Self-controlled methods for postmarketing drug safety surveillance in large-scale longitudinal data

Shawn E. Simpson

Submitted in partial fulfillment of the

requirements for the degree

of Doctor of Philosophy

in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2011

ABSTRACT

Self-controlled methods for postmarketing drug safety surveillance in large-scale longitudinal data

Shawn E. Simpson

A primary objective in postmarketing drug safety surveillance is to ascertain the relationship between time-varying drug exposures and adverse events (AEs) related to health outcomes. Surveillance can be based on longitudinal observational databases (LODs), which contain time-stamped patient-level medical information including periods of drug exposure and dates of diagnoses. Due to its desirable properties, we focus on the self-controlled case series (SCCS) method for analysis in this context. SCCS implicitly controls for fixed multiplicative baseline covariates since each individual acts as their own control. In addition, only exposed cases are required for the analysis, which is computationally advantageous. In the first part of this work we present how the simple SCCS model can be applied to the surveillance problem, and compare the results of simple SCCS to those of existing methods.

Many current surveillance methods are based on marginal associations between drug exposures and AEs. Such analyses ignore confounding drugs and interactions and have the potential to give misleading results. In order to avoid these difficulties, it is desirable for an analysis strategy to incorporate large numbers of time-varying potential confounders such as other drugs. In the second part of this work we propose the Bayesian multiple SCCS approach, which deals with high dimensionality and can provide a sparse solution via a Laplacian prior. We present details of the model and optimization procedure, as well as results of empirical investigations.

SCCS is based on a conditional Poisson regression model, which assumes that events at different time points are conditionally independent given the covariate process. This requirement is problematic when the occurrence of an event can alter the future event risk. In a clinical setting, for example, patients who have a first myocardial infarction (MI) may be at higher subsequent risk for a second. In the third part of this work we propose the positive dependence self-controlled case series (PD-SCCS) method: a generalization of SCCS that allows the occurrence of an event to increase the future event risk, yet maintains the advantages of the original by controlling for fixed baseline covariates and relying solely on data from cases. We develop the model and compare the results of PD-SCCS and SCCS on example drug-AE pairs.

Table of Contents

List of Figures

List of Tables

List of Acronyms

- **AE** = adverse event
- **AIC** = Akaike information criterion
- **AUC** = area under ROC curve
- **BBR** = Bayesian binary regression
- **BCPNN** = Bayesian confidence propagation neural network
- **BIC** = Bayesian information criterion
- **BLR** = Bayesian logistic regression
- **BSCCS** = Bayesian multiple SCCS
- **CCAE** = Thomson Reuters Commercial Claims and Encounters database
- **CCD** = cyclic coordinate descent
- **CI** = confidence interval
- **DOI** = drug exposure of interest
- **DP analysis** = disproportionality analysis
- **EB05** = 5th percentile of the posterior distribution for the RR
- **EBGM** = empirical Bayesian geometric mean
- **EHR** = electronic health record
- **FEP model** = fixed effects Poisson model
- **FPR** = false positive rate

- **HOI** = health outcome of interest
- **LOD** = longitudinal observational database
- **MAD** = mean absolute deviation
- **MAP estimate** = maximum a posteriori estimate
- **MAP score** = mean average precision score
- **MDCD** = Thomson Multistate Medicaid database
- **MDCR** = Thomson Medicare Supplemental database
- **MGPS** = multi-item gamma-Poisson shrinker
- **MI** = myocardial infarction
- **MSE** = mean squared error
- **MSLR** = MarketScan Lab Results database
- **OMOP** = Observational Medical Outcomes Partnership
- **OS** = observational screening
- **PAUC** = partial area under ROC curve
- **PD-SCCS** = positive dependence SCCS
- **PRR** = proportional reporting ratio
- **ROC curve** = reciever operating characteristic curve
- **ROR** = reporting odds ratio
- **RR** = reporting ratio
- **SCCS** = self-controlled case series
- **SE** = empirical standard error
- **SM** = estimated standard error (square root of the mean of the variance estimates)
- **SRS** = spontaneous reporting system

Acknowledgments

There are many people who I am greatly indebted to for making contributions to this work; they have my sincere apprecation and uttermost thanks.

First and foremost, my advisor, David Madigan, for his guidance and insight over the years and for making this work possible. Discussions with David always led to a greater sense of clarity and renewed energy for me — a remarkable feat for which I owe him a debt of gratitude. I hope that as my statistical career develops I will have even a portion of the enthusiasm and knowledge that David has, and I have been truly lucky to have him as an advisor.

George Hripscak, Tony Jebara, Martin Lindquist, David Madigan, and Zhiliang Ying for being on my defense committee and for their very helpful feedback on this work.

Everyone at the OMOP project for making their data and resources available to us and for facilitating our analyses in many ways.

Patrick Ryan and Ivan Zorych for their work on OMOP and for being so forthcoming about answering questions on method specifications, implementations, the server and many other things. The results section of this work would not exist without them.

Marc Suchard at UCLA for his remarkable speed in devising efficient CPU and GPU implementations for our methods, his help with OMOP packages, and for dynamic discussions. I think that our work with Marc on GPU-based methods has a lot of potential and I look forward to gaining experience there.

Carlos Vallarino for data that was very helpful for method comparisons.

Zhiliang Ying for providing insightful perspectives and conversations, suggesting books and references that were key in developing the PD-SCCS model, and telling many great stories.

Jonathan Campbell, Kristen Gore, Tyler McCormick, and John Zech for fruitful discussions of drafts of the PD-SCCS model at various stages.

The professors in our department for their kind words and advice over the years, particularly Richard Davis, Regina Dolgoarshinnykh, Chris Heyde, Ioannis Karatzas, Shaw-Hwa Lo, Martin Lindquist, Philip Protter, Dan Rabinowitz, Michael Sobel, Frank Wood, Zhiliang Ying and Tian Zheng; also Hammou Elbarmi, Birol Emir, Irene Hueter, and Michael Shnaidman.

The classmates in my cohort: Ivor Cribben, Xiaodong Li, Amal Moussa, Kamiar Rahnama Rad, Li Song, Johannes Ruf, Xiaoru Wu, and Chun-Yip Yau for being there along the way (we made it!), and all of the other students in the department for being great colleagues and friends.

Xinzhao Chu and Farzad Kamalabadi for taking me on as an undergraduate research assistant in the UIUC Electrical Engineering Department — their advice on research, presentations, posters, and conferences (and exposure to missing data and principal components analysis) undoubtedly helped lead me to graduate school in statistics.

The Intel Research Scholars Program for funding my undergraduate research and having a wonderful program — I hope that it continues for future students.

Hernando Ombao and Alexey Stepanov for teaching great undergraduate statistics courses at UIUC, and Jason Zych for his programming classes.

Friends Annie, Aurora, Emilia, Helen, Kelly, Kristen, and Ruth for fielding many phone calls and providing breaks from the calamity.

And finally, my parents Douglas and Fung-Yin Simpson and my brother Eric for their continual love and support — I love you all very much.

Chapter 1

Introduction

Increasing scientific, regulatory, and public scrutiny focuses on the obligation of the medical community, pharmaceutical industry, and health authorities to confirm that marketed medical products and prescription drugs have acceptable benefit-risk profiles. Ensuring drug safety is an intricate and ongoing process that begins with carefully designed randomized clinical trials prior to approval, and continues after regulatory market authorization when drugs are in widespread clinical use. This is the setting of *postmarketing surveillance*, which refers to the process of monitoring the safety of prescription drugs that have already been approved for the marketplace. Once drugs are out on the market they are taken by many more individuals and for longer periods of time than they were during pre-approval trials. Due to the large amount of data that is available, in postmarket analyses it may be possible to identify adverse events (AEs) related to drug exposure that had not been previously detected.

Both the complexity and scale of the problem of postmarketing surveillance present major challenges for statistical analysis and computation. Patients often take multiple drugs concurrently, or may take drugs in ways that differ from how they were prescribed. It is also difficult to determine patients' overall health status and whether or not they may be taking other actions that influence the frequency and/or timing of AEs. In addition to these challenges, data sources used for surveillance often contain

tens of millions of individuals who are observed over the course of several years. These individuals may experience tens of thousands of possible AEs. Furthermore, there are tens of thousands of possible drugs that patients can be exposed to, which results in millions of potential drug interactions. We are thus dealing with a large-scale problem that is also subject to the complications of high-dimensionality.

The current surveillance approach of the U.S. Food and Drug Administration (FDA) relies on its Adverse Event Reporting System (AERS), which is one of the several spontaneous reporting systems (SRSs) that are in use for surveillance. SRSs are based on data collected from spontaneous reports, which are voluntarily submitted by consumers and healthcare professionals. These reports contain self-reported information on suspected links between drug exposures and AE occurrences. Other prominent examples of SRSs include the Yellow Card Scheme of the Medicines and Healthcare Products Regulatory Agency (MHRA), and the international pharmacovigilance program of the World Health Organization (the WHO Uppsala Monitoring Center). In the post-approval environment, surveillance schemes based on SRSs represent a cornerstone for the early detection of hazards related to drug exposure.

Some of the limitations of SRSs such as AERS are that the temporal information may be unreliable, there may be bias due to under-reporting, over-reporting, or duplicate reports, and no denominator or control group is available to provide a comparison. The FDA relies primarily on adjusted 2×2 table summaries from these reports in order to deduce the relationship between drugs and AEs. Such analyses may give misleading results since they focus on marginal associations and do not adjust for the presence of confounding drug exposures. In addition, methods based on 2×2 summaries ignore potential drug interactions and do not incorporate information on the timing of AEs relative to exposures. Despite these limitations, analytic methods for spontaneous reports have attracted considerable attention in the last decade, and several different methods have become well established both in commercial software

products and in the medical literature.

In recent years, several high-profile drug safety cases have raised concerns about the strategies that are currently in use for postmarketing surveillance. One of the most prominent cases was that of Vioxx (rofecoxib), which was an anti-inflammatory drug manufactured by Merck & Co, Inc. Merck made a voluntary worldwide withdrawal of Vioxx in September 2004 due to evidence of an increased risk of cardiovascular events. At a U.S. Senate Finance Committee hearing on Vioxx in November 2004, an officer from the FDA estimated that "nearly 28,000 excess cases of heart attack or sudden cardiac death were caused by Vioxx" (Graham, 2004). As a response to the revelations from the Vioxx case and other drug safety concerns, U.S. Congress mandated that the FDA establish an *active* surveillance system to prospectively monitor the safety of marketed medical products, including marketed prescription drugs, as part of the 2007 FDA Amendments Act (FDAAA) (FDA, 2007c). The FDA responded to the passage of the FDAAA by launching the Sentinel Initiative, a long-term program whose goal is to implement a nation-wide active surveillance system. The FDAAA intends for healthcare information on at least 100 million people to be accessible by the year 2012 (Platt et al., 2009).

The AERS of the FDA is a *passive* surveillance system in which safety information is obtained only once reports are voluntarily submitted. In contrast, under an active surveillance system regulators can initiate their own safety investigations in existing data sources held by hospitals, insurance companies, and other healthcare information providers. In active surveillance, a key difference is that information in data sources is recorded automatically rather than submitted voluntarily. Data sources include medical insurance claims databases and electronic health records (EHRs), which are examples of longitudinal observational databases (LODs). These LODs contain time-stamped patient-level medical information, including periods of drug exposure and dates of diagnoses that can be used to examine safety risks.

In analyses for postmarketing surveillance in LODs, the observed outcomes are AEs related to diagnosed health conditions, e.g. myocardial infarction (MI), stroke, asthma attack, etc. Our objective in the surveillance problem is to ascertain drug safety by estimating the strength of the association between AEs and drug exposures that vary over time. We would like our analysis approach to make use of information on the frequency and timing of events, as well as control for potentially confounding drugs. In our work we focus on the *self-controlled case series* (SCCS) method, which is a model for the analysis of recurrent events whose features are well-suited for the surveillance context.

The self-controlled case series method was proposed by Farrington (1995) in order to estimate the relative incidence of AEs for assessing vaccine safety. SCCS is used to analyze recurrent events and determine their association with time-varying covariates. This method is based on a conditional Poisson regression in which the conditioning step produces two beneficial properties. First, it automatically controls for all multiplicative fixed baseline covariates without having to explicitly include them in the model. This feature is particularly appealing in situations where data provide too few baseline covariates to effectively adjust for confounding. Second, only exposed cases (individuals who were both exposed to the drug and had at least one event) need to be included in the analysis, which reduces the amount of data required and is computationally advantageous. With SCCS, each individual serves as their own control. In other words, SCCS compares outcome event rates during times when a person is exposed to outcome event rates during times when the same person is unexposed. In effect, the cases' unexposed time lets us infer expectations about what would have occurred during their exposed time had they not been exposed. Whitaker et al. (2005) provides a tutorial of SCCS and Farrington and Whitaker (2006) introduce a semi-parametric approach to the model.

Other related case-based methods have been developed within the epidemiological

literature (e.g. Navidi, 1998; Suissa, 1995), many of which are variants of the case-crossover design (Maclure, 1991). Case-crossover requires the choice of a comparator time period to serve as a control for each individual, and models exposures in case and control time intervals via conditional logistic regression. In case-crossover the exposures are outcomes and analysis is conditional on events, whereas in SCCS the events are modeled as random and exposures are conditioned upon. SCCS also differs from case-crossover in that it makes use of information available over the entire observation interval for each individual without the need to select a control period.

The problem of postmarketing surveillance in LODs must contend with millions of individuals and millions of potential drug exposure variables. In contrast, the epidemiological applications of SCCS have exclusively focused on situations with relatively small sample sizes and a manageable number of exposure variables of interest. The size of the surveillance problem presents a major computational challenge, and ensuring the availability of an efficient optimization procedure is essential for a feasible implementation.

The construction and development of the SCCS model that we present throughout this work differs from that in the work done by Farrington (e.g. Farrington, 1995; Farrington and Whitaker, 2006). In Farrington's framework, one would begin with data that is limited to case-series and, as a result, analysis proceeds conditional on the fact that only cases are being used. In our approach we argue that individual baseline rates are nuisance parameters, which can be eliminated by conditioning on their sufficient statistics. The fact that only case-series data are needed for the analysis comes through as a by-product of conditioning.

Due to this difference in development, we uncovered a connection between the SCCS method and the *fixed effects Poisson* (FEP) model for panel (longitudinal) data, which is well-established in econometrics literature (e.g. Hausman et al., 1984; Cameron and Trivedi, 1998; Winkelmann, 2000; Wooldridge, 2002). Many economet-

ric applications of the FEP model have been to panel data on patents; for example, Hausman et al. (1984) examined the relationship between time-varying R&D expenditures and firm-level patent applications over time. The FEP model differs from SCCS only in that all individuals have a common observation period length under FEP, whereas in SCCS observation length may vary by individual. Establishing this link between FEP and SCCS may help provide insight into properties of the SCCS model; for example Wooldridge (1999) gives results for consistency of the FEP estimator that hold as long as the conditional mean is correctly specified, and Hausman et al. (1984) presents a variant based on the negative binomial distribution.

The remainder of this work is organized as follows. In Chapter 2 we introduce the simple self-controlled case series model for surveillance and present analysis results based on data from LODs. In Chapter 3 we extend the SCCS model to include multiple drug exposures. We propose a regularization approach and an efficient optimization procedure that enable analysis for surveillance in the presence of very large numbers of drug exposure predictors. In Chapter 4 we propose a positive event dependence model, which avoids a restrictive conditional independence assumption that is required by the original SCCS model. Chapter 5 we provide a discussion and concluding remarks.

Chapter 2

Self-controlled case series for postmarketing surveillance

2.1 Introduction

The objective in postmarketing surveillance is to estimate the strength of the association between adverse events (AEs) and drug exposures that vary over time. The longitudinal observational databases (LODs) used in an active surveillance system contain individual-level data on both the timing and frequency of AE occurrences, and we would like to incorporate this information into our analysis approach. We focus on the use of the self-controlled case series (SCCS) model for this purpose. SCCS is based on a conditional Poisson model for recurrent events, where the conditioning step yields two main benefits. First, the model automatically controls for all multiplicative individual-level confounders that are fixed in time. Second, inference in SCCS depends only on exposed cases (individuals who are both exposed to the relevant drug and have at least one occurrence of the AE of interest), which substantially reduces the amount of data required in the analysis.

SCCS has been used primarily in the context of vaccine safety studies and epi-

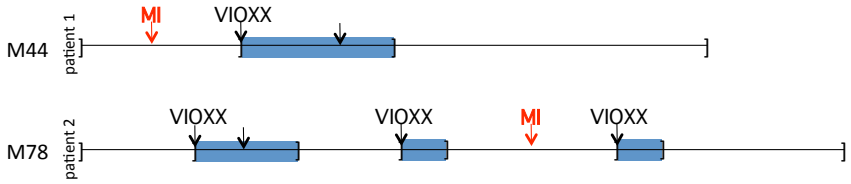
demiological applications to provide further assessment of previously established associations. In contrast, postmarketing surveillance centers around *risk identification*, which refers to the identification of new drug-AE associations. Risk identification presents a novel area of application for the SCCS model. Furthermore, an active surveillance system examines the strength of the potential relationship between all possible pairs of drug exposures and AEs, rather than for a specific few. The SCCS model has never been applied within this framework of large-scale surveillance, nor has its use for this purpose been empirically evaluated. In this Chapter we endeavor to take on both of these objectives. We present how SCCS can be applied to the problem of active surveillance in LODs, where the goal is to identify new assocations on a large scale. We compare the results of applying the simple SCCS model on a claims database to those of disproportionality analysis (DP) applied to AERS data. DP encompasses the primary class of analytic methods that are currently used for surveillance in spontaneous reporting systems (SRSs) such as AERS. We examine five example drug-AE pairs, and investigate whether using SCCS to analyze claims data would have led to earlier detection of drug-AE associations than implementing a DP analysis on SRS data. We also give results of using SCCS to discriminate between drug-AE pairs with and without associations based on data from several LODs. We compare these SCCS pair classification results to those obtained from other surveillance methods.

2.2 Notation and model framework

We will focus here on the *simple* SCCS model in which there is one drug exposure and one outcome AE of interest; this is analogous to the case of a simple linear regression with one predictor and one response variable. To set up notation, let $i = 1, \ldots, N$ index individuals. Events and exposures in LODs are recorded by date, so temporal information is available down to the level of days. Days are indexed by $d = 1, \ldots, \tau_i$,

where τ_i is the total number of days that person i is observed and (i, d) is their dth day of observation. The number of events on day (i, d) is denoted by y_{id}, and the drug exposure status on that day is indicated by x_{id}, where $x_{id} = 1$ if i is exposed to the drug on day d, and 0 otherwise. The observed sequence of exposures for i is given by the vector $\mathbf{x}_i = (x_{i1}, \ldots, x_{i\tau_i})^{\mathrm{T}}$, and observed event counts on each day are represented by $\mathbf{y}_i = (y_{i1}, \ldots, y_{i\tau_i})^{\mathrm{T}}$. Figure 2.1 below shows example longitudinal data for two individuals where Vioxx is the relevant drug exposure and myocardial infarction (MI) is the AE of interest.

Figure 2.1: Longitudinal traces for a 44 year-old male and a 78 year-old male who both experienced MI and were exposed to Vioxx during the course of their observation.



SCCS assumes that AEs arise according to a non-homogeneous Poisson process, where the time-varying event rate is modulated based on drug exposure status. The model presumes that person i has an individual baseline event rate of e^{ϕ_i}, which is constant over time, and that periods of drug exposure result in a multiplicative effect of e^{β} on the individual baseline rate. In other words, the Poisson event rate for person i on day d can be written as

$$\lambda_{id} = e^{\phi_i + \beta x_{id}},$$

and e^{β} gives the relative risk of an AE during exposure. The number of events y_{id} conditional on the current exposure status x_{id} is distributed as a Poisson random variable with density

$$p(y_{id} \mid x_{id}) = \frac{e^{-\lambda_{id}} \lambda_{id}^{y_{id}}}{y_{id}!}.$$

The Poisson likelihood for person i over their entire observation period is the joint

density of their observed events, given observed exposures. This can be written as

$$L_i = p(\mathbf{y}_i \mid \mathbf{x}_i) = \prod_{d=1}^{\tau_i} p(y_{id} \mid x_{id}) = e^{-e^{\phi_i}\sum_d e^{\beta x_{id}}} e^{\phi_i \sum_d y_{id}} \prod_{d=1}^{\tau_i} \frac{1}{y_{id}!} e^{\beta x_{id} y_{id}}. \tag{2.1}$$

There are two assumptions implicit in the Poisson model that allow us to write the likelihood expression in the factorized form in (2.1), namely:

(i) Events are conditionally independent given exposures

$$y_{id} \perp\!\!\!\perp y_{id'} \mid \mathbf{x}_i \quad \text{for } d \neq d'$$

(ii) Past events are conditionally independent of future exposures given the current exposure status

$$y_{id} \perp\!\!\!\perp x_{id'} \mid x_{id} \quad \text{for } d \neq d'$$

These assumptions are likely to be violated in practice, e.g. one might expect that having an MI may increase the future risk of an MI, which would violate (i), or impact future drug usage, which would violate (ii). In Chapter 4 we propose a model that circumvents assumption (i) by allowing for a positive dependence parameter to be added to the baseline event rate each time that an AE occurs. In the discussion in Chapter 5 we will consider other methods and future directions of work that deal with these assumptions and other modeling difficulties.

If individuals are independent, the full likelihood is the product of the individual L_i likelihood contributions. One could then maximize the full log-likelihood over all individuals in order to estimate the parameters. However since our main objective is to assess drug safety, the drug effect β is of primary interest and the person-specific ϕ_i effects can be treated as *nuisance parameters*. A further complication is that the LODs used for postmarketing surveillance can contain well over 10 million patients. Since the dimension of the vector of person-specific parameters $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_N)^{\mathrm{T}}$ is equal to the number of individuals N, estimation of $\boldsymbol{\phi}$ would call for optimization in an ultra high-dimensional space and presumably would be computationally prohibitive.

In order to avoid estimating these individual nuisance parameters, we can condition on their sufficient statistics. This will remove the dependence on ϕ_i from the likelihood expression. Under the Poisson model likelihood given in (2.1), the total number of events observed for person i, denoted by $n_i = \sum_d y_{id}$, is a sufficient statistic for the ϕ_i parameter. Since events follow a non-homogeneous Poisson process, the number of events n_i is also distributed as Poisson with mean equal to the cumulative event rate over the entire observation period, i.e.

$$n_i \mid \mathbf{x}_i \sim \text{Poisson}(e^{\phi_i} \sum_{d=1}^{\tau_i} e^{\beta x_{id}}).$$

In our case the cumulative intensity is a sum (rather than an integral) because we assume a constant event rate within each day. Conditioning on n_i yields the SCCS likelihood contribution for i, which is

$$L_i^c = P(\,\mathbf{y}_i \mid \mathbf{x_i},\, n_i\,) = \frac{P(\,\mathbf{y}_i \mid \mathbf{x}_i\,)}{P(\,n_i \mid \mathbf{x}_i\,)} \propto \prod_{d=1}^{\tau_i} \left(\frac{e^{\,\beta x_{id}}}{\sum_{d'} e^{\,\beta x_{id'}}} \right)^{y_{id}}. \tag{2.2}$$

Notice that because n_i is sufficient, the individual likelihood in the above expression no longer contains ϕ_i. This conditional likelihood takes the form of a multinomial, but it differs from a typical multinomial regression (e.g. see Agresti, 2002) in that the number of "bins" (observed days τ_i) varies by person, the β parameter is constant across days, and the covariates x_{id} vary by day.

The full SCCS likelihood is the product of the individual likelihoods, which is

$$L^c \propto \prod_{i=1}^{N} \prod_{d=1}^{\tau_i} \left(\frac{e^{\,\beta x_{id}}}{\sum_{d'} e^{\,\beta x_{id'}}} \right)^{y_{id}}. \tag{2.3}$$

In the SCCS model the drug effect parameter is estimated by maximizing the conditional log-likelihood corresponding to (2.3) to obtain $\hat{\beta}$. It is clear from the expresson for the SCCS likelihood in (2.3) that if person i has no observed events ($\mathbf{y}_i = \mathbf{0}$), they will have a likelihood contribution of $L_i^c = 1$. Such a person would have no effect on the estimation, and as a consequence only cases ($n_i \geq 1$) need

to be included in the analysis. SCCS does a within-person comparison of the event rate during exposure to the event rate while unexposed, and thus the method is "self-controlled". Intuitively it follows that if i has no events they cannot provide any information about the relative rate at which they have events. That the SCCS analysis relies solely on cases is a substantial computational advantage: since the incidence rate of most AEs is relatively low, typical SCCS analyses will utilize only a modest fraction of the total number of individuals. In fact, even from the limited pool of patients who had AEs, only those that experienced drug exposure during their observation need to be included; this feature will be discussed in Section 3.2.2 of the subsequent Chapter 3.

