Published in final edited form as:

Stat Med. 2008 May 20; 27(11): 1814–1833.

Mixed-effects Poisson regression analysis of adverse event reports:

The relationship between antidepressants and suicide

Robert D. Gibbons^{1,*†}, Eisuke Segawa¹, George Karabatsos², Anup K. Amatya¹, Dulal K. Bhaumik¹, C. Hendricks Brown^{1,3}, Kush Kapur¹, Sue M. Marcus^{1,4}, Kwan Hur^{1,5}, and J. John Mann^{6,7}

¹Center for Health Statistics, University of Illinois at Chicago, Rooms 455-457 (MC 912), 1601 W. Taylor, Chicago, IL 60612, U.S.A.

²Department of Education, University of Illinois at Chicago, (MC 147), 1040 W. Harrison St., Chicago, IL 60607, U.S.A.

³Prevention Science and Methodology Group, Department of Epidemiology and Biostatistics, College of Public Health MDC-56, University of South Florida, 13201 Bruce B Downs Blvd, Tampa, FL 33612, U.S.A.

⁴Department of Psychiatry, Mount Sinai School of Medicine, One Gustave L. Levy Place, P.O. Box 1230, New York, NY 10029, U.S.A.

⁵Cooperate Studies Program Coordinating Center, Hines VA Hospital, Hines, IL 60141, U.S.A.

⁶Department of Molecular Imaging and Neuropathology, New York State Psychiatric Institute, 1051 Riverside Drive, New York, NY 10032, U.S.A.

⁷Department of Psychiatry, Columbia University College of Physicians & Surgeons, 1051 Riverside Drive, New York, NY 10032, U.S.A.

SUMMARY

A new statistical methodology is developed for the analysis of spontaneous adverse event (AE) reports from post-marketing drug surveillance data. The method involves both empirical Bayes (EB) and fully Bayes estimation of rate multipliers for each drug within a class of drugs, for a particular AE, based on a mixed-effects Poisson regression model. Both parametric and semiparametric models for the random-effect distribution are examined. The method is applied to data from Food and Drug Administration (FDA)'s Adverse Event Reporting System (AERS) on the relationship between antidepressants and suicide. We obtain point estimates and 95 per cent confidence (posterior) intervals for the rate multiplier for each drug (e.g. antidepressants), which can be used to determine whether a particular drug has an increased risk of association with a particular AE (e.g. suicide). Confidence (posterior) intervals that do not include 1.0 provide evidence for either significant protective or harmful associations of the drug and the adverse effect. We also examine EB, parametric Bayes, and semiparametric Bayes estimators of the rate multipliers and associated confidence (posterior) intervals. Results of our analysis of the FDA AERS data revealed that newer antidepressants are associated with lower rates of suicide adverse event reports compared with older antidepressants. We recommend improvements to the existing AERS system, which are likely to improve its public health value as an early warning system.

*Correspondence to: Robert D. Gibbons, Center for Health Statistics, University of Illinois at Chicago, 1601 W. Taylor, Chicago, IL 60614, U.S.A.

†E-mail: rdgib@uic.edu, robert.gibbons@uic.edu

Keywords

Poisson regression; mixed-effects models; adverse event reports; empirical Bayes; suicide

INTRODUCTION

The primary tool available to the U.S. Food and Drug Administration (FDA) for post-marketing drug adverse event (AE) surveillance is the Adverse Event Reporting System (AERS). The FDA receives AE reports from manufacturers as required by regulation. Healthcare providers and consumers also send reports of AEs voluntarily through FDA's MedWatch program. These AE reports are evaluated by clinical reviewers in the Center for Drug Evaluation and Research and the Center for Biologics Evaluation and Research to detect safety signals and to monitor post-marketing drug safety. As a result, the FDA may take regulatory actions to improve product safety and protect the public health, such as revising a product's labeling information, sending out a 'Dear Health Care Professional' letter, or re-evaluating an approval decision. In October 1997, the earlier Spontaneous Reporting System (SRS) was replaced by the expanded AERS. Annually, FDA receives approximately 250 000 adverse experience reports possibly associated with drug use.

The AERS is a passive reporting system and is therefore subject to numerous sources of selection bias, over- and under-reporting [1]. Most notable among these are (1) a lack of a denominator indicating the population at risk for a particular drug-AE interaction, (2) lack of control or comparison groups, (3) recall bias of patients and reporters, (4) poor case documentation in the reports, (5) large amounts of missing data on patient characteristics, (6) media sensitivity, and (7) substantial under-reporting. To deal with the absence of a denominator, DuMouchel [2] has introduced a series of estimators that compare the expected number of events based on the marginal frequencies for each drug and AE with the observed frequency for that particular drug-AE combination. An empirical Bayes (EB) estimate of this risk ratio is then used to identify those drug-AE interactions that may be of concern. This is the principal data mining method used by FDA.

We propose an alternative approach, based on a mixed-effects Poisson regression model that specifically includes denominators adjusting for drug usage. This method yields rate multipliers for each drug in a class of drugs (e.g. antidepressants) that describe the deviation of the rate for a specific AE (e.g. suicide) from that for the drug class as a whole. The rate multiplier has a straightforward interpretation. A multiplier of 1.0 indicates that the drug has the average rate of the AE for the entire drug class, a multiplier of 2.0 indicates that the drug has twice the rate of the AE relative to the class average, and a multiplier of 0.5 indicates that the drug has one-half of the rate of the overall class. In this way, individual drugs can be simultaneously screened for a relevant AE. Statistical inference is provided by large sample confidence (posterior) intervals for the rate multipliers that can be screened to determine if the confidence (posterior) interval for the rate multiplier contains 1.0. Unlike DuMouchel's method that does not require a denominator, the current procedure requires either an exact denominator or a proxy denominator. For example, if we use a complete medical records database such as the Veterans Administration (VA) database as an alternative to the AERS system, we can determine the number of subjects who were prescribed a particular drug. This is an example of an exact denominator for a specific cohort. No such denominator is available for the AERS data; however, a proxy national denominator such as national projections of the number of prescriptions of a particular drug during a time interval of interest can be obtained from a variety of sources including the National Ambulatory Medical Care Evaluation Survey (NAMCES), the National Hospital Ambulatory Medical Care Evaluation Survey (NHAMCES) and IMS databases. In general, large sample sizes are required to justify the use of these methods.

One of the most debated questions of AEs in drugs is whether antidepressant usage causes suicide. In September 2004, the FDA's Neuro-Psychopharmacologic Advisory Committee and Pediatric Advisory Committee concluded that there was sufficient evidence to support a causal link between antidepressants and pediatric suicidality. One month later in October 2004, the FDA ordered pharmaceutical companies to add to antidepressant advertisements, package inserts, and information sheets developed for patients and clinicians, a 'black box' warning (a statement in prominent, boldfaced type and framed by a black border) regarding pediatric use. As pointed out by Leslie *et al.* [3], by issuing this warning, the federal drug regulators imposed one of their toughest requirements, short of banning a medication. The evidence that led to this conclusion was based on a retrospective review of clinical data (e.g. medical records) from a series of 25 randomized clinical trials (RCTs) of newer antidepressants (selective serotonin reuptake inhibitors (SSRIs), non-selective serotonin reuptake inhibitors (NSSRIs), in contrast to older tricyclic antidepressants (TCAs)) in children, showing more medication-related AE reports of suicidal ideation and behavior, namely 2 per cent in placebo compared with 4 per cent in actively treated children. There were no reports of completed suicide in any of these studies [4]. More recently, the FDA has extended the black box warning to young adults aged 18-25 [5]. Although the black box warning is for all antidepressants, the psychiatric literature has focused primarily on the newer antidepressants, SSRIs and NSSRIs, compared with older antidepressants such as TCAs [6].

