LQAS in Health Monitoring – Insights from a Bayesian Perspective

David Kwamena Mensah[1] and Paul Hewson[2,*]

[1]*African Institute for Mathematical Sciences, Currently National University of Singapore*

[2]*School of Computing and Mathematics, Plymouth University, Plymouth PL4 8AA, UK*

Abstract: Lot Quality Assurance Sampling (LQAS) is strongly advocated for use in monitoring the health status of populations, largely in the developing world. It is advocated both for the monitoring of small areas as well as for making global assessments of the health status of a larger region. This paper contrasts the interpretation offered by LQAS methods to that offered by Bayesian hierarchical models. It considers applications to previously reported local area data and presents a reanalysis of published data on vaccine coverage in Peru as well as HTLV-1 prevalence in Benin. The desirability of using Bayesian methods in the field may be challenged; nevertheless this work amplifies previously expressed concerns about the way the LQAS method can be used. It raises questions about the ability of the LQAS approach to make, sufficiently often, the correct decisions in order to be useful in monitoring health programmes at the local level.

Keywords: Cluster Sampling, Bayesian Hierarchical Model, Overdisperson, Hypergeometric distribution, Classification.

BACKGROUND

In the last few decades, Lot Quality Assurance Sampling (LQAS) has been applied in a wide range of healthcare settings [1]. LQAS is a well-established industrial technique intended for processes that are under statistical control [2]. Early work on LQAS predates Neyman and Pearson's work on hypothesis testing [3]. LQAS aims to make a correct classification as to whether a batch of manufactured products can be determined as "acceptable" or "not acceptable" in an industrial setting where effort and expense have been dedicated to ensuring that the process being monitored is under statistical control. This means that features of the process being monitored, such as the variance, are well characterized before the sampling scheme is established. Indeed, in the manufacturing context considerable effort may already have been expended to minimize variation in the process. Another feature of manufacturing is that the accuracy of classifying sampled objects as "acceptable" or "not acceptable" may be much higher than is possible in a healthcare setting.

There are currently many references which describe LQAS designs being used within healthcare, some of which refer to making local assessments of healthcare status (vaccine coverage, disease prevalence) in a manner analogous to the way in which LQAS is used in industry [1]. This paper will consider a previously published application to disease prevalence in some

detail [4]. There are further references which describe the use of LQAS collected data as a particular cluster sampling approach and the analysis is conducted to give a global estimate of prevalence [5]. Thus it is possible to see applications of the basic LQAS sample being used either in a classification way or in a survey sampling way. There have been extensive contributions to the literature in the last few years regarding LQAS including editorial comment in leading journals [6]. This paper seeks a further exploration of the closeness of the healthcare and industrial conceptualisations of the LQAS strategy. It is interesting to note that there has been some debate in the literature as to whether LQAS can be regarded as a classification or an hypothesis testing method.

This paper will approach the use of LQAS from a Bayesian perspective. This is motivated by the desire to obtain a posterior distribution for the quantity of interest which can be interrogated to give a direct estimate of the supplier and client risks when classifying a small area. In LQAS, these risks have to be pre-specified in the design and selection of sample size and decision criteria. There have already been intriguing suggestions in the literature for a Bayesian interpretation of an LQAS design [7]. The advantage of using a modeling approach (whether Bayesian or not) is that additional contextual information can be included. This is clearly important when estimating the prevalence of a condition that can vary considerably by strata where it is not possible to obtain LQAS samples that have been balanced with respect to these strata. The second case study (HTLV-1 prevalence in Benin) clearly shows an example where population strata are important. Another potential advantage of modeling is

*Address correspondence to this author at the School of Computing and Mathematics, Plymouth University, Plymouth PL4 8AA, UK; Tel: ++44(0)1752 586870; Fax: 0++44(0)1752 586300; E-mail: paul.hewson@plymouth.ac.uk

that we can account for uncertainty in missing covariates, for example by including hierarchical, temporal or spatial random effects. The second case study clearly shows that the HTLV-1 data are overdispersed relative to basic probability models such as the Hypergeometric and Binomial. Hence part of the discussion will review the implications of overdispersion for the correct performance of LQAS decision rules.

Derivation of LQAS Sampling Plans

LQAS sampling plans are based upon classifying a whole batch as "acceptable" by determining whether the number of "marks" from that batch exceeds a pre-determined amount. The original proposals for LQAS assumed that these marks could be considered a random variable following a Hypergeometric distribution. Here, the idea is that a sample of size *n* has been taken without replacement from a population of size *N* where that population consists of *K* individuals of state 1 (having the mark) and *N-K* individuals of state 0 (not having the mark). Letting *X* denote a random variable that counts the number of individuals in state 1 in a sample of size *n*, then *X* is said to have a Hypergeometric distribution if the probability mass function is:

$$P(X = x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}.$$

LQAS then informs the construction of decision rules based on maintaining the probability of misclassifying a batch below pre-specified levels for a population with different proportions of

$$\frac{(N-K)}{N}.$$

It could be possible to consider this decision as a kind of composite hypothesis test where the classification is based on selecting values of *n* and *k* such that the risk is maintained below a threshold for a range of values $\frac{N-K}{N}$ to $\frac{N}{N}$.

A key point to note about the use of the Hypergeometric distribution is that it makes a very strong statement about the variance, namely that

$$Var(X) = \frac{nk(N-k)(N-n)}{N^2(N-1)}$$

In a sample survey setting it is possible to model the number of marks in a sample using a Binomial distribution. Both the Binomial and Hypergeometric distributions are generalizations of a Bernoulli distribution for finite populations; the former appropriate for sampling with replacement and the latter appropriate for sampling without replacement. Indeed, under certain conditions a Binomial random variable *Y* has a similar distribution to a Hypergeometric random variable X in that $P(X \le k) = P(Y \le k)$. This point will be expanded in the discussion but for now it should be noted that many observed datasets feature overdispersion – a situation where the variance of the data is much larger than that predicted by a chosen probability model [8]. Common solutions to his include using a slightly different probability model; for example substituting the Binomial distribution by a Quasi-Binomial or a Beta-Binomial. Alternatively, it is common to use random effect models as a means of accounting for this failure of model assumptions. This paper will demonstrate that overdispersion exists in previously published LQAS data and illustrate how to account for this by using a carefully constructed model.

