SUPPLEMENT ARTICLE

Cluster Lot Quality Assurance Sampling: Effect of Increasing the Number of Clusters on Classification Precision and Operational Feasibility

Hiromasa Okayasu,[1] Alexandra E. Brown,[1] Michael M. Nzioki,[2] Alex N. Gasasira,[2] Marina Takane,[1] Pascal Mkanda,[2] Steven G. F. Wassilak,[3] and Roland W. Sutter[1]

[1]Research, Policy and Product Development, Global Polio Eradication Department, World Health Organization, Geneva, Switzerland; [2]Country Office, World Health Organization, Abuja, Nigeria; and [3]Global Immunization Division, Centers for Disease Control and Prevention, Atlanta, Georgia

Background. To assess the quality of supplementary immunization activities (SIAs), the Global Polio Eradication Initiative (GPEI) has used cluster lot quality assurance sampling (C-LQAS) methods since 2009. However, since the inception of C-LQAS, questions have been raised about the optimal balance between operational feasibility and precision of classification of lots to identify areas with low SIA quality that require corrective programmatic action.

Methods. To determine if an increased precision in classification would result in differential programmatic decision making, we conducted a pilot evaluation in 4 local government areas (LGAs) in Nigeria with an expanded LQAS sample size of 16 clusters (instead of the standard 6 clusters) of 10 subjects each.

Results. The results showed greater heterogeneity between clusters than the assumed standard deviation of 10%, ranging from 12% to 23%. Comparing the distribution of 4-outcome classifications obtained from all possible combinations of 6-cluster subsamples to the observed classification of the 16-cluster sample, we obtained an exact match in classification in 56% to 85% of instances.

Conclusions. We concluded that the 6-cluster C-LQAS provides acceptable classification precision for programmatic action. Considering the greater resources required to implement an expanded C-LQAS, the improvement in precision was deemed insufficient to warrant the effort.

Keywords. cluster lot quality assurance sampling; Nigeria; polio eradication; supplementary immunization activities; survey methods.

The Global Polio Eradication Initiative (GPEI) has been using cluster lot quality assurance sampling (C-LQAS) since 2009 as a method to rapidly assess supplementary immunization activities (SIAs). Although C-LQAS has proven to be a practical tool to assess the quality of immunization campaigns at a relatively low cost, questions have been raised over the optimal trade-off between operational feasibility and precision (reproducibility) of classification of lots to identify areas with poor immunization quality that require corrective programmatic action. In particular, concerns have arisen that the sample size recommended by GPEI is too small given potential heterogeneity between clusters within a lot. To address this issue, we conducted a pilot study using an expanded sample size to determine if the increase in precision would affect decision making.

LQAS was originally developed as a low-cost method for quality assurance testing in manufacturing, in which a small sample of goods from a production unit ("lot") is inspected for production quality: if the number of defective goods in this sample exceeds a predetermined number (decision value), then the entire lot is deemed to be of unacceptable quality. LQAS has been increasingly applied in the health context (see [1] for an overview of health surveys using LQAS). To monitor the quality of vaccination campaigns, LQAS can be used

Correspondence: Hiromasa Okayasu, MD, MBA, Polio Eradication Initiative, World Health Organization, 20 Ave Appia, CH-1211 Geneva 27, Switzerland (okayasuhi@who.int).

The Journal of Infectious Diseases® 2014;210(S1):S341–6

DOI: 10.1093/infdis/jiu162

to classify geographical areas of interest, such as districts or subdistricts, called "lots", as having acceptable or unacceptable SIA quality based on the number of unvaccinated children in the sample. The decision value *d*, the maximum allowable number of unvaccinated children in the sample, is determined based on a programmatically acceptable minimum underlying coverage rate, or lower threshold (LT), and is selected so that lots that have coverage lower than this threshold are unlikely to be accepted. Concerning programmatic action, any lot that is not accepted is recommended to be targeted for mop-up immunization activities or other corrective action. An upper threshold (UT) is selected to control the probability that a lot is erroneously rejected even though the true underlying proportion of children vaccinated is higher than the UT.

