MCW Biostatistics Technical Report 64
Nonparametric recurrent events analysis with BART and an application to the hospital admissions of patients with diabetes

RODNEY A. SPARAPANI*, LISA E. REIN, SERGEY S. TARIMA,
TOURETTE A. JACKSON, JOHN R. MEURER
Institute for Health and Society, Medical College of Wisconsin,
8701 Watertown Plank Road, Milwaukee, Wisconsin 53226, USA
*rsparapa@mcw.edu

Summary
Much of survival analysis is concerned with absorbing events, i.e., events for which only a single event can be experienced. Here we focus on recurrent events: events which subjects are capable of experiencing multiple times. Recurrent events have been studied by many, however, most rely on the linear proportional intensity assumption. We propose a new method for recurrent events based on Bayesian Additive Regression Trees (BART) which does not employ restrictive assumptions such as linearity or proportionality. We explore this new method via a motivating example of hospital admissions for diabetes patients and simulated data sets.

Key words: Bayesian Additive Regression Trees; cumulative intensity; Decoupling Shrinkage and Selection (DSS); electronic health records (EHR); nonlinear; nonproportional; variable selection.

*To whom correspondence should be addressed.
1. Introduction

Survival analysis is typically concerned with so-called absorbing events, i.e., events for which only a single event can be experienced such as mortality. This article is focused on non-absorbing or recurrent events, i.e., events which subjects are capable of experiencing multiple times such as hospital admissions. We advocate using all of the event information rather than restricting to the first event, if any; although, the first event will seemingly produce a valid inference, but only at the great cost of lost power and estimation efficiency. Recurrent events survival analysis has been studied by many, however, much of the literature relies on the linear proportional intensity assumption (Kalbfleisch and Prentice, 1980; Andersen and Gill, 1982; Lawless, 1987; Wei and others, 1989; Hosmer Jr and others, 2008). In this article, we propose a new method for recurrent events that does not rely on restrictive assumptions such as linearity or proportionality. We explore this new method via a motivating example of hospital admissions for diabetes patients. Before describing our new method, we provide some background for our example.

Excess health care utilization is a cause of concern in the United States. For example, the World Bank estimates that health care expenditures in the US were approximately 17% of Gross Domestic Product in 2014 while no other developed country was above 12% (World Bank, 2015). Furthermore, the US health care system is prone to extremes: some patients are receiving too much care while others are not receiving enough (Fisher and others, 2009; Smith and others, 2013). This state of affairs has produced a series of reactions by the US federal government including the Affordable Care Act of 2010 (ACA) with provisions such as the Medicare Hospital Readmissions Reduction Program (Singh and others, 2014; Centers for Medicare and Medicaid Services, 2016; Zuckerman and others, 2016).

Our purpose in studying health care utilization is to determine if there are ways to reduce unneeded services via preventive measures. In this undertaking, we look at the entire record of health care utilization. We analyze all hospital admissions (including readmissions) and, in doing
so, we make the following distinction between prudent and excessive health care utilization. In this investigation, we are not disputing whether a given patient should have been admitted or not, i.e., admission practices are not assumed to have been inappropriate or unnecessary. Rather, we want to identify risk factors which foreshadow preventable or avoidable hospital admissions; these factors could be prudently managed via outpatient clinical care. Not taking advantage of these mitigation opportunities will most likely lead to excessive utilization such as future admissions which otherwise could have been avoided.

In this article, we focus on the Electronic Health Record (EHR) databank of health care utilization for a cohort of patients with diabetes in a unified health care system including a network of hospitals and affiliated outpatient clinics. EHR databank systems have been used by hospitals for decades. More recently, the American Recovery and Reinvestment Act of 2009 (ARRA) incentivized the usage of EHR systems in outpatient clinics and physician offices transforming it into a nearly complete health ledger for patients treated in a health system.

The modern EHR is an omnibus of current and historical digital health information. In addition, EHR data are often paired with billing information; therefore, we use the term EHR for both health and billing information taken together. We intend to estimate the risk of a hospital admission based on data routinely available in the EHR such as demographics (age, gender, race and ZIP code), body size (height, weight and body mass index), primary insurance provider (if any), billing, vital signs, comorbidities/complications, procedures/surgeries, and pharmaceutical prescriptions to treat diabetes; note that some of these covariates are static like race while some are dynamic and may change or be modified over time.

In Section 2, we discuss our new BART methodology for recurrent events and our motivating application. In Section D.3, we display the results of our motivating example. In Section D.6, we discuss our findings and conclusions. In Section E, we provide the details of free software that we created for nonparametric recurrent events analysis with BART.
2. Methods

Hospital admissions can be handled by survival analysis, specifically via recurrent events. Recurrent events follow many of the same traits of usual time-to-event data such as right censoring; the main difference being, rather than a patient potentially experiencing one absorbing event, each patient is free to experience multiple events during their observation period. The observation period of hospital admission recurrent events can be censored by the end of follow-up or from a death. A death may entail either independent censoring or dependent censoring; independent censoring is ignorable while dependent censoring can be very challenging to deal with (Cook and Lawless, 1997; Ghosh and Lin, 2000; Wang and others, 2001; Ghosh and Lin, 2003). When deaths are relatively rare (as is the case in our application), we assume independent censoring which seems like a reasonable compromise.

2.1 Bayesian Additive Regression Trees (BART)

In this article, we propose to model health care utilization with a new nonparametric recurrent events framework based on Bayesian Additive Regression Trees (BART) which extends earlier work on survival analysis with BART (Sparapani and others, 2016). BART is a Bayesian non-parametric, sum of trees, ensemble method with out-of-sample predictive performance generally equivalent to, or exceeding that, of Lasso, Gradient boosting, Neural nets with one hidden layer and Random forests (Chipman and others, 2010). Over-fitting is the tendency to overly fit a model to the particular given data set at the expense of the predictive performance on another data set. Typically, BART does not over-fit to the data at hand due to the regularization tree-branching penalty built into the prior, i.e., generally, each tree has few branches and plays a small part in the overall fit yielding desirable properties. Essentially, BART is a Bayesian non-linear model with all the advantages of the Bayesian paradigm such as posterior inference including point and interval estimation. Conveniently, BART naturally scales to large numbers of covariates and variable selection; it does not require the covariates to be scaled; nor does it require the
covariate functional relationship, nor the interactions considered, to be pre-specified.

For the purpose of this article, the intimate details of how BART works are not relevant and are best described elsewhere (Chipman and others, 2010); only a brief introduction follows. The BART prior provides a flexible class of unknown functions of covariates from which we can gather randomly generated fits to the given data via the posterior. For example, for a continuous outcome, $y$, we have the following BART regression, on the vector of covariates, $x$: $y = f(x) + \epsilon$ where $\epsilon \sim iid N(\mu,\sigma^2)$; the unknown random function, $f$, and the error variance, $\sigma^2$, follow the BART prior expressed notationally as $(f,\sigma^2) \sim BART$. For a dichotomous outcome, which is the form most relevant to this discussion, the result is: $y \sim iid B(\Phi(f(x)))$ where $f \sim BART$, $B(.)$ is the Bernoulli distribution and $\Phi(.)$ is the standard Normal cumulative distribution function (in this case, the error variance is fixed: $\sigma^2 = 1$).

### 2.2 Recurrent events and BART

Due to the capabilities of BART, we have great flexibility in modeling the dependence of recurrent events on covariates. To elaborate, consider data in the form: $\delta_i, t_i, u_i, x_i(t)$ where $i = 1, \ldots, n$ indexes subjects, $\delta_i$ is a count of the number of events during the observation period, $t_i = [t_{0i}, \ldots, t_{\delta_i}]'$ and $t_{hi}$ is the event start time of the $h$th event ($h = 0$ for the end of follow-up which is typically right-censored), $u_i = [u_{0i}, \ldots, u_{\delta_i}]'$ and $u_{hi}$ is the event end time of the $h$th event and $x_i(t)$ is a vector of time-dependent covariates. In order to address non-instantaneous events such as hospital admissions, we need event end times, $u_{hi}$, for each corresponding start time, $t_{hi}$ (for convenience suppose that $u_{0i} = 0$); i.e., a subject is not at risk for a new hospital admission event until they are discharged from their current admission event. For instantaneous events (or roughly instantaneous events such as emergency department visits with time measured in days), the end times are assumed to be equal to the start times: $u_{hi} = t_{hi}$ for $h = 1, \ldots, \delta_i$.

