would also be unusual as vaccination and other coverage surveys usually use techniques to avoid a full census. Of these techniques, the literature best supports compact segment sampling. In general, the applied literature suggests that the likelihood of CSS leading to a sample unrepresentative to the point of biasing the overall study is very low. CSS will simplify survey logistics massively reducing operational risks to survey quality while contributing to the risk of a biased impact estimate only minimally if at all.

### ***Bibliography for Annex 5***

- Atijosan, O., Rischewski, D., Simms, V., Kuper, H., Linganwa, B., Nuhi, A., Foster, A. and Lavy, C., 2008. A national survey of musculoskeletal impairment in Rwanda: prevalence, causes and service implications. *PLoS One*, 3(7), p.e2851.
- Bennett, S., Radalowicz, A., Vella, V. and Tomkins, A., 1994. A computer simulation of household sampling schemes for health surveys in developing countries. *International journal of epidemiology*, 23(6), pp.1282-1291.
- Chao, L.W., Szrek, H., Peltzer, K., Ramlagan, S., Fleming, P., Leite, R., Magerman, J., Ngwenya, G.B., Pereira, N.S. and Behrman, J., 2012. A comparison of EPI sampling, probability sampling, and compact segment sampling methods for micro and small enterprises. *Journal of development economics*, 98(1), pp.94-107.
- Heeringa, S.G. and Connor, J.H., 1995. Technical description of the Health and Retirement Survey sample design. *Ann Arbor: University of Michigan*.
- ICF International. 2012. Demographic and Health Survey Sampling and Household Listing Manual. MEASURE DHS, Calverton, Maryland, U.S.A.: ICF International.
- J-Pal South Asia, 2016. Replication RCT Evaluating Immunization Incentives and SMS Reminders Program at Scale. *GiveWell Incubation Grant*.
- Luman, E.T., Worku, A., Berhane, Y., Martin, R. and Cairns, L., 2007. Comparison of two survey methodologies to assess vaccination coverage. *International Journal of Epidemiology*, 36(3), pp.633-641.
- Marmamula, S., Keeffe, J.E. and Rao, G.N., 2012. Rapid assessment methods in eye care: An overview. *Indian journal of ophthalmology*, 60(5), p.416.
- Mathenge, W., Kuper, H., Limburg, H., Polack, S., Onyango, O., Nyaga, G. and Foster, A., 2007. Rapid assessment of avoidable blindness in Nakuru district, Kenya. *Ophthalmology*, 114(3), pp.599-605.
- Milligan, P. and Bennett, S., 2004. Comparison of two cluster sampling methods for health surveys in developing countries. *International Journal of Epidemiology*, 33(3), pp.469-476.
- Mörchen, M., Langdon, T., Ormsby, G.M., Meng, N., Seiha, D., Piseth, K. and Keeffe, J.E., 2015. Prevalence of Blindness and Cataract Surgical Outcomes in Takeo Province, Cambodia. *The Asia-Pacific Journal of Ophthalmology*, 4(1), pp.25-31.
- Turner, A.G., Magnani, R.J. and Shuaib, M., 1996. A not quite as quick but much cleaner alternative to the Expanded Programme on Immunization (EPI) Cluster Survey design. *International journal of epidemiology*, 25(1), pp.198-203.

Wadud, Z., Kuper, H., Polack, S., Lindfield, R., Akm, M.R., Choudhury, K.A., Lindfield, T., Limburg, H. and Foster, A., 2006. Rapid assessment of avoidable blindness and needs assessment of cataract surgical services in Satkhira District, Bangladesh. *British Journal of Ophthalmology*, 90(10), pp.1225-1229.

## Annex 6: Timing of the Baseline

IDinsight recommended beginning baseline measurement in August in order to ensure the majority of baseline data collection completes prior to a potential measles campaign currently slated for October/ November this year. While we can adapt survey tools to try to disaggregate campaign vaccination from vaccinations received during the course of routine immunization, the reliability of the disaggregation will decrease with time between the campaign and a mother being surveyed.<sup>29</sup> We will continue to work with GiveWell, New Incentives, and the selected survey firm to balance finishing data collection before the campaign with ensuring high quality data collection. Final work planning decisions around the duration of baseline surveying can be made once more information is known about final campaign timing. Targeting an August baseline does pose some risks in terms of forcing an accelerated survey firm contracting process. However, the June fieldwork should provide a good indication of survey firm capacity, and whether a delay is necessary to contract another firm. One advantage of targeting an August start is that the consequences of any delays resulting from working with a survey firm will be less dramatic. The table below illustrates the tradeoffs involved in deciding when to target baseline data collection.

