Revista Colombiana de Estadística - Applied Statistics
July 2022, volume 45, issue 2, pp. 275–300
http://doi.org/10.15446/rce.v45n2.101597

Measuring Individual Benefits of Medical Treatments Using Longitudinal Hospital Data with Non-Ignorable Missing Responses Caused by Patient Discharge: Application to the Study of Benefits of Pain Management Post Spinal Fusion

Medición de los beneficios individuales de tratamientos médicos a partir de datos hospitalarios longitudinales con respuestas faltantes no ignorables causadas por la alta del paciente: Aplicación al estudio de los beneficios del tratamiento contra el dolor después de una fusión espinal

Xuan Zhang[1,a], Nikos Pantazis[2,b], Jose de Leon[3,c], Francisco J. Diaz[4,d]

[1]Boston Strategic Partners, Inc., Boston, MA, United States

[2]Department of Hygiene, Epidemiology and Medical Statistics, National and Kapodistrian University of Athens, Medical School, Athens, Greece

[3]Mental Health Research Center at Eastern State Hospital, University of Kentucky, Lexington, KY, United States

[4]Department of Biostatistics & Data Science, The University of Kansas Medical Center, Kansas City, KS, United States

Abstract

Electronic health records (EHR) provide valuable resources for longitudinal studies and understanding risk factors associated with poor clinical outcomes. However, they may not contain complete follow-ups, and the missing data may not be at random since hospital discharge may depend in part on expected but unrecorded clinical outcomes that occur after patient discharge. These non-ignorable missing data requires appropriate analysis methods. Here, we are interested in measuring and analyzing individual treatment benefits of medical treatments in patients recorded in EHR databases. We present a method for predicting individual

[a]Ph.D. E-mail: xuanzhang2001@yahoo.com
[b]Ph.D. E-mail: npantaz@med.uoa.gr
[c]M.D. E-mail: jdeleon@uky.edu
[d]Ph.D. E-mail: fdiaz@kumc.edu

benefits that handles non-ignorable missingness due to hospital discharge. The longitudinal clinical outcome of interest is modeled simultaneously with the hospital length of stay using a joint mixed-effects model, and individual benefits are predicted through a frequentist approach: the empirical Bayesian approach. We illustrate our approach by assessing individual pain management benefits to patients who underwent spinal fusion surgery. By calculating sample percentiles of empirical Bayes predictors of individual benefits, we examine the evolution of individual benefits over time. We additionally compare these percentiles with percentiles calculated with a Monte Carlo approach. We showed that empirical Bayes predictors of individual benefits do not only allow examining benefits in specific patients but also reflect overall population trends reliably.

***Key words*:** Electronic health records; Empirical Bayesian prediction; Joint mixed models; Non-ignorable missing data; Observational data; Random effects.

Resumen

Los registros de salud electrónicos (RSE) suministran recursos valiosos para estudios longitudinales y para comprender los factores de riesgo asociados con pobres resultados clínicos. Sin embargo, estos podrían no contener seguimientos completos, y los datos faltantes podrían no ser al azar, debido a que el alta hospitalaria puede depender en parte de resultados clínicos esperados pero no registrados que ocurren después de dar de alta al paciente. Esta ausencia de datos no ignorables requiere métodos apropiados de análisis. Aquí estamos interesados en medir y analizar beneficios individuales de tratamientos médicos en pacientes consignados en bases de datos RSE. Proponemos un método para predecir beneficios individuales el cual maneja los datos faltantes debidos al alta hospitalaria. La respuesta clínica longitudinal de interés se modela junto con el tiempo de estadía en el hospital usando un modelo conjunto de efectos mixtos, y los beneficios individuales se predicen por medio de un enfoque frecuentista: el enfoque Bayesiano empírico. Nuestro enfoque es ilustrado evaluando los beneficios individuales del tratamiendo para el dolor en pacientes que fueron sometidos a cirugía de fusión espinal. Aquí examinamos la evolución de los beneficios individuales a través del tiempo mediante el cálculo de los percentiles muestrales de los predictores de Bayes empíricos de los beneficios individuales. También comparamos estos percentiles con percentiles calculados mediante un enfoque Monte Carlo. Los resultados mostraron que los predictores de Bayes empíricos de beneficios individuales no sólo permiten examinar beneficios en pacientes específicos sino que también reflejan confiablemente las tendencias poblacionales globales.

***Palabras clave*:** Datos faltantes no ignorables; Datos observacionales; Efectos aleatorios; Modelos mixtos conjuntos; Predicción Bayesiana empírica;Registros de salud electrónicos.

1. Introduction

It is increasingly recognized that a patient's response to a medical treatment is a statistically heterogeneous phenomenon (de Leon, 2012). Average treatment effects may not accurately represent a heterogeneous population of patients (Ruberg et al. 2010; Gewandter et al. 2019). The benefits each patient receives from the treatment could differ; which implies that treatment benefits should be measured at the patient level (Diaz 2016; 2019; 2021; Zhang et al. 2020; Wang & Diaz 2020). Generalized linear mixed-effects models allow identifying the various sources of variation of patients' responses (Andrews & Cho 2018; Botts et al. 2008; Cho et al. 2017; Diaz & de Leon 2013; Diaz et al. 2007; 2008; 2012a; 2012b; 2013; 2014; Diaz 2018; Senn 2016; Zhu & Qu 2016; Shirafkan et al. 2020), offering an excellent tool for analyzing individual benefits and applications in personalized medicine. In the context of clinical trials, Diaz (2016) has proposed a methodology based on mixed models that allows quantifying and analyzing the individual benefits of medical treatments. He defines an individual benefit as the reduction in disease severity produced by the treatment, whereas disease severity is defined as the probability that the clinical outcome of interest is outside of the therapeutic target. Individual benefits are thus measured on a probability scale. In practice, to assess the benefits of a treatment for a specific patient, empirical Bayes (EB) predictors of the individual benefits are calculated. This frequentist approach has been implemented with both continuous (Diaz, 2019) and binary (Zhang et al., 2020) clinical outcomes and has been proposed as the foundation of treatment individualization methods in N-of-1 clinical trials (Diaz, 2021).

It is also recognized that real-world data such as electronic health records (EHR) collected in a non-randomized setting hold critical value for clinical evidence generation and play a complementary role to clinical trial data (Miksad & Abernethy, 2018). EHR data provide longitudinal follow-up for patient's outcomes. One limitation of EHR data, however, is that there is usually incomplete follow-up due to hospital discharge. Since hospital discharge often depends on the patient's response, non-recorded responses after discharge are nonignorable missing data (Little & Rubin 2002; Pantazis & Touloumi 2010; Albers et al. 2018). This creates a problem for building models to predict treatment benefits because generalized mixed effects models assume missingness at random (Hedeker & Gibbons 2006; Laird 1998). When the missingness is non-ignorable, statistical analyses with these models may be seriously biased if the model does not appropriately incorporate the variable causing the missingness (Touloumi et al., 1999).

This study has three objectives. The first is to extend the methodology for measuring and analyzing individual treatment benefits proposed by Diaz (2016; 2019; 2021) to longitudinal hospital data with non-ignorable missingness. The second objective is to illustrate the methodology with a detailed data analysis of the individual benefits of postoperative pain management in 330 patients who underwent spinal fusion surgery, using EHR data (Cerner HealthFacts®; Kansas City, MO). In our data analysis, we examine estimates of population percentiles of individual benefits to understand population time trends using sample percentiles

of EB predictions of individual benefits. The third objective is to compare our approach based on EB predictors with an alternative approach based on Monte Carlo approximations to the probability distribution of the individual benefits. The goal of this comparison is to illustrate that, in addition to allowing measuring the individual benefits to specific patients, EB predictors allow representing the estimated overall population trends (Zhang et al., 2020).

Here, we propose to analyze individual treatment benefits with hospital data by jointly modeling the patients' responses to the medical treatment and their hospital length of stay (LOS). LOS is also called patient's discharge time. Joint mixed-effects models that combine a generalized linear mixed effects model with a survival model have been used to handle longitudinal clinical trial data with informative drop-outs that produce non-ignorable missings (Schluchter 1992; De Gruttola & Tu 1994; Touloumi et al. 1999; Pantazis & Touloumi 2010; Crowther et al. 2012; Armero et al. 2018; Hickey et al. 2018; Shardell & Ferrucci 2018; Schluchter & Piccorelli 2019; Papageorgiou et al. 2019). In particular, Touloumi, Pocock, Babiker & Darbyshire (1999) developed a method of parameter estimation for joint models that combines restricted iterative generalized least-squares with a nested expectation-maximization algorithm. To our knowledge, these models have not been used to model hospital data, which are unavoidably biased by non-ignorable missingness due to the dependence of hospital discharge on the patient's response.

