12-17-2008

LEARNING FROM NEAR MISSES IN MEDICATION ERRORS: A BAYESIAN APPROACH

Jessica A. Myers
Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics, jamyers@jhsph.edu

Francesca Dominici
The Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics, fdominic@jhsph.edu

Laura Morlock
Department of Health, Policy and Management, Johns Hopkins University

Learning from Near Misses in Medication Errors: A Bayesian Approach

JESSICA A MYERS*
Department of Biostatistics,
Johns Hopkins University, Baltimore, MD 21205, USA
jamyers@jhsph.edu

FRANCESCA DOMINICI
Department of Biostatistics,
Johns Hopkins University, Baltimore, MD 21205, USA

LAURA MORLOCK
Department of Health Policy and Management,
Johns Hopkins University, Baltimore, MD 21205, USA

December 9, 2008

[1]To whom correspondence should be addressed.

Abstract

Medical errors originating in healthcare facilities are a significant source of preventable morbidity, mortality, and healthcare costs. Voluntary error report systems that collect information on the causes and contributing factors of medical errors regardless of the resulting harm may be useful for developing effective harm prevention strategies. Some patient safety experts question the utility of data from errors that did not lead to harm to the patient, also called near misses. A near miss (a.k.a. close call) is an unplanned event that did not result in injury to the patient. Only a fortunate break in the chain of events prevented injury. We use data from a large voluntary reporting system of 836,174 medication errors from 1999 to 2005 to provide evidence that the causes and contributing factors of errors that result in harm are similar to the causes and contributing factors of near misses. We develop Bayesian hierarchical models for estimating the log odds of selecting a given cause (or contributing factor) of error given harm has occurred and the log odds of selecting the same cause given that harm did not occur. The posterior distribution of the correlation between these two vectors of log-odds is used as a measure of the evidence supporting the use of data from near misses and their causes and contributing factors to prevent medical errors. In addition, we identify the causes and contributing factors that have the highest or lowest log-odds ratio of harm versus no harm. These causes and contributing factors should also be a focus in the design of prevention strategies. This paper provides important evidence on the utility of data from near misses, which constitute the vast majority of errors in our data.

KEYWORDS: Bayesian hierarchical models; Correlation; Medical error; Voluntary error reports

1 Introduction

Medical errors originating in healthcare facilities are a significant source of preventable morbidity, mortality, and healthcare costs (Brennan *and others* 1991, Leape *and others* 1991). The Institute of Medicine has estimated that 44,000 to 98,000 hospital deaths each year may be attributed to medical error (Kohn *and others* 1999). In the past, error prevention efforts have focused on examining the root causes of errors that resulted in serious harm, such as the death of the patient (Aspden *and others* 2003). This approach ignores all the information related to errors that did not result in harm, but had the potential to cause serious harm, referred to as near misses. Near misses occur much more frequently than harmful errors and, therefore, may be useful for informing prevention strategies, particularly for errors that occur rarely within a single healthcare facility.

Recently, several anonymous, voluntary reporting systems have been created to collect detailed information on the causes and contributing factors of medical errors (Wu *and others* 2002, Webb *and others* 1993). These systems combine reports of error across many healthcare facilities and include reports of both adverse events and near misses. The largest reporting system is MEDMARX®, a national, Internet-accessible database of medication error reports, created and maintained by the United States Pharmacopoeia (USP). An error report submitted to MEDMARX includes a detailed list of the types of medication errors that have occurred (e.g. Wrong administration technique, Wrong patient, Wrong time), the causes of the error (e.g. Calculation error, Communication, Handwriting illegible/unclear) and contributing factors of the error (e.g. Distractions, Fatigue, Poor lighting). In addition, an error report provides information on the degree of severity of the harm caused by the error.

Recently, the field of healthcare safety has begun adopting ideas developed by other

high-risk industries, including aviation, nuclear power technology, and petrochemical processing, for developing more effective safety strategies (Hudson 2003). These high risk industries have relied upon the "causal continuum hypothesis" (CCH), which states that *the causes and contributing factors of errors that lead to harm are similar to the causes and contributing factors of errors that do not lead to harm*. The CCH implies that, because near misses occur much more frequently than harmful errors, but are similar in their causes and contributing factors, data on near misses are useful for understanding error prevention, recovery from errors, and harm reduction.

In the transportation industry, Wright & Van der Schaaf (2004) studied the causes of 240 train incidents in the UK that were reported through a combination of mandatory and voluntary reporting systems. They first estimated $\hat{p}_1^c$ and $\hat{p}_0^c$, defined as the probabilities of citing cause c in adverse events (death or serious harm to individuals or train damage without harm to individuals) and near misses (no damage or injury), respectively. For each of the 21 possible causes, χ^2 tests were used to test for differences between $\hat{p}_0^c$ and $\hat{p}_1^c$. The authors argued that because of the lack of evidence for statistically significant differences between p_0^c and p_1^c, the CCH was validated.

We argue that conducting a series of tests of hypotheses of the type $H_0 : p_0^c = p_1^c$, one for each cause, to assess the validity of the CCH has limitations. First, lack of evidence of statistically significant differences in probabilities does not imply that the null hypothesis is true. Second, testing $p_0^c = p_1^c$ for each cause c independently does not provide a global quantitative measure of the evidence and its associated uncertainty for the validity of the CCH. Third, equality of probabilities ($p_0^c = p_1^c$ for all c) is not a necessary condition for the CCH. The CCH only requires that the causes and contributing factors are similar between adverse events and near misses. We interpret this similarity as saying that the causes that are most often identified in adverse events

are also the causes that are most often identified in near misses.

Since the 1930's, the CCH has been widely accepted in most areas of safety research (Heinrich 1931), despite a lack of evidence for its validity. Because of this lack of evidence, some experts in healthcare quality have questioned the utility of data on near misses (Layde *and others* 2002). Providing a global, quantitative measure of the evidence and its uncertainty supporting the CCH in healthcare is of primary importance. In addition, the majority of the statistical analyses of medical error reporting systems conducted so far have been descriptive with no formal statistical framework for inference (Hicks *and others* 2004, Hicks & Becker 2006, Morris & Morris 2000). Developing a statistical approach for quantifying the evidence for the CCH that can overcome the limitations identified above is very important. An appropriate analysis can inform future collection and analysis of data on medical errors and encourage the use of data on near misses, which constitute 98% of the data, to prevent harm.

In this paper, we use a voluntary error report database of 836,174 medication errors to provide evidence toward the causal continuum hypothesis. We develop a global statistical procedure for assessing the degree of similarity between causes and contributing factors of harmful errors and causes and contributing factors of near misses. More specifically, we define the correlation (ρ) between the log odds of reporting a specific cause of error given that harm has occurred and the log odds of reporting the same cause of error in a near miss. We use a Bayesian hierarchical model to estimate the posterior distribution of ρ as a measure of the evidence for the CCH. Under this approach, we can rely upon the correlation as a continuous measure of the evidence for the CCH and its uncertainty, rather than conducting a large number of hypothesis tests. Also, the hierarchical model estimates the correlation as a natural by-product, accounting for varying levels of precision in the log-odds estimates. We

compare our models to the analysis of Wright & Van der Schaaf. We also identify the causes and contributing factors that are most or least likely to be identified when the error resulted in harm than in a near miss.

In Section 2 we describe the large database of medication errors that is used in this analysis and illustrate some data characteristics. In Section 3 we present the methods used in this analysis. In Section 4 we present results from the various analyses, and in Section 5 we discuss the analyses and their impact on medication safety.

2 Data

We analyze data on medication errors collected through MEDMARX, one of the largest voluntary error report collection systems, containing more than 1.4 million reports submitted by more than 880 healthcare facilities. Each medication error reported to MEDMARX is categorized according to the "harm score," developed by the National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP n.d.). Table 1 summarizes the definitions of the harm score and the percentage of the reported errors that fall into each harm category. Not surprisingly, only 0.01% of reported errors resulted in death (category I) and less than 2% caused some harm to the patient (categories E,F,G,H,I).