The problem of overdispersion may be more likely in observational datasets such as those used in LQAS healthcare settings. There are some notable differences between industrial process control and monitoring health outcomes. In an industrial setting, the parameters of the process being monitored are often well understood and usually some effort has been expended towards the goal of reducing variation. LQAS is therefore used to confirm that a process is still operating within acceptable limits. In many published healthcare applications, even those working under a Monitoring and Evaluation framework the goal is different – monitoring service delivery or prevalence in order to reduce variation in a process is not an aspect of the context. Indeed, it seems likely that a human system (where different localities may operate in a different manner or where the prevalence may alter among different subsets of the population) the process is far more variable than the industrial systems where LQAS is used. This will be specifically illustrated in the second case study where there are strong age-gender effects in prevalence. The point of this caveat is that the LQAS decision rules have not been constructed to account for this stratification.

The structure of this paper is therefore to set out the modeling methods and then apply to these to case studies; two previously published applications of LQAS to healthcare monitoring. The first case study concerns

vaccine coverage in Peru. Whilst this is a very early exposition of the LQAS method in healthcare it is one where the authors report that the technique was "successful' [9]. LQAS practice and training materials have undoubtedly developed considerably since this study. The second case study concerns the prevalence of HTLV-1 in Benin, where LQAS was used to determine areas of low infection rate [4]. The implication of finding areas of low prevalence is either that healthcare resources can be diverted to areas of higher prevalence or it can be used to identify areas of low prevalence from which high prevalence areas might learn best practice. In the first case study, "high" vaccine coverage is the goal whereas in the second "low" prevalence is the goal. Clearly though, the case studies are otherwise symmetric. Part of the reason for selecting these two case studies is that they are two of the few studies that publish sufficient data to enable re-analysis. The paper will then apply what are now very well established Bayesian models to these data, and then discus the implications of the modeling.

METHODS

As briefly mentioned, the first case study is based on an early report of LQAS used to confirm vaccine non-coverage in Peru [9]. Baseline coverage was assessed and then the target was to classify areas that had achieved greater than 80% coverage. In making this determination, the goal was to do this with no more than an 8% risk of being misclassified as having high coverage when the coverage was in fact less than 80%. An LQAS scheme was selected based on *n=9* and *d=3*; in other words, based on a sample size of 9 areas where identified as being at risk of below target coverage if three or less children appeared not to have been vaccinated. Using this decision rule, eight of twelve small areas were found to be below 80% coverage. Subsequently a follow up survey was conducted using a slightly different LQAS rule targeting 90% coverage with a sampling plan *n* = 9 and *d* = 2, i.e., if two or more children in a sample size of 12 appeared not to be vaccine protected then the area was regarded as inadequately protected.

The second case study is based on a re-analysis of data on HTLV-1 prevalence taken from communes in the Atacora region in the north west of Benin in 1998 [4]. The authors state that whilst cluster sampling can provide information on a region of interest, they wished to identify small areas with higher prevalence levels in order to inform health policy planning. At the time of the study, Atacora region was divided into 17 sub-

prefectures and 73 communes. In total, 36 communes were examined. The use of LQAS manuals lead to a sample size of n = 65 with a decision rule of d = 0, in order to detect prevalence greater than 4% (this being twice the prevalence seen in previous studies). In other words, an area was classified as being above 4% prevalence if more than 1 person in a sample of 65 was observed to be HTLV-1 sero-positive. In the original publication, 25 communes were classified as high right and 11 were classified as acceptable risk.

The methods used in this paper are standard applications of Bayesian hierarchical models [10]. It is possible to use a conventional generalised linear model formulation for these disease/vaccination counts, where y_i denotes either the number of children with vaccine coverage or the number of respondents with HTLV-1. These can be modelled as a Binomial random variable with n_i=9,12,65 for the first and second vaccine samples and the HTLV-1 sample respectively. Therefore for areas *i*=1,…*N* (where *n=9,12* for the first and second Peru vaccine coverage study and *n*=65 for the Benin HTLV-1 study the observed data model can be given as:

$$Y_i \sim Binom(n_i, p_i)$$

Interest then surrounds fitting a suitable model to p_i. A conventional logistic link would therefore model

$$\text{logit}(p_i) = \beta_0 + \beta_i,$$

where β_0 is an intercept term and β_i is a random effect for area *i*. For a Bayesian implementation priors have to be selected for these parameters and it is convenient to assume a non-informative prior for the intercept so that

$$\beta_0 \sim N(0,1000).$$

A little more attention needs to be paid to the structure of the random effect. For the Peru vaccine coverage study there is little other information and it is only possible to have a random effect for each area. For the Benin HTLV-1 data however it is possible to allow for hierarchical structuring (communes within sub-arrondissements) and even to allow for spatial correlation. The choice found suitable for these particular data will be explained in the results section; here it suffices to note that there are rich possibilities for structuring these models based on the available meta-data. More importantly, it is possible to model both local and global prevalence simultaneously.

Ideally, these models could be extended to model individuals (as a Bernoulli response) and hence make use of relevant individual covariates such as age and gender. It is unlikely that any small sample will be perfectly balanced with respect to potentially informative covariates. Unfortunately, such individual data has not been made available. The problem of stratification is however clear from these data. An alternative method for allowing for overdispersion is to use a mixture model. It is possible to fit a simple mixture model to the intercept so that

$$\beta_{1z} \sim Normal(\mu_z, \sigma_z),$$

where *z* denotes membership of mixture group 1,2,… with an identifiability constraint $\mu_1 < \mu_2$ and $\sigma_1 = \sigma_2$.

All models have been fitted using Markov Chain Monte Carlo as implemented by the pymc software [11]. Details on validation of the model fitting are given in the results section.