LQAS offers a number of advantages compared with other methods for monitoring the quality of vaccination campaigns. Survey methods used to obtain point estimates of vaccination coverage (such as the Expanded Programme for Immunization [EPI] cluster surveys [2]) are labor intensive and cannot be considered for SIA assessments, particularly given that GPEI SIAs are conducted 2 to 12 times per year. LQAS provides a statistically sound method to assess whether SIA quality is adequate or not, as opposed to methods based on targeted or convenience sampling [3]. In addition, field implementation of LQAS surveys is straightforward. This ease of application makes the LQAS a valuable operational tool to detect pockets of low SIA quality and therefore redirect vaccination efforts.

In the GPEI C-LQAS method, 6 settlements are sampled within a lot from a defined list of all settlements, followed by sampling 10 children randomly per settlement, for a total sample size of 60 [4]. This is referred to hereafter as the 6-cluster standard GPEI C-LQAS method. For programmatic reasons, GPEI uses a cluster sample method in which the sampled children are concentrated in a small number of settlements, rather than using a simple random sample scattered geographically across the lot (eg, local government area (LGA) in Nigeria). Having the sampled persons clustered geographically greatly reduces the resource-intensiveness of the fieldwork and increases the operational feasibility of the method as a regular monitoring tool [5]. A November 2009 pilot for C-LQAS in 20 high-risk LGAs in Nigeria demonstrated the programmatic feasibility and value of the method as a tool for the GPEI [6]. Since the success of the pilot in Nigeria, LQAS has been used widely in other polio-affected countries, including India, Pakistan, Nigeria, Afghanistan, Chad, the Democratic Republic of the Congo, and Angola. As a result, the GPEI Strategic Plan 2013–2018 indicated that polio-endemic countries should adopt LQAS as the "gold standard" for gauging campaign quality and to track trends over time in high-risk areas [7].

Currently, LQAS surveys are conducted after each campaign in all polio-endemic countries and on an ad hoc basis in other countries to assess SIA quality quickly (see [3] for an overview of the C-LQAS method and field implementation by the GPEI). They are recommended to be conducted within a week of the completion of the round, starting 1 or 2 days after the SIA. The target population of the C-LQAS is the same as that targeted by the SIA, which is usually children 0–59 months of age living in the defined area during the campaign. Although the GPEI guidelines for LQAS specify a lower threshold of 90%, the single binary test at 90% provides limited information in high-risk areas with many failed lots. For such high-risk areas, GPEI has adjusted the LQAS method to use multiple decision values to classify lots into 4 bands of SIA quality: good (accepted at a lower threshold of 90%), intermediate (accepted at a lower threshold of 80% if not accepted at a threshold of 90%), poor (accepted at a lower threshold of 60% if not accepted at a threshold of 80%), and very poor (not accepted based on testing at a lower threshold of 60%). Each classification has corresponding classification errors α (probability of accepting a lot with inadequate SIA quality) and β (probability of not accepting a lot with adequate SIA quality) by decision value.

Variability in the proportion of children vaccinated among clusters within a lot has a significant impact on α and β values, and high variability increases the probability of error, compromising the robustness of the "pass/fail" determination. When the GPEI guidelines were written, the standard deviation (SD) in cluster-level coverage within an LGA was assumed to be relatively low at 10%. However, in an evaluation of the observed standard deviations from 220 clusters in LQASs from Nigeria in 2010 and 2011, the median standard deviation was found to be 19% (Wannemuehler, CDC, unpublished data). If the standard deviation among clusters is higher than anticipated, the classification of SIA quality (eg, accepted by testing at lower threshold of 90%) becomes less accurate and leads to an unacceptably high α (0.38 for LT = 90% and decision value = 3 when SD is 19%). A priori, it is unclear if having a variance much greater than expected would invalidate the technique in programmatic action to be taken as a result of the C-LQAS. To address these concerns, we proposed a study to assess whether increasing the number of clusters sampled reduces the effect of intercluster variability and leads to more robust classification.