Error reports also include information on the causes of the error (e.g. Documentation, Knowledge deficit, Workflow disruption) and contributing factors to the error (e.g. Distractions, Shift change, Patient transfer). Both the cause and contributing factor fields contain a list of choices, defined by USP in consultation with medication safety experts. More than one cause or contributing factor may be identified in the same error report. Tables 2 & 3 list all causes and contributing factors considered in this study and the observed frequencies of being cited in adverse events and in near

misses, respectively. In general, the data contain much more information on the causes of error than on the contributing factors of error, as 75.43% of the error reports did not identify any contributing factors. Throughout this paper, we will use the term causes to refer to either causes or contributing factors.

We analyze medication errors reported from 1999, the first complete year of data collection, through the end of 2005. During this time period, there were a total of 836,174 reports of error from 677 hospitals, including 16,052 (1.92%) reports involving some level of harm (Harm categories E-I). Figure 1 displays the number of reports submitted by the 100 hospitals with the largest total reporting volume separately for harm and no harm. In both plots the facilities are sorted by the total number of submitted reports, so that facility j in the left panel corresponds to facility j in the right panel. Each facility submitted at least one report to be included in the database, with a maximum of 18,145 reports submitted by a single facility. Note that facilities that report the largest number of near misses do not necessarily report the largest number of harmful errors. In addition, there is considerable variability across hospitals in the proportion of the reported errors that result in harm. This variability may be due to the different methods of error detection utilized, varying levels of emphasis within facilities on error detection and reporting, and the variety of facility types included in the database, which may have inherent differences in the rate of harm among errors.

Figure 2 illustrates the proportion of facilities that submit at least n reports of adverse events and the proportion of facilities that submit at least n reports of near misses, where n varies from zero up to 25. The curve for adverse events drops off quickly as n increases from zero, while the curve for near misses remains above 90% even at $n = 25$. Approximately 50% of facilities in the database submitted less than seven reports of harmful error, which equates with less than one harmful

report per year for facilities enrolled in the program for the entire time period under consideration. This lack of reporting of adverse events underscores the need to learn from near-miss data. Adverse events occur rarely compared to near misses, and those adverse events that do occur may be more likely to be concealed or underreported by staff members that fear retribution. With so little information available on the errors that result in patient harm, hospitals may find it difficult to draw conclusions and design interventions to reduce error. Therefore, data from near misses may provide an important additional source of information for hospitals that want to learn about how to reduce error and harm.

3 Methods

In this section, we introduce a Bayesian hierarchical model (BHM) to quantify the strength of the evidence and its uncertainty that the causes and contributing factors of adverse events are similar to the causes and contributing factors of near misses. Let X_i^c be the indicator of reporting cause c on event report i. Let z_i be the indicator of harm on report i ($z_i = 1$ when the error is in harmscore categories E-I and $z_i = 0$ when the error is in harmscore categories B-D). Our goal is to develop a global procedure for providing evidence toward the hypothesis that the causes that are most often identified in harmful errors are also most often identified in near misses. To accomplish this

objective, we introduce the following hierarchical model:

$$\text{Stage I:} \qquad X_i^c | z_i, \beta_0^c, \beta_1^c \sim Binom\left(\frac{exp\{\beta_0^c + (\beta_1^c - \beta_0^c) z_i\}}{1 + exp\{\beta_0^c + (\beta_1^c - \beta_0^c) z_i\}}\right); \tag{3.1}$$

$$i \in \{1, \ldots, N\}, c \in \{1, \ldots, C\} \ \ independent$$

$$\text{Stage II:} \qquad \begin{pmatrix} \beta_0^c \\ \beta_1^c \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mu_0 \\ \mu_1 \end{pmatrix}, \Sigma = \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{bmatrix}\right) \tag{3.2}$$

$$\text{Stage III:} \qquad \begin{pmatrix} \mu_0 \\ \mu_1 \end{pmatrix} \sim N_2\left(\begin{pmatrix} a_0 \\ a_1 \end{pmatrix}, \begin{bmatrix} A_0 & 0 \\ 0 & A_1 \end{bmatrix}\right) \tag{3.3}$$

$$\Sigma \sim IW_2(\omega, D)$$

In this model, $\boldsymbol{\beta_0} = (\beta_0^1, \ldots, \beta_0^C)'$ and $\boldsymbol{\beta_1} = (\beta_1^1, \ldots, \beta_1^C)'$ are vectors of log odds of citing each cause in reports of near misses and reports of adverse events, respectively. The parameters μ_0 and μ_1 are nuisance parameters denoting the average of β_0^c and β_1^c across causes.

The parameter of interest is $\rho = \sigma_{01}/(\sigma_0\sigma_1)$, denoting the correlation between β_0^c and β_1^c across c. If $\rho = 1$, then there is a perfect linear relation between β_0^c and β_1^c. In addition, because this relation is positive and monotonic, it would imply that for all causes, the relative rank of the log odds of identifying a given cause in an adverse event is identical to the relative rank of the log odds of identifying the same cause in near misses. Therefore, a large, positive correlation indicates that the causes and contributing factors most frequently involved in adverse events are among the causes and contributing factors most frequently involved in near misses. We argue that the posterior distribution $P(\rho|data)$ can be used as a measure of evidence in favor of the CCH.

We choose noninformative priors. In particular, in the hyperprior for Σ, we set

$D = I_2$ and $\omega = 3$. These values yield a marginal prior for ρ that is uniform on [-1,1] and a sufficiently noninformative prior for the variances, σ_0^2 and σ_1^2 (Barnard *and others* 2000). In the hyperprior for $\boldsymbol{\mu}$, $a_0 = a_1 = logit(1/C)$, and $A_0 = A_1 = 1000$. These values are intended to provide a prior that is approximately flat over the interval of reasonable values for μ_0 and μ_1. All posterior distributions were estimated via Markov Chain Monte Carlo (MCMC) sampling, as described in the Appendix.

Fitting a model using MCMC with many parameters (139) and a very large data set (836,174 $\times$ 67), as in this example, is very computationally intensive. Generating posterior samples of size 10,000 from the model described above took approximately 7 days. To simplify computation, as a sensitivity analysis, we used the normal approximation of the likelihood function and defined the following hierarchical model:

$$\text{Stage I:} \quad \begin{pmatrix} \hat{\beta}_0^c \\ \hat{\beta}_1^c \end{pmatrix} \sim N_2 \left(\begin{pmatrix} \beta_0^c \\ \beta_1^c \end{pmatrix}, \hat{V}^c \right) \tag{3.4}$$

$$\text{Stage II:} \quad \begin{pmatrix} \beta_0^c \\ \beta_1^c \end{pmatrix} \sim N_2 \left(\begin{pmatrix} \mu_0 \\ \mu_1 \end{pmatrix}, \Sigma \right)$$

where $\hat{V}^c$ is estimated from the data and assumed known. All other parameters are the same as in the previous model.

We used Two Level Normal independent sampling estimation (TLNise) to produce Bayesian posterior samples of model parameters, including ρ (Everson & Morris 2000). We compared the results from this model fit via TLNise to the previous model fit with MCMC. The TLNise sampling method specifies a uniform prior on Σ and noninformative prior on μ. It is much more efficient than MCMC, reducing computation time to approximately one minute.

The model proposed above allows for estimation of the heterogeneity matrix and

correlation among the true log odds as a natural by-product of the model. In addition, under the BHMs proposed here, we can estimate $P(\rho|data)$ accounting for the differing levels of precision in the log odds estimates. Finally, combining information across causes yields estimates of β_0^c and β_1^c with lower mean-square error, especially for the least frequently cited, and thus least precisely estimated, causes. Therefore, in addition to providing evidence for the CCH, our model provides improved estimates of β_0^c and β_1^c for identifying causes most likely to be aberrant with respect to the CCH.

