

# **HHS Public Access**

Author manuscript *J Infect Dis.* Author manuscript; available in PMC 2018 July 01.

Published in final edited form as:

J Infect Dis. 2017 July 01; 216(Suppl 1): S237–S243. doi:10.1093/infdis/jix125.

## Individually Linked Household and Health Facility Vaccination Survey in 12 At-Risk Districts in Kinshasa Province, Democratic Republic of Congo: Methods and Meta-data

Eleanor Burnett<sup>1</sup>, Kathleen Wannemuehler<sup>1</sup>, Guillaume Ngoie Mwamba<sup>2</sup>, Masembe Yolande<sup>3</sup>, Kaya Guylain<sup>2</sup>, Nzazi Nsambu Muriel<sup>3</sup>, Nzuzi Cathy<sup>2</sup>, Tshekoya Patrice<sup>2</sup>, Karen Wilkins<sup>1,3</sup>, and Norbert Yoloyolo<sup>2</sup>

<sup>1</sup>US Centers for Disease Control and Prevention, Atlanta, USA

<sup>2</sup>Ministry of Health, Kinshasa, DRC

<sup>3</sup>World Health Organization, Kinshasa, DRC

## Background

Household (HH) and health facility (HF) data can complement each other to provide a better understanding of the demand- and supply-side factors that contribute to infant vaccination status [1, 2, 3]. Many surveys of vaccination services are limited to either HH or HF, limiting conclusions to only HH or HF factors and not allowing for interaction between HH and HF factors. Alternatively, surveys can be designed to capture both HH and HF in the same population, but there are currently few surveys that record and present data from both settings together [4]. It is not uncommon to supplement home-based vaccination records with HF records for individual children in vaccination coverage surveys [5]; however, other information from the HF and vaccinators about service provision practices and knowledge is often not solicited. Several methods can be used to link HF data to a HH survey, including linking contemporary independent surveys in the same geographic areas [6, 7], surveying all HFs in HH survey sampling units [8], and individually linking HHs to visited HF [9, 10]. Each method has benefits and limitations in terms of feasibility, representativeness, and assumptions of geography and temporality.

The Province of Kinshasa includes the capital of the Democratic Republic of Congo (DRC), a megacity with an estimated population of 11.6 million [11]. In 12 of 36 administrative districts (zones), 2014 administrative data estimated coverage as <80% for the first dose of diphtheria-tetanus-pertussis-hepatitis B-*Haemophilus influenzae* type b (pentavalent) vaccine; drop-out (i.e., children who received the first dose of pentavalent vaccine but did not receive the third dose) was estimated as >10% [12]. Recent measles outbreaks [13] and

**Corresponding author:** Eleanor Burnett, Centers for Disease Control and Prevention, 1600 Clifton Rd NE, MS A-04, Atlanta, GA 30329-4027, wwg7@cdc.gov, Phone: +1-404-639-8252.

Guillaume Ngoie Mwamba is currently affiliated with the Agence de Medecine Preventative.

The authors either do not have a commercial or other association that might pose a conflict of interest;

Funding for this survey was provided by the IMG;

These results have not previously been presented;

the introduction of inactivated polio vaccine (IPV) into the routine infant immunization schedule in April 2015 [14] accentuated the need to identify reasons for sub-optimal coverage and drop-out for programmatic decision making. To meet this need, we planned a multi-faceted program evaluation that included objectives for infants 6–11 months of age and children 12–23 months of age in the 12 zones with reported low vaccination coverage within Kinshasa Province [Figure 1]. The surveys were done concurrently to maximize efficiency of resource use and to capture relevant data for the Expanded Programme on Immunization (EPI). For the 6–11 month old infants, we chose to use an individual linking method to assess HH and HF factors associated with up-to-date for age vaccination. In this paper, we discuss the design, implementation, analytical methods, and descriptive meta-data results for this survey, including challenges and lessons learned; the data analysis and interpretation of coverage survey results will be presented in a forthcoming publication.