In the LODs used for surveillance, AEs are identified on the basis of diagnosis codes associated with a particular date. This means that at most one AE can occur on a given day, and so in reality $y_{id} = 0$ or 1 for all days d. Since in practice y_{id} is a binary variable, one could apply a logistic (Bernoulli) model within each day rather than a Poisson model and perform a conditional analysis in a similar manner. In Appendix B we show that the SCCS model is equivalent to this conditional Bernoulli within-day model, which is a result of days being treated as the smallest increment of time in which an event can occur in the non-homogeneous Poisson process that SCCS is based upon (Cook and Lawless, 2007).

2.3 Comparison with disproportionality analysis

2.3.1 Background on DP analysis

Current global postmarketing surveillance efforts rely predominantly on SRSs, including the Adverse Event Reporting System (AERS) of the U.S. FDA. These systems were designed in order to provide early warnings for safety problems that were not detected during pre-approval clinical development (Hauben et al., 2005). A number

of methods have been applied to the analysis of SRSs, e.g. Praus et al. (1993); van Puijenbroek et al. (2000); Orre et al. (2005), however the most widely used methods are examples of disproportionality analysis (DP) methods. These methods quantify pairs of drug exposures and AEs based on the degree to which the drug-AE combination co-occurs "disproportionally" often compared to what would be expected if there were no association (Bate and Evans, 2009).

Individual records in SRS databases typically include limited demographic information (e.g. age and sex), date of report, one or more drugs and one or more AEs. Well-known examples of DP methods include the multi-item gamma-Poisson shrinker (MGPS) (DuMouchel, 1999), proportional reporting ratio (PRR) (Evans et al., 2001), reporting odds ratio (ROR) (Rothman et al., 2004), and the Bayesian confidence propagation neural network (BCPNN) (Bate et al., 1998). DP methods are based on 2×2 contingency tables for all possible pairs of drugs and AEs, where each table classfies reports according to the presence of a given drug-AE combination. An example 2×2 table categorizing the number of reports with co-occurrences of drug x and AE y is shown in Table 2.1.

Table 2.1: A 2×2 table for drug x and AE y

		Drug x: Yes	Drug x: No
AE y	Yes	w_{00}	w_{01}
	No	w_{10}	w_{11}

In our data analysis we focus on one widely-used DP method, MGPS, to provide a comparison for the simple SCCS surveillance approach. MGPS focuses on the reporting ratio (RR), which is the number of co-occurrences of the drug-AE combination

(where both are present) divided by the expected number of occurrences if the drug and AE were independent. In Table 2.1 the reporting ratio for drug x and AE y is

$$RR = \frac{w_{00}(w_{00} + w_{01} + w_{10} + w_{11})}{(w_{00} + w_{01})(w_{10} + w_{11})}.$$

RR is a sensible measure of disproportionality since if a drug is more likely to cause a particular AE than another drug, it will typically receive a higher score. If a drug and AE are stochastically independent, the measure will return a value of one. In the SRS context, however, the number of co-occurrences of a drug and AE (i.e. the count w_{00} in Table 2.1) is often small due to low AE incidence rates. This can lead to substantial variability in disproportionality measures even though there are often large numbers of reports overall. MGPS adopts a Bayesian approach in order to address the issue of variability. MGPS imposes a prior distribution on RRs that shrinks estimates toward their average value, which is typically close to one. As a result, strong evidence is required from the data in order to return an MGPS estimate of RR that is substantially larger than one, thereby reducing the chance that an association will be falsely identified due to high variability (Zorych et al., 2011). The MGPS estimate of RR is the Empirical Bayesian Geometric Mean (EBGM), which is the geometric mean of the posterior distribution of the true RR. It is suggested in DuMouchel (1999) that the 5th percentile of the posterior distribution (EB05) be used to provide a more conservative measure than EBGM, where an EB05 ≥ 2 provides evidence of an association.

2.3.2 Results

We will focus on five drug-AE combinations in order to compare the use of DP and SCCS for surveillance. To analyze these five combinations, we applied the simple SCCS method on a large-scale claims database and the MGPS method on SRS data from the AERS database. The drug-AE pairs that we examined were:

1. rofecoxib (Vioxx) and myocardial infarction (MI)

2. exenatide and acute pancreatitis

3. lapatinib and hepatotoxicity

4. estrogen and MI

5. simvastatin and cataracts

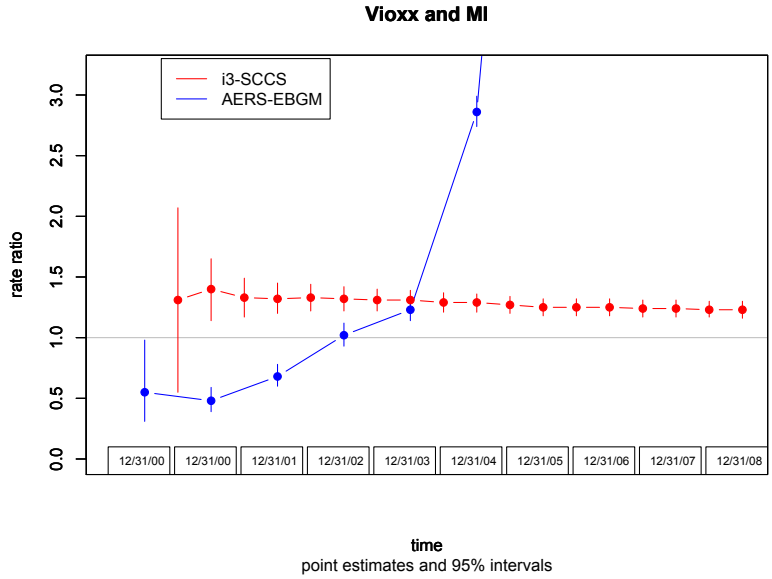
Evidence of an assocation exists for the first four analysis pairs. Rofecoxib (Vioxx) was withdrawn worldwide in 2004 due to an increased risk of cardiovascular events, including MI (Bresalier et al., 2005). In October 2007 the exenatide label was altered to include a warning for acute pancreatitis, and the FDA issued a corresponding alert (FDA, 2007b). However several previous SRS-based DP analyses yielded false negative findings for this pair (Hauben and Hochberg, 2008). Lapatinib recieved a label warning for hepatotoxicity in July 2008, but has not yet been confirmed as a signal of suspected causality (FDA, 2008). An association between estrogen and MI has been identified in some clinical trials, e.g. Manson et al. (2003). We included the last pair, simvastatin and cataracts, as a false positive that was previously identified as an association but was later discounted (Hauben et al., 2006). We would like to compare how SCCS and the MGPS DP method perform in the face of this challenge.

The data source for our SCCS analysis was the i3 InVision Data Mart^{TM} database, which contains longitudinal claims records from the United HealthCare insurance plans, including insurance eligibility, pharmacy claims, medical claims, inpatient and outpatient services utilization, and procedures with the associated diagnoses and costs. The enrollees of the United HealthCare plans have both medical and prescription benefit coverage, and are primarily a population of employees and their dependents. All data are de-identified to comply with the Health Insurance Portability and Accountability Act (HIPAA), a federal law that establishes standards for

protecting the privacy and security of health information. The data covers over 36 million individuals in the United States between May 1, 2000 and December 31, 2008. Further details are given in Madigan et al. (2011).

The plots in Figure 2.2–2.6 show relative risk estimates and confidence intervals from both the SCCS and MGPS approaches for the 5 drug-AE pairs over time. The values shown for MGPS are EBGM estimates of the RR based on AERS data, and the SCCS values are relative risks (e^{β}) based on the i3 claims data. The SCCS relative risk and the EBGM reporting ratio are on comparable scales (the RR acts as a surrogate for relative risk). Each risk estimate in the plots was generated using only the portion of the data that was available up to the specified time point, and bars correspond to 95% confidence intervals. We compare the estimates from SCCS and MGPS over time in order to determine if an SCCS analysis in claims data could yield earlier detection of an association than a DP analysis in AERS data.

Figure 2.2: Vioxx and MI



For the Vioxx-MI pair in Figure 2.2, we can see that SCCS estimates a relative risk that is statistically significant (i.e. the 95% confidence interval exceeds one)

Figure 2.3: Exenatide and pancreatitis

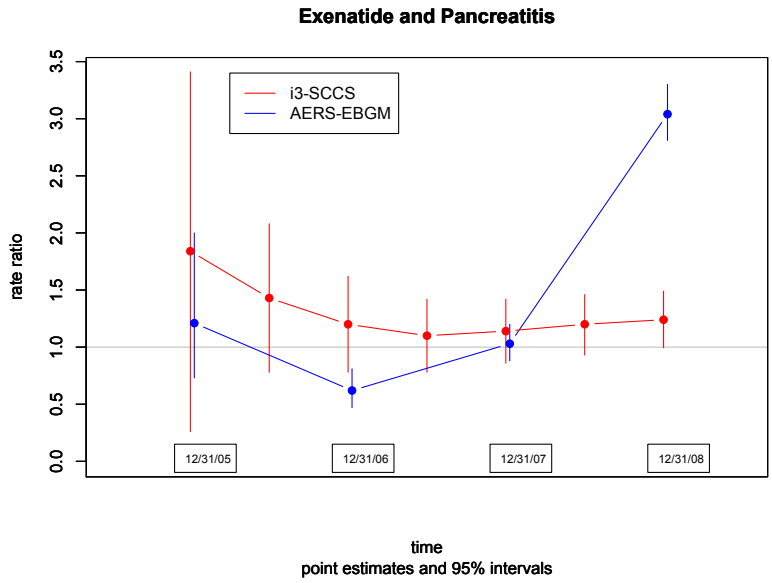
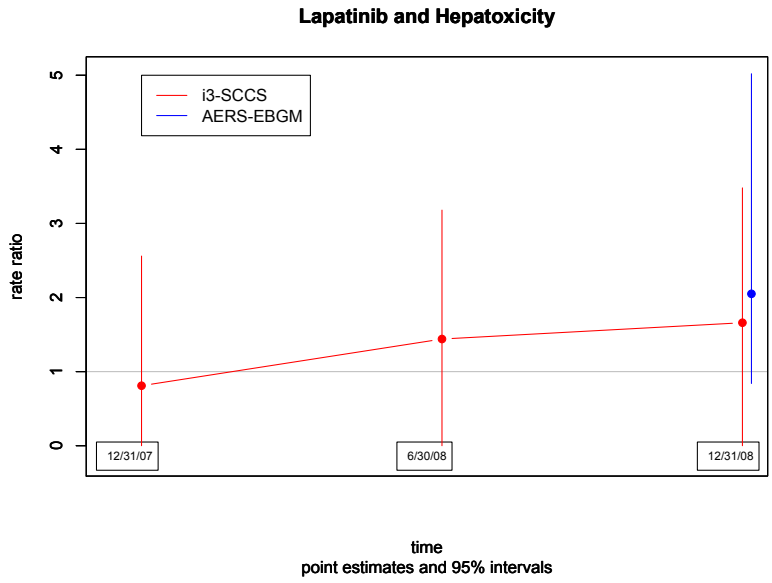


Figure 2.4: Lapatinib and hepatotoxicity



starting at the beginning of 2001. Beyond 2005, the EBGM estimate rises sharply and plateaus at $RR \approx 13$ in 2007 and 2008 (not shown). This increase in signal in AERS is likely due to the high-profile safety hearing on Vioxx in late 2004, which

Figure 2.5: Estrogen and MI

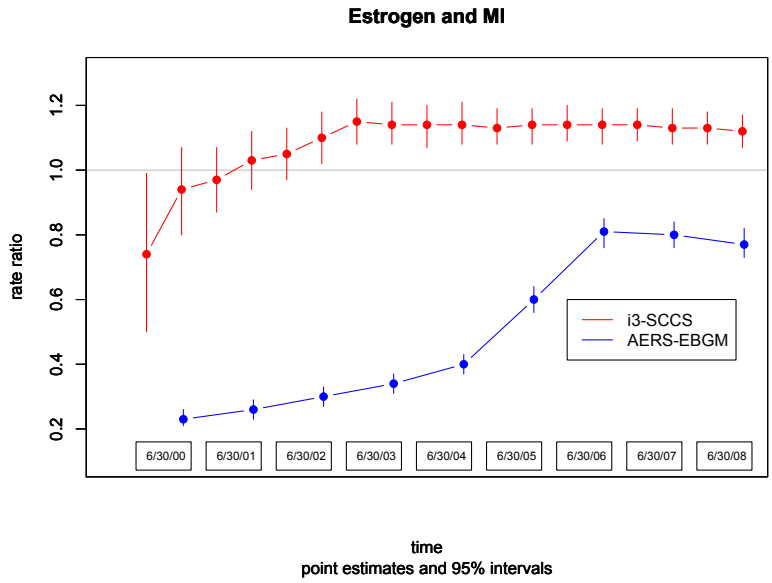
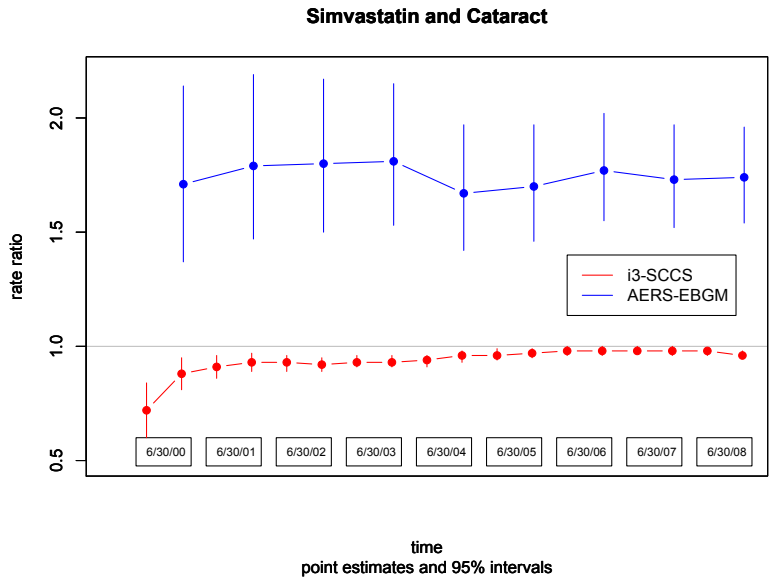


Figure 2.6: Simvastatin and cataracts



may have caused submission of a large number of associated reports to AERS. If we use the criterion EB05 $\geq$ 2, the EBGM estimate is not significant until the start of 2004. These results suggest that applying the simple SCCS analysis to claims data

may have lead to detection of the Vioxx-MI assocation 3 years prior to when the drug was pulled off the market in 2004.

For the exenatide-pancreatitis pair in Figure 2.3, MGPS shows an association at the beginning of 2008 and SCCS estimates a slightly elevated relative risk at the beginning of 2009. The label warning for exenatide was altered to include acute pancreatitis in 2007 (FDA, 2007b), so it is conceivable that public awareness led to an increase in AERS reports submissions and a larger subsequent risk estimate for MGPS.

The lapatinib-hepatotoxicity results are given in Figure 2.4. Although the point estimates were somewhat elevated for SCCS, there are too few cases for this pair to be able to determine significance. The MGPS estimate is in a similar range at the beginning of 2009, but again significance cannot be determined likely due to the small amount of data available for this pair.

In Figure 2.5 we can see that SCCS estimates a significant estrogen-MI association starting at the beginning of 2003. The relative risk from DP is steadily increasing over time, however by the end of 2008 it still shows a somewhat negative assocation.

Figure 2.6 shows results for the simvastatin-cataract pair, which was included as a false positive assocation. EBGM estimates for MGPS demonstrate an elevated risk, but the EB05 does not pass the significance threshold. SCCS relative risk estimates start with a slightly negative association, but are very close to one across time.

The 2001 detection of the Vioxx-MI association exhibits the potential of the simple SCCS-claims approach to provide substantially earlier risk identification than the standard DP-AERs analysis. For the pairs of exenatide-pancreatitis, simvastatin-cataract, and lapatinib-hepatotoxicity the simple SCCS-claims analysis and the DP-AERS method yielded similar results. SCCS-claims identifies an estrogen-MI risk that is suggested by literature (Manson et al., 2003), but is not found using the DP-AERS method and has not fully been established. In addition, neither the SCCS-claims

nor the DP-AERS analysis identified the false positive pair of simvastatin-cataract. Overall these results suggest that SCCS-claims performs similarly to or at least as well as DP-AERS, but the early detection of Vioxx-MI for SCCS-claims demonstrates its potential.

2.4 OMOP classification for drug-AE pairs

2.4.1 OMOP Background

The simple SCCS model for surveillance that we have introduced is part of a larger library of methods that are being systematically evaluated by the Observational Medical Outcomes Partnership (OMOP) project. Our work forms part of the overall OMOP effort to develop and test analytic methods for postmarketing surveillance based on LODs. OMOP's centralized network of large-scale LODs provides a rich source of real-world data for testing surveillance approaches, and acts as a prototype for the structure of an actual nation-wide active surveillance system that would be based on multiple data sources.

The OMOP project is a multi-year initiative that was established in order to evaluate and make recommendations regarding potential analysis methods for active surveillance in LODs (Stang et al., 2010). OMOP is assessing these methods on the basis of systematic empirical investigations. The project is a public-private partnership between the pharmaceutical industry, the FDA, academic institutions, and data source owners, and is administered by the Foundation for the National Institutes of Health. The objectives for OMOP are to determine the structure that is necessary for an active surveillance system and to develop and test methods that can effectively ascertain the benefit-risk profiles of marketed prescription drugs.

OMOP has access to a network of 10 large-scale LODs that cover 130 million lives in total. Data sources in this network include medical insurance claims databases,

provided by SDI Health, Humana, and Thomson Reuters, and several EHRs sourced by the Regenstrief Institute, Partners Healthcare System, GE Centricity, and VA MedSAFE. Performing analyses in these disparate databases is complicated by the fact that information is originally recorded using differing coding schemes. OMOP has greatly facilitated analysis efforts by mapping data sources to a common data model, where information on drug exposure and AE occurrence is represented in a consistent fashion across databases (Stang et al., 2010).

We use four of the OMOP databases (CCAE, MDCD, MDCR, and MSLR, whose descriptions are given in Table 2.2 below) to evaluate the performance of the simple SCCS approach when compared against other surveillance methods. Together these databases cover approximately 75 million individuals. We used these four data sources because they had full results available at the time of analysis; results are currently being generated for the remaining OMOP databases.

Table 2.2: OMOP collaborator databases for SCCS evaluation (Stang et al., 2010)

Database	Collaborator	Insurance type	Individuals (millions)
CCAE	Thomson Reuters Commercial Claims and Encounters	Multiple private insurers	58
MDCD	Thomson Multistate Medicaid	Medicaid	11.1
MDCR	Thomson Medicare Supplemental	Medicare Supplement	4.4
MSLR	MarketScan Lab Results	Multiple	1.5

We will compare the results of the SCCS method to those of two other approaches: disproportionality analysis (DP), and observational screening (OS). These methods are currently in use for surveillance in different capacities. DP is one of the most widely-used approaches for surveillance in SRSs, as discussed in Section 2.3.1. It can be adapted for surveillance in LODs such as the OMOP databases; details of adapting DP for this context are given in Zorych et al. (2011). OS is available in

commercial safety software marketed by the ProSanos Corporation, which is one of the OMOP collaborators. OS is based on screening rates, which give the number of AEs that occur during drug exposure divided by the cumulative time spent exposed to the drug. The measure of association for OS is the screening rate ratio, which is the ratio of the screening rate in a target group to the screening rate in a comparator group.

2.4.2 Health outcomes and drug exposures of interest

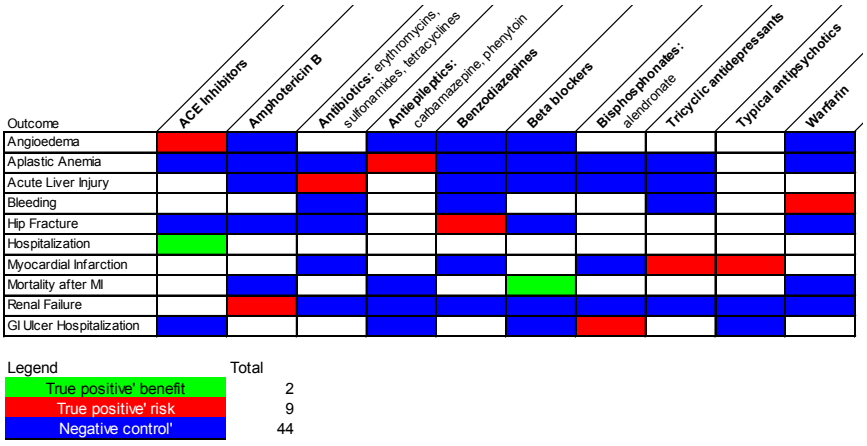
OMOP has defined 10 health outcomes of interest (HOIs), corresponding to outcome AEs, that can be used to evaluate method performance. These outcomes represent a portion of the health conditions that are of significant concern due to their medical severity, past association with drug exposures, or impact on public health. The selected HOIs are: angioedema, aplastic anemia, acute liver injury, bleeding, hospitalization due to gastrointestinal (GI) ulcer, hip fracture, hospitalization, acute MI, mortality after MI, and acute renal failure. Definitions for these health outcomes were created using a rigorous literature review process; based on the evidence tables and the investigators experience, a series of definitions for each outcome were created ranging from broad diagnosis code-based definitions to precise definitions based on a combination of coded diagnoses, coded procedures, and clinical laboratory results (OMOP, 2011). In our analyses we used the broadest definition for each health outcome based on diagnosis codes from previously published evaluation studies.

In conjunction with these HOIs, OMOP has identified 10 drug exposures of interest (DOIs). A portion of these DOIs are associated with the 10 HOIs defined above. The DOIs are: angiotensin converting enzyme (ACE) inhibitors, amphotericin B, antibiotics, antiepileptics, benzodiazepines, beta blockers, bisphosphonates, tricyclic antidepressants, typical antipsychotics, and warfarin.

Based on these HOIs and DOIs specified by OMOP, there are 10×10 possible

pairs of drugs and AEs that can be used to evaluate method performance. Of the possible drug-AE pairs, 9 have been classified as *positive associations* where there is believed to be an elevated risk of the outcome AE due to drug exposure. There are 44 remaining drug-AE pairs that serve as *negative controls*, where it has been determined that there is no relationship between the drug and the outcome. These 53 pairs were classified by consensus of an expert panel, who made their judgements on the basis of drug product labels, published literature, and previous observational studies of association (Stang et al., 2010). Of the 100 possible drug-AE combinations, the 47 remaining pairs were not classified due to mixed evidence of an association from the literature and expert review. Surveillance methods are assessed according to their ability to distinguish the 9 positive association pairs from the 44 that are negative controls. Figure 2.7 shows the 10×10 matrix of chosen drug-AE pairs with positive and negative pairs indicated.

Figure 2.7: OMOP's 10 HOIs and 10 DOIs, with 9 positive associations and 44 negative controls.



We can use these 53 pairs to gauge performance by treating this as a classfication problem on drug-AE pairs. Each method results in a particular categorization for each of the pairs as either a positive (+) association or negative (–) control. Table 2.3 shows the confusion matrix associated with the classification, which tabulates the

number of pairs that are correctly (true positive, true negative) or incorrectly (false positive, false negative) classified by a method. The performance metrics that we will use to compare different methods are based on the categories in the confusion matrix. A method's true positive rate, or *sensitivity*, refers to the number of positive associations that are correctly classified out of the total number of positives, i.e. $a/(a+c)$ in Table 2.3. The false positive rate, or *1–specificity*, is the number of negative controls that are incorrectly classified out of the total number of negatives, i.e. $b/(b+d)$.

