Research Article

Statistics in Medicine

Received 17 April 2009, Accepted 20 April 2010 Published online 16 June 2010 in Wiley Interscience

(www.interscience.wiley.com) DOI: 10.1002/sim.3979

# Approximate models for aggregate data when individual-level data sets are very large or unavailable

**Erol A. Peköz,[a*†] Michael Shwartz,[a,b] Cindy L. Christiansen[c] and Dan Berlowitz[d]**

**In this article, we study a Bayesian hierarchical model for profiling health-care facilities using approximately sufficient statistics for aggregate facility-level data when the patient-level data sets are very large or unavailable. Starting with a desired patient-level model, we give several approximate models and the corresponding summary statistics necessary to implement the approximations. The key idea is to use sufficient statistics from an approximate model fitted by matching up derivatives of the models' log-likelihood functions. This derivative matching approach leads to an approximation that performs better than the commonly used approximation given in the literature. The performance of several approximation approaches is compared using data on 5 quality indicators from 32 Veterans Administration nursing homes. Copyright © 2010 John Wiley & Sons, Ltd.**

**Keywords:** approximate Bayesian models; confidential data; Poisson binomial

## 1. Introduction

The field of approximate Bayesian computation (ABC) has developed in reaction to the computational demands of the growing size of data sets used in biology [1]. It is often necessary to summarize a huge amount of data by a few well-chosen summary statistics in situations where there are no exact sufficient statistics available. The recent article by Joyce and Marjoram [2] develops a framework for assessing whether an approximately sufficient statistics is useful; Le Cam [3] also addresses the concept of approximate sufficiency.

In this article, we look at a common statistical model often used to profile health-care facilities and develop customized approximately sufficient statistics. The setting of the problem is that there is a large amount of patient-level data located at many different facilities. A research group has developed risk-adjustment models to compute from the individual-level patient data a predicted probability that each patient experiences each of the different adverse events of interest and has developed a software that allows individual providers to calculate expected rates of adverse events; providers can compare these expected rates to observed rates as a way of monitoring their performance over time. In addition, a policy group would like to profile facilities. However, due to the volume of data and possible concerns about patient confidentiality, it is not feasible to transmit patient-level data to the policy group.

A realistic context for this problem arises out of software developed by the Agency for Healthcare Research and Quality (AHRQ) that hospitals can download in order to calculate risk-adjusted rates of patient safety indicators and of inpatient quality indicators from their own administrative data. As other measures of quality become acceptable (e.g. hospital re-admission rates), AHRQ is likely to continue to develop the software for calculation of relevant risk-adjusted rates. This software makes it easy for hospitals to use the quality indicators as part of their quality improvement programs. A useful service that AHRQ, other federal or state agencies, or private organizations could provide is to profile hospitals

[a]*Boston University School of Management, 595 Commonwealth Avenue, Boston, MA 02215, U.S.A.*
[b]*Center for Organization, Leadership and Management Research, VA Boston Healthcare System, Boston, MA, U.S.A.*
[c]*Boston University School of Public Health, 715 Albany Street, T-3W, Boston, MA 02118, U.S.A.*
[d]*Center for Health Quality, Outcomes and Economic Research, Bedford VA Hospital, 200 Springs Road, Bedford, MA 01730, U.S.A.*
[*]*Correspondence to: Erol A. Peköz, Boston University School of Management, 595 Commonwealth Avenue, Boston, MA 02215, U.S.A.*
[†]*E-mail: pekoz@bu.edu*

*Contract/grant sponsor: Department of Veterans Affairs Health Services Research and Development Services; contract/grant number: IIR 06-260*

based on the data generated from the software. Such an organization would encourage hospitals to submit results from use of the risk-adjustment software (much as AHRQ encourages hospitals to submit the results from use of its Hospital Survey on Patient Safety Culture, from which it then prepares comparative reports). To encourage flexibility in profiling, it would be most desirable for each hospital to submit, for each eligible patient for each adverse event of interest, the predicted probability of the outcome and whether or not the adverse event occurred. However, this would involve the transfer of large volumes of information. In addition, because patient-level information is being transferred, a variety of IRB considerations are likely to arise.

The goal of the research we present here is to identify useful summary statistics on quality indicators at each facility that can be conveniently transmitted to the profiling organization so that estimation approaches involving 'shrinkage' across facilities can be done. It is natural to consider sufficient statistics for this purpose, but there are none in our setting. We evaluate several alternate approaches to deal with the lack of sufficient statistics and identify one that performs well and does not introduce any computational complexity.

The organization of this article is as follows: In Section 2, we describe the statistical model and discuss the fact that sufficient statistics are not available. In Section 3, we describe the data to which we will apply our approach. In Section 4, we discuss two approaches for fitting approximate models for the data. In Section 5, we discuss the approximate models, in Section 6 we illustrate our approach using the data, in Section 7 we give a summary of the results, and Section 8 is an appendix containing the proofs.

## 2. The model

Let $X_{ij}$ be a binary data variable that equals 1 if a particular adverse event happens to patient $i$, $1 \leqslant i \leqslant N_j$, in health-care facility $j$, $1 \leqslant j \leqslant M$, and it equals zero otherwise. We are interested in the following hierarchical model, which we subsequently refer to as the 'exact' model, where $X_{ij}$ are the data values, $p_{ij}$ are known constants, $f$ is a given (known) link function and $\theta_j$, $\mu$, $\sigma$ are unknown parameters that we would like to estimate:

$$X_{ij}|\theta_j, p_{ij} \sim \text{Bernoulli}(f(\theta_j, p_{ij})) \quad \text{and} \quad \theta_j|\mu, \sigma \sim \text{Normal}(\mu, \sigma^2). \tag{1}$$

In this model, $p_{ij}$ is a risk-adjusted probability that has been previously calculated by taking into account various patient-specific characteristics, and it represents the chance patient $i$ at facility $j$ would have an adverse event if they were at an average facility. The parameter $\theta_j$ is a facility-specific factor, representing the quality at facility $j$, that increases or decreases the probability of an adverse event for the facility's patients. Note that this is a random effect model and therefore there should be shrinkage across facilities when estimating the model parameters. Our goal is to estimate the parameters $\theta_j$ for a function $f$ such as

$$f(\theta, p) = \text{logit}^{-1}(\theta + \text{logit}(p)) = (1 + (1-p)e^{-\theta}/p)^{-1},$$

which we refer to as the logit link function, or

$$f(\theta, p) = \min(1, pe^{\theta}),$$

which we refer to as the log link function. These functions are both sensible to use, as they both equal $p$ when $\theta = 0$ and they are increasing in $\theta$. This means $\theta = 0$ corresponds approximately to an average facility and higher values of $\theta$ represent worse facilities. Typically, the function $f$ will be increasing in $\theta$ and will have $f(0, p) = p$.