These findings appear to be inconsistent with findings from several psychological autopsy studies of completed suicide, which suggest that very few pediatric suicides are exposed to antidepressants [6], and from ecological studies on suicide completion [7] and attempts that indicate an overall protective effect of SSRIs in pediatric and young adult populations [8-10]. Gibbons *et al.* [11] also found a significant inverse relationship between county-level SSRI prescription levels and adult suicide rates both within counties over time and between counties. Ludwig and Marcotte [12] demonstrated an inverse association between SSRI prescription rates and SSRIs from 1980 to 1999 in 26 European countries, and similar results have been found in the U.S. [13]. Most recently, Bramness *et al.* [14] illustrated a significant inverse association between SSRI sales and suicide rates from 1980 to 2004 in a county-level analysis of Norway.

Posner *et al.* [15] point out that the results of FDA's meta-analysis could be explained by ascertainment bias, whereby patient suicidality is more likely to be noted in medical record reviews of patients (children or adults) randomly assigned to an active medication condition compared with placebo controls, due to increased contact with medical staff because of increased side effects in general with an active medication compared with placebo. Similarly, suicide attempts may also be subject to ascertainment bias in that overdoses of placebo may not be identified by study staff. In further support of this possible explanation, no significant differences between active and placebo conditions were found by FDA for prospective measures of suicidal ideation or behavior using the specific items on the Hamilton Depression Rating Scale or the Childhood Depression Rating Scale [4].

Most recently, Gibbons *et al.* [16] analyzed person-level suicide attempt data from the VA in a cohort of 226 866 newly diagnosed depressed patients with longitudinal records for all inpatient and outpatient visits, and complete pharmacy records. A significant 3-fold drop in suicide attempts was found in subjects treated with an SSRI only (monotherapy) compared with those without any antidepressant treatment (123/100 000 for SSRIs *versus* 335/100 000 for no antidepressant, OR = 0.37, $p < 0.0001$). Similar protective effects were found throughout the age range, including the 18-25 year old young adult cohort. In subjects treated with an SSRI only, the rate of suicide attempts was significantly greater prior to treatment (221/100 000) *versus* after treatment (123/100 000), (RR=0.56, $p < 0.0001$). Similar results were found by Simon and coworkers, who found that the highest rate of suicide attempts was immediately

prior to initiation of treatment with antidepressants [17,18]. Although some of these patients may have initiated pharmacotherapy because of a suicide attempt, these data clearly indicate that the rate of suicide attempts following initiation of treatment is not increasing.

In this paper, we derive parametric EB and parametric as well as semiparametric full Bayes (FB) estimators from a mixed-effects Poisson regression model appropriate for analyzing AEs from drugs. These methods are used to present another perspective on the question of the relationship between antidepressants and suicide, by examining all available spontaneous reports of suicide associated with antidepressant use in the U.S., between 1998 and 2004. The focus of this analysis is to illustrate this new statistical methodology for detecting higher or lower AE interactions (e.g. suicide) with a single drug or a specific class of drugs (e.g. SSRIs) compared with all drugs of that category (e.g. antidepressants).

METHODS

Data from the AERS from 1998 to 2004 were the focus of our analysis. Antidepressants (generic, brand name(s)) were categorized into three classes: SSRIs (citalopram, celexa; paroxetine, paxil; fluoxetine, prozac, sarafem; fluvoxamine, luvox; sertraline, zoloft), NSSRIs (nefazodone, serzone; mirtazapine, remeron; bupropion, wellbutrin; venlafaxine, effexor), and TCAs (amitriptyline, elavil, endep; amoxapine, asendin; clomipramine, anafranil; desipramine, norpramine, pertofrane; doxepin, adapin, sinequan; imipramine, janimine, tofranil; nortriptyline, aventyl, pamelor; protriptyline, triptil, vivactil; trimipramine, rhotrimine, surmontil). The AE of interest was completed suicide, as reported by a hospital, a healthcare provider, a pharmaceutical company, or a patient. The data set consisted of a total of 28 317 382 records which included all reported AEs and drug combinations. To obtain a denominator for our analyses (i.e. population at risk as measured by estimates of the national number of prescriptions for each drug or drug class by year), we used estimates of the national prescription rates from the NAMCES and NHAMCES. Each year 3000 physicians are randomly selected to provide data on approximately 30 patient visits over a randomly selected 1-week period (NAMCES). Each year 500 nationally representative hospitals are surveyed for a randomly selected 4-week period (NHAMCES). All prescriptions that are written during these periods are recorded and extrapolated to obtain estimates of national rates. These estimates do not provide information on whether the prescriptions were either filled or taken. Furthermore, the NAMCES data are a survey of office visits and not individual patients; therefore, the estimates may over-represent frequent users of the healthcare system. Finally, estimates of prescription rates for older drugs that are infrequently prescribed may have increased uncertainty given the small number of physicians sampled (i.e. 3000).

Mixed-effects Poisson regression models

In the analysis of AE data in general, and suicide rate data in particular, Poisson regression models are a natural choice. Suppose that there are c drugs under study and the adverse effect (i.e. suicide) of each drug is counted at T time points (in our case $T = 7$ and time is coded as -3, -2, -1 0,1,2, and 3). Let t_j be the j th time point and y_{ij} be the number of suicide counts for drug i at time point j . Also, let n_{ij} be the number of subjects who took drug i at time point j and $\log(n_{ij})$ is an offset. Denote the mean number of suicides on drug i at time j by λ_{ij} . The mixed-effects Poisson regression model that we propose is

$$f(y_{ij}; \lambda_{ij}) = \frac{\exp(-\lambda_{ij}) (\lambda_{ij}^{y_{ij}})}{y_{ij}!} \quad (1)$$

where the mean parameter λ_{ij} is linked to linear predictors through the log link,

$$\begin{aligned}
 \eta_{ij} &= \log(\lambda_{ij}) \\
 &= \log(n_{ij}) + \mathbf{x}_j^T \mathbf{b}_i \\
 &= \log(n_{ij}) + \mathbf{x}_j^T \boldsymbol{\beta} + \mathbf{x}_j^T \mathbf{u}_i
 \end{aligned} \tag{2}$$

where $\mathbf{x}_j^T$ is a $1 \times p$ vector of the time covariate, $[t_j^0, t_j^1, \dots, t_j^{p-1}]^T$, $\mathbf{b}_i$ is the uncentered random coefficient vector with $\mathbf{b}_i \sim G(\Phi)$, $\mathbf{b}_i$'s are independent and G is the assumed distribution for $\mathbf{b}_i$, which depends on a vector of hyperparameters, Φ . We assume that $\boldsymbol{\beta}$ is a fixed parameter vector and $\mathbf{u}_i$ is a random variable. The interpretation of $\mathbf{x}_j^T \boldsymbol{\beta}$ in (2) is the time trajectory of overall suicidal reports in the log mean scale and $\mathbf{x}_j^T \mathbf{u}_i$ measures the variation of the time trajectory across drugs.

As we have described the model in (2), it is a random trend model in which the $\mathbf{b}_i$'s represent the random polynomial trend coefficients that can take on drug-specific values. Alternatively, time can be nominally coded to reflect individual time-point differences from baseline or the midpoint of the time interval as desired. In the present context, we have used a random slope model to adjust for the observed gradually increasing trend in spontaneous suicide reports over time. The model can easily be expanded to include the effects of time-invariant covariates (e.g. sex, age, and race of the patient); however, the large amount of missing covariate information in the AERS database prevented us from including these effects in our illustration. Future applications of this methodology to large-scale medical records or medical claims databases (e.g. the VA database) can add these terms to the model to adjust for case mix.