## Survey objectives and sample size

#### Linked household and health care facility survey: 6-11 month old infants

For the survey of 6–11 month old infants, the primary objective was to assess the associations between vaccination status, specifically up-to-date for age, and both demandside (HH survey) and supply-side (HF survey) factors by individually linking the records from a HH survey with data from a survey of HF where the infants received their most recent vaccinations. This age group was chosen because these infants accessed, or should have accessed, DRC EPI services in the last 3–6 months; this timeframe ensures that data collected at HFs was contemporaneous to the infant's experiences and limits possible bias because of local record keeping practices, staff turnover, and the limitations of recall. Our choice of this age group was intended to increase the potential for timely remedial programmatic action.

Estimating target sample size without a priori knowledge of the strength of any associations is not straightforward in a multilevel model context [15, 16]. We simplified the calculation by powering the study for bivariate associations at the zonal level, with the objective of pooling across all 12 zones to build a multivariable model. For this age group, the sample size was calculated to detect a 20% difference in the proportion of infants up-to-date for age between two sub-populations of equal size. We estimated that 180 infants per zone were needed for the univariate analysis at the zonal level assuming 60% and 80% coverage in the two sub-populations, 80% power, alpha = 0.05, and ~9% non-response. (Pearson Chi-square test of independence, SAS v 9.3, PROC POWER). From the member line list available from the 2013–14 Demographic and Health Survey (DHS) [17], we estimated that 8% of HHs would have a 6–11 month old infant. Using a binomial distribution with probability = 8%, and 80% power to achieve the target sample size of 180, we estimated we would need to approach 2,400 HHs per zone. We planned to sample 15 neighborhoods per zone, therefore needed to approach 160 HHs per cluster. In houses with more than one 6–11 month old infant, one was randomly selected.

A HF was included if it was identified by the caregiver of a 6–11 month old as the child's most recent place of vaccination. Because of feasibility concerns, we decided a priori to limit the HF survey to those in the 12 study zones, accepting a loss to the linked sample size.

The EPI system in Kinshasa consists of public, private, and faith-based HFs. The available lists of the number, the type, or the locations of HFs within the 12 study zones were deemed incomplete and unreliable, making an independent survey of HFs impossible. The high density urban setting also made it difficult to identify all HFs.

#### Household survey: 12-23 month old children

The primary objective of the survey of 12–23 month old children was to estimate vaccination coverage in the combined geographic area of the survey. If we identified an average of 6 children 12–23 months of age from each of the 180 clusters, the estimated precision for a coverage estimate of 50% would be 4–4.5%, assuming an intra-class correlation of 0.167, a 95% probability of achieving the desired precision, and 10% non-response (95% Wald confidence interval). We expected 16% of HHs would have a child 12–23 months of age, so determined that we would select every fourth house approached for the 6–11 month old survey. In HHs with more than one 12–23 month old child, one was randomly selected.

## Survey design and implementation

The surveys were designed as stratified 3-stage cluster surveys. Surveyors had previously participated in polio campaigns and monitoring in Kinshasa and were trained for 5 days prior to beginning field work.

#### **Cluster selection**

Zones were defined using administrative boundaries provided with a sampling frame from the DRC EPI. A list of neighborhoods with estimated target population were provided for each of the 12 zones of interest based on polio micro-plan data compiled in 2013–2014. Within each of the 12 zones, 15 neighborhoods were sampled using systematic probability proportional to estimated size (PPES) methods, with the list sorted by *aire de santé* (HF catchment area) to spread the sample geographically across the zone.

#### **Household selection**

After selection of the neighborhoods, and a short time before field work was scheduled to begin, it was determined that the neighborhoods listed in the sampling frame had unknown boundaries, making it difficult to use neighborhoods to inform sampling of HHs. Therefore, an alternative strategy was created whereby random starting points were selected using Arc GIS v10.2, to correspond to the number of selected neighborhoods in each *aire de santé*. The completion of each neighborhood was scheduled to take 3 days, limited to daylight hours for reasons of practicality and security. A HH was defined as a group of people that eat together and sleep together. For HHs where occupants were absent, attempts were made to schedule 2 revisits through neighbors during the 3 scheduled days; revisits were not made after the end of the scheduled time.

Verbal informed consent was obtained from each selected child's parent or guardian; the selected child's parent or guardian was interviewed. Vaccination administration information

was recorded from verbal report and written documentation from home-based records when available.