Table 2.3: Confusion matrix

		actual class: +	actual class: –	
predicted class	+	true positive (a)	false positive (b)	
	–	false negative (c)	true negative (d)	
		9	44	53 total

For the simple SCCS method and the other comparison approaches the level of association for each drug-AE pair is determined by the value of a corresponding point estimate. Pairs are classified as positive or negative based on whether or not the point estimate for the pair surpasses a specific threshold. In the simple SCCS model, for example, the relative risk estimates for each drug-AE pair can be ranked as $e^{\hat{\beta}_{(1)}} < \cdots < e^{\hat{\beta}_{(M)}}$ with M being the total number of pairs. A threshold level c is set such that any pair with an estimate $e^{\hat{\beta}_{(m)}}$ where $m > c$ is classified as positive, and pairs that have $m \leq c$ are negatives. For the $M = 53$ drug-AE pairs in question

there are $M+1 = 54$ possible choices for the threshold c, each choice of which results in a different classification outcome.

A portion of our performance metrics are based on the reciever operating characteristic (ROC) curve, which plots sensitivity versus 1–specificity for a particular method. In our classification problem each point on the plot corresponds to a unique choice for the threshold c, and these $M + 1$ points determine the shape of the ROC curve. A good classifier should maximize the sensitivity while minimizing 1–specificity, so an ideal ROC curve would have a sharp corner close to the upper left-hand corner of the plot. A ROC curve of this type would also have a large corresponding area under curve (AUC), which is the first metric we will use in our assessment. AUC gives the probability that the method will rank a randomly chosen positive association higher than a randomly chosen negative control, so a value of AUC = 1 coincides with a perfect classification model and AUC = 0.5 is equivalent to random guessing. Methods that are better at classifying drug-AE pairs will have higher values of AUC, and it has been suggested that AUC is a better summary measure of classification performance than overall accuracy (Bradley, 1997).

Large values of 1–specificity are undesirable because they correspond to a high proportion of false positives, and it is unlikely that in practice we would choose a threshold c that yields such values. Desirable thresholds result in a low proportion of false positives, so it is really the left-hand portion of the ROC curve that reflects method performance for a realistic range of thresholds. In order to assess methods based on workable thresholds, we measure the partial area under curve (PAUC). This gives the area under the ROC curve up to a cut-off at a certain level of 1–specificity; in our case we chose a cut-off of 30%, so the maximum possible PAUC would be 0.3. For the 30% cut-off, random prediction corresponds to PAUC = 0.045.

Our next metric is the mean average precision (MAP) score. MAP is based on rank ordering the estimates for all pairs, placing cut-off points at the locations of

each of the positive associations, and averaging over the precision achieved at each of these cut-offs. For example, for the ordered estimates $e^{\hat{\beta}_{(1)}} < \cdots < e^{\hat{\beta}_{(M)}}$ in the simple SCCS model, cut-offs $c_1, \ldots, c_9$ are set to match the indices of the nine drug-AE pairs with positive associations. The precision at cut-off c is calculated as the proportion of true positives out of the total number of pairs classified as positive, i.e. $a/(a+b)$ in Table 2.3. The MAP score is then simply the average of the precision values at the thresholds $c_1, \ldots, c_9$.

There are two final performance metrics. The sensitivity at a 5% false positive rate (Sensitivity at 5%) is the sensitivity level when c is set to achieve a 5% false positive rate, i.e. it gives the height of the ROC curve at 5% 1–specificity. The average false positive rate (Average FPR) is the simple average of the false positive rates that are obtained by setting the thresholds $c_1, \ldots, c_9$ to match the indices of the positive association pairs. Unlike our other metrics, smaller values of Average FPR correspond to a lower proportion of false positives and better method performance.

2.4.3 Results

The plots in Figure 2.8 show classification results for the 53 chosen drug-AE pairs for the three comparison methods over all four OMOP databases. The numerical values used to generate Figure 2.8 are given in Table 3.1 in the subsequent Chapter 3. Each method has multiple parameter settings that correspond to various design decisions, including definition of time-at-risk, identification of AEs based on first occurrence or all occurrences of diagnosis codes, and choice of comparator group (Ryan et al., 2011). In our analyses we used the parameter settings for each method that were source optimal based on AUC, i.e. that yielded the highest AUC for the particular data source under investigation. Since the parameter settings were chosen to maximize AUC, the performance metrics in our investigation are optimistically biased. For a more accurate performance assessment one would need a larger number

of drug-AE pairs, and parameter settings would be chosen based on hold-out data or cross-validation. These directions are discussed in Chapter 5.

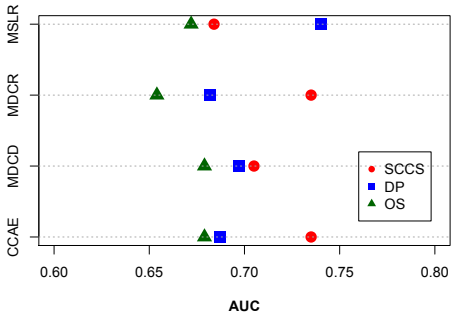
For four of the performance metrics (AUC, PAUC at 30%, Sensitivity at 5% and Average FPR), the simple SCCS model was the highest performing method for three out of the four databases. For the remaining metric, MAP score, SCCS outperforms the other methods on all four of the databases. In addition, simple SCCS is consistently best across all five metrics for both the CCAE and MDCD databases, which are the two largest LODs in our study in terms of number of lives covered. This suggests that SCCS is doing well on the pair classification problem compared with the other methods in our investigation.

AUC values for the simple SCCS model ranged from 0.7 to 0.74 on CCAE, MDCD, and MDCR, which suggests reasonable ability to discriminate between positive and negative pairs in these databases. For MSLR the AUC for SCCS was 0.68. In general SCCS had lower performance for the smaller databases (MDCR and MSLR), which are subject to a higher degree of variability. This effect may be mitigated by meta-analyses or combining information across databases (e.g. see Ryan et al., 2011), which is an avenue for future investigation. When using PAUC in place of AUC, SCCS does better on MSLR than the other methods (PAUC = 0.098), and maintains the highest scores for CCAE and MDCD (PAUC = 0.099 and 0.107, respectively). DP outperforms SCCS on MDCR, where for DP the PAUC = 0.119 and for SCCS the PAUC = 0.099. Simple SCCS is consistently highest in terms of MAP score, with values ranging from 0.39 to 0.45 across all databases. SCCS was also consistently highest for Sensitivity at 5%, however it is difficult to base an assessment on this metric since the small number of positive pairs leads to there being only a few possible values for sensitivity (true positive rate). For CCAE, MDCD, and MDCR, simple SCCS had the lowest range for Average FPR between 0.27 and 0.3, and for other methods values ranged from 0.3 to 0.35. On MSLR, DP was the best performer with

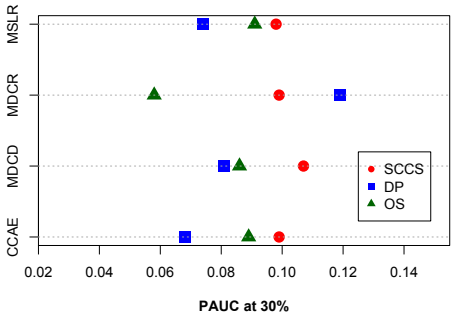
Average FPR = 0.26, whereas for SCCS the Average FPR = 0.32

The results of the drug-AE pair classification problem provide a preliminary assessment of the simple SCCS model for postmarketing surveillance. As would be the case for any evaluation procedure, the pair classification problem has some limitations that should be noted. First, evaluation results are contingent upon the correct categorization of pairs into positive associations and negative controls. Positive pairs were identified on the basis of product labels, previous observational studies, and expert consensus, however it is concievable that some of these assocations either do not exist or are not identifiable within the LODs in our analysis. Furthermore, since the drugs under investigation have been on the market for a significant amount of time, it may be the case that clinical practice has adapted in order to mitigate the occurrence of AEs known to be associated with these drugs. The precision of the performance measures is also limited by the small number of 53 drug-AE pairs available for testing, and a more accurate judgement of performance would require a larger number of test cases. OMOP is currently working to expand the number of available drug-AE pairs for testing, and assessing the simple SCCS model and other analysis approaches on this expanded test set will be part of our future work.

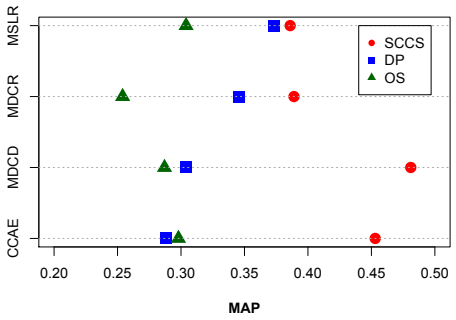
Figure 2.8: OMOP evaluation metrics for simple SCCS and comparison methods on 4 OMOP databases



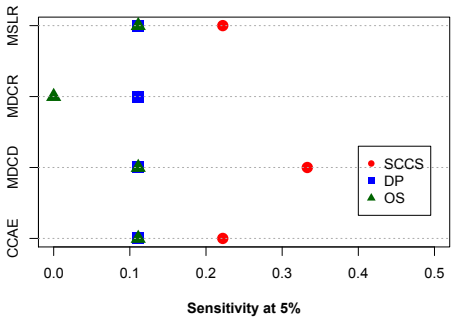
(a) AUC



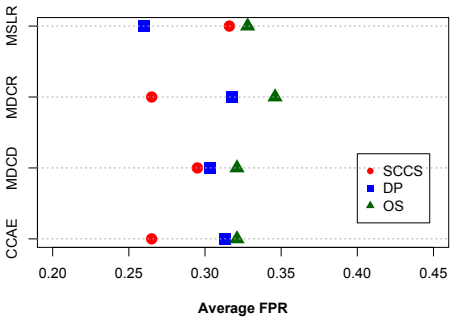
(b) PAUC at 30%



(c) MAP scores



(d) Sensitivity at 5%



(e) Average FPR

Chapter 3

Bayesian multiple SCCS for surveillance in large-scale LODs

3.1 Introduction

In the previous Chapter 2 we focused on the simple SCCS model in which there was one AE and one drug of interest. In reality, however, patients in LODs typically take multiple drugs throughout the course of their observation period. Simple SCCS only examines marginal associations between drugs and AEs, which is analogous to the perspective of DP methods that depend on 2×2 table summaries for each drug-AE combination. Basing analysis on marginal associations is problematic because it ignores the presense of other drug exposures that may be potential confounders.

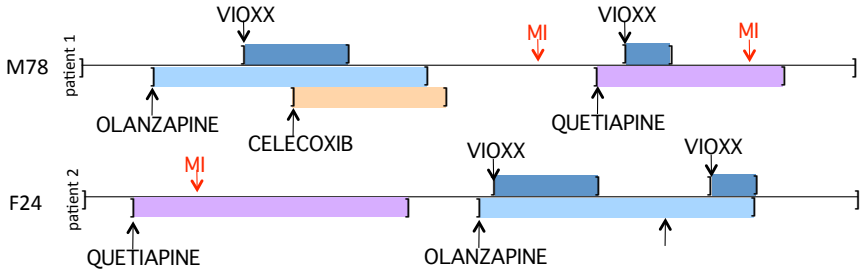
For instance, one particular kind of confounding risk has come to be known as the *innocent bystander* effect in the drug safety literature. As an example, consider two drugs, A and B, where A causes nausea and B does not. Suppose that drug A can also cause a rare infection, and that patients on drug A are often prescribed drug B to treat these infections. Since the infection is rare, drug B is not often prescribed for patients who are not on drug A. Marginally, it may appear that drug B is associated

with nausea when in fact this is not the case; here drug B is the innocent bystander. To avoid such spurious associations due to confounding, it is desirable for an analysis strategy to be able to control for the presence of other drug exposures when estimating the effect for a particular drug.

In addition, patients in LODs are often prescribed to multiple drugs concurrently, so there is a potential for interaction effects during intervals of time where different drug exposures overlap. In order to incorporate multiple drug exposures and interactions, the Poisson event rate in the SCCS model can be extended in a natural way. We refer to this extension as the *multiple* SCCS model to differentiate it from the simple SCCS analysis that was introduced in Chapter 2.

The multiple regression version of SCCS is referred to in work done by Farrington (e.g. see Farrington, 1995; Whitaker et al., 2005; Farrington and Whitaker, 2006) in the context of epidemiological and vaccine safety studies. However the problem of postmarketing surveillance presents a unique challenge due to the large number of drug exposure predictors that are involved. In order to handle this challenge of high-dimensionality, we take a Bayesian approach to regularize the regression and avoid the difficulties of overfitting that can result from using standard maximum likelihood estimation. In our approach inference is based on *maximum a posteriori* (MAP) estimation, which requires maximizing the log posterior for the drug effect parameter vector. We propose using the cyclic coordinate descent (CCD) algorithm for optimization due to its simplicity and efficiency for large problems. We present results of our Bayesian multiple SCCS analysis for the drug-AE pair classification problem on five OMOP databases, and compare its performance with that of several other existing analysis methods for surveillance.

Figure 3.1: Longitudinal traces for a 78 year-old male and a 24 year-old female who both experienced an MI and were exposed to multiple drugs (celecoxib, olanzapine, quetiapine, and Vioxx) during observation. Periods where drug exposures overlap may result in drug interaction effects.



3.2 Multiple SCCS

3.2.1 Modeling framework

To set up notation for the multiple SCCS model, suppose that there are p different drugs of interest indexed by $j = 1, \ldots, p$. On day (i, d) drug j has a corresponding exposure status indicator where $x_{idj} = 1$ if i is exposed to j on day d, and 0 otherwise. The vector of drug exposures for (i, d) can be written as $\mathbf{x}_{id} = (x_{id1}, \ldots, x_{idp})^{\mathrm{T}}$, and the vector of observed event counts over all days is $\mathbf{y}_i = (y_{id}, \ldots, y_{i\tau_i})^{\mathrm{T}}$ as before. Let e^{β_j} be the relative risk due to drug exposure j, which acts multiplicatively on the baseline rate e^{ϕ_i} during periods of time where i is exposed to j. Figure 3.1 shows example data for two patients with multiple drug exposures and MI as the outcome AE of interest.

The Poisson rate on day (i, d) takes the form

$$\lambda_{id} = e^{\phi_i + \boldsymbol{\beta}^{\mathrm{T}} \mathbf{x}_{id}} = e^{\phi_i \, + \, \beta_1 \, x_{id1} \, + \, \cdots \, + \, \beta_p \, x_{idp}},$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^{\mathrm{T}}$ is the vector of drug effect parameters. The total number of events n_i is again sufficient for ϕ_i under this multiplicative model, so the person-specific nuisance parameters will once again drop out of the likelihood upon

conditioning. One can derive the full conditional likelihood in a similar manner as described in Section 2.2, which results in

$$L^c = \prod_{i=1}^{N} p(\mathbf{y}_i \mid \mathbf{x}_{i1}, \ldots, \mathbf{x}_{i\tau_i}, n_i) \propto \prod_{i=1}^{N} \prod_{d=1}^{\tau_i} \left(\frac{e^{\boldsymbol{\beta}^{\mathrm{T}} \mathbf{x}_{id}}}{\sum_{d'} e^{\boldsymbol{\beta}^{\mathrm{T}} \mathbf{x}_{id'}}} \right)^{y_{id}}. \tag{3.1}$$

To simplify the summation in the denominator, days with common drug exposures can be grouped together within person. Each day (i, d) can be categorized into an exposure group based on the combination of drugs that i is taking on that day. Suppose that i experiences G_i distinct combinations of drug exposures, and that these exposure groups are indexed by $g = 1, \ldots, G_i$. We define G_i new exposure vectors $\mathbf{x}^*_{ig}$ corresponding to these groups, where day d falls in group g if $\mathbf{x}_{id} = \mathbf{x}^*_{ig}$. To calculate the multiple SCCS likelihood we need to know the number of events i has while in exposure group g, denoted by $y^*_{ig} = \sum_{d \in g} y_{id}$, along with the number of days d in group g, denoted by l_{ig}. Information is required only at the level of the G_i groups, rather than for each of the τ_i days, which allows for more efficient storage and computation; patients tend to take drugs over extended periods of time, so G_i is typically much smaller than τ_i. The multiple SCCS conditional likelihood for i based on grouping exposures is

$$L^c \propto \prod_{i=1}^{N} \prod_{g=1}^{G_i} \left(\frac{e^{\boldsymbol{\beta}^{\mathrm{T}} \mathbf{x}^*_{ig}}}{\sum_{g'=1}^{G_i} l_{ig'} e^{\boldsymbol{\beta}^{\mathrm{T}} \mathbf{x}^*_{ig'}}} \right)^{y^*_{ig}}. \tag{3.2}$$

We can also build drug interactions and other relevant time-varying covariates (e.g. age groups, measures of health taken over time, etc.) into the multiple SCCS model in addition to drug main effects. For instance, suppose that γ_{rs} is the two-way interaction effect between drugs r and s, and that $\mathbf{z}_{id}$ is a vector of additional time-varying covariates for (i, d) that are believed to effect AE occurrence. Then the Poisson event rate on (i, d) can be written as

$$\lambda_{id} = e^{\phi_i + \boldsymbol{\beta}^{\mathrm{T}} \mathbf{x}_{id} + \sum_{r \neq s} \gamma_{rs} x_{idr} x_{ids} + \boldsymbol{\alpha}^{\mathrm{T}} \mathbf{z}_{id}}. \tag{3.3}$$

The event rate in (3.3) can be inserted into the conditional likelihood in (3.1) in a straightforward manner, and the individual ϕ_i parameters drop out of the expression as before. Under this expanded model, the grouping of days would be based on unique combinations of drug exposures, interactions, and values of the $\mathbf{z}_{id}$ covariates, and the likelihood can be written in a grouped form similar to (3.2).

3.2.2 Inference and maximum likelihood estimation

Inference in the multiple SCCS model can be based on maximum likelihood estimation. The SCCS log-likelihood is concave (Appendix A) so optimization can be done in a variety of ways, e.g. by taking a multidimensional Newton-Raphson approach. Standard asymptotic theory for maximum likelihood is applicable for multiple SCCS, where the estimate $\hat{\boldsymbol{\beta}}$ is asymptotically distributed as $(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \sim \mathbf{N}(\mathbf{0}, \mathbf{I}^{-1}(\boldsymbol{\beta}))$ and $\mathbf{I}(\boldsymbol{\beta})$ is the information matrix given in (3.6). We can estimate the covariance matrix for $\hat{\boldsymbol{\beta}}$ by taking $\hat{\text{cov}}(\hat{\boldsymbol{\beta}}) = \mathbf{I}^{-1}(\hat{\boldsymbol{\beta}})$. Asymptotic Normality allows us to calculate $(1-\alpha)\%$ confidence intervals for the jth drug effect parameter β_j by using $\hat{\beta}_j \pm z^{\alpha/2}\sqrt{\hat{\text{var}}(\hat{\beta}_j)}$, where $z^{\alpha/2}$ is the $(1-\alpha/2)$ quantile of the standard Normal distribution and $\hat{\text{var}}(\hat{\beta}_j)$ is taken to be the jth diagonal element of the inverse information matrix $\mathbf{I}^{-1}(\hat{\boldsymbol{\beta}})$. We can use the delta method to obtain confidence intervals for the corresponding relative risk e^{β_j} due to exposure j, where the $(1-\alpha)\%$ confidence interval takes the form $e^{\hat{\beta}_j} \pm z_{\alpha/2} \times e^{\hat{\beta}_j}\sqrt{\hat{\text{var}}(\hat{\beta}_j)}$.

The log-likelihood for multiple SCCS that coincides with (3.1) is

$$l^c = \sum_{i=1}^{N}\left[\boldsymbol{\beta}^{\mathrm{T}}\sum_{d=1}^{\tau_i} y_{id}\mathbf{x}_{id} - n_i \log\Big(\sum_{d'=1}^{\tau_i} e^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_{id'}}\Big)\right]. \tag{3.4}$$

The corresponding score (gradient) vector of this log-likelihood is

$$\mathbf{U}(\boldsymbol{\beta}) = \frac{\partial l^c}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{N}\sum_{d=1}^{\tau_i}\mathbf{x}_{id}\left[y_{id} - n_i\left(\frac{e^{\mathbf{x}_{id}^{\mathrm{T}}\boldsymbol{\beta}}}{\sum_{d'=1}^{\tau_i} e^{\mathbf{x}_{id'}^{\mathrm{T}}\boldsymbol{\beta}}}\right)\right]. \tag{3.5}$$

The expected information matrix, i.e. the expected value of the negative of the Hessian matrix, takes the form

$$\mathbf{I}(\boldsymbol{\beta}) = E\left[-\frac{\partial^2 l^c}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\mathrm{T}}}\right] = \sum_{i=1}^{N} n_i \left[\frac{\sum_{d=1}^{\tau_i} e^{\mathbf{x}_{id}^{\mathrm{T}}\boldsymbol{\beta}} \left(\mathbf{x}_{id}^{\otimes 2}\right)}{\sum_{d'=1}^{\tau_i} e^{\mathbf{x}_{id'}^{\mathrm{T}}\boldsymbol{\beta}}} - \left(\frac{\sum_{d=1}^{\tau_i} e^{\mathbf{x}_{id}^{\mathrm{T}}\boldsymbol{\beta}} \mathbf{x}_{id}}{\sum_{d'=1}^{\tau_i} e^{\mathbf{x}_{id'}^{\mathrm{T}}\boldsymbol{\beta}}} \right)^{\otimes 2} \right], \quad (3.6)$$

where $\mathbf{x}^{\otimes 2} = \mathbf{x}\mathbf{x}^{\mathrm{T}}$ denotes the outer product. In this case the observed and expected information matrices are equal since the Hessian depends only on n_i and $\mathbf{x}_{id}$, which are fixed through conditioning, and is independent of the random y_{id} outcomes.

We noted in Section 2.2 that the analysis is dependent only on data from cases. From the expressions for the score vector $\mathbf{U}(\boldsymbol{\beta})$ in (3.5) and information matrix $\mathbf{I}(\boldsymbol{\beta})$ in (3.6), we can see that *unexposed cases* (i.e. individuals i where both $n_i \geq 1$ and $\mathbf{x}_{id} \equiv \mathbf{0}$ for all d) will not contribute to the estimation of $\hat{\boldsymbol{\beta}}$ or $\hat{\mathrm{var}}(\hat{\boldsymbol{\beta}})$ since they add a zero element to the summations in $\mathbf{U}(\boldsymbol{\beta})$ and $\mathbf{I}(\boldsymbol{\beta})$. Thus only exposed cases need to be included in the analysis, which leads to a substantial reduction in the amount of data required. Of the already modest number of cases, we are able to exclude those who were not exposed to at least one of the drugs of interest during the course of their observation.

3.3 Bayesian multiple SCCS approach

Inference and analysis results are typically based on maximum likelihood estimates of the drug parameter vector $\boldsymbol{\beta}$, as discussed in Section 3.2.2. However, in the setting of postmarketing surveillance there are millions of potential exposure predictors made up of tens of thousands of drug exposure main effects along with their interactions. We are thus dealing with a high-dimensional regression problem, where the maximum likelihood approach can result in overfitting, large estimated coefficient variances, numerical instability, and spurious effect estimates.

In order to avoid the difficulties related to high-dimensionality, we take a Bayesian approach to regularization by imposing a prior distribution on the drug exposure parameter $\boldsymbol{\beta}$. Priors are specified such that $\boldsymbol{\beta}$ is most likely to be near zero and large parameter values will be more heavily penalized, which has the effect of shrinking the posterior distribution toward zero. We perform inference based on MAP estimates, which are equivalent to constrained maximum likelihood estimates in the regularized regression framework where a restriction is placed on the norm of the parameter vector. Well-known examples of regularized regression are *ridge regression* (Hoerl and Kennard, 1970) and the *lasso* (Tibshirani, 1996). Efficient algorithms to find MAP estimates are available, which makes the optimization tractable even in the high-dimensional setting. There are many possible choices for priors on $\boldsymbol{\beta}$, however we focus on the Normal and Laplacian priors for simplicity.

