Title
Optimal design of multiple-objective Lot Quality Assurance Sampling (LQAS) plans

Authors
Duarte, Belmiro P.M.
Wong, Weng Kee

Publication Date
2018-11-14

DOI
10.1111/biom.13008

Peer reviewed

BIOMETRIC METHODOLOGY

Optimal design of multiple-objective Lot Quality Assurance Sampling (LQAS) plans†

Belmiro P.M. Duarte[1;2]

[1]Centro de Investigação em Processos Químicos e Produtos da Floresta, Department of Chemical Engineering, University of Coimbra, Pólo II, R. Sílvio Lima, 3030-790 Coimbra, Portugal

[2]Department of Chemical and Biological Engineering, ISEC, Polytechnic Institute of Coimbra, R. Pedro Nunes, 3030-199 Coimbra, Portugal.

email: bduarte@isec.pt

and

Weng Kee Wong

Department of Biostatistics, Fielding School of Public Health, UCLA, 10833 Le Conte Ave., Los Angeles, California 90095-1772, U.S.A.

email: wkwong@ucla.edu

†This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: [10.1111/biom.13008]

Additional Supporting Information may be found in the online version of this article.

Received 22 August 2017; Revised 12 November 2018; Accepted 14 November 2018
Biometrics

DOI 10.1111/biom.13008

Summary: Lot Quality Assurance Sampling (LQAS) plans are widely used for health monitoring purposes. We propose a systematic approach to design multiple-objective LQAS plans that meet user-specified type 1 and 2 error rates and targets for selected diagnostic accuracy metrics. These metrics may include sensitivity, specificity, positive predictive value, and negative predictive value in high or low anticipated prevalence rate populations. We use Mixed Integer Nonlinear Programming (MINLP) tools to implement our design methodology. Our approach is flexible in that it can directly generate classic LQAS plans that control error rates only and find optimal LQAS plans that meet multiple objectives in terms of diagnostic metrics. We give examples, compare results with the classic LQAS and provide an application using a malaria outcome indicator survey in Mozambique.

Key words: Diagnostic measures; Disease monitoring programs; Lot Quality Assurance Sampling; Mixed Integer Nonlinear Programming; Optimal design; Public Health policy.

1. Introduction

Lot Quality Assurance Sampling (LQAS) plans were developed in the 1950's using statistical tools developed around 1920-30 to control the quality at a production line. At its core, a LQAS plan is a decision making process whether to accept a lot or not based on the quality of the selected sample. The methodology quickly found wide-spread applications in health care surveys. For example, public health authorities want to monitor whether people in a region are benefiting sufficiently from a health care community program or people in a village or district have been exposed to an infectious agent. The sampling plans vary in sophistication in design and execution. Robertson and Valadez (2006) reviews health care surveys using LQAS and Lemeshow and Taber (1991) reviews sampling methods and compares merits of having a single or double-sampling plan.

LQAS is commonly used in both public health research and global health research to determine if a health policy or a community program is working for the intended purpose. Some specific applications of LQAS are monitoring immunization programmes to ascertain their cost effectiveness (Sandiford, 1993), monitoring elimination leprosy programs in a region (Gupte et al., 2004) and examining effectiveness of community intervention programs of capital and management systems on maternal and child health behavior change (Valadez et al., 2005). Vanamail et al. (2006) discussed operation feasibility and implementation of LQAS as a tool for routine monitoring in filariasis control programs. Deitchler et al. (2007), Olives et al. (2009), Olives and Pagano (2010) used LQAS to assess the prevalence of acute malnutrition and Biedron et al. (2010) applied LQAS plans to monitor malaria outcome indicators programs in Mozambique. Applications of LQAS continue to this day with Hund (2014) noting that they are increasingly popular in global health care applications. For example, Olives et al. (2012) applied ideas to incorporate ordinal outcomes with more than two categories (i.e. not just exposed or not to a disease) with application to Schistosomiasis

control and Brown et al. (2013) reported a LQAS plan to monitor quality of supplemental immunization activities as a tool for polio eradication in Nigeria. Interestingly, in 2013, the Medina Gates Foundation funded an innovative study that used LQAS-based methodology to estimate rates of adoption of agricultural technologies in Ethiopia.

The key statistical question in designing a LQAS plan is choosing the sample size that allows estimate the characteristics of the whole population and the threshold value for the acceptance so that pre-specified type 1 and type 2 error rates are attained. Practically, the LQAS sets the sample size required from the population and the required number of subjects testing positive in the sample to declare the population is exposed (or not) to an infectious disease or is in compliance with certain health care guidelines.

The design of LQAS plans for health monitoring is similar to Acceptance Sampling plans by variables or attributes for quality control purposes. Both can be formulated as optimization problems and frequently the sought solution is the smallest sample size to control cost, see for example, Duarte and Saraiva (2008, 2013). Alternatively, one may seek to minimize the Average Sampling Number (ASN) that meets user-specified constraints at the controlled points of the *operating characteristic* (OC) curve. We recall the OC curve describes the probability of accepting a lot as a function of the quality of the lot, measured by the proportion of nonconforming items, and the curve increases as the proportion of conforming cases increases.

Some common test diagnostic performance metrics in biomedical studies are (i) *Specificity*; (ii) *Sensitivity*; (iii) *Negative Predicted Value* (NPV); and (iv) *Positive Predictive Value* (PPV). Interestingly, such measures have not been directly incorporated at the design stage of LQAS plans. Our aim in this paper is to provide a systematic approach to construct LQAS plans that meet both error rates and accuracy requirements in the diagnostic performance metrics. This differs from the classic LQAS plans where they are designed to guarantee to

meet type 1 and 2 error rates only and the various diagnostic metrics are then computed a posteriori, see for example, Olives and Pagano (2013); Hund (2014) among many others. Motivation for this work comes partly from the literature that seems to suggest that PPV and specificity tend to under-perform compared to other measures for LQAS; see for example, Olives and Pagano (2013). This prompted us to design a LQAS that meets selected diagnostic accuracy metrics requirements at the onset and forms the basis of our work here.