The standard Bayesian hierarchical modeling approach is to put prior distributions on the unspecified parameters and estimate the means of all the parameters conditional on the data. The difficulty in this situation is that the values of $X_{ij}$ and $p_{ij}$ are both confidential and are too numerous to conveniently transmit from each of the facilities to the main organization that performs the profiling. We need a method for summarizing these sets of numbers so that each facility only needs to report a small number of summary statistics. The usual approach is to use sufficient statistics, but, as we discuss in Section 4 below, there are no sufficient statistics we can use because the $p_{ij}$ parameters are different across the $N_j$ patients—if they were identical, we could use the total number of adverse events as a sufficient statistic which, conditional on the parameters, would have a binomial distribution. When they are not identical, the sum has what is called a Poisson-binomial distribution: the sum of independent Bernoulli trials with different success probabilities.

To implement the model in (1), it may seem as though each facility could simply transmit the maximum likelihood estimate of $\theta_j$, but once these are gathered together there would be no obvious way to accurately allow for 'shrinkage' of the estimators across facilities. In fact, we test the approach of having each facility transmit the posterior mean and standard deviation of $\theta_j$ using separate (fixed-effect) models and then 'shrinking' them at the central organization using the assumption that they have a normal distribution. This approach does not work well on the data to which we applied it;

we call this the 'two-stage normal' approximation and illustrate it in Table III below. What is really necessary is for each facility to transmit enough information so that the central organization can, as best as possible, re-create the entire likelihood function for the data from each facility as a function of $\theta$.

As there are no sufficient statistics available, the key idea we propose in this article is to use sufficient statistics from an approximate model fitted by matching moments or derivatives of the models' log-likelihood functions. In our recommended approach, described in Proposition 2 below, we use a binomial distribution to approximate the sum of non-identically distributed Bernoulli random variables where the two parameters of this distribution (the number of trials and the probability of success) are fitted using the first two moments of the probabilities from each facility. As a result, each health-care facility will simply need to report the total number of adverse events, as well as the first two moments of the predicted probabilities. For a different approach to approximating a similar model, see [4].

## 3. About the data

The data for this study were originally collected in 1998 from 35 Department of Veterans Affairs (VA) nursing homes that were selected to represent a balanced sample of different sizes, locations, and quality of care [5]. Complete data were available from 32 of the 35 nursing homes on five binary quality indicators (QIs) that reflect changes in patients' status over time. These 32 facilities are the ones included in our study. All the QIs have been used previously as measures of nursing home quality: pressure ulcer development [6–9]; functional decline [7–9]; behavioral decline [7, 10]; mortality [7, 11]; and preventable hospitalizations [12].

Data used in calculating QIs were from semi-annual patient assessments performed for case-mix-based reimbursements. Pressure ulcer development was recorded if a patient who was ulcer free at one assessment had a stage 2 or deeper pressure ulcer at the subsequent assessment. Functional decline was measured by a change between assessments in a score measuring limitations in eating, toileting, and transferring. Behavioral decline was measured by a change in a score measuring extent of verbal disruption, physical aggression, and socially inappropriate behavior. Mortality was recorded if there is a death within 6 months of an assessment regardless of location. Preventative hospitalizations occurred if the patient was admitted to an acute medical unit within 6 months of an assessment for one of 13 conditions identified as a potentially preventable hospitalization [13]. Risk-adjustment models have been developed for these QIs: pressure ulcer development [6], functional decline [14], behavioral decline and mortality [15], and preventable hospitalizations [16].

For each patient, the risk-adjustment models give a predicted probability of the adverse event in a six-month period based on the risk factors at the time of initial assessment. Thus, for each patient, we know whether each adverse event occurred (indicated by a 0 or 1) and the predicted probability of the adverse event from the relevant risk-adjustment model. Table I shows percentiles of the distribution of the predicted probabilities for each of the QIs. For 4 of the 5 QIs, individual predictions range from close to 0 to around 0.50 to 0.60; for mortality, there is a longer right tail, with predictions ranging almost to 1.

Table II shows percentiles of the distribution of the number of eligible cases (the total number of people who could potentially have the given adverse event) at each facility for the pressure ulcer and mortality QIs. The number of eligible cases ranges from 32 to over 400 for pressure ulcers (and for functional decline and behavioral decline, not shown in the table); and from slightly over 80 to almost 1200 for mortality (and preventable hospitalizations, also not shown). The second part of Table II shows percentiles of the distribution of the ratio of the observed number of adverse events at each facility to the predicted number for each of the QIs. The distribution is quite wide for 3 of the QIs, ranging from zero to over 2; it is tightest for mortality, ranging from 0.65 to 1.29.