We consider three approaches to estimating the Poisson regression model, namely, the EB approach, the parametric FB approach, and the semiparametric FB approach. The main differences between the three approaches relate to how G is specified and estimated. In the EB and parametric FB approaches, G is assumed to be a normal distribution, with Φ defined by mean vector and covariance-matrix parameters. In the EB approach, Φ is inferred by point estimation using maximum marginal likelihood, whereas in the parametric FB approach, a prior is placed on Φ , and inference focuses on the posterior distribution of Φ . In the semiparametric FB approach, a 'non-parametric' prior is placed directly on G , with support on the set of all (measurable) distributions $\{G\}$. This prior depends on a hyperparameter vector Φ (described later).

Our main focus in statistical inference is the rate multiplier for each drug. It is an exponential of the centered random intercept, $\exp(u_{0i})$, as defined below. Posterior distributions are used in all FB approaches for the inference of the random effect (rate multiplier). EB treats the variance of the random effect as known in obtaining the posterior distribution of u_{0i} although in fact it is estimated with imprecision, whereas the FB approaches do not impose such assumption. We compare the results of these three approaches, and we also evaluate whether the normality assumptions in EB and parametric FB appreciably affect the inference of the rate multiplier. Such sensitivity analyses are potentially critical, because the assumptions of known and normally distributed random-effect distributions are not likely to be satisfied, due to the small number of units (i.e. 18 drugs).

Rate multiplier

A mean of the linear predictor over time for drug i is

$$\begin{aligned}
\bar{\eta}_i &= \frac{1}{J} \sum_{j=1}^J \eta_{ij} \\
&= \frac{1}{J} \sum_{j=1}^J \log(n_{ij}) + \frac{1}{J} \sum_{j=1}^J \mathbf{x}_j^T \beta + \frac{1}{J} \sum_{j=1}^J \mathbf{x}_j^T \mathbf{u}_i \\
&= \bar{n}_i + \bar{\beta} + \bar{\mathbf{u}}_i
\end{aligned} \tag{3}$$

where $\bar{n}_i = (1/J) \sum_{j=1}^J \log(n_{ij})$, $\bar{\beta} = (1/J) \sum_{j=1}^J \mathbf{x}_j^T \beta$, and $\bar{\mathbf{u}}_i = (1/J) \sum_{j=1}^J \mathbf{x}_j^T \mathbf{u}_i$. Taking exponentials of both sides of (3),

$$\begin{aligned}
\bar{\lambda}_i &= \exp(\bar{\eta}_i) \\
&= \exp(\bar{n}_i) \exp(\bar{\beta}) \exp(\bar{\mathbf{u}}_i) \\
&= \exp(\bar{n}_i) \bar{\lambda} m_i
\end{aligned} \tag{4}$$

where $\bar{\lambda} = \exp(\bar{\beta})$ and $m_i = \exp(\bar{\mathbf{u}}_i)$. The terms $\bar{\lambda}$, $\bar{\lambda}$, and m_i are called the rate of drug i , a mean rate (over drugs), and a rate multiplier for drug i , respectively. A multiplier of 1 reflects a drug with a suicide AE report rate equal to the overall rate for antidepressants, a multiplier of 2 reflects a doubling of the rate, and a multiplier of 0.5, half of the rate. Although λ is called a mean rate, it is not a mean of rates over drugs but an exponential of a mean of log rates. Because the coding of time is centered and balanced in our study, covariates with odd-powered time points, e.g. t_j^1 and t_j^3 , are canceled out in the computation of means. In the following sections, an outline of each approach is presented, and limitations of EB and parametric FB are addressed.

Parametric EB estimates of drug-specific rate multipliers—In this section we assume that $\mathbf{u}_i$ follows a normal distribution with mean vector $\mathbf{0}$ and variance matrix Σ . Let $\boldsymbol{\theta}_i = \Sigma^{-1/2} \mathbf{u}_i$, hence $\boldsymbol{\theta}_i \sim N(\mathbf{0}, \mathbf{I})$ where $\mathbf{I}$ is the identity matrix. The marginal density of $\mathbf{y}_i$ can be expressed as

$$h(\beta, \Sigma, \mathbf{y}_i) = h(\mathbf{y}_i) = \int_{\boldsymbol{\theta}_i} f(\mathbf{y}_i | \boldsymbol{\theta}_i) g(\boldsymbol{\theta}_i) d\boldsymbol{\theta}_i$$

where the density $g(\boldsymbol{\theta}_i)$ is the multivariate standard normal density with mean vector $\mathbf{0}$ and variance matrix $\mathbf{I}$. Hence, the log likelihood corresponding to Equations (1) and (2) for the i th drug is

$$\begin{aligned}
\log L(\mathbf{y}_i | \boldsymbol{\theta}_i) &= \log \left[\prod_j f(y_{ij} | \boldsymbol{\theta}_i) \right] \\
&= \sum_j \left[-(\log n_{ij} + \mathbf{x}_j^T \beta + \mathbf{x}_j^T \Sigma^{1/2} \boldsymbol{\theta}_i) + y_{ij} \log(\log n_{ij} + \mathbf{x}_j^T \beta + \mathbf{x}_j^T \Sigma^{1/2} \boldsymbol{\theta}_i) - \log(y_{ij}!) \right]
\end{aligned}$$

with $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{in_i})^T$.

For a more general approach to maximum marginal likelihood estimation of the parameters of this mixed-effects Poisson regression model, see Hedeker and Gibbons [19].

In the present context, we are primarily interested in obtaining estimates of the values of the random effects $\boldsymbol{\theta}_i$. In our example, these estimates represent the suicide rate for a given drug, conditional on year. Other model covariates such as age, race, and sex can be added either as fixed or random effects. If included as a fixed effect, the covariate exerts the same effect for all drugs. If included as a random effect, the covariate has a drug-specific effect. In our analysis of the AERS data, only year was included as a covariate due to the considerable amount of

missing data for demographic information in the AERS database. Adjustment for case mix using numerous relevant covariates could be accomplished using a large-scale medical records or claims database such as the VA or PHARMetrics databases.

The expected *a posteriori* (EAP) or EB estimator $\bar{\theta}_i$ has been suggested by Bock and Aitkin [20]. The estimator $\bar{\theta}_i$ given $\mathbf{y}_i$ and $\Sigma = \widehat{\Sigma}$ (i.e. the random-effects covariance matrix for which we obtain the maximum likelihood estimate $\widehat{\Sigma}_{ML}$) for drug i can be obtained as

$$\bar{\theta}_i = E(\theta_i | \mathbf{y}_i, \widehat{\Sigma}_{ML}) = \frac{1}{h(\mathbf{y}_i)} \int_{\theta_i} \theta_i f(\mathbf{y}_i | \theta_i, \widehat{\Sigma}_{ML}) g(\theta_i | \widehat{\Sigma}_{ML}) d\theta_i \quad (5)$$

Similarly, the variance matrix of the posterior distribution of θ_i , which may be used to make inferences regarding the EAP estimator (i.e. the posterior variance), an estimate of precision of the estimate of θ_i , is given by

$$\text{Var}(\theta_i | \mathbf{y}_i, \widehat{\Sigma}_{ML}) = \frac{1}{h(\mathbf{y}_i)} \int_{\theta_i} (\theta_i - \bar{\theta}_i)(\theta_i - \bar{\theta}_i)' f(\mathbf{y}_i | \theta_i, \widehat{\Sigma}_{ML}) g(\theta_i | \widehat{\Sigma}_{ML}) d\theta_i \quad (6)$$

The asymptotic confidence interval for the k th component of θ_i (e.g. $k = 1$ denotes the random intercept, and $k = 2$ is the random slope), say θ_{ki} , can be obtained directly from the EAP estimate and its standard deviation as $\bar{\theta}_{ki} \pm z_{\alpha} s(\theta_{ki})$, where z_{α} is the desired percentage point of the normal distribution and $s(\theta_{ki})$ is the posterior standard deviation of θ_{ki} , which is the square root of the kk th diagonal element of $\text{Var}(\theta_i | \mathbf{y}_i, \widehat{\Sigma}_{ML})$.