Section 2 provides background for our approach and Section 3 introduces the Mixed Integer Nonlinear Linear Programming (MINLP) formulation for designing LQAS plans that satisfy the OC-curve constraints and diagnostic metrics criteria. Section 4 presents our LQAS plans and discusses the case when all the target requirements are too stringent and not all objectives can be attained simultaneously. This issue arises because the objectives are frequently competitive and much of one may have to be given up for a small gain in another. Under such circumstance, the user needs to arrive at a compromised LQAS plan by relaxing some the metrics requirements or allow for higher values in one or more of the error rates. We also construct and compare our LQAS plan with a classic plan implemented to monitor malaria control in Mozambique. Section 5 concludes with a summary.

2. Background

This Section reviews LQAS plans, diagnostic metrics and how to formulate and solve the design problem via MINLP. Subsections 2.1 and 2.2 discuss LQAS plans and diagnostic performance metrics, respectively. In Subsection 2.3, we briefly review the fundamentals of MINLP and how they can be applied to solve two types of LQAS design problems. In Section 6 we provide information about the computer codes for the user to input pre-defined objectives and find the multiple-objective optimal LQAS, and when necessary, vary the requirements to arrive at a compromised LQAS plan.

2.1 *LQAS plans*

Throughout, we assume that (i) the probability of selecting a nonconforming or conforming individual is independent of the sampling method; and (ii) the population that we sample from is large enough that it is not impacted by the sample drawn.

We represent a sampling plan by $\mathcal{S}(n, r)$, where n is the sample size and r is the acceptance threshold to declare whether a lot/population is acceptable or not. If there are more than r individuals in the sample with the sought characteristic after testing, the population is deem to be acceptable; otherwise, it is unacceptable. Depending on the study, the sought characteristic may be whether the individual is tested negative for an infectious disease or is in compliance with certain health care guidelines. If there are r or more such subjects in the sample, the population is deemed disease free or is in compliance with the guidelines; otherwise, the population is declared unacceptable.

In practice, there are risks to misclassifying the population and we control them using user-specified error rates α and β. For example, we pre-specify that the probability of incorrectly classifying a population with high proportion of individuals with the sought characteristic, say p_U, as unacceptable should be smaller than α (type 1 error). Similarly, the probability of incorrectly classifying a population with small proportion of individuals having the sought characteristic, p_L, as acceptable should be smaller than β (type 2 error). In monitoring studies p_L is the unacceptable coverage level, and p_U is the acceptable coverage level. If x is the number of conforming individuals in the sample and $P(x \geqslant r|p)$ is the probability of accepting the population when the proportion of individuals of the population having the sought characteristic is p, we want

$$P(x \geqslant r|p_L) \leqslant \beta \tag{1a}$$

$$\text{and} \quad P(x \geqslant r|p_U) \geqslant 1 - \alpha. \tag{1b}$$

To fix ideas, we assume from now on our research concerns public health and the sought

characteristic is binary so that it models whether an individual is disease free or not. We use the binomial model to study the probability of obtaining a fixed number x of disease free individuals in a sample of size n from a population. If p is the unknown proportion of disease free individuals in the population, we denote such a probability by

$$\mathbb{P}(x) = \binom{n}{x} p^x \ (1-p)^{n-x}, \quad x = 0, \cdots, n. \tag{2}$$

The OC curve represents the probability of acceptance of the population as a function of p. For the binomial distribution (2), the OC curve is given by

$$F(p|n,r) = \sum_{x=r}^{n} \mathbb{P}(x) = \sum_{x=r}^{n} \binom{n}{x} p^x \ (1-p)^{n-x} = I(p, n-r+1, r), \tag{3}$$

where $I(p, n-r+1, r) = \int_0^p z^{n-r} \ (1-z)^{r-1} \mathrm{d}z / \int_0^1 z^{n-r} \ (1-z)^{r-1} \mathrm{d}z$ is the cumulative regularized Beta function, see DiDonato and Jarnagin (1967) and Press et al. (1996, Chap. 6) for details on the numerical computation.

An early strategy used to design LQAS plans is based on the OC curve. The sample size and the acceptance constant are determined so that the user-stipulated risk constraints $F(p_L|n,r) \leqslant \beta$ and $F(p_U|n,r) \geqslant 1-\alpha$ are met for pre-defined levels p_L, p_U, α and β. The interval $p_L \leqslant p \leqslant p_U$ represents a *grey region* of indecision, where consequences of a misclassification error are expected to have a lower impact. However, there are recognized risks incompletely accounted for in the classic LQAS plans, see for example, Olives and Pagano (2010, 2013), and it is helpful to have diagnostic measures to further evaluate the usefulness of a LQAS plan.

2.2 *Diagnostic performance metrics*

Results from the LQAS can be summarized in a 2×2 table with columns and rows representing disease status and test results from the LQAS plan. The 2×3 Table 1, a version of the 2×2 table including the grey region, displays the definitions of the expected risks in our

plan, where entries in the table are probabilities of accepting (or not) the populations based on the sample results.

[Table 1 about here.]

Our work considers the following diagnostic metrics expressed in terms of the entries displayed in Table 2 and broken down into two categories, LAPR and HAPR populations, discussed in Lemeshow and Taber (1991). Populations with a low proportion of disease free individuals, i.e. $p \leqslant p_L$, and are called low anticipated prevalence rate (LAPR) populations, and populations with a high proportion of disease free individuals, i.e. $p \geqslant p_U$, are called high anticipated prevalence rate (HAPR) populations. Given a uniform prior distribution on the prevalence rate p on $[0, 1]$, the following diagnostic metrics can be defined and expressed in terms of the joint probabilities in Table 2.

$$\text{Sens}(p \leqslant p_L) = P(\text{lot declared acceptable}|p \leqslant p_L) = P(x \geqslant r|p \leqslant p_L) = \frac{a}{p_L} \tag{4a}$$

$$\text{Sens}(p \geqslant p_U) = P(\text{lot declared unacceptable}|p \geqslant p_U) = P(x \leqslant r-1|p \geqslant p_U) = \frac{f}{1-p_U} \tag{4b}$$

$$\text{Spec}(p \leqslant p_L) = P(\text{lot declared unacceptable}|p \geqslant p_L) = P(x \leqslant r-1|p \geqslant p_L) = \frac{e+f}{1-p_L} \tag{4c}$$

$$\text{Spec}(p \geqslant p_U) = P(\text{lot declared acceptable}|p \leqslant p_U) = P(x \geqslant r|p \leqslant p_U) = \frac{a+b}{p_U} \tag{4d}$$

