

RESEARCH ARTICLE

Choosing a Cluster Sampling Design for Lot Quality Assurance Sampling Surveys

Lauren Hund^{1*}, Edward J. Bedrick², Marcello Pagano³

1 Department of Family and Community Medicine, University of New Mexico, Albuquerque, NM, USA, **2** Department of Biostatistics and Informatics, University of Colorado, Aurora, CO, USA, **3** Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA

* lbhund@gmail.com



Abstract

Lot quality assurance sampling (LQAS) surveys are commonly used for monitoring and evaluation in resource-limited settings. Recently several methods have been proposed to combine LQAS with cluster sampling for more timely and cost-effective data collection. For some of these methods, the standard binomial model can be used for constructing decision rules as the clustering can be ignored. For other designs, considered here, clustering is accommodated in the design phase. In this paper, we compare these latter cluster LQAS methodologies and provide recommendations for choosing a cluster LQAS design. We compare technical differences in the three methods and determine situations in which the choice of method results in a substantively different design. We consider two different aspects of the methods: the distributional assumptions and the clustering parameterization. Further, we provide software tools for implementing each method and clarify misconceptions about these designs in the literature. We illustrate the differences in these methods using vaccination and nutrition cluster LQAS surveys as example designs. The cluster methods are not sensitive to the distributional assumptions but can result in substantially different designs (sample sizes) depending on the clustering parameterization. However, none of the clustering parameterizations used in the existing methods appears to be consistent with the observed data, and, consequently, choice between the cluster LQAS methods is not straightforward. Further research should attempt to characterize clustering patterns in specific applications and provide suggestions for best-practice cluster LQAS designs on a setting-specific basis.

OPEN ACCESS

Citation: Hund L, Bedrick EJ, Pagano M (2015) Choosing a Cluster Sampling Design for Lot Quality Assurance Sampling Surveys. PLoS ONE 10(6): e0129564. doi:10.1371/journal.pone.0129564

Editor: Eric Brian Faragher, Liverpool School of Tropical Medicine, UNITED KINGDOM

Received: September 8, 2014

Accepted: May 11, 2015

Published: June 30, 2015

Copyright: © 2015 Hund et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All files are available in a GitHub repository maintained by LH, located at https://github.com/lbhund/ClusterLQAS_PLOS.

Funding: LH and MP were supported through grant R01 AI097015-01A1 from the National Institutes of Health. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

When designing global health monitoring and evaluation (M&E) surveys in resource limited settings, the key objective is to balance cost, precision, and feasibility. Lot quality assurance sampling (LQAS) surveys were originally proposed for industrial quality control [1] and have more recently been applied in global health M&E for classification of health indicators [2]. In LQAS surveys, samples of size n are collected from supervision areas (SAs), and the number of

successes X is compared to a decision rule d to classify an SA as acceptable or unacceptable [3]. Because LQAS is a classification procedure, LQAS surveys typically require smaller sample sizes than estimation-based surveys. Further, the simplicity of classification facilitates local-level data analysis, reducing the time between data collection and data-driven actions.

LQAS designs are typically determined by four parameters p_b, p_u, α , and β . Specifically, the sample size n and rule d are selected such that the risk of misclassifying an SA as acceptable or unacceptable is small, namely:

$$\begin{aligned} P(X \leq d | p \geq p_u) &\leq \alpha \\ P(X > d | p \leq p_l) &\leq \beta. \end{aligned} \tag{1}$$

The risk α is the probability of classifying as low when $p = p_u$, i.e. coverage is truly high; and the risk β is the probability of classifying as high when the health indicator coverage $p = p_b$, i.e. coverage is truly low. The risk α can then be interpreted as the ‘service provider risk,’ namely the risk of intervening in an area that actually has high coverage; and the risk β interpreted as the ‘population risk,’ namely the risk that a truly under-performing area is classified as having adequate coverage. It is important to note that the LQAS equations in 1 are often parameterized differently in the literature depending on the application, and the interpretation of the risks changes with the parameterization. The α and β risks are not restricted for SAs with $p_l < p < p_u$, i.e. in the “grey region.” Choice of the design parameters p_b, p_u, α , and β has been discussed elsewhere [3–5] and is not addressed in this paper.

Typically, the probabilities in Eq 1 are calculated by assuming data are collected using simple random sampling (SRS) from the target population and then modeling $X \sim \text{Binomial}(n, p)$. Implementing SRS requires enumerating and randomly sampling individuals from the target population, a challenging feat in geographically disperse and hard to reach populations. Cluster sampling surveys are popular in M&E, where enumerating clusters (often villages) and sampling within clusters is less costly [6]. Commonly used rapid M&E cluster surveys include 30×30 emergency nutrition surveys [7, 8] and 30×7 Expanded Program of Immunization (EPI) surveys [9].

Recently, three methods have been proposed to adjust LQAS sample sizes and decision rules to accommodate cluster sampling. We refer to these methods as the Pezzoli [10–13], Hedt [14], and Hund [15] methods. In this paper, we contrast these methods and provide recommendations for choosing a cluster LQAS design. First, we review two-stage cluster sampling surveys and describe the three methods for melding cluster sampling with LQAS. We then evaluate these methods using example designs from vaccination and nutrition surveys. Finally, we provide recommendations for choosing a cluster LQAS design based on these results.

Methods

Cluster surveys

Global health M&E surveys often implement two-stage cluster sampling, with clusters sampled using probability proportional to size (PPS) and individuals sampled using an approximation to SRS. When clusters are sampled with replacement (or the probability of sampling the same cluster more than once is small), this design is ‘self-weighting’. That is, every individual in the target population has an equal probability of being sampled and the data do not have to be weighted during the analysis. Cluster LQAS decision rules rely on use of self-weighting designs, and throughout this paper we assume that data are collected using a self-weighted two-stage cluster sampling design.

When individuals within a cluster are more similar than those between clusters, the data cannot be analyzed as a simple random sample. Consider a population of size N consisting of K total clusters of size M_j for cluster j . Define X_{ij} as the health indicator status (1 success, 0 failure) of subject i in cluster j ; $X_j = \sum_i X_{ij}$ as the number of ‘successes’ in cluster j ; and $X = \sum_j X_j$ as the number of ‘successes’ in the sample. Suppose k clusters are sampled and m individuals are sampled per cluster ($n = km$) and define p_j as the coverage in cluster j . In the presence of clustering, the cluster-level coverages $\{p_j\}$ differ from the overall coverage p ; one measure of clustering is the standard deviation of p_j across clusters, denoted σ . Another measure of clustering is the intraclass correlation, ρ . For two-stage designs in which cluster coverages $\{p_j\}$ have mean p and standard deviation σ and $X_j|p_j \sim \text{Binomial}(m, p_j)$, the intraclass correlation is defined as [16]:

$$\rho = \frac{\sigma^2}{p(1-p)}. \tag{2}$$

The intraclass correlation ρ varies between 0 and 1 with higher values denoting more clustering (technically, under alternative models, ρ can be negative when individuals within a cluster are less similar than between clusters, though this setting is typically not relevant to M&E surveys and is not considered here).