#### Field implementation of two target populations

From a design perspective, including two unique target populations with different target numbers was not complicated. For implementation, however, it was challenging to train data collectors to become familiar with carrying and completing two different forms for the two target populations, and to concurrently enroll 12–23 month olds in every fourth house. In 37 cases, the calculated age of the child based on recorded data during data analysis revealed that the wrong form was completed; 43 children aged 12–23 months were interviewed in a HH that was not the fourth; and in 14 HHs, a child from each age group was identified for interview, taking longer to complete the interview.

#### Linked health facility selection

Respondents in the 6-11 month old surveys were asked to identify the HF where the selected infant received their most recent vaccinations. Respondents for infants who were unvaccinated were asked to identify the HF where they had received curative care services most recently. The identified health facilities were compiled in an Excel spreadsheet; an attempt was made to collate a list of all infants attending the same HF to streamline data collection at the HF and to ensure unique identifiers for linking. The number of HFs, large geographic area of Kinshasa, existence of formal and informal names for the same HF, and non-standard orthography made this task challenging. Mail Merge Wizard in Microsoft Word was then used to generate register abstraction forms from the Excel spreadsheet for each infant; the register abstraction form included both the HH and HF unique identifiers. We attempted to survey all identified, operational HFs located in one of the 12 zones of interest at the time of data collection. At each HF, informed consent was obtained and the individual who led vaccination services was interviewed using a structured questionnaire. Where a vaccination register was available and was organized by infant, vaccination administration records for the associated infants were documented from the vaccination register.

## Meta-data Results

The HH and HF surveys were conducted from August 31 to September 22, 2015; remaining health facilities were surveyed from November 23 to December 3, 2015.

#### Household enrollment and data collection

Across the 12 zones, 2,409 HHs with one or more children 6–23 months of age were identified out of 28,800 total HHs approached. In total, residents of 1,920 HHs (80%) participated in the surveys and 86 HHs (4%) refused to participate. Additional HHs were not included because the family had lived in the neighborhood for fewer than 3 months or the child's primary caregiver were unavailable after 2 visits by an interview team. We excluded additional HHs that were interviewed but subsequently determined to be ineligible because the child was not in the target age group (71 children) or the child's month or year of birth were unknown (11 children). Thirty children were interviewed as 12–23 month olds, but

were later determined to be 6–11 months of age at the time of the interview; seven children were interviewed as 6–11 month olds, but were later determined to be 12–23 month olds. These children are categorized as 'interviewed' by their actual age group, though the 12–23 month old questionnaire had only a subset of questions.

By zone, the number of HHs identified with a 6–11 month old infant ranged from 112 to 138 and the number of interviewed HHs ranged from 90 to 127, much lower than the target of 180 [Table 1]. Overall, caregivers in 81% of HHs identified for the 6–11 month old survey were interviewed, providing data for 1,224 infants, 62% of the protocol's sample size.

By zone, the number of households identified with a 12–23 month old child ranged from 57 to 102 and the number of interviewed HHs ranged from 42 to 76 [Table 1]. Overall, caregivers in 77% of HHs identified for the 12–23 month old survey were interviewed, providing data for 710 children, 66% of the protocol's sample size.

Home-based vaccination cards were available and seen by the interviewer at the time of the survey for 63% of all children surveyed, 73% of 6–11 month old infants (n=845), and 59% of 12–23 month old children (n=368) [Figure 2]. The percentage of infants and children with a home based health card that had been seen by interviewers was 68% in the three zones that were surveyed during the first three days of data collection (earliest zones) and 58% in the three zones that were surveyed during the last three days of data collection (latest zones). The percentage of infants and children reported to have ever been issued a health card was 93% in the earliest zones and 95% in the latest zones; the percentage of infants and children reported to have at least one health card at the time of the survey was 88% in the earliest zones and 91% in the latest zones.