$$\text{PPV}(p \leqslant p_L) = P(p \leqslant p_L|\text{lot declared acceptable}) = P(p \leqslant p_L|x \geqslant r) = \frac{a}{a+b+c} \tag{4e}$$

$$\text{PPV}(p \geqslant p_U) = P(p \geqslant p_U|\text{lot declared unacceptable}) = P(p \geqslant p_U|x \leqslant r-1) = \frac{f}{d+e+f} \tag{4f}$$

$$\text{NPV}(p \leqslant p_L) = P(p \geqslant p_L|\text{lot declared unacceptable}) = P(p \geqslant p_L|x \leqslant r-1) = \frac{e+f}{d+e+f} \tag{4g}$$

$$\text{NPV}(p \geqslant p_U) = P(p \leqslant p_U|\text{lot declared acceptable}) = P(p \leqslant p_U|x \geqslant r) = \frac{a+b}{a+b+c}, \tag{4h}$$

where (4a), (4c), (4e) and (4g) are for LAPR populations and (4b), (4d), (4f) and (4h) are for HAPR populations, respectively.

[Table 2 about here.]

The metrics in (4) are conditional in nature and its expectation can be constructed using Bayes theorem, where $\pi(p)$ is a user-selected prior distribution for the proportion of disease free individuals in the population. It is convenient to work with uniform prior distributions for p, i.e. $\pi(p) = \mathbb{U}[0,1]$. This prior distribution is appropriate when the decision maker has no or little prior knowledge of p. Another reason for the choice of a uniform prior distribution is that closed formulae become available for the entries in Table 2 and they are shown below. When the prior distribution is not a uniform distribution, the entries in (4) are computed numerically using Gaussian quadratures to evaluate the integrals in (5) below. The terms in (5) can be interpreted as selected areas under or above a OC curve. Figure 1 displays them using the OC curve for a specific set of choices of p_L, p_U and target requirements. The figure also shows geometrically that $a + d = p_L$, $b + e = p_U - p_L$, $c + f = 1 - p_U$ and $a + d + b + e + c + f = 1$.

$$a = \int_0^{p_L} \sum_{x=r}^{n} \binom{n}{x} p^x (1-p)^{n-x} \pi(p)\, \mathrm{d}p = \sum_{x=r}^{n} \sum_{k=0}^{n-x} (-1)^k \binom{n}{x} \binom{n-x}{k} \frac{p_L^{x+k+1}}{x+k+1} \tag{5a}$$

$$b = \int_{p_L}^{p_U} \sum_{x=r}^{n} \binom{n}{x} p^x (1-p)^{n-x} \pi(p)\, \mathrm{d}p = \sum_{x=r}^{n} \sum_{k=0}^{n-x} (-1)^k \binom{n}{x} \binom{n-x}{k} \frac{p_U^{x+k+1} - p_L^{x+k+1}}{x+k+1} \tag{5b}$$

$$c = \int_{p_U}^{1} \sum_{x=r}^{n} \binom{n}{x} p^x (1-p)^{n-x} \pi(p)\, \mathrm{d}p = \sum_{x=r}^{n} \sum_{k=0}^{n-x} (-1)^k \binom{n}{x} \binom{n-x}{k} \frac{1 - p_U^{x+k+1}}{x+k+1} \tag{5c}$$

$$d = \int_0^{p_L} \sum_{x=0}^{r-1} \binom{n}{x} p^x (1-p)^{n-x} \pi(p)\, \mathrm{d}p = \sum_{x=0}^{r-1} \sum_{k=0}^{n-x} (-1)^k \binom{n}{x} \binom{n-x}{k} \frac{p_L^{x+k+1}}{x+k+1} \tag{5d}$$

$$e = \int_{p_L}^{p_U} \sum_{x=0}^{r-1} \binom{n}{x} p^x (1-p)^{n-x} \pi(p) \, \mathrm{d}p = \sum_{x=0}^{r-1} \sum_{k=0}^{n-x} (-1)^k \binom{n}{x} \binom{n-x}{k} \frac{p_U^{x+k+1} - p_L^{x+k+1}}{x+k+1} \tag{5e}$$

$$f = \int_{p_U}^{1} \sum_{x=0}^{r-1} \binom{n}{x} p^x (1-p)^{n-x} \pi(p) \, \mathrm{d}p = \sum_{x=0}^{r-1} \sum_{k=0}^{n-x} (-1)^k \binom{n}{x} \binom{n-x}{k} \frac{1 - p_U^{x+k+1}}{x+k+1} \tag{5f}$$

2.3 *Mixed Integer Nonlinear Programming*

Mixed Integer Nonlinear Programming methods aim to optimize a nonlinear function with both continuous and discrete variables under user-specified constraints. Such optimization problems occur commonly in real applications, such as in chemical engineering, finance, and management and are typically solved by MINLP solvers. The website of NEOS Server Team (2018) gives an introduction to MINLP and has a link to various types of state-of-the art solvers for numerical optimization. The general form of a MINLP is

$$\min_{\boldsymbol{x},\boldsymbol{y}} \; f(\boldsymbol{x}, \boldsymbol{y}) \tag{6a}$$

$$\text{subject to } h_i(\boldsymbol{x}, \boldsymbol{y}) = 0, \quad \forall i \in \mathcal{E} \tag{6b}$$

$$g_i(\boldsymbol{x}, \boldsymbol{y}) \leqslant 0, \quad \forall i \in \mathcal{I} \tag{6c}$$

$$\boldsymbol{x} \in \mathbf{X}, \; \boldsymbol{y} \in \mathbf{Y}, \tag{6d}$$

where each function $h_i(\boldsymbol{x}, \boldsymbol{y})$ and $g_i(\boldsymbol{x}, \boldsymbol{y})$ is a map from $\mathbb{R}^{n_x} \times \mathbb{R}^{n_y}$ to $\mathbb{R}$ where n_x and n_y stand for the number of continuous and integer variables, respectively, $\mathcal{E}$ is the set of equality constraints, $\mathcal{I}$ the set of inequalities, $\mathbf{X} \in \mathbb{R}^{n_x}$ is a continuous compact domain, $\mathbf{Y} \in \mathbb{N}_0^{n_y}$ is a discrete domain containing integer values, $\boldsymbol{x}$ represents the continuous variables and $\boldsymbol{y}$ represents the non-negative integer variables. In our design context the variables a, b, c, d, e and f are treated as continuous variables, and both n and r are positive integers.