We also compare the hierarchical models introduced here to a series of χ^2 tests, as was used in the study of transportation safety (Wright & Van der Schaaf 2004). That analysis focused on testing for statistically significant differences in the probability of citing each cause between events involving injury or damage and near misses. More specifically, we applied this approach to our data, using a χ^2 test for each cause $c \in \{1, \ldots, C\}$ to test the null hypothesis $H_0 : p_1^c = p_0^c$, where p_1^c is the probability of citing cause c in a report of an adverse event, and p_0^c is the probability of citing cause c in a report of a near miss. In the analysis of causes, $C = 67$, and in the analysis of contributing factors, $C = 20$. A conservative Bonferroni correction was used to adjust for the multiple tests. Causes with p-values less than $.05/C$ were identified as having a statistically significant difference between p_0^c and p_1^c.

In addition to providing evidence on the CCH, it is important to identify causes that have the highest or lowest log odds-ratios of being cited in adverse events compared to near misses. Bayes rule indicates that this quantity is equal to the log-odds ratio of harm occurring in errors citing the cause versus errors that did not cite the cause, as shown by the relation

$$\frac{P(X_i^c = 1|z_i = 1)/P(X_i^c = 0|z_i = 1)}{P(X_i^c = 1|z_i = 0)/P(X_i^c = 0|z_i = 0)} = \frac{P(z_i = 1|X_i =^c 1)/P(z_i = 0|X_i^c = 1)}{P(z_i = 1|X_i^c = 0)/P(z_i = 0|X_i^c = 0)}.$$

Therefore, causes with estimated log-odds ratios that are very high or very low may be of interest. We define the standardized log-odds difference as

$$\theta^c = \frac{\beta_1^c - \beta_0^c - (\mu_1 - \mu_0)}{\sqrt{\sigma_0^2 + \sigma_1^2 - 2\tau}}. \quad (3.5)$$

A priori, the θ^c are independent and identically distributed as $N(0,1)$, and $\theta^c = 0$ implies that the difference in log-odds estimates for cause c is equal to their average across causes, that is, $\beta_1^c - \beta_0^c = \mu_1 - \mu_0$. Therefore, the posterior distribution of θ^c can be used to flag causes with a log-odds difference that deviates from the average in excess of what would be expected, given the unexplained variability across causes. A cause or contributing factor is labeled as aberrant if $P(|\theta^c| > t|data) > K$, where t is a quantile of the standard normal distribution. We chose to use $t = 1.96$ and $K = .5$ so that we identify causes and contributing factors with a posterior median in the 2.5% tail area of the prior distribution.

The proposed approach for identifying causes that are aberrant with respect to the CCH has two main advantages. First, we have found in exploratory analyses that on average, a larger number of causes are cited simultaneously on reports of errors resulting in harm than on reports of near misses. By removing the overall difference in means $(\mu_1 - \mu_0)$ we can account for this reporting differential. Second, by dividing by the variance of $\beta_1^c - \beta_0^c$ across causes $(\sigma_0^2 + \sigma_1^2 - 2\tau)$, the standardized log-odds difference accounts for the heterogeneity across causes and, therefore, identifies causes with a difference in the log-odds of being cited between adverse events and near misses that is large enough to be considered abnormal compared to other causes.

4 Results

Table 2 summarizes the estimated probabilities, $\hat{p}_0^c$ and $\hat{p}_1^c$, of reporting each cause c among near misses and adverse events, respectively. For each cause we give the p-value from a χ^2 test of the null hypothesis $H_0 : p_0^c = p_1^c$. Causes marked with a "*" are those where we reject H_0 at a Bonferroni-corrected 0.05 level. In addition, we report the relative ranks of $\hat{p}_0^c$ and $\hat{p}_1^c$. Table 3 summarizes the same information for contributing factors.

Tables 2 and 3 show that, even when using the very conservative Bonferroni correction, in 34 out of 67 causes and 15 out of 20 contributing factors we reject $H_0 : p_0^c = p_1^c$. However, the empirical correlation between $\hat{p}_0^c$ and $\hat{p}_1^c$ is 0.941 (95% CI: 0.906, 0.964) for causes and 0.939 (95% CI: 0.85, 0.976) for contributing factors. Therefore, despite many causes and contributing factors with probabilities of being cited that are statistically significantly different between adverse events and near misses, $\hat{p}_0^c$ and $\hat{p}_1^c$ are highly correlated.

The left panel of Figure 3 displays the maximum likelihood estimates (top) and Bayesian posterior means (bottom) of the log-odds of citing each cause in a report of an adverse event plotted against the log-odds of citing the same cause in a report of a near miss. Causes that were identified as having the highest or lowest log odds-ratio of being cited in harm versus no harm by the standardized log-odds difference are plotted with a solid symbol. The diameter of the plotting symbols are scaled inversely according to the maximum variance of the two estimates of log-odds for each point. The right panel of Figure 3 displays the same information for contributing factors. Because of the large sample size, the log-odds of reporting a cause or contributing factor is estimated with high precision. Therefore, there is very little shrinkage in the Bayesian estimates, except for the least frequently reported causes. Overall, there

appears to be a strong positive linear relation between the two vectors of log-odds for both causes and contributing factors. This relation is reflected in the estimated posterior mean of ρ from the Bayesian model using MCMC, 0.943 (95% PI: 0.911, 0.966) for causes and 0.922 (95% PI: 0.837, 0.97) for contributing factors.

Figure 4 compares the estimated posterior of ρ from the MCMC sample to that of the TLNise sample, as well as the empirical estimate of ρ calculated from the log-odds MLEs. This figure shows that the posterior distribution of ρ estimated from TLNise is very similar to the results obtained from the MCMC sample. The posterior mean of ρ estimated using TLNise is 0.946 (95% PI: 0.911, 0.969) for causes and 0.942 (95% PI: 0.864, 0.982) for contributing factors.

Figure 5 displays boxplots of the posterior distributions of θ^c for each cause and contributing factor. Dashed lines mark the chosen cutoff, $t = 1.96$. The standardized log-odds difference criterion identifies the causes and contributing factors with the largest Bayesian estimates of the log-odds ratio of being cited in adverse events versus near misses. More specifically, we found that cause 47 (pump failure/malfunction) has an estimated log-odds ratio of harm of 1.461 (95% PI: 1.285, 1.626) and cause 32 (pump, improper use) has an estimated log-odds ratio of harm of 1.555 (95% PI: 1.477, 1.633). These two causes were cited more frequently in reports of adverse events than in reports of near misses. Cause 17 (computerized prescriber order entry) has an estimated log-odds ratio of harm of -1.945 (95% PI: -2.223, -1.682); cause 14 (workflow disruption) has an estimated log-odds ratio of harm of -1.478 (95% PI: -1.682, -1.286); and cause 12 (abbreviations) has an estimated log-odds ratio of harm of -1.305 (95% PI: -1.488, -1.127). These three causes were identified as being cited more frequently in reports of near misses than in reports of adverse events. No contributing factors were identified using the given threshholds.

Note that the causes identified in Figure 3 are not those with the smallest p-values from the tests reported in Tables 2 and 3, so simply reducing the α-level of the χ^2 tests would not produce the same results. Interestingly, the causes and contributing factors identified by the posterior for θ^c happen to be those with the largest magnitude differences in relative ranks. These causes tend to have log odds of being cited that are near the center of the distribution in both adverse events and near misses, where the majority of the causes lie. Because causes are clustered more tightly in this portion of the distribution, a difference between $\hat{p}_0^c$ and $\hat{p}_1^c$ that is only moderately large can lead to a very large difference in ranks $(r_1^c - r_0^c)$. For example, cause 17 identified above has $\hat{p}_0^{17} = 0.00204$ and $\hat{p}_1^{17} = 0.00027$. However, $r_0^{17} = 17$ $r_1^{17} = 45$. This large difference in relative ranks would be important to investigators that want to prioritize causes according to frequency. The standardized log-odds difference identifies causes like these because they have at least moderately large differences in log-odds with sufficiently small standard errors that prevent shrinkage toward equality.