*A Comparison of Baseline timing and Sampling Options*

Scenario	Target August baseline with compact segment sampling	Target September baseline with compact segment sampling	Conduct full census with 20-30 enumerators	Conduct full census with 80 enumerators
<b>Theoretical Rigor<sup>30</sup></b>	Medium: Accepted practice, but small risk of an unrepresentative sample	Medium: Accepted practice, but small risk of an unrepresentative sample	High: Full census minimizes risk of an unrepresentative sample	High: Full census minimizes risk of an unrepresentative sample
<b>Expected data quality</b>	High: good scope for oversight, and limited possibility of measles campaign confusion.	Medium: risk of campaign confusing recall and clinic records.	Medium: good oversight, but data collection will overlap campaign	Low: poor scope for oversight, and for option 5 high risk of campaign confusion
<b>Additional Factors</b>	Greater flexibility to delay if there are issues with the survey firm or piloting	Starting later will ensure more time to vet the survey firm and pilot different instrument.	Extremely expensive (~\$100-150K more) and will take a long time, delaying NI's expansion and the RCT results	Higher likelihood of delays if need to hire 60 enumerators. May be more difficult to find high quality <sup>31</sup> enumerators as well.

<sup>29</sup> Conversations with Hanovia about the 2015 post-measles campaign survey revealed that with a short time lag, qualitatively, most mothers were able to recall the campaign clearly. However, card retention was only 60% nationally.

<sup>30</sup> Increasing precision of baseline estimates will only have limited impact on the study's overall balance which is largely guaranteed through randomness. Furthermore, for an unrepresentative sample to bias the results the sample must be unrepresentative in a way that interacts with a program's treatment effect. The characteristics that simply effect vaccination coverage in general will be captured by the baseline coverage estimate for any given compact segment.

<sup>31</sup> Due to cultural norms in Zamfara and Katsina states we will need to hire female enumerators to conduct interviews with mothers. This constraint could make hiring large numbers of qualified local enumerators challenging.