In clustered data, the EB estimator θ_i of cluster-specific effects can be used similar to the method used by Thomas *et al.* [21] in hospital mortality rate analysis. In their presentation, these cluster-specific effects or hospital-specific effects represent how much the death rates for patients at hospital i differ (on the logit scale) from the national rates for patients with the same covariate values. In the present context, cluster-specific effects may be linked to pharmacological characteristics of the different drugs and demographic and diagnostic characteristics of the patients that are prescribed them. Longford [22] gives extensive references to applications involving EB estimates of random effects.

FB approaches

Hierarchical Bayesian models—A disadvantage of the EB approach in obtaining point estimates of the random effects is that it assumes that the random-effect covariance matrix is known (see [23], for details). FB approaches, including the parametric and semiparametric FB approaches considered in this paper, overcome this limitation by introducing a prior distribution for the parameters that govern the distribution of the random effects. We proceed to describe the basic hierarchical structure of the parametric and semiparametric FB models.

Let us denote the likelihood of all the observations $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)^T$ by

$$f(\mathbf{y} | \mathbf{b}) = \prod_i \prod_j f(y_{ij} | \lambda_{ij}, \mathbf{b}_i)$$

The prior distribution of the random-effect parameter, $g(\mathbf{b} | \Phi) = \prod_i g(\mathbf{b}_i | \Phi)$, with $\mathbf{b} = (\mathbf{b}_1, \dots, \mathbf{b}_n)^T$ is specified.

Recall that in the parametric FB approach, like the parametric EB approach, it is assumed that $g(\cdot | \Phi)$ is a normal density. In contrast, the semiparametric approach supports all measurable distributions G of the random effects, including normal distributions, where Φ are the parameters of a Dirichlet process [24]. Finally, to complete the model specification, a prior

density on the hyperparameters, $p(\Phi)$, is specified. The joint posterior distribution of random effects, $\mathbf{b}$, and parameters in the prior, Φ , is given up to a constant of proportionality, by

$$f(\mathbf{b}, \Phi | \mathbf{y}) \propto f(\mathbf{y} | \mathbf{b}) g(\mathbf{b} | \Phi) p(\Phi) \quad (7)$$

The posterior density of $\mathbf{b}_i$ for the 18 drugs

$$f(\mathbf{b}_i | \mathbf{y}) \propto f(\mathbf{y} | \mathbf{b}_i) p(\mathbf{b}_i), i = 1, \dots, 18 \quad (8)$$

is our main interest. The specific implementations in parametric and semiparametric FB approaches are explained below.

Parametric FB—Several authors have considered the parametric FB approach for statistical inferences for mixed-effects Poisson regression models. El-Sayyad [25] considered an improper prior for $\mathbf{b}$, in the context of a mixed-effects Poisson regression model. Using a well-known result of a gamma distribution that when its shape parameter is large, the distribution of the logarithm of the gamma variable can be approximated by a normal distribution, El-Sayyad proved that the posterior density follows a multivariate normal distribution. Lee and Nelder considered several prior distributions for $\mathbf{b}$ [26]. For a single random component $\mathbf{b}$, Lee and Nelder assumed that $\mathbf{b}$ has a log-gamma distribution and this model is called a Poisson-gamma hierarchical generalized linear model (HGLM). They also considered the lognormal distribution for the random component $\mathbf{b}$ and called the model a Poisson-lognormal HGLM. Note that the Poisson-gamma model is one of the classes of conjugate HGLMs. Albert [27] also considered a Bayesian analysis of a Poisson random-effects model assuming a vague prior distribution for the mean parameter. Cooil [28] considered the hierarchical Poisson regression models with random components having any distribution in the gamma family to predict the frequency of claims using medical malpractice data. Normand *et al.* [29] considered a hierarchical logistic regression model for profiling providers of medical care under the assumption that the random vector $\mathbf{b}$ follows a multivariate normal distribution with a mean vector $\boldsymbol{\eta}$ and covariance matrix $\mathbf{C}$. They further assumed vague prior distributions on both $\boldsymbol{\eta}$ and $\mathbf{C}$. To fit their model they used Gibbs sampling and the sampler was implemented using the BUGS software.

Christiansen and Morris [30] considered a very general hierarchical Poisson regression model assuming conjugate gamma distributions for the mean parameters of the Poisson distribution and improper prior distributions of the hyperparameters related to the gamma distributions. The justification for the choice of an improper prior distribution for the structural hyperparameters is to provide good repeated sampling properties to the resulting rules. Our parametric FB approach is similar to the approach of Normand *et al.* [29]. In what follows we show that this limiting effect of the priors is achieved by imposing some conditions on the parameters of the distributions of hyperpriors.

For the FB parametric mixed-effects Poisson model, we assume that the prior distribution of $\mathbf{b}_i$ is normal with mean $\boldsymbol{\beta}$ and variance $\boldsymbol{\Sigma}$ as specified in EB. Normal₂ and Wishart₂ distributions are used for hyperpriors of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}^{-1}$, respectively,

$$\boldsymbol{\beta} \sim N(\boldsymbol{\eta}, \mathbf{C}) \quad (9)$$

and

$$\boldsymbol{\Sigma}^{-1} \sim \text{Wishart}(\Lambda, \nu) \quad (10)$$

The parametric FB approach is implemented in WinBUGS [31]. The posterior distribution is obtained by the Markov chain Monte Carlo (MCMC) method. In particular, we specified $\boldsymbol{\eta}^T = (0, 0)$ and $\mathbf{C}$ is a diagonal matrix with 10^4 for its diagonal elements for the normal prior for $\boldsymbol{\beta}$. Thus, we assume that the regression parameters are independent *a priori*, with each individual regression parameter having a rather flat, vague normal prior. As a consequence,

the resulting posterior density of β will be mostly determined by the data, especially given the very large number of observations in the data we will analyze.

As will be shown, this is empirically supported by the similarity between EB and FB estimates of β . For the Wishart its degrees of freedom are set equal to its dimension, $\nu = 2$. We specified

$\Lambda = \begin{pmatrix} 0.72 & 0.36 \\ 0.36 & 42 \end{pmatrix}$, which is obtained by substituting estimates of variance component from EB and $\nu = 2$ in the following relation: $E(\Sigma^{-1}) = \nu\Lambda$.

Semiparametric FB—Using the semiparametric approach, we extend the flexibility of the parametric Poisson mixed-effects models by specifying a non-parametric, Dirichlet process prior that supports the entire space of distributions of the random effects, $\{G\}$, including all normal distributions. This consideration is motivated by the fact that inferences of the random effects may be inaccurate, if we incorrectly assume that the random effects are normally distributed (see [32]). This non-parametric modeling can be achieved with ideas of Bayesian non-parametrics, where a ‘non-parametric’ prior distribution is specified to support all possible (measurable) distributions of the random effects. See [33] for a recent review of Bayesian non-parametric inference, and see [34] for a discussion of how maximum likelihood estimators are obtained under a specific choice of a non-parametric prior. A very general non-parametric prior is provided by the Dirichlet process. Given a space V and a σ -field B of subsets of V , a random distribution function G on (V, B) follows a Dirichlet process $DP(\alpha, G_0)$, where $\alpha > 0$ is a scalar parameter and G_0 is a specified distribution on V , if, given pairwise disjoint partitions of $B_1, \dots, B_s$ of B , the joint distribution of the vector $(G(B_1), \dots, G(B_s))$ is a Dirichlet distribution with parameters $\alpha G_0(B_1), \dots, \alpha G_0(B_s)$ (see [24]).

