Research Article

Statistics in Medicine

Received 1 November 2008, Accepted 15 September 2009 Published online in Wiley Interscience

(www.interscience.wiley.com) DOI: 10.1002/sim.3756

Establishing the relationship between nurse staffing and hospital mortality using a clustered discrete-time logistic model

L. Diya,[a]*† E. Lesaffre,[a,c] K. Van den Heede,[b] W. Sermeus[b] and A. Vleugels[b]

Studies based on aggregated hospital outcome data have established that there is a relationship between nurse staffing and adverse events. However, this result could not be confirmed in Belgium where 96 per cent of the variability of nurse staffing levels over nursing units (belonging to different hospitals) is explained by within-hospital variability. To better appreciate the possible impact of nurse staffing levels on adverse events, we propose a multilevel approach reflecting the complex nature of the data. In particular we suggest a clustered discrete-time logistic model that captures the risks associated with a given unit in the patient's trajectory through the hospital. The model also allows for nurse staffing levels to affect the current and subsequent nursing unit (carry-over effect). In the model 'time' is represented by the sequential number of the nursing unit that the patient is passing through. The model incorporates hospital and nursing unit random effects to express that patients treated in the same hospital and taken care of by nurses of the same unit share a common environment. In this study we used Belgian national administrative databases for the year 2003 to assess the relationship between nurse staffing levels and nurse education variables with in-hospital mortality. The analysis was restricted to elective cardiac surgery patients. Lower nursing unit staffing levels in the general nursing units were associated with high in-hospital mortality in units past the traditional cardiac surgery nursing units.

Keywords: multilevel; non-hierarchical; logistic; discrete time; clustered data

1. Introduction

Patient safety issues reached worldwide prominence with the publication of the 'To Err is Human' report by the Institute of Medicine in 1999 [1]. The increased attention toward this topic is in particular owed to the publication of the estimates that between 44 000 and 98 000 deaths occur as a consequence of medical errors in U.S. hospitals each year. Patient safety is defined as 'the absence of the potential for, or occurrence of, health care-associated injury to patients created by avoiding medical errors as well as taking action to prevent errors from causing injury' [2]. Adverse events that are 'injuries caused by medical management rather than by the underlying disease or condition' [1] can be used as proxies of patient safety.

In the middle of the nineteenth century, Florence Nightingale pioneered epidemiological and statistical methods to study hospital death rates. Using registration data she observed that death rates in hospitals could be reduced by investing in better sanitation (nursing) [3]. In the early twentieth century, Ernest Codman was the first to systematically follow all patients to record the 'end result' of the surgical care they received. He linked care, errors and 'end results' with the purpose of patient care improvement [4]. Despite the pioneering work of Nightingale and Codman, the measurement of adverse events to improve health

[a] *Interuniversity Institute for Biostatistics and Statistical Bioinformatics, Katholieke Universiteit Leuven, Kapucijnenvoer 35, Blok D, Bus 7001, B3000 Leuven, and Universiteit Hasselt, Belgium*

[b] *Centre for Health Services and Nursing Research, Kapucijnenvoer 35, B-3000 Leuven, Belgium*

[c] *Department of Biostatistics, Erasmus Medical Center, PO Box 2040, 3000 CA Rotterdam, The Netherlands*

* *Correspondence to: L. Diya, Interuniversity Institute for Biostatistics and Statistical Bioinformatics, Katholieke Universiteit Leuven, Kapucijnenvoer 35, Blok D, Bus 7001, B3000 Leuven, and Universiteit Hasselt, Belgium.*

† *E-mail: luwis.diya@med.kuleuven.be*

Contract/grant sponsor: Belgian Ministry of Science Policy, in collaboration with the Belgian Ministry of Public Health; contract/grant number: AG/01/123
Contract/grant sponsor: Interuniversity Attraction Poles Program P6/03_Belgian State_Federal Office for Scientific, Technical and Cultural Affairs.

care organization only gained momentum with the publication of the 'Harvard Medical Practice Study' in 1991, one of the two studies on which the 'To Err is Human' report estimates were based. Furthermore, research conducted in seven different countries revealed that between 4 and 17 per cent of the hospitalized patients experience at least one adverse event during a hospital episode [5, 6].

The Institute of Medicine recommended in 1996 to urgently invest in research determining the relationship between hospital outcomes and nurse staffing levels, since a literature review illustrated that there was no systematic evidence supporting the relationship [7]. More than 100 papers reporting on the association between nurse staffing levels and patient safety were published in the last decade using a variety of study approaches and data sources. Most of these studies support the notion that nurse staffing levels are related to patient safety in acute hospitals. The undisputed landmark study in this study domain is that of Aiken *et al.* [8], where nursing information was aggregated at the hospital level and established that nurse's workload is associated with 30-day mortality and failure-to-rescue (death after a complication). A second paper from this group also concluded that hospitals with higher educated nurses are associated with low 30-day mortality and failure-to-rescue rates [9].