This study was motivated by the fact that many outcomes of clinical procedures and pharmacological therapies, as well as patient-reported outcomes recorded in longitudinal EHR data are associated with hospital LOS. For instance, laboratory results such as biological markers of acute myocardial infarction (Gronski et al., 2012) or acute kidney injury (Edelstein, 2008), as well as physical or behavioral scores (Shaw et al., 2018), are often measured only during hospital stay and are used in discharge planning and decision making. Another example is patients' self-reported measurements of pain scores to monitor the effectiveness of a surgical procedure, which are available before surgery, or during the hospital stay after the surgery, but are no longer recorded after discharge.

Here, we apply our novel benefit analysis methodology to the quantification and comparison of individual benefits of postoperative pain management. We evaluate the effects of depression and geriatric age (age $>$ 65 years) on patient-reported pain levels. Although lumbar spinal fusion is the top treatment for chronic low back pain and the second most common lower back operation overall, we need a better understanding of how patients' characteristics influence postoperative outcomes (Gaudin et al. 2017; Gerbershagen et al. 2014). Depression is known to be associated with chronic pain such as back pain and is a negative predictor of spinal fusion outcomes (Gaudin et al., 2017). Retrospective cohort studies have found that: 1) patients with pre-existing depression were absent from work for more days after spinal fusion surgery compared to those without depression (Anderson et al., 2015), and 2) preoperative depression influences patient satisfaction independent of the surgery's effectiveness (Adogwa et al., 2013). Patient-reported maximum pain levels on a scale from 0 to 10 are often used as postoperative quality measures to monitor pain relief and track patients' progress after spinal fusion. Studies of risk factors for severe postoperative pain have provided varying results. The risk factors

could be procedure-specific; however, preoperative chronic pain and younger age were associated with higher postoperative pain level independent of the type and extent of the surgery in pooled data from 150 German hospitals (Gerbershagen et al., 2014). In a German registry of knee replacement, older age was associated with lower reported maximum pain levels. On the other hand, the elderly patients did not report less functional impairment caused by pain, suggesting that they tend to underreport their pain levels (Weinmann et al., 2017).

2. Methods

2.1. Joint Model for Observational Longitudinal Continuous Outcomes with Non-ignorable Missingness

Next, we describe a joint multivariate random effects model that simultaneously models a continuous clinical outcome and the hospital LOS. In the application presented in this article, the outcome is a transformed self-reported pain score from patients who underwent lumbar spinal fusion surgery. In general, we assume that we have data from N patients that underwent a medical treatment in the hospital and that we want to measure the individual benefits of the treatment for each patient after t days on treatment. Suppose patient i provided n_i outcome measurements on days $t_{i,1} < \cdots < t_{i,n_i}$ counted from treatment initiation. In our application, day 0 is the day of the surgery and "treatment" refers to pain management after surgery. Let $\boldsymbol{y}_i^* = (y_{i1}, \ldots, y_{in_i})^T$ be a vector containing the outcome measurements for patient i in time order, where y_{i1} is assumed to be measured right before treatment (the baseline measure at day 0), and y_{ij} , $j > 1$, is measured after treatment initiation.

For patient i, let $\boldsymbol{x}_i(t) = (g_0(t),\ g_1(t), \ldots, g_K(t),\ x_{i,1}, \ldots, x_{i,p}\ )^T$ be a vector of covariates corresponding to time t, where $g_0(t), g_1(t), \ldots, g_K(t)$ are functions of time and $x_{i,1}, \ldots, x_{i,p}$ are patient-level covariates (i.e, patient's characteristics). Let $\boldsymbol{z}(t)$ be a vector containing some or all the functions $g_0(t)$, $g_1(t)$,..., $g_K(t)$. Denote $\boldsymbol{x}_{ij} = \boldsymbol{x}_i(t_{ij})$ and $\boldsymbol{z}_{ij} = \boldsymbol{z}(t_{ij})$. For instance, to model the evolution of the outcome over time with a polynomial trend, we may use $g_k(t) = t^k$, $k = 0,\ 1, \ldots, K$. Alternatively, the g_k can be orthogonal polynomials (Zhang et al., 2020).

For patient i, the design matrices for the fixed and random effects of the outcome model are $\boldsymbol{X}_i^* = \left(\boldsymbol{x}_{i,1}, \ldots, \boldsymbol{x}_{i,n_i}\right)^T$ and $\boldsymbol{Z}_i^* = \left(\boldsymbol{z}_{i,1}, \ldots, \boldsymbol{z}_{i,n_i}\right)^T$, respectively (Pantazis & Touloumi, 2010). The model for the clinical outcome variable is

$$\boldsymbol{y}_i^* = \boldsymbol{X}_i^* \boldsymbol{\beta} + \boldsymbol{Z}_i^* \boldsymbol{b}_i + \boldsymbol{e}_i\,, \tag{1}$$

where $\boldsymbol{\beta}$ is a vector of fixed regression coefficients, $\boldsymbol{b}_i$ is a normally distributed vector of random effects with mean 0, and $\boldsymbol{e}_i$ is a vector of residuals for patient i that are assumed to be independent between patients and normally distributed with mean 0 and variance-covariance matrix $\boldsymbol{R}_i^* = \sigma_e^2 I_{n_i}$. Moreover, $\boldsymbol{b}_i$ and $\boldsymbol{e}_i$ are assumed to be independent.

Let L_i^d be the hospital LOS in days for patient i and let $\boldsymbol{x}_i^d = \left(1,\ x_{i1}^d, \ldots, x_{ir}^d\right)^T$ be time-independent patient's characteristics possibly related to LOS. The LOS model is

$$\log\left(L_i^d\right) = \boldsymbol{x}_i^{dT}\boldsymbol{\beta}^d + e_i^d\,, \tag{2}$$

where $\boldsymbol{\beta}^d$ is a vector of fixed regression coefficients and $e_i^d \sim N(0, \sigma_d^2)$ is a residual.

The joint multivariate random effects model combines equations (1) and (2) into

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta}^J + \boldsymbol{Z}_i\boldsymbol{b}_i^J + \boldsymbol{\varepsilon}_i\,, \tag{3}$$

where $\boldsymbol{y}_i = \begin{bmatrix} \boldsymbol{y}_i^* \\ \log\left(L_i^d\right) \end{bmatrix}$, $\boldsymbol{X}_i = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{X}_i^* \\ \boldsymbol{x}_i^{dT} & \boldsymbol{0}^T \end{bmatrix}$, $\boldsymbol{\beta}^J = \begin{bmatrix} \boldsymbol{\beta}^d \\ \boldsymbol{\beta} \end{bmatrix}$, $\boldsymbol{Z}_i = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{Z}_i^* \\ 1 & \boldsymbol{0}^T \end{pmatrix}$, $\boldsymbol{b}_i^J = \begin{bmatrix} e_i^d \\ \boldsymbol{b}_i \end{bmatrix}$, and $\boldsymbol{\varepsilon}_i = \begin{bmatrix} \boldsymbol{e}_i \\ 0 \end{bmatrix}$.

Note that the residual of the LOS model (e_i^d) is treated as a random effect in the joint model. This allows accounting for correlations between this residual and the random effects of the outcome model, which are essentially the cause of non-ignorable missingness.

The model assumes that hospital discharge always occurs right after the last available outcome measurement. Thus, if L_i^d was available in the EHR dataset and $t_{n_i} = L_i^d$ we add a small offset (i.e. 0.01 days) to make discharge time slightly larger than the last outcome measurement time (Pantazis & Touloumi, 2010). But the discharge time is considered censored at $t_{n_i} + 0.01$ if L_i^d is missing in the dataset or if L_i^d is available but $t_{n_i} \leq L_i^d - 1$.

By fitting the joint model (3) to the data provided by the N patients, we calculate maximum likelihood estimators $\hat{\boldsymbol{D}}$, $\hat{\boldsymbol{\beta}}$, $\hat{\sigma}_e$ and $\hat{\sigma}_d$ for $\boldsymbol{D}$, $\boldsymbol{\beta}$, σ_e and σ_d, respectively, where $\boldsymbol{D} = \mathrm{Var}\left(\boldsymbol{b}_i^J\right)$ (Touloumi et al. 1999; Pantazis & Touloumi 2010).