The fundamentals of MINLP are well described in the literature, see for example, Floudas (2002), who also provided illustrative examples. Typical algorithms in MINLPs are the outer

approximation method (Duran and Grossmann, 1986), the extended cutting plane method (Westerlund and Pettersson, 1995) and the branch and bound method (Fletcher and Leyffer, 1998). In our paper, we apply the branch and bound based algorithm using the solver `SBB` (GAMS Development Corporation, 2013b) from the General Modeling System `GAMS 24.2.1` (GAMS Development Corporation, 2013a). The solver `SBB` combines the standard branch and bound method from Mixed Integer Linear Programming and a standard NLP solver supported by `GAMS 24.2.1` to solve the relaxed optimization problems. It uses the solver `CONOPT` for solving the relaxed nonlinear programs (Drud, 1985) and `CPLEX` for solving the local integer linear programs (GAMS Development Corporation, 2013b). Because our optimization problem is not convex by definition, `SBB` only guarantees local optimal solutions. However, in our problems, there is usually a single optimum and this suggests the solution may be the global optimum. The algorithm requires the gradient and hessian information to solve the local NLP problems, and automatic differentiation techniques are used to internally compute them. In all our problems, we use absolute and relative tolerances of 1×10^{-5} for converging the solution.

All computations in this paper are carried out using an Intel Core i7 computer (Intel Corporation, Santa Clara, CA) running 64 bits Windows 10 operating system with 2.80 GHz.

3. Optimal LQAS formulations

We now present the MINLP formulations for designing LQAS plans. In Section 3.1 we address the problem of finding a LQAS plan that assures the conditions in (1) at the controlled points of the OC curve are satisfied, and in Subsection 3.2, we consider the problem of designing plans that incorporate a combination of diagnosis performance criteria where lower bound thresholds for the metrics are pre-specified. The former problem is designated as a *OC curve-constrained design problem* and the latter as a *performance criteria-constrained design*

problem. In both cases the main objective is to minimize the sample size and reduce cost. In §3.3 we compare the goals for using both schemes in practical applications.

3.1 *Formulation for OC curve-constrained LQAS plans*

The typical algorithms used to design *OC curve-constrained* LQAS plans are enumerative procedures where n and r are successively iterated until the constraints at the controlled points of the OC curve are met (Lemeshow and Taber, 1991). Here, given the risks α and β and the target proportions required for *LAPR* and *HAPR* populations, the goal is to use MINLP to determine the minimum sample size that meet conditions (1) with Equation (3) as the defining OC curve. The optimization problem is

$$\min_{n,r} n \tag{7a}$$

$$\text{subject to } I(p_U, n-r, r-1) \geqslant 1-\alpha \tag{7b}$$

$$I(p_L, n-r, r-1) \leqslant \beta \tag{7c}$$

$$n \geqslant 2 \tag{7d}$$

$$r \geqslant 1 \tag{7e}$$

$$n,\ r \in \mathbf{N}, \tag{7f}$$

where equations (7b-7c) are the constraints at the OC points, and (7d) and (7e) are natural lower bounds for n and r, respectively. The problem (7) allows finding designs that have specific (imposed) type 1 and type 2 errors at p_L and p_U, respectively, as primary constraints.

3.2 *Formulation for performance-constrained LQAS plans*

Suppose we have known targets for one or more of the diagnostic performance criteria listed in Subsection 2.2. Let us designate the Sensitivity target for the diseased subjects in the populations as σ_L for *LAPR* and the corresponding target for the diseased subjects in the population as σ_U for *HAPR*. Similarly, let the corresponding Specificity targets for the two groups be θ_L and θ_U, respectively; the corresponding PPV targets for the two groups be π_L

and π_U, respectively, and the corresponding NPV targets for the two groups be ϱ_L and ϱ_U, respectively. Using Table 2 and user-specified targets, our goal is to design a LQAS plan that satisfies one or more of the following performance constraints:

$$\text{Sensitivity for } LAPR \qquad a \geqslant \sigma_L p_L \tag{8a}$$

$$\text{Sensitivity for } HAPR \qquad f \geqslant \sigma_U(1-p_U) \tag{8b}$$

$$\text{Specificity for } LAPR \qquad e+f \geqslant \theta_L(1-p_L) \tag{8c}$$

$$\text{Specificity for } HAPR \qquad a+b \geqslant \theta_U p_U \tag{8d}$$

$$\text{PPV for } LAPR \qquad a \geqslant \pi_L(a+b+c) \tag{8e}$$

$$\text{PPV for } HAPR \qquad f \geqslant \pi_U(d+e+f) \tag{8f}$$

$$\text{NPV for } LAPR \qquad e+f \geqslant \varrho_L(d+e+f) \tag{8g}$$

$$\text{NPV for } HAPR \qquad a+b \geqslant \varrho_U(a+b+c). \tag{8h}$$

Clearly, the above formulations comes from Table 2 and the MINLP formulation for the *performance criteria-constrained design problem* is:

$$\min_{n,r} n \tag{9a}$$

$$\text{subject to Equations (5)} \tag{9b}$$

$$\text{Combination of equations (8)} \tag{9c}$$

$$n \geqslant 2 \tag{9d}$$

$$r \geqslant 1 \tag{9e}$$

$$n,\ r \in \mathbf{N} \tag{9f}$$

$$a,\ b,\ c,\ d,\ e,\ f \in [0,1]. \tag{9g}$$

This formulation is quite general as it can also incorporate in the plan the constraints in §3.1 to control the points at the OC curve and the diagnostic performance requirements in (8). The complete problem formulation to include the user-specified type 1 and 2 error rates

is shown below and its solution can be found similarly using MINLP.

$$\min_{n,r} n \tag{10a}$$

$$\text{subject to Equations (5)} \tag{10b}$$

$$\text{Combination of equations (8)} \tag{10c}$$

$$\text{Equations (7b) and (7c)} \tag{10d}$$

$$n \geqslant 2 \tag{10e}$$

$$r \geqslant 1 \tag{10f}$$

$$n,\ r \in \mathbf{N} \tag{10g}$$

$$a,\ b,\ c,\ d,\ e,\ f \in [0,1]. \tag{10h}$$

We note that this more inclusive problem may not result in a feasible LQAS plan because the demands may be too stringent and no plan can satisfy all the requirements simultaneously.