Cluster sampling in LQAS

Initially, cluster sampling LQAS surveys were designed to minimize the amount of variance inflation due to clustering without directly inflating the sample size for clustering. For instance, within-cluster sample sizes of $m = 1$ have been used in LQAS surveys (e.g. [17]; [18]) to achieve the same precision as a SRS survey. In 67×3 and 33×6 LQAS nutrition surveys, small within-cluster sample sizes were selected to minimize cluster sampling variance inflation [19–21]. Stroh and Birmingham (2009) use cluster sampling for LQAS neonatal tetanus elimination surveys and ignore the cluster sampling design, noting that cluster sampling approximates SRS ‘when an attribute to be measured in a population is rare or is homogeneously distributed’ [22]. The common theme between these approaches is that the cluster sampling design can be ignored if the within cluster sample size is small or if clustering is negligible. In these situations, the LQAS sample size and decision rule can be calculated using the standard binomial model with no adjustments required to accommodate the use of cluster sampling.

In practice, clustering is often not negligible, and visiting many clusters and selecting few individuals per cluster may be less logistically feasible or cost-effective than sampling more individuals per cluster. To address these limitations, three methods have recently been developed to explicitly accommodate clustering sampling at the design-phase in cluster LQAS surveys: the Pezzoli [10–13], Hedt [14], and Hund [15] methods. In this paper, we compare these three methods (Hund, Hedt, and Pezzoli) for explicitly accounting for the cluster sampling design and do not discuss designs that ignore the cluster sampling design. We describe these methods below.

Pezzoli method. Pezzoli *et. al* [10] proposed the first cluster LQAS method, developed to assess vaccination coverage. The mathematical representation of the Pezzoli model described in [11, 12] is:

$$\begin{aligned} p_j &\sim \text{Binomial}(\eta, p)/\eta \\ \text{sd}(p_j) &= \sigma, \quad E(p_j) = p \\ X_j &\sim \text{Binomial}(m, p_j). \end{aligned} \tag{3}$$

We refer to this model as a binomial-scaled model, because p_j is modeled as a binomial random

variable scaled by η to ensure that $p_j \in (0,1)$. Clustering is incorporated through specifying a value for σ , the standard deviation of the cluster level coverages. To apply the Pezzoli design, η is selected such that $\sigma^2 = p(1 - p)/\eta$. Using the relationship between σ and ρ , it follows that $\eta = 1/\rho$. This method is subject to some rounding error, since η is rounded to an integer.

To select designs, the authors choose $p = \{p_l, p_u\}$ and $\sigma \in (0,1)$ and determine decision rules via simulation based on the distribution of $X = \sum_j X_j$. There is currently no user-friendly software for implementing the Pezzoli method that allows users to select their own design, though Stata simulation code is available from the authors of [11]. We have written R functions for designing these surveys and provide examples using this code at https://github.com/lbhund/ClusterLQAS_PLOS.

Hedt method. Hedt *et. al* [14] propose modeling X_j as a beta-binomial random variable with mean p and intraclass correlation ρ :

$$X_j \sim \text{Betabinomial}(m, p, \rho) \tag{4}$$

The betabinomial model can also be written as a two-stage model, with $X_j \sim \text{Binomial}(m, p_j)$ and $p_j \sim \text{Beta}(p, \rho)$ (where the beta distribution is parameterized based on the mean p and intraclass correlation ρ). The beta distribution has support on $(0,1)$, and this betabinomial model is common for clustered binary data. The final decision rule is based on the distribution of $X = \sum_j X_j$. Hedt *et. al* [14] provide R code for selecting sample sizes and decision rules for this method; the R package in [15] can also be used to calculate sample sizes and decision rules using the Hedt method.

Hund method. Hund and Pagano (2014) do not specify a distribution for p_j and instead use an overdispersed binomial model for X [15]:

$$\begin{aligned} X &\sim \text{Quasibinomial}(n, p) \\ E(X) &= np, \quad \text{Var}(X) = Dnp(1 - p) \end{aligned} \tag{5}$$

where D is the survey design effect allowing variability in X above that of the binomial distribution. The design effect is the ratio of the variances using cluster sampling compared to SRS and is the multiplicative factor by which the sample size should be inflated to achieve the same precision as SRS when using cluster sampling [23]. For the two-stage cluster sample with large K , the design effect is $D = 1 + (m - 1)\rho$.

The Hund method can also be written in terms of the ‘effective sample size’ as $X^* \sim \text{Binomial}(n^*, p)$, where $X^* = X/D$ and $n^* = n/D$ is the effective sample size. Thus, the sample size and decision rule can be calculated under SRS and inflated by the design effect to accommodate clustering. The R package in [15] can be used to calculate sample sizes and decision rules for this method. A limitation of the Hund method is that α and β errors are inexact due to rounding error when transforming from the effective sample size scale.

Evaluating the designs

There are two primary differences between the methods above: the distributional assumptions (binomial-scaled, beta, and quasi-binomial models) and clustering parameterization using σ versus ρ (that is, the difference between treating σ or ρ as known and the same for p_l and p_u). These differences are summarized in Table 1. We compare the impacts of these differences across the Pezzoli, Hedt, and Hund methods and determine situations in which the choice of method results in a substantively different design. As example designs, we use three couplets for p_l and p_u : 55–70%; 75–90%; and 90–95%. The first two couplets are from the 6×10 two-stage cluster designs for vaccination coverage proposed in [12]. The last couplet is consistent with the 5–10% couplet used in the 33×6 or 67×3 malnutrition LQAS surveys. We

Table 1. Differences between the 3 methods.

Method	Distribution for p_j	Clustering parameterization
Pezzoli	Binomial-scaled	σ fixed
Hedt	Beta	ρ fixed
Hund	None specified	ρ fixed

doi:10.1371/journal.pone.0129564.t001

parameterize these designs according to Eq 1 and specify X as the number of successes (vaccinated children or children with no evidence of malnutrition) out of n . As noted above, both the Pezzoli and Hund models are subject to some rounding error; we use nearest integer rounding for all calculations. All analyses have been conducted using R 3.1.1 [24], and code for reproducing all figures and tables, along with the R package from [15], is available at https://github.com/lbhund/ClusterLQAS_PLOS.

Comparing the distributional assumptions. First, we conduct a simulation study to evaluate whether the choice of distributional model (binomial-scaled, beta, and quasi-binomial) makes a substantive difference in the classification precision of the cluster LQAS designs. For the 2 p_l and p_u threshold couplets from [12], we use the 6×10 design with decision rules $d = 38$ and 50; for the 90–95% couplet, we use the 33×6 design with $d = 13$. We fix $\sigma = .1$ for all models (we repeat the analysis fixing $\sigma = .05$; results are similar and are not presented). That is, we select the clustering parameters at p_l and p_u such that $\sigma = .1$ for all models and thresholds (deviating from the specific methods described above) to isolate the impact of the probability distribution choice. For each p_b, p_u couplet, we generate 10,000 draws from each of the three distributions fixing m, k, d , and $\sigma = .1$ and examine how α and β change with the distributional assumption. For the quasi-binomial and betabinomial models, we calculate ρ using Eq 2. We then repeat the simulation fixing $\rho = .1$ (rather than fixing $\sigma = .1$) and calculate σ for the binomial-scaled model using Eq 2.

Fixing σ versus ρ . Historically, survey sample size calculations rely on the design effect and thus assume ρ is known and constant across levels of coverage p [16, 25]. The Hund and Hedt methods also model clustering assuming ρ is known and fixed as a function of p . The Pezzoli method parameterizes clustering differently, assuming σ is fixed as a function of p . Fixing ρ results in less cluster-level variance (smaller σ) as p moves away from .5. Fixing σ results in the same variance of p_j for all levels of p (and larger ρ as moves away from .5). Intuitively, variability in p_j may decrease as p moves away from .5 (e.g. less variability in the coverage of a rare event), though there is also evidence that ρ decreases as p moves away from .5 [25, 26]. It is thus not immediately clear as to whether fixing σ, ρ , or neither is appropriate.