#### Health facility enrollment, linkage, and data collection

The expected number of HFs was unknown before the survey; 182 HFs located within the 12 zones of interest and cited by caregivers of the enrolled 6-11 month old infants as the most recent facility attended for vaccination or curative care were enrolled in the survey. Two hundred and seventy nine infants were linked to a HF that was located outside the 12 zones of interest, which was no longer operational at the time of the survey, or where the staff declined to participate. The number of surveyed HFs located in each zone ranged from 7 to 21. Of the 1,224 infants 6–11 months of age who participated in this survey, 879 (72%) infants were linked to an interviewed HF located in one of the 12 zones of interest [Figure 3], with a range of 54 to 101 infants by infant zone of residence. The median number of surveyed infants who were linked to each interviewed facility was 4 (range 1 to 33). Caregivers of six infants reported that their child never received curative or preventative healthcare services. For 51 infants, including the 30 that were interviewed using the 12-23 month old form, there was not adequate information to identify the HF; 9 infants were linked to a HF in the 12 zones of interest and had vaccination history abstracted from the register, but the facility was not interviewed. The total number of unique health facilities ranged from 20 to 53 by infant zone of residence. Overall, 55% of infants attended a HF located in their zone of residence. All of the interviewed HF reported providing vaccination services.

Of the 879 infants ages 6–11 months linked to a HF in the 12 Zones, 518 (59%) were found in the register at their facility and vaccination history was abstracted. Of the remaining infants, 316 (36%) were not found in the register. The register abstraction forms for 45 infants were not available.

## **Analytic methods**

#### Linked analysis

To account for clustering of children within neighborhoods (survey cluster) and clustering of children within HFs, and to incorporate covariates at the child level (child, mother, HH characteristics) and the HF level, the method of choice is generalized linear mixed models [18]. The data from Kinshasa consist of cross-classification, whereby children within the same neighborhood go to different HFs and HFs provide services to children from different neighborhoods. The cross-classification and binary outcome (up-to-date for age) can be handled by a random effects logistic regression model, including random effects in the model for both the survey cluster and the HF [16, 19]. In this study there are 879 children 6–11 months old from 177 survey clusters (median [min, max]: 5 [1, 10] children per cluster), linked to 179 HFs in the 12 study zones (median [min, max]: 3 [1, 32] children per HF). There are 577 cross-classification cells (median [min, max]: 1[1, 7] child per cell; 50% of the cells have 1 child).

#### Estimating coverage

The updated World Health Organization (WHO) guidelines on vaccination coverage surveys [20] recommend adhering to probability sampling methods and calculation of sampling weights. However, the first stage selection probability of neighborhoods could not be calculated because of having to rely on random starting points within *aire de santé* rather than having known geographical clusters, and this meant that the second stage selection probability of HHs could not be calculated because the denominator was unknown.

## Discussion

This household and health facility linked survey was designed to answer research questions about factors contributing to vaccination status for infants and vaccination coverage for older children in 12 zones of Kinshasa with low vaccination coverage. Although the linked survey was resource intensive and we found fewer households than expected with children of the target ages, we were able to enroll a large number of children in both age groups. We also successfully obtained home-based records for a high percentage of the children.

Two factors likely played a role in not reaching the expected sample size. The first was the use of the DHS household line list because this list only includes HHs with an adult female, the target population for the DHS survey. The second factor was that response rate was lower than anticipated because many households were not available during the survey period. This could have been because of the 3-day time limit per cluster, or that survey teams did not return in the evening when more people would have been home. It is also possible that internal migration may have caused changes in the number of households with children. Completing a HH census in the selected areas before data collection could have ensured we

met our sample size. Although card retention in this survey was higher than previously documented in DRC [17], the number of home-based health records seen decreased over the course of the data collection period. We speculate that this decrease may be because of interviewers spending less time at each HH in the latter part of the survey, a possibility that highlights the need for strong supervision or a shorter implementation period to mitigate interviewer fatigue.

We chose to individually link an infant's HH information with information collected from their most recently visited HF; this approach may allow us to explore the associations of HH and HF factors to children's vaccination status more accurately than if linking at the ecological level, at least in part because many surveyed infants were vaccinated outside of their immediate neighborhoods. With the individually linked method, the number of HFs in the sample and the average number of infants linked to a HF were not known in advance. The sample of HFs captured through the individual linking approach may not be representative of all HFs in the 12 zones of interest, though is potentially more representative than a sample obtained through purposive selection; a census of HFs could have facilitated the individual linkage of our HF and HH surveys as well as provided context to our results. Additionally, we identified children who had never accessed the healthcare system; if this subset were a larger proportion of the sample it could lead to biased estimates of associations of HF factors. Our linked analyses, which will exclude these children, will represent only those HHs that have accessed the healthcare system. The demand-side reasons for never accessing the system could be explored in separate analyses focused on the HH data.