**Table I**. Percentiles of the distribution of predicted probabilities by type of adverse event (pu—pressure ulcer development; fd—functional decline; bd—behavioral decline; mort—mortality; ph—preventable hospitalization).

| Percentile | pu | fd | bd | mort | ph |
|---|---|---|---|---|---|
| Maximum | 0.469 | 0.512 | 0.545 | 0.978 | 0.573 |
| 95th | 0.107 | 0.348 | 0.320 | 0.759 | 0.197 |
| 90th | 0.084 | 0.307 | 0.275 | 0.451 | 0.158 |
| 75th | 0.055 | 0.250 | 0.206 | 0.221 | 0.109 |
| 50th | 0.038 | 0.170 | 0.158 | 0.127 | 0.064 |
| 25th | 0.021 | 0.081 | 0.127 | 0.074 | 0.027 |
| 10th | 0.013 | 0.037 | 0.101 | 0.046 | 0.017 |
| Minimum | 0.011 | 0.011 | 0.062 | 0.006 | 0.004 |

**Table II**. Percentiles of the distribution of the number of eligible cases and of the ratio of observed to predicted number of adverse events in each of the 32 facilities (pu—pressure ulcer development; fd—functional decline; bd—behavioral decline; mort - mortality; ph—preventable hospitalization. The numbers of eligible cases for fd and bd are approximately the same as for pu; the number of eligible cases for ph is approximately the same as for mort).

| | Number of eligible cases | | Observed/predicted | | | | |
|---|---|---|---|---|---|---|---|
| Percentile | pu | mort | pu | fd | bd | mort | ph |
| Minimum | 32 | 83 | 0.00 | 0.15 | 0.00 | 0.65 | 0.23 |
| 25th | 85 | 228 | 0.44 | 0.60 | 0.39 | 0.90 | 0.62 |
| 50th | 159 | 324 | 0.90 | 0.96 | 0.81 | 0.99 | 0.96 |
| 75th | 236 | 521 | 1.41 | 1.10 | 1.16 | 1.08 | 1.23 |
| Maximum | 408 | 1193 | 2.07 | 2.26 | 2.78 | 1.29 | 1.59 |

## 4. Two approaches for finding an approximate model

The log-likelihood function for the data at facility $j$ under the model in (1) is

$$LL_j(\theta)=\sum_i X_{ij}\log f(\theta, p_{ij})+(1-X_{ij})\log(1-f(\theta, p_{ij})). \tag{2}$$

In general, there are no sufficient statistics for the individual-level data $X_{ij}$ under this model. For example, with $M=1$ facility having $N_1=2$ patients with probabilities $p_1=\frac{1}{3}$, $p_2=\frac{1}{4}$ (omitting the subscript $j$) and $f(\theta, p)=\theta p$, we simplify (2) to get

$$LL(\theta)=X_1\log\frac{\theta}{3-\theta}+X_2\log\frac{\theta}{4-\theta}+\log\frac{\theta}{3-\theta}+\log\frac{\theta}{4-\theta}.$$

Since this log-likelihood is of the form $aX_1+bX_2+a+b$, it can be seen that this cannot be expressed in terms of two quantities where one depends on $a$, $b$ but not $X_1$, $X_2$ and the other depends on $X_1$, $X_2$ but not $a$, $b$.

In the absence of a sufficient statistic, we therefore propose several alternative approximate models for the data for which the sum of the data values

$$X_j=\sum_i X_{ij}$$

is a sufficient statistic for a facility. Notice, we omit the subscript $i$ to indicate that the binary variables have been summed over $i$ for a given facility.

We say that an approximate model for the data is a $k$ moment approximation if, when $\theta=0$, the first $k$ moments of $X_j$ conditional on all other parameters under the exact model coincide with the corresponding conditional moments under the approximate model. We also say that an approximate model for the data is a $k$ derivative approximation for the exact model if the first $k$ derivatives of its log-likelihood function with respect to $\theta$ evaluated at $\theta=0$ match the first $k$ derivatives for the exact model $LL_j(\theta)$.

These definitions can be formalized by saying that an approximate model for the data $X_j$ given $\theta_j$ along with some set of parameters $S$ is a $k$ moment approximation if, with the above definitions

$$E[(X_j)^m|\theta_j=0, p_{1j}, p_{2j}, \ldots]=E^a[(X_j)^m|\theta_j=0, S]$$

for $m=1, 2, \ldots, k$. We use the superscript $a$ to denote expectations with respect to the approximate model. An approximate model with log-likelihood function $LL_j^a(\theta)$ for the data $X_j$ is a $k$ derivative approximation with respect to $\theta$ if

$$\frac{\partial^m}{\partial\theta^m}LL_j^a(\theta)|_{\theta=0}=\frac{\partial^m}{\partial\theta^m}LL_j(\theta)|_{\theta=0} \quad \text{for } m=1, 2, \ldots, k. \tag{3}$$

We say that an approximate model with multiple parameters is a $k$ derivative approximation with respect to a given set of parameters if (3) holds with respect to each parameter separately.

Both these approaches seem sensible. The moment approximation is simpler in the sense that it does not depend on the link function $f$, whereas the derivative approximation does depend on $f$. As we show empirically later, when they are different, the derivative approximation performs better. For some choices of $f$ the two approximations will coincide, whereas for others they will be different.

The intuition for why the derivative approximation performs better is the following: To estimate the distribution of $X_j$ conditional on the parameters, it seems best to match moments to directly approximate the distribution. To approximate

the posterior distribution of $\theta$ conditional on the data requires the likelihood function and so it seems natural that the best performance is obtained when the likelihood function is directly approximated using an approach such as the derivative approach.

In order to find a good approximation for this model, it is important for practical reasons that it be easy to implement in standard Bayesian software packages like WINBUGS. Thus, it would be best to use only common distributions such as the binomial, Poisson or normal. In the article by Peköz *et al.* [17], several different approximations to the distribution of the sum of independent Bernoulli random variables are considered and evaluated. A Poisson approximation performs well if the Bernoulli probabilities are all very small. The normal approximation performs well if the probabilities are mostly near $\frac{1}{2}$. The binomial distribution performs well if the probabilities are either very small or very large overall—or very similar to each other. In practical applications, the probabilities will be quite diverse: some facilities will have very low numbers while others have probabilities spread more evenly between zero and one. It is therefore important to have an approximation that can flexibly adapt across all the situations. With a standard binomial approximation, it makes sense to pick the probability parameter to match the mean of the distribution to be approximated. One way of improving the binomial approximation is to let both the probability and the number of trials be two free parameters that can be fitted. This means in principle it may be possible to match the first two moments. Another idea for improving the Poisson approximation is to consider a translated Poisson approximation where a constant is added to a Poisson random variable. The constant and the Poisson rate are then two free parameters that can be adjusted to approximately match the first two moments.