We denote the $J$ collectively distinct event start and end times for all subjects by $0 < t_{(1)} < \cdots < t_{(J)} < \infty$ thus taking $t_{(j)}$ to be the $j^{th}$ order statistic among distinct observation times and,
for convenience, \( t_{(j')} = 0 \) where \( j' \leq 0 \) (note that \( t_{(j)} \) are constructed from all event start/end times, but they may be a censoring time for any given subject). Now consider binary event indicators \( y_{ij} \) for each subject \( i \) at each distinct time \( t_{(j)} \) up to the subject’s last observation time \( t_{(n_i)} \leq t_{0i} \) with \( n_i = \arg \max_j [t_{(j)} \leq t_{0i}] \), i.e., \( y_{i1}, \ldots, y_{in_i} \in \{0, 1\} \). We then denote by \( p_{ij} \) the probability of an event at time \( t_{(j)} \) conditional on \( (t_{(j)}, N_i(t_{(j)}-), v_i(t_{(j)}), x_i(t_{(j)})) \) where \( N_i(t-) \) is the counting process of events for subject \( i \) just prior to time \( t \) and \( v_i(t) = t - u_{N_i(t-)} \) is the sojourn time for subject \( i \), i.e., time since last event, if any. For simplicity of presentation, we let

\[
\tilde{x}_i(t_{(j)}) = (N_i(t_{(j-1)}), v_i(t_{(j)}), x_i(t_{(j)})) \quad \text{(notice that we have replaced } N_i(t_{(j)}) - \text{ with } N_i(t_{(j-1)}) \text{ since, by design, the state of information available at time } t_{(j)} - \text{ is the same as that available at } t_{(j-1)})
\]

Assuming a constant intensity and constant covariates, \( \tilde{x}_i(t_{(j)}) \), in the interval \((t_{(j-1)}, t_{(j)}]\) allows us to define the cumulative intensity process as:

\[
\Lambda(t_{(j)}, \tilde{x}_i(t_{(j)})) = \int_0^{t_{(j)}} d\Lambda(t, \tilde{x}_i(t)) = \sum_{j'=1}^{j} \int_{t_{(j'-1)}}^{t_{(j')}} \Pr(dN_i(t) = 1 \mid t, \tilde{x}_i(t))
\]

\[
= \sum_{j'=1}^{j} \Pr(N_i(t_{(j'-1)}) - N_i(t_{(j'-1)}) = 1 \mid t_{(j')}, \tilde{x}_i(t_{(j')})) = \sum_{j'=1}^{j} p_{ij'}
\]

(the terms “intensity” and “hazard” are generally interchangeable).

Conditioning on the number of events just prior to the current time, \( N_i(t-) \), is not the same as conditioning on the entire event history, \( N_i(s) \) where \( 0 \leq s \leq t- \), which would be the most general, or least stringent, model. However, conditioning on the entire event history is often problematic as it is in this case. Conditioning on \( N_i(t-) \) creates a very general model and it is a reasonable compromise (Kalbfleisch and Prentice, 2002). Whether to model conditioning on time, \( t \), or the most recent sojourn time, \( v_i(t) \), is debatable. We avoid the debate by including both and we rely on BART to determine which is the most informative.

We now write the model for \( y_{ij} \) as a nonparametric probit regression of \( y_{ij} \) on \((t_{(j)}, \tilde{x}_i(t_{(j)}))\) akin to parametric models of the discrete hazard (Thompson Jr., 1977; Arjas and Haara, 1987; Czado, 1992; Fahrmeir and Tutz, 1994). Specifically, with temporal data converted from \( \delta_i, t_i, u_i, x_i(t) \)
to a sequence of longitudinal binary events as follows.
\[ y_{ij} = \sum_{h=1}^{\delta_i} I(t_{hi} = t_{(j)}) , \quad j = 1, \ldots, n_i \] (2.1)

This definition of \( y_{ij} \) is correct; however, there is a caveat in translating this into the likelihood (the likelihood is presented below in equation (2.4)) which we address by a vignette. Suppose that we have two subjects with the following values:

\[ \delta_1 = 1, t_1 = (10, 3)', u_1 = (0, 7)' \Rightarrow y_{11} = 1, y_{12} = 0, y_{13} = 0 \] (2.2)

\[ \delta_2 = 1, t_2 = (10, 4)', u_2 = (0, 7)' \Rightarrow y_{21} = 0, y_{22} = 1, y_{23} = 0 \]

which implies \( t_{(1)} = 3, t_{(2)} = 4, t_{(3)} = 7 \). Notice that \( y_{12} = 0 \) as it should be since no event occurred for subject 1 at time \( t = t_{(2)} = 4 \). However, no event occurred because this subject’s first event had not ended yet; therefore, we call this a forced zero since this subject was not currently at risk for an event. These forced zeros do not contribute to the likelihood since there is no random behavior to account for, i.e., a forced zero, \( y_{ij} = 0 \), implies \( p_{ij} = 0 \). Therefore, the only \( y_{ij} \) that contribute to the likelihood are those corresponding to \( j \) which are in the at-risk set, \( j \in R_i(t_{(j)}) \), where either the at-risk set for subject \( i \) at time \( t_{(j)} \) contains \( j \) or is empty. For notational convenience from here on, we write \( j \in R_i \) suppressing the relationship with time, but note that it is implicit. For simplicity, rather than the directly defining the at-risk set, we define the complement of the at-risk set as \( R_i^C = \{ j : t_{hi} < t_{(j)} < u_{hi}, \ h = 1, \ldots, \delta_i \} \) and we denote the binary outcomes for subject \( i \) by the longitudinally ordered vector \( y_i = [y_{ij}] \) where \( j \in R_i \).

We carry out the probit regression via truncated Normal latent variables \( z_{ij} \) to reduce it to a continuous outcome BART model like so (Albert and Chib, 1993; Chipman and others, 2010).

\[
\begin{align*}
y_{ij} | p_{ij} & \sim \mathcal{B}(p_{ij}) \quad \text{where } i = 1, \ldots, n; j = 1, \ldots, n_i \\
p_{ij} | f & \sim \begin{cases} 
\Phi(\mu_{ij}), \mu_{ij} = \mu_0 + f(t_{(j)}, \tilde{x}_i(t_{(j)})) & \text{if } j \in R_i \\
0 & \text{otherwise} 
\end{cases} \\
f & \sim \text{BART} \\
z_{ij} | y_{ij}, f & \sim \begin{cases} 
\mathcal{N}(\mu_{ij}, 1) I(-\infty, 0) & \text{if } y_{ij} = 0 \\
\mathcal{N}(\mu_{ij}, 1) I(0, \infty) & \text{if } y_{ij} = 1 
\end{cases}
\end{align*}
\] (2.3)
This model, as displayed in (2.3), corresponds to $\mathbf{y}$ (the entire collection of $y_i$’s) given $\mathbf{p}$ (the entire collection of $p_{ij}$’s). Consequently, we have the following Bernoulli likelihood.

$$
[y|p] = \prod_{i=1}^{n} \prod_{j \in R_i} p_{ij}^{y_{ij}} (1 - p_{ij})^{1-y_{ij}}
$$

(2.4)

For binary data, $\mu_0 = \Phi^{-1}(p_0)$ can be used for centering the latent $z$’s around the probability of an event $p_0$ where $\Phi^{-1}(.)$ is the inverse standard Normal cumulative distribution function.

For recurrent event data, we can similarly center the latent $z$’s by assuming the times of recurrent events follow an exponential distribution and the covariates, $\tilde{x}$, have no impact, i.e.,

$$
\mu_0 = \Phi^{-1}\left(1 - \exp\left(\sum_i \delta_i \frac{1}{\sum_i t_i}\right)\right).
$$

The model just described can be readily estimated using existing software for binary BART such as the R package dbarts (Dorie and others, 2016). It allows one to estimate the functions $f(t, \tilde{x}(t))$ and $p(t, \tilde{x}(t)) = \Phi(\mu_0 + f(t, \tilde{x}(t)))$.