To examine design differences according to the clustering parameterization, we compare the Pezzoli, Hund, and Hedt designs when $\sigma = .1$ and when $\rho = .1$. That is, we first calculate ρ_l and ρ_u such that $\sigma = .1$ at p_l and p_u , respectively, and compare fixing both ρ_l and ρ_u for the Hedt and Hund designs to the Pezzoli design with $\sigma = .1$. We then repeat this exercise, calculating σ_l and σ_u such that $\rho = .1$ at p_l and p_u , respectively, and compare fixing both σ_l and σ_u for the Pezzoli design to the Hedt and Hund designs with $\rho = .1$. We again evaluate the three couplets 55–70%, 75–90%, and 90–95% and specify $\alpha = \beta = .1$ and $m = 10$.

We then examine reported estimates of p, σ , and ρ from [12] and [27] to assess how estimates of ρ and σ vary as a function of estimated vaccination coverage in practice. Before delving into this data example, we note that the standard deviation of the estimated \hat{p}_j s, $s =$

$\sqrt{(k-1)^{-1} \sum_j (\hat{p}_j - \hat{p})^2}$ where \hat{p} is the estimated coverage, is an overestimate of σ . Some have criticized the Pezzoli method for using an upper bound of $\sigma = .1$ in the designs and cite examples of $\sigma > .1$ in practice. Using s as an estimate of σ with small m likely accounts for some of this phenomenon. While s estimates $sd(\hat{p}_j)$, $sd(\hat{p}_j)$ is higher than σ , because $sd(\hat{p}_j)$ includes additional variability associated with the estimation of \hat{p}_j . That is, the standard deviation of the estimated cluster coverages is not an accurate estimate of σ . In the Appendix, we derive an estimator for σ^2 , denoted $\hat{\sigma}^2$:

$$\hat{\sigma}^2 = \frac{ms^2 - \hat{p}(1 - \hat{p})}{m - 1}. \tag{6}$$

From Eq 6, it is clear that using s to estimate σ is only appropriate when m is large (which is seldom the case in cluster LQAS surveys).

We estimate $\hat{\sigma}$ using reported \hat{p} and standard errors s/\sqrt{k} for 20 different 6×10 surveys in [12]. Minetti *et. al* [27] report \hat{p} , s , and intraclass correlation estimates $\hat{\rho}$ from 41 different 10×15 vaccination coverage surveys. Rather than using Eq 6, we estimate $\hat{\sigma} = \hat{\rho}\hat{p}(1 - \hat{p})$ for the data in [27]. For each dataset, we estimate the relationship between $\hat{\rho}$ and \hat{p} and between $\hat{\sigma}$ and \hat{p} using a nonparametric loess smooth [28] based on the `loess` function in R. We use the default parameters in the `loess` function, using a weighted least squares fit based on 75% of the data for the local smoothing.

For each data example, we then estimate operating characteristic (OC) curves for the 6×10 design with $d = 50$ for the 75–90% couplet using the binomial-scaled distribution but varying the values of ρ and σ . First, we assume ρ is fixed at the value corresponding to the mean of $\hat{\rho}$ across all SAs; next, we allow ρ to vary as a function of p , estimating ρ from the nonparametric loess smooth for $\hat{\rho}$. We then repeat for σ , resulting in 4 OC curves for each data example.

Results

Distributional assumptions

In Fig 1, we contrast the shapes of the beta and binomial-scaled distributions for p_j (the Hund method does not impose a distribution on p_j). We plot the shape of these distributions with $\sigma = .1$ and $p = .5, .7$ and $.9$. When $p = .5$, these distributions are symmetric and similar. When p

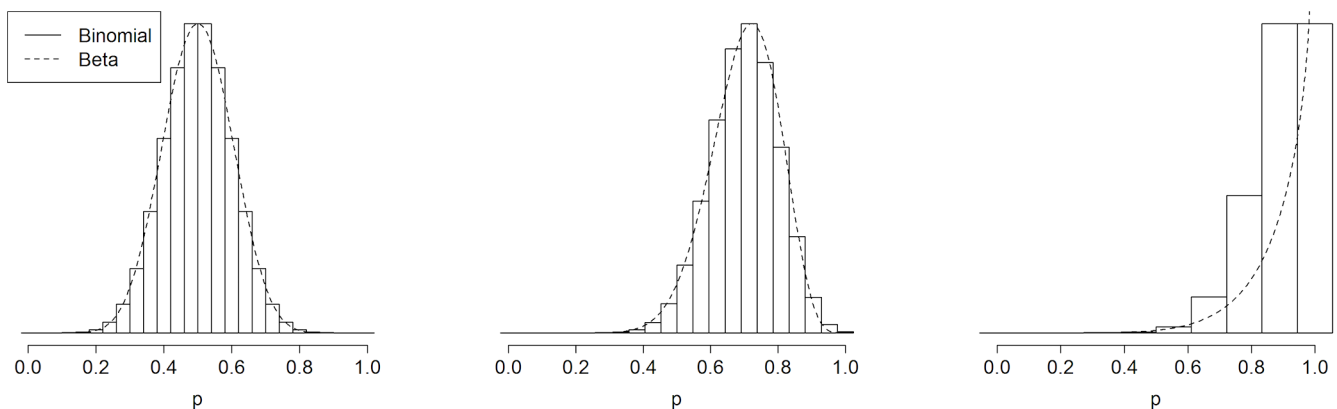


Fig 1. Shape of the beta and binomial-scaled distributions with standard deviation $\sigma = .1$ and means $p = .5, p = .7$, and $p = .9$.

doi:10.1371/journal.pone.0129564.g001

Table 2. Comparing distributional assumptions between the methods when $\sigma = .1$ and $\rho = .1$.

	$p_l = .55, p_u = .7$		$p_l = .75, p_u = .9$		$p_l = .9, p_u = .95$	
	α	β	α	β	α	β
$\sigma = .1$						
Binomial-scaled	.203	.107	.147	.081	.203	.103
Betabinomial	.195	.109	.143	.088	.203	.107
Quasi-binomial	.159	.160	.072	.169	.204	.103
$\rho = .1$						
Binomial-scaled	.234	.149	.147	.111	.170	.107
Betabinomial	.233	.148	.131	.118	.168	.106
Quasi-binomial	.222	.149	.092	.153	.126	.143

For the 55–70% and 75–90% couplets, the risks for the 6×10 design with $d = 38$ and $d = 50$, respectively, are shown. For the 90–95% couplet, the risks for the 33×6 design with $d = 13$ are shown. We compare the binomial-scaled (Pezzoli), beta (Hedt), and quasi-binomial (Hund) models.

doi:10.1371/journal.pone.0129564.t002

$\neq .5$, both distributions are skewed, with the beta distribution more skewed than the binomial. Additionally, as $p \rightarrow 1$ and as σ increases, the binomial distribution becomes more discrete. When $p = .9$ and $\sigma = .1$, $\eta = 9$ and p_j can only take on 10 distinct values: $(0, 1/9, \dots, 8/9, 1)$.

While Fig 1 highlights important differences in beta and binomial-scaled distributions, our simulations suggest that the distribution choice does not have a large impact on the α and β risk levels (Table 2). Specifically, when the clustering levels are fixed (to either $\sigma = .1$ or $\rho = .1$), the differences in risks between the binomial-scaled and beta model are negligible for these designs, even with the relatively high levels of clustering specified. The quasi-binomial model generally results in similar α and β errors to the other models, though differences occur due to rounding error in small sample sizes. Results of these simulations thus suggest that specification of the distribution for p_j does not markedly impact the α and β errors.