METHODS

The primary objective of this study was to investigate how increasing the number of clusters improves the precision of the C-LQAS technique compared with the current standard method of 6 clusters of 10 children per lot. Another objective was to investigate the feasibility and operational implications of increasing the sample size to potentially obtain more robust classifications of SIA quality for high-risk LGAs in Nigeria.

Given the higher between-cluster standard deviation of 19%, we determined that a C-LQAS design of 16 clusters of 10 children each would be adequate to assess a lower threshold of

Table 1. Parameters for Multithreshold C-LQAS for the 6-Cluster and 16-Cluster Samples

Classification[a]		6-Cluster Sample (SD = 10%)[a]	16-Cluster Sample (SD = 19%)[a]
Good Accepted at 90%	No. of missed children	0–3	0–7
	α	24%	14%
	β	–13%	. . .[b]
Intermediate Accepted at 80% but not at 90%	No. of missed children	4–8	8–22
	α	19%	14%
	β	–21%	20%
Poor Accepted at 60% but not at 80%	No. of missed children	9–19	23–48
	α	16%	5%
	β	–35%	4%
Very Poor Not accepted at 60%	No. of missed children	20+	49+

Abbreviations: α, probability of type I error (probability that an area with coverage below the lower threshold is accepted); β, probability of type II error (probability that an area with coverage above the upper threshold is not accepted); C-LQAS, cluster lot quality assurance sampling; GPEI, Global Polio Eradication Initiative; No., number; SD, standard deviation; UT, upper threshold.

[a] The 16-cluster sample assumes a between-cluster SD of 19%, and parameter values are computed by simulation using a beta distribution. A 10% SD is assumed for the 6-cluster sample, and parameter values are consistent with the standard GPEI methodology.

[b] β could not be simulated for UT = 98% and SD = 19%. For UT = 95%, β = 0.49.

90%, 80%, and 60% with decision values of 7, 22, and 48 missed children, respectively, at an α of 0.15 or less at each threshold and a β of 0.20 or less where possible. This is referred to hereafter as the 16-cluster or expanded sample size C-LQAS. Child-level data from the 16 clusters were analyzed to determine the following: (1) the observed standard deviations between clusters, as well as point estimates and 95% confidence intervals (CIs) taking into account the cluster sampling; and (2) the variability in decisions that would have been made under the 6-cluster standard design compared with the 16-cluster expanded sample C-LQAS. In particular, the assessment of the 6-cluster sample was conducted by determining the distribution of lot classifications for all possible (${}_{16}C_6 = 8008$) 6-cluster subsamples within the 16-cluster sample and comparing with the observed classification of the 16-cluster sample. This assessed the "robustness" of the classification under the 6-cluster sample that forms the current standard GPEI LQAS design, and provided implications for the use of the results for decision making in the field.

We used a multithreshold cutoff for the analysis, comparing classifications resulting from the 6-cluster samples with the outcome of the 16-cluster sample. A multithreshold cutoff was used in order to be consistent with GPEI field implementation of C-LQAS and because a binary classification (pass/fail) would provide limited information about the probability of misclassification. Table 1 shows the parameters, including the maximum allowable number of missed children per lot and ranges for α and β for the 6-cluster and 16-cluster samples, for the 3 pairs of thresholds [4]. The parameters for the 6-cluster sample are the standard values under the GPEI guidelines based on a SD of 10%. For the 16-cluster sample, a between-cluster SD of 19% was assumed based on the observed standard deviations from previous LQASs in Nigeria in 2010 and 2011.

Taking the classification of the 4 LGAs based on the 16-cluster samples and the decision values based on high variability between clusters (SD = 19%) as an approximation of the "true" classification, we examined the frequency of misclassification for all 8008 possible 6-cluster subsample combinations based on the decision values as specified under the GPEI guidelines. This should provide a conservative estimate of the probability and direction of misclassification under the GPEI-recommended 6-cluster design, assuming that the classification observed from the sample of 16 clusters is representative of the true classification. As the 16-cluster sample is selected from all settlements in the LGA using probability proportional to estimated size (PPS), the sampling of 6 clusters from among the 16 already takes into consideration the size of the cluster as a prior probability. Thus, each 6-cluster combination can be considered to have equal weight in this analysis.