The aim of the present study is to determine the impact of nurse staffing levels and/or nurse education on in-hospital mortality exploiting the multilevel structure of the data. Though the data structure is multilevel, it is not completely hierarchical; hence, the classical multilevel modeling is not appropriate. Instead we developed an appropriate model that mimics the patient trajectory and hence accommodates the true structure of the data. This paper is organized as follows. Section 2 gives a general description of the data, the study population and the variables. In Section 3 we first discuss the pros and cons of some standard modeling approaches to analyze the data at hand and then move to the suggested clustered discrete-time logistic model. In Section 4 we apply our model to the Belgian data. The paper ends with a discussion in Section 5.

2. The Belgian databases

A variety of data sources are used in the nurse staffing and patient safety research. However, most large multicenter studies use hospital discharge data sets to obtain adverse events records and nurse surveys or administrative data to measure nurse staffing levels. Hospital discharge databases generally contain information about demographic characteristics, discharge status, length of stay, diagnoses based on International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM), and procedures. In Belgium all acute hospitals contribute to the Belgian Hospital Discharge Data set (B-HDDS). Individual nursing unit data on staffing patterns and the nature of nursing care can be obtained from the Belgian Nursing Minimum Data set (B-NMDS) [10]. Although these databases contain a large population of nurses, there are some limitations, e.g. (a) for B-HDDS: some diagnoses are not assigned a time stamp (date) and in some cases it might be difficult to distinguish comorbidity from an adverse event (complication), (b) for B-NMDS: there might be problems in distinguishing staff involved in in-patient and out-patient activities.

In this study we used the B-HDDS and the B-NMDS for the year 2003 to assess the relationship between nurse staffing variables (nurse staffing levels and nurse education) and in-hospital mortality.

The study sample is restricted to patients, aged between 20 and 85 years, who were electively admitted to a Belgian acute hospital for a coronary artery bypass graft (CABG) or heart valve procedure. We considered patients as having a CABG or heart valve procedure if they were assigned to one of the following categories from All Patients Refined Diagnosis-Related-Group (APR-DRG): 162, 163, 165 and 166 [11]. The selected cardiac procedures involved specialized care that is delivered in 29 of the 115 Belgian acute care facilities. One hospital was dropped from the analysis because there was no link between the two administrative databases. Since we are interested in how nurse staffing variables affect mortality during the postoperative period, patients who died before their surgical procedure were removed from our analysis (0.4 per cent). Also excluded from analysis were patients whose records did not permit linkage to a specific nursing unit postoperatively (2.5 per cent). The final sample comprised 9054 patients.

The response used in this paper is in-hospital mortality, which could be considered as a surrogate of patient safety. Unlike the other adverse events (e.g pressure ulcers and pneumonia) one can pinpoint the nursing unit where the death occurred. However, a downside of using in-hospital mortality is that we are diluting mortality caused by medical errors and 'naturally' occurring mortality. To establish the relationship between nurse staffing variables and in-hospital mortality, there is a need to control for confounders. The considered covariates (including confounders) in this study were taken at (a) patient level, (b) nursing unit level and (c) hospital level.

The patient characteristics considered are age of patient (age), gender of patient (male $=1$ for males or 0 otherwise), type of surgical procedure i.e. APR-DRG (162, 163, 165 or 166) and Risk of Mortality (ROM: $1=$Minor, $2=$Moderate, $3=$Major, $4=$Extreme). ROM categories represent the likelihood for a given patient to die and were based initially on all available secondary diagnoses [12] and calculated using the 3M grouper software [13]. Unfortunately secondary diagnoses may represent co-morbidities as well as complications. This might lead to an over-correction of the relationship between staffing variables and in-hospital mortality. Therefore, the final ROM categories were calculated without in-hospital complications. But, there is no best solution to the problem since by doing so, there is now a danger of under-correction.

The nursing unit characteristics considered include: nurse staffing levels, educational level of nurses and intensity of nursing care (measure to quantify how much care patients need) [10]. In this paper intensive care will be denoted as IC and general nursing unit as *G*. In particular, the variables are intensive care nurse staffing levels (IC-Staff), general nursing unit nurse staffing levels (*G*-Staff), percentage of nurses with a bachelor degree in intensive care nursing unit (IC-%Bat), percentage of nurses with a bachelor degree in general nursing unit (*G*-%Bat), intensity of nursing care for intensive care nursing unit (IC-Intensity) and intensity of nursing care for general nursing unit (*G*-Intensity). The nursing unit characteristics are allowed to influence current and subsequent units (carry-over effect).