For the first approach to Bayesian multiple SCCS we place an independent univariate Normal prior with mean zero, variance σ_β^2 on each of the parameter components β_j, which has the effect of pulling the estimates toward zero. A smaller choice of σ_β^2 will lead to a higher degree of shrinkage. The prior is specified as

$$\beta_j \mid \sigma_\beta^2 \sim \mathrm{N}(0, \sigma_\beta^2) \quad \text{for} \quad j = 1, \ldots, p.$$

This choice of prior is equivalent to placing a ridge regression penalty on the likelihood, which constrains the L_2-norm of the parameter vector $\boldsymbol{\beta}$. From the expression for the corresponding posterior distribution in (3.7), we can see that for this prior MAP estimates are the same as maximum likelihood estimates subject to a constraint on $\sum_{j=1}^{p} \beta_j^2$.

As a second approach, we use an independent univariate Laplace distribution with mean zero, variance $2/\lambda^2$ as the prior on each of the β_j parameter components. This Laplacian prior has the added advantage of inducing sparsity in the model: a portion of the components of the MAP estimate of $\boldsymbol{\beta}$ will shrink all the way to zero and, as a consequence, the drug exposure predictors associated with those components will not

enter into the model. Larger values of λ lead to a greater degree of shrinkage and a higher number of variables being selected out of the model. The prior is specified as

$$\beta_j \mid \lambda \sim \text{Laplace}(0, 2/\lambda^2) \quad \text{for} \quad j = 1, \ldots, p.$$

From the corresponding posterior distribution given in (3.8) we can see that MAP estimates will be equivalent to maximum likelihood estimates with a lasso penalty. Here the L_1-norm of the parameter vector $\sum_{j=1}^{p} |\beta_j|$ is constrained in order to prevent parameter estimates from getting too large.

Under the Normal and Laplace priors, respectively, expressions for the full log posterior distributions for $\boldsymbol{\beta}$ in our Bayesian multiple SCCS model are

$$p(\boldsymbol{\beta} \mid \mathbf{y}_i, \mathbf{x}_{i1}, \ldots, \mathbf{x}_{i\tau_i}, n_i) \propto \sum_{i=1}^{N} \left[\sum_{d=1}^{\tau_i} y_{id} \, \boldsymbol{\beta}^{\mathrm{T}} \mathbf{x}_{id} - n_i \log \Big(\sum_{d'=1}^{\tau_i} e^{\boldsymbol{\beta}^{\mathrm{T}} \mathbf{x}_{id'}} \Big) \right] - \frac{N}{2} \log \sigma_\beta^2 - \frac{1}{2\sigma_\beta^2} \sum_{j=1}^{p} \beta_j^2 \tag{3.7}$$

and

$$p(\boldsymbol{\beta} \mid \mathbf{y}_i, \mathbf{x}_{i1}, \ldots, \mathbf{x}_{i\tau_i}, n_i) \propto \sum_{i=1}^{N} \left[\sum_{d=1}^{\tau_i} y_{id} \, \boldsymbol{\beta}^{\mathrm{T}} \mathbf{x}_{id} - n_i \log \Big(\sum_{d'=1}^{\tau_i} e^{\boldsymbol{\beta}^{\mathrm{T}} \mathbf{x}_{id'}} \Big) \right] - N \log \lambda - \lambda \sum_{j=1}^{p} |\beta_j|. \tag{3.8}$$

Estimating $\boldsymbol{\beta}$ via MAP requires maximizing (3.7) or (3.8), depending on the prior that is imposed. Our strategy for optimization is discussed in the subsequent Section 3.4. For a fully Bayesian analysis, ideally one would use Markov chain Monte Carlo (MCMC) or other sampling techniques to obtain the full posterior distribution for the $\boldsymbol{\beta}$ parameter. However, in our application the large number of drug exposure predictors renders such approaches intractable. To estimate the variability and other second order features of the posterior for $\boldsymbol{\beta}$, it may be possible to make use of methods that rely on approximations and are less computationally costly than MCMC. These possibilities are discussed in the final Chapter 5 as a direction for future work.

3.4 Optimization and implementation

Finding the MAP estimate for the $\boldsymbol{\beta}$ parameter depends on maximizing the SCCS log posteriors given in (3.7) and (3.8). If we frame this optimization as a minimization problem, the objective function for Bayesian multiple SCCS is the negative of the log posterior. In Appendix A we show that the multiple SCCS likelihood is log concave in $\boldsymbol{\beta}$. Since the density functions for both the Normal and Laplacian priors on $\boldsymbol{\beta}$ are also log concave, it follows that the negated log posterior is a convex function. The objective function we are using for MAP estimation is thus amenable to algorithms that are specialized for convex problems.

There are many possible strategies for convex optimization, however we choose to use the method of *cyclic coordinate descent* (CCD). The advantages of CCD are that it is straightforward to implement and it has been demonstrated to have high speed for large problems (Wu and Lange, 2008). Convergence properties of CCD are discussed in Tseng (2001) and Saha and Tewari (2010). The CCD algorithm is comparable to approaches used in state-of-the-art software packages for regularized regression, namely the open-source Bayesian binary regression (BBR) software of Genkin et al. (2007) and the `glmnet` R package of Friedman et al. (2010). Our development follows that of Genkin et al. (2007), which is based on the CLG algorithm of Zhang and Oles (2001).

CCD works by cycling through each component of the parameter vector and doing a one-dimensional minimization update for the chosen component while holding the values of the other components constant. The process continues for each parameter component in turn, and the algorithm makes multiple passes over the full vector until a suitable convergence criterion is met. This procedure enjoys substantially reduced computation time when compared with the multidimensional Newton-Raphson approach or similar methods; CCD relies on one-dimensional updates, which circumvents the need for processing and storage of a high-dimensional Hessian matrix.

The details of the algorithm are as follows. Let $k = 0, 1, \ldots$ index the number of times that CCD has cycled through all $j = 1, \ldots, p$ components of the drug parameter vector $\boldsymbol{\beta}$. First, the value $\boldsymbol{\beta}^{(0)}$ is initialized. For the jth component, $g(z)$ is the negated log posterior function that varies only along the dimension of β_j while the values of the other parameter components are held constant. On the kth iteration of the algorithm, we would like $\beta_j^{(k)}$ to be updated such that the new value $\beta_j^{(k+1)}$ minimizes the objective function with respect to the jth component, i.e.

$$\beta_j^{(k+1)} = \arg\min_z g(z).$$

This one-dimensional minimization is done via Newton's method, which depends on the updating equation $\beta_j^{(new)} = \beta_j - \frac{g'(\beta_j)}{g''(\beta_j)}$. Expressions for $g'(\beta_j)$ and $g''(\beta_j)$, the first and second derivatives of the objective function $g(z)$ evaluated at β_j, are given in (3.9–3.12). As described in Genkin et al. (2007), it is sufficient to take a single minimization update step for the current component on each cycle of the algorithm rather than doing the full minimization. Component values are updated on each cycle of CCD, so there is little benefit to spending time finding the exact minimizer for one component on a single cycle.

CCD uses a trust region approach in which the allowable update step size for each component is thresholded in order to improve convergence. This type of approach constrains jump sizes so that parameter updates stay within the region where the quadratic function used in the Newton-Raphson step is a reasonable approximation to the objective (Dennis and Schnabel, 1996). On iteration k, we calculate $\beta_j^{(k+1)}$ by adding $\Delta\beta_j^{(k)} = -\frac{g'(\beta_j^{(k)})}{g''(\beta_j^{(k)})}$ to $\beta_j^{(k)}$. The magnitude of $\Delta\beta_j^{(k)}$ is thresholded by $\Delta_j^{(k)}$, which is the upper limit for the jump size for β_j on the kth iteration. This limit is itself updated so that $\Delta_j^{(k+1)} = \max(2|\Delta\beta_j^{(k)}|, \Delta_j^{(k)}/2)$ for iteration $k + 1$.

Expressions for the first and second derivatives of the component-wise objective function $g(z)$ are given in (3.9–3.12). We let $r_{id} = \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_{id}$ for person i on day d. This inner product depends on the current value of the parameter vector and is thus

constant with respect to z. The first and second derivatives of the negative of the log posterior in (3.7) in the case of the Normal prior are

$$g'(\beta_j) = \frac{\partial g(z)}{\partial z}\bigg|_{z=\beta_j} = \sum_{i=1}^{N}\sum_{d=1}^{\tau_i} y_{id}x_{idj} - \sum_{i=1}^{N} n_i\left(\frac{\sum_{q=1}^{\tau_i} x_{iqj}e^{r_{iq}}}{\sum_{s=1}^{\tau_i} e^{r_{is}}}\right) - \frac{\beta_j}{\sigma_\beta^2} \tag{3.9}$$

and

$$g''(\beta_j) = \frac{\partial^2 g(z)}{\partial^2 z}\bigg|_{z=\beta_j} = -\sum_{i=1}^{N} n_i\left[1 - \frac{(\sum_{q=1}^{\tau_i} e^{r_{iq}})(\sum_{v=1}^{\tau_i} x_{ivj}^2 e^{r_{iv}})}{(\sum_{s=1}^{\tau_i} x_{isj}e^{r_{is}})^2}\right] - \frac{1}{\sigma_\beta^2}. \tag{3.10}$$

Analogous expressions for the Laplace prior based on the log posterior in (3.8) are

$$g'(\beta_j) = \frac{\partial g(z)}{\partial z}\bigg|_{z=\beta_j} = \sum_{i=1}^{N}\sum_{d=1}^{\tau_i} y_{id}x_{idj} = -\sum_{i=1}^{N} n_i\left[\frac{\sum_{q=1}^{\tau_i} x_{iqj}e^{r_{iq}}}{\sum_{s=1}^{\tau_i} e^{r_{is}}}\right] - \lambda\ \mathrm{sgn}(\beta_j) \tag{3.11}$$

and

$$g''(\beta_j) = \frac{\partial^2 g(z)}{\partial^2 z}\bigg|_{z=\beta_j} = -\sum_{i=1}^{N} n_i\left[1 - \frac{(\sum_{q=1}^{\tau_i} e^{r_{iq}})(\sum_{v=1}^{\tau_i} x_{ivj}^2 e^{r_{iv}})}{(\sum_{s=1}^{\tau_i} x_{isj}e^{r_{is}})^2}\right] \quad \text{for } \beta_j \neq 0, \tag{3.12}$$

where $\mathrm{sgn}(x) = +1$ if $x > 0$, and -1 if $x < 0$.

When $\beta_j^{(k)}$ is updated on the kth iteration, the inner products r_{id} are updated via the equation $r_{id}^{(k+1)} = r_{id}^{(k)} + x_{idj}(\beta_j^{(k+1)} - \beta_j^{(k)})$. In postmarketing surveillance the exposure vectors $\mathbf{x}_{id}$ are sparse in that $x_{idj} = 0$ for the majority of vector components j. This sparsity is a result of individuals typically being exposed only to a small subset of all possible drugs, and for that subset the exposures may not occur on the majority of observation days. For efficient computation, the CCD implementation keeps a current version of the inner product r_{id} stored. This saves processing time and takes advantage of the sparsity in the exposure vectors since if $x_{idj} = 0$ then r_{id} does not need to be updated.

Under the Laplace prior, the values of $g'(\beta_j)$ and $g''(\beta_j)$ are not defined at $\beta_j = 0$ since the L_1-norm constraint is not differentiable at that point. We follow the approach of Genkin et al. (2007), which modifies the updates in the CCD algorithm to

handle this special case. With this modification, $\beta_j^{(k+1)}$ is set to zero on iteration $k+1$ if the calculated jump $\Delta\beta_j^{(k)}$ would change the current sign of the parameter component, i.e. if $\text{sgn}(\beta_j^{(k+1)}) \neq \text{sgn}(\beta_j^{(k)})$. In the case that the parameter component $\beta_j^{(k)} = 0$, we attempt to update β_j in both directions, $\text{sgn}(\beta_j) = +1$ and $\text{sgn}(\beta_j) = -1$, and proceed with the update if the calculated $\Delta\beta_j^{(k)}$ would result in an update in the same direction, i.e. $\Delta\beta_j^{(k)} > 0$ for $\text{sgn}(\beta_j^{(k+1)}) = +1$ and $\Delta\beta_j^{(k)} < 0$ for $\text{sgn}(\beta_j^{(k+1)}) = -1$. The value is only updated if this criterion is passed; otherwise the component is kept at zero. Since the objective $g(z)$ is convex, either the -1 or the $+1$ direction will result in a successful update, but not both. Parameters are selected out of the model when components become locked to zero prior to the convergence of the algorithm.

3.5 OMOP data analysis

3.5.1 Background

In order to evaluate our Bayesian multiple SCCS approach, we would like our investigation to target two primary objectives: (1) determine whether Bayesian multiple SCCS outperforms the simple SCCS analysis presented in Chapter 2, which only considers marginal associations, and (2) compare the performance of the Bayesian multiple and simple SCCS approaches to standard alternative surveillance methods. We address these questions through empirical evaluation. Our empirical results are based on the drug-AE pair classification problem described in Section 2.4.2 for the health outcomes and drug exposure outcomes of interest defined by OMOP.

We will compare the results of applying Bayesian multiple SCCS for pair classification to those of the simple SCCS method and three other approaches: disproportionality analysis (DP), observational screening (OS), and Bayesian logistic regression (BLR). Open source software for these methods is available on the OMOP website, `http://omop.fnih.org`. The DP and OS methods were referred to in Section 2.4.1.

We chose to include BLR as an additional method in our comparison since it is a regularized regression model that is similar in spirit to our Bayesian multiple SCCS approach, and it is also well-suited to high-dimensional problems. BLR adjusts for the presence of drug exposures and other potential confounders in a similar manner as the multiple SCCS model. Multiple drug exposures can be included as predictors in BLR, unlike in DP or simple SCCS where we are limited to marginal associations.

The BLR model is as follows. Let y_i represent the presence or absence of the AE of interest within the observation period for person i, i.e. $y_i = 1$ if the outcome AE occurs during observation and 0 otherwise. There are p drugs of interest, and $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^{\mathrm{T}}$ is the corresponding vector of drug exposure indicators. For inviduals with $y_i = 1$, the exposure indicator is $x_{ij} = 1$ if i is exposed to drug j during a window of time prior to the first occurrence of the condition, and 0 otherwise. The length of this window of time is a parameter to be specified by the user. In our BLR implementation for the HOI/DOI pair classification problem, the control group (where $y_i = 0$) is made up of a randomly chosen subset of individuals who did not have the AE of interest, but did have an occurrence of at least one of the other HOIs. For the group where $y_i = 0$, one of the other HOI occurrences for i is randomly selected. The drug exposure indicator is $x_{ij} = 1$ if i is exposed to drug j during a window of time prior to the selected HOI occurrence, and is 0 otherwise. The values of other pertinent fixed covariates such as age or sex could also be incorporated into the $\mathbf{x}_i$ vector. The outcomes y_i are modeled via a logistic regression on $\mathbf{x}_i$, where it is assumed that

$$\log\left[\frac{p(y_i = 1 \mid \boldsymbol{\beta}, \mathbf{x}_i)}{p(y_i = 0 \mid \boldsymbol{\beta}, \mathbf{x}_i)}\right] = \beta_1 x_{i1} + \cdots + \beta_p x_{ip} = \boldsymbol{\beta}^{\mathrm{T}} \mathbf{x}_i, \tag{3.13}$$

and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^{\mathrm{T}}$ is the parameter vector of log odds ratios associated with the p drug exposures. In BLR there is a prior imposed on the $\boldsymbol{\beta}$ drug effect parameter, which can take the form of an independent univariate Normal or Laplace prior on each of the β_j parameter components. This shrinks MAP estimates and reduces overfitting

in a similar manner as our Bayesian multiple SCCS approach, and helps mitigate the number of false positive associations.

3.5.2 Evaluation results

In our evaluation we examine the same set of five performance metrics that we used previously for the drug-AE pair classification problem: AUC, PAUC at 30%, MAP score, sensitivity at 5%, and average FPR. We applied our comparison methods for this problem to the four OMOP databases of CCAE, MCDC, MDCR, and MSLR. A detailed description of the performance metrics and OMOP data sources is given in Section 2.4.1.

Figure 3.2 shows results for the five analysis methods implemented on the four OMOP databases. In the plot, SCCS refers to the simple SCCS method of Chapter 2 and BSCCS refers to the Bayesian multiple SCCS approach. For each analysis method we chose parameter settings that were source optimal based on AUC, as described in Section 2.4.3. Table 3.1 gives the full list of numerical values that were used to generate Figure 3.2, with the best performers highlighted for each combination of evaluation metric and database.

For both AUC and Average FPR, the BSCCS approach is consistently the best performer out of the five methods across all four databases. The AUC values for BSCCS range from 0.75 to 0.82, suggesting that the method does relatively well in discriminating between a randomly chosen positive and negative pair. SCCS had the second highest AUC values for CCAE and MDCR (0.74 for both), with BLR being second highest for MDCD (AUC = 0.8) and DP being second for MSLR (AUC = 0.74). For PAUC at 30%, BSCCS has the highest values on CCAE and MDCR (PAUC = 0.12 for both), and BLR has the highest for MDCD and MSLR (PAUC = 0.17 and 0.12, respectively.

BSCCS had the best MAP score performance for the three largest databases

(CCAE, MDCD, and MDCR). BLR had the highest MAP for MSLR, with BSCCS being second. BLR has the highest Sensitivity at 5% values for CCAE and MSLR, and BSCCS was highest for MDCD and MDCR. As noted in Section 2.4.3, this metric is difficult to use for comparison here since there is a small number of positive pairs. In general BSCCS outperformed SCCS across all metrics and data sources, the only exceptions being PAUC at 5% and Specificity at 5% for MSLR. This suggests that adjusting for multiple drug exposures and regularizing the regression leads to improved performance in terms of drug-AE pair classification.

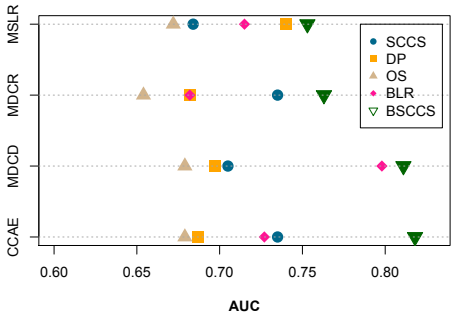
The ROC curves for the five methods across all four databases are shown in Figure 3.3. The larger points on the curves correspond to the locations of the 9 positive associations. The sensitivity value rises at each of these locations since passing that threshold increases the number of true positives detected by one. Based on the ROC curves, we can see differences in method performance in more detail. For PAUC at 30%, for example, we can see that BLR is outperforming BSCCS on MDCD and MSLR. The ROC curves demonstrate that BLR is able to detect a higher number of true associations at lower levels of 1–specificity than BSCCS on these databases (i.e. the ROC curves for BLR rise more steeply for values of 1–specificity near zero). Despite this, BSCCS tends to place true positives at low levels of 1–specificity overall, resulting in low Average FPR compared to other methods. We can also see that in general OS has lower AUC values, since its ROC curves lie closer to the diagonal (AUC = 0.5, which corresponds to random guessing) than the other methods.

Figure 3.4 shows BSCCS relative risk estimates for the 53 drug-AE pairs plotted against those for simple SCCS on the four OMOP databases. The enlarged red squares correspond to positive association pairs and the smaller black squares are the negative control pairs. The dashed diagonal line is where BSCCS and SCCS estimates are equal. We can see that estimates tend to lie below the diagonal, so the shrinkage in BSCCS typically leads to relative risk estimates that are smaller than

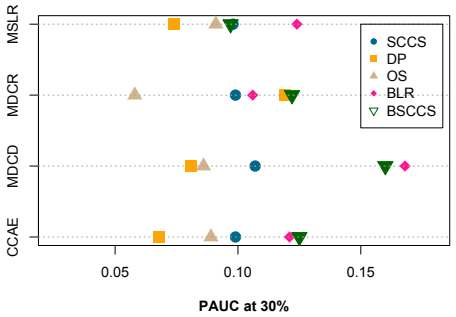
those from SCCS. For MDCR and MSLR, there is one positive assocation pair where BSCCS estimates are higher than those from SCCS; this difference may be due to the adjustment for confounding drug exposures in the BSCCS model. In CCAE, MDCD, and MDCR a noticeable portion of the negative control pairs tend to get a higher degree of shrinkage from BSCCS than the positive associations since they lie farther below the diagonal line. For MSLR it appears that, with the single positive pair as an exception, both negative and positive pairs lie close to the diagonal. The lack of shrinkage for the negative pairs for BSCCS compared to SCCS on MSLR may be one of the reasons that BSCCS generally does not do as well on this database as it does for the other three.

These evaluation results are subject to the same limitations that were laid out in Section 2.4.3, however they provide a means of preliminary assessment for our BSCCS model. These results demonstrate that BSCCS often is the highest performer compared to the other methods on our 4 OMOP databases. In particular, BSCCS has better performance than the simple SCCS model across the majority of metrics and databases, which suggests that adjusting for confounding drug exposures and regularizing the regression leads to improved ability to discriminate between positive and negative drug-AE pairs in our experiment. BSCCS also tends to score higher than BLR, which indicates that having the recurrent nature of outcome AEs incorporated into the model may also improve method performance.