### 2.3 Data construction

As we have seen, we need to transform the temporal data $(\delta, t, u)$ to longitudinal data suitable for the BART model (2.3). While the description of this is contained in equation (2.1) and the definitions preceding it, for additional clarification we provide a very simple example of a data set based on the vignette with two observations (note that the forced zero is absent as described in the vignette; see (2.2)): $[\delta_1 = 1, t_1 = (10, 3), u_1 = (0, 7)]$ and $[\delta_2 = 1, t_2 = (10, 4), u_2 = (0, 7)]$ where $t_{(1)} = 3, t_{(2)} = 4, t_{(3)} = 7$. Putting these together leads to

$$
\mathbf{y}' = (y_{11}, y_{13}, y_{21}, y_{22}, y_{23}) = (1, 0, 0, 1, 0)
$$

$$
\mathbf{t}' = (t_{(1)}, t_{(3)}, t_{(1)}, t_{(2)}, t_{(3)}) = (3, 7, 3, 4, 7)
$$

where $\mathbf{y}$ is the binary response vector and $\mathbf{t}$ makes up the first column of the matrix of covariates (of course, the remaining columns contain $\tilde{x}_i(t)$).

### 2.4 Targets for inference

With the data prepared as described above, the BART model for binary data treats the probability of an event in an interval as a nonparametric function of the time $t$ and the covariates.
\(x(t)\). Conditioned on the data, our R package BART (see Section E) provides samples from the posterior distribution of \(f\). For any \(t\) and \(\tilde{x}(t)\), we can obtain the posterior distribution of 
\[
p(t, \tilde{x}(t)) = \Phi(\mu_0 + f(t, \tilde{x}(t)))
\]
For the purposes of recurrent events survival analysis, we are typically interested in estimating the cumulative intensity function, so, we write the following expression to estimate this function at any time \(t\).

\[
\hat{\Lambda}(t, \tilde{x}(t)) = \sum_{j=1}^{J} \hat{Pr}(N_i(t(j)) - N_i(t(j-1)) = 1|t(j), \tilde{x}_i(t(j))) \Delta_j(t)
\]
where \(\Delta_j(t) = \min \left[ 1, \frac{t - t(j-1)}{t(j) - t(j-1)} \right] I(t(j-1) < t)\)
\[
= \sum_{j=1}^{J} \hat{p}(t(j), \tilde{x}_i(t(j)))\Delta_j(t)
\]
With these estimates, one can accomplish inference from the posterior via means, quantiles or other functions of \(p(t, \tilde{x}_i(t))\) or \(\Lambda(t, \tilde{x}(t))\) as needed such as the relative intensity, i.e.,
\[
RI(t, \tilde{x}_n(t), \tilde{x}_d(t)) = \frac{p(t, \tilde{x}_n(t))}{p(t, \tilde{x}_d(t))}
\]
where \(\tilde{x}_n(t)\) and \(\tilde{x}_d(t)\) are two settings we wish to compare like two treatments.

2.5 Marginal effects

The model, (2.3), does not directly provide a summary of the effect of a single covariate or a subset of covariates. In general, this is true for nonparametric regression models. We follow (Chipman and others, 2010) and use Friedman’s partial dependence function (Friedman, 2001) with BART to summarize the marginal effect due to a subset of the covariates, \(x_S\), by aggregating over the complement covariates, \(x_C\), i.e., \(x' = [x'_S, x'_C]\). The marginal dependence function is defined as
\[
f_S(x_S) = n^{-1} \sum_{i=1}^{n} f(x_S, x_{iC})\text{ via aggregating over the settings of the complement in the cohort.}
\]
For example, consider the cumulative intensity function: \(\Lambda_S(t, x_S) = n^{-1} \sum_{i=1}^{n} \Lambda(t, x_S, x_{iC})\).
Other marginal functions can be obtained in a similar fashion. Estimates can be derived via functions of the samples from the posterior such as means, quantiles, etc.
2.6 Variable selection

Various methods for variable selection have been proposed for BART (Chipman and others, 2010; Bleich and others, 2014; McCulloch and others, 2015). We will utilize the computationally friendly approach of Decoupling Shrinkage and Selection (DSS) which was proposed by Hahn and Carvalho (2015) for Bayesian linear models and extended to BART by McCulloch and others (2015). DSS attempts to find the subset of variables that best describe the mean function, \( f(\tilde{x}) \), rather than \( y \). Step 1 consists of fitting \( y \) by \( \tilde{x} \) with BART as proposed (resulting in \( f(t, \tilde{x}) \) in our notation). Step 2 fits \( f(t, \tilde{x}) \) by \( \tilde{x} \) via a non-linear method of choice. Since over-fitting is not an issue in Step 2, we are free to select any practical non-linear method; preferably a non-demanding computational method since we will be repeating the fit many times with different subsets and choosing the best subset, \( \tilde{x}_S \), via some criteria. In this instance, we choose the method of Classification and Regression Trees (CART) and the criteria of \( R^2 \) following the lead of McCulloch and others (2015). Step 3 is optional; we may choose to re-fit \( y \) by \( \tilde{x}_S \) with BART arriving at \( f(t, \tilde{x}_S) \).

2.7 Handling missing data with BART

BART can handle missing data (Kapelner and Bleich, 2016; Xu and others, 2016). We utilize the missing data framework developed by Xu and others (2016) which they call Sequential BART. Sequential BART assumes that the missing covariates are missing at random, i.e., missingness only depends on what has been observed. Specifically, Sequential BART assumes that a missing covariate can be imputed by BART from the rest of the covariates, and so on, sequentially for all missing covariates. A brief description of this method as employed in this study can be found in Appendix C of the Supplementary Materials.

2.8 Motivating example

We explored the hospital admissions for a cohort of patients with diabetes cared for by the Froedtert and Medical College of Wisconsin health network which is comprised of three hospitals.
and 16 primary care sites. These patients were identified via their EHR. This human subjects research was approved by the Medical College of Wisconsin and Froedtert Hospital joint Institutional Review Board. To maintain patient privacy, roughly one fourth of patients were randomly sampled for inclusion as well as other de-identification procedures.

We identified likely incident diabetes mellitus type 2 patients by tabulating their first diagnosis code of primary diabetes (ICD-9 codes 250.x0 and 250.x2) in 2006 or 2007, i.e., no such codes were found for these patients prior to 2006 for as far back as each patient’s records go which is variable. ICD-9 codes 250.x0 and 250.x2 are not specific for type 2 and we suspect that there are some type 1 patients among them; based on the diabetes population at large we assume that 80%-90% of our cohort are type 2. We restricted the population to adults aged 21 to 90 by 01/01/2008. Among the patients treated in this health system, the vast majority were racially self-identified as either white or black so our inclusion criteria is restricted to these groups. Since the health care data is de-identified, we only have the first 3 digits of the ZIP code (ZIP3) for their current address circa 2013. Most of the patients treated in this health care system reside in central or northern Milwaukee county, 532xx; or suburban communities in adjacent western and northwestern counties, 530xx (Waukesha county and Washington county respectively); therefore, we restricted the population to these ZIP3 areas.

We want to identify those who are primarily cared for by this health system (rather than those who may have been admitted to one of the system hospitals, but were clinically treated elsewhere) so that the EHR would roughly include their entire health care experience for a five-year window from 2008 to 2012. Therefore, we required that they have an Evaluation and Management (E&M) physician office visit as a new or established patient in 2006 or 2007; as well as a later visit in 2012 or 2013; for those who passed away before 12/31/2012, their second E&M visit had to be within one year prior to death. Since our interest is in patients with primary diabetes, we excluded those patients who were diagnosed with either secondary diabetes or gestational diabetes.
For this cohort, we identified every hospital admission between 01/01/2008 and 12/31/2012. For convenience, follow-up begins on 01/01/2008, rather than from each patient’s actual incident diagnosis date which varied over the preceding 2 years. Following all patients concurrently allows us to temporally adapt, via our model, for seasonal/epidemic hospital admissions such as the H1N1 influenza outbreak in the US from April to June 2009 (Jain and Kamimoto, 2009).