3.3 *Comparison of plans*

The design of classic LQAS plans, addressed in §3.1, find designs that only meet preset type 1 and type 2 error rates. Our proposed LQAS plans in §3.2 provide flexibility by allowing the user to incorporate prior information on the unknown proportion of disease free individuals in the population and construct LQAS plans that meet user-selected diagnostic metrics requirements. We note that some of these plans in §3.2 reduce the expected value of the risk of wrong decision when $p \in [0, p_L]$ and increase it when $p \in [p_U, 1]$. Consequently, such plans may require smaller sample size, and the reduction may be substantial. LQAS plans that emphasize only on diagnostic accuracy metrics are likely less discriminant than the classic LQAS plans because they do not control the α and β risks directly. Our proposed method can also control these risks directly but there may be no solution because the set of requirements becomes too competitive. When this happens, we recommend a compromise

among the competing goals in the study by lowering the requirements in some of the metrics or error rates.

4. Results

We now demonstrate our approach to find LQAS plans that meet various diagnostic accuracy metrics for the *LAPR* and *HAPR* populations. Other design problems with a combination of criteria can be similarly formulated and solved. We also show a case when there is no feasible solution to the problem. Specifically, we solve some *OC curve-constrained design problems* and compare them with those obtained for the *performance criteria-constrained design problem* using different combinations of performance criteria.

To solve problem (7) and arrive at an optimally designed LQAS plan, we consider a scenario where $\alpha = \beta = 0.10$ and both p_L and p_U vary in a user-selected region. To fix ideas, we fix the difference between p_U and p_L but let their individual values vary. From Table 3, we observe that as the proportions for *LAPR* and *HAPR* populations increase, the ratio r/n becomes larger and the plan becomes more discriminant corresponding to steeper OC curves. Our computational results validate what is often used in the field, which is a design with $n = 19$, and $r = 13$ (Valadez, 1991). All our examples require less than 1.0 second of CPU time to generate the plans suggesting that our proposed algorithm is quite efficient.

[Table 3 about here.]

Table 4 presents LQAS plans for several combinations of diagnostic performance criteria. The same trend observed for the OC curve constrained plans applies here; as the proportions assumed for *LAPR* and *HAPR* increase, the ratio r/n increases as well as the discrimination ability of the plans. The *performance criteria-constrained* LQAS plans have smaller sample sizes and acceptance constants but larger type 1 and 2 error rates than those imposed to *OC curve-constrained* LQAS plans. This is expected as our plans focus on the diagnostic

metrics. Mathematically, by changing from $p_L = 0.4$ fixed to $p_L \sim \mathbb{U}[0, 0.4]$ we have effectively increased the risk of wrong decision for LAPR population, i.e. accept the population. Similarly, by changing from $p_U = 0.8$ fixed to $p_U \sim \mathbb{U}[0.8, 1.0]$ the risk of unacceptance of HAPR population increases, thereby making the situation more "extreme", which means that we can detect differences with smaller sample sizes.

To analyze the optimality of the solutions of the *performance criteria-constrained design problem*, we revisit the LQAS plan for scenario S5 when both the criteria Sensitivity for *LAPR* and for *HAPR* populations are to be met (line 5 of Table 4) that requires a plan $\mathcal{S}(6, 4)$. The corresponding OC curve in Figure 1 depicts the regions of risk a, b, c, d, e and f. Here, it is used to relate the terms (5) to an OC curve and check the satisfaction of the diagnosis accuracy metrics imposed. Table 5 lists their values obtained by equations (5), along with the Specificity for both groups, and shows that the constraints ($a/p_L \geqslant \sigma_L$ and $f/(1 - p_U) \geqslant \sigma_U$) are both satisfied.

[Table 4 about here.]

[Figure 1 about here.]

[Table 5 about here.]

We note that, depending on the constraints imposed on the optimization problem, our optimization problems may or may not have a feasible solution. This is attributable to the antagonistic nature of the constraints. When there is no solution for the LQAS plan, we either relax the various metric requirements or convert the metric inequalities (8) into a set of equalities and solve a square system of 6 algebraic equations with respect to a, b, c, d, e and f using the solver `CNS` in `GAMS 24.2.1`. The abbreviation `CNS` stands for constrained nonlinear systems and uses the nonlinear programming solver `CONOPT`, which initially checks the feasibility of the problem; some details are available GAMS Development Corporation

(2013b). Below is an example of a program that we have implemented to check feasibility of a LQAS plan. We consider the setup where the design criteria are the Specificity for *LAPR* and *HAPR* populations (second line in Table 4) and the threshold level is 0.95 for both criteria, i.e. $\theta_U = \theta_L = 0.95$, $p_L = 0.6$ and $p_U = 0.9$. The formulation of the problem is

$$\text{find } a,\ b,\ c,\ d,\ e,\ f \tag{11a}$$

$$\text{subject to } a + b = \theta_L(a + b + d + e) \tag{11b}$$

$$e + f = \theta_U(b + c + e + f) \tag{11c}$$

$$a + d = p_L \tag{11d}$$

$$b + e = p_U - p_L \tag{11e}$$

$$c + f = 1 - p_U \tag{11f}$$

$$a + b + c + d + e + f = 1 \tag{11g}$$

$$a,\ b,\ c,\ d,\ e,\ f \in [0, 1]. \tag{11h}$$

Running our code shows the problem (11) does not have feasible solution, and consequently, a LQAS plan that satisfies all the constraints cannot be found. It is straightforward to verify that the maximum value of θ_U and θ_L that produce a feasible solution is to have $\theta_L = \theta_U$ is 0.7692. If we allow $\theta_U \neq \theta_L$ and $\theta_U = 0.95$, it can be shown that the maximum value of θ_L that produces a feasible plan is 0.6888. Clearly, these findings are problem-specific and depend on the values assumed for p_L and p_U.

4.1 *LQAS plan for monitoring the Malaria Eradication Program in Mozambique*

We applied our algorithms in §3.1 and §3.2 to design new LQAS plans applied for monitoring the malaria eradication program in Mozambique (Biedron et al., 2010). For comparison purposes, our setup is that $\alpha + \beta < 0.2$, $p_U = 0.7$ and $p_L = 0.4$, equal to that used in the application. Next, we compute the diagnostic metrics of the LQAS plans used when applied

to LAPR and HAPR populations, see Table 6. Some of the figures are poor and *performance criteria-constrained* LQAS plans can be designed to specifically increase them.