For hospital characteristics we considered the yearly volume of the cardiac procedures (volume), the average of IC-Staff caring for cardiac surgery patients in a hospital ($\overline{\text{IC-Staff}}$), the average of *G*-Staff caring for cardiac surgery patients in a hospital ($\overline{G\text{-Staff}}$), the percentage of IC unit nurses caring for cardiac surgery patients with a bachelor degree in hospital ($\overline{\text{IC-\%Bat}}$) and the percentage of G unit nurses caring for cardiac surgery patients with a bachelor degree in hospital ($\overline{G\text{-\%Bat}}$). Standard methods used in analyzing clustered data model the relationship between the covariates and the response without partitioning the covariate effect into the between- and within-cluster effects. That is, they implicitly assume that the effects are identical, which can lead to misleading inferences [14]. By including the nurse staffing variables at nursing unit level and averaging them over the nursing units to create mean nurse staffing variables at the hospital level, we have taken these effects into account. This can also reduce the bias in the parameter estimates induced when the random effects are correlated with some of the covariates [15].

3. Statistical modeling

In order to adequately assess the impact of nurse staffing variables on in-hospital mortality, the statistical models should reflect the data structure. However, many studies on hospital outcome data have not adequately addressed the complex nature of the data and the statistical models are sometimes based on strong unjustified assumptions [8, 9, 16]. In the first part of this section, we describe some simple approaches that have been used in the literature to analyze the hospital outcome data and highlight the limitations of these approaches. In the second part, we introduce the clustered discrete-time logistic model.

3.1. Some possible approaches

The data set in this study is clustered in nature, that is, patients are clustered in nursing units and nursing units are clustered within hospitals. These levels of clustering lead to dependencies within hospitals and within nursing units, that is, patients in the same nursing unit are exposed to the same nursing care and nursing units in the same hospital are directed by the same management. Some studies were based on aggregated nursing unit information resulting in a two-level data structure: patient level and hospital level [8]. The Generalized Estimating Equation approach (GEE) can then be used to evaluate the relationship between nursing variables and adverse events. GEE parameter estimates have a marginal or population average interpretation, whereas those of multilevel models have a conditional interpretation. This approach correctly adjusts the standard errors of the parameter estimates for dependencies in the data without necessarily modeling the dependency structure. We used such an approach on the Belgian data set and thereby replayed the analysis of Aiken *et al.* [8]. However, our analysis did not show a significant relationship between the nurse staffing variables and various adverse events. The reason for the failure to confirm the results of the study of Aiken *et al.* [8] appeared to be that there is much more variability within hospitals than between hospitals with respect to nurse staffing variables. By aggregating covariate information from the nursing unit level to the hospital level, a lot of valuable detailed information is lost.

To examine the relationship between the nurse staffing variables and the occurrence of adverse events, one must realize that the data show a multilevel but not truly hierarchical nature. Indeed, patients can reside in more than one nursing unit during their hospital stay. One way to get around this is to pick one nursing unit. Moreover, there are serious problems with this approach, since it is not clear which nursing unit to choose. Indeed, the entire patient trajectory is important since an adverse event may develop in one of the nursing units but manifest itself in another unit. That is, there may be carry-over effects that cannot be addressed by a classical hierarchical model. Also the estimated variance components coming from the hierarchical model may be biased [17]. An alternative approach is to use non-hierarchical multilevel models such as the Multiple Membership model (MM model). In an MM model, the lower-level units are influenced by more than one higher-level unit. In this approach the nursing units are assigned weights proportional to the patient's length of stay. However, the length of stay is not a good measure to base these weights on as the length of stay in a nursing unit could be influenced to a large extent by the development of adverse events.

3.2. The clustered discrete-time logistic model

3.2.1. Introduction. We propose the clustered discrete-time logistic model to analyze the relationship between the nurse staffing variables and in-hospital mortality. This model is an extension of the discrete-time logistic regression model [18]. Applied to the current settings, the discrete-time logistic model assumes that the logit of the probability of death in a given nursing unit, conditional that death has not occurred prior to that nursing unit, is a linear function of the covariates (risk adjusters included) and time-specific intercepts. Here, time represents the sequential number of the nursing unit that a patient passes through. In the clustered discrete-time logistic model, we further adjust for the clustering of patients within nursing units and within hospitals through random intercepts. The entire patient trajectory consists of the preoperative nursing units, the operation theater (OT) and the postoperative nursing units. The preoperative nursing unit was not considered in the model as the incidence of adverse events is quite low (0.4 per cent) and we believe that adding this unit to the model offers little gain in understanding the relationship between the nurse staffing variables and in-hospital mortality. This gives the OT (unit 1), first postoperative IC nursing unit (unit 2), first postoperative *G* nursing unit (unit 3) and an artificial unit (unit 4) representing all postoperative nursing units that the patient might travel through after the first postoperative general nursing unit. We evaluate the immediate and carry-over effects of nurse staffing variables on the patient's outcome.

