Journal of Statistical Planning and Inference 160 (2015) 51–59

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Inference for identifying outlying health care providers

Michael J. Racz [a,*], J. Sedransk [b]

[a] Department of Basic and Social Sciences, Albany College of Pharmacy and Health Sciences, Albany, NY 12208, United States

[b] Department of Statistics, Case Western Reserve University, Cleveland, OH 44106-7054, United States

ARTICLE INFO

Article history:
Received 5 June 2013
Received in revised form 3 October 2014
Accepted 5 December 2014
Available online 19 December 2014

Keywords:
Coronary artery bypass graft surgery
Credible regions
Posterior predictive *p*-value

ABSTRACT

Provider profiling is the evaluation of the performance of hospitals, doctors, and other medical practitioners to enhance the quality of care. Many jurisdictions have released public report cards comparing hospital or physician-specific outcomes. Given the importance of provider profiling, studying the methodology used and providing enhancements are essential. Ohlssen, Sharples and Spiegelhalter (2007) give a thoughtful evaluation of provider profiling methodology. In particular they are concerned about whether a putative outlier is really an outlier or an observation in the tail of the common distribution for all practitioners, and present methodology to address this issue. In this paper we suggest as an alternative the use of a $100(1-\alpha)\%$ credible region for provider fixed effects to identify outliers. Using both New York State bypass surgery data and simulated data of the same type as that used in Racz and Sedransk (2010), we compare the Ohlssen et al. (2007) approach with standard methodology, and with use of the credible region.

1. Introduction

Provider profiling is the evaluation of the performance of hospitals, doctors, and other medical practitioners to enhance the quality of care. Many jurisdictions have released public report cards comparing hospital or physician-specific outcomes. The objectives of provider profiling may be to seek corrective measures when a provider's performance is deemed unsatisfactory, or to increase public awareness so that individuals can make an informed decision about selection of a medical practitioner. Given the importance of provider profiling, studying the methodology used and providing enhancements are essential. There is a large literature on this subject: The articles most relevant to our work are those by Christiansen and Morris (1997), Goldstein and Spiegelhalter (1996), Normand et al. (1997), Ohlssen et al. (2007),Thomas et al. (1994), Austin et al. (2001), Austin (2002), Austin et al. (2003), Austin (2005) and Shahian et al. (2005). The last five articles provide comparisons of alternative methods.

The New York State Department of Health (NYS DOH) is a leader in provider profiling. Since 1989 the NYS DOH has evaluated the performance of the hospitals in New York that are licensed to perform coronary artery bypass graft surgery (CABG) (New York State Department of Health, 2009). Racz and Sedransk (2010) presented a Bayesian analogue of the NYS DOH methodology. In both approaches a hospital's actual measure of performance is compared with its expected performance, assuming the same model for all hospitals but using this hospital's own patient case mix to obtain the expected performance measure. Authors like Normand et al. (1997) and Shahian et al. (2005) have advocated adding hospital random effects to the models used by the NYS DOH (see (1) and (2) below). However, with a thoroughly risk-adjusted data set such as

* Correspondence to: Albany College of Pharmacy and Health Sciences, 106 New Scotland Avenue, Albany, NY 12208, United States. Tel.: +1 518 694 7182; fax: +1 518 694 7382.

E-mail addresses: Michael.Racz@acphs.edu (M.J. Racz), jxs123@cwru.edu, jxs123@po.cwru.edu (J. Sedransk).

http://dx.doi.org/10.1016/j.jspi.2014.12.003
0378-3758/

that used by the NYS DOH, using a random effects model rather than (2) is not advantageous. Depending on the criterion used to detect an outlier we will either identify many fewer outliers or somewhat fewer outliers by assuming a random effects model (Racz and Sedransk, 2010). This occurs, mainly, because of inappropriate shrinkage. Section 4 has a brief discussion of methodology when the data set is not thoroughly risk-adjusted.

Ohlssen et al. (2007) are concerned about whether a putative outlier is really an outlier or an observation in the tail of the distribution, and present methodology to address this issue. This is an important concern for applications such as provider profiling where decisions about outlying status may have serious practical consequences. While a potentially useful methodology, it does not provide a way to determine the number and identity of the providers to flag as outliers. Thus, we present an alternative method, i.e., one that provides a simultaneous $100(1-\alpha)\%$ credible region for the ensemble of provider effects, thus addressing the multiple comparisons problem directly. In this paper we use both Cardiac Surgery Reporting System (CSRS) data and simulated data of the same type as that used in Racz and Sedransk (2010) to evaluate the Ohlssen et al. (2007) (hereafter OSS) methodology together with the Bayesian analogue of the NYS DOH method, and use of a credible region. Note that this is also the first systematic evaluation of the OSS approach.