[Table 6 about here.]

We next apply our methodology to construct multiple-objective optimal LQAS plans that meet specific targets in the diagnosis metrics. First, we compute the optimal classic plan using (7) with $\alpha = 0.03$ and $\beta = 0.1$ ($\alpha + \beta = 0.13$), and obtained $\mathcal{S}(29, 15)$, i.e. the population is accepted when more than 15 individuals out of 29 is found to be disease free.

Table 7 displays optimal LQAS plans for various scenarios of *performance criteria-constrained designs*. First, in scenario S1' we consider constraints for Sensitivity in both populations, then constraints for Specificity (S2'), constraints for PPV (S3') and constraints for NPV (S4') so they can overcome the inefficiencies of the plans used. Recall that diagnosis metrics constrained plans should be used when one or various diagnosis metrics are the focus. The plans designed to assure 75 % of Specificity (i.e. $\theta_U = \theta_L$=0.75) and 65 % of PPV ($\pi_U = \pi_L$=0.65) for both populations are those displayed for S2' and S3', and we observe better performances than those offered by classic plans. However, the plans obtained for S2' and S3' have sums of risk ($\alpha + \beta$) above the limit set for implementation.

[Table 7 about here.]

Finally, we use our formulation (10) to construct a LQAS plan where the risks at p_L and p_U are constrained and a given combination of diagnosis metrics is simultaneously meet. We also consider the case when $p_L = 0.4, p_U = 0.7, \alpha = 0.03$ and $\beta = 0.1$ in order to compare the resulting LQAS plan with the one obtained from formulation (7). Table 8 shows the plans, where the first column lists the requirements imposed on the diagnostic metrics in addition to the constraints at the points $(p_U, 1-\alpha)$ and (p_L, β). The results show that all the plans are at least as discriminant as the one that only uses the constraints at the OC curve controlled

points (scenario S6'), and the tool is successful in finding designs where different kinds of constraints are imposed.

[Table 8 about here.]

5. Conclusions

We are the first to use MINLP to find multiple-objective LQAS plans for monitoring health care programmes. Our general goal is to minimize the sample size subject to a set of user-specified requirement targets for a combination of diagnostic performance criteria and type 1 and 2 error rates. Our method is flexible in that (i) it allows designing classic LQAS plans where only constraints on the type I and II error rates are considered; (ii) designing *performance criteria-constrained* LQAS plan that are required to meet a specific combination of diagnosis metrics. Besides the metrics considered, one can also include targets for false omission rate, positive likelihood ratio, diagnostic odds ratio; and (iii) designing classic plans that must also allow satisfy targets of diagnosis metrics.

We apply our method to a range of setups with different requirements on the various combinations of diagnostic accuracy criteria, and compare results. We considered a real application of LQAS to test our tool and obtained multiple-objective optimal LQAS plans ranging from classic to diagnostic constrained plans. To deal with scenarios where the constraints are too stringent, we propose a linear programming tool for checking the feasibility of the LQAS design problem for that combination of anticipated prevalence proportions and threshold bounds.

6. Supplementary Materials

The code referenced in Sections 3 and 4 is available at the Biometrics website on Wiley Online Library.

Acknowledgments

Wong was partially supported by a grant from the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R01GM107639. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

Biedron, C., Pagano, M., Hedt, B., Kilian, A., Ratcliffe, A., Mabunda, S., and Valadez, J. (2010). An assessment of lot quality assurance sampling to evaluate malaria outcome indicators: extending malaria indicator surveys. *International Journal of Epidemiology* **39,** 72–79.

Brown, A. E., Okayasu, O., Nziok, M. M., Wadood, M. Z., Chabot-Couture, G., Quddus, A., Walker, G., and Sutter, R. W. (2013). Lot quality assurance sampling to monitor suppemental immunization activity quality: an essential tool for improving performance in polio endemic countries. *The Journal of Infectious Diseases* **210,** S333–340.

Deitchler, M., Valadez, J., Egge, K., Fernandez, S., and Hennigan, M. (2007). A field test of three LQAS designs to assess the prevalence of acute malnutrition. *International Journal of Epidemiology* **36,** 858–864.

DiDonato, A. R. and Jarnagin, M. P. (1967). The efficient calculation of the incomplete beta function ratio for half-integer values of the parameters a, b. *Math. Comp.* **21,** 652 – 662.

Drud, A. (1985). `CONOPT`: A GRG code for large sparse dynamic nonlinear optimization problems. *Mathematical Programming* **31,** 153–191.

Duarte, B. and Saraiva, P. (2008). An optimization based approach for designing attribute acceptance sampling plans. *International Journal of Quality and Reliability Management* **25,** 824–841.

Duarte, B. and Saraiva, P. (2013). Optimal design of acceptance sampling plans by variables for nonconforming proportions when the standard deviation is unknown. *Communications in Statistics - Simulation and Computation* **42,** 1318–1342.

Duran, M. A. and Grossmann, I. E. (1986). An outer-approximation algorithm for a class of mixed-integer nonlinear programs. *Mathematical Programming* **36,** 307–339.

Fletcher, R. and Leyffer, S. (1998). Numerical experience with lower bounds for MIQP branch-and-bound. *SIAM Journal on Optimization* **8,** 604–616.

Floudas, C. (2002). Mixed-Integer Nonlinear Optimization. In Pardalos, P. and Resende, M., editors, *Handbook of Applied Optimization*, pages 451–475. Oxford University Press, Oxford.

GAMS Development Corporation (2013a). *GAMS - A User's Guide, GAMS Release 24.2.1.* GAMS Development Corporation, Washington, DC, USA.

GAMS Development Corporation (2013b). *GAMS - The Solver Manuals, GAMS Release 24.2.1.* GAMS Development Corporation, Washington, DC, USA.

Gupte, M., Murthy, B., Mahood, K., Meeralakshmi, B., Nagaraju, B., and Prabhakaran, R. (2004). Application of lot quality assurance sampling for leprosy elimination monitoring–examination of some critical factors. *International Journal of Epidemiology* **33,** 344–348.

Hund, L. (2014). New tools for evaluating LQAS survey designs. *Emerging Themes in Epidemiology* **11,** 2+.