RESULTS

The first set of results concerns the re-analysis of the Peru vaccination data [9]. Firstly, results are given for a simple model fitted to the first time point only, in other words

$$Y_i \sim Binom(p_i, 9)$$

and

$$\text{logit}(p_i) = \beta_0 + \beta_{1area}$$

where β_0 is assumed a *Normal* (0, 1000) prior and β_{1area} are 12 area specific random effects also with *Normal* (0, 1000) priors. The McMC sampler was run for 5,000 iterations as a burn in and then 5,000 samples were retained. Standard checks were made to ensure that the McMC algorithm had converged [13]. The posterior distribution for p_i is summarised by directly computing the posterior probability that p_i was greater than 80%. These results are shown in Table **1** below. It can be seen for example that Lot 1, which was accepted by LQAS has a posterior probability of acceptable vaccine coverage of only 40%. Conversely, lot 12 (which was also accepted by LQAS) has a posterior probability of having acceptable vaccine coverage of 0.62.

One reason for discrepant results between the LQAS decision and the decision based upon the posterior distribution from a Bayesian model is that in

Table 1: Summary of Peru Vaccination Data at Time Point 1: Original LQAS Decision and Posterior Probability that Vaccine Coverage is Greater than 80%

Lot Number	Accepted by LQAS	Posterior probability of acceptable coverage
1	Yes	0.40
2	Yes	0.07
3	No	0.00
4	Yes	0.19
5	No	0.00
6	Yes	0.20
7	No	0.00
8	Yes	0.07
9	Yes	0.21
10	Yes	0.21
11	No	0.02
12	Yes	0.62

fitting the entire data are used essentially to estimate a global prevalence and then the individual lots are allowed to vary from this assuming each to be an exchangeable sample from the same population. This means that individual samples have their posterior distributions shrunk towards the population values. This may or may not seem sensible. In industrial LQAS, the population has been well characterised prior to establishment of decision rules and decision rules are intended to detect a change in operating conditions. In the case of many healthcare applications, the population have not been so characterised and it may be appropriate to conduct a modeling exercise which aims to estimate the overall population characteristics.

As stated, several technical checks were made on the performance of the McMC model fitting algorithm essentially looking for convergence [12]. However, it is also important to check the statistical fit of the model. A convenient way of doing this is *via* the posterior predictive distribution; to generate simulated data

$$Y_i^{(*)} \sim Binom(p_i, 9 | M)$$

from the model M fitted to the data. Given simulations of the posterior predictive distribution it is trivial to check for example that the observed data are reasonable given the 95% credible intervals for the posterior predictive distribution. These results are given in Table **2**.

Table 2:   Summary of Posterior Predictive Distribution for Time Point 1 Data Fitted to Peru Vaccination Data. Three Columns Show 95% Interval and Median. Final Column Denotes the Proportion of Posterior Predictions which Yield an LQAS "Accept" Signal

Lot	2.5%	50%	97.5%	Proportion of posterior predictions yielding LQAS accept signal
1	0	2	6	0.793
2	0	3	7	0.524
3	3	6	9	0.076
4	0	3	7	0.672
5	2	6	9	0.137
6	0	3	7	0.658
7	1	5	8	0.234
8	0	3	7	0.521
9	0	3	7	0.667
10	0	3	7	0.656
11	1	4	8	0.366
12	0	1	5	0.891

What Table **2** shows is that for example lot 5 has a 95% interval for the posterior predictive distribution of (2,9) which is consistent with the observed data. If the observed data were indeed one realisation from this distribution, this distribution yields 14% of samples which have an accept signal despite the fact that the LQAS decision made on the original data was reject.

Clearly the next step is to extend the model so that both time points can be considered. The appropriate model can be denoted:

$$Y_{it} \sim Binom(p_{it}, n_{it}),$$

with $n_{i1} = 9$ and $n_{i2} = 12$. Then

$$\text{logit}(p_{it}) = \beta_0 + \beta_{1area} + \delta x_t$$

where x_t is an indicator variable equal to zero for the first time point and 1 for the second time point. In the Bayesian model δ is assumed to follow a *Normal(0,1000)* distribution. Models were fitted exactly as before and the posterior estimate for δ had a 95% credible interval (-0.74,0.16). This is represented in Figure **1** (along with a trace plot suggestive of convergence).