5 Discussion

The proliferation of large and complex data sets providing information on the causes and contributing factors of medical errors has many implications for the future of health care research, practice, and policy. Careful analyses of these data could yield great insight into the prevention of error and the reduction of harm due to error. In order to draw reliable inferences, however, we must carefully consider the use of data from near-miss reports and its ability to inform investigators about the causes and contributing factors of errors resulting in patient harm. This analysis contributes substantially to the field of medication safety by providing evidence in support of the use of near-miss data for learning about the causes and contributing factors of medication errors.

Utilizing near-miss data in future analyses of medication error will increase the ability of investigators to make inference about the causes and contributing factors of error and may allow for the identification of sources of error within healthcare practice before they are able to cause significant harm.

Evaluating the evidence for the CCH in medication error involves comparing the presence/absence of the various causes and contributing factors between two samples of reports: adverse events and near misses. The transportation industry performed this comparison by testing for a statistically significant difference in the probability of citing each cause between train accidents resulting in injury or damage and accidents with no damage or injury. A similar approach is used routinely in microarray differential expression analysis (Lonnstedt & Speed 2002, Ideker *and others* 2000, Roberts *and others* 2000), where investigators wish to compare the expression of various genes between two samples of cells (e.g. diseased and non-diseased).

We argue that testing for difference in the probability of identifying a given cause between adverse events and near misses is not optimal for providing evidence for the CCH. First, a lack of evidence for significant differences in probabilities does not provide evidence for equality of probabilities. Second, testing for difference in the probabilities of each cause independently does not provide an overall assessment of the similarity between the probability of citing a cause or contributing factor in an adverse event and the probability of citing the same cause or contributing factor in a near miss. Evaluating the evidence for the causal continuum hypothesis requires quantifying this similarity. Third, equality of probabilities is an unnecessarily strong condition for assessing the validity of the CCH. Therefore, rejecting the null hypothesis that the probabilities are equal for one or more causes does not necessarily provide evidence against the CCH.

In this analysis, we used a BHM to measure the agreement of causal profiles between reports of errors that resulted in harm and reports of errors that did not result in patient harm. Our approach assumes a more appropriate statistical definition of the CCH based on posterior inference of a single parameter, ρ, which is estimated accounting for varying levels of precision in the first level parameters. Because we do not focus on strict acceptance or rejection of the CCH, but instead aim to measure the strength of evidence for the hypothesis, our model does not require adjustment for multiple comparisons. Furthermore, the multiple parallel measurements of distinct but related outcomes lends itself naturally to a hierarchical modeling approach, such as the one we have used. Extensions of this model to include other levels of clustering, such as clustering at the facility level, are possible as well. We found that, despite many causes and contributing factors with unequal probabilities of being cited between adverse events and near misses, our BHM estimated a correlation near one with very high posterior probability. This high correlation indicates strong evidence for the CCH in medication errors reported through the MEDMARX system.

In addition, we used a simple criterion for identifying the causes and contributing factors with the highest and lowest log odds-ratios of being cited in adverse events versus near misses. The causes with the highest log odds-ratios occur more frequently in adverse events, indicating they may warrant special attention from investigators as a priority for intervention. The two causes identified as having unusually high log odds-ratios are both related to pumps used in the administration of medication, where errors have little chance of being recovered before they are able to cause harm. The causes with the lowest log odds-ratios occur more frequently in near misses, indicating that there may be effective barriers in place that prevent them from leading to harm. These causes may be useful for the study of recovery from error and harm

reduction. For example, computerized prescriber order entry (CPOE) was implemented with the intention of reducing opportunities for error in medication use (Schiff & Rucker 1998). However, incorrect or incomplete entries in a CPOE system may cause errors as well, including approximately 2% of the near misses in this study. A much smaller proportion of adverse events were caused by CPOE (0.27%), likely due to the error checking software built in to most COPE systems. The small log odds-ratio observed in this study for errors caused by CPOE indicates that the CPOE systems are actually functioning as an effective barriers to stop errors from causing patient harm once they have occurred.

One limitation of our model is that the assumption of independence between causes implied by our model is likely to be unrealistic. The number of causes cited simultaneously on a report ranged from 0 up to 35 out of 67 distinct possible causes with 297,683 (35.60%) error reports citing more than one cause. Between 0 and 15 contributing factors were cited simultaneously out of 21 possible with 47,848 (5.72%) reports citing more than one contributing factor simultaneously. Certain groups of causes or contributing factors likely tend to be cited together when reporting an error. Although we attempted several methods of modeling the correlation among causes, at least on a pairwise basis, none were found to be reasonably efficient due to the extremely large size of the data. The cluster size in this application is too large for the methods available for modeling complex correlation structures. However, the presence of correlation among causes and contributing factors is unlikely to reverse the conclusions reached in this analysis because modeling the correlation would primarily influence the efficiency of the estimates of log-odds, not the estimates themselves.

We also note that the types, locations, and circumstances of events included in the MEDMARX database are widely varying. Certain causes or contributing factors

may be associated with many different types of events, while others may participate in only one type of event. Some causes may be relevant only in a portion of the reporting facilities, for example those with on-site pharmacies, or in particular locations within a facility, for example, a surgical unit. Considering the huge amount of variability that is represented in the standardized report, it is not surprising that we observe such high heterogeneity between reports within a facility, and between facilities.

Finally, the statistical methods developed in this research could be applied to the expanding number and variety of patient safety databases and used to address other questions of scientific importance. As these databases grow in size and complexity, more sophisticated statistical methods are necessary to efficiently extract information so that healthcare organizations may learn from their own and one another's errors and quickly implement change to reduce patient harm.

APPENDIX: Details of model estimation

Model (1) from section 3.2 was estimated by an MCMC Gibbs sampler with a Metropolis-Hastings step for the first level parameters, $\boldsymbol{\beta_0}$ and $\boldsymbol{\beta_1}$. The initial values for these parameters in both the MCMC for model (1) and the TLNise for model (3) were taken from MLEs estimated by the models in (2). The initial values for μ_0 and μ_1 were the mean of the MLEs for $\boldsymbol{\beta_0}$ and $\boldsymbol{\beta_1}$, respectively, and the initial value for Σ is the sample covariance matrix of the MLEs.

The MCMC was run for 10,000 iterations and checked for convergence. Mixing occurred very fast, and there was little autocorrelation even at a small lag. For ρ in the analysis of causes, the estimated autocorrelation at lag 1 was 0.086, and in the analysis of contributing factors, the estimated autocorrelation at lag 1 was 0.075. Therefore, chains were not thinned, and all results reported in the body of the paper

use the last 5000 samples from the posterior with the first 5000 discarded as burn-in.

References

Aspden, P., Corrigan, J. M., Wolcott, J. & Erickson, S. M., eds (2003), *Patient Safety: Achieving a New Standard for Care*, The National Academy Press: Washington DC.

Barnard, J., McCulloch, R. & Meng, X. (2000), 'Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage', *Statistica Sinica* **10**(4), 1281–1312.

Brennan, T., Leape, L., Laird, N., Hebert, L., Localio, A., Lawthers, A., Newhouse, J., Weiler, P. & Hiatt, H. (1991), 'Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard Medical Practice Study I.', *New England Journal of Medicine* **324**(6), 370–6.

Everson, P. J. & Morris, C. N. (2000), 'Inference for multivariate normal hierarchical models', *Journal of the Royal Statistical Society: Series B* **62**, 399–412.

Heinrich, H. (1931), *Industrial Accident Prevention*, McGraw Hill.