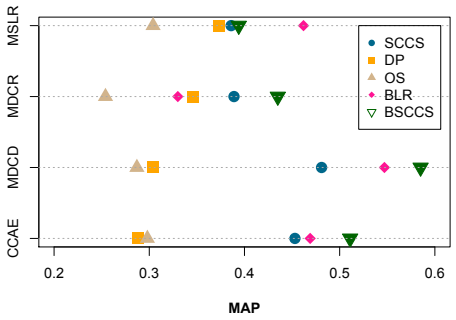
Figure 3.2: OMOP evaluation metrics for Bayesian multiple SCCS and comparison methods on 4 OMOP databases



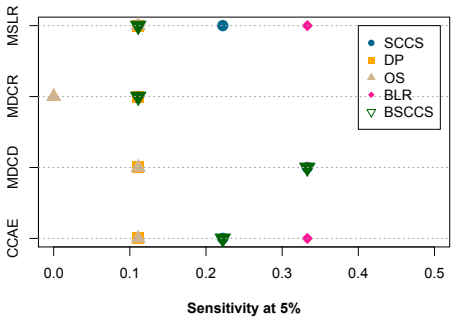
(a) AUC



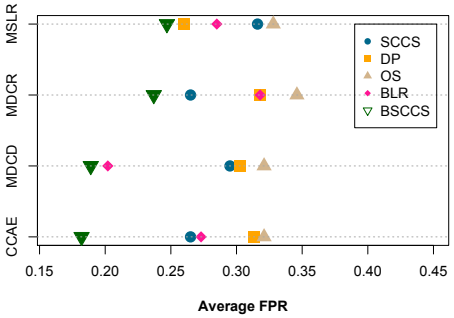
(b) PAUC at 30%



(c) MAP scores



(d) Sensitivity at 5%



(e) Average FPR

Figure 3.3: ROC curves

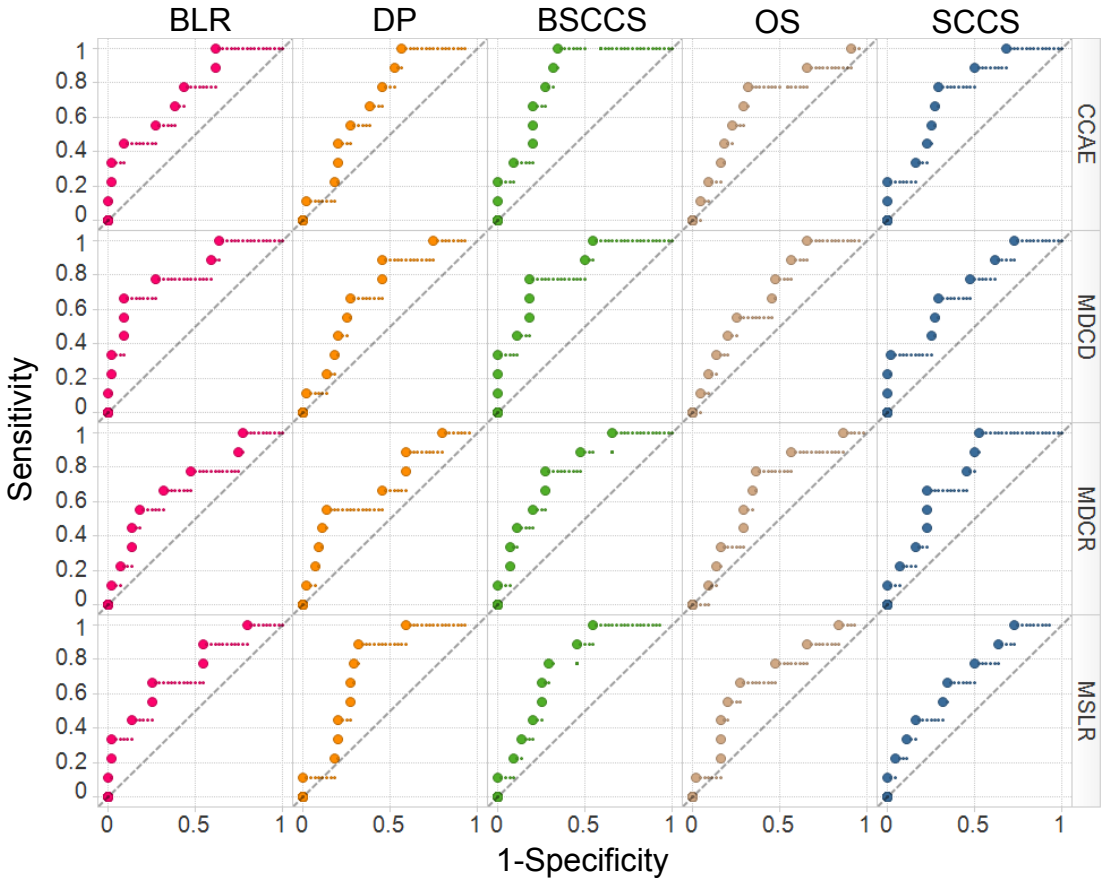


Figure 3.4: BSCCS vs. SCCS pair estimates for four OMOP databases

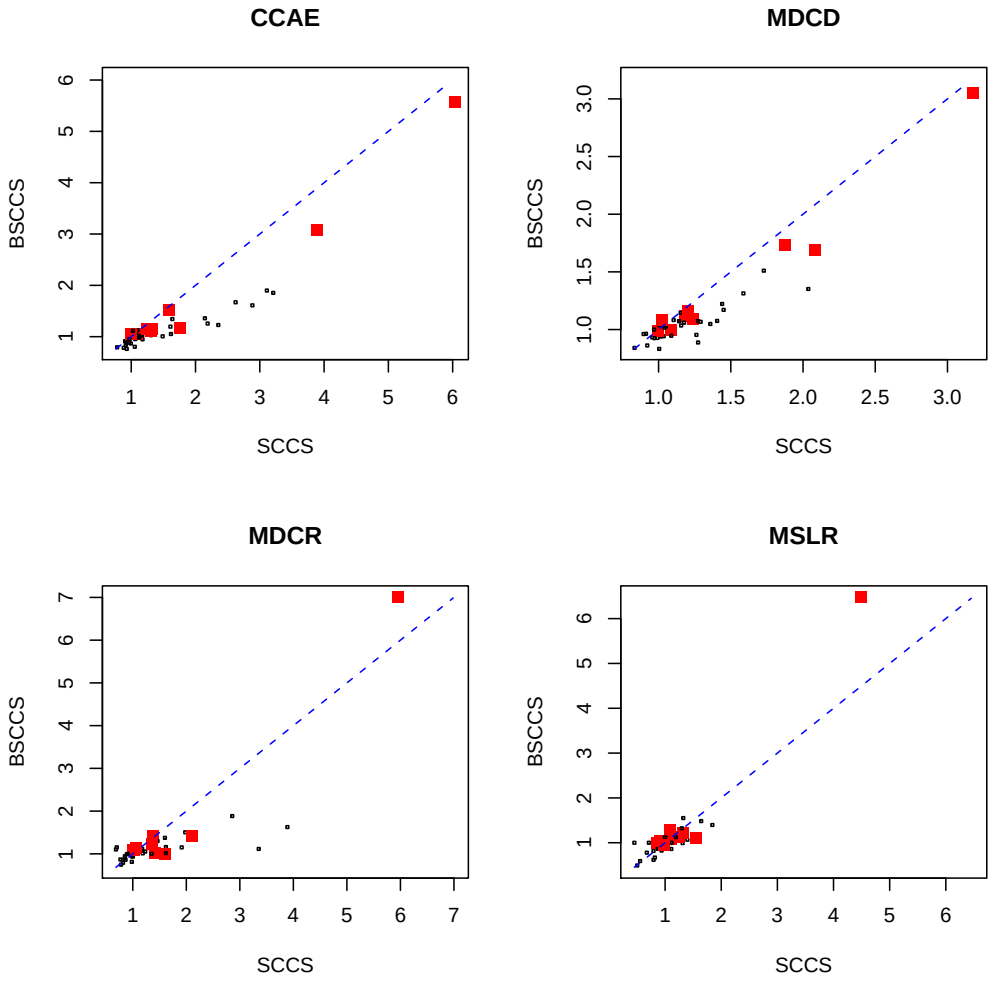


Table 3.1: Results of method comparison on OMOP databases

Database	Method	AUC	PAUC at 30%	MAP	Specificity at 5%	Average FPR
CCAE	DP	0.687	0.068	0.288	0.111	0.313
	OS	0.679	0.089	0.298	0.111	0.321
	BLR	0.727	0.121	0.469	**0.333**	0.273
	SCCS	0.735	0.099	0.453	0.222	0.265
	BSCCS	**0.818**	**0.125**	**0.511**	0.222	**0.182**
MDCD	DP	0.697	0.081	0.304	0.111	0.303
	OS	0.679	0.086	0.287	0.111	0.321
	BLR	0.798	**0.168**	0.547	**0.333**	0.202
	SCCS	0.705	0.107	0.481	**0.333**	0.295
	BSCCS	**0.811**	0.160	**0.585**	**0.333**	**0.189**
MDCR	DP	0.682	0.119	0.346	**0.111**	0.318
	OS	0.654	0.058	0.254	0.000	0.346
	BLR	0.682	0.106	0.330	**0.111**	0.318
	SCCS	0.735	0.099	0.389	**0.111**	0.265
	BSCCS	**0.763**	**0.122**	**0.435**	**0.111**	**0.237**
MSLR	DP	0.740	0.074	0.373	0.111	0.260
	OS	0.672	0.091	0.304	0.111	0.328
	BLR	0.715	**0.124**	**0.462**	**0.333**	0.285
	SCCS	0.684	0.098	0.386	0.222	0.316
	BSCCS	**0.753**	0.097	0.394	0.111	**0.247**

Chapter 4

A positive event dependence model

4.1 Introduction

Events in the SCCS model arise from a non-homogeneous Poisson process, which relies on implicit assumptions regarding conditional independence. The model assumes that, within individuals, event occurrences are conditionally independent given exposures. This is conditional independence assumption (i) discussed in Section 2.2. This requirement is clearly violated when the occurrence of an event can alter the risk of future events. In a clinical setting, for example, once a patient has a first MI they may be at higher subsequent risk for a second MI. In the application of postmarketing surveillance many AEs exhibit this type of behavior, so we would like be able to analyze the association between drug exposures and AEs without requiring conditional independence.

General intensity-based models for recurrent events can incorporate functions of the event history, such as cumulative event count or time elapsed since the most recent event, into the process intensity in order to represent dependencies between event occurrences (Cook and Lawless, 2007). Although these models allow for dependence, they lack the desirable properties that result from conditioning in SCCS. There have

been recent efforts to model and investigate event dependence within the framework of the SCCS model. Hocine et al. (2005) developed a method to test for independence between different types of events in a bivariate setting, and Farrington and Hocine (2010) present a method that extends SCCS by imposing a dependence function in the intensity of a multidimensional Poisson model.

In this Chapter we propose a model that does not require the assumption of conditional independence between event occurrences. We refer to our generalization as the *positive dependence self-controlled case series* (PD-SCCS) method in order to emphasize that dependence is due to a positive term added to the baseline event risk. PD-SCCS lets individual risks increase additively through the inclusion of a birth process component and, as such, models a type of dependence that cannot be incorporated in Poisson process-based methods. At the same time our model preserves the key advantages of the original SCCS model: it automatically adjusts for multiplicative fixed baseline covariates without explicitly including them in the model, and data requirements are reduced since the analysis depends only on cases. Fixed individual-level baseline parameters drop out of the likelihood, as they do in the SCCS model. Data sources used for postmarketing surveillance can contain upwards of tens of millions of people, so in this context it is particularly advantageous that PD-SCCS avoids doing a costly estimation of these individual parameters. We will develop expressions for large sample inference and optimization based on the PD-SCCS conditional likelihood and compare the results of our generalized model with the more restrictive SCCS approach. The PD-SCCS model is developed and given a detailed treatment in Simpson (2011).

4.2 Model specification

4.2.1 Notation and framework

Suppose that there are N individuals in the data, indexed by $i = 1, \ldots, N$. Person i is observed over an interval of time $(a_i,\ b_i]$ of length $\tau_i = b_i - a_i$, where the unit of duration may be taken to be calendar time, age, or another appropriate measure. Let $N_i(a_i, t)$ represent the number of events that i experiences during the time interval $(a_i,\ t]$. For ease of notation we will write $N_i(t) = N_i(a_i, t)$ where the time interval is understood to begin at a_i and $\{N_i(t), a_i \leq t\}$ is the counting process for events experienced by i. Assume that n_i is the total number of events observed for i, i.e. $N_i(b_i) = n_i$. These events occur at times $t_{i1} < \cdots < t_{in_i}$, where t_{ij} corresponds to the time of the jth event for i. We assume that time is continuous and that there are no ties between event times.

Our objective is to determine the association between the event process and a set of relevant time-varying covariates. Suppose that there are p such covariates, indexed by $k = 1, \ldots, p$, whose values for person i at time t are represented by the vector $\mathbf{x}_i(t) = (x_{i1}(t), \ldots, x_{ip}(t))^{\mathrm{T}}$. We focus on the case where covariates represent exposure indicators, so that the kth element of the vector may be written as $x_{ik}(t)$ $= 1$ if i experiences exposure k at time t and $x_{ik}(t) = 0$ otherwise. This leads to simplifications in storage and computation, and is the form we will use in our examples. It is also valid, however, for the covariate vector $\mathbf{x}_i(t)$ to be made up of general (non-indicator) functions of time.

The process history at time t is denoted by $H_i(t) = \{N_i(s) : a_i \leq s < t\}$, which comprises all past values of the counting process for i until time t^-. An event process such as $N_i(t)$ can be defined in terms of its intensity function $\lambda_i(t \mid H_i(t))$, which gives the instantaneous probability that an event occurs at time t given the process history (Cook and Lawless, 2007). In our development the intensity function at t is

also conditional on the current covariate history $\{\mathbf{x}_i(s) : a_i \leq s \leq t\}$, but in order to simplify notation this will not be explicitly stated. In addition, all likelihoods and probability statements should be assumed to be conditional on covariate history.

4.2.2 Poisson-based continuous time SCCS model

To allow for person-level heterogeneity, each person i is assumed to have an individual baseline intensity of e^{ϕ_i}. During any intervals of time where i is exposed to variable k, there is a multiplicative effect of e^{β_k} on the baseline intensity, i.e. e^{β_k} gives the relative risk of an event due to exposure k. Letting $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^{\mathrm{T}}$ denote the vector of effects for the p exposures, the SCCS intensity for i at time t can be written as

$$\lambda_i(t \mid H_i(t)) = e^{\phi_i + \mathbf{x}_i(t)^{\mathrm{T}}\boldsymbol{\beta}}. \tag{4.1}$$

The intensity function in (4.1) is that of a Poisson process because at any time point t it is independent of the current process history $H_i(t)$. The likelihood contribution of i follows from the probability density of observing n_i events at times $t_{i1} < \cdots < t_{in_i} \in (a_i, b_i]$. In the case of the Poisson intensity this takes the form

$$L_i = e^{n_i\phi_i} \prod_{j=1}^{n_i} e^{\mathbf{x}_i(t_{ij})^{\mathrm{T}}\boldsymbol{\beta}} \times \exp\Big\{ - e^{\phi_i} \int_{a_i}^{b_i} e^{\mathbf{x}_i(u)^{\mathrm{T}}\boldsymbol{\beta}}\, du \Big\}.$$

In many applications, $\boldsymbol{\beta}$ is the parameter of interest because it determines the relationship between events and exposures. On the other hand the individual parameter ϕ_i provides no information about this relationship and can be treated as a nuisance parameter. The factorization theorem implies that the observed number of events n_i is a sufficient statistic for ϕ_i. As a consequence, conditioning on n_i will remove ϕ_i from the likelihood expression for i.

The corresponding conditional likelihood contribution for i is

$$L_i^c = \prod_{j=1}^{n_i} \left(\frac{e^{\mathbf{x}_i(t_{ij})^{\mathrm{T}}\boldsymbol{\beta}}}{\int_{a_i}^{b_i} e^{\mathbf{x}_i(u)^{\mathrm{T}}\boldsymbol{\beta}}\, du} \right). \tag{4.2}$$

Since $N_i(t)$ follows a non-homogeneous Poisson process, the total count $N_i(b_i)$ is a Poisson random variable with mean $\int_{a_i}^{b_i} \lambda_i(u \mid \mathbf{x}_i(u))du$. The value of this Poisson density for $N_i(b_i)$ at n_i yields the term in the denominator.

4.2.3 Our proposed PD-SCCS model

The intensity function for our proposed PD-SCCS model for person i at time t takes the form

$$\lambda_i(t \mid H_i(t)) = (e^{\phi_i} + \delta\, N_i(t^-))\, e^{\mathbf{x}_i(t)^{\mathrm{T}}\boldsymbol{\beta}}. \tag{4.3}$$

The intensity in (4.3) varies according to the event count $N_i(t^-)$ just prior to time t, which is part of the process history $H_i(t)$ that is conditioned upon. Due to this dependence on $H_i(t)$, events no longer follow a Poisson model, but instead arise from a non-homogeneous pure birth process with immigration (Parzen, 1964). At time t an event occurs either as an "immigrant" from a non-homogeneous Poisson process with intensity $e^{\phi_i + \mathbf{x}_i(t)^{\mathrm{T}}\boldsymbol{\beta}}$ or as a "birth" with intensity $\delta N_i(t^-) e^{\mathbf{x}_i(t)^{\mathrm{T}}\boldsymbol{\beta}}$. Since a process intensity must be non-negative, we restrict the space of the dependence parameter to $\delta \geq 0$. When $\delta = 0$ the PD-SCCS intensity model in (4.3) reduces to that of the SCCS model in (4.1).

For a general intensity-based process, the likelihood contribution for individual i is (Cook and Lawless, 2007)

$$L_i = \prod_{j=1}^{n_i} \lambda_i(t_{ij} \mid H_i(t_{ij})) \times \exp\Big\{ - \int_{a_i}^{b_i} \lambda_i(u \mid H_i(u))du \Big\}.$$

For our proposed intensity in (4.3), the likelihood for i is

$$\begin{aligned} L_{PD,i} = \frac{\Gamma(\frac{e^{\phi_i}}{\delta} + n_i)}{\Gamma(\frac{e^{\phi_i}}{\delta})} \exp\Big\{ - e^{\phi_i} \int_{a_i}^{b_i} e^{\mathbf{x}_i(u)^{\mathrm{T}}\boldsymbol{\beta}} du \Big\} \\ \times \exp\Big\{ - \delta \int_{a_i}^{b_i} N_i(v^-) e^{\mathbf{x}_i(v)^{\mathrm{T}}\boldsymbol{\beta}} dv \Big\} \prod_{j=1}^{n_i} \delta e^{\mathbf{x}_i(t_{ij})^{\mathrm{T}}\boldsymbol{\beta}}. \end{aligned} \tag{4.4}$$

From this likelihood expression we can see that n_i is sufficient for the person-specific parameter ϕ_i. This is a result of $N_i(t^-)$ having an additive effect on the baseline e^{ϕ_i}; if the event count were included as a multiplicative effect sufficiency would not hold. Conditioning on n_i removes the nuisance parameter ϕ_i from the likelihood, as it did in the SCCS model.

Determining the conditional likelihood requires an expression for the density of the total event count n_i, which is found by integrating $L_{PD,i}$ over all possible ways for n_i events to occur in $(a_i, b_i]$. This density takes the form

$$p(n_i) \propto \int_{\substack{\{\mathbf{t}_i \in (a_i,b_i]^{n_i}: \\ t_{i1}<\cdots<t_{in_i}\}}} \cdots \int \exp\Big\{ -\delta \int_{a_i}^{b_i} N_i(u^-) e^{\mathbf{x}_i(u)^{\mathrm{T}}\boldsymbol{\beta}} du \Big\} \prod_{j=1}^{n_i} e^{\mathbf{x}_i(t_{ij})^{\mathrm{T}}\boldsymbol{\beta}} dt_{i1} \cdots dt_{in_i}. \quad (4.5)$$

The integral inside the exponential term is a function of the ordered times $t_{i1} < \cdots < t_{in_i}$ since the step function $N_i(t^-)$ jumps at successive events. However this integral can also be expressed as

$$\int_{a_i}^{b_i} N_i(u^-) e^{\mathbf{x}_i(u)^{\mathrm{T}}\boldsymbol{\beta}} du = \sum_{j=1}^{n_i} \int_{t_{ij}}^{b_i} e^{\mathbf{x}_i(u)^{\mathrm{T}}\boldsymbol{\beta}} du. \quad (4.6)$$

Substituting in the expression on the right-hand side of (4.6), the density in (4.5) becomes

$$p(n_i) \propto \int_{\substack{\{\mathbf{t}_i \in (a_i,b_i]^{n_i}: \\ t_{i1}<\cdots<t_{in_i}\}}} \cdots \int \prod_{j=1}^{n_i} e^{\mathbf{x}_i(t_{ij})^{\mathrm{T}}\boldsymbol{\beta}} \exp\Big\{ -\delta \int_{t_{ij}}^{b_i} e^{\mathbf{x}_i(u)^{\mathrm{T}}\boldsymbol{\beta}} du \Big\} dt_{i1} \cdots dt_{in_i}. \quad (4.7)$$

The integrand in (4.7) is a product in j, which is invariant to the ordering of the t_{ij} variables. For a point $\mathbf{t}_i = (t_{i1}, \ldots, t_{in_i})^{\mathrm{T}} \in (a_i, b_i]^{n_i}$ whose elements are not necessarily in ascending order, let $\{t_{i(1)}, \ldots, t_{i(n_i)}\}$ be its set of order statistics where $t_{i(1)} < \cdots < t_{i(n_i)}$. For any such set of order statistics there are $n_i!$ corresponding points that can be generated via permutation. It follows that $(a_i, b_i]^{n_i}$ can be partitioned into $n_i!$ regions, each of which imposes an ordering on the elements of its

points. The integration region in (4.7) corresponds to one such ordering and the integrand is symmetric with respect to permutation. As a result (4.7) can be evaluated as

$$\begin{aligned} p(n_i) &\propto \frac{1}{n_i!}\prod_{j=1}^{n_i}\int_{a_i}^{b_i} e^{\mathbf{x}_i(t_{ij})^{\mathrm{T}}\boldsymbol{\beta}} \exp\Big\{-\delta\int_{t_{ij}}^{b_i} e^{\mathbf{x}_i(u)^{\mathrm{T}}\boldsymbol{\beta}}du\Big\}dt_{ij} \\ &= \frac{1}{n_i!\delta^{n_i}}\Big(1-\exp\Big\{-\delta\int_{a_i}^{b_i} e^{\mathbf{x}_i(u)^{\mathrm{T}}\boldsymbol{\beta}}du\Big\}\Big)^{n_i}. \end{aligned} \tag{4.8}$$

From the likelihood in (4.4) and density in (4.8), the full conditional likelihood for the PD-SCCS model can be written as

$$\begin{aligned} L^c_{PD} &= \prod_{i=1}^{N}\frac{p(n_i \text{ events at times } t_{i1}<\cdots<t_{in_i}\in(a_i,b_i]^{n_i})}{p(n_i)} \\ &= \prod_{i=1}^{N} n_i!\exp\Big\{-\delta\int_{a_i}^{b_i} N_i(u^-)e^{\mathbf{x}_i(u)^{\mathrm{T}}\boldsymbol{\beta}}du\Big\}\prod_{j=1}^{n_i}\left(\frac{\delta e^{\mathbf{x}_i(t_{ij})^{\mathrm{T}}\boldsymbol{\beta}}}{1-e^{-\delta\int_{a_i}^{b_i}e^{\mathbf{x}_i(v)^{\mathrm{T}}\boldsymbol{\beta}}dv}}\right). \end{aligned} \tag{4.9}$$

In the limit as $\delta \to 0$, this PD-SCCS conditional likelihood converges to the SCCS likelihood in (4.2). Inference for PD-SCCS is based on the conditional likelihood in (4.9). A key outcome of conditioning on n_i is that individuals with no events ($n_i = 0$) have a likelihood of $L^c_{PD,i} = 1$, and thus make no contribution to the estimation. It follows that only cases ($n_i \geq 1$) need to be included in the analysis.

Under the original SCCS model, only exposed cases are needed since unexposed cases ($\mathbf{x}_i(t) \equiv \mathbf{0}$ for all $t \in (a_i, b_i]$) have no effect on $\widehat{\boldsymbol{\beta}}$ or parameter inference. We discussed this feature of the SCCS model in Section 3.2.2. For PD-SCCS, however, unexposed cases can impact the $\hat{\delta}$ estimate, which in turn influences $\widehat{\boldsymbol{\beta}}$ and parameter inference. Thus, if unexposed cases are available in the data they should be included in the PD-SCCS analysis. Nonetheless, there is still a substantial reduction in computational cost compared to an unconditional analysis since the number of cases is often much smaller than the total number of individuals. For example, data sources used for postmarketing surveillance can contain tens of millions of individuals, but for a particular AE the number of cases is typically on the order of one hundred thousand.

4.3 Optimization and large sample inference

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}^{\mathrm{T}}, \delta)^{\mathrm{T}}$ denote the full parameter vector for our PD-SCCS model. We use maximum likelihood based on the expression in (4.9) in order to estimate the components of $\boldsymbol{\theta}$. Parameter variances can be estimated from the observed information matrix $\mathbf{I}(\boldsymbol{\theta})$ (the negative of the Hessian matrix) derived from the PD-SCCS conditional log-likelihood. We perform optimization via the Newton-Raphson algorithm, which relies on the score (gradient) vector $\mathbf{U}(\boldsymbol{\theta})$ and observed information matrix $\mathbf{I}(\boldsymbol{\theta})$. Expressions for the components of $\mathbf{U}(\boldsymbol{\theta})$ and $\mathbf{I}(\boldsymbol{\theta})$ are given in (4.10) and (4.11), respectively. If $\widehat{\boldsymbol{\theta}}^{(0)}$ is taken to be the initial estimate of the full parameter vector, optimization proceeds via the updating formula $\widehat{\boldsymbol{\theta}}^{(k+1)} = \widehat{\boldsymbol{\theta}}^{(k)} + \mathbf{I}^{-1}(\widehat{\boldsymbol{\theta}}^{(k)})\mathbf{U}(\widehat{\boldsymbol{\theta}}^{(k)})$ for $k = 0, 1, \ldots$ until suitable convergence criteria are met.