The expanded LQAS was implemented in 4 LGAs as part of a larger LQAS assessment conducted from 6 to 11 February 2013, covering 85 LGAs in 6 states to assess the quality of the February 2013 campaign. One urban and 1 rural LGA each were selected from 2 high-risk states for a total of 4 lots with expanded sample size: Kankara and Katsina LGAs in Katsina State, and Ikara and Kaduna North LGAs in Kaduna State. Table 2 shows the LGAs selected, including the number of settlements, total and target populations, and the total estimated population of the sampled settlements. The sampling method of the expanded LQAS within each of the selected LGAs followed the GPEI field manual. In each LGA, 16 settlements (clusters) were sampled from the available (presumed complete) list of settlements based on PPS. This procedure gives a higher probability to the larger localities to be selected as clusters. In each settlement, surveyors use a "spin-the-bottle" procedure from the center of the sector to select the household that will serve as the starting point. The remaining households are visited by exiting the first household to the right and selecting the 9 subsequent households at a predetermined interval of 1 or 2 houses, depending on the size of the settlement. In each of the 10 households, 1 child in the target age group (under 5 years of age) is selected at random, and their immunization status is checked based on the presence of a

Table 2. Background Characteristics of the 4 Local Government Areas

State	LGA	Number of Settlements	Total Population	Target Population (Children Under 5 y of Age)	Total Population of Sampled Settlements
Kaduna	Ikara	889	635 305	127 061	29 540
Kaduna	Kaduna North	1419	554 005	110 801	8345
Katsina	Kankara	501	254 819	50 964	12 185
Katsina	Katsina	604	639 915	127 983	18 070

Abbreviation: LGA, local government area.

finger mark. Further details of the standard C-LQAS method used by GPEI can be found in the field manual [4].

We collected data electronically in the field using Magpi [8] on mobile phones. To assess the adequacy of SIA quality, we analyzed data in the field to identify and provide immediate feedback to LGAs that failed to meet a lower threshold of 90% (corresponding to a decision value of 7 missed children in the sample). To assess the operational feasibility of the expanded LQAS sample size, data on time in the field were compiled and sent back to the national office. Time required for field data collection was taken as a proxy measure for resource-intensiveness and was analyzed by taking the difference between the first and last data submission times per data collector per LGA. For the 16-cluster LGAs, the sum of data collection times for the 2 data collectors covering 8 clusters each was taken to allow direct comparison with the 6-cluster samples in resources required.

RESULTS

Results for the 4 LGAs selected for the expanded sample size are shown in Table 3. All 4 LGAs failed the binary test at 90%, as they had more than 7 children unmarked. The observed variability in the proportion of children marked in a cluster is greater than the assumed SD of 10% in all 4 LGAs. The degree of variability differs across LGAs, with Ikara exhibiting considerable variability (SD = 23%), while Kankara, Kaduna North, and Katsina had between-cluster SDs between 10% and 15%.

The observed proportion of children with finger marking ranged from 67% in Katsina to 91% in Kankara. The width of the CIs ranged from 12 percentage points (pp) for Kankara to approximately 25 pp for Ikara. The determination of the point estimate is not the objective of the LQAS as currently implemented by GPEI, only the assessment of SIA quality to direct programmatic action to areas that fail standards.

Table 4 shows the distribution of classifications of the 8008 6-cluster subsamples for each of the 4 LGAs, as well as the classification based on the 16-cluster sample. The results indicate that the majority (56%–84%) of 6-cluster subsamples have a classification that matches that of the 16-cluster sample. The LGA with the highest proportion of mismatched classification was Katsina at 44%, but smaller in the other 3 LGAs (less than 30%). In all 4 LGAs, a mismatch in classification is limited to the adjacent categories with a tendency toward misclassification into the higher SIA quality category (eg, "good" instead of "intermediate"), except for the 2-category mismatch for 1% of subsamples for Ikara, which had the highest intercluster SD.