We investigated the following risk factors: gender, race, age, ZIP3 area, insurance status (commercial, government or other), diabetes therapy (insulin, metformin and/or sulfonylurea), health care charges, relative value units (RVU) (Federal Register, 2010), vital signs, laboratory values, comorbidity/complication diagnoses and procedures/surgeries (we will refer to vital signs and laboratory values collectively as signs; and comorbidity/complication diagnoses and procedures/surgeries collectively as conditions). In total, we considered 85 covariates of which 82 are external factors as described above and three are temporal factors: time, $t$, the counting process, $N_i(t-)$, and the sojourn time, $v_i(t)$. Among these potential predictors only gender, race and ZIP3 area are time-independent. Age was calculated as a function of time. The rest are defined as last value carried forward.

For insulin, metformin and sulfonylurea, we only had access to prescription orders (rather than prescription fills) and self-reported current status of prescription therapy during clinic office visits. Since, generally, orders are only required after every three fills, and each fill can be for up to 90 days, we define insulin, metformin and sulfonylurea as binary indicators which are one if there exists an order or current status indication within the prior 270 days; otherwise zero.

Health care charges and relative value units (RVU) are measures related to the services and procedures delivered. However, they are so closely related that recent charges/RVUs are of no practical value in this analysis. Since we are interested in preventive opportunities, we ignore recent charges/RVUs. Alternatively, we investigate charges/RVUs that are the sum total of the following moving windows of days prior to any given date: 31 to 90, 91 to 180, 181 to 300.
For many patients, some signs were not available for a given date so they were set to missing; similarly, if a sign was not observed within the last 180 days, then it was set to missing (except for height which never expires, weight extended to 365 days and body mass index which is a deterministic function of the two). We utilized the Sequential BART missing imputation method as described in Appendix C of the Supplementary Materials. However, instead of creating several imputed data sets, we imputed a new sign at each date when it was missing, i.e., in order to properly address uncertainty within one data set, a new value was imputed for each date that it was missing and never carried forward.

Conditions are binary indicators which are zero until the date of the first coding and then they are one from then on. Based on clinical rationale, we identified 26 conditions (23 comorbidities and 3 procedures/surgeries) which are potential risk factors for a hospital admission many of which are possible complications of diabetes; besides clinical merit, these conditions were chosen since they are present in more than just a few subjects so that they may be informative (see Table 3 which includes the diagnosis codes utilized). Similarly, we employed 15 general conditions which are the Charlson diagnoses (Charlson and others, 1987; Quan and others, 2005) (see Table 4 which includes the diagnosis codes utilized) and 18 general conditions from the RxRisk adult diagnoses which are defined by prescription orders (Fishman and others, 2003; Johnson and others, 2006) (see Table 5 which includes the pharmaceutical class codes utilized). Seven conditions are a composite of diagnosis codes and prescription orders: these codes are denoted by an asterisk in Tables 3, 4 and 5.

Notice that the conditions in Table 3 and Table 4 are not independent; for example, renal disease is a superset of chronic kidney disease which is also a superset of kidney failure. These hierarchical definitions are intentional since we would like to identify the narrowest risk factor definition wherever possible. Also, note, the following conditions are mutually exclusive so necessarily dependent: mild liver disease and moderate/severe liver disease; and malignancy and
metastatic solid tumor.

3. Results

Here we present the results of our study. First, we summarize the comparison of our new BART method with the counting process Cox model for recurrent events based on simulated data sets with known cumulative intensity (the full details are contained in Appendix D of the Supplementary Materials). Then, we apply our new method to the motivating example of hospital admissions among newly diagnosed diabetes patients.

3.1 BART vs. Cox Comparison Summary

Comparisons between our new BART method and the counting process Cox model are described in Appendix D of the Supplementary Materials. The results of these comparisons based on simulated data sets are as we anticipated. We assumed that the BART (Cox) model would generally outperform the Cox (BART) model for the nonproportional (proportional) setting. The reasons for these results are obvious. In the proportional setting, the Cox model has an advantage since it assumes linear proportionality. In the nonproportional setting, the BART model has an advantage since it does not assume linear proportionality.

In practice with real recurrent events data, we advocate for an approach with fewer assumptions like BART. BART does not require you to pre-specify the functional form and/or interactions, instead, it learns the functional form of the cumulative intensity from the data with respect to covariates, time, sojourn time, previous events, etc. On the other hand, Cox models require you to assume linear proportionality while pre-specifying interactions and your choice of model including the functional form of the cumulative intensity with respect to time, sojourn time, previous events, etc. (common Cox model choices are discussed in Appendix B of the Supplementary Materials). Of course, you can test your Cox model assumptions and choose alternative models based on those diagnostics; however, this process is time-consuming, repetitive, non-trivial, prone to error and often requires further assumptions. These realistic simulations lend credence to the
contention that our new BART model can reliably analyze recurrent events data without adopting restrictive assumptions.

3.2 Motivating example results

Utilizing all 85 temporal and external covariates is a challenge for traditional methods, but is rather routine for BART. Nevertheless, it is difficult to tabulate information on all 82 external covariates in a convenient manner for the readers of this article. Therefore, we mainly restrict our attention to the four covariates that were deemed to be the most important by variable selection with BART and DSS (see Section 3.2.2): insulin therapy, serum calcium, peripheral vascular disease (PVD) and peptic ulcer disease (PUD). We tabulate these important covariates and a select few others in Table 7.

3.2.1 Aggregate description We describe the Table 7 summary as follows: in the first (second) column of numbers, we have summarized the covariates as they appear at the beginning (end) of follow-up; and in the third column, as they appear at the beginning of hospital admissions. For the fourth column, we summarize via a simple unadjusted counting process Cox model of time (we describe this model in Appendix B of the Supplementary Materials) by providing relative intensity and its corresponding 95% confidence interval. As we have previously discussed, the counting process Cox model has the disadvantage of assuming proportionality, it is only one of several possible Cox model choices, etc. However, there are also some advantages. It is convenient to compute and we can calculate the counting process Cox model on the same grid that we employ for BART which allows us to use the same data including the same time-dependent covariates like age, insulin therapy, etc. And, we can easily incorporate our forced zero correction, as shown for BART via the vignette (see 2.2).

The first thing to notice is the relatively high admission rate: 525 admissions for 488 patients in roughly five years of follow-up. In this cohort, there are more women than men, roughly a
56:44 split and a similar hospital admission rate with a relative intensity near 1. There are about
twice as many whites as blacks, however, the hospital admission rate is about 50:50 with whites
under- and blacks over-represented and a significant unadjusted relative intensity of 1.8.

Among the oldest members of the population, there is a non-significant trend towards more
hospital admissions. About 78% of this cohort resided in Milwaukee county which is poorer
than the surrounding suburbs where about 22% resided. Milwaukee county had a higher hospital
admission rate than the suburbs with a significant unadjusted relative intensity of 1.9.

Virtually everyone in the cohort aged 65 and older was enrolled in the Medicare insurance
program and that increased throughout the study from roughly 29% to 39%. About 28% of the
cohort was under 65 and had some form of public health insurance which consisted mainly of
Medicaid for the poor and disabled. Over the course of the study, commercial health insurance
coverage of those under 65 dropped from roughly 39% to 29%. Public health insurance for those
under 65 had the highest rate of hospital admissions, but only commercial insurance for those
under 65 had a statistically significant unadjusted relative intensity: a relative intensity of 0.3 for
fewer hospital admissions.

During the course of the study, insulin therapy increased slightly. Insulin recipients were much
more likely to have a hospital admission than those not taking insulin; a statistically significant
unadjusted relative intensity of 4.6. Metformin therapy was constant: 50 patients added Met-
formin and 50 patients dropped it resulting in the same rate. Patients taking metformin had a
non-statistically significant relative intensity of 0.8 as opposed to those not taking metformin.