Clustering parameterization

In Tables 3 and 4, designs are compared based on the clustering parameterization (fixed σ versus fixed ρ). For the 55–70% and 75–90% couplets, the designs are similar regardless of how clustering is parameterized. For the 90–95% couplet, the sample size is much larger in the Hedt and Hund methods with fixed ρ_u versus fixed ρ_l ; similarly, the sample size is much larger in the Pezzoli method with fixed σ_l versus fixed σ_u . Recall that the examples in Tables 3 and 4 assume that m , the sample size within clusters, is fixed and increments the overall sample size by adding additional clusters. As noted above, clustering variance inflation increases as the within-cluster sample size increases. Subsequently, discrepancies between the three methods are more pronounced when the number of sampled clusters k is fixed, rather than m . As an example, if we were to fix the number of clusters to $k = 6$ rather than fixing the within-cluster sample size $m = 10$, the total sample size n would range between 72 and 150 when $\sigma = .1$ (analogous to the 70–90 sample size range in Table 3 with m fixed).

Eq 2 helps explain the sensitivity of the sample size to the clustering parameterization when p_l and p_u are near 0 or 1. The parameters ρ and σ will be similar for both p_l and p_u when $p_l(1 - p_l) \approx p_u(1 - p_u)$. Hence, ρ_l and ρ_u are negligibly different when $\sigma = .1$ for the 55–70% couplet, whereas ρ_u is nearly double ρ_l for the 75–90% and 90–95% couplets. To illustrate this difference

Table 3. Comparing the Pezzoli, Hund, and Hedt designs for $\sigma = .1$, $\alpha = \beta = .1$, and $m = 10$.

		n	d	k	m	α	β
<i>$\rho_I = .55, \rho_U = .7$</i>							
SRS		71	44			.091	.096
Pezzoli	$\sigma = .1$	100	62	10	10	.090	.096
Hund	$\rho_I = .04$	110	68	11	10	.069	.091
Hedt	$\rho_I = .04$	110	68	11	10	.071	.097
Hund	$\rho_U = .05$	110	69	11	10	.091	.079
Hedt	$\rho_U = .05$	110	68	11	10	.071	.099
<i>$\rho_I = .75, \rho_U = .9$</i>							
SRS		40	33			.100	.096
Pezzoli	$\sigma = .1$	80	66	8	10	.088	.076
Hund	$\rho_I = .05$	70	58	7	10	.093	.071
Hedt	$\rho_I = .05$	70	58	7	10	.081	.077
Hund	$\rho_U = .11$	80	66	8	10	.100	.096
Hedt	$\rho_U = .11$	90	75	9	10	.096	.074
<i>$\rho_I = .9, \rho_U = .95$</i>							
SRS		187	173			.087	.098
Pezzoli	$\sigma = .1$	440	407	44	10	.095	.091
Hund	$\rho_I = .11$	380	352	38	10	.096	.087
Hedt	$\rho_I = .11$	370	343	37	10	.100	.093
Hund	$\rho_U = .21$	540	501	54	10	.087	.098
Hedt	$\rho_U = .21$	550	510	55	10	.088	.091

doi:10.1371/journal.pone.0129564.t003

Table 4. Comparing the Pezzoli, Hund, and Hedt designs for $\rho = .1$, $\alpha = \beta = .1$, and $m = 10$.

		n	d	k	m	α	β
<i>$\rho_I = .55, \rho_U = .7$</i>							
SRS		71	44			.091	.096
Pezzoli	$\sigma_I = .16$	140	87	14	10	.098	.098
Pezzoli	$\sigma_U = .14$	130	81	13	10	.099	.087
Hund	$\rho = .1$	140	87	14	10	.091	.087
Hedt	$\rho = .1$	140	87	14	10	.087	.099
<i>$\rho_I = .75, \rho_U = .9$</i>							
SRS		40	33			.100	.096
Pezzoli	$\sigma_I = .14$	110	90	11	10	.063	.095
Pezzoli	$\sigma_U = .09$	80	66	8	10	.075	.071
Hund	$\rho = .1$	90	74	9	10	.093	.071
Hedt	$\rho = .1$	90	75	9	10	.088	.073
<i>$\rho_I = .9, \rho_U = .95$</i>							
SRS		187	173			.087	.098
Pezzoli	$\sigma_I = .09$	410	379	41	10	.093	.098
Pezzoli	$\sigma_U = .07$	320	296	32	10	.088	.094
Hund	$\rho = .1$	360	332	36	10	.093	.090
Hedt	$\rho = .1$	370	343	37	10	.089	.085

doi:10.1371/journal.pone.0129564.t004

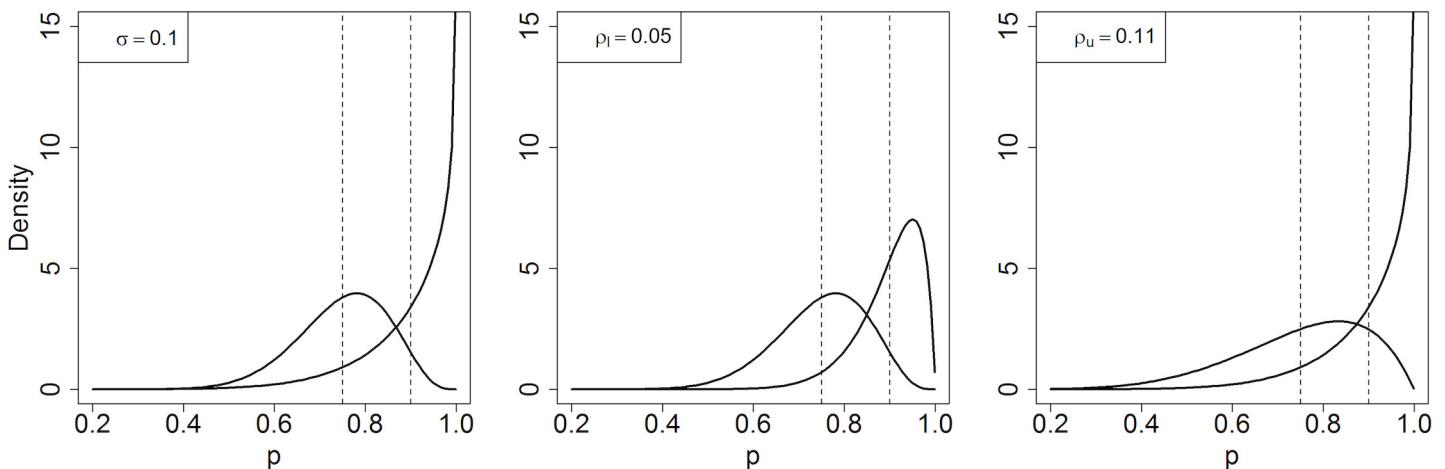


Fig 2. Plot of beta distributions with fixed mean coverages and varying standard deviation and/or intraclass correlation. The dotted lines represent the mean coverages $p_l = .75$ and $p_u = .9$, which are the same in all 3 panels. Left: standard deviation is fixed for both distributions at $\sigma = .1$; middle: intraclass correlation is fixed at $\rho_l = .05$ (note: $\rho_l = .1^2 / (.75 * .25)$); right: intraclass correlation is fixed at $\rho_u = .11$ (note: $\rho_u = .1^2 / (.9 * .1)$).

doi:10.1371/journal.pone.0129564.g002

for the 75–90% couplet, we plot beta distributions fixing $\sigma = .1$, fixing ρ_l , and fixing ρ_u in Fig 2. Unlike the distributional assumptions, choice of clustering parameterization can substantially impact the final sample size for the survey when p_l and p_u are near 0 or 1.