Hicks, R. & Becker, S. (2006), 'An overview of intravenous-related medication administration errors as reported to MEDMARX®, a national medication error-reporting program', *Journal of Infusion Nursing* **29**(1), 20–27.

Hicks, R. W., Becker, S. C., Krenzischeck, D. & Beyea, S. C. (2004), 'Medication errors in the PACU: a secondary analysis of MEDMARX findings', *Journal of PeriAnesthesia Nursing* **19**(1), 18–28.

Hudson, P. (2003), 'Applying the lessons of high risk industries to health care', *Quality and Safety in Health Care* **12**(90001), i7–i12.

Ideker, T., Thorsson, V., Siegel, A. & Hood, L. (2000), 'Testing for differentially-expressed genes by maximum-likelihood analysis of microarray data', *Journal of Computational Biology* **7**(6), 805–817.

Kohn, L., Corrigan, J. & Donaldson, M., eds (1999), *To Err is Human: Building a Safer Health System*, National Academy Press: Washington, DC.

Layde, P., Maas, L., Teret, S., Brasel, K., Kuhn, E., Mercy, J. & Hargarten, S. (2002), 'Patient Safety Efforts Should Focus on Medical Injuries', *JAMA* **287**(15), 1993–1997.

Leape, L., Brennan, T., Laird, N., Lawthers, A., Localio, A., Barnes, B., Hebert, L., Newhouse, J., Weiler, P. & Hiatt, H. (1991), 'The nature of adverse events in hospitalized patients. Results of the Harvard Medical Practice Study II', *New England Journal of Medicine* **324**(6), 377–384.

Lonnstedt, I. & Speed, T. (2002), 'Replicated microarray data', *Statistica Sinica* **12**(1), 31–46.

Morris, G. & Morris, R. (2000), 'Anaesthesia and fatigue: an analysis of the first 10 years of the Australian Incident Monitoring Study 1987-1997', *Anaesthesia Intensive Care* **28**(3), 300–4.

NCC MERP (n.d.), 'Medication error category index'. [Online; accessed 08-Dec-2008]. `http://www.nccmerp.org/medErrorCatIndex.html`

Roberts, C., Nelson, B., Marton, M., Stoughton, R., Meyer, M., Bennett, H., He, Y., Dai, H., Walker, W. & Hughes, T. (2000), 'Signaling and circuitry of multiple

MAPK pathways revealed by a matrix of global gene expression profiles', *Science* **287**(5454), 873.

Schiff, G. & Rucker, T. (1998), 'Computerized prescribing building the electronic infrastructure for better medication usage', *Journal of the American Medical Association* **279**(13), 1024–1029.

Webb, R., Currie, M., Morgan, C., Williamson, J., Mackay, P., Russell, W. & Runciman, W. (1993), 'The Australian Incident Monitoring Study: an analysis of 2000 incident reports.', *Anaesthesia Intensive Care* **21**(5), 520–8.

Wright, L. & Van der Schaaf, T. (2004), 'Accident versus near miss causation: a critical review of the literature, an empirical test in the UK railway domain, and their implications for other sectors', *Journal of Hazardous Materials* **111**(1), 105–110.

Wu, A., Pronovost, P. & Morlock, L. (2002), 'ICU incident reporting systems', *Journal of Critical Care* **17**(2), 86–94.

Table 1: Number and percent of error reports in each harmscore category, defined by The National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP). We do not include reports in category A because no error has occurred. Near misses include reports in the categories B, C, and D. Adverse events include reports in the categories E, F, G, H, and I.

Category	Description	Number	Percent
No Error			
A	Circumstances or events that have the capacity to cause error.	0	0.00
Error, no harm			
B	An error occurred but the error did not reach the patient.	371083	44.38
C	An error occurred that reached the patient but did not cause patient harm.	384553	45.99
D	An error occurred that reached the patient and required monitoring to confirm that it resulted in no harm to the patient and/or required intervention to preclude harm.	64486	7.71
Error, harm			
E	An error occurred that may have contributed to or resulted in temporary harm to the patient and required intervention.	12847	1.54
F	An error occurred that may have contributed to or resulted in temporary harm to the patient and required initial or prolonged hospitalization.	2672	0.32
G	An error occurred that may have contributed to or resulted in permanent patient harm.	139	0.02
H	An error occurred that required intervention necessary to sustain life.	285	0.03
Error, death			
I	An error occurred that may have contributed to or resulted in the patient's death.	109	0.01
Total		836174	100

Table 2: List of causes that can be cited when a medication error has occurred. Definitions can be found at www.biostat.jhsph.edu/ jamyers/research. More than one cause can be checked for each error. For each cause, $\hat{p}_0^c$ and $\hat{p}_1^c$ are the observed proportions of citing cause c among adverse events and near misses, respectively. The p-value for testing $H_0 : p_0^c = p_1^c$ versus $H_a : p_0^c \neq p_1^c$ is given, and a * indicates p<.05/67. r_0^c and r_1^c are the ranks of $\hat{p}_0^c$ and $\hat{p}_1^c$, respectively.

No.	Cause	$100\times\hat{p}_0^c$	$100\times\hat{p}_1^c$	p-value	r_0^c	r_1^c	
1	Performance (human) deficit	38.64	44.45	1.5×10^{-50}	1	1	*
2	Procedure/protocol not followed	17.12	25.69	1.0×10^{-177}	2	2	*
3	Transcription inaccurate/omitted	12.97	10.36	1.8×10^{-22}	3	5	*
4	Computer entry	11.32	7.31	5.9×10^{-57}	4	7	*
5	Documentation	11.30	8.88	8.8×10^{-22}	5	6	*
6	Communication	9.93	15.90	2.5×10^{-137}	6	4	*
7	Knowledge deficit	9.14	16.00	1.0×10^{-193}	7	3	*
8	Written order	5.06	4.52	2.2×10^{-03}	8	11	
9	System safeguard(s)	3.10	6.49	1.5×10^{-129}	9	9	*
10	Monitoring inadequate/lacking	2.93	7.26	2.0×10^{-221}	10	8	*
11	Drug distribution system	2.80	3.31	9.7×10^{-05}	11	13	*
12	Abbreviations	2.73	0.74	1.7×10^{-53}	12	31	*
13	Handwriting illegible/unclear	2.49	2.04	3.8×10^{-04}	13	17	*
14	Workflow disruption	2.41	0.54	3.5×10^{-53}	14	38	*
15	Dispensing device involved	2.28	3.10	7.7×10^{-12}	15	15	*
16	Calculation error	2.10	4.75	7.3×10^{-116}	16	10	*
17	Computerized prescriber order entry	2.04	0.27	2.6×10^{-56}	17	45	*
18	Dosage form confusion	1.93	1.89	7.4×10^{-01}	18	18	
19	Fax/scanner involved	1.67	1.19	3.4×10^{-06}	19	22	*
20	Verbal order	1.63	2.39	5.1×10^{-14}	20	16	*
21	Contraindicated, drug allergy	1.46	3.15	1.3×10^{-67}	21	14	*
22	Brand/generic names look alike	1.35	0.95	1.2×10^{-05}	22	26	*
23	Preprinted medication order form	1.22	1.35	1.7×10^{-01}	23	20	
24	Incorrect medication activation	1.12	1.36	4.8×10^{-03}	24	19	
25	Brand names look alike	1.04	0.88	6.1×10^{-02}	25	30	
26	Generic names look alike	1.00	0.67	3.7×10^{-05}	26	33	*
27	Labeling (your facility's)	0.98	0.92	4.4×10^{-01}	27	27	
28	Similar packaging/labeling	0.96	1.34	1.5×10^{-06}	28	21	*
29	Patient identification failure	0.94	0.51	2.5×10^{-08}	29	40	*
30	Computer software	0.93	1.06	1.0×10^{-01}	30	23	
31	Brand/generic names sound alike	0.92	0.70	4.6×10^{-03}	31	32	
32	Pump, improper use	0.90	4.16	$0.0\times 10^{+00}$	32	12	*
33	Brand names sound alike	0.76	0.65	1.3×10^{-01}	33	34	
34	Packaging/container design	0.76	1.05	2.4×10^{-05}	34	24	*
35	Generic names sound alike	0.73	0.60	5.5×10^{-02}	35	36	