It is well known that any continuous distribution can be well approximated by a suitable mixture of normal distributions with a large-enough number of components. Along these lines, we adopt Kleinman and Ibrahim’s [32] Bayesian semiparametric mixed-effects model, which is based on modeling the distribution of the random effects (G) with a mixture of Dirichlet process prior with baseline distribution G_0 defined by a normal distribution. In this approach, the distribution of random effects is modeled by a countably infinite mixture of normal distributions. Denote $\mathbf{x}_j = (1, j)^T$ as the vector of random-effect regressors (j is time). Let $\mathbf{b}_i = \beta + \mathbf{u}_i$, where β denotes the average of $\mathbf{b}_i$ and $\mathbf{u}_i$ denotes the random effect centered at $\mathbf{0}$. With this Dirichlet process prior on the distribution of random effects, G , and suitable prior distributions on the rest of the parameters of the mixed-effects model, the semiparametric mixed-effects model is defined in the hierarchical form as follows:

$$\begin{aligned} \mathbf{y}_i &\sim \prod_{j=1}^J \text{Poisson}(\mathbf{y}_{ij} | \mathbf{x}_{ij}^T \mathbf{b}_i), i = 1, \dots, c \\ \mathbf{b}_1, \dots, \mathbf{b}_c &\stackrel{\text{iid}}{\sim} G \\ G &\sim \text{DP}(\alpha, \text{Normal}_2(\mu, \Sigma)) \\ \alpha &\sim \text{Gamma}(0.001, 1000) \\ \mu &\sim \text{Normal}_2(0, \text{diag}(1000)) \\ \Sigma^{-1} &\sim \text{Wishart}_2(\text{diag}(10), 4) \end{aligned}$$

In the above set of equations, $\text{DP}(\alpha, \text{Normal}_2(\mu, \Sigma))$ denotes a Dirichlet process prior on G , with prior mean (G_0) defined by the bivariate normal distribution with parameters μ and Σ , and the parameter α reflects the degree of belief that this normal distribution is the true distribution of the random effects. Even though this $\text{Normal}_2(\mu, \Sigma)$ distribution has highest prior weight in the Dirichlet process prior, this prior supports all possible distributions of the random effects, provided that α is low enough. Interestingly, in the limit $\alpha \rightarrow \infty$, the semiparametric model presented above reduces to the normal Poisson mixed-effects model. Thus, the normal model is a special case of the semiparametric model, under the very strong

prior belief that the random effects are normally distributed. In order to allow the posterior inference of all the parameters of the semiparametric model to be mostly determined by the data, as shown above, we specify relatively non-informative priors $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}$, with extremely high variances, along with a rather vague Gamma prior for α with shape 0.001 and scale 1000. Gamma prior is a rather flat prior that supports a large range of values of α but places more support for lower values of α in the range between 0 and 60.

Standard MCMC algorithms, first presented in this journal by Kleinman and Ibrahim [32], can be implemented to infer the posterior distribution of this semiparametric mixed model, given the data. In particular, a conditional posterior sample of $\boldsymbol{\mu}$ is generated from a multivariate normal distribution, and $\boldsymbol{\Sigma}$ is drawn from a Wishart distribution (see [32]), while a conditional posterior sample of the precision parameter α is generated by a beta mixture of gamma distributions (see [35]). Finally, a conditional posterior sample of the uncentered random-effect parameters $\mathbf{b}_1, \dots, \mathbf{b}_{18}$ can be efficiently generated using MCMC methods for non-conjugate priors (see [36, Algorithm 8, setting $m=1$]). We used the DPpackage of the R software to estimate the full posterior distribution of our semiparametric mixed model presented above, which can be downloaded from the Web site <http://cran.r-project.org/src/contrib/Descriptions/DPpackage.html>.

RESULTS

Spontaneous reports and prescription frequencies

Figure 1 displays the rate of suicide AE reports per million prescriptions for each drug class. Reported suicide rates have been increasing for all drug classes, especially since 2002. Figure 1 reveals that when adjusted for numbers of prescriptions, it is the TCAs in adults that have most suicide-related AE reports (see Figure 1(b)), relative to the newer antidepressant classes, with little difference between drug classes for youth (see Figure 1(a)). The increased noise in the rate of TCA prescriptions over time may be due in part to the smaller absolute number of TCA prescriptions during this time period. Figure 2 illustrates that these effects are preserved for the individual drugs within each drug class. Random slope Poisson regression (EB model) with drug-type indicators reveals that the overall rate of suicide for both SSRIs ($p < 0.0001$) and NSSRIs ($p < 0.0001$) are significantly lower than that for TCAs. No difference in spontaneously reported suicide rates between SSRIs and NSSRIs was found ($p > 0.05$).

Figure 2 reveals that in general, the results for individual drugs within classes are consistent with the overall class-level rates. SSRIs (labeled 1-5) have uniformly low suicide AE report rates for all age groups, NSSRIs (labeled 6-9) also have low rates, with possible minor elevations for mirtazapine and bupropion in children, and TCAs (labeled 10-18) in general have much higher suicide AE report rates, where amoxapine and trimipramine in adults and desipramine in children have exceptionally high suicide AE report rates. By contrast, amitriptyline and protriptyline have low overall suicide AE report rates for both youth and adults among the class of TCAs.

EB Estimates

Figure 3 presents a plot of the EB rate multiplier estimates and their confidence limits for each drug, for all ages combined.

Figure 3 reveals that as a class, SSRIs and NSSRIs have rate multipliers that are significantly less than 1.0, which means lower than the national average suicide AE report rate for antidepressants. By contrast, as a class, TCAs have rate multipliers that are significantly above the national average suicide rate for antidepressants. Most notable are the TCAs amoxapine (asendin), desipramine (norpramine), and trimipramine (surmontil) that have rates 4-5 times

higher than the average for antidepressants. The TCAs clomipramine (anafranil) and doxepin (sinequan) also have rate multipliers that are significantly above the average rate for the class of antidepressants ($\exp(\text{EB}) = 1.5$ or 50 per cent increase). Based on these data, the safest TCA is amitriptyline (elavil). By contrast, all SSRIs and NSSRIs with the exception of mirtazapine (remeron) and fluvoxamine (luvox) are significantly below the average suicide adverse event report rate for antidepressants. The NSSRI nefazodone (serzone) had the lowest spontaneously reported rate for completed suicide. As in the previous analyses, these data do not support the hypothesis that SSRIs and NSSRIs are related to suicide. When adjusted for numbers of prescriptions, TCAs appear to have the highest risk of suicide AE reports, compared with SSRIs and NSSRIs, and these differences are statistically significant.

FB estimates

The results of the parametric FB model were based on 9000 posterior samples, taken from three parallel MCMC chains of length 3000, with the first 1000 MCMC samples of each chain discarded as burn-in. Using the convergence diagnostics that are routinely calculated in WinBUGS, the choice of burn-in was justified using the Gelman-Rubin convergence statistic (<1.1). Also, we found that our choice of the number of samples (9000) led to a very small ratio of the Monte Carlo error over the posterior standard deviation of each model parameter. The results of the semiparametric FB model were based on 20 000 posterior samples, taken from a total of 200 000 MCMC samples, where every 10th sample is taken after the first 2999 burn-in MCMC samples. We found that by retaining only every 10th sample after burn-in, the autocorrelation of the samples of each model parameter was near zero, as desired.

Both parametric and semiparametric FB posterior intervals for the rate multipliers are consistently wider than the EB posterior intervals (lower and higher limits of FB are lower and higher respectively than EB, see Table I), with the semiparametric confidence being the widest among the three methods. However, in general, conclusions regarding significant deviations from the overall suicide rate for the entire class of antidepressant drugs were quite similar for the EB and the two FB methods. Tofranil was borderline significant for EB but not significant for the two FB methods. Prozac, anafranil, and sinequan were significant for both EB and parametric FB methods but borderline for the semiparametric FB method (see Table I).