Next, we examine whether existing data support one of the two proposed parameterizations. In Figs 3 and 4, $\hat{\sigma}$ and $\hat{\rho}$ are plotted as a function of \hat{p} for the data in [12] and [27], respectively. Both $\hat{\rho}$ and $\hat{\sigma}$ decrease as a function of estimated vaccination coverage (less clustering occurs in more highly vaccinated areas), consistent with [25] and [26]. Fixing either σ or ρ likely results in some level of misspecification of α and β . The OC curves corresponding to the different clustering specifications are also shown in Figs 3 and 4. In the example from [12], the α and β risks are substantially higher when the mean versus the lowest smooth estimate of σ and ρ are used. This inflation of the risks occurs because σ and ρ decrease as a function of p but the average coverage across SAS is lower than 75%. The risks are not inflated in the example from [27] because the range of vaccination coverages is similar to the 75–90% couplet thresholds.

Discussion

After reviewing three methods for implementing cluster sampling in LQAS surveys, we conclude that the cluster LQAS methods result in similar designs when p_l and p_u are far from 0 or 1 or when clustering is minimal. In these situations, the user should not be concerned about which of the cluster methods to use, as the methods should all result in similar designs. Our simulation study suggests that the choice of distributional assumptions also does not appear to substantially impact the survey design.

Choice of the clustering parameterization is the major factor that can impact the final cluster LQAS survey design. Specifically, if p_l or p_u is near 0 or 1, the design (sample size) changes substantially depending on the clustering parameterization, namely whether ρ or σ is fixed as a function of p . That is, when $p_l \leq .1$ and $p_u \geq .9$, clustering parameters should be selected carefully to avoid severe misspecification of classification risks. As an example, assume $p_l, p_u > .5$. When σ and ρ result in the same amount of clustering for $p = p_l$ (that is, $\sigma = \rho p_l (1 - p_l)$), the

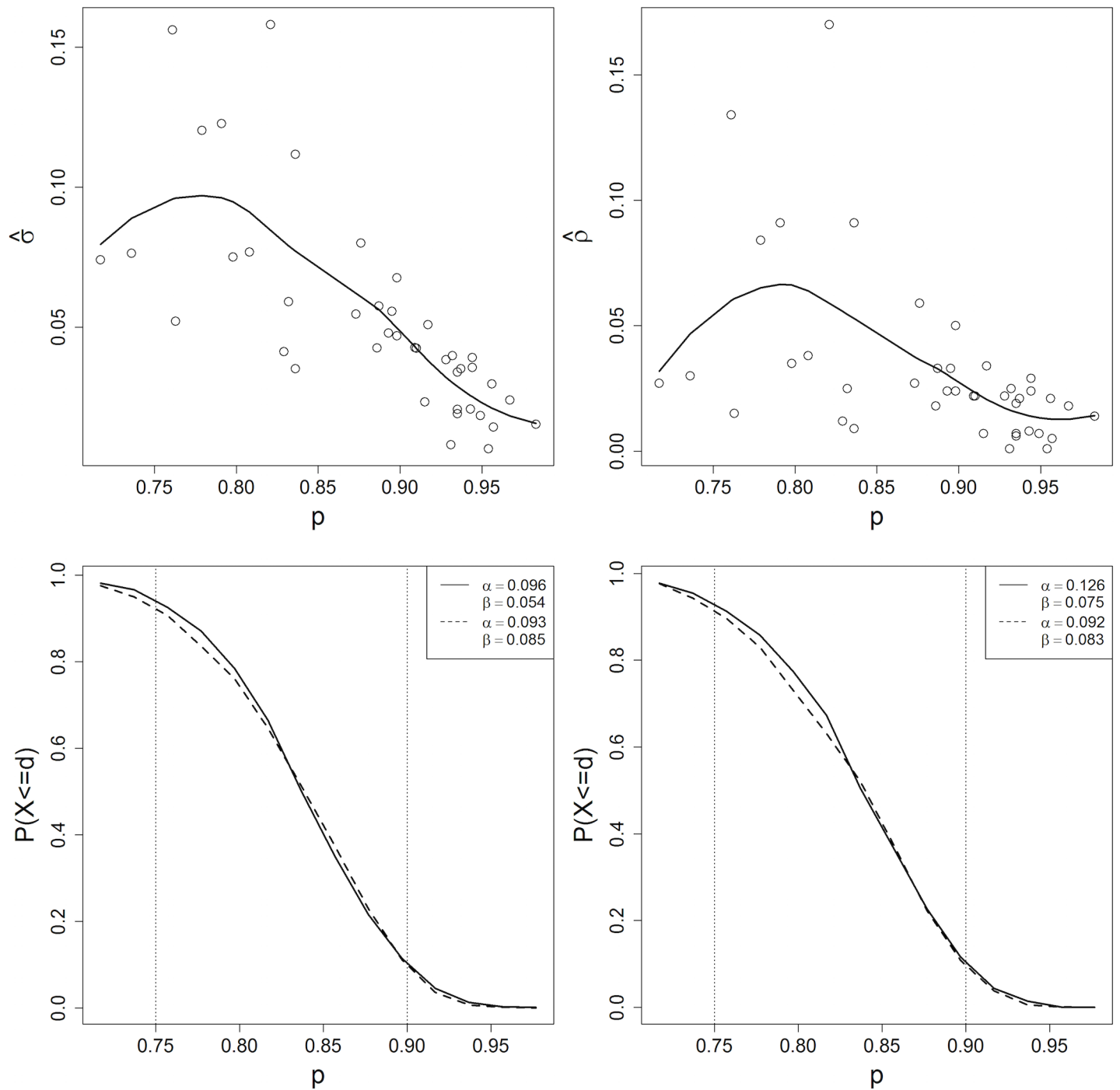


Fig 3. Top panel: Estimated σ and ρ as a function of $\hat{\rho}_j$ for 37 areas in Minetti *et. al* [27], with loess smooth overlaid. Bottom panel: Solid line represents OC curve for $n = 60, d = 50$ for $p_l = .75, p_u = .9$ design when σ and ρ are fixed at the mean value of $\hat{\rho}_j$ (left) and $\hat{\rho}_j$ (right). Dashed line represents OC curve when σ and ρ vary over ρ according to the predicted loess smooth of $\hat{\rho}_j$ (left) and $\hat{\rho}_j$ (right).