3.2.2. Model description. Suppose an event can occur in one of J nursing units and that a patient is followed up to the unit where the event occurs or up to the last unit if the event does not occur. Let Y_{ijk} be the variable denoting the occurrence of death for the kth ($1 \leqslant k \leqslant n_{ij}$) patient in the jth ($1 \leqslant j \leqslant n_i$) nursing unit of the ith ($1 \leqslant i \leqslant N$) hospital. Let $\mathbf{x}_{ijk}$ denote the vector of covariates, that is the history of patient k in hospital i up to unit j. This vector contains risk adjusters and the nurse staffing variables. The effect of the risk adjusters is considered to be the same for all nursing units but for nursing intensity and the nurse staffing variables their effects are allowed to depend on the nursing unit. Carry-over effects of IC and G nurse staffing variables are also taken into account. Let $\mathbf{b}_{ik}$ represent the vector of random effects associated with the kth patient in the ith hospital. Further, let $\pi(\mathbf{x}_{ijk}, \mathbf{b}_{ik})$ denote the conditional probability that the kth patient in the ith hospital dies in the jth (post)operative nursing unit, i.e.

$$\pi(\mathbf{x}_{ijk}, \mathbf{b}_{ik}) = P(Y_{ijk} = 1 | Y_{ij'k} = 0 \ \forall \ j' < j, \mathbf{x}_{ijk}, \mathbf{b}_{ik}) \tag{1}$$

The last nursing unit that a patient stays in (j_{ik}) can be unit 1 (*patient dies in OT*), unit 2 (*patient dies in intensive care nursing unit*), unit 3 (*patient dies/is discharged in general nursing unit*) or unit 4 (*patient dies/is discharged after the general nursing unit*).

The clustered discrete-time logistic model used here is defined as follows:

$$\operatorname{logit}(\pi(\mathbf{x}_{ijk}, \mathbf{b}_{ik})) = \begin{cases} \beta' \mathbf{x}_{ijk} + b_i + b_{1ik} & \text{for OT} \\ \beta' \mathbf{x}_{ijk} + b_i + b_{2ik} & \text{for IC} \\ \beta' \mathbf{x}_{ijk} + b_i + b_{3ik} & \text{for } G \\ \beta' \mathbf{x}_{ijk} + b_i + b_{4ik} & \text{for fourth unit} \end{cases} \tag{2}$$

where $\mathbf{b}_{ik} = (b_i, b_{1ik}, b_{2ik}, b_{3ik}, b_{4ik})^{\mathrm{T}}$ represents the vector of random effects, with b_i, b_{1ik}, b_{2ik}, b_{3ik} and b_{4ik} representing the hospital, OT, IC nursing unit, G nursing unit and fourth unit random effects, respectively [$\mathbf{b}_{ik} \sim \mathrm{N}(\mathbf{0}, \mathbf{D})$ where $\mathbf{D} = \operatorname{diag}(\sigma_1^2, \sigma_{21}^2, \sigma_{22}^2, \sigma_{23}^2, \sigma_{24}^2)$]. Typically a patient will stay in unit 4 if the complications occurred in unit 3.

The conditional likelihood contribution (conditional on the random effects) for the kth patient in the ith hospital who stayed up to nursing unit j_{ik} (last unit for that patient) is

$$L_{ik}(\boldsymbol{\beta}, \mathbf{b}_{ik} | \mathbf{y}_{ik}) \equiv p(\mathbf{y}_{ik} | \boldsymbol{\beta}, \mathbf{b}_{ik}) = \pi(\mathbf{x}_{ij_{ik}k}, \mathbf{b}_{ik})^{y_{ij_{ik}k}} \prod_{j=1}^{j_{ik}} (1 - \pi(\mathbf{x}_{ijk}, \mathbf{b}_{ik}))^{(1-y_{ijk})} \tag{3}$$

where $\mathbf{y}_{ik}$ is a vector of responses for the kth patient in the ith hospital.