Lemeshow, S. and Taber, S. (1991). Lot quality assurance sampling plans: single- and double-sampling plans. *World Health Stat. Quart.* **44,** 115–132.

NEOS Server Team (2018). *Mixed Integer Nonlinear Programming.* University of Wisconsin - Madison, Madison, USA. Available at https://neos-guide.org/content/mixed-integer-nonlinear-programming.

Olives, C. and Pagano, M. (2010). Bayes-LQAS: classifying the prevalence of global acute

malnutrition. *Emerging Themes in Epidemiology* **7,** 3+.

Olives, C. and Pagano, M. (2013). Choosing a design to fit the situation: how to improve specificity and positive predictive value using bayesian lot quality assurance sampling. *International Journal of Epidemiology* **42,** 346–355.

Olives, C., Pagano, M., Deitchler, M., Hedt, B., Egge, K., and Valadez, J. (2009). Cluster designs to assess the prevalence of acute malnutrition by lot quality assurance sampling: a validation study by computer simulation. *Journal Royal Stat. Soc.: Ser A* **172,** 495–510.

Olives, C., Valadez, J., Brooker, S., and Pagano, M. (2012). Multiple category-lot quality assurance sampling: A new classification system with application to schistosomiasis control. *PLoS Neglected Tropical Diseases* **6,** 1–8.

Press, W., Teukolsky, S., Vetterling, W., Flannery, B., and Metcalf, M. (1996). *Numerical Recipes in Fortran 90: Volume 2.* Cambridge University Press, Cambridge.

Robertson, S. and Valadez, J. (2006). Global review of health care surveys using lot quality assurance sampling (LQAS), 1984-2004. *Social Science & Medicine* **63,** 1648–1660.

Sandiford, P. (1993). Lot quality assurance sampling for monitoring immunization programmes: cost-efficient or quick and dirty? *Health Policy Plan* **8,** 217–223.

Valadez, J., Hage, J., and Vargas, W. (2005). Understanding the relationship of maternal health behavior change and intervention strategies in a nicaraguan ngo network. *Social Science and Medicine* **61,** 1356–1368.

Valadez, J. J. (1991). *Assessing child survival programs in developing countries.* Harvard University Press, Cambridge, Mass.

Vanamail, P., Subramanian, S., Srividya, A., Ravi, R., Krishnamoorthy, K., and Das, P. K. (2006). Operational feasibililtiy of lot quality assurance sampling (LQAS) as a tool in routine process monitoring of filariasis control programmes. *Tropical Medicine and International Health* **11,** 1256–1263.

Westerlund, T. and Pettersson, F. (1995). An extended cutting plane method for solving convex MINLP problems. *Computers & Chemical Engineering* **19,** 131–136.

Received October 2016. *Revised February* 2008. *Accepted March* 2008.

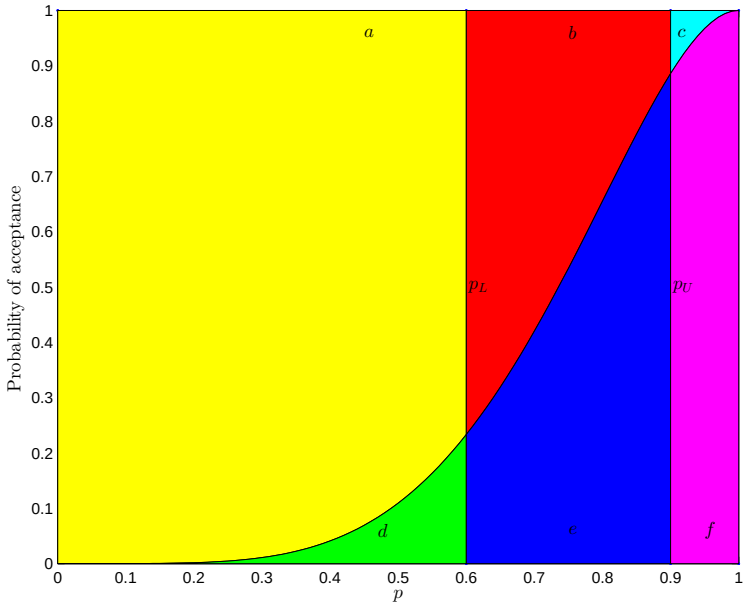

Figure 1. OC curve of the LQAS plan for the *performance criteria-constrained design problem* for scenario S5 with imposed Sensitivity targets of $\sigma_U = \sigma_L = 0.95, p_L = 0.6$ and $p_U = 0.9$ for the *LAPR* and *HAPR* populations. This figure appears in color in the electronic version of this article.

Table 1

Expected risks of the LQAS plan.

		Proportion of non-diseased individuals in the population.		
		$0 \leqslant p \leqslant p_L$	$p_L \leqslant p \leqslant p_U$	$p_U \leqslant p \leqslant 1.0$
Test Outcomes	acceptable	a	b	c
	unacceptable	d	e	f

Table 2

Diagnostic metrics for different prevalence rates of diseased individuals in the population.

		Anticipated prevalence	
		Low rate ($p \leqslant p_L$)	High rate ($p \geqslant p_U$)
Performance metrics	Sens.	$\frac{a}{p_L}$	$\frac{f}{1-p_U}$
	Spec.	$\frac{e+f}{1-p_L}$	$\frac{a+b}{p_U}$
	PPV	$\frac{a}{a+b+c}$	$\frac{f}{d+e+f}$
	NPV	$\frac{e+f}{d+e+f}$	$\frac{a+b}{a+b+c}$

Table 3
LQAS plans for the OC curve-constrained design problem *with* $\alpha = \beta = 0.1$.

Scenario	p_L	p_U	n	r
S1	0.40	0.70	25	13
S2	0.45	0.75	23	13
S3	0.50	0.80	21	13
S4	0.55	0.85	21	14
S5	0.60	0.90	18	13

Table 4

LQAS plans for selected performance criteria-constrained design problem.