Glycohemoglobin $\text{A}_1c$ levels $< 9\%$ vs. $\geq 9\%$ did not produce a statistically significant
relative intensity. Serum calcium levels $\leq 9.1\text{mg/dL}$ vs. $> 9.1\text{mg/dL}$ produced a statistically significant
unadjusted relative intensity of 7.3. Peripheral vascular disease (PVD) and peptic ulcer disease
(PUD) both produced a statistically significant increase in hospital admissions with relative
intensities of 16.2 and 7.5 respectively.
3.2.2 Model-based analysis with BART  In order to assess the degree of over-fitting, we randomly divided our data set into two parts: a training cohort with 235 patients and a validation cohort with 253 patients. We performed variable selection via DSS. In Step 1, we fit a full model including 82 external covariates and 3 temporal variables with BART. In addition to the four previously mentioned external covariates, DSS also identified $N_i(t-)$ which is the number of prior hospital admissions. So, we will restrict ourselves to assessing the effects of just the important covariates with the comforting reassurance that the analysis is adjusted for the rest.

Since we have fit this full model to the data, now we use our model to understand the data via the model’s consequences. In order to evaluate the effects of the important covariates, we estimate our function of interest at chosen settings, $x_S$, while aggregating over the other covariates, $x_C$ (see Section 2.5 for a complete description). For a binary covariate like insulin, the chosen settings are easy to enumerate: 1 for being treated and 0 otherwise. However, for $N_i(t-)$ and serum calcium, it is not so simple.

For $N_i(t-)$, it is challenging since $N_i(t-)$, $t$ and $v_i(t)$ are intimately related. Therefore, we created 6 profiles where $N_i(t-)$ has the potential to increment at $t$ nearest the time points of 6, 12, ..., 54 months. Table 8 shows the value of $N_i(t-)$ for each of the 6 profiles (note that $v_i(t)$ is also set in a corresponding manner). In Figures 1 and 2, we can see that previous hospital admissions substantially increase the chance of new hospital admissions. In Figure 1, as the number of previous hospital admissions increase, the cumulative intensity function for a new hospital admission increases respectively. Let’s consider three profiles: 1, the lowest risk group; 6, the highest risk group; and 3, which is about midway in between. In Figure 2, we show the relative intensities for profiles 3 to 1 (black) and 6 to 1 (gray). For 3 to 1, from 24 to 60 months, the relative intensity fluctuates around 1.3 (the relative intensity is one until 24 months since both profiles are identical) and the 95% credible interval goes from 1 to fluctuating around 2.5. For 6 to 1, from 6 to 36 months, the relative intensity rises to around 1.5, flattens and then
attenuates at 48 months; the 95% credible interval is much wider than for 3 to 1 since this is a more extreme setting: only 4 patients in the training cohort reached 9 hospital admissions hence the wider interval.

For serum calcium, we created 20 profiles: 8.4, 8.5, ..., 10.3 (mg/dL); to explore its effect, for each profile we fixed it at these values for all 60 months. In Figure 3, we can see that lower levels of calcium substantially increase the cumulative intensity for new hospital admissions. The relative intensities are approximately constant with respect time; therefore, we have tabulated them at select values in Table 9 by aggregating over time, i.e., $RI(x_n, x_d) = J^{-1} \sum_j RI(t_{(j)}, \tilde{x}_n(t_{(j)}), \tilde{x}_d(t_{(j)}))$. In the second column, we have the relative intensity and, in the third column, the 95% credible interval. Notice as the serum calcium level drops the relative intensity of hospital admission increases dramatically. But, also, notice that all of the credible intervals include one which would seem to indicate considerable uncertainty; however, this interpretation is not entirely correct.

Recall that we have the sum of regression tree functions, $f$, which we are sampling via MCMC and the relative intensity is a function of $f$ and the patient data (see Section 2.4 for the definition of relative intensity). At each MCMC step, the calcium variable may, or may not, be chosen for a decision branch in any one of the $m$ regression trees. If calcium is not chosen at all, then by definition all relative intensities with respect to calcium will be exactly one for that step. A similar phenomenon occurs if calcium is chosen; for example, suppose only one tree has selected the calcium variable for a decision branch with the split value of $\leq 8.6$ and the reference group (in the denominator of the relative intensity) is $> 8.6$; then, for all patients with a calcium level above 8.6, their relative intensity will be exactly one (you can extend the same argument to more than one tree choosing calcium for a branch, calcium appearing on multiple branches in one tree, etc.). Therefore, whether one is in the interval (or not) is an inadequate measure of uncertainty; rather, the proportion of relative intensities that are one is a better measure which
we have tabulated in the fourth column in Table 9. Note that the proportion of ones decreases as the calcium level falls.

In Table 10, we display the relative intensities for the other important variables which are all binary; the relative intensities are present vs. absent, for all 60 months, and aggregated over time since they did not appear to vary temporally. Insulin therapy roughly doubles a patient’s risk of hospital admission. The 95% credible interval for insulin contains one, but we can also see that the relative intensity is rarely one which is a sign of a very important variable. The relative intensity for peripheral vascular disease (PVD) has a similar profile indicating it is also a very important variable with respect to hospital admission. Peptic ulcer disease (PUD) has a smaller relative intensity than insulin and PVD. The credible interval for PUD not only includes one, but it also includes values less than one which is a sign that it is less likely to be important than the others.

4. Discussion

Here we have outlined our new methodology for recurrent events and we have applied it to a motivating example of the hospital admissions experienced by diabetes patients. For investigations such as this, we are advocates for using as much information as is practically available. For example, with recurrent events we can employ all of the hospital admissions; if we had restricted to only the first admission in this cohort, we would have been limited to only 182 (34.7%) admissions rather than the 525 total that we actually analyzed. In the same vein, we brought in a variety of data from the EHR. From the clinical perspective, we are not aware of similar work assessing the risk of hospital admissions for diabetes patients with rich data sources like EHR; previous efforts on patients with diabetes such as Kaplan and Feinstein (1974) assessed the risk of complications and mortality from comorbidities, and Ou and others (2012) predicted health behavior and utilization from comorbidities.

Observational studies have limitations; for example, the patients in this study were not ran-
domized to insulin therapy vs. no insulin. However, our inferences are causal if we can assume that all confounders are measured due to BART’s uncanny capability to account for treatment heterogeneity (Hill, 2011); an assumption which we feel has some merit due to the omnibus of health care data enlisted from the EHR in this study. But, the causal inference assertion must be tempered for several reasons. 1) Although we went to great lengths to include a wide variety of EHR data (as described in Section 2.8), we employed only a portion of the EHR data available in this study, e.g., we ignored imaging, free text notes written by physicians and nurses, etc. 2) We assumed the information available to us was accurate and complete. 3) We assumed that the patients had equal access to, and received, high quality health care. Nevertheless, we strongly recommend mitigating the likely causal elevated risk of hospital admission that we foresee for the following groups of diabetes patients (by closely monitoring them in the outpatient clinic setting): those previously admitted to the hospital, with low levels of serum calcium, treated with insulin and/or suffering from peripheral vascular disease.

5. Software

We have created an R package called BART to perform nonparametric survival analysis, including recurrent events, with BART. BART is open source, free software and publicly available online (McCulloch and Sparapani, 2016). In Appendix E of the Supplementary Materials, we provide a brief introduction.

APPENDIX

A. Introduction

Herein contains the Appendices. In Appendix B, we review Cox proportional intensity models commonly employed for recurrent events. In Appendix C, we describe the Sequential BART missing imputation method which was employed in our motivating example. In Appendix D, we describe the simulation study we employed to compare our new BART method vs. counting
process Cox models for recurrent events. In Appendix E, we provide a brief introduction to the BART R package.