It is convenient to summarise the posterior distribution for the various values of p_{it} visually where the values of p_i from the single time point model can be superimposed. These are given in Figure **2**. This illustrates the natural way in which Bayesian models can be updated sequentially given additional data. The intervals for p_{i1} in the two time point model have been shifted relative to those of p_i the single time point model because they have been shrunk towards the assumed overall value based on evidence from two time points. Some narrowing of the credible intervals is also seen as the vaccine coverage prevalence is being estimated from more data.
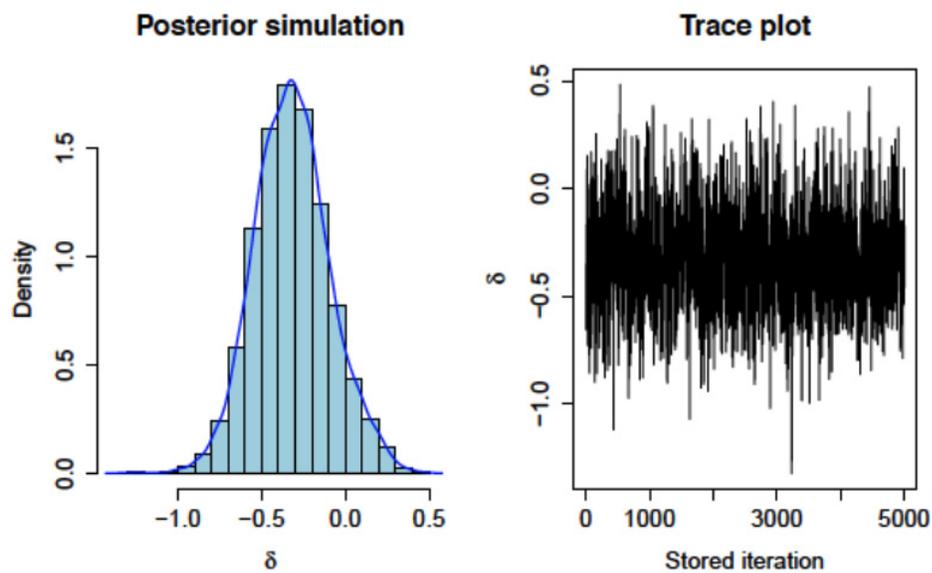
Figure 1: Summary of posterior distribution of parameter δ from combined Peru vaccination data model. Left plot is a histogram/density plot of the posterior distribution, right plot is a trace plot showing values by iteration for a subset of the interations.
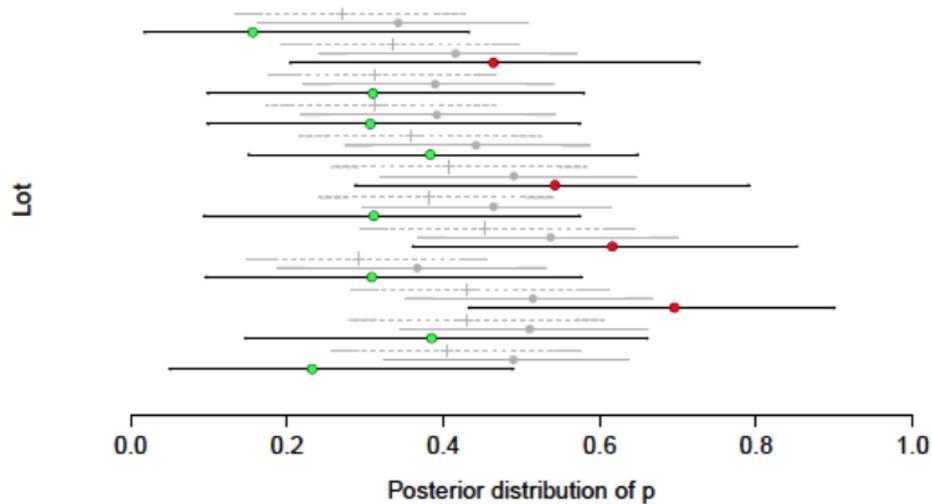
Figure 2: Summary of posterior distribution (posterior median as dots, lines showing 95% credible interval). Three summaries are presented. (a) solid black lines show the posterior distribution for the initial survey based on nine data points in each area whereas grey lines show the posterior for the combined model. (b) Solid grey lines show the posterior distribution for p_{i1} (based on nine data points in each area in a combined model) and dashed grey lines show the posterior distribution for p_{i2} (based on 12 data points in each area in a combined model). Dots coloured green denote an LQAS accept decision, dots coloured red denote an LQAS reject decision.

As with the single time point model it is also possible to generate a posterior predictive distribution. Results are given in Table **3**. They show for example that in Lot 12, if the model is correct, it is not unreasonable to see a value of 0 children unvaccinated, neither is it unreasonable to see a value of 8 children unvaccinated. In all, 28% of the posterior predictive simulations yielded values that would lead to an LQAS accept decision. Yet the posterior predictive distribution does not place much probability that the vaccine prevalence is in fact greater than 90%.

Some authorities have expressed the concern that LQAS decision rules are not being correctly applied in situations such as this one [13]. In other words the decision to accept a lot based on evidence that the underlying p_i is greater than 80/90% not by finding evidence it is higher but by failing to find evidence that it is lower. However, LQAS methodology and training materials have certainly developed since this early publication. What has been noted in this reanalysis is that statistical models consider sample values as realisations from some underlying population and that when a model assuming exchangeability is fitted the posterior probability of prevalence is seen in a different light. The OC characteristics of the LQAS rule may not be as anticipated given the variation in the data. This can be more specifically illustrated in the second case study.

Having established that the point of modelling is that we no longer need to assume a well-controlled,

homogenous-variance industrial process it is interesting to consider the role of stratification. Whilst monitoring HTLV-1 prevalence may not be a mainstream application of LQAS it does represent a published account of the technique. As before, full data are not available but there is at least data on localities and it is possible to obtain administrative maps of Benin. The original paper also gave a table indicating prevalence by strata which is informative. As before, we model the count of HTLV-1 sero-positive respondents y_{ij} in each commune $I = 1, \ldots, 36$ in each sub-prefecture $j=1,\ldots,7$ as

$$Y_{ij} \sim Binom(p_{ij}, 65)$$

with

$$\text{logit}(p_{ij}) = \beta_0 + \beta_{1area} + \beta_2 x_{i2}$$

where β_0 is the intercept and assumed a *Normal(0,1000)* prior in the Bayesian model, β_{1area} is a vector of seven independent *Normal(0,1000)* random effects for the sub-prefecture from which the communes were drawn. β_2 is also assumed *Normal(0,1000)* and x_{i2} is an indicator variable set to 1 if the commune is urban and zero otherwise. It should be noted that this formulation treats commune as exchangeable given their parent arrondissement. Whilst different structures are possible (including spatial correlation between sub-arrondissements, or random effects for each commune) this was the

Table 3:

Lot	2.5%	50%	97.5%	Proportion of samples giving LQAS accept signal
Time point 1 (targeting 80% coverage)				
1	1	4	8	0.292
2	2	5	8	0.241
3	1	5	8	0.249
4	0	3	7	0.554
5	2	5	8	0.224
6	1	4	7	0.354
7	1	4	8	0.297
8	1	4	7	0.406
9	1	3	7	0.511
10	0	3	7	0.526
11	1	4	7	0.448
12	0	3	6	0.635
Time point 2 (targeting 90% coverage)				
1	2	6	10	0.062
2	2	6	10	0.045
3	2	6	10	0.046
4	1	4	8	0.213
5	2	6	11	0.035
6	2	5	10	0.076
7	2	6	10	0.064
8	1	5	9	0.106
9	1	4	9	0.175
10	1	4	9	0.170
11	1	5	9	0.134
12	0	4	8	0.280

Table 4: DIC Measure for the Hierarchical Model Fitted to the Benin HTLV-1 Data with Random Effects for Sub-Prefecture and an Indicator Variable for "Urban" or Otherwise

Posterior mean deviance	pD	Penalised deviance
142.6	7.64	150.2

preferred model based on the Deviance Information Criteria (given in Table **4**) [14].