Combination of criteria	Scenario	p_L	p_U	n	r	α	β
Sensitivity for *LAPR* &	S1	0.40	0.70	8	4	0.1941	0.1737
Sensitivity for *HAPR*	S2	0.45	0.75	9	5	0.2703	0.0994
(Constraints: (8a,8b))	S3	0.50	0.80	7	4	0.3529	0.0963
($\sigma_U = \sigma_L = 0.95$)	S4	0.55	0.85	8	5	0.4482	0.0498
	S5	0.60	0.90	6	4	0.5798	0.0410
Specificity for *LAPR* &	S1	0.40	0.70	12	6	0.1178	0.1582
Specificity for *HAPR*	S2	0.45	0.75	13	7	0.1654	0.0977
(Constraints: (8c,8d))	S3	0.50	0.80	17	10	0.2248	0.0348
($\theta_U = \theta_L = 0.75$)	S4	0.55	0.85	13	8	0.3457	0.0321
	S5	0.60	0.90	18	12	0.4656	0.0058
Sensitivity for *LAPR* &	S1	0.40	0.70	3	2	0.6570	0.0640
Specificity for *HAPR*	S2	0.45	0.75	3	2	0.6570	0.0640
(Constraints: (8a,8d))	S3	0.50	0.80	3	2	0.6570	0.0640
($\theta_U = 0.75,\ \sigma_L = 0.95$)	S4	0.55	0.85	3	2	0.6570	0.0640
	S5	0.60	0.90	4	3	0.7599	0.0256
PPV for *LAPR* &	S1	0.40	0.70	4	2	0.3483	0.1792
PPV for *HAPR*	S2	0.45	0.75	5	3	0.4718	0.0870
(Constraints: (8e,8f))	S3	0.50	0.80	4	3	0.7599	0.0256
($\pi_U = \pi_L = 0.6$)	S4	0.55	0.85	5	4	0.8319	0.0102
	S5	0.60	0.90	8	7	0.9424	0.0007
NPV for *LAPR* &	S1	0.40	0.70	6	3	0.2557	0.1792
NPV for *HAPR*	S2	0.45	0.75	7	4	0.3529	0.0963
(Constraints: (8g,8h))	S3	0.50	0.80	6	4	0.5798	0.0410
($\varrho_U = \varrho_L = 0.95$)	S4	0.55	0.85	7	5	0.6706	0.0188
	S5	0.60	0.90	5	4	0.8319	0.0102
PPV for *LAPR* &	S1	0.40	0.70	4	2	0.3483	0.1792
NPV for *LAPR*	S2	0.45	0.75	5	3	0.4718	0.0870
(Constraints: (8e,8g))	S3	0.50	0.80	4	3	0.7599	0.0256
($\pi_L = 0.6,\ \varrho_L = 0.95$)	S4	0.55	0.85	5	4	0.8319	0.0102
	S5	0.60	0.90	5	4	0.8319	0.0102

Table 5

Areas of the OC curve for LQAS plan for the performance criteria-constrained design problem *for scenario S5 when we impose Sensitivity targets for* LAPR *and* HAPR *populations are* $\sigma_U = \sigma_L = 0.95, p_L = 0.6$ *and* $p_U = 0.9$.

						Sensitivity for *LAPR*	Sensitivity for *HAPR*
a	b	c	d	e	f	$\frac{a}{p_L}$	$\frac{f}{1-p_U}$
0.5733	0.1369	0.0041	0.0266	0.1631	0.0959	0.9557	0.9590

Table 6

LQAS plans used for the malaria eradication monitoring program (Biedron et al., 2010).

n	r	Sens (LAPR)	Sens (HAPR)	Spec (LAPR)	Spec (HAPR)	PPV (LAPR)	PPV (HAPR)	NPV (LAPR)	NPV (HAPR)
21	12	0.9902	0.6598	0.7511	0.7754	0.7262	0.6541	0.9914	0.9951
20	12	0.9938	0.6922	0.7102	0.8091	0.6957	0.6882	0.9942	0.9912
19	11	0.9892	0.6650	0.7428	0.7806	0.7195	0.6587	0.9904	0.9935
18	10	0.9817	0.6380	0.7773	0.7484	0.7461	0.6281	0.9845	0.9953
17	10	0.9882	0.6711	0.7329	0.7866	0.7115	0.6640	0.9893	0.9912
16	9	0.9794	0.6416	0.7706	0.7515	0.7400	0.6304	0.9825	0.9937
15	9	0.9870	0.6782	0.7205	0.7938	0.7018	0.6701	0.9881	0.9879
14	8	0.9766	0.6459	0.7622	0.7553	0.7324	0.6329	0.9799	0.9913

Table 7

Alternative performance criteria-constrained *LQAS plans for scenarios where* p_L*=0.4 and* $p_U = 0.7$.

Combination of criteria	Scenario	p_L	p_U	n	r	α	β	$\alpha+\beta$
Sensitivity for *LAPR* & Sensitivity for *HAPR* (Constraints: (8a,8b)) ($\sigma_U = \sigma_L = 0.99$)	S1'	0.40	0.70	21	11	0.0676	0.0849	0.1525
Specificity for *LAPR* & Specificity for *HAPR* (Constraints: (8c,8d)) ($\theta_U = \theta_L = 0.75$)	S2'	0.40	0.70	12	6	0.1178	0.1582	0.2761
PPV for *LAPR* & PPV for *HAPR* (Constraints: (8e,8f)) ($\pi_U = \pi_L = 0.65$)	S3'	0.40	0.70	9	5	0.2703	0.0994	0.3697
NPV for *LAPR* & NPV for *HAPR* (Constraints: (8g,8h)) ($\varrho_U = \varrho_L = 0.99$)	S4'	0.40	0.70	19	10	0.0839	0.0885	0.1724

Table 8

OC curve-constrained LQAS plans for various scenarios where p_L=0.4, $p_U = 0.7$, $\alpha = 0.03$ and $\beta = 0.1$ that meet multiple diagnostic metric requirements in the first column.

Combination of criteria	Scenario	p_L	p_U	n	r	α	β	$\alpha+\beta$
Specificity for *LAPR* & Sensitivity for *HAPR* (Constraints: (8c,8b)) ($\theta_L = 0.75$, $\sigma_U = 0.99$)	S5'	0.40	0.70	32	16	0.0138	0.0920	0.1058
PPV for *LAPR* & NPV for *HAPR* (Constraints: (8e,8h)) ($\pi_L = 0.65$, $\varrho_U = 0.99$)	S6'	0.40	0.70	29	15	0.0293	0.0710	0.1002
Specificity for *HAPR* & PPV for *HAPR* (Constraints: (8d,8f)) ($\theta_U = 0.75$, $\pi_U = 0.65$)	S7'	0.40	0.70	34	18	0.0268	0.0444	0.0712