B. Recurrent events and Cox proportional intensity models

Recurrent events are often analyzed via Cox proportional intensity models (Kalbfleisch and Prentice, 2002; Hosmer Jr and others, 2008). We will briefly outline four Cox models commonly employed. We adopt the following notation: \((\kappa_i, t_i, x_i(t))\) where \(i = 1, \ldots, n\) indexes subjects; \(\kappa_i\) is a count of the number of events experienced during the observation period; \(t_i = [t_{0i}, t_{1i}, \ldots, t_{\kappa_i}]\) is a vector of the censoring time, \(t_{0i}\), and the event times, \(t_{1i}, \ldots, t_{\kappa_i}\); and \(x_i(t)\) is vector of covariates which may be time-dependent. For single event survival analysis, i.e. \(\kappa_i = \delta_i \in \{0, 1\}\) for all \(i\), the general form of the Cox proportional intensity model is the following: \(\lambda(t, x_i(t)) = \lambda_0(t) \exp(\beta' x_i(t))\) where \(\lambda(t, x_i(t))\) is the intensity, \(\lambda_0(t)\) is a nonparametric baseline intensity and \(\exp(\beta' x_i(t))\) is a parametric multiplier which we call linear proportionality. The cumulative intensity is \(\Lambda(t, x_i(t)) = \int_0^t \lambda(s, x_i(s))ds\) and the survival probability is \(S(t, x_i(t)) = \Pr(s > t | x_i(t)) = \exp(-\Lambda(t, x_i(t)))\). The likelihood contribution for each subject is \(\lambda(t_i, x_i(t_i))\delta_i S(t_i, x_i(t_i))\). The four Cox models we present differ in how they adapt the single event Cox model to the recurrent events setting.

B.1 Counting process Cox model of time (CPC)

In the counting process model of time, each subject’s experience is broken up into independent observation time intervals: \((0, t_{1i}], \ldots, (t_{\kappa_i-1,i}, t_{\kappa_i}], (t_{\kappa_i}, t_{0i}]\) (which collapses to a single interval \((0, t_{0i}]\) for a subject having no events). Each of these intervals is associated with a corresponding event indicator, \(\delta_{hi}\). Note that only the first interval starts at time zero; therefore, the remaining intervals are left truncated, i.e., delayed entry. This arrangement results in the following likelihood contribution for each subject:
\[ \lambda(t_{1i}, x_i(t_{1i}))^\delta_{hi} S(t_{1i}, x_i(t_{1i})) \left[ \prod_{h=2}^{\kappa_i} \lambda(t_{hi}, x_i(t_{hi}))^\delta_{hi} \frac{S(t_{hi}, x_i(t_{hi}))}{S(t_{h-1,i}, x_i(t_{h-1,i}))} \right] \frac{S(t_{0i}, x_i(t_{0i}))}{S(t_{\kappa_i,i}, x_i(t_{\kappa_i,i}))} \] which simplifies to \( S(t_{0i}, x_i(t_{0i})) \prod_{h=1}^{\kappa_i} \lambda(t_{hi}, x_i(t_{hi}))^\delta_{hi} \). This model employs a robust variance which goes by various names such as the Huber sandwich estimator. So, this is one way to adapt the recurrent events data into something akin to a single event survival analysis.

\section*{B.2 Counting process Cox model of time stratified by prior events}

The counting process model of time stratified by prior events is a simple extension of the previous model. The only difference is that the baseline intensity varies by the number of prior events, i.e., redefine \( \lambda(t, x_i(t)) = \lambda_{0h}(t) \exp(\beta' x_i(t)) \) where \( h = N_i(t-) \) where \( N_i(t-) \) is the counting process of prior events for subject \( i \) just prior to time \( t \). This model employs a robust variance via the Huber sandwich estimator.

\section*{B.3 Counting process Cox model of sojourn time stratified by prior events}

The counting process model of sojourn time stratified by prior events is similar to the previous model. The only difference is how the baseline intensity is parameterized with respect to time. The baseline intensity is constructed as a function of the sojourn time rather than time, i.e., redefine \( \lambda(t, x_i(t)) = \lambda_{0h}(v_i(t)) \exp(\beta' x_i(t)) \) where the sojourn time is \( v_i(t) = t - t_{hi} \) and \( h = N_i(t-) \). This model employs a robust variance via the Huber sandwich estimator.

\section*{B.4 Marginal Cox model of time}

The marginal model is a departure from the previous Cox models. We represent the maximum number of events experienced by any subject as \( \kappa = \max_i \kappa_i \). We assume that every subject is followed from time zero and has \( \kappa + 1 \) observation periods as follows: an event for each of intervals \( (0, t_{1i}], \ldots, (0, t_{\kappa_i}] \) with respective strata \( h = N_i(t-) = 0, \ldots, \kappa_i - 1 \); and \( \kappa - \kappa_i + 1 \) repeated non-events for the interval \( (0, t_{0i}] \) for strata \( h = \kappa_i, \ldots, \kappa + 1 \). This model employs a
robust variance via the Huber sandwich estimator.

B.5 Cox model summary

By no means is this an exhaustive list of the types of Cox models which might be considered for recurrent events. In our view, this is one of the issues with using Cox models for recurrent events, i.e., how should we decide which model to use with real data?

C. Handling missing data with BART

BART can handle missing data (Kapelner and Bleich, 2016; Xu and others, 2016). We utilize the missing data framework developed by Xu and others (2016) which they call Sequential BART (coincidentally, they applied it to a study of hyperglycemia with EHR). Sequential BART assumes that the missing covariates are missing at random, i.e., missingness only depends on what has been observed. Specifically, Sequential BART assumes that a missing covariate can be imputed by BART from the rest of the covariates, and so on, sequentially for all missing covariates.

A brief description of this method follows where we assume that all missing covariates are continuous (which is adequate for this study, although, Sequential BART can be extended to binary and categorical covariates as well).

Suppose that the covariates which are always observed for all subjects are denoted by $x_{i}^{\text{obs}}$ and the covariates which may be missing by $x_{i}^{\text{mis}} = (x_{i1}^{\text{mis}}, \ldots, x_{iK}^{\text{mis}})$ where $x_{ik}^{\text{mis}}$ are ordered from the least missing overall, $k = 1$, to the most missing, $k = K$, for computational efficiency. If for subject $i$ the covariate $x_{ik}^{\text{mis}}$ is missing, then its value can be imputed by Metropolis-Hastings
sampling (Hastings, 1970) at the \(i^{th}\) step as follows.

\[
x_{ik}^* | x_{i1}^{\text{obs}}, x_{i1}^{\text{mis}}, \ldots, x_{i(k-1)}^{\text{mis}}, f_k, \sigma_k^2 \sim \mathcal{N} \left( f_k(x_{i1}^{\text{obs}}, x_{i1}^{\text{mis}}, \ldots, x_{i(k-1)}^{\text{mis}}), \sigma_k^2 \right)
\]

where \(\theta_k = (f_k, \sigma_k^2) \sim \text{BART}\)

\[
\alpha_{k(l)} = \left[ x_{i(k+1)}^{\text{mis}} | x_{i1}^{\text{obs}}, x_{i1}^{\text{mis}}, \ldots, x_{ik}^*, \theta_{k+1} \right] \cdots \left[ x_{iK}^{\text{mis}} | x_{i1}^{\text{obs}}, x_{i1}^{\text{mis}}, \ldots, x_{i(l-1)}^{\text{mis}}, \theta_K \right] \]

Sample \(x_{ik}^*\) from the proposal density, (C.1), and accept the proposal with the probability 
\(\min(\alpha_{k(l)}, 1)\). For more details, see (Xu and others, 2016).

### D. Simulated data set scenarios

Here is a brief review of the Weibull distribution employed in these simulated data sets. The
Weibull distribution with “shape” parameter \(\alpha\) and “scale” parameter \(\beta\) has the following den-
sity \([t] = \frac{\alpha}{\beta} t^{\alpha-1} \exp \left[-\left(\frac{t}{\beta}\right)^\alpha\right]\). The survival function is \(S(t) = \exp \left[-\left(\frac{t}{\beta}\right)^\alpha\right]\), therefore, the
intensity (or hazard) is \(\lambda(t) = \frac{\alpha}{\beta} t^{\alpha-1}\) and the cumulative intensity is \(\Lambda(t) = \left[\frac{t}{\beta}\right]^\alpha\).