The marginal likelihood is obtained as a product of $L_{ik}(\boldsymbol{\beta}, \mathbf{b}_{ik} | \mathbf{y}_{ik})$ over all patients and whereby the random intercepts have been integrated out. Maximizing this likelihood yields the Maximum Likelihood Estimates of the fixed effects parameters as well as the variance components. Another approach is to use a Bayesian model fitting technique. This employs Markov Chain Monte Carlo (MCMC) techniques to estimate the parameters of model (2) using WinBUGS [19]. The R package CODA [20] was used to assess the MCMC convergence through various convergence diagnostics. Given the prior distributions of $\boldsymbol{\beta}$ ($p(\boldsymbol{\beta})$) and of $\mathbf{D}$ ($p(\mathbf{D})$) representing the model parameters of the random effects distribution, the posterior distribution can be computed as follows:

$$p(\boldsymbol{\beta}, \mathbf{b} | \mathbf{y}) \propto \prod_{i}^{N} \prod_{k}^{n_i} p(\mathbf{y}_{ik} | \boldsymbol{\beta}, \mathbf{b}_{ik}) \prod_{i}^{N} \prod_{k}^{n_i} p(\mathbf{b}_{ik} | \mathbf{D}) p(\mathbf{D}) p(\boldsymbol{\beta}) \tag{4}$$

For model assessment, we noted how sensitive our analysis is with respect to varying the link function and the distribution of the random effects. More specifically, (a) we replaced the logit link function by the complementary log(-log) link and (b) we assumed a Student-t distribution (with varying degrees of freedom but also estimating the degrees of freedom) for the random effects distribution. The alternative models were compared with model (2) using the deviance information criterion (DIC) [21]. Posterior predictive checks (PPC) were employed to assess the goodness-of-fit of model (2) using discrepancies. The discrepancy (T) used here is the sum of in-hospital mortality over a subset of the observations. Different versions of T might be considered:

$$T_{1,i}(\mathbf{y}_i; \boldsymbol{\theta}) = \sum_{k=1}^{n_i} \sum_{j=1}^{j_{ik}} y_{ijk} \quad \text{where } i = 1, 2, \ldots, 27 \text{ or } 28 \tag{5}$$

$$T_{2,t}(\mathbf{y}_t; \boldsymbol{\theta}) = \sum_{i=1}^{N} \sum_{k=1}^{n_i} y_{ij(t)k} \quad \text{where } t = 1, 2, 3 \text{ or } 4 \tag{6}$$

where $j(t)$ signifies that the jth nursing unit is in position t of the patient trajectory; $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \mathbf{D}\}$ represents the total parameter vector; $\mathbf{y}$, $\mathbf{y}_i$ and $\mathbf{y}_t$ represent the total vector of responses, the total vector of responses for the ith hospital and the total vector of responses for the tth discrete 'time', respectively. The above discrepancies are inspired by the Hosmer and Lemeshow goodness-of-fit test [22]. They partition the total set of observations into deciles of risk, but here we partition the total set of observations according to hospitals and according to units (discrete-time points).

From the PPC we calculated the Bayesian p-value [23] as

$$p_{\mathrm{B}} = P(T(\mathbf{y}^{\mathrm{rep}};\boldsymbol{\theta}) \geqslant T(\mathbf{y};\boldsymbol{\theta})|\mathbf{y}) \tag{7}$$

$$= \frac{1}{M}\sum_{m=1}^{M} I(T(\mathbf{y}_m^{\mathrm{rep}},\boldsymbol{\theta}_m) \geqslant T(\mathbf{y},\boldsymbol{\theta}_m)) \tag{8}$$

where $I(.)$ takes the value 1 if the condition is true and 0 otherwise, $\mathbf{y}_m^{\mathrm{rep}}$ is the mth replication from the posterior predictive distribution. If the model is appropriate the posterior predictive replications should look similar to the observed data. We will consider Bayesian p-values close to 0 ($\leqslant 0.1$) and 1 ($\geqslant 0.9$) as extreme.

4. Results

Most in-hospital deaths were observed in the group of patients categorized as ROM 4 (247 deaths), followed by ROM 3 (24 deaths), then ROM 2 (5 deaths) and ROM 1 (4 deaths). Table I shows the percentage of in-hospital deaths for some of the covariates. The proportion of deaths is highest for the group of patients who stayed past the traditional cardiac surgery nursing units (Unit 4).

We ran three MCMC chains each for 50 000 iterations and employed a thinning ratio of 50. There was no evidence suggesting that the MCMC had not converged as all the diagnostic measures were within acceptable limits, e.g. the Brooks, Gelman and Rubin's diagnostics ($\hat{R}$) were close to 1 and there was a quick mixing of MCMC chains. The continuous covariates were centered around their mean and standardized.