Population parameters for all three methods are displayed in Table II. In terms of fixed effects, the posterior mean estimates of both intercept and slope are similar between the three approaches. The 95 per cent confidence (posterior) intervals of intercepts are also similar between the three methods. For the 95 per cent interval of slope (i.e. test of temporal homogeneity), both parametric and semiparametric FB posterior interval are wider than the EB confidence interval and the parametric is larger than semiparametric. Interestingly, the 95 per cent confidence interval for the slope under the EB estimate does not envelope the value 0, and thus with this estimate, the conclusion is that time is a significant predictor of suicide rate. In contrast, for both the parametric and semiparametric FB approaches, 95 per cent posterior intervals for the slope envelop (or nearly envelop) zero, and thus under either of these approaches, the conclusion is that time is not a significant predictor (or a borderline predictor) of suicide rate.

In terms of random effects, the covariance of random intercepts and slopes was not significantly different from zero for any of the methods. For the intercept variance, both posterior mean estimates and 95 per cent posterior intervals for the two FB approaches were substantially larger than for the EB posterior mean estimates, with the semiparametric model having the highest posterior mean and widest 95 per cent posterior interval. By contrast, for the slope variance, the two FB approaches were substantially larger than for the EB estimates, and the parametric estimates and intervals were larger than the semiparametric FB estimates and intervals.

In terms of the semiparametric model, the 95 per cent posterior interval of the precision parameter α was (1.55, 28.19), and this result suggests that there is evidence against the assumption that the random effects are normally distributed. Of course, the semiparametric model is able to describe all possible distributions of random effects, including all normal distributions. Thus, especially given the posterior distribution of the precision parameter α , it would seem that of the three Poisson mixed-effects regression models, the semiparametric model seems to provide the most realistic model for the suicide rates. Furthermore, we see that there were 10 distinct latent classes of the intercept and time slope among the 18 drugs. Figure 4 displays the estimated posterior densities of the rate multipliers for each of the 18 drugs. Looking at Figure 4, there appear to be two major groups that differ in terms of their mean values, drugs 1-9 (the SSRIs and NSSRIs) and drugs 10-18 (the TCAs). Within the TCAs there are two classes, one with a very high overall mean and variance (asendin, norpramine, and surmontil) and the other, TCAs that have intermediate mean values with varying degrees of skewness. The variance of the posterior distributions are proportional to the rate at which the drug is prescribed and as such describes the uncertainty of the rate multiplier for the drug.

DuMouchel's method

Finally, we applied the method of DuMouchel [2], which is the method used by FDA to identify which drugs were associated with specific serious AEs. The method computes the expected number of AEs based on the marginal effects for each drug and AE, and rank orders drug-AE interactions in terms of an EB-adjusted estimate of the deviation between the observed and expected drug-AE rates. In terms of completed suicides, SSRIs as a class had a ranking of 332 438 out of all drug-AE combinations, meaning that there were nearly one-third of a million stronger associations than those of SSRIs and suicide. Of course, there are several million possible combinations of drugs and AEs. This relationship had an EB-estimated relative risk of 4.69. TCAs had a slightly higher ranking of 314 373 and corresponding relative risk of 5.09. NSSRIs had the lowest ranking of 368 680, and a corresponding relative risk of 4.06. These findings reveal that relative to other drug-AE interactions, the relationship between antidepressants and suicide is quite modest. Nevertheless, this risk of spontaneously reporting a suicide on an antidepressant is 4-5 times higher than spontaneously reporting a suicide on the combination of all other drugs. This effect is actually less than what has been previously observed for increases in the rate of completed suicide in depressed populations relative to the general population [37]. Although this is a very different approach to the one presented in this paper, and the differences in suicide risk across the three classes of drugs using DuMouchel's method are much smaller than our method found, it is reassuring to note that both approaches reveal the same ordering with increased risk of suicide for TCAs relative to more modern antidepressants (i.e. SSRIs and NSSRIs).

DISCUSSION

We have developed a new approach to drug safety monitoring, which provides a method of comparing the rate of an AE simultaneously for all individual drugs within a class of drugs (e.g. antidepressants) with the overall rate of AE of the class as a whole. Using this method, we can determine whether certain drugs within a particular class of drugs are associated with an AE of interest. The exploration of differential AE rates for individual drugs within a class of drugs may hold a key to the etiology of a particular drug-AE relationship. The method can be applied to any class of drugs and to any AE. Furthermore, the methodology is not restricted to SRSs but could just as easily and perhaps more fruitfully be applied to longitudinal medical records and/or claims databases. In this case, the added information of the number of patients at risk will eliminate bias produced by restricting the analysis to only those patients who report the AE.

With respect to suicide, we have found that the older antidepressants, the TCAs, have significantly higher overall suicide adverse report rates per million prescriptions than the newer antidepressant medications (i.e. SSRIs and NSSRIs). When examined for individual drugs within classes, a clear picture of higher risk of completed suicide AE reports for TCAs and lower rates for SSRIs and NSSRIs was found using our new methods. These findings are supported by previous findings [10-14], all showing that SSRIs are inversely associated with completed suicide rates in both children and adults, and that in adults, TCAs are positively related to completed suicide. These results do not necessarily contradict findings by the FDA [4] showing that SSRIs and NSSRIs are associated with a 2 per cent increased risk of suicidal ideation relative to placebo in children, since suicidal thinking and suicide completion are quite different endpoints. It may be possible for a drug to increase suicidal thinking but decrease suicide completion. Nevertheless, these data do not support the view that the newer antidepressants differentially lead to the emergence of suicidal thinking and behavior that ultimately leads to completed suicide [38]. Finally, these types of analyses do not support causal inferences. As pointed out in the following, there are numerous possible selection effects that could produce the types of differences between drug classes that we have observed here. These methods are for the purpose of surveillance only, and once a signal is identified, results must be verified using additional studies including large-scale RCTs if feasible and warranted.

There are numerous limitations of any analysis of the AERS data. First, there is confounding by selection. For example, patients with higher suicide risk may be more likely to be placed on an SSRI or NSSRI, which has a better safety profile than the TCA. However, if this were the case, we would expect higher rates of AE reports for SSRIs and NSSRIs relative to TCAs, and in fact, we find just the opposite. An alternative explanation is that treatment with the older class of antidepressant medications is an indicator of limited access to quality mental health care, resulting in poorer medical outcomes. In addition, overdose with TCAs is often fatal, whereas overdose with SSRIs and NSSRIs is generally not. Second, the rate of spontaneous AE reports should in general dramatically underestimate the true rate of the AEs in the population. This may be a smaller problem for suicide completion, where a life-ending AE would be reported to the FDA with higher probability than a non-fatal event. Third, the AERS provides no denominator, reflecting the number at risk in the population. By adding prescription rate data to the analysis, we have at least added a proxy for the denominator in several of the analyses that we performed. Finally, it is unclear how representative patients who end up in the AERS are of patients in the general population. Again, given the severity of the AE, it is unlikely that only a select subset of suicides would be identified by FDA's AERS. Many modeling approaches have been tested to mitigate the limitations of this database with different strengths and weaknesses [39], but clearly, factors such as publicizing suspected associations will increase reports that are partly a potential false signal [40]. Publicity tends to affect reports from consumers more than reports from professionals but clearly limits the usefulness of AERS as a way of detecting meaningful associations once the public is aware of a suspected association.

While this is the first comprehensive analysis of the AERS data for the relationship between antidepressants and suicides, the most novel aspect of the analysis is the development of rate multipliers and corresponding confidence intervals under EB- and FB-modeling frameworks. These rate multipliers allow us to determine whether individual drugs within a drug class exhibit elevated risk of a particular AE. Indeed this methodology can be applied to any potential drug-AE interaction. There is no limitation on the number of drugs that can be examined, and in fact, as the number of drugs increase, uncertainty in the estimated variance components will decrease, and the EB and FB methods should produce more similar results (although this will not compensate for true non-normality of the random-effect distribution). In the present example, all three methods of estimating the rate multipliers and associated confidence regions yielded rather similar substantive conclusions; however, the EB estimates appeared to provide

narrower confidence regions because they do not fully account for uncertainty in the variance component estimates, and perhaps because these estimates hinge on the assumption that the random effects are normally distributed. This limitation is addressed in the FB treatment of the problem, and the contrast between parametric and semiparametric methods nicely explicates the effect of assuming normality of the random-effect distribution. Indeed, the observed posterior distributions of the drug-specific effects appear to be well characterized by a mixture of 10 distinct distributions characterized by shifts in mean, variance, and skewness.