These scenarios have some major similarities, and minor differences, to/from our real data
example which we spell out here. We only use the covariates as measured at time zero with the
following exceptions: 1) \(t, N_i(t-)\) and \(v_i(t);\) and 2) for signs that are missing at time zero, we
use their imputed time zero value. Since death was rare in our real example, there are no deaths
in these simulated data sets; all patients are followed for the fixed period of five years per the
example. We assume that all events are instantaneous so that a new event can occur immediately
following the previous event, i.e., no forced zeros. For simplicity, we assume that there can be up
to three recurrent events, but no more.

#### D.1 Settings

For each scenario, we utilize 78 actual time zero covariates observed in this study (3 conditions
were excluded since they were all absent at time zero and not informative; similarly, another
condition was excluded since it was present for only one subject at time zero). We generated failure times as described above based on two settings: a proportional setting and a nonproportional setting. For each setting, we generated 400 data sets.

In each setting, there are only four active covariates each with a moderate to large impact on the outcomes. We do not intend to create scenarios that actually generated the hospital admissions at hand; rather, these scenarios are purely to test our methods. However, we do want to replicate the very important actual inter-relationships among the observed covariates since inter-dependent covariates present a challenge that is often over-looked in simulated data sets.

D.1.1 Proportional Setting We define the active covariates as follows.

\[ x_1 = I(\text{PUD}) \quad x_3 = I(\text{insulin}) \]
\[ x_2 = I(\text{PVD}) \quad x_4 = I(\text{Ca} \leq 9.1) \]

For each subject, we simulate three event times as follows (in months).

\[ t_1, t_2, t_3 \sim \text{Wei}(\alpha, \beta) \]
\[ \alpha = 4 \]
\[ \beta = \exp(x' \gamma) \quad \text{where} \quad x = [1, x_1, x_2, x_3, x_4]' \]
\[ = \exp(4.3 - 0.1x_1 - 0.2x_2 - 0.3x_3 - 0.35x_4) \]

Based on these settings, the relative intensity, \( RI(t, x_n, x_d) \), is as follows.

\[
RI(t, x_n, x_d) = \frac{\lambda(t, x_n)}{\lambda(t, x_d)} = \frac{\exp(x_n' \gamma)}{\exp(x_d' \gamma)} = \exp[\alpha(x_d - x_n)' \gamma]
\]

where \( x_n = [x_{n1}, x_{n2}, x_{n3}, x_{n4}]' \) and \( x_d = [x_{d1}, x_{d2}, x_{d3}, x_{d4}]' \)

This relative intensity is constant with respect to time and, therefore, these settings are propor-
tional as desired.

Let’s display these intensities graphically. Since \((x_1, x_2, x_3, x_4)\) are all binary, we can easily generate all sixteen configurations in Supplementary Figure 4 (note: there are only fourteen distinct \(\beta\)'s, therefore, only fourteen curves). In comparisons, we evaluate the cumulative intensity for these settings at the following periodic time points which roughly correspond to the percentiles of the distribution, averaged over the cohort, in parentheses: 12 (0.5), 24 (7), 36 (25) and 48 (48). The censoring time of 60 corresponds to roughly the 69th percentile, i.e., 31% censoring.

D.1.2 Nonproportional Setting The active covariates are defined in the same way as they are for the proportional setting. For each subject, we simulate three event times as follows (in months).

\[
t_1, t_2, t_3 \sim \text{Wei}(\alpha, \beta)
\]

\[
\alpha = x'\nu = 0.5 + 3.5x_4
\]

\[
\beta = x'\gamma = 66 - 3x_1 - 6x_2 - 9x_3 - 4.5x_4
\]

Based on these settings, the relative intensity, \(RI(t, x_n, x_d)\), is as follows.

\[
RI(t, x_n, x_d) = \frac{\lambda(t, x_n)}{\lambda(t, x_d)} = \frac{(x_n'\nu)(x_d'\gamma)(x_n'\nu)}{(x_d'\nu)(x_n'\gamma)(x_n'\nu)}t(x_n - x_d)'\nu
\]

where \(x_n = [x_{n1}, x_{n2}, x_{n3}, x_{n4}]'\) and \(x_d = [x_{d1}, x_{d2}, x_{d3}, x_{d4}]'\)

This relative intensity is not constant with respect to time and, therefore, these settings are nonproportional as desired.

Let’s display these intensities graphically. Since \((x_1, x_2, x_3, x_4)\) are all binary, we can easily generate all sixteen configurations in Supplementary Figure 5 (note: there are only fourteen distinct \(\beta\)'s, therefore, only fourteen curves). In comparisons, we evaluate the cumulative intensity for these settings at the following periodic time points which roughly correspond to the percentiles.
of the distribution, averaged over the cohort, in parentheses: 12 (23), 24 (31), 36 (42) and 48 (55).
The censoring time of 60 corresponds to roughly the 68\textsuperscript{th} percentile, i.e., 32\% censoring which is calibrated to be about the same for both proportional and nonproportional settings.

D.2 Comparisons

We compare our new BART model to a Counting Process Cox (CPC) model (see Section B.1) via 400 simulated data sets from each of the proportional and nonproportional settings. For the proportional setting, the CPC is the correct model. For the nonproportional setting, the CPC is not correct since it assumes proportionality. However, under the nonproportional setting, we still compare BART to CPC since it is the “right” model. By “right”, we mean that there are ad hoc Cox techniques analysts can use to mitigate the lack of proportionality such as stratification and time-dependent coefficients. But, these techniques require correctly diagnosing and implementing the counter-measures which require judgement and experience. This process is not certain to work, error prone and may not even be attempted. Therefore, we will compare BART to CPC in both settings.

The focus of the comparison is the cumulative intensity which is central to the recurrent events framework. For each method, we estimate the cumulative intensity and compare it to the actual known cumulative intensity. Since the cumulative intensity is a function of time, we compare at a set of time points (note that the variance of the cumulative intensity increases with time necessitating comparing at multiple time points). BART and CPC each estimate the cumulative intensity on a grid of time points, but these two grids are not the same. You could force the grids to be the same, but we do not believe that is the way these methods should be used in practice. Therefore, the grids are different, yet capable of answering the same questions. We compare in two ways. For each method, we measure the $R^2$ between the estimated and actual cumulative intensity on each method’s own grid. Also, we measure root mean square error (RMSE), bias,
95% interval coverage and 95% interval length via the estimated and actual cumulative intensity at the following time points: 12, 24, 36 and 48 months since all times are censored at 60 months in accordance with the real data example.

D.3 Results of comparisons

Here we summarize the $R^2$, root mean square error (RMSE), bias, 95% interval coverage and 95% interval length. N.B., technically, comparing the frequentist 95% confidence intervals from the Cox model with the Bayesian 95% credible intervals from the BART model is not an apples to apples comparison. However, as a practical matter, these comparisons are often done, and, we believe, they provide useful insights into the two methods.

These summaries are statistically compared via simple hypothesis testing. The $R^2$ results are compared by a paired z-test, $z = \frac{\bar{z}}{s_2/\sqrt{400}} \sim N(\mu, 1)$ where $\Delta_i = x_{i,BART} - x_{i,Cox}$, with the null hypothesis representing no difference, $H_0 : \mu = 0$, vs. the alternative hypotheses, $H_B : \mu > 0$ (BART is superior), and $H_C : \mu < 0$ (Cox is superior). Similarly, the bias, RMSE and 95% interval coverage will be compared by a paired ANOVA where time is the main effect, i.e., $\mu_t$ where $t \in \{12, 24, 36, 48\}$ with corresponding null hypotheses, $H_0 : \mu_t = 0$, vs. the alternative hypotheses for bias and RMSE, $H_B : \mu_t < 0$ (BART is superior) and $H_C : \mu_t > 0$ (Cox is superior); or for 95% interval coverage $H_B : \mu_t > 0$ (BART is superior) and $H_C : \mu < 0$ (Cox is superior). We did not statistically compare 95% interval length since it is unclear whether a shorter interval or a longer interval is preferable; such a decision is necessarily dependent on the 95% interval coverage comparison.