A good combination of these ideas is to try a translated binomial approximation, where a constant is added to a binomial random variable. This then allows three parameters to be adjusted and, in most cases, three moments or three derivatives could be matched. The article by Peköz *et al.* [17] evaluates all the approximations discussed here (and derives error bounds) over a range of different settings and finds that a translated binomial approximation performs the best. As we show later, our analysis demonstrates that a binomial approximation with two fitted parameters performs much better than the usual binomial approximation with one fitted parameter, and the three parameter binomial approximation does not perform much better than the two parameter binomial approximation—though it requires additional computations and data to be transmitted and stored. We will therefore recommend the two-parameter binomial approximation, though we will illustrate all three approximations below. Once an approximation is fit to the data, the unknown parameters can be computed using either a Bayesian approach in WINBUGS, maximum likelihood estimation or other transform methods [18].

## 5. The approximate models

The first approximation we propose is one that is commonly used and is both a one-moment approximation and one-derivative approximation. It assumes that $X_j$, conditional on the parameters, has a binomial distribution, where the number of trials is the number of eligible cases and the binomial probability is chosen to be the average Bernoulli probability at that facility. To implement this approximation, each facility must simply transmit the total number of adverse events $X_j$ along with the total number of eligible cases $N_j$ and the average probability $p_j = \sum_i p_{ij}/N_j$ at that facility. This first approximation is summarized next.

*Proposition 1*
The one moment, one-derivative approximation. With the above definitions and the model in (1) with

$$f(\theta, p) = \text{logit}^{-1}(\theta + \text{logit}(p)),$$

and $p_j = \sum_i p_{ij}/N_j$, the model

$$X_j|\theta_j, p_j \sim \text{Binomial}(N_j, f(\theta_j, p_j))$$

is both a one-moment approximation and a one-derivative approximation with respect to $\theta_j$, $j = 1, \ldots, M$ for the exact model (1).

*Proof*
See Appendix.

The second approximation we propose is both a two-moment approximation and a two-derivative approximation for the exact model. It assumes that $X_j$, conditional on the parameters has a binomial distribution, where the number of trials and the probability of success are both chosen to match the first two moments. This means that the number of trials in the binomial distribution $n_j$ will not necessarily equal the number of eligible cases $N_j$ in the facility. To implement this

approximation, each facility must simply transmit the total number of adverse events $X_j$ along with the first two moments of the probabilities $\sum_i p_{ij}^2$ and $\sum_i p_{ij}$. We will recommend this approximation based on the numerical illustrations below.

*Proposition 2*
The two-moment, two-derivative approximation. With the above definitions and the model in (1) with

$$f(\theta, p) = \text{logit}^{-1}(\theta + \text{logit}(p)),$$

and $p_j = \sum_i p_{ij}^2 / \sum_i p_{ij}$, $n_j = \sum_i p_{ij} / p_j$, the model

$$X_j | \theta_j, n_j, p_j \sim \text{Binomial}(n_j, f(\theta_j, p_j))$$

is both a two-moment approximation and a two-derivative approximation with respect to $\theta_j$, $j = 1, \ldots, M$ for the exact model (1).

*Proof*
See Appendix.

*Remark 1*
With the above definition, $n_j$ may not necessarily be an integer, though the number of trials in a binomial distribution must be an integer. In principle, we could round off the number of trials to the nearest integer. In practice, we are using the likelihood as a function of $\theta$ and so the use of non-integer values in some sense represents an interpolation of the likelihood functions at the two adjacent integer numbers of trials. Standard software packages such as WINBUGS allow non-integer values for the number of trials in a binomial distribution and so a non-integer $n_j$ is not a limitation of the approach. So that the propositions make sense, we can assume that $n_j$ happens to be an integer—though the approximations can be improved using non-integer values.

Our third approximation is both a three-moment approximation and a three-derivative approximation. It assumes that $X_j$, conditional on the parameters, has a translated binomial distribution. A translated binomial distribution is simply a binomial distribution plus a constant. The three parameters, which consist of the number of trials, the probability of success and the amount translated, are chosen to match the first three moments. To implement this approximation, each facility must transmit the total number of adverse events $X_j$ as well as the first three moments of the probabilities $p_{ij}$.

*Proposition 3*
The three-moment, three-derivative approximation. With the above definitions and the model in (1) with

$$f(\theta, p) = \text{logit}^{-1}(\theta + \text{logit}(p)),$$

and $\lambda_{kj} = \sum_i (p_{ij})^k$ along with

$$p_j = \frac{\lambda_{2j} - \lambda_{3j}}{\lambda_{1j} - \lambda_{2j}},$$

$$n_j = \frac{\lambda_{1j} - \lambda_{2j}}{p_j(1 - p_j)},$$

$$s_j = \lambda_{1j} - n_j p_j,$$

the model

$$(X_j - s_j | \theta_j, s_j, n_j, p_j) \sim \text{Binomial}(n_j, f(\theta_j, p_j))$$

is both a three-moment approximation and a three-derivative approximation with respect to $\theta_j$, $j = 1, \ldots, M$ for the exact model (1).

*Proof*
See Appendix.

*Remark 2*
It may be that the quantity $s_j$ is not an integer. However, as discussed in Remark 1 above, this is not a limitation.

Our next result is that for the log link function $f(\theta, p) = \min(1, pe^{\theta})$, the $k$ moment approximations above do not coincide with the $k$ derivative approximations. We can derive a different $k$ derivative approximation. Empirical evidence shows that this performs much better than the $k$ moment approximation.

*Proposition 4*
The two-derivative approximation differs from the two-moment approximation with the log link function. With the above definitions and the exact model in (1) with

$$f(\theta, p) = \min(1, pe^{\theta}),$$

and $p_j = \sum_i p_{ij}^2 / \sum_i p_{ij}$, $n_j = \sum_i p_{ij} / p_j$, the model

$$X_j | \theta_j, n_j, p_j \sim \text{Binomial}(n_j, f(\theta_j, p_j))$$

is a two-moment approximation but is not a two-derivative approximation with respect to $\theta_j$, $j = 1, \ldots, M$ for the exact model (1).

*Proof*
See Appendix.

We next present the two-derivative model for the log link function.