Nevertheless, fairly similar posterior mean estimates of the rate multipliers were observed for both the normal and Dirichlet process priors in the FB models, indicating that the method is fairly robust to the exact specification of the prior distribution. Even the EB method generally led to fairly similar posterior mean estimates as the two FB alternatives. However, we note that the 95 per cent posterior intervals of the rate multipliers did substantially differ across the parametric EB, parametric FB, and semiparametric FB approaches, with the semiparametric FB approach consistently yielding the widest posterior intervals, and the parametric EB approach consistently yielding the smallest intervals. The parametric EB approach yielded smaller 95 per cent intervals than the parametric FB, because the latter approach better accounts for the uncertainty of the rate multipliers through a prior distribution on the normal hyperparameters of the random effects. Also, the parametric FB approach yielded smaller 95 per cent intervals than the semiparametric FB. This is because the semiparametric approach better accounts for the uncertainty in the distribution of the random effects, with the former approach only supporting normal random-effect distributions and the latter approach supporting all random-effect distributions including the normal. In practice, the semiparametric approach is probably the method of choice, since among the three methods, it best accounts for the uncertainty in all the parameters of the Poisson mixed-effects model. Furthermore, the non-parametric characterization of the prior distribution can be informative in and of itself.

While the use of the AERS data in this paper is an important illustration and has enhanced our understanding of drug-specific antidepressant effects on completed suicide, this work also points to improvements in the existing AERS-based surveillance system that can help protect our nation's health. First, the data from AERS are severely limited in terms of their quality and the numerous selection effects that can bias their interpretation. Despite these known limitations, these are the data that FDA primarily relies upon for the purpose of post-marketing surveillance [1]. We recommend that systematic examination of AE rates for specific classes of drugs also make use of large-scale longitudinal records databases such as the VA database [16]. These data are longitudinally collected, cohorts can be identified based on being diagnosed with a specific disease (e.g. major depressive disorder), and appropriate denominators are directly available (i.e. number of patients at risk). There may still be underreporting of morbidity data, but quite complete cause of death information could be obtained from the National Death Index. While the VA data are largely for older adult males, good alternative databases that cover a demographically broader population include the PHARMetrics database in the U.S. [8] and the PHARMO database in the Netherlands [41]. Furthermore, unlike the AERS data in which information on relevant covariates is sparse, more complete information on patient characteristics, concomitant drug use and prior history of the AE are systematically available and can be directly incorporated into the model.

Second, we recommend that the existing data mining procedures be extended. The surveillance system that searches for drug-AE interactions should use analytic methods that correct for denominators and include mixed-effects modeling examining comparison of individual and classes of drugs within broad categories of drugs as we have done in this paper.

ACKNOWLEDGEMENTS

This work was supported by NIMH grants MH46745 (J.J.M.), MH65556 (R.D.G., D.K.B., K.H.), R56 MH078580 (R.D.G., C.H.B., D.K.B., K.H.), and MH40859 (C.H.B.), and AHRQ ('the Agency for Healthcare Research and Quality') CERT grant 1U18HS016973-01 (R.D.G., D.K.B., K.H.).

Contract/grant sponsor: NIMH; contract/grant numbers: MH46745, MH65556, R56 MH078580, MH40859 Contract/grant sponsor: AHRQ; contract/grant number: 1U18HS016973-01

REFERENCES

1. Institute of Medicine. The Future of Drug Safety: Promoting and Protecting the Health of the Public. 2006
2. DuMouchel W. Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system (with discussion). *The American Statistician* 1999;53:177–202.
3. Leslie LK, Newman TB, Chesney PJ, Perrin JM. The food and drug administration's deliberations on antidepressant use in pediatric patients. *Pediatrics* 2005;116:195–204. [PubMed: 15995053]
4. Hamad TA, Laughren T, Racoosin J. Suicidality in pediatric patients treated with antidepressant drugs. *Archives of General Psychiatry* 2006;63:332–339. [PubMed: 16520440]
5. U.S. Food and Drug Administration Clinical Review. Relationship Between Antidepressant Drugs and Suicidality in Adults. [accessed on 22 July 2007]. Available from: <http://www.fda.gov/ohrms/dockets/ac/06/briefing/2006-4272b1-01-FDA.pdf>
6. Mann JJ, Emslie G, Baldessarini RJ, Beardslee W, Fawcett JA, Goodwin FK, Leon AC, Meltzer HY, Ryan ND, Shaffer D, Wagner KD. ACNP Task Force Report on SSRIs and Suicidal Behavior in Youth. *Neuropsychopharmacology* 2006;31:473–492. [PubMed: 16319919]
7. Isacson G, Boëthius G, Bergman U. Low level of antidepressant prescription for people who later commit suicide: 15 years of experience from a population-based drug database in Sweden. *Acta Psychiatrica Scandinavica* 1992;85:444–448. [PubMed: 1642127]
8. Valuck RJ, Libby AM, Sills MR, Giese AA, Allen RR. Antidepressant treatment and risk of suicide attempt by adolescents with major depressive disorder: a propensity-adjusted retrospective cohort study. *CNS Drugs* 2004;18:1119–1132. [PubMed: 15581382]
9. Olfson M, Shaffer D, Marcus SC, Greenberg T. Relationship between antidepressant medication treatment and suicide in adolescents. *Archives of General Psychiatry* 2003;60:978–982. [PubMed: 14557142]
10. Gibbons RD, Hur K, Bhaumik DK, Mann JJ. The relationship between antidepressant prescription rates and rate of early adolescent suicide. *American Journal of Psychiatry* 2006;163:1898–1904. [PubMed: 17074941]
11. Gibbons RD, Hur K, Bhaumik DK, Mann JJ. The relationship between antidepressant medication use and rate of suicide. *Archives of General Psychiatry* 2005;62:165–172. [PubMed: 15699293]
12. Ludwig J, Marcotte D. Anti-depressants, suicide and drug regulation. *Journal of Policy Analysis and Management* 2005;24:249–272. [PubMed: 15776534]
13. Grunebaum M, Ellis S, Li S, Oquendo M, Mann J. Antidepressants and suicide risk in the United States, 1985-1999. *Journal of Clinical Psychiatry* 2004;65:1456–1462. [PubMed: 15554756]
14. Bramness JG, Walby FA, Tverdal A. The sales of antidepressants and suicide rates in Norway and its counties 1980-2004. *Journal of Affective Disorders* 2007;102:1–9. [PubMed: 17223200]
15. Posner K, Oquendo MA, Stanley B, Davies M, Gould M. Columbia classification algorithm of suicide assessment (C-CASA): classification of suicidal events in the FDA's pediatric suicidal risk analysis of antidepressants. *American Journal of Psychiatry* 2007;164:1035–1043. [PubMed: 17606655]
16. Gibbons RD, Brown CH, Hur K, Marcus SM, Bhaumik DK, Mann JJ. The relationship between antidepressants and suicide: results of analysis of the veterans health administration datasets. *American Journal of Psychiatry* 2007;164:1044–1049. [PubMed: 17606656]
17. Simon GE, Savarino J, Operskalski B. Suicide risk during antidepressant treatment. *American Journal of Psychiatry* 2006;163:41–47. [PubMed: 16390887]