doi:10.1371/journal.pone.0129564.g003

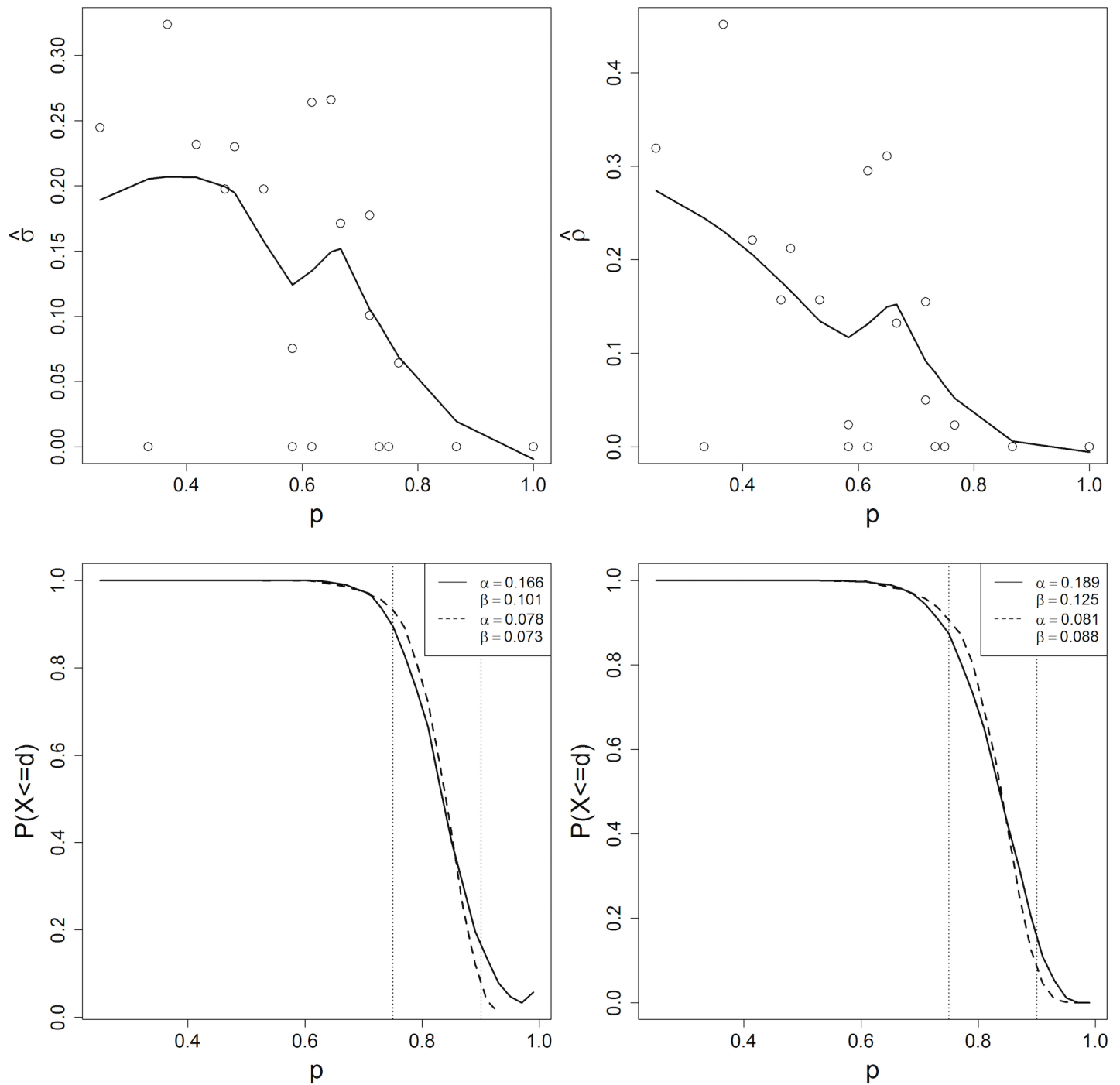


Fig 4. Top panel: Estimated σ and ρ as a function of \hat{p}_j for 20 areas in Greenland et. al [12], with loess smooth overlaid. Bottom panel: Solid line represents OC curve for $n = 60, d = 50$ for $p_l = .75, p_u = .9$ design when σ and ρ are fixed at the mean value of $\hat{\sigma}$ (left) and $\hat{\rho}$ (right). Dashed line represents OC curve when σ and ρ vary over p according to the predicted loess smooth of $\hat{\sigma}$ (left) and $\hat{\rho}$ (right).

doi:10.1371/journal.pone.0129564.g004

design with fixed σ will require a larger n ; for $p = p_u$, the design with fixed ρ will require a larger n . Differences in n between these options can be substantial when p_u is close to 1 or $p_u - p_l$ is large.

We did not find evidence that ρ is constant as a function of p (assumed by Hedt and Hund) nor that σ is constant as a function of p (assumed by Pezzoli). In two vaccination coverage examples, estimates of both ρ and σ decrease as coverage increases, consistent with [25, 26]. Using historical data to plot clustering estimates $\hat{\sigma}$ and $\hat{\rho}$ as a function of estimated coverage \hat{p} can help evaluate the clustering parameterization assumptions. Our data analysis suggests that, if clustering design parameters (σ or ρ) are determined from historical data, only estimates of $\hat{\sigma}$ and $\hat{\rho}$ from SAS with estimated coverages near the range of p_l and p_u should be incorporated. To facilitate incorporation of historical data, we provide R functions that estimate σ using summary statistics commonly available in the literature. We strongly encourage cluster LQAS users to publish their estimates of these summary statistics (\hat{p} , s , and $\hat{\rho}$) for each SA to provide more historical data for specifying these design parameters in the future.

When the Hund or Hedt methods are used, we propose the following steps to choose a design:

1. Define p_m as whichever of p_l or p_u is closer to .5. Estimate ρ using existing data or existing literature using SAS with coverage estimates near p_m . If data are not available, consider how much variability is expected around p_m to select ρ .
2. Determine the cost of sampling a cluster and sampling each individual per cluster.
3. Iterate through choices of m and k to minimize cost for α , β , p_b , p_u , and ρ .

The same steps could be applied to the Pezzoli method, replacing ρ with σ . The R package accompanying this paper can help achieve Steps 1 and 3. Additionally, users can examine the sensitivity of their designs to the selected model (as in Tables 3 and 4) by calculating different possible designs with varied clustering parameterizations (after selecting the relevant values of p_b , p_u , α and β) using the R package. Knowledge of the design's sensitivity to the clustering parameterization contributes to understanding the robustness of the design.

In Step 1 above, determining ρ or σ based on p_m has two advantages: (1) σ and ρ are likely to decrease as p moves away from .5, so this clustering estimate is conservative (note that, using this algorithm, fixing σ is more conservative than fixing ρ); and (2) visualizing a highly skewed distribution in terms of a mean and standard deviation is not straightforward, and the skewness of the beta and binomial-scaled distributions increases as p moves away from .5. Returning to Fig 1, the beta and binomial-scaled distributions with mean $p = .5$ are symmetric, and one can easily visually check that $\sigma = .1$ by examining the spread of the distribution. With mean $p = .9$, there is no way to visually inspect these highly skewed distributions and conclude that $\sigma = .1$.

If the clustering parameters are underestimated in practice, the survey design will not have the desired precision; that is, the true α and β risks will be higher than those specified by design. One way to protect against this situation is to employ a sequential design [22, 29]. In a sequential design, the data are collected in multiple stages. In theory, the clustering parameters could be estimated as the data are collected and the design could then be changed to accommodate the observed clustering values. However, the variance of clustering parameters is high in small samples. Further, sequential designs are logistically more challenging. Future research should evaluate such sequential designs in simulation and in practice. Users should keep in mind that overestimating clustering parameters at the design phase is always more conservative and adds to the cost of the survey.