Table 2: *(continued)*

No.	Cause	$100\times\hat{p}_0^c$	$100\times\hat{p}_1^c$	p-value	r_0^c	r_1^c	
36	Information management system	0.60	0.90	9.3×10^{-07}	36	29	*
37	Storage proximity	0.49	0.39	8.0×10^{-02}	37	44	
38	Label (your facility's) design	0.46	0.56	7.1×10^{-02}	38	37	
39	Contraindicated, drug/drug	0.45	0.45	$1.0\times 10^{+00}$	39	41	
40	Decimal point	0.42	1.02	3.9×10^{-30}	40	25	*
41	Label (manufacturer's) design	0.36	0.39	4.8×10^{-01}	41	43	
42	Non-formulary drug	0.29	0.17	4.7×10^{-03}	42	49	
43	Reconciliation-transition	0.23	0.17	1.5×10^{-01}	43	50	
44	MAR variance	0.22	0.16	1.2×10^{-01}	44	52	
45	Similar products	0.22	0.07	9.3×10^{-05}	45	59	*
46	Diluent wrong	0.21	0.43	1.9×10^{-09}	46	42	*
47	Pump failure/malfunction	0.21	0.91	3.6×10^{-79}	47	28	*
48	Equipment design	0.20	0.62	1.5×10^{-29}	48	35	*
49	Leading zero missing	0.19	0.09	7.9×10^{-03}	49	56	
50	Contraindicated in disease	0.19	0.52	2.6×10^{-21}	50	39	*
51	Drug shortage	0.14	0.19	1.2×10^{-01}	51	48	
52	Prefix/suffix misinterpreted	0.11	0.17	5.0×10^{-02}	52	51	
53	Reconciliation-admission	0.11	0.15	1.9×10^{-01}	53	53	
54	Repackaging by your facility	0.11	0.03	4.2×10^{-03}	54	62	
55	Blanket orders	0.09	0.07	4.0×10^{-01}	55	58	
56	Trailing/terminal zero	0.09	0.09	9.5×10^{-01}	56	57	
57	Reference material	0.08	0.20	9.6×10^{-08}	57	47	*
58	Equipment failure/malfunction	0.08	0.11	1.3×10^{-01}	58	55	
59	Non-metric units used	0.08	0.11	2.1×10^{-01}	59	54	
60	Measuring device	0.07	0.20	2.1×10^{-08}	60	46	*
61	Reconciliation-discharge	0.05	0.04	8.8×10^{-01}	61	60	
62	Override	0.03	0.02	4.1×10^{-01}	62	63	
63	Contraindicated, drug/food	0.02	0.03	7.6×10^{-01}	63	61	
64	Repackaging by other facility	0.02	0.01	3.6×10^{-01}	64	64	
65	Weight	0.01	0.01	9.2×10^{-01}	65	67	
66	Contraindicated in pregnancy	0.01	0.01	7.0×10^{-01}	66	65	
67	Unlabeled syringe/container	0.00	0.01	6.9×10^{-01}	67	66	

Table 3: List of contributing factors that can be cited when a medication error has occurred. Definitions can be found at www.biostat.jhsph.edu/ jamyers/research. More than one contributing factor can be checked for each error. For each cause, $\hat{p}_0^c$ and $\hat{p}_1^c$ are the observed proportions of citing contributing factor c among adverse events and near misses, respectively. The p-value for testing $H_0 : p_0^c = p_1^c$ versus $H_a : p_0^c \neq p_1^c$ is given, and a * indicates p<.05/20. r_0^c and r_1^c are the ranks of $\hat{p}_0^c$ and $\hat{p}_1^c$, respectively.

No.	Contributing Factor	$100\times\hat{p}_0^c$	$100\times\hat{p}_1^c$	p-value	r_0^c	r_1^c	
1	Distractions	10.95	9.91	3.0×10^{-05}	1	1	*
2	Workload increase	6.00	4.77	6.4×10^{-11}	2	3	*
3	Staff, inexperienced	4.19	8.07	1.3×10^{-127}	3	2	*
4	Staffing, insufficient	2.32	2.26	6.6×10^{-01}	4	4	
5	Shift change	1.38	2.25	3.9×10^{-20}	5	5	*
6	Cross coverage	1.25	1.40	9.7×10^{-02}	6	9	
7	Staff, agency/temporary	0.95	2.19	3.7×10^{-56}	7	6	*
8	Emergency situation	0.91	2.07	5.4×10^{-52}	8	7	*
9	No 24-hour pharmacy	0.77	0.99	2.2×10^{-03}	9	12	*
10	Patient transfer	0.75	1.51	2.7×10^{-28}	10	8	*
11	Staff, floating	0.60	1.25	2.4×10^{-25}	11	10	*
12	No access to patient information	0.60	1.13	2.7×10^{-17}	12	11	*
13	Imprint, identification failure	0.40	0.18	1.9×10^{-05}	13	17	*
14	Staffing, alternative hours	0.32	0.54	1.8×10^{-06}	14	13	*
15	Code situation	0.26	0.53	3.4×10^{-11}	15	14	*
16	Computer system/network down	0.23	0.23	9.4×10^{-01}	16	16	
17	Patient names similar/same	0.19	0.05	8.9×10^{-05}	17	18	*
18	Poor lighting	0.13	0.29	9.2×10^{-08}	18	15	*
19	Range orders	0.05	0.04	6.3×10^{-01}	19	19	
20	Fatigue	0.03	0.02	4.9×10^{-01}	20	20	

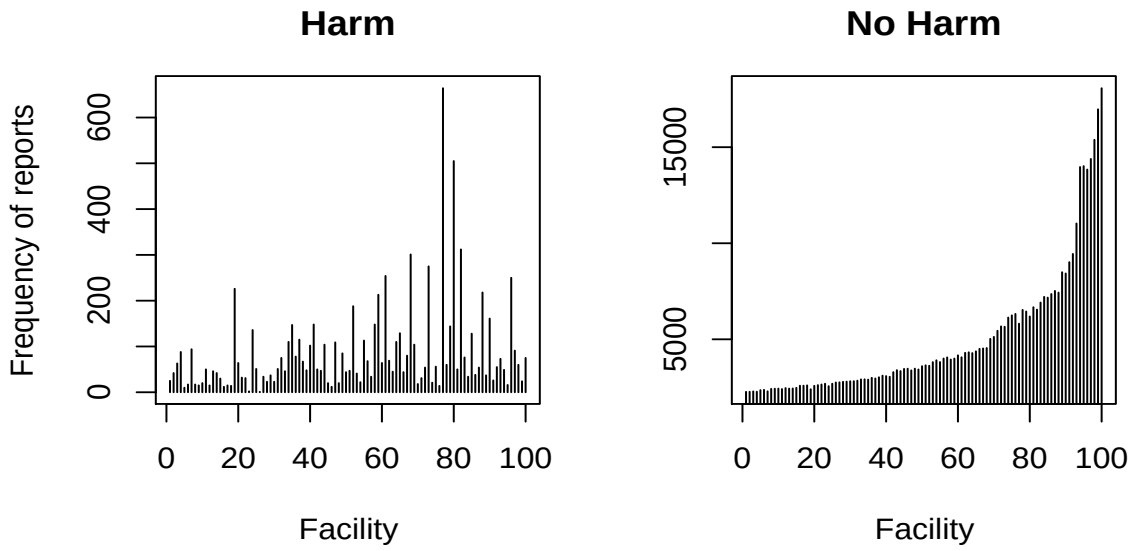

Figure 1: Number of reports submitted from each of the 100 facilities with the highest total number of reports, plotted separately for the number of harmful reports and the number of nonharmful reports. The facilities in each plot are sorted on total number of reports (harmful + nonharmful).