In Section 2 we present the Bayesian NYS DOH methodology, the OSS technique and the procedure to obtain the credible regions. Section 3 contains a description of the CSRS data and our simulation procedure followed by the results of our investigation. Additional discussion and a brief summary are provided in Section 4.

2. Methodology

The first phase of the analysis by the NYS DOH is to perform stepwise logistic regression to find significant predictors of in-hospital mortality from approximately 40 pre-operative risk factors from CSRS. In detail, the data base is split into a development and validation group with roughly equal mortality rates. The risk factors are entered as candidate independent variables in a stepwise logistic regression model with in-hospital mortality as the binary dependent variable. Significant variables in this model are then validated in a stepwise model on the validation group. Significant variables from the validation model are then used to develop a stepwise model with the entire data set. The fit of the final logistic regression model is measured in terms of its discrimination (c-statistic). This method has been used since 1992 and the c-statistic is about 0.80 in each year. For additional details, see Hannan et al. (2006).

Assuming r hospitals and with n_i patients in hospital i, define $Y_{ij} = 1$ if the jth patient in the ith hospital died and $Y_{ij} = 0$ otherwise. Also, $\underline{X}'_{ij} = \left(1, X_{ij1}, \ldots, X_{ijk}\right)$ and $\underline{\beta}' = (\beta_0, \beta_1, \ldots, \beta_k)$ denote the regression structure. Let

$$p_{ij} = Pr\left\{Y_{ij} = 1\right\} \quad i = 1, \ldots, r,\ j = 1, \ldots, n_i, \tag{1}$$

and assume that

$$\operatorname{logit}\left(p_{ij}\right) = \underline{X}'_{ij}\underline{\beta}. \tag{2}$$

It is assumed throughout that, conditional on the p_{ij}, the Y_{ij} are mutually independent. The pre-operative risk factors, $\underline{X}$, are described in NYS DOH (2009).

For hospital i we use Bayesian inference to obtain the predictive distribution of $\sum_{j=1}^{n_i} Y_{ij}$, assuming the statewide model in (1) and (2). We then obtain a $100(1-\alpha)\%$ predictive interval for $\sum_{j=1}^{n_i} Y_{ij}$. Hospital i is considered to be an outlier if, and only if, $\sum_{j=1}^{n_i} y_{ij}^{obs}$ is outside this interval. Define $\underline{Y} = \left\{Y_{ij} : i = 1, \ldots, r,\ j = 1, \ldots, n_i\right\}$ and $\underline{p} = \left\{p_{ij} : i = 1, \ldots, r, j = 1, \ldots, n_i\right\}$, and assume the locally uniform prior distribution on $\underline{\beta}$, $\underline{\beta} \sim N\left(0, 10^4 I\right)$.

Let Y_{ij}^{Pred} denote a predicted value of Y for patient j in hospital i. In applications of this type the posterior predictive distribution of the total number of deaths at hospital i under (2) is typically taken as

$$f\left(\sum_{j=1}^{n_i} Y_{ij}^{Pred} \mid \underline{y}_{-i}^{obs}\right) = \int g\left(\sum_{j=1}^{n_i} Y_{ij}^{Pred} \mid \underline{p}\right) h\left(\underline{p} \mid \underline{y}_{-i}^{obs}\right) d\underline{p} \tag{3}$$

where $\underline{y}_{-i}^{obs}$ is the set of observed values except for those in hospital i. To reduce computations (3) is often approximated by

$$f\left(\sum_{j=1}^{n_i} Y_{ij}^{Pred} \mid \underline{y}^{obs}\right) = \int g\left(\sum_{j=1}^{n_i} Y_{ij}^{Pred} \mid \underline{p}\right) h\left(\underline{p} \mid \underline{y}^{obs}\right) d\underline{p}.$$

Using the Gibbs sampler it is straightforward to make draws from $f\left(\sum_{j=1}^{n_i} Y_{ij}^{Pred} \mid \underline{y}_{-i}^{obs}\right)$ or $f\left(\sum_{j=1}^{n_i} Y_{ij}^{Pred} \mid \underline{y}^{obs}\right)$ and obtain a $100(1-\alpha)\%$ interval for $\sum_{j=1}^{n_i} Y_{ij}^{Pred}$ using the $(\alpha/2)100$th and $(1-(\alpha/2))100$th percentiles of the draws. Hospital i is considered a high outlier if $\sum_{j=1}^{n_i} y_{ij}^{obs}$ is greater than the $(1-(\alpha/2))100$th percentile of the predictive distribution of $\sum_{j=1}^{n_i} Y_{ij}^{Pred}$.