*Proposition 5*
With the above definitions and the exact model in (1) with

$$f(\theta, p) = \min(1, pe^{\theta}),$$

and

$$\alpha_j = \sum_i \frac{p_{ij}}{1 - p_{ij}} (1 - X_{ij}),$$

$$\beta_j = \sum_i \frac{p_{ij}}{(1 - p_{ij})^2} (1 - X_{ij}),$$

and letting

$$p_j = 1 - \alpha_j / \beta_j,$$

and

$$n_j = \frac{\alpha_j^2}{\beta_j - \alpha_j} + X_j$$

the model

$$X_j | \theta_j, n_j, p_j \sim \text{Binomial}(n_j, f(\theta_j, p_j))$$

is a two-derivative approximation with respect to $\theta_j$, $j = 1, \ldots, M$ for the exact model (1).

*Proof*
See Appendix.

The final result in this section is a straightforward extension to multiple types of adverse events. Note in this model that $\theta_j$ is an underlying latent measure of quality at facility $j$ that is reflected in the data for each of the quality indicators.

*Proposition 6*
A model for multiple adverse events. Suppose there are $m$ different types of adverse events. Let $X_{ijk}$ equal 1 if person number $i$ in facility number $j$ has an adverse event of type $k$, $1 \leqslant k \leqslant m$, and let it equals zero otherwise. Let $p_{ijk}$ be the given probability associated with this variable. For the $k$th type of adverse event, we define

$$f_k(\theta, p) = \text{logit}^{-1}(d_k + c_k \theta + \text{logit}(p)).$$

Suppose that

$$X_{ijk} | \theta_j, p_{ijk} \sim \text{Bernoulli}(f_k(\theta_j, p_{ijk})) \quad \text{and} \quad \theta_j | \mu, \sigma \sim \text{Normal}(0, 1) \tag{4}$$

is the exact model for some unknown parameters $c_k$, $d_k$ that are to be estimated. With $X_{jk} = \sum_i X_{ijk}$ along with $p_{jk} = \sum_i (p_{ijk})^2 / \sum_i p_{ijk}$, $n_{jk} = \sum_i p_{ijk} / p_{jk}$, the model

$$(X_{jk} | \theta_j, n_{jk}, p_{jk}) \sim \text{Binomial}(n_{jk}, f_k(\theta_j, p_{jk}))$$

**Table III**. The approximate models.

| Proposition | Approximation (link function) | Summary statistics necessary |
|---|---|---|
| 1 | 1-moment, 1-derivative (logit) | $N_j, \sum_i p_{ij}, \sum_i X_{ij}, \forall j$ |
| 2 | 2-moment, 2-derivative (logit) | $\sum_i p_{ij}^2, \sum_i p_{ij}, \sum_i X_{ij}, \forall j$ |
| 3 | 3-moment, 3-derivative (logit) | $\sum_i p_{ij}^3, \sum_i p_{ij}^2, \sum_i p_{ij}, \sum_i X_{ij}, \forall j$ |
| 4 | 2-moment (log) | $\sum_i p_{ij}^2, \sum_i p_{ij}, \sum_i X_{ij}, \forall j$ |
| 5 | 2-derivative (log) | $\sum_i \frac{p_{ij}}{1-p_{ij}}(1-X_{ij}), \sum_i \frac{p_{ij}}{(1-p_{ij})^2}(1-X_{ij}), \sum_i X_{ij}, \forall j$ |
| 6 | 2-moment, multiple QIs (logit) | $\sum_i p_{ijk}^2, \sum_i p_{ijk}, \sum_i X_{ijk}, \forall j,k$ |

is both a two-moment approximation and a two-derivative approximation with respect to $\theta_j$, $j=1,\ldots,M$ for the exact model in (4) when $c_k=1$, $d_k=0$ for $1\leqslant k\leqslant m$.

*Proof*
See Appendix.

We summarize all the approximations, along with the necessary summary statistics, in Table III.

## 6. Numerical results and comparison of the approaches

In this section, we illustrate the different approximations. We will see that the two-derivative approximations (for both link functions we consider) seem to perform much better than the one-derivative approximations, and (for the logit link function) the three-derivative approximation performs quite similarly to the two-derivative approximation—though all the approximations are fairly reasonable. We will illustrate the approximations using two different approaches. In the first approach, we test the approximations on data from the Veterans Health Administration—see Section 3 for a description of the data.

To do the computations, we put flat priors on the top level unknown parameters and then estimate the posterior means of the facility effects conditional on the data under the exact model and various approximate models. We use the notation $\bar{\theta}_j$ to denote our numerical estimate (computed by averaging together samples from the posterior distribution using the software package WINBUGS) of the posterior mean $E[\theta_j|X_{ij},\forall i,j]$ under the exact model and we use the notation $\bar{\theta}_j^a$ to denote our numerical estimate of the posterior mean $E^a[\theta_j|X_j,\forall j]$ under some given approximate model. For the illustrations below, we measure the accuracy of an approximation using the average of the absolute value of the difference between our estimate of the facility-effect parameters under the approximate and exact models. We compute this using

$$\text{Error}=\frac{1}{M}\sum_{j=1}^{M}|\bar{\theta}_j-\bar{\theta}_j^a| \tag{5}$$

except for the multiple quality indicators in Proposition 6, where we instead use

$$\text{Error}=\frac{1}{m}\sum_{k=1}^{m}\frac{1}{M}\sum_{j=1}^{M}|\bar{\theta}_{j,k}-\bar{\theta}_{j,k}^a|, \tag{6}$$

where $\theta_{j,k}=d_k+c_k\theta_j$ and $\theta_{j,k}^a=d_k^a+c_k^a\theta_j^a$.