The time required for field data collection was 3.2 days for the 16-cluster sample compared with 1.6 days for the 6-cluster sample. This indicates that a single data collector would take approximately twice as long in the field for the expanded sample. In absolute terms, taking into consideration that 2 data collectors were in the field simultaneously, the data collection time for the 16-cluster sample was slightly longer (1.8 days) compared with the 6-cluster sample.

Table 3. Results of Expanded C-LQAS in 4 LGAs: 16 Clusters of 10 Children Each, Nigeria, 2013

LGA (Lot)	Total No. Unmarked	Classification	Proportion Marked (%)[a]	95% CI[a]	SD in Proportion Marked in a Cluster[b]
Ikara	35	Poor	78	66–90	22.9
Kaduna North	29	Poor	82	75–88	12.2
Kankara	14	Intermediate	91	85–97	11.5
Katsina	53	Very Poor	67	60–74	13.5

Abbreviations: CI, confidence interval; LGA, local government area; C-LQAS, cluster lot quality assurance sampling; No., number; SD, standard deviation.

[a] Point estimates and CIs account for the clustering. Possible sources of bias include incomplete sampling frame and nonresponse.

[b] It is possible that observed variability could be an underestimate if the sampling frame used for the sampling of settlements was incomplete because of settlements that are unknown or unrecognized in the microplans.

Table 4. Distribution of Classifications of all Possible Combinations of 6-Cluster Subsamples for the 4 LGAs[a]

Classification	Ikara (%)	Kaduna North (%)	Kankara (%)	Katsina (%)
Good (accepted at 90%)	1	0	23	0
Intermediate (accepted at 80% but not at 90%)	14	16	***70***	0
Poor (accepted at 60% but not at 80%)	***77***	***84***	7	44
Very poor (not accepted at 60%)	8	0	0	***56***
Total	100	100	100	100

Abbreviation: LGAs, local government areas.

[a] Percentages show the proportion of 6-cluster subsample combinations falling into each category. Percentages representing the category with the highest proportions of combinations are shown in bold. The classification of the LGA based on the 16-cluster sample is indicated with italics. For all 4 LGAs, the category representing the highest proportion of 6-cluster combinations matches the classification resulting from the 16-cluster sample.

DISCUSSION

The advantage of the LQAS method is that, with a relatively small sample drawn randomly from the target population, one can make a rapid assessment on whether SIA quality standards are met in the target population. The original LQAS theory is based on the assumption of drawing a simple random sample. The C-LQAS method extends the theory to account for cluster sampling; however, to implement the method, one must make an a priori assumption about how much between-cluster variability exists. In the 2010–2011 Nigeria surveys, the median observed SD was 19%, indicating considerably more heterogeneity in coverage than had been assumed. The results in this study also indicate somewhat greater heterogeneity than the assumed 10%. The sample size in this study is too small to ascertain the true extent of deviations from 10%; however, the magnitude of the heterogeneity in coverage likely varies considerably across LGAs.

The estimated CI in the 16-cluster surveys was calculated based on the observed standard deviation; given the small sample size and design factor, there is substantial uncertainty around point estimates of LGA-level coverage for LGAs with high between-cluster SD, as seen in the CI width of close to 25 pp for Ikara. There is also a possibility that the observed variability is underestimated if the list of settlements used for the sampling was incomplete. The results of this study show that increasing the number of clusters within an LGA to 16 is insufficient to obtain adequate precision of SIA coverage point estimates; if the program requires a more precise estimate of LGA-level coverage, an ad hoc coverage survey with a larger sample size (eg, cluster sampling with 30 clusters of 7 persons [2]) would be more appropriate in order to reliably obtain a point estimate within ±10% [9]. However, this method is unsuitable for routine assessment of GPEI SIAs.