In order to simplify notation, let $g_i(\boldsymbol{\theta}) = \exp\{-\delta \int_{a_i}^{b_i} e^{\mathbf{x}_i(u)^{\mathrm{T}}\boldsymbol{\beta}} du\}/(1 - \exp\{-\delta \int_{a_i}^{b_i} e^{\mathbf{x}_i(v)^{\mathrm{T}}\boldsymbol{\beta}} dv\})$. Then the elements of the score vector $\mathbf{U}(\boldsymbol{\theta}) = (\mathbf{U}_{\beta}^{\mathrm{T}}(\boldsymbol{\theta}), \mathbf{U}_{\delta}(\boldsymbol{\theta}))^{\mathrm{T}}$ can be written as

$$\begin{aligned} \mathbf{U}_{\beta}(\boldsymbol{\theta}) &= \sum_{i=1}^{N} \left[\sum_{j=1}^{n_i} \mathbf{x}_i(t_{ij}) - \delta \int_{a_i}^{b_i} \mathbf{x}_i(u) \big(N_i(u^-) + n_i g_i(\boldsymbol{\theta}) \big) e^{\mathbf{x}_i(u)^{\mathrm{T}}\boldsymbol{\beta}} du \right] \\ \mathbf{U}_{\delta}(\boldsymbol{\theta}) &= \sum_{i=1}^{N} \left[\frac{n_i}{\delta} - \int_{a_i}^{b_i} \big(N_i(u^-) + n_i g_i(\boldsymbol{\theta}) \big) e^{\mathbf{x}_i(u)^{\mathrm{T}}\boldsymbol{\beta}} du \right]. \end{aligned} \tag{4.10}$$

It follows that the components of the observed PD-SCCS information matrix $\mathbf{I}(\boldsymbol{\theta})$

are

$$\begin{aligned}
\mathbf{I}_{\beta\beta}(\boldsymbol{\theta}) &= \delta \sum_{i=1}^{N} \int_{a_i}^{b_i} \mathbf{x}_i^{\otimes 2}(v)\{N_i(v^-) + n_i g_i(\boldsymbol{\theta})\} e^{\mathbf{x}_i(v)^{\mathrm{T}}\boldsymbol{\beta}} dv \\
&\qquad - \delta^2 \sum_{i=1}^{N} n_i g_i(\boldsymbol{\theta})\{1 + g_i(\boldsymbol{\theta})\}\Big(\int_{a_i}^{b_i} \mathbf{x}_i(u) e^{\mathbf{x}_i(u)^{\mathrm{T}}\boldsymbol{\beta}}\Big)^{\otimes 2} \\
\mathbf{I}_{\delta\beta}(\boldsymbol{\theta}) &= \sum_{i=1}^{N} \int_{a_i}^{b_i} \mathbf{x}_i(u) \Bigg(N_i(u^-) + n_i g_i(\boldsymbol{\theta}) \big[1 - \delta\{1 + g_i(\boldsymbol{\theta})\} \int_{a_i}^{b_i} e^{\mathbf{x}_i(v)^{\mathrm{T}}\boldsymbol{\beta}} dv\big] \Bigg) e^{\mathbf{x}_i(u)^{\mathrm{T}}\boldsymbol{\beta}} du \\
\mathbf{I}_{\delta\delta}(\boldsymbol{\theta}) &= \sum_{i=1}^{N} n_i \Bigg[\frac{1}{\delta^2} - g_i(\boldsymbol{\theta})\{1 + g_i(\boldsymbol{\theta})\}\Big(\int_{a_i}^{b_i} e^{\mathbf{x}_i(u)^{\mathrm{T}}\boldsymbol{\beta}} du\Big)^2 \Bigg],
\end{aligned} \tag{4.11}$$

where $\mathbf{x}^{\otimes 2} = \mathbf{x}\mathbf{x}^{\mathrm{T}}$ and

$$\mathbf{I}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{I}_{\beta\beta}(\boldsymbol{\theta}) & \mathbf{I}_{\delta\beta}^{\mathrm{T}}(\boldsymbol{\theta}) \\ \mathbf{I}_{\delta\beta}(\boldsymbol{\theta}) & \mathbf{I}_{\delta\delta}(\boldsymbol{\theta}) \end{pmatrix} \quad \text{and} \quad \mathbf{I}^{-1}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{I}^{\beta\beta}(\boldsymbol{\theta}) & \mathbf{I}^{\delta\beta}(\boldsymbol{\theta})^{\mathrm{T}} \\ \mathbf{I}^{\delta\beta}(\boldsymbol{\theta}) & \mathbf{I}^{\delta\delta}(\boldsymbol{\theta}) \end{pmatrix}.$$

Standard asymptotic theory for maximum likelihood is applicable when $\boldsymbol{\theta}$ lies on the interior of its parameter space ($\boldsymbol{\beta} \in \Re^p$, $\delta > 0$), in which case the estimate $\widehat{\boldsymbol{\theta}}$ can be treated as if it is distributed as $(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \sim N(\mathbf{0}, \mathbf{I}^{-1}(\boldsymbol{\theta}))$. We can use $\mathbf{I}^{-1}(\widehat{\boldsymbol{\theta}})$ to estimate the covariance matrix of $\widehat{\boldsymbol{\theta}}$, where $\widehat{\mathrm{cov}}(\widehat{\boldsymbol{\beta}}) = \mathbf{I}^{\beta\beta}(\widehat{\boldsymbol{\theta}})$ and $\widehat{\mathrm{var}}(\hat{\delta}) = \mathbf{I}^{\delta\delta}(\widehat{\boldsymbol{\theta}})$.

In the PD-SCCS model, the value $\delta = 0$ lies at the boundary of the parameter space. Moran (1971) considers the asymptotic behavior of maximum likelihood estimates in the boundary case, which results in nonstandard distributions. Under the conditions given in Moran (1971), the distribution of $\hat{\delta}$ converges to an equal mixture of a positive half-normal distribution and a point mass at zero. When the true parameter is $\delta = 0$, the observed information $\mathbf{I}^{\delta\delta}(\widehat{\boldsymbol{\theta}})$ estimates the variance of the full-normal distribution, rather than that of the half-normal mixture distribution. This adjustment needs to be accounted for when hypothesis testing at $\delta = 0$, which will be discussed in Section 4.4.2.

4.4 Simulation experiments

In this section we assess the performance of PD-SCCS based on simulation experiments. In our investigation we simulate data from both the PD-SCCS ($\delta > 0$) and SCCS ($\delta = 0$) models. We examine PD-SCCS estimation results under a variety of parameter settings, and compare these results to those obtained from fitting SCCS. Simulations also allow us to evaluate the accuracy of a one-sided Wald test for $\delta = 0$ and of AIC and BIC model selection criteria.

4.4.1 True model is PD-SCCS

Tables 4.1-4.3 show results of simulating from the PD-SCCS model. In every iteration we created $N = 1000$ individuals, all with an observation period of $\tau = 2500$ days, and ran 2000 iterations for each setting of the parameters. We used a single time-varying covariate that was an indicator of exposure over time. Values of β were varied to yield relative event rates of $e^{\beta} = 0.75$, 1, and 1.25 for exposure versus baseline. We assigned a common baseline rate of e^{ϕ} to all individuals, which we set to 0.00025, 0.0005, and 0.001. These values correspond to expected total numbers of events of 0.625, 1.25, and 2.5, respectively, in an unconditional Poisson model where the event rate is e^{ϕ}. For each combination of e^{ϕ} and e^{β}, the positive dependence parameter varied between $\delta = 0.00025$, 0.0005, and 0.001. Exposures for each individual were generated by splitting the observation period randomly into between 3 and 8 intervals. Days on which to start and end these intervals were chosen uniformly throughout the observation period, and exposure status for each interval alternated between exposed and unexposed.

Table 4.1 shows MSE and MAD for $\hat{\beta}$ when the true model is PD-SCCS. The SCCS and PD-SCCS models yield similar performance in terms of MSE and MAD when the level of dependence is small ($\delta = 0.00025$), however the level of error for

SCCS grows with increasing δ. The difference in accuracy between the models is most pronounced for large values of the relative event rate e^{β} and baseline rate e^{ϕ}.

Table 4.2 shows variance estimation and interval coverage results for β when the true model is PD-SCCS. SM $\hat{\beta}$ denotes the square root of the mean of the $\widehat{\text{var}}(\hat{\beta})$ estimates over all iterations, where $\widehat{\text{var}}(\hat{\beta}) = \mathbf{I}^{\beta\beta}(\widehat{\boldsymbol{\theta}})$. SE $\hat{\beta}$ is the empirical standard error of the $\hat{\beta}$ estimates. Close agreement of SM $\hat{\beta}$ and SE $\hat{\beta}$ suggests that the estimates $\widehat{\text{var}}(\hat{\beta})$ are performing well. When fitting the PD-SCCS model, SM $\hat{\beta}$ and SE $\hat{\beta}$ are close in value and interval coverage levels are near 95%. On the other hand, fitting SCCS results in less accurate $\widehat{\text{var}}(\hat{\beta})$ estimates and lower confidence interval coverage. In particular, when the dependence parameter value is high ($\delta = 0.001$) observed interval coverage levels all lie below 84%.

Table 4.3 shows results for $\hat{\delta}$ when PD-SCCS is the true model. The Mean $\hat{\delta}$ column shows the mean of the δ estimates over all iterations. In general the mean bias is less than 5% of the estimated standard error SM $\hat{\delta}$, except in cases where the dependence parameter is high ($\delta = 0.001$). In these cases the value of δ tends to be underestimated. The estimated SM $\hat{\delta}$ and empirical SE $\hat{\delta}$ standard errors are close in value in most cases, with a few exceptions. Interval coverage levels are generally close to 95% and lie above 88% in all cases. In our simulations, there was no observed difference between AIC and BIC for the proportion of times that the criteria favored PD-SCCS over SCCS. Observed proportions for AIC and BIC were greater than 0.998 for all parameter settings, which demonstrates that both criteria are able to correctly select the PD-SCCS model for the given parameter values.

Table 4.1: Estimation error for $\hat{\beta}$ when PD-SCCS is true model

e^{β}	e^{ϕ} ($\times 10^{-4}$)	δ ($\times 10^{-4}$)	MSE $\hat{\beta}$ ($\times 10^{-3}$)		MAD $\hat{\beta}$ ($\times 10^{-2}$)	
			SCCS	PD-SCCS	SCCS	PD-SCCS
		2.5	3.02	3.00	4.37	4.36
	2.5	5	2.23	2.03	3.81	3.61
		10	2.14	1.04	3.70	2.57
		2.5	2.33	2.28	3.84	3.80
0.75	5	5	1.91	1.69	3.43	3.25
		10	1.81	0.95	3.40	2.30
		2.5	1.66	1.75	3.24	3.23
	10	5	1.34	1.13	2.91	2.68
		10	1.44	0.52	3.06	1.83
		2.5	2.62	2.61	4.06	4.04
	2.5	5	1.94	1.76	3.52	3.35
		10	1.64	0.77	3.24	2.22
		2.5	2.09	2.08	3.65	3.60
1	5	5	1.57	1.35	3.13	2.90
		10	1.55	0.58	3.15	1.92
		2.5	1.36	1.34	2.94	2.88
	10	5	1.07	9.46	2.63	2.45
		10	1.12	0.36	2.69	1.50
		2.5	2.49	2.46	4.00	3.97
	2.5	5	1.90	1.57	3.43	3.15
		10	2.28	0.58	3.82	1.93
		2.5	1.97	1.89	3.54	3.48
1.25	5	5	1.46	1.18	3.06	2.73
		10	2.02	0.45	3.64	1.70
		2.5	1.14	1.18	2.70	2.65
	10	5	0.98	0.71	2.50	2.12
		10	1.73	0.28	3.36	1.34

Table 4.2: Results for $\hat{\beta}$ and $\widehat{\text{var}}(\hat{\beta})$ when PD-SCCS is true model

e^{β}	e^{ϕ} ($\times 10^{-4}$)	δ ($\times 10^{-4}$)	SCCS ($\times 10^{-2}$)		PD-SCCS ($\times 10^{-2}$)		95% CI β (%)	
			SM $\hat{\beta}$	SE $\hat{\beta}$	SM $\hat{\beta}$	SE $\hat{\beta}$	SCCS	PD-SCCS
		2.5	5.39	5.49	5.37	5.48	94.30	94.25
	2.5	5	4.61	4.70	4.57	4.51	94.85	94.80
		10	3.28	4.26	3.17	3.23	83.80	94.45
		2.5	4.79	4.83	4.77	4.77	94.60	95.25
0.75	5	5	4.10	4.35	4.06	4.12	93.15	94.30
		10	2.91	3.80	2.81	3.07	81.20	94.45
		2.5	3.91	4.08	3.90	4.18	94.25	93.85
	10	5	3.35	3.59	3.32	3.35	92.40	94.75
		10	2.38	3.25	2.30	2.28	78.50	95.35
		2.5	5.14	5.11	5.12	5.11	95.30	95.30
	2.5	5	4.29	4.40	4.24	4.19	94.50	95.30
		10	2.88	4.05	2.76	2.77	82.50	95.05
		2.5	4.50	4.58	4.49	4.57	94.85	94.90
1	5	5	3.76	3.97	3.71	3.68	93.25	94.90
		10	2.52	3.94	2.42	2.41	78.95	95.25
		2.5	3.61	3.69	3.60	3.66	93.85	94.55
	10	5	3.02	3.27	2.98	3.08	93.45	94.95
		10	2.02	3.35	1.94	1.90	75.90	95.05
		2.5	4.97	4.99	4.95	4.96	94.90	94.85
	2.5	5	4.05	4.31	3.98	3.96	93.20	95.15
		10	2.55	4.20	2.43	2.40	70.00	95.75
		2.5	4.29	4.44	4.28	4.35	94.40	95.05
1.25	5	5	3.50	3.77	3.44	3.43	92.85	95.05
		10	2.21	3.88	2.10	2.12	65.00	95.00
		2.5	3.39	3.37	3.38	3.44	95.20	95.05
	10	5	2.76	3.07	2.72	2.67	91.35	94.90
		10	1.74	3.37	1.66	1.65	57.65	94.40

Table 4.3: Results for $\hat{\delta}$ when PD-SCCS is true model ($\delta > 0$)

e^{β}	e^{ϕ} ($\times 10^{-4}$)	δ ($\times 10^{-4}$)	Mean $\hat{\delta}$ ($\times 10^{-4}$)	SM $\hat{\delta}$ ($\times 10^{-5}$)	SE $\hat{\delta}$ ($\times 10^{-5}$)	95% CI δ (%)	AIC	BIC
		2.5	2.50	3.87	3.86	95.10	1	1
	2.5	5	5.01	3.48	3.47	95.50	1	1
		10	9.96	2.88	2.82	94.85	1	1
		2.5	2.51	3.43	3.38	95.55	1	1
0.75	5	5	4.99	3.08	3.02	95.10	1	1
		10	9.95	2.55	3.42	94.20	0.9995	0.9995
		2.5	2.50	2.80	2.82	95.45	0.9995	0.9995
	10	5	5.00	2.52	2.51	94.90	1	1
		10	9.94	2.08	2.03	94.65	1	1
		2.5	2.50	3.30	3.34	95.20	0.999	0.999
	2.5	5	5.00	2.98	3.74	94.25	0.998	0.998
		10	9.95	2.48	2.41	94.95	1	1
		2.5	2.51	2.90	2.98	94.65	0.999	0.999
1	5	5	5.01	2.61	2.62	95.60	1	1
		10	9.94	2.18	2.15	94.15	1	1
		2.5	2.50	2.32	2.38	94.60	0.9995	0.9995
	10	5	4.99	2.09	2.37	94.60	0.9995	0.9995
		10	9.91	1.74	1.74	91.45	1	1
		2.5	2.50	2.86	2.91	94.85	1	1
	2.5	5	5.00	2.60	2.54	95.40	1	1
		10	9.93	2.17	2.18	93.25	1	1
		2.5	2.50	2.47	2.43	95.05	1	1
1.25	5	5	5.00	2.24	2.29	94.65	1	1
		10	9.91	1.87	1.87	91.85	1	1
		2.5	2.50	1.95	2.11	94.90	0.999	0.999
	10	5	5.00	1.77	1.79	94.85	1	1
		10	9.89	1.48	1.48	88.05	1	1

4.4.2 SCCS simulations

Tables 4.4-4.6 show results of simulating from the SCCS model. In every iteration we generated $N = 1000$ individuals, all of whom had observation period lengths of $\tau = 2500$ days. We created exposures in the same manner as in the PD-SCCS simulations. Relative risk due to exposure was set to $e^{\beta} = 0.75$, 1, and 1.25, and the initial baseline rate for all individuals was set to $e^{\phi} = 0.00025$, 0.0005, 0.001, and 0.002.

Table 4.4 shows MSE and MAD for $\hat{\beta}$ when the true model is SCCS. PD-SCCS and SCCS error levels are closer for higher values of e^{β}. SCCS generally has a lower level of error than PD-SCCS according to both metrics, which is sensible since SCCS is the true model in this case.

Table 4.4: Estimation error for $\hat{\beta}$ when SCCS is true model

e^{β}	e^{ϕ} ($\times 10^{-4}$)	MSE $\hat{\beta}$ ($\times 10^{-3}$)		MAD $\hat{\beta}$ ($\times 10^{-2}$)	
		SCCS	PD-SCCS	SCCS	PD-SCCS
0.75	2.5	3.91	4.35	4.98	5.26
	5	3.17	5.38	4.51	5.17
	10	2.11	3.78	3.63	4.02
	20	1.15	3.01	2.72	3.26
1	2.5	3.53	3.80	4.77	4.95
	5	2.77	3.15	4.19	4.45
	10	1.74	2.06	3.34	3.52
	20	0.99	1.31	2.47	2.88
1.25	2.5	3.70	3.46	4.87	4.66
	5	2.68	2.70	4.11	4.19
	10	1.64	1.70	3.21	3.23
	20	0.93	1.18	2.43	2.66

Table 4.5 shows variance estimation and confidence interval coverage results for β when SCCS is the true model. SM $\hat{\beta}$ and SE $\hat{\beta}$ are close in value and observed interval coverage levels are near 95% for SCCS, which is expected since SCCS is the true model. PD-SCCS tends to underestimate the variability of $\hat{\beta}$ when the relative

risk is low ($e^{\beta} = 0.75$). This leads to lower interval coverage, however observed levels all lie above 91% for the given parameter settings.

Table 4.5: Results for $\hat{\beta}$ and $\widehat{\text{var}}(\hat{\beta})$ when SCCS is true model

e^{β}	e^{ϕ} ($\times 10^{-4}$)	SCCS ($\times 10^{-2}$)		PD-SCCS ($\times 10^{-2}$)		95% CI β (%)	
		SM $\hat{\beta}$	SE $\hat{\beta}$	SM $\hat{\beta}$	SE $\hat{\beta}$	SCCS	PD-SCCS
0.75	2.5	6.21	6.26	6.21	6.58	95.10	93.65
	5	5.52	5.63	5.52	7.26	94.75	92.15
	10	4.51	4.59	4.51	6.12	94.00	92.55
	20	3.37	3.39	3.37	5.47	95.15	92.45
1	2.5	6.06	5.94	6.06	6.09	95.80	95.00
	5	5.31	5.26	5.31	5.53	95.05	93.05
	10	4.26	4.17	4.26	4.50	95.55	94.75
	20	3.13	3.14	3.13	3.63	94.95	91.35
1.25	2.5	5.98	6.08	5.98	5.87	94.80	94.95
	5	5.17	5.18	5.17	5.09	95.00	95.95
	10	4.08	4.06	4.08	4.13	95.30	95.25
	20	2.97	3.05	2.97	3.44	93.95	91.15

Table 4.6 shows simulation results for δ when fitting PD-SCCS to the SCCS model, where the true PD-SCCS dependence parameter δ is equal to zero. The first column gives the mean of the $\hat{\delta}$ estimates over all iterations, which for $\delta = 0$ is equivalent to MAD $\hat{\delta}$. We can see that fitting PD-SCCS yields small positive $\hat{\delta}$ estimates, and that the estimates approach the correct value of zero as the event rate increases.

As discussed in Section 4.3, when the true parameter lies at the boundary $\delta = 0$ the distribution of $\hat{\delta}$ approaches an equal mixture of a point mass at zero and a positive half-normal. The variance of a half-normal is $\sigma^2(1 - 2/\pi)$, where σ^2 is the variance of the analogous full-normal distribution. This suggests that the estimate $\widehat{\sigma}^2 = \mathbf{I}^{\delta\delta}(\widehat{\boldsymbol{\theta}})$ should be scaled by a factor of $1/2 \times (1 - 1/\pi))$ for accurate comparison with empirical standard errors. Table 4.6 shows that the adjusted estimates (Adj. SM $\hat{\delta}$) and empirical standard errors (SE $\hat{\delta}$) are close in value, which indicates that $\mathbf{I}^{\delta\delta}(\widehat{\boldsymbol{\theta}})$ is an accurate estimate for σ^2.

For the boundary parameter mixture distribution, the $1-\alpha$ quantile lies at $\sqrt{2}\sigma\ \text{erf}^{-1}(1-2\alpha)$. Thus to test the hypothesis $H_0 : \delta = 0$ (SCCS) against $H_1 : \delta > 0$ (PD-SCCS), one may use a one-sided Wald test and reject at the α-level if

$$\frac{\hat{\delta}}{\widehat{\sigma}} \geq \sqrt{2}\ \text{erf}^{-1}(1-2\alpha) = \Phi^{-1}(1-\alpha),$$

where $\Phi(.)$ is the standard normal CDF. Observed type I errors for this test procedure (shown in Table 4.6) are close to the desired $\alpha = 0.05$ level. To select between the PD-SCCS and SCCS models, one method would be to fit the PD-SCCS model and test $H_0 : \delta = 0$ as above, where failing to reject H_0 suggests that positive dependence is not present and SCCS may be used. Alternatively one could evaluate the models based on AIC or BIC, with BIC being more conservative in how often it will favor the PD-SCCS model. The last two columns of Table 4.6 give the observed proportion of times that AIC and BIC selected PD-SCCS over SCCS. In this case BIC correctly favors SCCS more often, whereas AIC has higher error than the 0.05-level Wald test of $\delta = 0$.

Table 4.6: Results for $\hat{\delta}$ when SCCS is true model ($\delta = 0$)

e^{β}	e^{ϕ} ($\times 10^{-4}$)	Mean $\hat{\delta}$ ($\times 10^{-5}$)	Adj SM $\hat{\delta}$ ($\times 10^{-5}$)	SE $\hat{\delta}$ ($\times 10^{-5}$)	Type I err (%)	AIC ($\times 10^{-2}$)	BIC ($\times 10^{-3}$)
0.75	2.5	1.85	2.56	2.65	5.10	8.25	6.00
	5	1.57	2.26	2.30	5.20	8.30	5.00
	10	1.25	1.85	1.84	5.00	7.75	4.50
	20	1.00	1.38	1.42	5.25	8.50	5.00
1	2.5	1.54	2.20	2.21	5.10	8.05	4.50
	5	1.34	1.93	1.91	5.20	8.00	2.50
	10	1.10	1.55	1.55	5.20	7.90	4.00
	20	0.81	1.14	1.14	5.00	8.05	3.50
1.25	2.5	1.39	1.92	2.01	6.15	9.10	4.50
	5	1.12	1.67	1.64	4.65	7.80	4.00
	10	0.95	1.31	1.34	5.15	8.10	6.00
	20	0.71	0.96	0.95	4.60	7.90	2.50

4.5 Example data analyses

We will consider two example analyses investigating the relationship between time-varying drug exposures and recurring MI events. For the outcome of MI, it is plausible that event occurrences exhibit positive dependence in a manner that is compatible with our PD-SCCS model. Data are composed of de-identified records from a longitudinal health insurance claims database. This database includes information on patient-level prescriptions and diagnoses, and covers approximately 36 million individuals between the years of 2000 and 2007.

4.5.1 Vioxx and myocardial infarction

Vioxx (rofecoxib) is a COX-2 selective nonsteroidal anti-inflammatory drug (NSAID) manufactured by Merck & Co, Inc. that went on the market in May 1999. Its primary indication was for the management of osteoarthritis symptoms and acute pain. Merck withdrew Vioxx worldwide in September 2004 due to evidence of an increased risk of cardiovascular events, including MI (Bresalier et al., 2005). Our objective is to estimate the strength of the association between Vioxx and MI while allowing for positive dependence between MI events.

In our database there were $N = 163387$ individuals who experienced at least one MI over the course of their observation, which qualifies them as cases for our analysis. MI events were required to be separated by a window of at least 60 days in order to minimize double counting. Under this definition, 137933 (84.42%) individuals had exactly one MI during their observation period, 18230 (11.16%) had two, 4430 (2.71%) had three, and 2794 (1.71%) had 4 or more. The median observation period length was 3.3 years, with a 25th percentile of 1.8 years and a 75th percentile of 5.4 years. For individuals who were exposed to Vioxx, we assumed that the physical exposure risk would persist during a window of 30 days following the last known availability

date of the drug. For our analyses we fit both the standard SCCS model and our PD-SCCS model with Vioxx exposure as the single time-varying covariate.