2.2. Individual Disease Severity and Individual Benefits

Here, we use the estimates of the model parameters to predict the individual treatment benefits of the patients that were used to fit the joint model. Following Diaz (2016; 2019; 2021), we define the severity of a patient's disease or condition at a given time point as the probability that the clinical outcome is outside of the therapeutic target. Here, we assume that the therapeutic target is to achieve an outcome $\leq y$, where y is a number prespecified by the clinician. (In the pain management application, the treatment target was defined as achieving a daily maximum pain score ≤ 6.) Therefore, for a specific patient i, the severity of the patient's condition at time t after treatment initiation is

$$s_i\left(t\right) = \Phi\left(\frac{y - y_i\left(t\right)}{\sigma_e}\right), \tag{4}$$

where Φ is the standard normal cumulative distribution function and $y_i(t)$ is the expected outcome at time t for patient i, that is, $y_i(t) = \boldsymbol{x}_i^T\left(t\right)\boldsymbol{\beta} + \boldsymbol{z}^T(t)\boldsymbol{b}_i$.

The individual benefit of the medical treatment for patient i after t days on the treatment is therefore defined as the reduction in disease severity from baseline (Diaz, 2016), that is,

$$b\left(t;\boldsymbol{\beta},\boldsymbol{b}_i,\sigma_e,x_{i,1},\ldots,x_{i,p}\right)=s_i\left(0\right)-s_i\left(t\right). \tag{5}$$

Note that the benefit depends on the number of days that the patient has been under treatment. We multiply this quantity by 100 for convenience. For instance, a benefit of 25% at a specific day after treatment initiation means that the probability that the patient has achieved the desired therapeutic target has increased by 25 units on a probability scale from 0 to 100, using the patient's condition before treatment as a reference point. This approach to measuring the individual benefit of a medical treatment allows modeling well-known clinical phenomena that are relevant to personalized medicine (Diaz 2016; 2019; 2021; Zhang et al. 2020).

2.3. Empirical Bayesian Prediction of Individual Benefits

In practice, to measure the individual benefits of the therapy for patient i, we need the EB predictor $\hat{\boldsymbol{b}}_i$ of the patient's random effects $\boldsymbol{b}_i$ (Diaz 2016; 2019; 2021). Let $\hat{e}_i^d$ denote the EB predictor of e_i^d. These two predictors are jointly obtained with

$$\boldsymbol{b}_{EB,i}^J=\begin{bmatrix}\hat{e}_i^d\\ \hat{\boldsymbol{b}}_i\end{bmatrix}=\hat{\boldsymbol{D}}\boldsymbol{Z}_i^T\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{\varepsilon}}_i\,,$$

where $\hat{\boldsymbol{V}}_i$ is an estimator of $\boldsymbol{V}_i=\mathrm{Var}\left(\boldsymbol{y}_i\right)=\boldsymbol{R}_i+\boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i^T$, with $\boldsymbol{R}_i=\mathrm{Var}\left(\boldsymbol{\varepsilon}_i\right)=\begin{pmatrix}\boldsymbol{R}_i^* & \boldsymbol{0}\\ \boldsymbol{0}^T & 0\end{pmatrix}$, and $\hat{\boldsymbol{\varepsilon}}_i=\begin{bmatrix}\boldsymbol{y}_i^*-\boldsymbol{X}_i^*\hat{\boldsymbol{\beta}}\\ 0\end{bmatrix}$ is the estimated residual vector for patient i.

The last row of $\hat{\boldsymbol{\varepsilon}}_i$ is set to 0 for the calculation of the random effects because the error term of the LOS model (e_i^d) is already included in $\boldsymbol{b}_i^J$. In other words, $\boldsymbol{b}_{EB,i}^J$ includes predictors of both the LOS-model residual and the outcome-model random effects for patient i.

Thus, the EB predictor of the individual benefit for patient i at time t is obtained by replacing the parameters in equation (5) with their estimates, that is, calculating

$$b\left(t;\ \hat{\boldsymbol{\beta}},\hat{\boldsymbol{b}}_i,\hat{\sigma}_e,x_{i,1},\ldots,x_{i,p}\right). \tag{6}$$

2.4. Application: Effects of Depression and Age on Individual Benefits of Pain Management Post Spinal Fusion

2.4.1. Data Source and Study Subjects

The EHR dataset (Cerner Health Facts®, Kansas City, MO) is deidentified and has been used in previously published articles (Shaw et al. 2018; Urman et al. 2018). An Institutional Review Board (IRB) exemption for this study was granted by Western IRB (Olympia, WA) and the IRB from the University of Kansas Medical Center agreed with this exemption. We selected adult inpatients undergoing spinal fusion surgery in a university general hospital between January 1, 2014, and December 31, 2015, using International Classification of Diseases (ICD) -9 codes 81.00 to 81.08 and corresponding ICD-10 codes. The hospital had more than 500 beds at that time and is in an urban area in the Northeast census region of the United States.

Additional inclusion criteria were 1) at least one pain score on and after the day of surgery (day 0); 2) a maximum score on day 0 between 7 and 10 inclusive; 3) 1 to 5 days post-surgical hospital stay; and 4) at least 6 months of history captured in the database prior to the surgery. We identified 330 patients who satisfied the inclusion criteria (Table 1).

2.4.2. Pain Assessments

Numerical patient-reported pain scores ranged from 0 to 10, with 0 indicating no pain and 10 indicating the most severe pain. The outcome of interest was the patient's maximum daily score, obtained at day 0 and during 1 to 5 days of post-surgical hospital stay (Table 2). Since patients' pain levels were not measured after discharge, this longitudinal observational study conveys the challenges of a highly unbalanced dataset caused by non-random missing data. Since pain scores are usually lower on or after the discharge day, the assumption that missing data would be random, which is required by standard longitudinal statistical models, is violated (Ibrahim & Molenberghs, 2009).

2.4.3. Depression Assessments

Depression comorbidity was defined as having an ICD-9 code (3004, 30112, 3090, 3091, or 311) or an ICD-10 code (F320, F321, F322, F323, F328, F3281, F3289, F329, F330, F331, F332, F333, F338, F339, F341, or F4321) during the hospital stay or within 6 months before admission, or having received antidepressants during the stay (Table 1).

TABLE 1: Demographics and clinical characteristics of 330 patients who underwent spinal fusion surgery.

	Mean	SD
Age (years)	53.9	12.4
	%	
GERIATRIC AGE (>65 years)		
Yes	18 (59/330)	
No	82 (271/330)	
GENDER		
Female	52 (173/330)	
Male	48 (157/330)	
RACE		
Caucasian	93 (308/330)	
African American	2 (7/330)	
Other	5 (15/330)	
PAIN MEDICATION		
Opioids and acetaminophen	78 (257/330)	
Opioids, NSAIDs and acetaminophen	18 (60/330)	
Opioids only	4 (12/330)	
Opioids and NSAIDs	<1 (1/330)	
DEPRESSION[a]		
Yes	46 (150/330)	
No	54 (180/330)	
ANTIDEPRESSANT MEDICATION		
Taking antidepressants	81 (121/150)	
SSRI	34 (51/150)	
SNRI	11 (17/150)	
Other[a]	9 (13/150)	
SSRI and other	7 (11/150)	
SSRI and TCA	5 (8/150)	
TCA	5 (7/150)	
SNRI and TCA	4 (6/150)	
SSRI and SNRI	1 (2/150)	
SNRI and other	1 (2/150)	
MAOI	<1 (1/150)	
Other and TCA	<1 (1/150)	
SSRI, SNRI and other	<1 (1/150)	
SSRI, other and TCA	<1 (1/150)	

Abbreviations: MAOI, monoamine oxidase inhibitor; NSAID, nonsteroidal anti-inflammatory drug; SD, standard deviation; SNRI, serotonin norepinephrine reuptake inhibitor; SSRI, serotonin selective reuptake inhibitor; TCA, tricyclic antidepressant.

[a]The EHR database did not itemize the medications in the "Other" category.

2.4.4. Hospital Length of Stay

The patients' hospital LOS after surgery may be affected by their characteristics and responses to postoperative pain management. Pain levels are usually not measured after discharge and even when patients are in the hospital their pain measurements may be terminated for various reasons. Thus, hospital LOS was included in the analyses and was defined as the number of days elapsed from surgery to hospital discharge.

In most patients (326/330), the last pain score was observed on the day of the discharge. In the remaining patients, it was determined before the day of discharge and the outcome was therefore considered censored on the day of the last pain score measurement. An offset of 0.01 was added to both LOS and censoring times

to make them slightly larger than the time of the last pain score (Pantazis & Touloumi, 2010).

TABLE 2: Mean and SD of stratified maximum baseline pain scores and hospital LOS after surgery in 330 patients who underwent spinal fusion surgery.