18. Simon GE, Savarino J. Suicide attempts among patients starting depression treatment with medications or psychotherapy. *American Journal of Psychiatry* 2007;164:1029–1034. [PubMed: 17606654]
19. Hedeker, D.; Gibbons, RD. *Longitudinal Data Analysis*. Wiley; New York: 2006.
20. Bock RD, Aitkin M. Marginal maximum likelihood estimation of item parameters: an application of the EM algorithm. *Psychometrika* 1981;46:443–459.
21. Thomas, N.; Longford, NT.; Rolph, J. A statistical framework for severity adjustment of hospital mortality rates. Working Paper; RAND, Santa Monica, CA: 1992.
22. Longford NT. Logistic regression with random coefficients. *Computational Statistics and Data Analysis* 1994;17:1–15.
23. Seltzer MH, Wong WH, Bryk AS. Bayesian inference in applications of hierarchical models: issues and methods. *Journal of Educational and Behavioral Statistics* 1996;21:131–167.
24. Ferguson TS. A Bayesian analysis of some nonparametric problems. *Annals of Statistics* 1973;1:209–230.
25. El-Sayyad GM. Bayesian and classical analysis of Poisson regression. *Journal of the Royal Statistical Society, Series B (Methodological)* 1973;3:445–451.
26. Lee Y, Nelder JA. Hierarchical generalized linear models. *Journal of the Royal Statistical Society, Series B (Methodological)* 1996;4:619–678.
27. Albert J. A Bayesian analysis of a Poisson random effect models for home run hitters. *The American Statistician* 1992;46:246–253.
28. Cooil B. Using medical malpractice data to predict the frequency of claims: a study of Poisson process models with random effects. *Journal of the American Statistical Association* 1991;86:285–295.
29. Normand ST, Glickman ME, Gatsonis CA. Statistical methods for profiling providers of medical care: issue and applications. *Journal of the American Statistical Association* 1997;92:803–814.
30. Christiansen C, Morris C. Hierarchical Poisson regression models. *Journal of the American Statistical Association* 1997;92:618–632.
31. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* 2000;10:325–337.
32. Kleinman KP, Ibrahim JG. A semi-parametric Bayesian approach to generalized linear mixed models. *Statistics in Medicine* 1998;17:2579–2596. [PubMed: 9839349]
33. Karabatsos G. Bayesian nonparametric model selection and model testing. *Journal of Mathematical Psychology* 2006;50:123–148.
34. Karabatsos G, Walker SG. On the normalized maximum-likelihood and Bayesian decision theory. *Journal of Mathematical Psychology* 2006;50:517–520.
35. Escobar MD, West M. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* 1995;90:577–588.
36. Neal RM. Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics* 2000;9:249–265.
37. Gould MS, Greenberg T, Velting DM, Shaffer D. Youth suicide risk and preventive interventions: a review of the past 10 years. *Journal of the American Academy of Child and Adolescent Psychiatry* 2003;42:386–405. [PubMed: 12649626]
38. Fergusson D, Doucette S, Glass KC, Shapiro S, Healy D, Herbert P, Hutton B. Association between suicide attempts and selective serotonin reuptake inhibitors: systemic review of randomized controlled trials. *British Medical Journal* 2005;330:396–399. [PubMed: 15718539]
39. Thomas EJ, Petersen L. Measuring errors and adverse events in health care. *Journal of General Internal Medicine* 2003;18:61–67. [PubMed: 12534766]
40. Mosholder AD, Pamer CA. Postmarketing surveillance of suicidal adverse events with pediatric use of antidepressants. *Journal of Child and Adolescent Psychopharmacology* 2006;16:33–36. [PubMed: 16553527]
41. Gibbons RD, Brown CH, Hur K, Marcus SM, Bhaumik DK, Erkens JA, Herings RMC, Mann JJ. Early evidence on the effects of the FDA black box warning on SSRI prescriptions and suicide in children and adolescents. *American Journal of Psychiatry* 2007;164:1356–1363. [PubMed: 17728420]

Figure 1.

(a) Rate of completed suicide adverse event reports per million prescriptions by year and drug class for children under age 18 and (b) rate of completed suicide adverse event reports per million prescriptions by year and drug class for adults 18 and over.

Figure 2.
Rate of completed suicide adverse event reports per million prescriptions by year and drug (1998-2004).

Figure 3.
Rate multipliers—Bayes estimates and confidence intervals—all ages.

Figure 4.
Posterior distributions of rate multipliers.

Table 1
Comparison of rate multiplier estimates from EB, parametric FB, and semiparametric FB approaches

Drug	EB				Parametric FB				Semiparametric FB			
	Mean	2.5 per cent	97.5 per cent	Width	Mean	2.5 per cent	97.5 per cent	Width	Mean	2.5 per cent	97.5 per cent	Width
Celexa	0.59	0.51	0.68	0.17	0.60	0.37	0.92	0.55	0.60	0.34	0.98	0.64
Paxil	0.51	0.46	0.57	0.11	0.54	0.34	0.82	0.48	0.54	0.31	0.88	0.57
Prozac	0.62	0.57	0.68	0.11	0.66	0.41	0.99	0.58	0.66	0.37	1.07	0.70
Luvax	1.06	0.80	1.39	0.59	1.15	0.66	1.85	1.19	1.06	0.60	1.73	1.13
Zoloft	0.36	0.32	0.40	0.08	0.38	0.23	0.58	0.35	0.38	0.21	0.61	0.40
Serzone	0.24	0.14	0.39	0.25	0.24	0.12	0.43	0.31	0.26	0.10	0.51	0.41
Remeron	1.03	0.85	1.24	0.39	1.09	0.67	1.68	1.01	1.04	0.59	1.70	1.11
Wellbutrin	0.54	0.47	0.62	0.15	0.57	0.35	0.87	0.52	0.55	0.31	0.89	0.58
Effexor	0.50	0.42	0.58	0.16	0.52	0.32	0.81	0.49	0.54	0.31	0.89	0.58
Elavil	0.95	0.84	1.07	0.23	1.00	0.61	1.51	0.90	1.03	0.59	1.68	1.09
Asendin	4.28	2.24	8.17	5.93	5.15	2.09	10.33	8.24	5.65	2.52	10.45	7.93
Anafranil	1.77	1.22	2.57	1.35	1.86	1.01	3.16	2.15	1.70	0.90	2.89	1.99
Norpramine	3.59	2.61	4.94	2.33	3.41	1.90	5.86	3.96	3.57	1.80	6.28	4.48
Sinequan	1.71	1.39	2.09	0.70	1.78	1.07	2.80	1.73	1.71	0.95	2.84	1.89
Tofranil	1.31	1.00	1.72	0.72	1.35	0.77	2.17	1.40	1.52	0.76	2.62	1.86
Aventyl	0.97	0.77	1.20	0.43	1.01	0.61	1.56	0.95	1.04	0.59	1.69	1.10
Triptil	0.84	0.29	2.39	2.10	0.86	0.19	2.13	1.94	0.90	0.24	2.12	1.88
Surmontil	4.72	3.00	7.43	4.43	4.85	2.42	8.65	6.23	5.61	2.78	10.07	7.29

Table II

Comparison of population-level parameters

	Empirical Bayes			Parametric Bayes			Semiparametric Bayes		
	Estimate	2.5 per cent	97.5 per cent	Mean	2.5 per cent	97.5 per cent	Mean	2.5 per cent	97.5 per cent
Intercept	-12.00	-12.39	-11.60	-12.01	-12.45	-11.57	-12.02	-12.53	-11.49
Slope	0.34	0.28	0.40	0.34	-0.44	1.12	0.34	0.01	0.67
var(int)	0.70	0.20	1.19	0.84	0.39	1.69	1.92	0.80	4.39
var(slp)	0.01	0.00	0.02	2.67	1.35	5.24	1.02	0.43	2.39
cov(int,slp)	-0.01	-0.07	0.05	-0.02	-0.77	0.76	-0.02	-1.00	0.97
Precision parameter							12.07	1.55	28.19
Number of classes							10.19	8.00	13.00