We successfully linked a large proportion of infants to a surveyed health facility but, as nearly half of children attended facilities outside of their residential zone, some HH and HF cross-classified clusters contain only one child and these may pose methodological challenges during analysis. Given the complexity of the linked approach, it may be more efficient to conduct a HH survey first and to use preliminary data analysis to determine if an individually linked HF survey is warranted. In our case, the additional efforts needed to do a linked survey may not have been required if an initial HH survey had found coverage to be higher than administrative coverage indicated. Efficiency in HF data collection may be increased by focusing on those visited by more than one enrolled child and not excluding those outside the zones where the HH were located. In rural areas where healthcare service options are limited the individually linked method may be easier than other linking strategies; however, an ecological link may be sufficient in that context.

Many of the challenges we encountered in implementing the HH surveys were not unique to linked surveys and the lessons learned can be applied to other population-based HH surveys in dense urban contexts. The challenges associated with implementing the new WHO survey recommendations are also relevant to other situations. In our project, conducting two surveys with different objectives and different target age groups concurrently posed additional challenges, which may have contributed to lower data quality in some instances. For example, the children who were interviewed using the incorrect survey tool for their age group but were subsequently analyzed in their correct age group, had missing data since the two survey tools were not identical.

This description of the methods, meta-data, and lessons learned is intended to provide insight into the challenges and benefits of implementing individually linked HH and HF surveys in a dense urban context for vaccination. Individual linking methodology may be useful for immunization system surveys in urban contexts where families may choose from a large number of healthcare service options. However, our experience highlighted several issues to be considered when designing the survey, planning and implementing data collection, and analyzing the results.

## Acknowledgments

We would like to thank Danni Daniels, Patrice Tchekoya, Pascal Sogbo, Dan Ehlman, Nandini Sreenivasan, Eddie Mwenge, Bile Nsaka, Jack Nkongolo, Therese Osomba, Gaston Tshefu Pongo, Rand Young, Laura Wright, Muriel Nzazi Nsambu, Richard Letshu Tsheke, Tiekoura Coulibaly, Aissata Diaha, Aaron Wallace, and Laura Conklin for their contributions in developing and implementing this survey. We are very appreciative of the supervisors, interviewers, and data enterers for their hard work.

## References

- Rainey JJ, Watkins M, Ryman TK, Sandhu P, Bo A, Banerjee K. Reasons related to non-vaccination and under-vaccination of children in low and middle income countries: findings from a systematic review of the published literature, 1999–2009. Vaccine. 2011; 29:8215–21. [PubMed: 21893149]
- 2. Favin M, Steinglass R, Fields R, Banerjee K, Sawhney M. Why children are not vaccinated: a review of the grey literature. Int Health. 2012; 4:229–38. [PubMed: 24029668]
- Hutchins SS, Jansen HA, Robertson SE, Evans P, Kim-Farley RJ. Studies of missed opportunities for immunization in developing and industrialized countries. Bull World Health Organ. 1993; 71:549–60. [PubMed: 8261558]
- Datar A, Mukherji A, Sood N. Health infrastructure & immunization coverage in rural India. Indian J Med Res. 2007; 125:31–42. [PubMed: 17332655]
- 5. Suarez-Castaneda E, Pezzoli L, Elas M, et al. Routine childhood vaccination programme coverage, El Salvador, 2011-In search of timeliness. Vaccine. 2014; 32:437–44. [PubMed: 24315884]
- Skiles MP, Burgert CR, Curtis SL, Spencer J. Geographically linking population and facility surveys: methodological considerations. Popul Health Metr. 2013; 11:14. [PubMed: 23926907]
- Burget, CR., Prosnitz, D. Linking DHS Household and SPA Facility Surveys: Data Considerations and Geospatial Methods. DHS Spatial Analysis Reports (10). Accessed 20 Sept 16 from http:// dhsprogram.com/pubs/pdf/SAR10/SAR10.pdf
- Chen, S., Guilkey, DK. The Effect of Facility Characteristics on Choice of Family Planning Facility in Rural Tanzania. Chapel Hill, NC: MEASURE Evaluation, Carolina Population Center, University of North Carolina at Chapel Hill; 2002.
- 9. Marchant, T., Schellenberg, JA. Measuring skilled attendance at birth using linked household, health facility, and health worker surveys in Ethiopia, North-East Nigeria, and Uttar Pradesh, India. Global Health Metrics and Evaluation Conference 2013; 15–17 June 2013; Seattle, USA.
- 10. Frankenberg E, Thomas D. The Indonesia Family Life Survey (IFLS): Study design and results from waves 1–2.
- 11. World Factbook. Democratic Republic of the Congo. 2016. Accessed 28 sept 2016 https:// www.cia.gov/library/publications/the-world-factbook/geos/cg.html
- WHO and UNICEF. Democratic Republic of the Congo: WHO and UNICEF estimates of immunization coverage: 2014 revision. 2014. Access http://www.who.int/immunization/ monitoring\_surveillance/data/cod.pdf
- Shidi, C. Situation rougeole Kinshasa janvier au 6 juin 2014 [PowerPoint slides]. Presented 13 June 2014, Expanded Program on Immunization Offices; Kinshasa, DRC.
- 14. Hampton L. Personal communication. 1 July 2015.