First, we illustrate the approximations given in Propositions 1, 2, 3. The top half of Table IV gives the errors for models using the logit link function $f(\theta,p)=\text{logit}^{-1}(\theta+\text{logit}(p))$. The standard deviation of the effect sizes in Table V, when compared with the first row of Table IV, shows that the error from the one-moment approximation is not bad; in many cases, it is of a smaller order of magnitude than the estimated standard deviation (of the effect size $\theta$ in Table V). The second line shows that the two-moment approximation is always much better than the one-moment approximation. In several cases, the error is an order of magnitude improvement on the one-moment approximation in the prior line. The third line of the table shows that the three-moment approximation does not perform consistently better than the two-moment approximation with the logit link function. The rightmost column (labeled as '3QIs') shows that when multiple quality indicators are included in the model (using the approach of Proposition 6 above, where only the two-moment case is illustrated—the one- and three-moment versions are analogous to those in Propositions 1 and 3) the

**Table IV**. Errors associated with different approximations for the different quality indicators.

| | pu | fd | bd | mort | ph | 3QIs |
|---|---|---|---|---|---|---|
| Logit link function | | | | | | |
| One-moment | 0.015 | 0.028 | 0.021 | 0.042 | 0.020 | 0.016 |
| Two-moment | 0.004 | 0.003 | 0.007 | 0.003 | 0.002 | 0.004 |
| Three-moment | 0.004 | 0.003 | 0.007 | 0.002 | 0.004 | 0.004 |
| Two-stage normal | 0.103 | 0.018 | 0.182 | 0.003 | 0.012 | |
| Log link function | | | | | | |
| Two-moment | 0.014 | 0.018 | 0.040 | 0.050 | 0.014 | |
| Two-derivative | 0.006 | 0.008 | 0.020 | 0.004 | 0.004 | |

**Table V**. Standard deviation of the effect size $\theta$ for the different quality indicators.

| pu | fd | bd | mort | ph |
|---|---|---|---|---|
| 0.420 | 0.465 | 0.982 | 0.123 | 0.422 |

two-moment approximation performs much better than the one-moment, and the three-moment approximation performs quite similarly to the two-moment approximation.

The fourth line of the table, labeled 'two-stage normal,' illustrates an approximation approach where each facility separately computes its best estimate of $\theta_j$, which is then transmitted to the organization and 'shrunk' using a hierarchical model. Specifically, first the posterior mean and standard deviation for $\theta_j$ is computed separately for each facility using a fixed-effects model with a flat prior for $\theta_j$. Then, these facility estimates are combined in a model where we assume the $\theta_j$ values are the sum of two draws from a normal distribution: one is a distribution with an unknown mean and standard deviation that is common to all facilities, and the other is an error term that is normally distributed with mean zero and standard deviation which is specific to each facility. The fourth line shows that this approximation approach performs better than the one-moment approximation in most cases—though in cases where there are facilities with zero observed adverse events (for pu and bd), it performs much worse. Some adjustments to the model would need to be made to account for zero values. But, the approximation is not nearly as good as the two-moment approach in most cases.

The last two rows of the table show, using the log link function $f(\theta, p)=\min(1, pe^{\theta})$, that the two-derivative approximation in Proposition 5 performs much better than the two-moment approximation. This is evidence for the theory that the derivative approach works better than the moment approach when they differ.

In the second approach, we use to illustrate the accuracy of the approximations, we generate simulated data and compare the maximum likelihood estimates of the unknown parameters across the different models. In particular, we create a single health-care facility with 100 patients having predetermined risk-adjusted probabilities. In the first simulation, we create these probabilities according to equally spaced percentiles of a beta distribution with a coefficient of variation equal to 1; in the second simulation, the coefficient of variation is equal to 0.4. We also vary the means of these distributions from 0 to 0.2. These ranges are chosen in order to mimic the summary statistics we saw in the actual data for the risk-adjusted probabilities. We also use different values of the unknown parameter $\theta$ ranging from 0 through 4. Given specific values of the parameters, we use the logit link function to generate simulated data and then compute the maximum likelihood estimate of $\theta$ under the exact model and then under the various approximations. Our simulations show that when $\theta$ is in the range from 0 to 2, the two-moment approximation generally performs better than the one-moment approximation. As the value of the $\theta$ parameter becomes very large (this corresponds to the case where there is a large amount of variation in the risk-adjusted probabilities across facilities and implies large differences in underlying quality), we see that a two-moment approximation performs worse than the one-moment approximation. As there is rarely such large variation in underlying quality across facilities, it is not a surprise that the two-moment approximation performs well when tested on real data.

## 7. Conclusion

In conclusion, we see that the two- and three-derivative approximations performed the best. The three-derivative approximation requires more data and performs similarly to the two-derivative approximation, and so we recommend the two-derivative approximation. For the logit link function, the two-derivative and two-moment approximations are identical though, for the log link function, the two-derivative approximation seems to perform better. To implement the

two-derivative approximation, each health-care facility needs to report three summary statistics to the central profiling organization: the first two moments of the predicted risk-adjusted probabilities, as well as the total number of adverse events occurring. Then, the two-parameter binomial distributional approximation described in Proposition 2 can be fitted and used to estimate the likelihood function.

The approach we recommend is nice because the approximation is quite simple and easy to implement. The standard (one-moment) binomial approximation is commonly used and this two-parameter binomial approximation is an improvement with essentially no additional computational costs or model complexity other than computing and transmitting one additional summary statistic.

## Appendix

In this section, we prove the propositions above. We drop the facility subscript $j$ in each proof.

*Proof of Proposition 1*
As the first moment of a binomial distribution equals $np$, the fact that this model is a one-moment estimator follows immediately. To see that it is also a one-derivative estimator, let

$$f(\theta, p)=\frac{1}{1+\dfrac{1-p}{p}e^{-\theta}},$$

and recall that the exact log-likelihood function is

$$LL(\theta)=\sum_i X_i \ln(f(\theta, p_i))+(1-X_i)\ln(1-f(\theta, p_i)).$$

The log-likelihood function for this approximation is

$$LL^a(\theta)=X\ln(f(\theta, p))+(N-X)\ln(1-f(\theta, p))+c,$$

where we write $X=\sum_i X_i$ and $c=\ln(N!/(X!(N-X)!)$. It then follows that

$$\frac{\partial}{\partial\theta}LL(\theta)=\sum_i(p_i e^{-\theta}-e^{-\theta}-p_i)^{-1}(p_i-X_i p_i-X_i e^{-\theta}+X_i p_i e^{-\theta})$$