The posterior distributions for p_i are given in Figure **3**.

Figure **3** instantly highlights the different interpretation given to the data by the Bayesian hierarchical model and the LQAS decision. The LQAS decision rule regards any observation of 0 as having "low" (below 4%) prevalence. This carries with it the implication that resources could be removed from green areas, or that red areas should learn from green areas. Therefore, in Boukoumbe there are three communes classified as having low prevalence; Dipoli, Tabota and Koronitere. In contrast, the Bayesian model regards these as exchangeable observations from within Boukoumbe and places some weight on the fact that there were 20 cases observed across the seven communes (and indeed places some weight on the overall prevalence in North West Benin). Consequently, the posterior distributions overlap the 4% decision rule (indeed, the
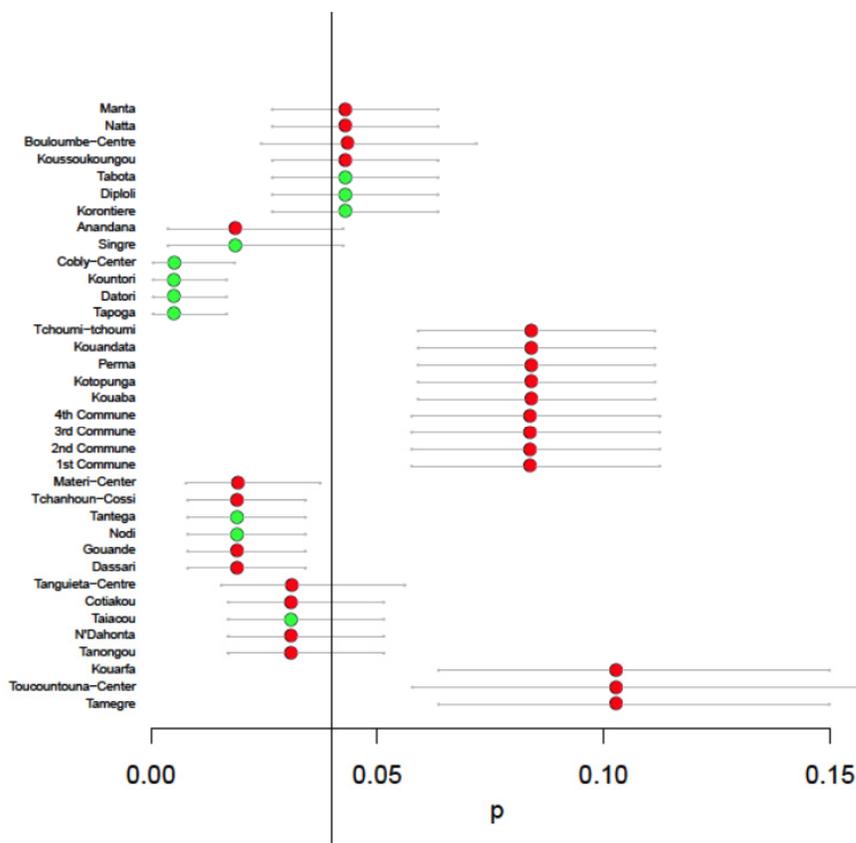
Figure 3: Posterior median and 95 percentiles for p_i for the first Benin HTLV-1 model. Dots denote median, lines denote 95% credible interval. Dots coloured red denote an LQAS reject decision and green an LQAS accept decision. The target prevalence (4%) is denoted by a vertical grey line.

posterior median prevalence is estimated as greater than 4%). Conversely, in Materi the opposite happens. A number of communes have been identified as having high prevalence (Materi-Central, Tohanhoun-Kossi, Gouande and Dassari). Yet the hierarchical model, by placing weight on the fact that a total of ten cases were observed in the six communes and hence the 95% credible interval for the posterior distribution for prevalence for each commune is entirely below the 4% target. Clearly, a wide range of model structures are possible; but the point is made that this is a very different interpretation of the data to that obtained from the use of an LQAS decision rule. There is obviously no way to reconcile the differences through the use of data analysis, but it is worth paying careful attention to the assumptions in each approach. The Bayesian hierarchical model depends on the assumption of exchangeability within each sub-prefecture the LQAS method depends on the assumption of *iid* observations from a Hypergeometric distribution.

Mixture Models

The introduction referred to the problem of overdispersion and highlighted the way a statistical model (Bayesian or otherwise) could deal with overdispersion in a way that is not immediate to the extant LQAS decision rules. The Benin data were chosen precisely because at least some information is given to illustrate the concern about stratification. Table **3** in the original paper gives information on HTLV-1 prevalence by age and sex [4]. At either extreme the authors report prevalence of around 2% (0.8 to 4.6%) for adults under 30 and for adults over 60 they report 4.8% for males and 18.2% for females. Whilst data on age / gender for each of the 36 samples is not available it is easy to see that the data are overdispersed. The sample standard deviation of the published data is 2.95 which is considerably higher than the theoretical standard deviation for either Binomial or Hypergeometric distribution with *n=65* and a *p* (directly or implied) of 0.04. Indeed, in the absence of raw data it is possible to perform Monte Carlo simulations where 36 samples of size 65 are drawn from a total sample of size 2340 with 111 seropositive individuals where the age gender labels are set up to match the original paper and permuted. In none of 1000 such simulations was it possible to find a sample standard deviation as high as 2.95 and indeed, a more direct simulation

Table 5:

Posterior mean deviance	pD	Penalised deviance
112.8	30.17	142.9

shows that finding 11 samples with zero cases is unusual. This may lead to a challenge of the exchangeability assumption. One answer to this could be to use a non-parametric mixture model where as before the observed data model is

$$Y_{ij} \sim Binom(p_{ij}, 65)$$

but now p_{ij} is modelled as:

$$\text{logit}(p_{ij}) = \sum_{k=1}^{K} \pi_k \beta_{0k} + \beta_{1j} + \beta_x x_{i2.}$$

This is essentially the same as the previous model except that there is now a mixture of intercepts. π_k denotes the mixing probabilities with $\sum_{k=1}^{K} \pi_k = 1$, and β_{0k} are now *K* independent *Normal(0,1000)* random variables with the constraint that $\beta_{0k} < \beta_{0(k+1)}$. In practice, it was only necessary to use two mixture components and the overall model fit results are given in Table **5**. Comparing Tables **4** and **5** shows that the latter model fits better overall, even when penalised (the penalised deviance is lower) although the "effective number of parameters" penalty is somewhat larger.