4.1. Model results

The variables volume, male, age, APR-DRG, ROM, IC-Intensity, G-Intensity, $\overline{\text{IC-Staff}}$ and $\overline{\text{G-Staff}}$ were considered as risk adjusters from the literature; hence, these variables were retained in the model irrespective of their statistical importance. As for the within-hospital nursing variables (nurse staffing levels and nurse education at the nursing unit level), they were dropped when the 95 per cent credible interval (CI) of their regression coefficient included 0 and the same is true for between-hospital nurse staffing variables provided that their associated within-hospital nurse staffing variables were not statistically important. Table II shows the posterior means and 95 per cent CIs of the parameter estimates and the odds ratios of the final model. From this model, one can conclude that there is a carry-over effect of general nursing unit staffing levels, i.e. cardiac surgery patients who stay past the postoperative general nursing unit where nurse staffing levels are relatively high, are associated with a lower risk of dying. Note that despite that the 95 per cent CI of the odds ratios for the average (between-hospital) nurse staffing level variables contains 1, they were retained in the model. The reason for this is that we have split up the effect of these variables into a within- and a between-hospital effect (see Section 2) and for interpretability reasons we preferred to keep both effects in the model. Further, there seems to be no major impact of the experience of the institution as expressed by volume (95 per cent CI of the odds ratio for *volume* contains 1). Further, the odds of death increases with 1.329 (1.112; 1.580) for an increase of 10 years in age. The odds of death is lower for male patients compared to female patients (0.857 (0.739; 0.992)).

4.2. Model checking

There is no noticeable difference between the model with a complementary log(-log) link (DIC value = 1822.0) and the model with a logit link (DIC value = 1823.7). Hence the logit link was retained. While changing the random effects distribution from

Table I. Percentage of in-hospital deaths.

Characteristic	Percentage of deaths				
Gender	Males			Females	
	2.5			4.6	
Age groups	20–34	35–49	50–64	65–79	80+
	0	0.4	1.6	3.5	7.6
APR-DRG	162	163	165	166	
	7.2	3.7	3.8	2.0	
ROM	1	2	3	4	
	0.1	0.2	1.1	20.2	
Unit	1	2	3	4	
	0.0	2.1	0.4	8.6	

Table II. Results of Bayesian analysis of the clustered discrete-time logistic regression model.

Effect	Mean (95 per cent CI)	Odds ratio (95 per cent CI)
Intercept	−11.883 (−14.081; −10.150)	— (—; —)
Unit 2	4.803 (3.596; 6.678)	330.331 (36.451; 794.928)
Unit 3	2.637 (1.147; 4.529)	42.907 (3.149; 92.673)
Unit 4	3.978 (2.102; 6.123)	186.768 (8.181; 456.243)
volume	−0.115 (−0.462; 0.229)	0.905 (0.630; 1.257)
male	−0.157 (−0.306; −0.011)	0.857 (0.736; 0.989)
age	0.414 (0.156; 0.675)	1.526 (1.169; 1.964)
APR-DRG 162	0.000 (0.000; 0.000)	1.000 (—; —)
APR-DRG 163	0.117 (−0.383; 0.647)	1.163 (0.682; 1.911)
APR-DRG 165	−0.055 (−0.548; 0.459)	0.977 (0.578; 1.582)
APR-DRG 166	−0.149 (−0.665; 0.394)	0.893 (0.515; 1.483)
ROM 1	0.000 (0.000; 0.000)	1.000 (—; —)
ROM 2	−0.052 (−1.566; 1.474)	1.258 (0.209; 4.367)
ROM 3	2.023 (0.923; 3.304)	9.238 (2.518; 27.223)
ROM 4	5.230 (4.249; 6.470)	224.349 (70.035; 645.565)
IC-Intensity	0.269 (−0.001; 0.573)	1.323 (0.999; 1.773)
Carry-over IC-Intensity	0.383 (−0.160; 0.955)	1.527 (0.852; 2.599)
G-Intensity	−0.150 (−0.680; 0.332)	0.888 (0.506; 1.394)
Carry-over *G*-Intensity	−0.058 (−0.802; 0.657)	1.013 (0.448; 1.929)
$\overline{\text{IC-Staff}}$	−0.155 (−0.496; 0.185)	0.869 (0.609; 1.203)
$\overline{\text{G-Staff}}$	−0.326 (−0.706; 0.049)	0.734 (0.494; 1.050)
Carry-over *G*-Staff	−1.075 (−2.175; −0.138)	0.389 (0.114; 0.871)
σ^2_{21}	0.491 (0.028; 2.737)	
σ^2_{22}	0.150 (0.030; 0.449)	
σ^2_{23}	0.821 (0.048; 2.963)	
σ^2_{24}	8.786 (2.881; 19.231)	
σ^2_{1}	0.428 (0.100; 1.030)	

normal to Student-*t* distribution there also was no major difference between these models. Indeed, for a model where all the random effects follow a Student-*t* distribution with four degrees DIC was 1823.4; for a model where all random effects follow a Student-*t* distribution but with equal unspecified degrees of freedom DIC was 1822.9 and the estimated degrees of freedom was 18.9; for a model where each random intercept follows a Student-*t* distribution having unspecified degrees of freedom DIC was 1822.7 and the estimated degrees of freedom were 25.0, 14.6 and 27.9 for hospital, IC and *G* nursing units, respectively. From these results we believe that the normality assumption on the random effects is justified.