	Baseline pain scores		LOS			
	Mean	SD	Mean	SD	Min	Max
All (N=330)	8.65	1.11	1.62	1.00	1.0	5.0
GERIATRIC AGE[a]						
Yes (N=59)	8.31	0.99	1.98	1.17	1.0	5.0
No (N=271)	8.72	1.12	1.54	0.94	1.0	5.0
GENDER						
Female (N=173)	8.66	1.11	1.64	1.03	1.0	5.0
Male (N=157)	8.62	1.11	1.59	0.97	1.0	5.0
RACE						
Caucasian (N=308)	8.62	1.11	1.61	1.00	1.0	5.0
African American (N=7)	8.71	1.25	1.71	1.50	1.0	5.0
Other (N=15)	9.20	1.21	1.67	0.62	1.0	5.0
DEPRESSION						
Yes (N=150)	8.81	1.13	1.67	1.03	1.0	5.0
No (N=180)	8.51	1.07	1.57	0.97	1.0	5.0

Abbreviations: LOS, length of stay; SD, standard deviation.
[a]Geriatric age was defined as age >65 years.

2.4.5. Transformation of Pain Scores

We transformed the maximum daily pain scores before data analyses using a transformation that we call "discrete logit transformation", given by

$$y_{ij} = T\left(\text{Pain Score}_{ij}\right) = \log\left(\frac{\text{Pain Score}_{ij} + 1}{m + 1 - \text{Pain Score}_{ij}}\right), \tag{7}$$

where m is the maximum value of the pain scale ($m = 10$ in this application), and Pain Score_{ij} is the maximum daily pain score for patient i at day t_{ij}. This transformation allows accounting for possible floor effects — we expected the pain scores of many patients to decrease over time and stabilize at a minimum value due to the pain management therapy. Moreover, since the distribution of the pain scores was highly skewed with higher frequencies for severe pain scores, this transformation helped improve the goodness-of-fit of the outcome model. After the transformation, the distribution of the EB predictor of the LOS model residuals and those of the random intercept and random effect of time for the pain score model were relatively normal, suggesting a good model fit.

The transformation in equation (7) can be viewed as an adaptation of the logistic transformation to a bounded continuous latent variable measured with *finite precision* equal to 1. In fact, here, we imagine a patient's pain level as a continuous latent variable, which has been measured with integer numbers from 0 to 10. The ability of the logistic transformation to produce approximate normal distributions was originally investigated by Johnson (1949) for unbounded and bounded outcomes. The bounded case was further investigated by Lesaffre, who also applied the transformation to discrete measures of continuous latent variables.

In the context of our study, using transformation (7) is essentially equivalent to assuming that, conditional on the random effects, the distribution of the pain scores can be approximated by a scaled distribution in the family of logit-normal distributions, which is a very flexible family that includes distributions of many different shapes including right-skewed, left-skewed, symmetric and bathtub shapes (Lesaffre et al., 2007).

To calculate individual benefits, we assumed that the clinician wants to achieve pain scores ≤ 6. Since daily maximum pain scores were transformed using the discrete logit transformation, the therapeutic target was therefore to reach a transformed pain score $\leq T(6) = 0.3365$. This was the value of y in equation (4).

2.4.6. Joint model for Pain Scores and Hospital Discharge Day

We built a joint multivariate random-effects model, which is a generalized linear mixed-effects model that accounts for non-ignorable missingness (Touloumi et al. 1999; Pantazis & Touloumi 2010). The model combined a model of daily maximum pain scores with a model of LOS to account for the missing pain measurements caused by hospital discharge (Table 3). The simultaneous modeling of LOS and pain took into consideration the correlations between them. It reduced the bias associated with the unbalanced data, providing more accurate estimation of the effects of the covariates on pain scores (Touloumi et al., 1999).

To build the joint model, we followed a backward selection procedure that locked the time variable and was based on both p-values from Wald tests and the examination of covariate effect sizes at each step (Woodward 2014; Hedeker & Gibbons 2006). The initial pain and LOS models included geriatric age (1 if age $>$ 65 years, 0 otherwise), depression (1 if the patient had depression comorbidity, 0 otherwise), and potential confounders available in the database including gender, race, and some comorbidities (rheumatoid arthritis, diabetes, liver disease, renal failure, peripheral vascular disease, and other neurological disorders). A variable was kept in both the pain and LOS models if it was significant in only one model or in both at a 0.05 significance level.

The variables included in the final pain model were geriatric age, depression, time as the number of days after surgery, and the interaction between depression and time. The transformed pain scores followed a linear time trend. The intercept and the time slope were considered random, meaning they were different for each patient (Diaz 2016; 2019).

Results of the LOS model are also provided in Table 3. It was assumed that the random residual of the LOS model was correlated with both the random intercept and random time slope of the pain model. Gender, race, and the examined comorbidities did not exhibit any significant effects on pain scores or LOS and were therefore not included in the final model.

In matrix notation, if x_{i1} and x_{i2} denote the dichotomous variables geriatric age and depression, respectively, then $\boldsymbol{X}_i^* = [\mathbf{1}\ X_{i1}\ X_{i2}\ \boldsymbol{t}_i\ X_{i3}]$, where $X_{i1} = (x_{i1}, \ldots, x_{i1})^T$, $X_{i2} = (x_{i2}, \ldots, x_{i2})^T$, $\boldsymbol{t}_i = (t_{i1}, \ldots, t_{in_i})^T$ is a vector containing

the days from surgery on which the pain scores were observed for patient i, $X_{i3} = (x_{i2}t_{i1}, \ldots, x_{i2}t_{in_i})^T$ is the interaction between depression and time, and $\mathbf{1}$ is an $n_i \times 1$ vector of ones. In the LOS model, the design vector was $\boldsymbol{x}_i^d = (1, x_{i1}, x_{i2})$.

In summary, the pain model included a random intercept and a random slope for time and had the form

$$\begin{aligned}\text{Transformed Scores}_{ij} &= \beta_0 + \beta_1 \times (\text{Geriatric Age})_i + \beta_2 \times \text{Depression}_i \\ &\quad + \beta_3 \times \text{Time}_{ij} + \beta_4 \times \text{Depression}_i \times \text{Time}_{ij} + b_{0i} + b_{1i} \times \text{Time}_{ij} + e_{ij}\,,\end{aligned}$$

where e_{ij} indicates the residuals for the pain score model for patient i at occasion j which has mean 0 and residual variance σ_e^2 (Table 3). The parameters β_k, $k = 1, \ldots, 4$, are population-average effects (the fixed effects), whereas b_{0i} and b_{1i} are parameters specific to patient i that represent deviations from the corresponding population averages (the random effects). And the model of LOS for patient i had the form

$$\log(\text{LOS}_i) = \beta_0^d + \beta_1^d \times (\text{Geriatric Age})_i + \beta_2^d \times \text{Depression}_i + e_i^d\,,$$

where e_i^d is a random residual following a normal distribution with mean 0.

We fitted the joint model using the `jmre1` command of Stata and calculated $\boldsymbol{b}_{EB,i}^J$ with the command "predict" (StataCorp LLC, College Station, TX) (Table 3) (Pantazis & Touloumi, 2010). To test the significance of a correlation between two random effects, we computed a likelihood ratio test that compared the final model with a model for which the corresponding covariance was set to 0.

2.4.7. Individual Benefits of Pain Management

The severity of the patient's disease was defined as the probability of being outside of the pain treatment target, which in turn was defined as a daily maximum pain score ≤ 6 (Diaz 2016; 2019). The patient's individual treatment benefit was therefore the decrease in disease severity from baseline ($\times 100$). To examine how much benefit patients received from postoperative pain management during the 5 days after spinal fusion, we predicted the patients' individual benefits. The estimated random effects of a patient were used to predict his/her benefits, combining the patient's available data with parameter estimates in Table 3, using formula (6).

For each patient, individual benefits were predicted from day 0.2 to day 5 by 0.2-day increments, by interpolation with formula (6). In fact, although the time variable in the EHR data was available only in the form of integer days $0, 1, \ldots, 5$, formula (6) allowed predicting benefits for any non-integer time point from 0 to 5 days (Diaz 2016; 2019). Median, 25^{th} and 75^{th} percentiles of individual benefits were calculated. For each of the 4 groups determined by age and depression status, these statistics were plotted in Figure 1. They are also presented in Table 4 for days 1 through 5.