As expected, some mismatches in classification were seen between the current 6-cluster GPEI method and the 16-cluster expanded samples, and an upward trend of misclassification occurred when the observed classification of the 16-cluster sample was assumed to be "true." However, from the programmatic perspective, as the 6-cluster sample gives the same classification up to 85% of the time, we conclude that this is sufficiently reliable and provides useful information for operational decision making and for following trends over time. Considering the logistical difficulties, the increase in time required for data collection (3.2 days compared with 1.6 days) and the increase in cost associated with a larger sample size, we propose that the improvement in precision is insufficient to warrant the increase in sample size. Both approaches assume that the list of settlements is complete and population estimates are accurate. One source of bias in this analysis is the assumption that the 16-cluster sample provides a classification that is correct (ie, accurately reflects the true SIA quality for the LGA). In reality, even the 16-cluster sample has certain probabilities of classification error: the probability of misclassifying a lot as "good" when the true SIA quality is "intermediate" or worse is approximately 15% with the 16-cluster sample.

We concluded that the current 6-cluster C-LQAS design is an appropriate tool to assess SIA quality to identify areas where SIA quality is poor and to monitor trends in campaign quality. C-LQAS does not provide point estimates of SIA coverage because of the small sample size and design factors; at the same time, it does not appear that a modest increase in precision would justify the increased financial and time burden of an expanded sample.

Notes

Acknowledgments. The authors thank Kathleen Wannemuehler (Global Immunization Division, Center for Global Health, Centers for Disease Control and Prevention) for contributions to study design and comments on data analysis, Ryo Ueno (School of Medicine, University of Tokyo) for the concept and the analysis of the classification distribution of all 6-cluster combinations, and the data collection teams and field supervisors in Nigeria.

Financial support. This work was supported by the World Health Organization, and Centers for Disease Control and Prevention.

Supplement sponsorship. This article is part of a supplement entitled "The Final Phase of Polio Eradication and Endgame Strategies for the Post-Eradication Era," which was sponsored by the Centers for Disease Control and Prevention.

Potential conflicts of interest. All authors: No reported conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Robertson SE, Valadez JJ. Global review of health care surveys using lot quality assurance sampling (LQAS), 1984–2004. Soc Sci Med 1982 **2006**; 63:1648–60.
2. Henderson RH, Sundaresan T. Cluster sampling to assess immunization coverage: a review of experience with a simplified sampling method. Bull World Health Organ **1982**; 60:253–60.

3. Brown AE, Okayasu H, Nzioki MM, et al. Lot Quality Assurance Sampling to monitor supplemental immunization activity quality: an essential tool for improving performance in polio endemic countries. J Infect Dis **2014**; 210(suppl 1):S333–40.
4. Global Polio Eradication Initiative. Assessing Vaccination Coverage Levels Using Clustered Lot Quality Assurance Sampling: Field Manual, **2012**. http://www.polioeradication.org/portals/0/document/research/opvdelivery/lqas.pdf. Accessed 11 November 2013.
5. Pezzoli L, Andrews N, Ronveaux O. Clustered lot quality assurance sampling to assess immunisation coverage: increasing rapidity and maintaining precision. Trop Med Int Health TM IH **2010**; 15:540–6.
6. Greenland K, Rondy M, Chevez A, et al. Clustered lot quality assurance sampling: a pragmatic tool for timely assessment of vaccination coverage. Trop Med Int Health TM IH **2011**; 16:863–8.
7. Global Polio Eradication Initiative. Polio Eradication & Endgame Strategic Plan 2013–2018, **2013**. http://www.polioeradication.org/Portals/0/Document/Resources/StrategyWork/PEESP_EN_A4.pdf. Accessed 11 November 2013.
8. Datadyne. Magpi, **2014**. https://www.magpi.com. Accessed 4 March 2014.
9. Hoshaw-Woodard S. Description and comparison of the methods of cluster sampling and lot quality assurance sampling to assess immunization coverage. Geneva: World Health Organization, **2001**.