In practice, to select an LQAS survey design, the user first must decide whether SRS is feasible and practical. If SRS is used, survey design can proceed using a standard binomial model and the simple steps described in the Introduction. If SRS is not used, two-stage cluster sampling is a commonly used practical alternative. In some situations, the cluster sampling design can be ignored when small within-cluster sample sizes are used [17, 19–21] or when clustering is known to be negligible [22]. Survey design is much simpler when the cluster sampling design is ignored, because the user can act as though the data are a simple random sample and use standard LQAS procedures for design; that is, the user does not need to specify a clustering model or specify the clustering amount (i.e. ρ or σ). When the cluster sampling design cannot be ignored, choice between the cluster LQAS methods (Pezzoli, Hedt, and Hund) is not straightforward. Historically, ρ , not σ , is fixed as a function of p in design calculations [16, 25]. Uniting the methodology for LQAS surveys with other M&E estimation surveys may help de-mystify LQAS. The binomial-scaled model is not used in the statistics or survey sampling literature, possibly due to the discrete nature of this distribution. However, based on the empirical calculations in this article, choice between the Pezzoli, Hedt, and Hund methods should depend more upon the anticipated clustering patterns in the study population than the distributional specification (though prior information on clustering is likely quite limited). Hence, further research should attempt to characterize clustering patterns in specific applications and provide suggestions for best-practice cluster LQAS designs on a setting-specific basis. We only considered data examples in the context of vaccination coverage cluster LQAS surveys in this paper, but the validity of different clustering parameterizations should be evaluated in other contexts, for instance nutrition.

Cluster sampling in LQAS has also been recently discussed in the context of Large Country LQAS (LC-LQAS) [18, 30, 31]. These LC-LQAS surveys select SAS using cluster sampling with the goals of estimating prevalence over a large geographic region and classifying indicators at the SA-level. The LC-LQAS surveys are distinct from cluster LQAS surveys discussed in this paper which sample clusters *within* SAS, rather than sampling SAS.

Future work could explore accommodating uncertainty in ρ or σ at the design phase. In the data examples in this paper, both ρ and σ vary with p , complicating selection of a cluster LQAS method. All three methods could be modified to allow σ or ρ to vary as a function of p , though this modification would introduce an additional design parameter into the LQAS procedure. Future work should investigate the merits of considering designs with different clustering parameters corresponding to p_l and p_u or exploring different clustering parameterizations [25]. Additionally, all of the cluster surveys discussed in this paper rely on using a self-weighting sampling design. Future work should explore how to extend LQAS designs for unequal probability weights.

Appendix

Assume we have a model where $p_j \sim F$ with $E(p_j) = p$ and $Var(p_j) = \sigma^2$ and $X_j \sim Binomial(m, p_j)$. Let $\hat{p}_j = X_j/m$. To estimate σ^2 from summary statistics, note that:

$$Var(\hat{p}_j) = E\{Var(\hat{p}_j|p_j)\} + Var\{E(\hat{p}_j|p_j)\} = E\left\{\frac{p_j(1-p_j)}{m}\right\} + \sigma^2.$$

Additionally, $E\{p_j(1-p_j)\} = E(p_j) - E(p_j^2) = p(1-p) - \sigma^2$.

Then, $Var(\hat{p}_j) = \frac{p(1-p)-\sigma^2}{m} + \sigma^2$.

Solving for σ^2 , we have that $\sigma^2 = \frac{mVar(\hat{p}_j) - p(1-p)}{m-1}$.

Plugging in the empirical variance of the \hat{p}_j s for $Var(\hat{p}_j)$ and \hat{p} for p , the final estimator is:

$$\hat{\sigma}^2 = \frac{ms^2 - \hat{p}(1 - \hat{p})}{m - 1}.$$

This estimator is unbiased for σ^2 as m , k , or mk gets large (i.e. $Var(\hat{p}) \rightarrow 0$ or $m \rightarrow \infty$), since $E(\hat{\sigma}^2) = \sigma^2 + \frac{Var(\hat{p})}{m-1}$.

Note that $\hat{\sigma}^2 < 0$ when $s^2 < \hat{p}(1 - \hat{p})/m$ (within-cluster variance > between cluster variance). When estimates of σ are negative, we use $\hat{\sigma}^2 = 0$.

Author Contributions

Conceived and designed the experiments: LH EJB MP. Performed the experiments: LH. Analyzed the data: LH. Wrote the paper: LH EJB MP. Designed the software used in analysis: LH.

References

1. Dodge HF, Romig H. A Method of Sampling Inspection. The Bell System Technical Journal. 1929; 8 (398):613–631. doi: [10.1002/j.1538-7305.1929.tb01240.x](https://doi.org/10.1002/j.1538-7305.1929.tb01240.x)
2. Robertson SE, Valadez JJ. Global Review of Health Care Surveys using Lot Quality Assurance Sampling (LQAS), 1984–2004. Social Science & Medicine. 2006; 63(6):1648–1660. doi: [10.1016/j.socscimed.2006.04.011](https://doi.org/10.1016/j.socscimed.2006.04.011)
3. Pagano M, Valadez JJ. Commentary: Understanding Practical Lot Quality Assurance Sampling. International Journal of Epidemiology. 2010; 39(1):69–71. doi: [10.1093/ije/dyp406](https://doi.org/10.1093/ije/dyp406) PMID: [20139434](https://pubmed.ncbi.nlm.nih.gov/20139434/)
4. Olives C, Pagano M. Choosing a Design to Fit the Situation: How to Improve Specificity and Positive Predictive Values using Bayesian Lot Quality Assurance Sampling. International Journal of Epidemiology. 2013; 42(1):346–355. doi: [10.1093/ije/dys230](https://doi.org/10.1093/ije/dys230) PMID: [23378151](https://pubmed.ncbi.nlm.nih.gov/23378151/)
5. Hund L. New Tools for Evaluating LQAS Survey Designs. Emerging themes in epidemiology. 2014; 11 (1). doi: [10.1186/1742-7622-11-2](https://doi.org/10.1186/1742-7622-11-2) PMID: [24528928](https://pubmed.ncbi.nlm.nih.gov/24528928/)
6. Bennett S, Woods T, Liyanage WM, Smith DL. A Simplified General Method for Cluster-sample Surveys of Health in Developing Countries. World Health Stat Q. 1991; 44(3):98–106. PMID: [1949887](https://pubmed.ncbi.nlm.nih.gov/1949887/)
7. Binkin N, Sullivan K, Staehling N, Nieburg P. Rapid Nutrition Surveys: How Many Clusters are Enough? Disasters. 1992; 16(2):97–103. doi: [10.1111/j.1467-7717.1992.tb00383.x](https://doi.org/10.1111/j.1467-7717.1992.tb00383.x) PMID: [20958740](https://pubmed.ncbi.nlm.nih.gov/20958740/)
8. Spiegel PB. Who Should Be Undertaking Population-based Surveys in Humanitarian Emergencies? Emerging Themes in Epidemiology. 2007; 4(12). doi: [10.1186/1742-7622-4-12](https://doi.org/10.1186/1742-7622-4-12)
9. Lemeshow S, Robinson D. Surveys to Measure Programme Coverage and Impact: a Review of the Methodology Used by the Expanded Programme on Immunization. World health statistics quarterly. 1985; 38(1):65–75.
10. Pezzoli L, Pineda S, Halkyer P, Crespo G, Andrews N, Ronveaux O. Cluster-sample Surveys and Lot Quality Assurance Sampling to Evaluate Yellow Fever Immunisation Coverage following a National Campaign, Bolivia, 2007. Tropical Medicine & International Health. 2009; 14(3):355–361. doi: [10.1111/j.1365-3156.2009.02231.x](https://doi.org/10.1111/j.1365-3156.2009.02231.x)
11. Pezzoli L, Andrews N, Ronveaux O. Clustered Lot Quality Assurance Sampling to Assess Immunisation Coverage: Increasing Rapidity and Maintaining Precision. Tropical Medicine & International Health. 2010; 15(5):540–546.
12. Greenland K, Rondy M, Chevez A, Sadozai N, Gasasira A, Abanida E, et al. Clustered Lot Quality Assurance Sampling: a Pragmatic Tool for Timely Assessment of Vaccination Coverage. Tropical Medicine & International Health. 2011; 16(7):863–868. doi: [10.1111/j.1365-3156.2011.02770.x](https://doi.org/10.1111/j.1365-3156.2011.02770.x)
13. Pezzoli L, Conteh I, Kamara W, Gacic-Dobo M, Ronveaux O, Perea WA, et al. Intervene before Leaving: Clustered Lot Quality Assurance Sampling to Monitor Vaccination Coverage at Health District Level before the End of a Yellow Fever and Measles Vaccination Campaign in Sierra Leone in 2009. BMC public health. 2012; 12(415). doi: [10.1186/1471-2458-12-415](https://doi.org/10.1186/1471-2458-12-415) PMID: [22676225](https://pubmed.ncbi.nlm.nih.gov/22676225/)
14. Hedt-Gauthier BL, Mitsunaga T, Hund L, Olives C, Pagano M. The Effect of Clustering on Lot Quality Assurance Sampling: a Probabilistic Model to Calculate Sample Sizes for Quality Assessments. Emerging themes in epidemiology. 2013; 10(1). doi: [10.1186/1742-7622-10-11](https://doi.org/10.1186/1742-7622-10-11) PMID: [24160725](https://pubmed.ncbi.nlm.nih.gov/24160725/)
15. Hund L, Pagano M. Extending Cluster Lot Quality Assurance Sampling Designs for Surveillance Programs. Statistics in medicine. 2014; 33(16):2746–2757. doi: [10.1002/sim.6145](https://doi.org/10.1002/sim.6145) PMID: [24633656](https://pubmed.ncbi.nlm.nih.gov/24633656/)