The posterior distribution for p_i for the mixture model is given in Figure **4** where it can be seen the results are largely similar to those for the simpler model. However, these small differences are important. Tabole, Dipoli and Kountori have posterior distributions shifted toward zero (due to borrowing strength from a mixture group more than from their parent sub-arrondisement) although in this case the 95% credible interval still overlaps the target 4% prevalence. Materi-Center, Tchanhoun-Cossi now have credible intervals which overlap the target – again because of balancing the way strength is borrowed from the mixture group and the parent sub-arrondissement. Whilst it is argued that the mixture structure is a non-parametric feature, the right panel of Figure **4** gives the posterior probability of belonging to the second group (with the larger parameter for β_{0k}.

Whilst conventional convergence checks have been carried out on these models, as with the Peru vaccine models it is sensible to compare the model fit. As before, the posterior predictive distribution can be plotted and Figure **5** presents these results visually for the mixture model.

The left panel suggests that the data might be compatible with the model, observations generally lying within the 95% posterior predictive distribution. Assuming this is a reasonable model, the right panel essentially shows the posterior probability of an LQAS
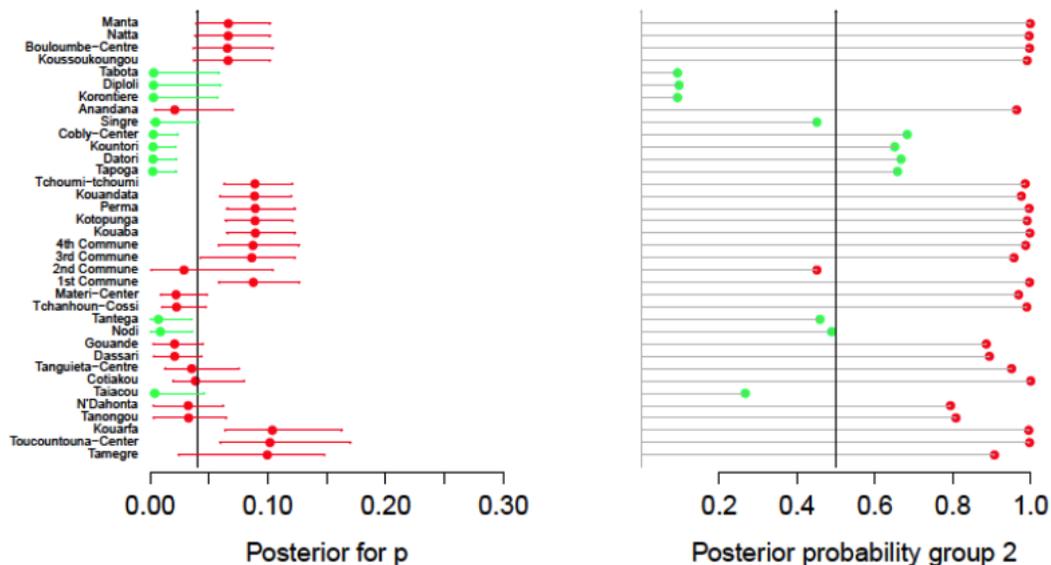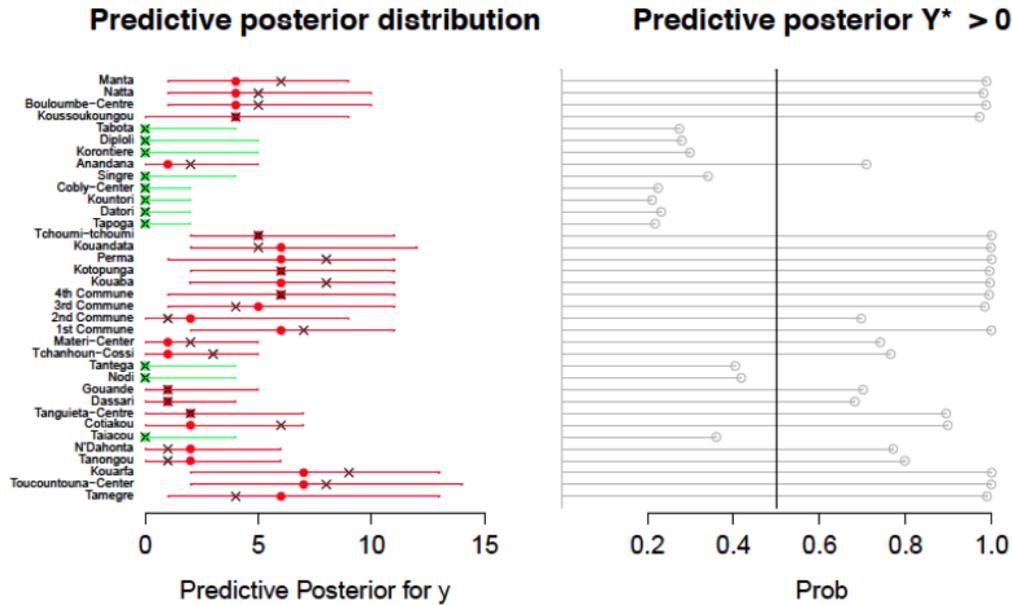

Figure 4:

Figure 5: Posterior predictive distribution for $Y^{(*)}$ for Benin HTLV-1. Left panel superimposes a 95% credible interval for the posterior with dots giving the posterior median. Dots coloured according to the original LQAS decision rule with red denoting reject and green denoting accept. The reported data (counts out of 65) are denoted by a cross. The right panel denotes the posterior predictive probability of the model data generating mechanism generating for $Y^{(*)} > 0$ (i.e., yielding an LQAS reject signal).

reject signal being generated from a sample taken from that commune. These posterior probabilities can be contrasted with the intention of limiting the error rate from LQAS decision rules. This finding will be considered in more detail in the discussion.