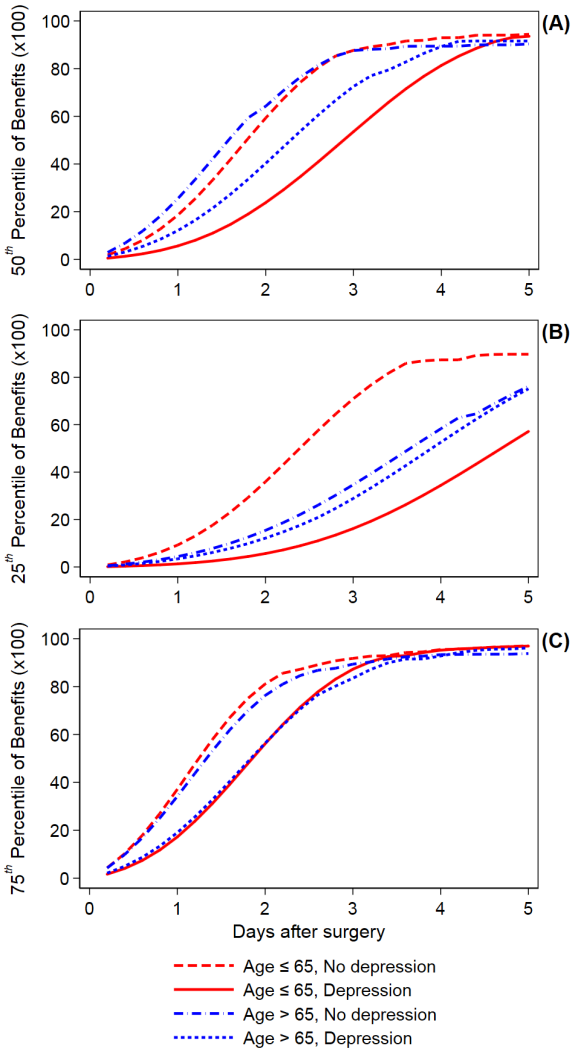

FIGURE 1: Predicted evolution of individual pain management benefits (x100) after spinal fusion surgery over days 1 through 5 for 330 patients. For a particular patient, predicted individual treatment benefits were obtained by combining the patient's data with the parameter estimates in Table 3. (A), (B), and (C) show 50th, 25th and 75th percentiles of individual benefits, respectively.

2.5. Monte Carlo Estimation of Quartiles of Individual Benefits

For comparison purposes, we implemented Monte Carlo computations as an alternative approach for estimating the population quartiles of the probability distribution of individual benefits for the four subpopulations determined by

TABLE 3: Joint random-effects model of transformed daily maximum pain scores and hospital length of stay for 330 patients after spinal fusion surgery.

Parameter name	Parameter estimate (SE)	p-value
Fixed effects for transformed pain scores		
Pain score intercept, β_0	1.4704 (0.0544)	<0.001
Geriatric age,[a] β_1	−0.1853 (0.0931)	0.047
Depression,[b] β_2	0.2478 (0.0742)	0.001
Time (days),[c] β_3	−0.6771 (0.0461)	<0.001
Interaction between depression and time, β_4	0.1327 (0.0663)	0.045
Fixed effects for LOS (days)		
LOS intercept, β_0^d	0.2465 (0.0385)	<0.001
Geriatric age,[a] β_1^d	0.2196 (0.0644)	0.001
Depression,[b] β_2^d	0.0886 (0.0532)	0.096
Correlations between random effects		
LOS residual and pain score intercept, $\mathrm{Corr}(e_i^d, b_{0i})$	0.4985	<0.001
LOS residual and time, $\mathrm{Corr}(e_i^d, b_{1i})$	0.6631	<0.001
Pain score intercept and time, $\mathrm{Corr}(b_{0i}, b_{1i})$	0.8471	<0.001
Variances of random effects		
LOS residual, σ_d^2	0.2281	--
Pain score intercept, $\mathrm{Var}(b_{0i})$	0.1384	--
Time,[c] $\mathrm{Var}(b_{1i})$	0.0916	--
Residual variance for pain model, σ_e^2	0.3835	--

Abbreviations: SE, standard error; LOS, length of stay.

[a]The dichotomous covariate geriatric age was defined as 1 if the age of the patient was >65, and 0 otherwise.

[b]The dichotomous covariate depression was defined as 1 if the patient had a record of depression diagnosis or was under antidepressants, and 0 otherwise.

[c]Time was defined as days post-spinal fusion surgery.

geriatric age and depression status. In contrast to the approach that we used to calculate Table 4 and Figure 1, this approach does not require EB prediction, but it is more computationally demanding and does not allow examining individual benefits for specific patients in the EHR data (Zhang et al., 2020). The Monte Carlo approach, however, is useful as a reference standard for investigating whether EB individual benefit predictors reflect estimated overall population trends reliably (Zhang et al., 2020). Ideally, sample quartiles of EB predicted benefits should be close to Monte Carlo estimates of population quartiles. The Monte Carlo approach to the analysis of individual benefits has been previously used to validate the EB approach in the context of longitudinal binary outcomes (Zhang et al., 2020).

TABLE 4: Sample medians (and first and third quartiles) of individual benefits (× 100) of postoperative pain management on days 1 through 5 for 330 patients after spinal fusion. a Empirical Bayesian predictors of the patient's random effects were used for predicting treatment benefits, combining the patient's data with the parameter estimates given in Table 3.

Study group	Day 1	Day 2	Day 3	Day 4	Day 5
Age ≤ 65 and no depression (N=191)	18.6 (9.3, 37.1)	59.3 (35.9, 81.1)	87.7 (70.9, 91.8)	93.0 (87.3, 95.5)	94.4 (89.7, 97.0)
Age ≤ 65 and depression (N=80)	5.6 (1.3, 17.3)	23.8 (5.7, 56.1)	53.6 (16.2, 87.3)	81.3 (34.4, 95.2)	93.6 (57.2, 96.9)
Age > 65 and no depression (N=49)	25.5 (4.4, 34.2)	64.2 (15.4, 76.3)	87.6 (34.6, 89.4)	89.4 (58.3, 93.3)	90.4 (76.2, 93.8)
Age > 65 and depression (N=10)	12.1 (3.5, 19.1)	40.3 (12.2, 56.4)	72.5 (28.9, 83.5)	89.3 (52.6, 92.8)	91.6 (75.1, 96.1)

[a]The individual benefit is the decrease in disease severity from baseline, where disease severity is the probability of being outside the treatment target (Diaz 2016, 2019). The treatment target was defined as achieving a maximum daily pain score ≤6. After transforming the scores with a discrete logit transformation, the treatment target corresponded to a transformed pain score ≤0.3365.

In the context of the joint model for pain scores and LOS, the Monte Carlo estimates of population quartiles are in Table 6 and were produced with the following algorithm:

1. Draw 1,000 random effects for each subpopulation of patients from the estimated distribution of the random effects $\boldsymbol{b}_i$ of the pain score model, using the parameter estimates in Table 3. Each of these random effects vectors represents a simulated *new* patient (Diaz 2016; 2019; Zhang et al. 2020).

2. Generate the random intercept and random coefficient of time of the pain score model for the 1,000 simulated patients in each subpopulation by adding up the random effects to their corresponding estimated fixed effects given in Table 3.

3. Calculate benefits for each of the 1,000 patients on days 1 through 5 post-surgery by using the formula $b(t; \hat{\boldsymbol{\beta}}, \boldsymbol{b}_i, \hat{\sigma}_e,\ x_{i,1}, \ldots, x_{i,p})$.

4. For each of the four subpopulations, calculate the median and the 25^{th} and 75^{th} percentiles of the benefits of the 1,000 patients for days 1 through 5 post-surgery.

The above Monte Carlo approach is a numerical method for calculating population percentiles of the distribution of individual benefits that utilizes the model parameter estimates in Table 3 (Zhang et al., 2020). By simulating many patients, we take advantage of the law of large numbers to numerically approximate the inverse of the cumulative distribution function of individual benefits. Thus,

with this approach, we solve a deterministic numerical problem using random number generation.

Note that the Monte Carlo approach is not an empirical Bayesian approach. In fact, the former is not mediated by the calculation of the EB predictors given by Formula (6). That is, we do not need to estimate the benefits of specific individuals to implement the Monte Carlo approach. In this sense, the Monte Carlo approach to the estimation of population percentiles is an alternative method that is useful only to analyze population trends and does not allow for examining the benefits of specific patients. In contrast, in the EB approach described in Section 2.4.7., we predict the individual benefits of each of the patients in the database and then calculate the *sample* percentiles of these benefits as if these were a random sample of true benefits. Showing that the Monte Carlo and EB approaches yield similar estimates of population percentiles of individual benefits can be regarded as evidence that the EB approach produces reliable estimates (Zhang et al., 2020).