along with

$$\frac{\partial}{\partial\theta}LL^a(\theta)=(pe^{-\theta}-e^{-\theta}-p)^{-1}(Np-Xp-Xe^{-\theta}+Xpe^{-\theta}),$$

and we obtain

$$\frac{\partial}{\partial\theta}LL(\theta)|_{\theta=0}=\sum_i(X_i-p_i),$$

and

$$\frac{\partial}{\partial\theta}LL^a(\theta)|_{\theta=0}=(X-Np).$$

Solving the equation

$$\frac{\partial}{\partial\theta}LL(\theta)|_{\theta=0}=\frac{\partial}{\partial\theta}LL^a(\theta)|_{\theta=0}$$

for $p$ yields the result of the proposition. □

*Proof of Proposition 2*
As the mean and variance of a binomial distribution equals $np$ and $np(1-p)$ respectively, the fact that this model is a two-moment estimator follows immediately. To see that it is also a two-derivative estimator, we again let

$$f(\theta, p)=\frac{1}{1+\dfrac{1-p}{p}e^{-\theta}},$$

and recall that the exact log-likelihood function is

$$LL(\theta)=\sum_i X_i \ln(f(\theta, p_i))+(1-X_i)\ln(1-f(\theta, p_i)).$$

The log-likelihood function for this approximation is

$$LL^a(\theta)=X\ln(f(\theta, p))+(n-X)\ln(1-f(\theta, p))+c,$$

where we write $X=\sum_i X_i$ and $c=\ln(N!/(X!(N-X)!)$ and now the parameters $n$, $p$ are both to be solved for. It then follows, using the results from the previous proposition, that

$$\frac{\partial^2}{\partial\theta^2}LL(\theta)=\sum_i(p_i e^{-\theta}-e^{-\theta}-p_i)^{-2}(p_i-1)(e^{-\theta})p_i,$$

and

$$\frac{\partial^2}{\partial\theta^2}LL^a(\theta)=(pe^{-\theta}-e^{-\theta}-p)^{-2}(p-1)(e^{-\theta})np,$$

and we obtain

$$\frac{\partial^2}{\partial\theta^2}LL(\theta)|_{\theta=0}=\sum_i(p_i^2-p_i),$$

and

$$\frac{\partial^2}{\partial\theta^2}LL^a(\theta)|_{\theta=0}=(np^2-np).$$

Using the results from the previous proposition, we can then solve the system of equations

$$\frac{\partial}{\partial\theta}LL(\theta)|_{\theta=0}=\frac{\partial}{\partial\theta}LL^a(\theta)|_{\theta=0},$$

and

$$\frac{\partial^2}{\partial\theta^2}LL(\theta)|_{\theta=0}=\frac{\partial^2}{\partial\theta^2}LL^a(\theta)|_{\theta=0}$$

for both $n$ and $p$ to obtain the result of the proposition. □

*Proof of Proposition 3*

The mean, variance, and third-central moment of a binomial distribution equals $np$, $np(1-p)$ and $np(1-p)(1-2p)$, respectively, and for a Poisson-binomial distribution $W$ with probabilities $p_1, p_2, \ldots$ and $\lambda_j=\sum_i(p_i)^j$, we have $E[W]=\lambda_1$, $\mathrm{Var}(W)=\lambda_1-\lambda_2$ and $E[(W-E[W])^3]=\lambda_1-3\lambda_2+2\lambda_3$. The fact that this model is a three-moment estimator follows from these facts. To see that it is also a three-derivative estimator, we again let

$$f(\theta, p)=\frac{1}{1+\dfrac{1-p}{p}e^{-\theta}},$$

and recall that the exact log-likelihood function is

$$LL(\theta)=\sum_i X_i \ln(f(\theta, p_i))+(1-X_i)\ln(1-f(\theta, p_i)).$$

The log-likelihood function for this approximation is

$$LL^a(\theta)=(X-s)\ln(f(\theta, p))+(n-X-s)\ln(1-f(\theta, p))+c,$$

where we write $X=\sum_i X_i$ and $c=\ln(N!/(X!(N-X)!)$ and now the parameters $s$, $n$, $p$ are all to be solved for. It then follows, using the results from the previous proposition, that

$$\frac{\partial^3}{\partial\theta^3}LL(\theta)=\sum_i(p_i e^{-\theta}-e^{-\theta}-p_i)^{-3}(p_i-e^{-\theta}+p_i e^{-\theta})(p_i-1)(e^{-\theta})p_i,$$

and

$$\frac{\partial^3}{\partial\theta^3}LL^a(\theta)=(pe^{-\theta}-e^{-\theta}-p)^{-3}(p-e^{-\theta}+pe^{-\theta})(p-1)(e^{-\theta})np,$$

and we obtain

$$\frac{\partial^3}{\partial\theta^3}LL(\theta)|_{\theta=0}=\sum_i(3p_i^2-p_i-2p_i^3),$$

and

$$\frac{\partial^3}{\partial\theta^3}LL^a(\theta)|_{\theta=0}=(3np^2-np-2np^3).$$

Using the results from the previous proposition, we can then solve the system of equations

$$\frac{\partial}{\partial\theta}LL(\theta)|_{\theta=0}=\frac{\partial}{\partial\theta}LL^a(\theta)|_{\theta=0},$$

$$\frac{\partial^2}{\partial\theta^2}LL(\theta)|_{\theta=0}=\frac{\partial^2}{\partial\theta^2}LL^a(\theta)|_{\theta=0},$$

and

$$\frac{\partial^3}{\partial\theta^3}LL(\theta)|_{\theta=0}=\frac{\partial^3}{\partial\theta^3}LL^a(\theta)|_{\theta=0}$$

for $s$, $n$ and $p$ to obtain the result of the proposition. □

*Proof of Propositions 4 and 5*

The fact that this approximation is a two-moment approximation follows from the argument in the proof of Proposition 2 above. Let

$$f(\theta,p)=pe^{\theta}$$

and recall that the exact log-likelihood function is

$$LL(\theta)=\sum_i X_i\ln(f(\theta,p_i))+(1-X_i)\ln(1-f(\theta,p_i)).$$

The log-likelihood function for this approximation is

$$LL^a(\theta)=X\ln(f(\theta,p))+(n-X)\ln(1-f(\theta,p))+c,$$

where we write $X=\sum_i X_i$ and $c=\ln(N!/(X!(N-X)!)$ and now the parameters $n$, $p$ are both to be solved for. It then follows that