From comparing the observed with the expected number of deaths in each 'time point' we obtained the following Bayesian *p*-values 0.348, 0.476, 0.447 and 0.468 for units 1 to 4, respectively. We performed the same calculations across hospitals and obtained Bayesian *p*-values ranging from 0.197 to 0.720. In summary, our sensitivity analysis does not suggest major discrepancies between the posited model and the data.

5. Discussion

The exploration of health outcome data is essential in comprehending patient safety since medical errors not only cause human suffering, if not loss of life, but also impact hospitals financially. Many studies have established a link between nurse staffing variables and adverse events. Belgium is unique by having two important databases that can be linked to assess this relationship. In-hospital mortality fails to capture patients who developed a complication but died post discharge from hospital. However, the above databases necessitated the use of in-hospital mortality instead of the preferred 30-day mortality. A further limitation of using mortality as a surrogate for patient safety is that (in-hospital or 30-day) mortality incorporates 'naturally' occurring mortality, i.e. deaths that have nothing to do with patient safety. In this regard other adverse events appear to be better proxies, e.g. hospital-acquired pneumonia, wound infection, etc. However, these measures lack a time stamp in the B-HDDS. But, even if time stamps were available the accuracy of these time stamps would be questionable. Finally, since the B-HDDS were established for hospital financing purposes not all necessary information was collected, such as information on the treating physician and surgeons.

The clustered discrete-time logistic model allows for proper modeling of the dependency structure in non-hierarchical multilevel data structures. Indeed, this model captures the trajectory of a patient in a hospital and the dependencies of patients in the same nursing unit and in the same hospital. The discrete 'time' points, number of nursing units, can be extended to accommodate all nursing units in a patient trajectory but then the model becomes less parsimonious. Extending the model to accommodate all possible paths a patient can take is currently under investigation. A strength of our model is that it can address the impact of carry-over effects, that is, nurse staffing variables may affect the current and subsequent nursing units.

PPC have been criticized in the statistical literature for their double usage of the data [24]. Other approaches have been suggested namely the partial predictive checks [24] and conflict *p*-values [25], but there is still debate about the best approach to assess multilevel models. The advantage of the PPC is that they are computationally easy to implement as compared with partial predictive checks and conflict *p*-values. Given the large size of the data set and the time-consuming model fitting procedures, we restricted ourselves to the PPC.

Another approach to assess the impact of nurse staffing variables on adverse events is the model proposed by Ecochard and Clayton [26] that was developed to model pregnancies and where the random effects are log-gamma distributed. We are currently investigating this model.

The nurse staffing levels of the general nursing unit were negatively associated with in-hospital mortality for patients staying past the first postoperative general nursing unit. However, we need to realize that the strength of the evidence is not great. Moreover, a lot of modeling has been involved thereby increasing the chance for spurious findings. Therefore our results need to be confirmed by additional studies.

Acknowledgements

The authors would like to thank the associate editor and the reviewers for constructive comments and suggestions. These have led to considerable improvements in this paper. This research was supported by Grant (AG/01/123) from the Belgian Ministry of Science Policy, in collaboration with the Belgian Ministry of Public Health. The first and second authors were partially supported by the interuniversity Attraction Poles Program P6/03_Belgian State_Federal Office for Scientific, Technical and Cultural Affairs.