Table 4.7: Vioxx and MI Results

Parameter (95% CI)	SCCS	PD-SCCS
$\hat{\beta}$	0.15 (0.10, 0.20)	0.25 (0.20, 0.30)
$e^{\hat{\beta}}$	1.17 (1.11, 1.22)	1.29 (1.22, 1.35)
$\hat{\delta}$ ($\times 10^{-4}$)	-	2.74 (2.65, 2.83)
AIC	2850144	2846761
BIC	2850154	2846781

From the results in Table 4.7, we can see that SCCS estimates an elevated relative risk of MI due to Vioxx exposure with a 95% confidence interval of (1.11, 1.22). The estimated relative risk increases under the PD-SCCS model, with a confidence interval of (1.22, 1.35). The estimated interval for δ under the PD-SCCS model is (2.65×10^{-4}, 2.83×10^{-4}), and both AIC and BIC favor the PD-SCCS model over SCCS for this analysis. These results suggest that positive dependence is present between MI events, which matches the clinical understanding that the occurrence of an MI can increase an individual's future risk of MI.

4.5.2 Proton pump inhibitors and myocardial infarction

As a second example, we will compare the SCCS and PD-SCCS models when estimating the strength of association between proton pump inhibitors (PPIs) and MI. The main action of PPIs is to reduce gastric acid production, which is helpful for treating heartburn, inflammation of the esophagus, and ulcers. In 2007 the FDA performed a safety review of two PPIs, Prilosec (omeprazole) and Nexium (esomeprazole), after data from two long-term studies raised concerns about an increased risk of heart

problems associated with this class of drugs. After a full review, the FDA stated that "long-term use of omeprazole or esomeprazole is not likely to be associated with an increased risk of heart problems" and recommended that patients continue use of these drugs as prescribed (FDA, 2007a).

The relevant individuals for our analysis are those who had at least one MI during their period of observation. This is the same group of individuals we considered in the previous example with Vioxx and MI.

Table 4.8: PPI and MI Results

Parameter (95% CI)	SCCS	PD-SCCS
$\hat{\beta}$	0.08 (0.06, 0.10)	-0.03 (-0.05, -0.01)
$e^{\hat{\beta}}$	1.08 (1.06, 1.10)	0.97 (0.95, 0.99)
$\hat{\delta}$ ($\times 10^{-4}$)	-	2.76 (2.66, 2.86)
AIC	2850107	2846830
BIC	2850117	2846850

From the results in Table 4.8 we see that the SCCS confidence interval for e^{β} is (1.06, 1.10), suggesting a slightly elevated relative risk of MI during periods of PPI exposure. With the PD-SCCS model, however, the estimated risk is reduced and the confidence interval (0.95, 0.99) for e^{β} is close to 1, which is more consistent with the findings of the FDA's 2007 review. Results for the PD-SCCS dependence parameter δ are similar to those in the Vioxx and MI example, and both AIC and BIC favor the use of PD-SCCS for this analysis.

4.6 PD-SCCS vs. Farrington and Hocine (2010)

Farrington and Hocine (2010) present an approach that extends SCCS to allow for within-individual event dependence. This method treats the vector of observed event

times $\mathbf{t}_i = (t_{i1}, \ldots, t_{in_i})^{\mathrm{T}}$ for each individual i as a single point in an n_i-dimensional region, where n_i denotes the number of events experienced by individual i. This region is restricted to $Q_i(n_i) = \{\mathbf{t}_i \in (a_i, b_i]^{n_i} : t_{i1} < \cdots < t_{in_i}\}$ since the components of $\mathbf{t}_i$ are ordered by time and no event times can occur outside of the observation window $(a_i, b_i]$. The standard SCCS model assumes that events are realizations of a one-dimensional Poisson process and conditions upon the observed number of events n_i. In Farrington and Hocine (2010), however, the event time vector $\mathbf{t}_i$ is treated as a single point arising from an n_i-dimensional Poisson process. In this multi-dimensional framework, conditioning on n_i is equivalent to conditioning on the occurrence of a single point in the region $Q_i(n_i)$.

We will let $\mathbf{x}_i$ be shorthand for the full covariate history for person i, i.e. $\mathbf{x}_i = \{\mathbf{x}_i(t) : t \in (a_i, b_i]\}$. If $\lambda_i(t_1, \ldots, t_{n_i} \mid \mathbf{x}_i)$ is the intensity of the n_i-dimensional Poisson process on $Q_i(n_i)$, the likelihood of a single multidimensional event $\mathbf{t}_i$ is (Cressie, 1993)

$$L_i = \lambda_i(t_{i1}, \ldots, t_{in_i} \mid \mathbf{x}_i) \exp\Big\{ - \int_{Q_i(n_i)} \cdots \int \lambda_i(s_{i1}, \ldots, s_{in_i} \mid \mathbf{x}_i) ds_{i1} \cdots ds_{in_i} \Big\}.$$

It follows that the conditional likelihood of $\mathbf{t}_i$ given the occurrence of one such point in $Q_i(n_i)$ takes the form of a normalized multidimensional intensity, i.e.

$$L_i^c = \frac{p(\text{an event at } \mathbf{t}_i)}{p(\text{one event in } Q_i(n_i))} = \frac{\lambda_i(t_{i1}, \ldots, t_{in_i} \mid \mathbf{x}_i)}{\int_{Q_i(n_i)} \cdots \int \lambda_i(u_{i1}, \ldots, u_{in_i} \mid \mathbf{x}_i) du_{i1} \cdots du_{in_i}}, \tag{4.12}$$

where the exponential term ends up dropping out of the model. Farrington and Hocine assume that the n_i-dimensional Poisson intensity can be written in the form

$$\lambda_i(t_{i1}, \ldots, t_{in_i} \mid \mathbf{x}_i) = \prod_{j=1}^{n_i} \lambda_i(t_{ij} \mid \mathbf{x}_i) \times H_{n_i}(t_{i1}, \ldots, t_{in_i}), \tag{4.13}$$

where the product term is made up of independent univariate intensities $\lambda_i(t \mid \mathbf{x}_i)$, and the $H_{n_i}(.)$ function determines the dependence between events. Under the intensity in (4.13), the conditional likelihood takes the form

$$L_i^c = \frac{\prod_{j=1}^{n_i} \lambda_i(t_{ij} \mid \mathbf{x}_i) \times H_{n_i}(t_{i1}, \ldots, t_{in_i})}{\int_{Q_i(n_i)} \cdots \int \prod_{k=1}^{n_i} \lambda_i(u_{ik} \mid \mathbf{x}_i) \times H_{n_i}(u_{i1}, \ldots, u_{in_i}) du_{i1} \cdots du_{in_i}}. \tag{4.14}$$

From (4.14) we can see that multiplicative terms of $\lambda_i(t \mid \mathbf{x}_i)$ that are fixed in time will cancel out of the conditional likelihood, as they do in the original SCCS model. Similarly, fixed multiplicative terms in $H_{n_i}(.)$ will also drop out of the conditional likelihood. Different possible choices for the dependence function $H_{n_i}(.)$ are explored in Farrington and Hocine (2010). In their extension, the vector of event times $\mathbf{t}_i$ is treated as a single realization of an n_i-dimensional process, which is analogous to the perspective we used to evaluate the integral in the denominator of the PD-SCCS conditional likelihood in (4.7). However, Farrington and Hocine's method does not accommodate the type of positive event dependence that is represented in the PD-SCCS model.

If we approach the PD-SCCS model from an analogous multidimensional perspective, we can write out the n_i-dimensional PD-SCCS intensity for i as

$$\lambda_{PD,i}(t_{i1}, \ldots, t_{in_i} \mid \mathbf{x}_i) = \prod_{j=1}^{n_i} (e^{\phi_i} + \delta(j-1))\, e^{\mathbf{x}_i(t_{ij})^{\mathrm{T}}\boldsymbol{\beta}}. \tag{4.15}$$

Our event vector $\mathbf{t}_i$ is also restricted to the region $Q_i(n_i)$, as it was for Farrington and Hocine's model. It follows from (4.15) that the conditional likelihood takes the form

$$L^c_{PD,i} = \frac{\prod_{j=1}^{n_i} e^{\mathbf{x}_i(t_{ij})^{\mathrm{T}}\boldsymbol{\beta}} \exp\Big\{ -\delta \int_{t_{ij}}^{\tau_i} e^{\mathbf{x}_i(u)^{\mathrm{T}}\boldsymbol{\beta}}\, du \Big\}}{\int \cdots \int_{Q_i(n_i)} \prod_{k=1}^{n_i} e^{\mathbf{x}_i(s_{ik})^{\mathrm{T}}\boldsymbol{\beta}} \exp\Big\{ -\delta \int_{s_{ik}}^{\tau_i} e^{\mathbf{x}_i(u)^{\mathrm{T}}\boldsymbol{\beta}}\, du \Big\}\, ds_{i1} \cdots ds_{in_i}}.$$

The portion of the intensity that is fixed in time, i.e. $\prod_{j=1}^{n_i}(e^{\phi_i} + \delta(j-1))$ drops out of the conditional likelihood as we would expect. However, the exponential term that involves the birth process component remains in the likelihood since it is dependent on the observed event times $t_{i1}, \ldots, t_{in_i}$. Due to this additional birth process component, our model cannot be written as a multidimensional Poisson process since in the Poisson case the exponential term would drop out of the likelihood, as it did in (4.12). As a consequence, Farrington and Hocine's method is not able to represent the positive event dependence in the PD-SCCS model.

Chapter 5

Discussion

In Chapter 2 we introduced how the simple SCCS model can be used to identify new drug-AE associations in an active surveillance system based on LODs. We presented a unique development of the SCCS model from a perspective that differs from Farrington's presentation and connects SCCS with the well-established FEP model in econometrics literature. This connection opens promising directions for future work; for example Hausman et al. (1984) discuss a negative binomial version of the model that can overcome problems of overdispersion in Poisson counts, and Wooldridge (1997) proposes method of moments estimators that can be applied without imposing a strict exogeneity assumption, i.e. without requiring assumption (ii) in which future exposure values must be determined independently of the past event process.

The features of SCCS make it an appealing choice for postmarketing surveillance. SCCS automatically controls for fixed multiplicative individual-level baseline covariates, which is an advantage in the context of surveillance since insurance claims databases provide a limited amount of baseline information. This also prevents the need to estimate individual parameters, which in our problem would require high-dimensional optimization. SCCS depends only on data from exposed cases, which saves a considerable amount of computation since in LODs there are typically a mod-

est number of exposed cases compared to the total number of individuals.

We identified five drug-AE pairs for analysis in order to investigate whether applying the simple SCCS aproach for surveillance in claims data could lead to earlier detection of suspected associations than the usual DP method based on AERS data. Results suggest that earlier detection may have been possible in the case of the Vioxx-MI and estrogen-MI pairs. In addition, the simple SCCS analysis did not pick up the false negative pair (simvastatin-cataract), which had previously been identified in AERs data. As further evaluation of the simple SCCS method we looked at results of the drug-AE pair classification problem, which is based on the HOIs and DOIs defined by OMOP. We compared simple SCCS to the DP and OS approaches, which are both currently in use for surveillance. Results showed that simple SCCS had good performance for our evaluation metrics across the majority of the databases. We are currently working with OMOP to compare simple SCCS to a broader group of surveillance methods, and results of this evaluation will soon be available (see Ryan et al. (2011) for preliminary assessments). Our analyses are part of the OMOP project's large scale effort to evaluate surveillance methods based on empirical results. Systematic assessment of methods for postmarketing surveillance is a considerable task that has not been undertaken before, and results of these evaluations will be important in informing a future nation-wide active surveillance system.

In Chapter 3 we focused on extensions of the SCCS model in order to adapt it for the high-dimensional problem of postmarketing surveillance. First, we introduced the multiple SCCS model in which multiple drug exposures are incorporated as time-varying covariates. The multiple SCCS framework also allows drug interaction effects to be included, which is beneficial since patients often take multiple drugs concurrently. Second, we proposed taking a Bayesian approach to regularize the regression model. In our Bayesian multiple SCCS model, regularization restricts the size of the parameter estimates and prevents the difficulties of overfitting that are encoun-

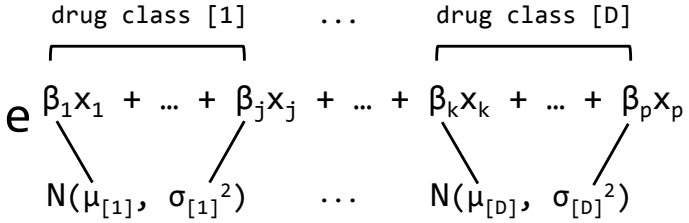
tered when using maximum likelihood for high-dimensional regression models. The Laplacian prior has the added benefit of inducing sparsity in the estimates so that covariates without an appreciable effect are selected out of the model. Lastly, we propose using the CCD algorithm for MAP estimation, which is an efficient approach for large problems and makes optimization tractable in high-dimensional settings. Our Bayesian multiple SCCS approach has demonstrated good performance for classifying drug-AE pairs into positive associations and negative controls based on our analyses of OMOP databases. Bayesian SCCS generally performs better than or as well as BLR; it is possible that these gains are due to AEs being modeled as recurrent outcomes or to the self-controlled (within-person comparison) aspects of SCCS. Both the Bayesian SCCS and BLR approaches tend to do better than methods based on marginal associations (simple SCCS, DP, and OS), which suggests that adjusting for potential confounders leads to improved performance. As noted in Section 3.5.2, our performance evaluation is subject to the limitations of the 53 drug-AE pair classification experiment, and further investigations may lead to greater insight regarding methods assessment.

There are many avenues for extensions and future work, and several approaches to modeling that can be built into the SCCS analysis in its current form. For instance, if it is believed that drug exposures may influence patients in different ways according to sex or age, one could include interaction effects between exposure and sex (or age) rather than having a common drug exposure parameter across all individuals. If certain drug effects are thought to accumulate over time, time-varying cumulative drug exposure covariates (e.g. that count the total number of days an individual has been exposed to the drug thus far) could also be directly incorporated. We may also want to adjust for the presence of certain health conditions in the patients in our analysis. For example, a condition may be indicative of deteriorating health over time, which would increase the frequency of AEs, drugs may have unusual effects on

people who have particular conditions, or patients may be more likely to experience AEs that are related to underlying health conditions, regardless of what drugs they are exposed to.

Another direction for future work is hierarchical modeling, which provides a way for us to incorporate knowledge about the relationship structure of both drug exposures and AE outcomes into our model. For example, drugs can be grouped into larger drug classes based on their chemical compounds. We would expect that drugs within a drug class would have exposure effects that are similar and that drugs in different classes would have effects that are unrelated. We can extend our Bayesian multiple SCCS approach in a natural way to build in hierarchical structure and borrow strength across drug classes. Common priors can be imposed on classes of drugs, rather than treating all drugs as *a priori* independent, which was our perspective in Chapter 3. Figure 5.1 shows a schematic of a possible hierarchical modeling approach for drug classes. Letting $d = 1, \ldots, D$ index drug classes, one strategy would be to place a Normal prior over the drug effects within a class such that $\beta_j \mid \mu_{[d]} \sim N(\mu_{[d]}, \sigma^2_{[d]})$ if drug j is in class d. This shrinks individual β_j's toward a common drug class effect $\mu_{[d]}$. A Normal hyperprior could then be placed on the drug class effects themselves, i.e. $\mu_{[d]} \sim N(0, \sigma^2_\mu)$, in order to shrink these class estimates toward zero.

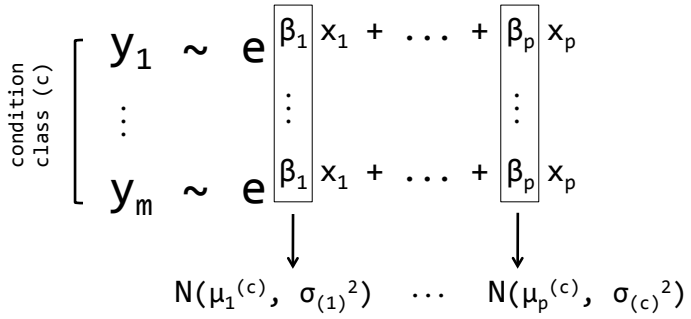
Figure 5.1: Hierarchical modeling approach with shrinkage of drug exposure effects β_j toward a common effect for the drug class $\mu_{[d]}$.



Following similar reasoning, we could group AE outcomes into larger classes based on which body systems are affected by the relevant health condition. For hierarchical

modeling of AE classes, one could put a common prior over exposure estimates for AE outcomes in a common class. This would tie together drug effect estimates across different AE outcomes, whereas previously we had focused on the case where there is a single outcome AE of interest. Figure 5.2 shows a schematic of a hierarchical modeling strategy based on classes of AEs. If $c = 1, \ldots, C$ indexes health conditions that correspond to AEs, one could place a prior on the jth drug exposure parameter so that $\beta_j \mid \mu_j^{(c)} \sim N(\mu_j^{(c)}, \sigma^2_{(c)})$ for all AEs that fall into condition class c. This would shrink individual drug j effects toward a common estimate for that drug within a condition class, and we could then impose a hyperprior $\mu_j^{(c)} \sim N(0, \sigma^2_\mu)$ on the condition class estimates for the jth drug exposure.

Figure 5.2: Hierarchical modeling with shrinkage of drug effect estimates toward a common drug effect for the relevant condition class.



We proposed the Bayesian multiple SCCS approach in Section 3.3 in order to deal with complications of high-dimensionality. In our approach inference is based on MAP estimates, but in a fully Bayesian analysis we would ideally obtain an estimate of the full posterior distribution via MCMC or another sampling procedure. In our high-dimensional surveillance problem such techniques are computationally prohibitive. In order to estimate the variability of and find confidence regions for our MAP estimates we may be able to rely on methods that approximate the posterior. For example, variational approximations are used for Bayesian analyses (Hinton and van Camp, 1993; Bishop, 2006) and the integrated nested Laplace approximations (INLA) introduced

in Rue and Martino (2009) provide another promising alternative. Further investigation is needed to determine how accurately these approximation-based methods can estimate posterior variability in our analysis.

We apply our Bayesian multiple SCCS analysis and CCD optimization using an efficient C++ implementation. To further reduce computation time, we are working with collaborators on an implementation that makes use of Graphics Processing Units (GPUs). These processors were originally developed to render high-end 3D graphics, but can be programmed for parallel processing at a fine scale. Each GPU contains multiple processor cores, which are used to apply a common operation to multiple elements of a data array. For procedures that fall into this *single instruction, multiple data* (SIMD) paradigm, this method of processing leads to substantial computational speed-ups (Suchard et al., 2010). CCD is an inherently sequential algorithm that is dependent on iterative updates. However GPU parallelization happens at the scale of the matrix operations that occur inside of these sequential iterations. The matrix operations involved in our algorithm lend themselves to the SIMD framework, and we have seen promising results with greater than an 80-fold speed-up using a basic implementation.

In our Bayesian multiple SCCS model, one needs to choose a value for the Normal σ_β^2 or Laplacian λ hyperparameters. One approach for setting this parameter is through cross-validation, which involves searching through a range of possible values and selecting the one that performs the best over disjoint random subsets of hold-out data. Cross-validation is a computationally costly procedure that would ordinarily be intractable for our problem. However due to its high speed, our GPU implementation makes cross-validation tractable even in the large-scale problem of surveillance. We are currently working on modifying our implementation so that results can be run on OMOP databases, which would allow us to determine whether selecting cross-validated hyperparameters improves method performance.

In Chapter 4 we presented the PD-SCCS model, which allows event risk to increase additively when an event occurs while maintaining the desirable properties that arise from conditioning in the SCCS model. This model circumvents assumption (i) for the SCCS model, which requires conditional independence between event occurrences. Our example analyses of the association between Vioxx and MI (Section 4.5.1) and PPIs and MI (Section 4.5.2) demonstrate that accounting for the presence of positive event dependence by using PD-SCCS rather than SCCS can lead to meaningful differences between relative risk estimates. It is thus important to consider whether positive event dependence may be present between event occurrences when choosing an analysis method. If no event dependence is present, the simulation results in Section 4.4.2 show that PD-SCCS will approximate SCCS by estimating δ to be close to zero. Our simulations in Section 4.4 indicate that one can select between PD-SCCS and SCCS models by using AIC or BIC, or by using the one-sided Wald test of $\delta = 0$ presented in Section 4.4.2.

In our PD-SCCS model development we assumed a common dependence parameter δ for all individuals. This assumption may be too restrictive for certain types of analyses, in which case δ can instead be permitted to vary with relevant fixed covariates such as sex or baseline health measures. Suppose there are L such covariates, and for person i the associated covariate vector is denoted by $\mathbf{z}_i = (z_{i1}, \ldots, z_{iL})^{\mathrm{T}}$. The dependence parameter can be written as a function of $\mathbf{z}_i$ in the form

$$\delta = \delta(\mathbf{z}_i) = e^{\boldsymbol{\psi}^{\mathrm{T}}\mathbf{z}_i},$$

for person i where $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_L)^{\mathrm{T}}$ is the vector of effects for $\mathbf{z}_i$. Exponentiation maintains the positivity of $\delta(\mathbf{z}_i)$ and results in $\mathbf{z}_i$ having a multiplicative effect on the dependence parameter. In order to make use of this more flexible form, the δ parameter may be replaced by $\delta(\mathbf{z}_i)$ in the PD-SCCS expressions for person i. If we use this $\delta(\mathbf{z}_i)$ form for dependence, the expressions that we have presented for the conditional likelihood in (4.9) and large sample inference presented in Section 4.3

carry through in a straightforward manner.

Future work extending our PD-SCCS model could investigate whether we can incorporate other types of event dependence and still maintain the advantages that arise from conditioning. For instance, there may be situations where the baseline risk is reduced after an event occurence (a negative event dependence model), the risk is only altered temporarily, or the effect on the risk decays over time.

The PD-SCCS model that we have presented circumvents conditional independence assumption (i) in a particular manner. As discussed in Section 2.2, assumption (ii) is also likely to be violated in practice. For example, if an individual experiences an MI or another traumatic health event on a particular day this may effect whether or not they fill their prescribed medications on subsequent days. Approaches that address this restriction within the framework of SCCS include Farrington et al. (2009); Kuhnert et al. (2010); Roy et al. (2006). Farrington et al. (2009) and Kuhnert et al. (2010) use methods based on consistent estimating equations. Roy et al. (2006) avoids the restriction by jointly modeling the distribution of exposures and AE outcomes. As mentioned previously, in econometrics literature assumption (ii) corresponds to strict exogeneity. Investigating how this is handled in FEP models may provide some insight into how we can account for this in our analysis.

The SCCS-based models that we have presented also make the implicit assumption that, for person i, the length τ_i of the observation period is determined independently of the event process. Under this assumption the distribution of observation lengths is ignorable, i.e. τ_i is conditionally independent of the event times given the covariate process, and thus we do not need to explicitly condition on or specify a distribution for τ_i in our model. Roy et al. (2006) and more recently Farrington et al. (2011) present methods for avoiding this assumption in the SCCS model. Non-random dropout can occur in the context of surveillance; some AEs may increase the risk of mortality following event, which would result in dependence between event times and

observation length τ_i. Incorporating non-random drop-out into our analysis is an area of future work.

Bibliography

Agresti, A. (2002). *Categorical Data Analysis*. New Jersey: John Wiley & Sons, Inc.

Bate, A. and Evans, S. J. (2009). Quantitative signal detection using spontaneous ADR reporting. *Pharmacoepidemiology and Drug Safety* **18,** 427–436.

Bate, A., Lindquist, M., Edwards, I. R., Olsson, S., Orre, R., Lansner, A., and De Freitas, R. M. (1998). A Bayesian neural network method for adverse drug reaction signal detection. *European Journal of Clinical Pharmacology* **54,** 315–321.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. New York: Springer.

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition* **30,** 1145–1159.

Bresalier, R. S., Sandler, R. S., Quan, H., Bolognese, J. A., Oxenius, B., Horgan, K., Lines, C., Riddell, R., Morton, D., Lanas, A., Konstam, M. A., and Baron, J. A. (2005). Cardiovascular events associated with rofecoxib in a colorectal adenoma chemoprevention trial. *New England Journal of Medicine* **352,** 1092–1102.