$$\frac{\partial}{\partial\theta}LL(\theta)=\sum_i X_i-e^{\theta}p_i\frac{1-X_i}{1-e^{\theta}p_i},$$

$$\frac{\partial^2}{\partial\theta^2}LL(\theta)=\sum_i -e^{\theta}p_i\frac{1-X_i}{1-e^{\theta}p_i}-e^{2\theta}p_i^2\frac{1-X_i}{(1-e^{\theta}p_i)^2}$$

along with

$$\frac{\partial}{\partial\theta}LL^a(\theta)=X-pe^{\theta}\frac{n-X}{1-pe^{\theta}},$$

and

$$\frac{\partial^2}{\partial\theta^2}LL^a(\theta)=-pe^{\theta}\frac{n-X}{1-pe^{\theta}}-p^2e^{2\theta}\frac{n-X}{(1-pe^{\theta})^2},$$

and we obtain

$$\frac{\partial}{\partial\theta}LL(\theta)|_{\theta=0}=\sum_i -p_i\frac{1-X_i}{1-p_i}-p_i^2\frac{1-X_i}{(1-p_i)^2},$$

and

$$\frac{\partial}{\partial\theta}LL^a(\theta)|_{\theta=0}=X-p\frac{n-X}{1-p},$$

and

$$\frac{\partial^2}{\partial\theta^2}LL^a(\theta)|_{\theta=0}=-p\frac{n-X}{1-p}-p^2\frac{n-X}{(1-p)^2}.$$

We can then solve the system of equations

$$\frac{\partial}{\partial\theta}LL(\theta)|_{\theta=0}=\frac{\partial}{\partial\theta}LL^a(\theta)|_{\theta=0}$$

along with

$$\frac{\partial^2}{\partial\theta^2}LL(\theta)|_{\theta=0}=\frac{\partial^2}{\partial\theta^2}LL^a(\theta)|_{\theta=0}$$

for both $n$ and $p$ to obtain the result of the proposition. □

*Proof of Proposition 6*
This proposition follows immediately from the same argument used to prove Proposition 2. □

## Acknowledgements

This research was supported in part by a grant from the Department of Veterans Affairs Health Services Research and Development Service, IIR 06-260. The authors thank the referees for many detailed suggestions that have improved the presentation of this article.

## References

1. Plagnol V, Tavare S. Approximate Bayesian Computation and MCMC 2002. In *Monte Carlo and Quasi-Monte Carlo Methods,* Niederreiter H (ed.) Springer: New York, 2004; 99–114.
2. Joyce P, Marjoram P. Approximately sufficient statistics and Bayesian computation. *Statistical Applications in Genetics and Molecular Biology* 2008; **7**:Article 26.
3. Le Cam L. Sufficiency and approximate sufficiency. *Annals of Mathematical Statistics* 1964; **35**:1419–1455.
4. Miller S, Bradlow E, Dayaratna K. Closed-form bayesian inferences for the logit model via polynomial expansions. *Quantitative Marketing and Economics* 2006; **4**:173–206.
5. Berlowitz DR, Young GJ, Hickey EC, Saliba D, Mittman BS, Czarnowski E, Simon B, Anderson JJ, Ash AS, Rubenstein LV. Quality improvement implementation in the nursing home. *Health Services Research* 2003; **38**:65–83.
6. Berlowitz DR, Ash AS, Brandeis GH, Brand HK, Halpern JL, Moskowitz MA. Rating long-term care facilities on pressure ulcer development: importance of case-mix adjustment. *Annals of Internal Medicine* 1996; **124**:557–563.
7. Porell F, Caro FG, Silva A, Monane M. A longitudinal analysis of nursing home outcomes. *Health Services Research* 1998; **33**:835–865.
8. Mukamel DB. Risk-adjusted outcome measures and quality of care in nursing homes. *Medical Care* 1997; **35**:367–385.
9. Zimmerman DR, Karon SL, Arling G, Clark BR, Collins T, Ross R, Sainfort F. Development and testing of nursing home quality indicators. *Health Care Financing Review* 1995; **16**:107–127.
10. Arling G, Karon SL, Sainfort F, Zimmerman DR, Ross R. Risk adjustment of nursing home quality indicators. *Gerontologist* 1997; **37**:757–766.
11. Braun BI. The effect of nursing home quality on patient outcome. *Journal of the American Geriatric Society* 1991; **39**:329–338.
12. Carter MW. Factors associated with ambulatory care-sensitive hospitalizations among nursing home residents. *Journal of Aging and Health* 2003; **15**:295–330.
13. Agency for Healthcare Research and Quality. Prevention Quality Indicators Overview, 2004. Available from: http://www.qualityindicators.ahrq.gov/pqi_overview.htm. [accessed November 4, 2008].

14. Rosen A, Wu J, Chang B, Berlowitz DB, Rakovski C, Ash AS, Moskowitz MA. Risk adjustment for measuring health outcomes: an application in VA long-term care. *American Journal of Medical Quality* 2001; **16**:118–127.
15. Berlowitz DR, Rosen AK, Wang F, Tsilimingras D, Tariot PN, Kader B, Mukamel DB. Purchasing or providing nursing home care: can quality of care data provide guidance. *Journal of the American Geriatric Society* 2005; **53**:603–608.
16. Pizer SD, Wang M, Comstock C. Preventable hospitalization as a measure of quality of care in nursing homes. *Health Care Finance and Economics*, VA Bedford, MA. 2003; *Working Paper #2003-01*.
17. Peköz E, Röllin A, Čekanavičius V, Shwartz M. A three-parameter binomial approximation. *Journal of Applied Probability* 2009; **46**(4):1073–1085.
18. Brown LD, Cai T, Zhou H. Nonparametric regression in exponential families. *Technical Report*, 2008. Available from: http://www-stat.wharton.upenn.edu/~tcai/paper/html/VST-Exp-Regression.html [accessed April 16, 2009].