References

1. Kohn LT, Corrigan JM, Donaldson MS. *To Err is Human: Building a Safer Health System*. National Academy Press: Washington, DC, 1999.
2. Agency for Health Research and Quality. AHRQ's patient safety initiative: building foundations, reducing risk. Interim Report to the Senate Committee on Appropriations. *AHQR Publication No. 04-Rg005*, 2004. Available from: http://www.ahrq.gov/qual/pscongrpt/.
3. Iezzoni LI. *Risk Adjustment for Measuring Health Care Outcomes* (3rd edn). Health Administration Press: Chicago, IL, 2003.
4. Neuhauser D. Ernest Amory Codman MD. *Quality in Safety and Health Care* 2002; **11**:104–105.
5. Wilson RM, Runciman WB, Gibberd RW, Harrison BT, Newby L, Hamilton JD. The quality in Australian health care study. *Medical Journal of Australia* 1995; **163**:458–471.
6. Baker RG, Norton PG, Flintoft V, Blais R, Brown A, Cox J, Etchells E, Ghali WA, Hebert P, Majumdar SR, O'Beirne M, Palacios-Derflingher L, Reid RJ, Sheps S, Tamblyn R. The Canadian adverse events study: the incidence of adverse events among hospital patients in Canada. *Canadian Medical Association Journal* 2004; **170**:1678–1686.
7. Wunderlich G, Sloan F, Davis C. *Nursing Staff in Hospitals and Nursing Homes: Is It Adequate?* National Academy Press: Washington, DC, 1996.
8. Aiken LH, Clarke SP, Sloane DM, Sochalski J, Silber JH. Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. *Journal of the American Medical Association* 2002; **288**:1987–1993.
9. Aiken LH, Clarke SP, Cheung RB, Sloane DM, Silber JH. Educational levels of hospital nurses and surgical patient mortality. *Journal of the American Medical Association* 2003; **290**:1617–1623.
10. Sermeus W, Delesie L, Van den Heede K, Diya L, Lesaffre E. Measuring the intensity of nursing care: making use of the Belgian nursing minimum data set. *International Journal of Nursing Studies* 2008; **45**(7):1011–1021. DOI: 10.1016/j.ijnurstu.2007.05.006.
11. van Den Heede K. Nurse-staffing levels and patient safety in acute hospitals: analyzing administrative data at the nursing unit level. *Ph.D. in Medical Sciences*, Centre for Health Services and Nursing Research, Department of Public Health, Katholieke Universiteit Leuven, Centre for Health Services and Nursing Research, Kapucijnenvoer, Leuven, Belgium, 2008. ISBN 9729961506.
12. Weingart SN, Iezzoni LI, Davis RB, Palmer RH, Cahalane M, Hamel MB, Mukamal K, Phillips RS, Davies DTJ, Banks NJ. Use of administrative data to find substandard care: validation of the complications screening program. *Medical Care* 2000; **38**(8):796–806.
13. Averil RF, Goldfield N, Hughes JS, Bonazelli J, McCullough EC, Steinbeck BA, Mullin R, Tang AM. Development of the all patient refined DRGs (APR-DRGs). 3 M HIS Research Report. *Technical Report*, 3M Health Information Systems, Wallingford, CT, 1997.
14. Neuhaus JM, Kalbfleisch JD. Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics* 1998; **54**(2):638–645. Available from: http://www.jstor.org/stable/3109770.
15. Neuhaus JM, Kalbfleisch JD. Separating between- and within-cluster covariate effects by using conditional and partitioning methods. *Journal of the Royal Statistical Society, Series B* 2006; **68**(5):859–872.
16. Needleman J, Buerhaus P, Mattke S, Stewart M, Zelevinsky K. Nurse-staffing levels and the quality of care in hospitals. *The New England Journal of Medicine* 2002; **346**(22):1715–1722.
17. Leyland AH, Goldstein H (eds). *Multilevel Modelling of Health Statistics*. Wiley: England, 2004.
18. Singer JD, Willett JB. It's about time: using discrete-time survival analysis to study duration and the timing of events. *Journal of Educational and Behavioral Statistics* 1993; **18**(2):155–195. DOI: 10.3102/10769986018002155. Available from: http://dx.doi.org/10.3102/10769986018002155.
19. Lunn DJ, Thomas A, Spiegelhalter D. WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* 2000; **10**:325–337.
20. Plummer M, Best N, Cowles K, Vines K. CODA: output analysis and diagnostics for MCMC. *R Package Version 0.13-3*, 2008.

21. Spiegelhalter DJ, Best NG, Carlin BP, Van der Linde A. Bayesian measures of model complexity and fit (with Discussion). *Journal of the Royal Statistical Society, Series B* 2002; **64**(4):583–616.
22. Hosmer DW, Lemeshow S. *Applied Logistic Regression*. Wiley: New York, 1989.
23. Gelman A, Meng X, Stern H. Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica* 1996; **6**:733–807.
24. Bayarri MJ, Castellanos ME. Bayesian checking of the second levels of hierarchical models. *Statistical Science* 2004; **22**(3):322–343.
25. Marshall EC, Spiegelhalter DJ. Approximate cross-validatory predictive checks in disease mapping models. *Statistical Modelling* 2003; **22**: 1649–1660.
26. Ecochard R, Clayton DG. Fitting complex random effect model with standard software using data augmentation: application to a study of male and female fecundability. *Statistical Modelling* 2002; **1**:319–331.