3. Results

3.1. Patient Characteristics

Patients' characteristics, pain and antidepressant medications are in Table 1. Almost half (46%) the patients had comorbid depression. Depression was more frequent in females (54%, 94/173) than in males (36%, 56/157) and in non-geriatric patients (49%, 132/271) than in geriatric patients (31%, 18/59). Baseline pain scores and hospital LOS are in Table 1.

3.2. Correlations Between LOS Model Residuals and Pain Model Random Effects

Table 3 shows the correlations between the random effects of the pain model and the residuals of the LOS model. The correlations were significantly different from 0. Specifically, the correlation between the LOS residuals and the random intercept of the pain model was 0.4985 ($p < 0.001$), whereas the correlation between the LOS residuals and the random effect of time was 0.6631 ($p < 0.001$). These significant correlations confirm the expected dependency of hospital discharge on pain levels and, therefore, suggest that the pain scores that were not recorded after discharge are nonignorable missing data. Our simultaneous modeling of pain scores and hospital LOS is therefore justified.

3.3. Joint Model and the Impact of Depression and Age on Pain Scores and LOS

We found significant correlations between 1) high baseline pain scores and longer postoperative LOS ($r = 0.4985$, $p < 0.001$; Table 3), 2) slower pain reduction and longer LOS ($r = 0.6631$, $p < 0.001$), and 3) high baseline pain

scores and slower pain reduction post-surgery ($r = 0.8471$, $p < 0.001$). These correlations indicated that 1) patients who had higher baseline pain scores tended to stay longer after surgery; 2) patients whose pain decreased more slowly after surgery tended to stay longer; and 3) patients with higher pain scores at baseline tended to have slower pain reduction after surgery.

The pain model suggested that, on average: 1) a preoperative record of depression was significantly associated with higher baseline pain scores ($p = 0.001$; Table 3); 2) geriatric age was significantly associated with lower baseline pain scores ($p = 0.047$); and 3) a significant interaction existed between depression and time (parameter estimate = 0.1327, $p = 0.045$), meaning that patients with depression had significantly slower pain reduction after surgery.

The LOS model suggested that, on average: 1) geriatric age was significantly associated with longer LOS ($p = 0.001$; Table 3); and 2) depression tended to be associated with a slightly longer LOS, although it did not reach significance ($p = 0.096$).

3.4. Impact of Depression and Age on Individual Benefits of Postoperative Pain Management

Although treatment benefits tended to increase over time for all four patient groups, the amount and rate of change of achieved benefits varied across groups (Table 4). For instance, at day 1, in non-geriatric patients without depression the median decrease in disease severity was 18.6% probability units compared to 5.6% in non-geriatric patients with depression.

By day 5, the median achieved benefits were comparable for patients with or without depression. However, the first quartiles for patients with depression tended to be smaller than those for non-depressed patients in the comparable age group at specific times, indicating that there were more patients with depression receiving small benefits than patients without depression after controlling for age and time.

For patients with depression, non-geriatric age was associated with slower individual benefit development. For instance, in geriatric patients with depression, the median decrease in disease severity was 12.1% probability units at day 1 compared to 5.6% for non-geriatric patients with depression. On day 5, the first quartile for geriatric patients with depression (75.1%) was higher than that for non-geriatric patients with depression (57.2%). Figure 1A illustrates that, for average patients with depression, non-geriatric age was associated with smaller benefits, compared to geriatric age. In general, average patients with depression had much smaller benefits after controlling for age.

In patients receiving the least benefit from pain management the combination of depression and non-geriatric age was associated with the slowest responses, whereas non-geriatric age without depression was associated with the fastest responses (Figure 1B). Interestingly, for the patients achieving the greatest benefits (Figure 1C), individual benefits were more clearly affected by depression comorbidity than by age.

TABLE 5: Hospital LOS (in days) for study patients grouped by quartiles of individual pain management benefits at day 1 post-spinal fusion surgery. Individual benefits were calculated with the empirical Bayes approach.

Individual Benefits	N	Mean (SD)	Median	Minimum	Maximum
1st quartile (0 to 3.28%)	82	2.44 (1.25)	2	1	5
2nd quartile (3.29 to 13.64%)	83	1.69 (0.96)	1	1	5
3rd quartile (13.65 to 27.92%)	84	1.19 (0.50)	1	1	3
4th quartile (≥27.93%)	81	1.15 (0.45)	1	1	3

Abbreviations: LOS, length of stay; SD, standard deviation.

3.5. Individual Benefits One Day after Surgery as Predictors of LOS

To examine whether levels of individual benefits from post-surgery pain management achieved after 1 day are predictive of hospital LOS, we compared the LOS of patients from among the 1^{st}, 2^{nd}, 3^{rd}, and 4^{th} quartiles for individual benefits (Table 5). Patients with higher immediate benefits tended to have shorter LOS.

3.6. Comparison of Quartiles of Empirical Bayesian Individual Estimates with Monte Carlo Quartile Estimates

The patterns of the evolution of individual antipsychotic benefits over time shown by the Monte Carlo quartile estimates (Table 6) were very similar to those observed when the sample quartiles of EB benefit predictors were used (Table 4). For instance, in patients with depression, those with non-geriatric age obtained the benefits of pain therapy more slowly than older patients. Moreover, other things being equal, patients with depression tended to obtain smaller benefits than patients without depression. The first and third quartiles presented in Table 6 also allow conclusions like those obtained from Table 4 described in Section 3.4. The similarities between Tables 4 and 6 suggest the adequacy of EB predictors of individual benefits for detecting overall clinical population trends (Zhang et al., 2020). EB predictors, however, have the additional advantage that they allow examining the benefits achieved by single patients, which makes them useful for potential applications in clinical practice (Diaz 2016; 2019; 2021).

Interestingly, the interquartile ranges of the EB predictions of individual benefits tended to be shorter than those calculated through the Monte Carlo approach (Tables 4 and 6). This should not be surprising, however, because EB predictors tend to reflect the shrinkage property of best linear unbiased predictors (BLUPs) of random effects (Frees, 2004).

3.7. Limitations of the Application

Our data analysis included patients with severe pain (≥7) at baseline and at least one pain score and who stayed at least 1 day in the hospital. These

TABLE 6: Estimates of medians (and first and third quartiles) of individual benefits ($\times$ 100) of postoperative pain management on days 1 through 5 after spinal fusion, obtained with Monte Carlo computation. The model in Table 3 was used for simulating 1,000 patients in each study group.

Study group	Day 1	Day 2	Day 3	Day 4	Day 5
Age $\leq$ 65, no depression	18.5 (5.4, 39.8)	57.6 (21.6, 79.1)	82.8 (49.2, 90.6)	89.5 (72.1, 94.4)	92.4 (82.0, 96.3)
Age $\leq$ 65, depression	5.8 (1.3, 21.8)	28.5 (5.4, 65.4)	62.3 (14.7, 88.7)	85.1 (29.8, 94.5)	92.2 (51.1, 96.7)
Age > 65, no depression	25.1 (8.9, 46.2)	63.7 (30.3, 77.6)	80.1 (56.3, 87.8)	86.0 (71.7, 92.1)	88.9 (78.2, 93.8)
Age > 65, depression	11.0 (2.5, 28.9)	38.0 (9.2, 70.8)	69.4 (22.2, 87.4)	84.9 (40.5, 92.8)	89.7 (60.9, 94.9)

criteria may have excluded less severe cases so our results cannot be extrapolated to them. Moreover, the pain scores used in this study were self-reported. Patient-reported measures such as pain scores and levels of satisfaction are important measures for evaluating treatment effects (Lotzke et al., 2016). They could be biased, however, since each patient may have different levels of sensitivity and expectation. However, predictors of individual benefits, which are on a probability scale, compare baseline pain severity with post-treatment severity within a patient, canceling out potential individual biases in the perception of pain.

We did not find evidence that gender or race had significant effects on pain scores or LOS. We cannot rule out, however, that other variables that were not available in the EHR database such as surgeon may have had potentially confounding effects. Moreover, 99% of the patients who had health insurance information had some form of insurance covering. Thus, the lack of insurance could not be a confounding variable in these analyses.

Unfortunately, it was not possible to include the type of pain treatment as a covariate in the joint model because all patients were on opioids and nearly all (226 out of 330) received an additional pain medication (Table 1). Thus, it was not possible to investigate whether pain treatment choice interacted with depression or antidepressant use and impacted pain or LOS.

4. Discussion

Electronic health records (EHR) provide valuable resources for longitudinal studies and understanding risk factors associated with poor clinical outcomes. However, they may not provide complete follow-up, and the missing data are not at random since hospital discharge may depend in part on expected but unrecorded clinical outcomes after discharge (Ibrahim & Molenberghs, 2009). This "non-ignorable missingness" requires appropriate statistical techniques (Pantazis

& Touloumi, 2010). Ignoring the unbalanced nature of longitudinal EHR data may lead to serious bias (Albers et al., 2018).