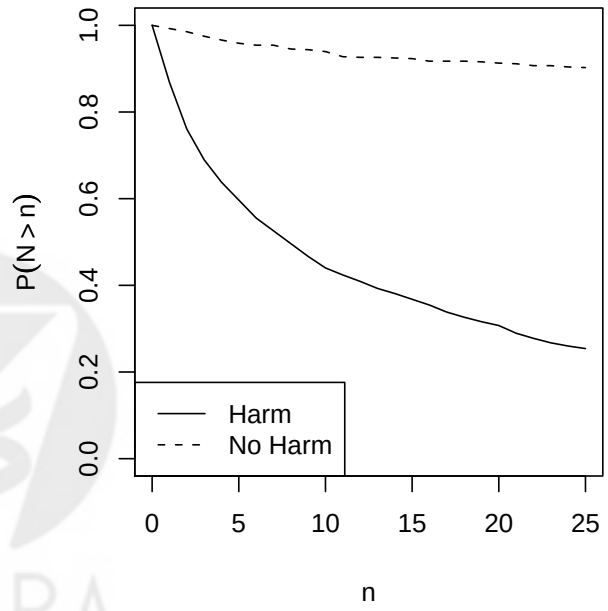

Figure 2: Proportion of hospitals with $N > n$, where N is the number of reports of adverse events (solid line) or near misses (dashed line) submitted by a given hospital.

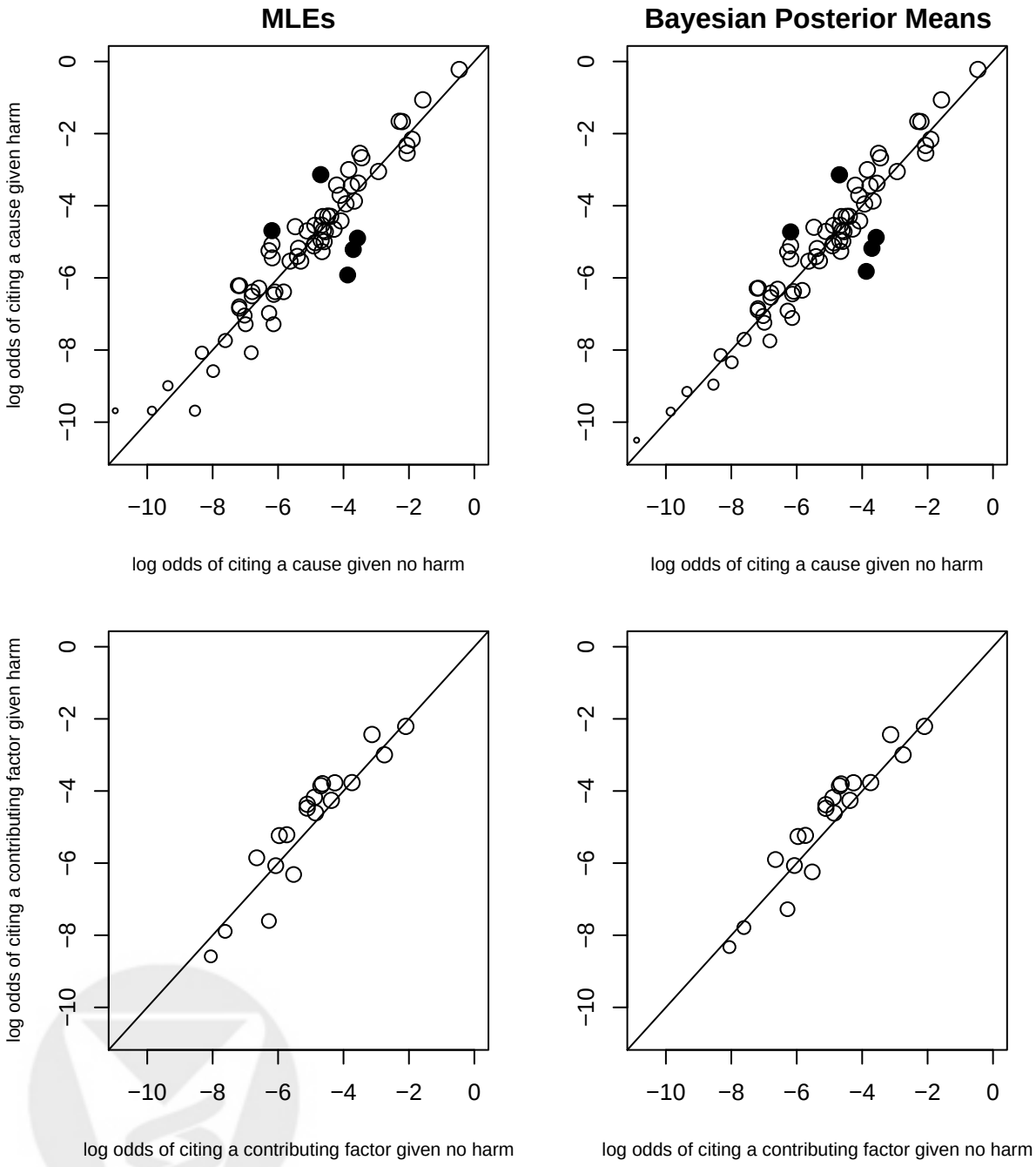

Figure 3: Maximum likelihood estimates (left) and Bayesian estimates (right) of the log-odds of citing a cause (or contributing factor) given that harm has occurred versus harm has not occurred with plotting symbols inversely scaled according to the maximum of the two associated variances (one for each log-odds estimate). Causes identified as having the highest or lowest log odds-ratio of being cited in adverse events versus near misses by the BHM are plotted with a solid circle. The top panel shows the results for the 67 causes, and the bottom panel shows the results for the 20 contributing factors.

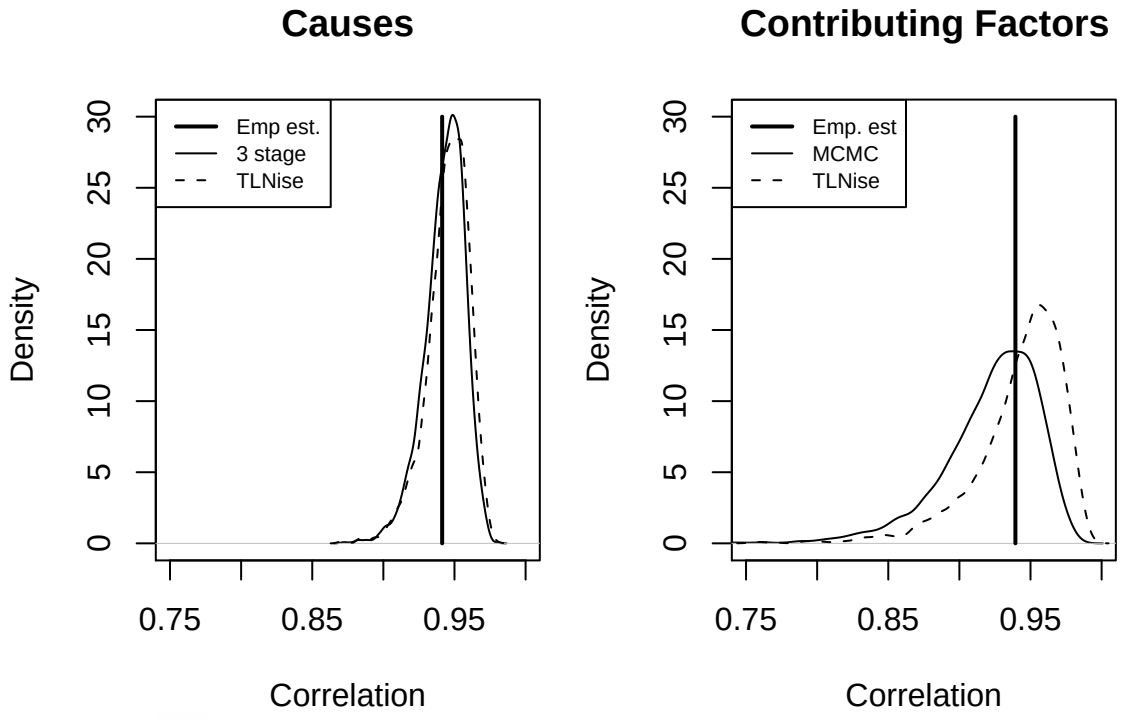

Figure 4: Posterior distribution of the correlation between the log odds of citing a cause given that harm has occurred and the log odds of citing the same cause given that harm has not occurred for causes (left) and contributing factors (right).