DISCUSSION

This paper has re-examined the way that LQAS can be used for local decision-making. It is important to be very clear about the way LQAS is used in healthcare. Firstly, there is the application to decision making in local areas in a healthcare system by classification. Secondly, there is application to prevalence estimation at a large area level by using the LQAS collected data as a cluster sample. The first clearly relates to the industrial application of LQAS except that industrial processes by their nature can be designed and refined so as to minimize variance. In human / environmental systems this control is not available and the detailed process information is rarely available. It could be posited that it may be sensible to consider prevalence / vaccine coverage as a process which is overdispersed relative to standard statistical theory. This paper demonstrates that data used for LQAS methods may indeed be overdispersed and hence provides a further insight into the use of LQAS decision rules. Certainly there have already been proposals for the use a Bayesian approach to constructing LQAS decision rules [7]. Indeed there have been proposals for careful

use of prior information in the construction of LQAS decision rules [15]. However, the proposal here is more subtle. Given that people use LQAS in a cluster sample survey approach, the suggestion is to use the data in that way to assess overall prevalence. Having done this, it is possible to extract information from relevant parameters (directly incorporated or derived) to estimate prevalence in small areas. The idea is that in the absence of knowledge of the overall process characteristics (vaccine coverage or disease prevalence) this has to be estimated. The model has to be structured in a way that is reasonable and this paper has demonstrated that there are differences between allowing areas to borrow strength from a parent area or to borrow strength from a non-parametric mixture component. This paper has demonstrated that it is possible to borrow strength in this way and indicated that different conclusions obtained to those given by application of the LQAS rule. This paper cannot definitively identify the correct approach – careful assessment of the assumptions is needed.

One key feature of the data which has not been adequately modelled is stratification. That there is an issue with stratification has been clearly shown, and that it has potential to alter conclusions has been indicated. This remains a point for further development. What has been clearly demonstrated, from a published application of the LQAS method is overdispersion in
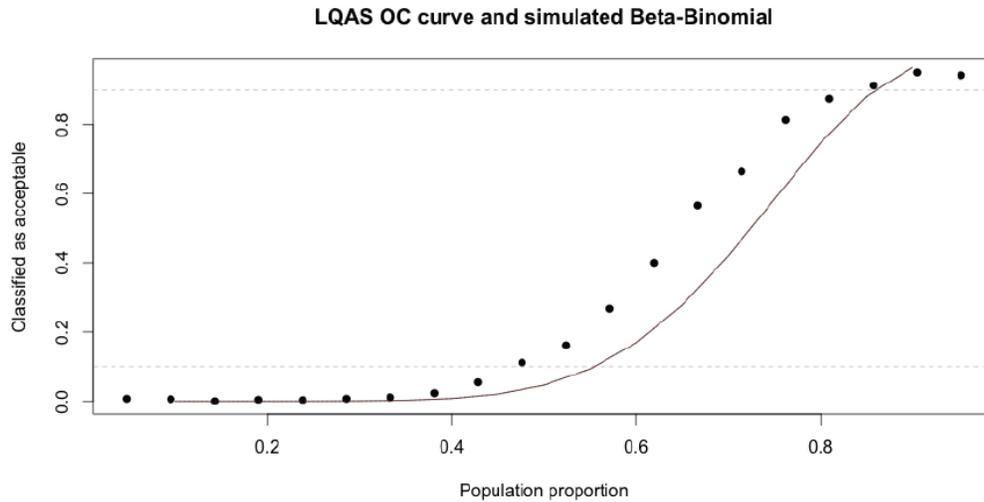
Figure 6: LQAS decision rule based on minimising error for a sample of size n=19 and d-13. The curve denotes estimates from Hypergeometric and Binomial (which are identical). The dots represent error rates from simulations of Beta-Binomial random variables having p given on the x axis and standard deviation of 2.9.

the data. This has actually been a recognised problem in applications of the method since the 1960s [16]. There are indeed proposals to replace the Hypergeometric distribution with a convolution of Binomials to allow for errors in the sample characterization [16]. This work was originally motivated by accounting problems where allowance had to be made for the possibility that a sampled item might have an error but an auditor might fail to notice it. Existence of overdispersion for analogous reasons in a healthcare setting seem highly plausible. The problem of overdisperion for the LQAS rule is perhaps not so obvious and is illustrated here from a small Monte Carlo study. Data were simulated from a Beta-Binomial distribution having standard deviation comparable to that seen in the Benin HTLV-1 data. Using a sample size of *n=19* and a *d=9* yields the OC curves shown in Figure **6**. The solid line shows an OC curve assuming a Hypergeometric distribution. The dots show the proportion of simulated samples accepted for different values of *p*. It can be seen for example that the error rate is much higher for data simulated from a Beta-Binomial than would be obtained were the data from a Hypergeometic distribution.

The OC curve given in Figure **6** shows that if data are not following a Hypergeometric distribution and are indeed overdispersed relative to that model then the error rate will be larger than that suggested by the standard OC curves. Conversely, fitting hierarchical models to the data (Bayesian or otherwise) allows a careful assessment of the overdispersion and should be less sensitive to the variance assumptions made by the LQAS method. This is incorporated naturally in an appropriate model and is seen as a widening of the posterior distribution for the parameters of interest.