In this paper, we extended the methods for individual treatment benefit prediction based on mixed-effects models proposed by Diaz (2016; 2019; 2021) to allow for non-ignorable missingness in the longitudinal data. Modeling informative drop-out in the analysis of clinical trial data with some patients dropping out of the study after randomization is common in statistical practice (Pantazis & Touloumi, 2010); however, our proposal of extending this concept to real-world hospital data for which the follow-up data are incomplete due to hospital discharge is novel. This is the first paper to analyze individual treatment benefits as defined by Diaz (2016) using EHR data. Since EHR data are becoming more and more important in clinical evidence generation, this work offers a useful way of analyzing treatment effects from a personalized medicine perspective.

Generalized linear mixed-effects modeling of longitudinal outcomes is a statistical approach useful for predicting individuated treatment benefits (Diaz 2016; 2019; 2021; Zhang et al. 2020; Wang & Diaz 2020), which takes into consideration the heterogeneity of patients' characteristics including unknown traits. While traditional statistical analyses focus on average treatment effects, mixed-effects modeling can analyze the variation of treatment effects in individual patients. Moreover, in the joint mixed modeling approach followed in this work, the correlations between longitudinal outcomes and hospital LOS that cause non-ignorable missingness are taken into consideration, leading to more reliable and accurate estimation of the parameters of the outcome model (Touloumi et al., 1999). More accurate parameter estimates in turn lead to better predictions of individual treatment benefits.

In our application, longitudinal pain score data from patients undergoing spinal fusion surgery were modeled simultaneously with post-surgical LOS, and then individual benefits of pain management were measured in accordance with the definition of treatment benefits proposed by Diaz (2016; 2019; 2021). Our application showed that it is important to further understand the impact of patients' characteristics such as preoperative depression and age upon individual benefits of pain management after spinal fusion surgery. We compared individual benefits of pain management among four groups of patients determined by depression diagnosis and age. An examination of median benefits was not enough, and other subgroups of individuals emerged (Figure 1). In "average" patients, age played an important role in those with depression, who were prone to receive less benefit (Figure 1A). In contrast, among patients tending to receive the least benefit (Figure 1B), younger patients without depression achieved some benefit quicker than geriatric patients with depression, whereas younger patients with depression benefitted least from pain management. Moreover, among patients achieving the highest benefits (Figure 1C), the effect of age on treatment benefits was negligible compared to the effect of depression. Our finding that the effect of age is unimportant in patients receiving high benefits is consistent with a previous study that found that, although elderly patients reported lower pain scores post-total knee replacement, their functional impairment caused by pain did not differ from younger patients (Weinmann et al., 2017).

Interestingly, patients who received less benefit from one day of post-surgery pain management tended to stay longer at the hospital (Table 5), suggesting that early benefit measurements may serve as predictors of hospital LOS after surgery.

To rule out the possibility that antidepressant medication explains the observed slower response to pain management in patients with depression, we fitted an additional joint mixed model using only patients with depression, similar to the model in Table 3 except that the depression variable was replaced by antidepressant use (data not shown). Antidepressants were not significantly associated with baseline pain scores or pain management response. Thus, the slow response to pain management in patients with depression may be due to the comorbidity itself instead of antidepressant medication.

Acknowledgements

The authors thank Lorraine Maw, M.A., from the Mental Health Research Center at Eastern State Hospital, Lexington, KY, USA, for editorial assistance. They also thank Drs. Jonathan Mahnken, Jo Wick, and Milind Phadnis from the University of Kansas Medical Center, KS, USA, for their comments and suggestions. The authors thank two anonymous reviewers for comments and suggestions that contributed to improve the quality and contents of the paper. No commercial organizations had any role in writing this paper for publication. Dr. Zhang is a consultant at Boston Strategic Partners, Inc. (BSP), Boston, MA, but received no financial support from BSP for this study. None of the authors had conflict of interest in the last 36 months. None of the authors is or has been an employee or a paid contractor of Cerner Corporation, the owner of the pain database used in this article, and none of them was involved in the collection of the data contained in the database.

Availability of Data

The pain data are available as the Cerner Health Facts® database from the Cerner Corporation, which is available to the public through licensure with Cerner Corporation (Kansas City, USA).

[Received: June 2021 — Accepted: February 2022]

References

Adogwa, O., Parker, S. L., Shau, D. N., Mendenhall, S. K., Bydon, A., Cheng, J. S., Asher, A. L. & McGirt, M. J. (2013), 'Preoperative Zung depression scale predicts patient satisfaction independent of the extent of improvement after revision lumbar surgery', *The Spine Journal* **13**, 501–506.

Albers, D. J., Elhadad, N., Claassen, J., Perotte, R., Goldstein, A. & Hripcsak, G. (2018), 'Estimating summary statistics for electronic health record laboratory data for use in high-throughput phenotyping algorithms', *Journal of Biomedical Informatics* **78**, 87–101.

Anderson, J. T., Haas, A. R., Percy, R., Woods, S. T., Ahn, U. M. & Ahn, N. U. (2015), 'Clinical depression is a strong predictor of poor lumbar fusion outcomes among workers' compensation subjects', *Spine* **40**, 748–756.

Andrews, N. & Cho, H. (2018), 'Validating effectiveness of subgroup identification for longitudinal data', *Statistics in Medicine* **37**, 98–106.

Armero, C., Forte, A., Perpiñán, H., Sanahuja, M. J. & Agustí, S. (2018), 'Bayesian joint modeling for assessing the progression of chronic kidney disease in children', *Statistical Methods in Medical Research* **27**, 298–311.

Arnold, L. M., Palmer, R. H., Gendreau, R. M. & Chen, W. (2012), 'Relationships among pain, depressed mood, and global status in fibromyalgia patients: post hoc analyses of a randomized, placebo-controlled trial of milnacipran', *Psychosomatics* **53**, 371–379.

Botts, S., Diaz, F. J., Santoro, V., Spina, E., Muscatello, M. R., Cogollo, M., Castro, F. E. & de Leon, J. (2008), 'Estimating the effects of co-medications on plasma olanzapine concentrations by using a mixed model', *Progress in Neuro-Psychopharmacology & Biological Psychiatry* **32**, 1453–1458.

Cho, H., Wang, P. & Qu, A. (2017), 'Personalized treatment for longitudinal data using unspecified random-effects model', *Statistica Sinica* **27**, 187–205.

Crowther, M. J., Abrams, K. R. & Lambert, P. C. (2012), 'Flexible parametric joint modelling of longitudinal and survival', *Statistics in Medicine* **31**, 4456–4471.

De Gruttola, V. & Tu, X. M. (1994), 'Modelling progression of CD4-lymphocyte count and its relationship to survival time', *Biometrics* **50**, 1003–1014.

de Leon, J. (2012), 'Evidence-based medicine versus personalized medicine: are they enemies?', *Journal of Clinical Psychopharmacology* **32**, 153–164.

Diaz, F. J. (2016), 'Measuring the individual benefit of a medical or behavioral treatment using generalized linear mixed-effects models', *Statistics in Medicine* **35**, 4077–4092.

Diaz, F. J. (2018), 'Construction of the design matrix for generalized linear mixed-effects models in the context of clinical trials of treatment sequences', *Revista Colombiana de Estadística* **41**, 191–233.

Diaz, F. J. (2019), 'Estimating individual benefits of medical or behavioral treatments in severely ill patients', *Statistical Methods in Medical Research* **28**, 911–927.

Diaz, F. J. (2021), 'Using population crossover trials to improve the decision process regarding treatment individualization in N-of-1 trials', *Statistics in Medicine* **40**, 4345–4361.

Diaz, F. J., Berg, M. J., Krebill, R., Welty, T., Gidal, B. E., Alloway, R. & Privitera, M. (2013), 'Random-effects linear modeling and sample size tables for two special crossover designs of average bioequivalence studies: the four-period, two-sequence, two-formulation and six-period, three-sequence, three-formulation designs', *Clinical Pharmacokinetics* **52**, 1033–1043.

Diaz, F. J., Cogollo, M. R., Spina, E., Santoro, V., Rendon, D. M. & de Leon, J. (2012b), 'Drug Dosage Individualization Based on a Random-Effects Linear Model', *Journal of Biopharmaceutical Statistics* **22**, 463–484.

Diaz, F. J. & de Leon, J. (2013), 'The mathematics of drug dose individualization should be built with random effects linear models', *Therapeutic Drug Monitoring* **35**, 276–277.