16. Eldridge SM, Ukoumunne OC, Carlin JB. The Intra-Cluster Correlation Coefficient in Cluster Randomized Trials: A Review of Definitions. *International Statistical Review*. 2009; 77(3):378–394. doi: [10.1111/j.1751-5823.2009.00092.x](https://doi.org/10.1111/j.1751-5823.2009.00092.x)
17. Valadez JJ, Hage J, Vargas W. Understanding the Relationship of Maternal Health Behavior Change and Intervention Strategies in a Nicaraguan NGO Network. *Social science & medicine*. 2005; 61(6):1356–1368. doi: [10.1016/j.socscimed.2005.02.002](https://doi.org/10.1016/j.socscimed.2005.02.002)
18. Hund L, Northrop-Clewes CA, Nazario R, Suleymanova D, Mirzoyan L, Irisova M, et al. A Novel Approach to Evaluating the Iron and Folate Status of Women of Reproductive Age in Uzbekistan after 3 Years of Flour Fortification with Micronutrients. *PloS one*. 2013; 8(11). doi: [10.1371/journal.pone.0079726](https://doi.org/10.1371/journal.pone.0079726)
19. Deitchler M, Valadez JJ, Egge K, Fernandez S, Hennigan M. A Field Test of Three LQAS Designs to Assess the Prevalence of Acute Malnutrition. *International Journal of Epidemiology*. 2007; 36(4):858–864. doi: [10.1093/ije/dym092](https://doi.org/10.1093/ije/dym092) PMID: [17517808](https://pubmed.ncbi.nlm.nih.gov/17517808/)
20. Deitchler M, Deconinck H, Bergeron G. Precision, Time, and Cost: a Comparison of Three Sampling Designs in an Emergency Setting. *Emerging Themes in Epidemiology*. 2008; 5(6). doi: [10.1186/1742-7622-5-6](https://doi.org/10.1186/1742-7622-5-6) PMID: [18454866](https://pubmed.ncbi.nlm.nih.gov/18454866/)
21. Olives C, Pagano M, Deitchler M, Hedt BL, Egge K, Valadez JJ. Cluster Designs to Assess the Prevalence of Acute Malnutrition by Lot Quality Assurance Sampling: a Validation Study by Computer Simulation. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. 2009; 172(2):495–510. doi: [10.1111/j.1467-985X.2008.00572.x](https://doi.org/10.1111/j.1467-985X.2008.00572.x)
22. Stroh G, Birmingham M. Protocol for Assessing Neonatal Tetanus Mortality in the Community using a Combination of Cluster and Lot Quality Assurance Sampling. Geneva: Department of Vaccines and Biologicals, World Health Organization; 2002. URL: <http://apps.who.int/iris/handle/10665/67193>, accessed 2/7/14.
23. Lohr S. Sampling: Design and Analysis. Boston: Cengage Learning; 2009.
24. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2014. Available from: <http://www.R-project.org/>.
25. Crespi CM, Wong WK, Wu S. A New Dependence Parameter Approach to Improve the Design of Cluster Randomized Trials with Binary Outcomes. *Clinical Trials*. 2011; 8(6):687–698. doi: [10.1177/1740774511423851](https://doi.org/10.1177/1740774511423851) PMID: [22049087](https://pubmed.ncbi.nlm.nih.gov/22049087/)
26. Gulliford M, Adams G, Ukoumunne O, Latinovic R, Chinn S, Campbell M. Intraclass Correlation Coefficient and Outcome Prevalence Are Associated in Clustered Binary Data. *Journal of Clinical Epidemiology*. 2005; 58(3):246–251. doi: [10.1016/j.jclinepi.2004.08.012](https://doi.org/10.1016/j.jclinepi.2004.08.012) PMID: [15718113](https://pubmed.ncbi.nlm.nih.gov/15718113/)
27. Minetti A, Riera-Montes M, Nackers F, Roederer T, Koudika MH, Sekkenes J, et al. Performance of Small Cluster Surveys and the Clustered LQAS Design to Estimate Local-level Vaccination Coverage in Mali. *Emerging Themes Epidemiology*. 2012; 9(6). doi: [10.1186/1742-7622-9-6](https://doi.org/10.1186/1742-7622-9-6)
28. Cleveland WS, Grosse E, Shyu WM. 8. In: *Statistical Models in S*. Boca Raton: Wadsworth & Brooks/Cole; 1992.
29. Myatt M, Bennett DE. A Novel Sequential Sampling Technique for the Surveillance of Transmitted HIV Drug Resistance by Cross-sectional Survey for Use in Low Resource Settings. *Antiviral therapy*. 2008; 13(2):37–48. PMID: [18575190](https://pubmed.ncbi.nlm.nih.gov/18575190/)
30. Hedt BL, Olives C, Pagano M, Valadez JJ. Large Country-Lot Quality Assurance Sampling: a New Method for Rapid Monitoring and Evaluation of Health, Nutrition and Population Programs at Sub-national Levels. Washington, DC: World Bank Group; 2008. <http://documents.worldbank.org/curated/en/2008/05/18673016/large-country-lot-quality-assurance-sampling-new-method-rapid-monitoring-evaluation-health-nutrition-population-programs-sub-national-levels>, accessed 1/12/2015.
31. Biedron C, Pagano M, Hedt BL, Kilian A, Ratcliffe A, Mabunda S, et al. An Assessment of Lot Quality Assurance Sampling to Evaluate Malaria Outcome Indicators: Extending Malaria Indicator Surveys. *International Journal of Epidemiology*. 2010; 39(1):72–79. doi: [10.1093/ije/dyp363](https://doi.org/10.1093/ije/dyp363) PMID: [20139435](https://pubmed.ncbi.nlm.nih.gov/20139435/)