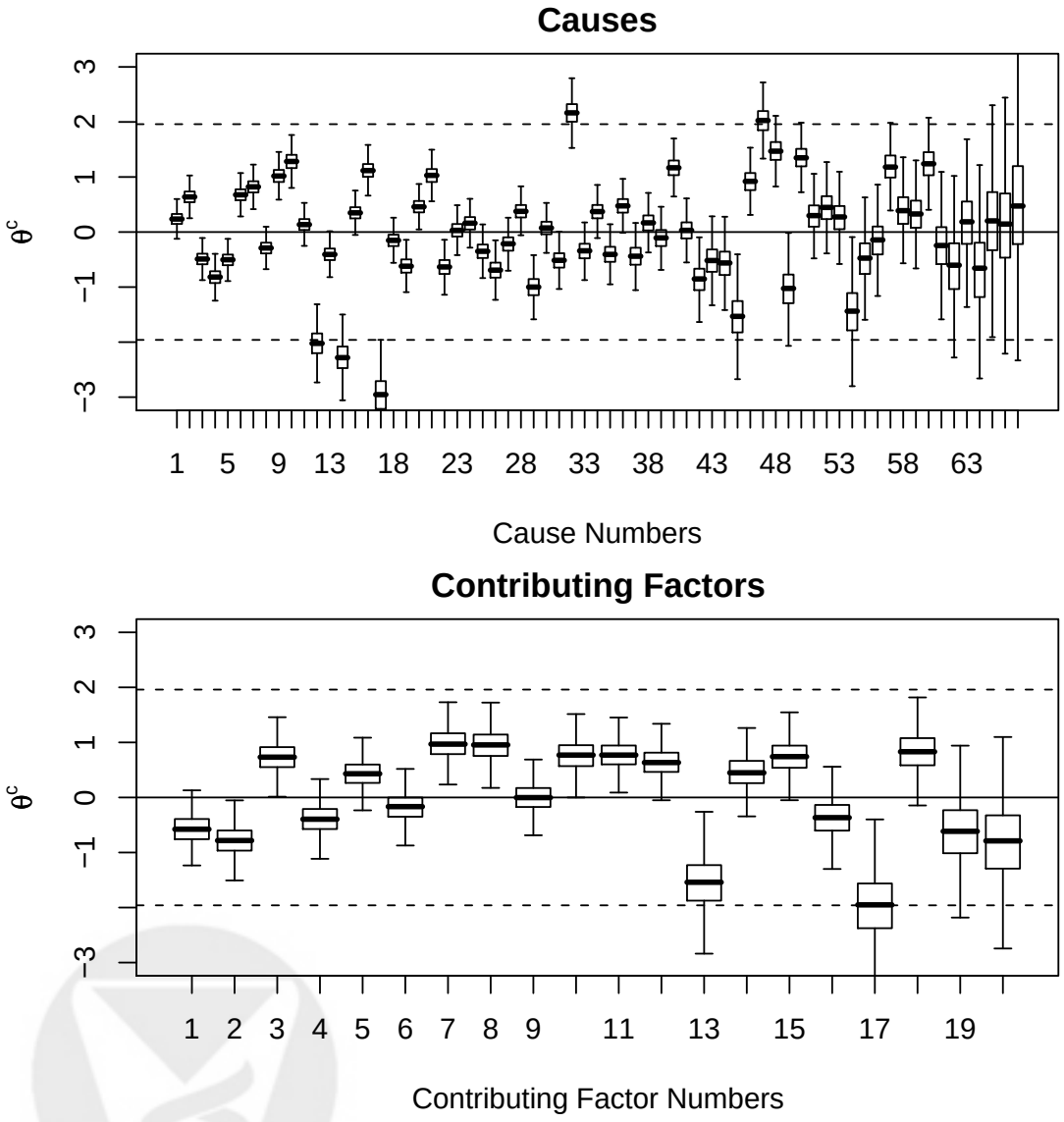

Figure 5: Boxplots of posterior distributions of the standardized log-odds difference θ^c, as estimated by the BHM for causes (top) and contributing factors (bottom). Dashed lines mark the chosen cutoff for identifying aberrant causes, $t = 1.96$

Funding

This work was supported by the National Institute of Environmental Health Sciences [5T32ES012871 to J.M.].

Captions

Table 1: Number and percent of error reports in each harmscore category, defined by The National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP). We do not include reports in category A because no error has occurred. Near misses include reports in the categories B, C, and D. Adverse events include reports in the categories E, F, G, H, and I.

Table 2: List of causes that can be cited when a medication error has occurred. Definitions can be found at www.biostat.jhsph.edu/ jamyers/research. More than one cause can be checked for each error. For each cause, $\hat{p}_0^c$ and $\hat{p}_1^c$ are the observed proportions of citing cause c among adverse events and near misses, respectively. The p-value for testing $H_0 : p_0^c = p_1^c$ versus $H_a : p_0^c \neq p_1^c$ is given, and a * indicates p<.05/67. r_0^c and r_1^c are the ranks of p_0^c and p_1^c, respectively.

Table 3: List of contributing factors that can be cited when a medication error has occurred. Definitions can be found at www.biostat.jhsph.edu/ jamyers/research. More than one contributing factor can be checked for each error. For each contributing factor, $\hat{p}_0^c$ and $\hat{p}_1^c$ are the observed proportions of citing contributing factor c among adverse events and near misses, respectively. The p-value for testing $H_0 : p_0^c = p_1^c$ versus $H_a : p_0^c \neq p_1^c$ is given, and a * indicates p<.05/67. r_0^c and r_1^c are the ranks of p_0^c and p_1^c, respectively.

Figure 1: Number of reports submitted from each of the 100 facilities with the highest total number of reports, plotted separately for the number of harmful reports and the number of nonharmful reports. The facilities in each plot are sorted on total number of reports (harmful + nonharmful).

Figure 2: Proportion of hospitals with $N > n$, where N is the number of reports of adverse events (solid line) or near misses (dashed line) submitted by a given hospital.

Figure 3: Maximum likelihood estimates (left) and Bayesian estimates (right) of the log- odds of citing a cause (or contributing factor) given that harm has occurred versus harm has not occurred with plotting symbols inversely scaled according to the maximum of the two associated variances (one for each log-odds estimate). Causes identified as having the highest or lowest log odds-ratio of being cited in adverse

events versus near misses by the BHM are plotted with a solid circle. The top panel shows the results for the 67 causes, and the bottom panel shows the results for the 20 contributing factors.

Figure 4: Posterior distribution of the correlation between the log odds of citing a cause given that harm has occurred and the log odds of citing the same cause given that harm has not occurred for causes (left) and contributing factors (right).

Figure 5: Boxplots of posterior distributions of the standardized log-odds difference θ^c, as estimated by the BHM for causes (top) and contributing factors (bottom). Dashed lines mark the chosen cutoff for identifying aberrant causes, $t = 1 : 96$.