## Bibliography

- Abhijit Vinayak Banerjee, Esther Duflo, Rachel Glennerster, and Dhruva Kothari, 2010. “Improving immunization coverage in rural India: clustered randomized controlled evaluation of immunization campaigns with and without incentives.” *BMJ*; 340: c2220
- Benjamin P. Loevinsohn, Michael E. Loevinsohn, 1986. “Improvement In Coverage Of Primary Health Care In A Developing Country Through Use Of Food Incentives.” *The Lancet* pp. 1314-1316
- Chandir, S; Khan, AJ; Hussain, H; Usman, HR; Khowaja, S, 2010. “Effect of food coupon incentives on timely completion of DTP immunization series in children from a low-income area in Karachi, Pakistan: A longitudinal intervention study.” *Vaccine*; 28 pp. 3473-3478.
- Mark A. Travassos, Berhane Beyene, Zenaw Adam, James D. Campbell, Nigisti Mulholland, Seydou S. Diarra ,Tassew Kassa, Lisa Oot, Jenny Sequeira, Mardi Reymann ,William C. Blackwelder, Yukun Wu, Inna Ruslanova, JayaGoswami1, Samba O. Sow, Marcela F. Pasetti, Robert Steinglass, Amha Kebede, Myron M. Levine, 2016. Immunization Coverage Surveys and Linked Biomarker Serosurveys in Three Regions in Ethiopia. *PLoS ONE* 11(3): e0149970
- Mylene Lagarde, Andy Haines, Natasha Palmer, 2007. “Conditional Cash Transfers for Improving Uptake of Health Interventions in Low- and Middle-Income Countries.” *JAMA*; 298, no. 16: 1900-1910
- National Population Commission (NPC) [Nigeria] and ICF International. 2014. *Nigeria Demographic and Health Survey (DHS) 2013*. Abuja, Nigeria, and Rockville, Maryland, USA: NPC and ICF International.
- National Routine Immunization Strategic Plan (NRISP) Advisory Committee. 2013. *National Routine Immunization Strategic Plan 2013-2015: Intensifying Reaching Every Ward through Accountability*. Abuja, Nigeria: Federal Ministry of Health – Nigeria.
- Nigeria Centre for Disease Control (NCDC). 2016. *Weekly Epidemiology Report*. Abuja, Nigeria: Federal Ministry of Health – Nigeria.
- Ryoko Sato, Yoshito Takasaki, 2016. “Peer Effects on Vaccination: Experimental Evidence from Rural Nigeria.” CIRJE Discussion Papers-F-1002.
- Rajni Gunnala, Ikechukwu U. Ogbuanu1, Oluwasegun J. Adegoke, Heather M. Scobie, Belinda V. Uba, Kathleen A. Wannemuehler, Alicia Ruiz, Hashim Elmousaad, Chima J. Oluabunwo, Mahmud Mustafa, Patrick Nguku, Ndadilnasiya Endie Waziri, John F. Vertefeuille, 2016. Routine Vaccination Coverage in Northern Nigeria: Results from 40 District-Level Cluster Surveys, 2014-2015. *PLoS ONE*
- UNICEF and National Bureau of Statistics. 2015. *National Nutrition and Health Survey (NNHS) 2015 Report on the Nutrition and Health Situation of Nigeria*. Abuja, Nigeria: National Bureau of Statistics
- Wang, PC, AL Connor,E Guo, M Nambao, P Chanda-Kapata, N Lambo, & C Phiri, 2016. “Measuring the impact of non-monetary incentives on facility delivery in rural Zambia: a clustered randomized controlled trial.” *Tropical Medicine and International Health*, 21(4), 515-524

- Wang PC, Mwango A, Moberley S, Brockman BJ, Connor AL, Kalesha-Masumbu P, Mutembo S, Bweupe M, Chanda-Kapata P, Biemba G, Hamer DH, Chibuye B, McCarthy E, 2015. “A cluster randomised trial on the impact of integrating early infant HIV diagnosis with the expanded programme on immunization and HIV testing rates in rural health facilities in southern Zambia.” *PLoS ONE*, 10(10), e0141455.
- Wang PC, Connor AL, Joudeh AS, Steinberg J, Ndhlovu K, Siyolwe M, Ntebeka B, Chibuye B and Hamainza B. 2016. “Community point distribution of insecticide-treated bed nets and community health worker hang-up visits in rural Zambia: a decision-focused evaluation.” *Malaria Journal*, 15, 140
- Leiby, K., Connor, A., Tsague, L., Sapele, C., Kaonga, A., Kakaire, J., and Wang, P. (2016). The Impact of SMS-Based Interventions on VMMC Uptake in Lusaka Province, Zambia: A Randomized Controlled Trial. *Journal of Acquired Immune Deficiency Syndromes* (1999), 72 (Suppl 4), S269–S277
- Hayford, Kyla T, Mohammed S Shomik, Hassan M Al-Emran, William J Moss, David Bishai, and Orin S Levine, 2013. Measles vaccination coverage estimates from surveys, clinic records, and immune markers in oral fluid and blood: a population-based cross-sectional study. *BMC Public Health* 2013, 13:1211
- Warrener, Lenessa, Rimantas Slibinskas, Kaw Bing Chua, Wondatir Nigatu, Kevin E Brown, Kestutis Sasnauskas, Dhanraj Samuel & David Brown, 2011. A point-of-care test for measles diagnosis: detection of measles-specific IgM antibodies and viral nucleic acid. *Bulletin of the World Health Organization*; 89:675–682.
- Ohuma E. O., E. A. Okiro, A. Bett, J. Abwao, S. Were, D. Samuel, A. Vyse, N. Gay, D. W. G. Brown and D. J. Nokes. 2009. Evaluation of a measles vaccine campaign by oral-fluid surveys in a rural Kenyan district: interpretation of antibody prevalence data using mixture models. *Epidemiological Infection*, 137, 227–233