CONCLUSION

This paper demonstrates that substantively different interpretations are obtained when constrasting Bayesian hierarchical modeling with the use of LQAS decision rules. This assertion is based on data taken from two published applications. Data analysis alone cannot determine which method (if any) is "correct" however it is suggested that the findings merit careful attention. The assumptions in the hierarchical models are subject to challenge and in presenting results from two different models it can be seen that interpretations are sensitive to model choice. What seems clear is that the data show overdispersion and a simple Monte Carlo simulation challenges whether the LQAS rules can work the way the OC curves suggest with real data. It is also clear that stratification may be important in prevalence estimates. It was not possible to incorporate this information in the model as the data were not published in a suitable format. However, the LQAS models assume that each observation making up the sample are *iid* and this seems like a problematic assumption. It is possible that prevalence being stratified by population subgroup contributes to overdispersion but again some simple Monte Carlo simulations suggest there is more to the overdispersion that missing covariates for age / gender. It is therefore possible to argue that a very standard industrial technique (LQAS) can be stretched when applied to healthcare systems where assumptions of well-controlled, homogenous industrial processes are no longer tenable.

The final point to make is that the Bayesian models produce a posterior distribution for the parameters of interest that can be directly interpreted by delivery professionals as a supplier/client risk. Making accept / reject decisions on healthcare systems has non-trivial implications. In the case of global acute malnutrition it has been shown that over-intervention causes problems for local famers and under-intervention clearly leaves people with malnutrition [17]. There can be no "always correct" decision in such circumstances and publication of the posterior risks affords the possibility of learning over time where the optimal intervention decision lies. Perhaps in more typical applications of LQAS the consequences are less severe. For example, perhaps areas requiring follow up after a vaccination are identified using LQAS [18]. Perhaps the risk that areas needing following up are identified less often than they should be has less severe consequences than the malnutrition example.

It is acknowledged that these are both dated examples and that LQAS methodology and training have developed. However, these are both published examples of the use of LQAS. LQAS is often used to assess trachoma [19]. Whilst this may well be a more common application than HTLV-1 prevalence, the data from Benin are rare in being available for re-analysis. Greater access to such datasets would allow other researchers to carry out more detailed modeling on more typical LQAS applications.

REFERENCES

[1] Robertson SE, Valadez JJ. Global review of healthcare surveys using Lot Quality Assurance Sampling (LQAS), 1984-2004. Soc Sci Med 2006; 63: 1648-1660.
http://dx.doi.org/10.1016/j.socscimed.2006.04.011

[2] Dodge HF, Romig H A method of sampling inspection. Bell System Technical Journal 1929; 8: 613-631.
http://dx.doi.org/10.1002/j.1538-7305.1929.tb01240.x

[3] Neyman J. Pearson: On the problem of the most efficient tests of statistical hypotheses Philos Trans R Soc Lond A 1933; 231: 289-337.
http://dx.doi.org/10.1098/rsta.1933.0009

[4] Houinato D, Preux PM, Charriere B, Massit B, Avode G, Denis F, Dumas M, Boutros-Toni F, Salamon R. Interest of LQAS method in a survey of HTLV-1 infection in Benin (West Africa). J Clin Epidemiol 2002; 55: 192-196.
http://dx.doi.org/10.1016/S0895-4356(01)00463-2

[5] Deitchler M, Valadez JJ, Egge K, Fernandez S, Hennigan M. A field test of three LQAS designs to assess the prevalence of acute malnutrition. Int J Epidemiol 2007; 36: 858-864.
http://dx.doi.org/10.1093/ije/dym092

[6] Garner P, Smith GD. Information for decision making: lot quality assurance sampling in the spotlight. Int J Epidemiol 2010; 39: 5-6.
http://dx.doi.org/10.1093/ije/dyq006

[7] Olives C, Pagano M. Bayes-LQAS: classifying the prevalence of global acute malnutrition. Emerg Themes Epidemiol 2010; 7: 3.
http://dx.doi.org/10.1186/1742-7622-7-3

[8] Hinde J, Demetrio CG. Overdispersion: models and estimation. Comput Stat Data Anal 1998; 27: 151-170.
http://dx.doi.org/10.1016/S0167-9473(98)00007-3

[9] Lanata CF, Stroh G, Black RE, Gonzales H. An evaluation of Lot Quality Assurance Sampling to monitor and improve immunization coverage. Int J Epidemiol 1990; 19: 1086-1090. [http://ije.oxfordjournals.org/content/19/4/1086.short. 00043]

[10] Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian Data Analysis. Boca Raton: Chapman and Hall / CRC 2nd edition 2003.

[11] Patil A, Huard D, Fonnesbeck CJ. PyMC: Bayesian Stochastic Modeling in Python. J Stat Softw 2010; 35: 1-81.

[12] Cowles MK, Carlin BP. Markov chain Monte Carlo convergence diagnostics: a comparative review. J Am Stat Assoc 1996; 91: 883-904.
http://dx.doi.org/10.1080/01621459.1996.10476956

[13] Rhoda DA, Fernandez SA, Fitch DJ, Lemeshow S. LQAS: user beware. Int J Epidemiol 2010; 39: 60-69.
http://dx.doi.org/10.1093/ije/dyn366

[14] Plummer M. Penalized loss functions for Bayesian model comparison. Biostatistics 2008; 9: 523-539.
http://dx.doi.org/10.1093/biostatistics/kxm049

[15] 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. Int J Epidemiol 2013. [http://ije.oxfordjournals.org/content/ early/2013/02/01/ije.dys230.short]

[16] Johnson NL, Kotz S, Wu XZ. Inspection errors for attributes in quality control. 1991; Volume 44 CRC Press.

[17] Bilukha O, Blanton C. Interpreting results of cluster surveys in emergency settings: is the LQAS test the best option Emerg Themes Epidemiol 2008; 5: 25.
http://dx.doi.org/10.1186/1742-7622-5-25

[18] Pezzoli L, Conteh K, Kamara W, Gacic-Dobo M, Ronveaux O, Perea WA, Lewis RF. 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.
http://dx.doi.org/10.1186/1471-2458-12-415

[19] Myatt M, Limburg H, Minassian D, Katyole D. Field trial for the applicability of Lot Quality Assurance Sampling survey method for rapid assessment of prevalence of active trachoma. Bulletin of the World Health Organisation 2003; 81: 877-885.