Diaz, F. J., Eap, C. B., Ansermot, N., Crettol, S., Spina, E. & de Leon, J. (2014), 'Can valproic acid be an inducer of clozapine metabolism?', *Pharmacopsychiatry* **47**, 89–96.

Diaz, F. J., Rivera, T. E., Josiassen, R. C. & de Leon, J. (2007), 'Individualizing drug dosage by using a random intercept linear model', *Statistics in Medicine* **26**, 2052–2073.

Diaz, F. J., Santoro, V., Spina, E., Cogollo, M., Rivera, T. E., Botts, S. & de Leon, J. (2008), 'Estimating the size of the effects of co-medications on plasma clozapine concentrations using a model that controls for clozapine doses and confounding variables', *Pharmacopsychiatry* **41**, 81–91.

Diaz, F. J., Yeh, H.-W. & de Leon, J. (2012a), 'Role of Statistical Random-Effects Linear Models in Personalized Medicine', *Current Pharmacogenomics and Personalized Medicine* **10**, 22–32.

Edelstein, C. L. (2008), 'Biomarkers of acute kidney injury', *Advances in Chronic Kidney Disease* **15**, 222–234.

Frees, E. W. (2004), *Longitudinal and Panel Data*, Cambridge University Press, Cambridge.

Gaudin, D., Krafcik, B. M., Mansour, T. R. & Alnemari, A. (2017), 'Considerations in spinal fusion surgery for chronic lumbar pain: psychosocial factors, rating scales, and perioperative patient education-a review of the literature', *World Neurosurgery* **98**, 21–27.

Gerbershagen, H. J., Pogatzki-Zahn, E., Aduckathil, S., Peelen, L. M., Kappen, T. H., van Wijck, A. J., Kalkman, C. J. & Meissner, W. (2014), 'Procedure-specific risk factor analysis for the development of severe postoperative pain', *Anesthesiology* **120**, 1237–1245.

Gewandter, J. S., McDermott, M. P., He, H., Gao, S., Cai, X., Farrar, J. T., Katz, N. P., Markman, J. D., Senn, S., Turk, D. C. & Dworkin, R. H. (2019), 'Demonstrating heterogeneity of treatment effects among patients: an overlooked but important step toward precision medicine', *Clinical Pharmacology & Therapeutics* **106**, 204–210.

Greden, J. F. (2009), 'Treating depression and pain', *Journal of Clinical Psychiatry* **70**(6), e16.

Gronski, L., Martinson, W., Singh, K. P. & Ryan, J. (2012), 'Utility of daily troponin orders for identifying acute myocardial infarction patients for quality improvement', *Critical Pathway in Cardiology* **11**, 74–76.

Hedeker, D. & Gibbons, R. D. (2006), *Longitudinal Data Analysis*, Wiley-Interscience, Hoboken, NJ.

Hickey, G. L., Philipson, P., Jorgensen, A. & Kolamunnage-Dona, R. (2018), 'joineRML: a joint model and software package for time-to-event and multivariate longitudinal outcomes', *BMC Medical Research Methodology* **18**, 50. [doi.org/10.1186/s12874-018-0502-1].

Ibrahim, J. G. & Molenberghs, G. (2009), 'Missing data methods in longitudinal studies: a review', *Test (Madr)* **18**, 1–43.

Johnson, N. L. (1949), 'Systems of Frequency Curves Generated by Methods of Translation', *Biometrika* **36**, 149–176.

Laird, N. M. (1998), 'Missing data in longitudinal studies', *Statistics in Medicine* **7**, 305–315.

Lesaffre, E., Rizopoulos, D. & Tsonaka, R. (2007), 'The logistic transform for bounded outcome scores', *Biostatistics* **8**, 72–85.

Little, R. J. A. & Rubin, D. B. (2002), *Statistical Analysis with Missing Data, Second Edition*, Wiley, New York.

Lotzke, H., Jakobsson, M., Brisby, H., Gutke, A., Hägg, O., Smeets, R., den Hollander, M., Olsson, L. E. & Lundberg, M. (2016), 'Use of the PREPARE (PREhabilitation, Physical Activity and exeRcisE) program to improve outcomes after lumbar fusion surgery for severe low back pain: a study protocol of a person-centred randomised controlled trial', *BMC Musculoskeletal Disorders* **17**(1), 349. [doi: 10.1186/s12891-016-1203-8].

Miksad, R. A. & Abernethy, A. P. (2018), 'Harnessing the power of real-world evidence (RWE): A checklist to ensure regulatory-grade Data Quality', *Clinical Pharmacology & Therapeutics* **103**, 202–205.

Pantazis, N. & Touloumi, G. (2010), 'Analyzing longitudinal data in the presence of informative drop-out: The jmre1 command', *Stata Journal* **10**, 226–251.

Papageorgiou, G., Mauff, K., Tomer, A. & Rizopoulos, D. (2019), 'An overview of joint modeling of time-to-event and longitudinal outcomes', *Annual Review of Statistics and its Application* **6**, 223–240.

Ruberg, S. J., Chen, L. & Wang, Y. (2010), 'The mean does not mean as much anymore: finding sub-groups for tailored therapeutics', *Clinical Trials* **7**, 574–583.

Schluchter, M. D. (1992), 'Methods for the analysis of informatively censored longitudinal data', *Statistics in Medicine* **11**, 1861–1870.

Schluchter, M. D. & Piccorelli, A. V. (2019), 'Shared parameter models for joint analysis of longitudinal and survival data with left truncation due to delayed entry - Applications to cystic fibrosis', *Statistical Methods in Medical Research* **28**, 1489–1507.

Senn, S. (2016), 'Mastering variation: variance components and personalised medicine', *Statistics in Medicine* **35**, 966–977.

Shardell, M. & Ferrucci, L. (2018), 'Joint mixed-effects models for causal inference with longitudinal data', *Statistics in Medicine* **37**, 829–846.

Shaw, A. D., Mythen, M. G., Shook, D., Hayashida, D. K., Zhang, X., Skaar, J. R., Iyengar, S. S. & Munson, S. H. (2018), 'Pulmonary artery catheter use in adult patients undergoing cardiac surgery: a retrospective, cohort study', *Perioperative Medicine (Lond)* **7**, 24. [doi.org/10.1186/s13741-018-0103-x].

Shirafkan, H., Mahmoudi-Gharaei, J., Fotouhi, A., Mozaffarpur, S. A., Yaseri, M. & Hoseini, M. (2020), 'Individualizing the dosage of Methylphenidate in children with attention deficit hyperactivity disorder', *BMC Medical Research Methodology* **20**, 56. [doi.org/10.1186/s12874-020-00934-y].

Touloumi, G., Pocock, S. J., Babiker, A. G. & Darbyshire, J. H. (1999), 'Estimation and comparison of rates of change in longitudinal studies with informative drop-outs', *Statistics in Medicine* **18**, 1215–1233.

Trivedi, M. H. (2004), 'The link between depression and physical symptoms', *Primary Care Companion of the Journal of Clinical Psychiatry* **6**(Suppl 1), 12–16.

Urman, R. D., Boing, E. A., Pham, A. T., Khangulov, V., Fain, R., Nathanson, B. H., Zhang, X., Wan, G. J., Lovelace, B. & Cirillo, J. (2018), 'Improved outcomes associated with the use of intravenous acetaminophen for management of acute post-surgical pain in cesarean sections and hysterectomies', *Journal of Clinical Medicine Research* **10**, 499–507.

Wang, Z. & Diaz, F. J. (2020), 'A graphical approach to assess the goodness-of-fit of random-effects linear models when the goal is to measure individual benefits of medical treatments in severely ill patients', *BMC Medical Research Methodology* **20**, 193. [doi.org/10.1186/s12874-020-01054-3].

Weinmann, C., Komann, M. & Meissner, W. (2017), 'Tough cookies: the older the patients, the more pain tolerating?', *European Journal of Anesthesiology* **34**(Suppl 55), 215.

Woodward, M. (2014), *Epidemiology: Study Design and Data Analysis, Third Edition*, Chapman & Hall/CRC, Boca Raton, FL.

Zhang, X., de Leon, J., Crespo-Facorro, B. & Diaz, F. J. (2020), 'Measuring individual benefits of psychiatric treatment using longitudinal binary outcomes: Application to antipsychotic benefits in non-cannabis and cannabis users', *Journal of Biopharmaceutical Statistics* **30**, 916–940.

Zhu, X. & Qu, A. (2016), 'Individualizing drug dosage with longitudinal data', *Statistics in Medicine* **35**, 4474–4488.