- 16. Hox JJ. Multilevel analysis: techniques and applications. 2002
- Ministère du Plan et Suivi de la Mise en oeuvre de la Révolution de la Modernité & MEASURE DHS. Democratic Republic of the Congo DHS 2013–14. Accessed 20 sept 2016 from http:// dhsprogram.com/pubs/pdf/FR300/FR300.pdf
- Multilevel Modeling of Hierarchical and Longitudinal Data Using SAS Course Notes. SAS Institute Inc; Cary NC, USA: 2012.
- Johnson BD. Cross-Classified Multilevel Models: An Application to the Criminal Case Processing of Indicted Terrorists. J Quant Criminol. 2012; 28:163–189.
- 20. World Health Organization. Vaccination Coverage Cluster Surveys: Reference Manual (V3 working draft 2015). Accessed 17 May 2016 from http://www.who.int/immunization/monitoring\_surveillance/Vaccination\_coverage\_cluster\_survey.pdf?ua=1



## Figure 1.

Map of 12 zones with administrative vaccination coverage <80% for the first dose of pentavalent vaccine. Kinshasa Province, Democratic Republic of Congo, 2014.



## Figure 2.

Card ownership, retention, availabilty, and history recorded by stage of data collection. Kinshasa Province, Democratic Republic of Congo 2015.



#### Figure 3.

Flow of linked survey records. Kinshasa Province, Democratic Republic of Congo, 2015.

Author Manuscript

Household level response rate by zone. Kinshasa, Democratic Republic of Congo, 2015.

								one						
	1	7	e	4	S	9	٢	×	6	10	11	12	Total	Target
Both age groups														
Households identified	185	232	204	234	202	186	208	166	210	210	165	207	2409	
Total household refusals	٢	15	10	2	5	9	6	3	33	8	11	٢	86	
Refusal rate	4%	6%	5%	1%	2%	3%	4%	2%	1%	4%	7%	3%	4%	
6–11 month old infants														
Households with infant(s) identified $^*$	123	132	133	153	123	122	116	112	126	138	106	129	1,513	
Households interviewed	90	104	101	129	94	96	76	90	106	106	76	114	1,224	1,966
Response rate	73%	%6L	76%	84%	76%	%6L	84%	80%	84%	%LL	92%	88%	81%	
12–23 month old children														
Households with child(ren) identified $^{*}$	67	102	73	82	82	64	94	57	85	73	60	62	918	
Households interviewed	56	76	48	61	54	55	81	42	60	59	53	65	710	1,080
Response rate	84%	75%	66%	74%	%99	86%	86%	74%	71%	81%	88%	82%	%LL	
* Includes children who were excluded bec	ause of	time in 1	neighboi	thood cr	iteria ar	id house	holds th	lat were	nnsucce	ssful in	reachin	g the pri	marv car	egiver