Cameron, A. C. and Trivedi, P. K. (1998). *Regression Analysis of Count Data*. New York: Cambridge University Press.

Cook, R. J. and Lawless, J. F. (2007). *The Statistical Analysis of Recurrent Events*. New York: Springer.

Cressie, N. A. (1993). *Statistics for Spatial Data*. John Wiley & Sons, Inc.

Dennis, J. E. and Schnabel, R. B. (1996). *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. New Jersey: Prentice-Hall, Inc.

DuMouchel, W. (1999). Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. *The American Statistician* **53,** 177–190.

Evans, S. J., Waller, P. C., and Davis, S. (2001). Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiology and Drug Safety* **10,** 483–486.

Farrington, C. P. (1995). Relative incidence estimation from case series for vaccine safety evaluation. *Biometrics* **51,** 228–235.

Farrington, C. P., Anaya-Izquierdo, K., Whitaker, H. J., Hocine, M. N., Douglas, I., and Smeeth, L. (2011). Self-controlled case series analysis with event-dependent observation periods. *Journal of the American Statistical Association.* DOI: 10.1198/jasa.2011.ap1010.

Farrington, C. P. and Hocine, M. N. (2010). Within-individual dependence in self-controlled case series models for recurrent events. *Journal of the Royal Statistical Society, Series C* **59,** 457–475.

Farrington, C. P. and Whitaker, H. J. (2006). Semiparametric analysis of case series data. *Applied Statistics* **53,** 553–594.

Farrington, C. P., Whitaker, H. J., and Hocine, M. N. (2009). Case series analysis for censored, perturbed or curtailed post-event exposures. *Biostatistics* **10,** 3–16.

FDA (2007a). Early communication about an ongoing safety review of omeprazole (Prilosec) and esomeprazole (Nexium). Retrieved from `http://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/ucm143258.htm`.

FDA (2007b). Exenatide and pancreatitis letter to healthcare professionals. Retrieved from `http://www.fda.gov/downloads/Safety/MedWatch/SafetyInformation/SafetyAlertsforHumanMedicalProducts/ucm126437.pdf`.

FDA (2007c). Food and Drug Administration Amendments Act of 2007. Pub. L. No. 110-85.

FDA (2008). FDA safety labeling changes to lapatinib tablets. Retrieved from `http://www.fda.gov/Safety/MedWatch/SafetyInformation/Safety-RelatedDrugLabelingChanges/ucm121837.htm`.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* **33,** 1–22.

Genkin, A., Lewis, D. D., and Madigan, D. (2007). Large-scale Bayesian logistic regression for text categorization. *Technometrics* **49,** 291–304.

Graham, D. J. (2004). Testimony on FDA, Merck and Vioxx: Putting patient safety first?: Hearing before the United States Senate Committee on Finance. (S. HRG. 108791). Text from `http://finance.senate.gov/imo/media/doc/111804dgtest.pdf`.

Hauben, M. and Hochberg, A. (2008). The importance of reporting negative findings in data mining: The example of exenatide and pancreatitis. *Pharmaceutical Medicine* **22,** 215–219.

Hauben, M., Madigan, D., Gerrits, C., and Meyboom, R. (2005). The role of data mining in pharmacovigilance. *Expert Opinion in Drug Safety* **4,** 929–948.

Hauben, M., Reich, L., van Puijenbrock, E., and et al. (2006). Data mining in pharmacovigilance: Lessons from phantom ships. *European Journal of Clinical Pharmacology* **62,** 967–970.

Hausman, J., Hall, B. H., and Griliches, Z. (1984). Econometric models for count data with an application to the patents-R&D relationship. *Econometrica* **52,** 909–938.

Hinton, G. E. and van Camp, D. (1993). Keeping the neural networks simple by minimizing the description length of the weights. In *Proc. 6th A. Conf. Computational Learning Theory*, pages 5–13.

Hocine, M., Guillemot, D., Tubert-Bitter, P., and Moreau, T. (2005). Testing independence between two Poisson-generated multinomial variables in case series and cohort studies. *Statistics in Medicine* **24,** 4035–4044.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **12,** 55–67.

Kuhnert, R., Hecker, H., Poethko-Müller, C., Schlaud, M., Vennemann, M., Whitaker, H. J., and Farrington, C. P. (2010). A modified self-controlled case series method to examine association between multidose vaccinations and death. *Statistics in Medicine* **30,** 666–677.

Maclure, M. (1991). The case-crossover design: A method for studying transient effects on the risk of acute events. *American Journal of Epidemiology* **133,** 144–153.

Madigan, D., Hauben, M., Vallarino, C., Patadia, V., Gerrits, C., and Simpson, S. E. (2011). The self-controlled case series method applied to a claims database for detection of signals in pharmacovigilance: A pilot study. Preprint.

Manson, J. E., Hsia, J., Johnson, K. C., Rossouw, J. E., Assaf, A. R., Lasser, N. L., Trevisan, M., Black, H. R., Heckbert, S. R., Detrano, R., Strickland, O. L., Wong, N. D., Crouse, J. R., Stein, E., and Cushman, M. (2003). Estrogen plus progestin and the risk of coronary heart disease. *New England Journal of Medicine* **349,** 523–534.

Moran, P. A. P. (1971). Maximum-likelihood estimation in non-standard conditions. *Proceedings of the Cambridge Philosophical Society* **70,** 441–450.

Navidi, W. (1998). Bidirectional case-crossover designs for exposures with time trends. *Biometrics* **54,** 596–605.

OMOP (2011). Observational Medical Outcomes Partnership, Health Outcomes of Interest library. Retrieved from `http://omop.fnih.org/HOI`.

Orre, R., Bate, A., Norn, G. N., Swahn, E., Arnborg, S., and Edwards, I. R. (2005). A Bayesian recurrent neural network for unsupervised pattern recognition in large incomplete data sets. *International Journal of Neural Systems* **15,** 207–222.

Platt, R., Wilson, M., Chan, K. A., Benner, J. S., Marchibroda, J., and McClellan, M. (2009). The new Sentinel network - Improving the evidence of medical-product safety. *The New England Journal of Medicine* **361,** 645–647.

Praus, M., Schindel, F., Fescharek, R., and Schwarz, S. (1993). Alert systems for postmarketing surveillance of adverse drug reactions. *Statistics in Medicine* **12,** 2382–2393.

Rothman, K. J., Lanes, S., and Sacks, S. T. (2004). The reporting odds ratio and its advantages over the proportional reporting ratio. *Pharmacoepidemiology and Drug Safety* **13,** 519–523.

Roy, J., Alderson, D., Hogan, J. W., and T., T. K. (2006). Conditional inference methods for incomplete Poisson data with endogenous time-varying covariates: Emergency department use among HIV-infected women. *Journal of the American Statistical Association* **101,** 424–434.

Rue, H. and Martino, S. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximation. *Journal of the Royal Statistical Society, Series B* **71,** 1–35.

Ryan, P. B., Madigan, D., Stang, P. E., Overhage, J. M., Racoosin, J. A., and Hartzema, A. G. (2011). Empirical assessment of analytic methods for risk identification in observational healthcare data: Results from the experiments of the Observational Medical Outcome Partnership. Preprint.

Saha, A. and Tewari, A. (2010). On the finite time convergence of cyclic coordinate descent methods. Preprint.

Simpson, S. E. (2011). A positive event dependence model for self-controlled case series with applications in postmarketing surveillance. Submitted.

Stang, P. E., Ryan, P. B., Racoosin, J. A., Overhage, J. A., Hartzema, A. G., Reich, C., Welebob, E., Scarnecchia, T., and Woodcock, J. (2010). Advancing the science for active surveillance: Rationale and design for the Observational Medical Outcomes Partnership. *Annals of Internal Medicine* **153,** 600–606.

Suchard, M. A., Holmes, C., and West, M. (2010). Some of the What?, Why?, How?, Who? and Where? of graphics processing unit computing for Bayesian analysis. *ISBA Bulletin* **17,** 12–17.

Suissa, S. (1995). The case-time-control design. *Epidemiology* **6,** 248–253.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B* **58,** 267–288.

Tseng, P. (2001). Convergence of a block coordinate descent method for nondifferentiable maximization. *Journal of Optimization Theory and Applications* **109,** 475–494.

van Puijenbroek, E. P., Egberts, A. C., Heerdink, E., and Leufkens, H. G. (2000). Detecting drug-drug interactions using a database for spontaneous adverse drug reactions: an example with diuretics and non-steroidal anti-inflammatory drugs. *European Journal of Clinical Pharmacology* **56,** 733–738.

Whitaker, H. J., Farrington, C. P., Spiessens, B., and Musonda, P. (2005). Tutorial in biostatistics: The self-controlled case series method. *Statistics in Medicine* **0,** 1–31.

Winkelmann, R. (2000). *Econometric Analysis of Count Data*. New York: Springer.

Wooldridge, J. M. (1997). Multiplicative panel data models without the strict exogeneity assumption. *Econometric Theory* **13,** 667–678.

Wooldridge, J. M. (1999). Distribution-free estimation of some nonlinear panel data models. *Journal of Econometrics* **90,** 77–97.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. MIT Press.

Wu, T. T. and Lange, K. (2008). Coordinate descent algorithms for lasso penalized regression. *The Annals of Applied Statistics* **2,** 224–244.

Zhang, T. and Oles, F. (2001). Text categorization based on regularized linear regression classifiers. *Information Retrieval* **4,** 5–31.

Zorych, I., Madigan, D., Ryan, P., and Bate, A. (2011). Disproportionality methods for pharmacovigilance in longitudinal observational databases. Submitted.

Appendix A

Concavity of the SCCS log-likelihood

The CCD optimization algorithm presented in Section 3.4 depends on the convexity of the objective function to be minimized. We can show that the multiple SCCS log-likelihood in (3.2) is concave, and thus the negated log posterior function is convex for both the Normal and Laplacian priors.

From (3.2), and dropping the *'s for convenience, the log-likelihood for multiple SCCS based on grouping exposures is

$$l^c = \sum_{i=1}^{N} \left[\sum_{g=1}^{G_i} \left(y_{ig}\, \mathbf{x}_{ig}^{\mathrm{T}} \boldsymbol{\beta} \right) - n_i \log \left(\sum_{g'=1}^{G_i} l_{ig'}\, e^{\mathbf{x}_{ig'}^{\mathrm{T}} \boldsymbol{\beta}} \right) \right].$$

The gradient of this log-likelihood is

$$\frac{\partial l^c}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{N} \sum_{g=1}^{G_i} \left[y_{ig} - n_i \left(\frac{l_{ig}\, e^{\mathbf{x}_{ig}^{\mathrm{T}} \boldsymbol{\beta}}}{\sum_{g'=1}^{G_i} l_{ig'}\, e^{\mathbf{x}_{ig'}^{\mathrm{T}} \boldsymbol{\beta}}} \right) \right] \mathbf{x}_{ig}.$$

The Hessian of the log-likelihood is

$$\mathbf{H} = \frac{\partial^2 l^c}{\partial \boldsymbol{\beta}\, \partial \boldsymbol{\beta}^{\mathrm{T}}} = \sum_{i=1}^{N} n_i \left[\left(\frac{\sum_{g=1}^{G_i} l_{ig} e^{\mathbf{x}_{ig}^{\mathrm{T}} \boldsymbol{\beta}} \mathbf{x}_{ig}}{\sum_{g'=1}^{G_i} l_{ig'}\, e^{\mathbf{x}_{ig'}^{\mathrm{T}} \boldsymbol{\beta}}} \right)^{\otimes 2} - \frac{\sum_{g'=1}^{G_i} l_{ig'}\, e^{\mathbf{x}_{ig'}^{\mathrm{T}} \boldsymbol{\beta}} \left(\mathbf{x}_{ig}^{\otimes 2} \right)}{\sum_{g''=1}^{G_i} l_{ig''}\, e^{\mathbf{x}_{ig''}^{\mathrm{T}} \boldsymbol{\beta}}} \right]$$

Let $\mathbf{X}_i^{\mathrm{T}} = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iG_i})_{(p \times G_i)}$

and $\mathbf{w}_i = \begin{bmatrix} l_{i1}\, e^{\mathbf{x}_{i1}^{\mathrm{T}}\boldsymbol{\beta}} \\ \vdots \\ l_{iG_i}\, e^{\mathbf{x}_{iG_i}^{\mathrm{T}}\boldsymbol{\beta}} \end{bmatrix}_{(G_i \times 1)}$

then $\mathbf{X}_i^{\mathrm{T}}\mathbf{w}_i = \sum_{g=1}^{G_i} l_{ig}\, e^{\mathbf{x}_{ig}^{\mathrm{T}}\boldsymbol{\beta}}\, \mathbf{x}_{ig}$ and $\mathbf{1}^{\mathrm{T}}\mathbf{w}_i = \sum_{g=1}^{G_i} l_{ik}\, e^{\mathbf{x}_{ig}^{\mathrm{T}}\boldsymbol{\beta}}$

where $\mathbf{1}$ is a $(G_i \times 1)$ column vector of ones. Then

$$\left(\frac{\sum_{g=1}^{G_i} l_{ig} e^{\mathbf{x}_{ig}^{\mathrm{T}}\boldsymbol{\beta}}\, \mathbf{x}_{ig}}{\sum_{g'=1}^{G_i} l_{ig'}\, e^{\mathbf{x}_{ig'}^{\mathrm{T}}\boldsymbol{\beta}}} \right)^{\otimes 2} = \frac{\left(\mathbf{X}_i^{\mathrm{T}}\mathbf{w}_i\right)\left(\mathbf{X}_i^{\mathrm{T}}\mathbf{w}_i\right)^{\mathrm{T}}}{(\mathbf{1}^{\mathrm{T}}\mathbf{w}_i)^2}$$

In order to verify that the Hessian of the log-likelihood is negative semidefinite, we need to show that for any $(p \times 1)$ vector $\mathbf{m} = (m_1, \ldots, m_p)^{\mathrm{T}} \neq \mathbf{0}$, $\mathbf{m}^{\mathrm{T}}\mathbf{H}\mathbf{m} \leq 0$.

The product $\mathbf{m}^{\mathrm{T}}\mathbf{H}\mathbf{m}$ can be written as

$$\mathbf{m}^{\mathrm{T}}\mathbf{H}\mathbf{m} = \sum_{i=1}^{N} \frac{n_i}{(\mathbf{1}^{\mathrm{T}}\mathbf{w}_i)^2} \left[\mathbf{m}^{\mathrm{T}} \left(\mathbf{X}_i^{\mathrm{T}}\mathbf{w}_i\right)\left(\mathbf{X}_i^{\mathrm{T}}\mathbf{w}_i\right)^{\mathrm{T}} \mathbf{m} - (\mathbf{1}^{\mathrm{T}}\mathbf{w}_i) \underbrace{\mathbf{m}^{\mathrm{T}} \left(\sum_{g=1}^{G_i} l_{ig}\, e^{\mathbf{x}_{ig}^{\mathrm{T}}\boldsymbol{\beta}} \mathbf{x}_{ig}\mathbf{x}_{ig}^{\mathrm{T}} \right) \mathbf{m}}_{=:\, t} \right]$$

Define two $(G_i \times 1)$ vectors $\mathbf{a}_i$ and $\mathbf{b}_i$, with the corresponding gth components

$$\begin{aligned} a_{ig} &= \sqrt{l_{ig}\, e^{\mathbf{x}_{ig}^{\mathrm{T}}\boldsymbol{\beta}}} \\ b_{ig} &= \sqrt{l_{ig}\, e^{\mathbf{x}_{ig}^{\mathrm{T}}\boldsymbol{\beta}}}\, \mathbf{m}^{\mathrm{T}}\mathbf{x}_{ig} \end{aligned}$$

Then the term t above can be written out as

$$\begin{aligned} t &= \sum_{q=1}^{p}\sum_{r=1}^{p} m_q\, m_r \sum_{g=1}^{G_i} l_{ig}\, e^{\mathbf{x}_{ig}^{\mathrm{T}}\boldsymbol{\beta}}\, x_{igq}\, x_{igr} \\ &= \sum_{g=1}^{G_i} l_{ig}\, e^{\mathbf{x}_{ig}^{\mathrm{T}}\boldsymbol{\beta}} \underbrace{\sum_{q=1}^{p} m_q\, x_{igq}}_{\mathbf{m}^{\mathrm{T}}\mathbf{x}_{ig}} \underbrace{\sum_{r=1}^{p} m_r\, x_{igr}}_{\mathbf{m}^{\mathrm{T}}\mathbf{x}_{ig}} \\ &= \mathbf{b}_i^{\mathrm{T}}\mathbf{b}_i \end{aligned}$$

Also note that

$$\begin{aligned} \mathbf{a}_i^{\mathrm{T}}\mathbf{b}_i &= \sum_{g=1}^{G_i} l_{ig}\, e^{\mathbf{x}_{ig}^{\mathrm{T}}\boldsymbol{\beta}}\, \mathbf{m}^{\mathrm{T}}\mathbf{x}_{ig} \\ &= \mathbf{m}^{\mathrm{T}}\mathbf{X}_i^{\mathrm{T}}\mathbf{w}_i \\ \mathbf{a}_i^{\mathrm{T}}\mathbf{a}_i &= \mathbf{1}^{\mathrm{T}}\mathbf{w}_i \end{aligned}$$

Thus,

$$\mathbf{m}^{\mathrm{T}}\mathbf{H}\mathbf{m} = \sum_{i=1}^{N} \frac{n_i}{(\mathbf{a}_i^{\mathrm{T}}\mathbf{a}_i)^2}\left[(\mathbf{a}_i^{\mathrm{T}}\mathbf{b}_i)^2 - (\mathbf{a}_i^{\mathrm{T}}\mathbf{a}_i)(\mathbf{b}_i^{\mathrm{T}}\mathbf{b}_i)\right]$$

The Cauchy-Schwarz inequality says that $(\mathbf{a}^{\mathrm{T}}\mathbf{a})(\mathbf{b}^{\mathrm{T}}\mathbf{b}) \geq (\mathbf{a}^{\mathrm{T}}\mathbf{b})^2$ for any vectors $\mathbf{a}$ and $\mathbf{b}$. From this inequality, we know that $(\mathbf{a}_i^{\mathrm{T}}\mathbf{b}_i)^2 - (\mathbf{a}_i^{\mathrm{T}}\mathbf{a}_i)(\mathbf{b}_i^{\mathrm{T}}\mathbf{b}_i) \leq 0$ for every i. We also know that $n_i/(\mathbf{a}_i^{\mathrm{T}}\mathbf{a}_i)^2 > 0$ for every i, so $\mathbf{m}^{\mathrm{T}}\mathbf{H}\mathbf{m} \leq 0$ and the Hessian $\mathbf{H}$ is negative semidefinite. This shows that the log-likelihood l^c is a concave function in the parameter vector $\boldsymbol{\beta}$.

Appendix B

Equivalence of SCCS and the conditional within-day logistic model

We start with the derivation of the within-day logistic (Bernoulli) model. As noted at the end of Section 2.2, the number of events y_{id} is either 0 or 1 on day (i, d), and the p drug exposure status indicators are given by $\mathbf{x}_{id} = (x_{id1}, \ldots, x_{idp})^{\mathrm{T}}$. The outcome y_{id} on day can be modeled as a logistic regression for day (i, d) which is specified as

$$\log\left[\frac{p(y_{id} = 1 \mid \mathbf{x}_{id})}{p(y_{id} = 0 \mid \mathbf{x}_{id})}\right] = \phi_i + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_{id}$$

where ϕ_i gives the baseline individual log-odds of an event for person i, and drug exposures x_{idj} act multiplicatively on the odds. It follows that the individual likelihood contribution for person i under this model is

$$p(\mathbf{y}_i \mid \mathbf{x}_{i1}, \ldots, \mathbf{x}_{i\tau_i}) = \prod_{d=1}^{\tau_i} \frac{(e^{\phi_i + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_{id}})^{y_{id}}}{1 + e^{\phi_i + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_{id}}} = \left[\prod_{d=1}^{\tau_i}(1 + e^{\phi_i + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_{id}})^{-1}\right] e^{\phi_i n_i} e^{\sum_{d'} y_{id'}\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_{id'}}$$

where $n_i = \sum_d y_{id}$. From the factorization theorem we can see that n_i is sufficient for ϕ_i, as it was in the Poisson-based SCCS model derivation in Section 2.2. By

conditioning on n_i, we can remove ϕ_i from the likelihood expression. This results in an individual conditional likelihood contribution for person i of the form

$$p(\mathbf{y}_i \mid \mathbf{x}_{i1}, \ldots, \mathbf{x}_{i\tau_i}, n_i) = \frac{p(\mathbf{y}_i \mid \mathbf{x}_{i1}, \ldots, \mathbf{x}_{i\tau_i})}{p(n_i \mid \mathbf{x}_{i1}, \ldots, \mathbf{x}_{i\tau_i})} = \frac{e^{\sum_d y_{id}\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_{id}}}{\sum_{\{\mathbf{y}_i|\sum_d y_{id}=n_i\}} e^{\sum_{d'} y_{id}\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_{id'}}}. \quad \text{(B.1)}$$

For the SCCS model, the Poisson likelihood contribution is

$$p(\mathbf{y}_i \mid \mathbf{x}_{i1}, \ldots, \mathbf{x}_{i\tau_i}) = \Big(\prod_{d=1}^{\tau_i} y_{id}!\Big)^{-1} e^{\phi_i n_i} e^{\sum_{d'} y_{id}\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_{id'}} e^{-e^{\phi_i}\sum_{d'} e^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_{id'}}}.$$

Since for each day $y_{id} = 1$ or 0, we have that $\big(\prod_{d=1}^{\tau_i} y_{id}!\big)^{-1} = 1$ and this term can be removed from the likelihood expression. Conditioning on the sufficient statistic n_i yields the conditional likelihood contribution for person i, which takes the form

$$p(\mathbf{y}_i \mid \mathbf{x}_{i1}, \ldots, \mathbf{x}_{i\tau_i}, n_i) = \frac{e^{\sum_d y_{id}\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_{id}}}{\sum_{\{\mathbf{y}_i|\sum_d y_{id}=n_i\}} e^{\sum_{d'} y_{id}\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_{id'}}}. \quad \text{(B.2)}$$

In this case the SCCS likelihood for person i in (B.2) is the same as the conditional logistic within-day likelihood for i given in (B.1), so the models are equivalent. In this setting we are discretizing time into days, where at most one event can occur per day. This is analogous to letting days be the smallest unit of time in which an event can occur in the derivation of the non-homogeneous Poisson process that SCCS is based on (e.g. see Cook and Lawless, 2007). Making this assumption is effectively equivalent to assuming that the process is Bernoulli within each day.

In our presentation of the SCCS model, the denominator in (B.2) was evaluated in closed form. To do this we can think of the process occurring in continuous time as in the derivation of SCCS in Section 4.2.2, and make use of the notation corresponding to that derivation. There the event rate was modeled as $\lambda_i(t \mid H_i(t)) = e^{\phi_i + \mathbf{x}_i(t)^{\mathrm{T}}\boldsymbol{\beta}}$, as given in equation (4.1). Thus the equivalent expression for the denominator term (i.e. the marginal density of n_i) in continuous time is

$$p(n_i) \propto \int_{\substack{\{\mathbf{t}_i \in (a_i,b_i]^{n_i}: \\ t_{i1}<\cdots<t_{in_i}\}}} \cdots \int \prod_{j=1}^{n_i} e^{\mathbf{x}_i(t_{ij})^{\mathrm{T}}\boldsymbol{\beta}} dt_{i1} \cdots dt_{in_i}.$$

Due to the symmetry of the integrand with respect to permutations of the t_{ij}'s (see the PD-SCCS model derivation in Section 4.2.3 for a detailed description), the integral can be evaluated in a closed form as

$$p(n_i) \propto \frac{1}{n_i!} \prod_{j=1}^{n_i} \int_{a_i}^{b_i} e^{\mathbf{x}_i(t_{ij})^{\mathrm{T}}\boldsymbol{\beta}} dt_{ij} = \frac{1}{n_i!} \left(\int_{a_i}^{b_i} e^{\mathbf{x}_i(t)^{\mathrm{T}}\boldsymbol{\beta}} dt \right)^{n_i}.$$

With this closed form expression for the denominator, the conditional likelihood expression matches the one given in (4.2) for SCCS in continuous time.