D.4 Proportional Setting

We compare the counting process Cox model (see Section B.1) to BART under the linear proportionality setting by analyzing 400 simulated data sets (see Section D.1.1).
Cox generally has a higher $R^2$ than BART between the estimated and active cumulative intensity; see Supplementary Figure 6. Since the $R^2$ estimates here do not approach the boundaries, a paired $z$-test is a legitimate hypothesis test. For $R^2$, the paired mean differential is -0.093, the $z$-test statistic is -63.6 and the p-value $< 1E-300$, therefore, we reject the null in favor of the hypothesis that the Cox model is superior.

Cox generally has smaller bias than BART in estimating the cumulative intensity; see Supplementary Figure 7. For bias from the paired ANOVA, Cox is superior at all time points: at 12 months, a 0.0019 difference, p-value=0.0012; at 24 months, a 0.027 difference, p-value $< 1E-300$; at 36 months, a 0.064 difference, p-value $< 1E-300$; and at 48 months, a 0.178 difference, p-value $< 1E-300$.

Cox generally has smaller RMSE than BART in estimating the cumulative intensity; see Supplementary Figure 8. For RMSE from the paired ANOVA, Cox is superior at 12 and 24 months, at 36 months we do not reject the null, and at 48 months BART is superior: at 12 months, a 0.0085 difference, p-value=8E-7; at 24 months, a 0.088 difference: p-value $< 1E-300$; at 36 months, a 0.00064 difference, p-value=0.71; and at 48 months, a -0.20 difference, p-value $< 1E-300$.

At 12 months, Cox has better 95% interval coverage than BART; at 24, 36 and 48 months, BART has better coverage than Cox; see Supplementary Figure 9. Since the 95% interval coverage estimates generally do not approach the boundaries, a paired ANOVA is a legitimate hypothesis test. For 95% interval coverage from the paired ANOVA, Cox is superior at 12; and at 24, 36 and 48 months BART is superior: at 12 months, a -0.59 difference, p-value $< 1E-300$; at 24 months, a 0.022 difference; p-value=0.0071; at 36 months, a 0.19 difference, p-value=9.3E-124; and at 48 months, 0.065 difference, p-value=2.2E-15.

BART generally has longer 95% interval length than Cox; see Supplementary Figure 10. As expected, Cox generally outperforms BART in this setting.
We compare the counting process Cox model (see Section B.1) to BART under the nonproportional setting by analyzing 400 simulated data sets (see Section D.1.2).

BART generally has a higher $R^2$ than Cox between the estimated and active cumulative intensity; see Supplementary Figure 11. Since the $R^2$ estimates here do not approach the boundaries, a paired z-test is a legitimate hypothesis test. For $R^2$, the paired mean differential is 0.087, the z-test statistic is 50.1 and the p-value $< 1E-300$, therefore, we reject the null in favor of the hypothesis that the BART model is superior.

BART generally has smaller bias than Cox in estimating the cumulative intensity; see Supplementary Figure 12. For bias from the paired ANOVA, BART is superior at all time points: at 12 months, a -0.012 difference, p-value=4.8E-52; at 24 months, a -0.026 difference, p-value=5.3E-250; at 36 months, a -0.025 difference, p-value=2.4E-218; and at 48 months, a -0.0054 difference, p-value=7.1E-12.

BART generally has smaller RMSE than Cox in estimating the cumulative intensity; see Supplementary Figure 13. For RMSE from the paired ANOVA, BART is superior at 12, 24 and 36 months, and at 48 months Cox is superior: at 12 months, a -0.063 difference, p-value $< 1E-300$; at 24 months, a -0.096 difference; p-value $< 1E-300$; at 36 months, a -0.065 difference, p-value $< 1E-300$ and at 48 months, a 0.017 difference, p-value=4.6E-31.

BART generally has better 95% interval coverage than Cox; see Supplementary Figure 14. Since the 95% interval coverage estimates generally do not approach the boundaries, a paired ANOVA is a legitimate hypothesis test. For 95% interval coverage from the paired ANOVA, BART is superior at all time points: at 12 months, a 0.033 difference, p-value=2.8E-48; at 24 months, a 0.059 difference: p-value=6.7E-148; at 36 months, a 0.13 difference, p-value $< 1E-300$; and at 48 months, 0.22 difference, p-value $< 1E-300$.

BART generally has longer 95% interval length than Cox; see Supplementary Figure 15. As
expected, BART generally outperforms Cox in this setting.

D.6 Discussion

We believe that the results of these simulations were predictable. We assumed that the BART (Cox) model would generally outperform the Cox (BART) model for the nonproportional (proportional) setting. The reasons for these results are obvious. In the proportional setting, the Cox model has an advantage since it assumes linear proportionality. In the nonproportional setting, the BART model has an advantage since it does not assume linear proportionality.

In practice with real data, we advocate for an approach with fewer assumptions like BART. BART does not require you to pre-specify the functional form and/or interactions, instead, it learns the functional form of the cumulative intensity from the data with respect to covariates, time, sojourn time, previous events, etc. On the other hand, Cox models require you to assume linear proportionality while pre-specifying interactions and your choice of model including the functional form of the cumulative intensity with respect to time, sojourn time, previous events, etc. (common Cox model choices listed in Section B). Of course, you can test your Cox model assumptions and choose alternative models based on those diagnostics; however, this process is time-consuming, repetitive, non-trivial, prone to error and often requires further assumptions. These realistic simulations lend credence to the contention that our new BART model can reliably analyze recurrent events data without adopting restrictive assumptions.

E. The BART R Package

We have created an R package called BART to perform nonparametric survival analysis, including recurrent events, with BART. BART is open source, free software and publicly available online (McCulloch and Sparapani, 2016). Since the computations involved for large data sets such as the motivating example can take some time, we provide both serial and parallel versions of
some functions which utilize the parallel R package function mcparallel. However, note that the mcparallel function depends on your operating system’s support for forking to perform parallel processing, e.g., this support is not available on Windows. Here is a transcript which shows off the capabilities of the package.

```r
## install BART
install.packages("BART")
## and load it
library(BART)
## peruse the help for recur.bart
?recur.bart
## notice that this help page documents both recur.bart and
## mc.recur.bart which are the serial and parallel versions
## respectively
## the motivating example shown below is at the bottom of the
## help page in the Examples section

## load 20 percent random sample
## however, in this report, we used the full 100 percent cohort
data(xdm20.train)
data(xdm20.test)
data(ydm20.train)
set.seed(99)
post <- recur.bart(x.train=xdm20.train, y.train=ydm20.train)
## larger data sets such as these can take some time
## if parallel processing is available, uncomment this block
```
# post <- mc.recur.bart(x.train=xdm20.train, y.train=ydm20.train,
#                         mc.cores=5, seed=99)
require(rplot)
require(rpart.plot)
dss <- rpart(post$yhat.train.mean~xdm20.train)
rpart.plot(dss)
## refer to Figure 16 below
## notice that all splits in the tree
## (except 1 at the bottom involving a small group)
## involve ca, cci_pud, cci_pvd, ins270 and n

## compare patients treated with insulin vs not
N.train <- 50
N.test <- 50
K <- post$K ## 798 unique time points
## only testing set, i.e., remove training set
xdm20.test. <- xdm20.test[N.train*K+(1:(N.test*K)), ]
xdm20.test. <- rbind(xdm20.test., xdm20.test.)
xdm20.test.[, 'ins270'] <- rep(0:1, each=N.test*K)
pred <- pwbart(xdm20.test., post$treedraws)
## larger data sets such as these can take some time
## if parallel processing is available, uncomment this block
## pred <- mc.pwbart(x.test., post$treedraws, mc.cores=5)
pred.haz.test <- pnorm(pred)
## create Friedman’s partial dependence function for the
## intensity/hazard by time and ins270

\[ NK_{test} \leftarrow N_{test}K \]

\[ M \leftarrow \text{nrow(pred.haz.test)} \] ## number of MCMC samples, typically 1000

\[ RI \leftarrow \text{matrix(0, M, K)} \]

for(i in 1:N.test)
    \[ RI \leftarrow RI+(\text{pred.haz.test}[ , (N.test+i-1)\star K+1:K])/\]
    \[ \text{pred.haz.test}[ , (i-1)\star K+1:K])/N_{test} \]

\[ RI_{lo} \leftarrow \text{apply(RI, 2, quantile, probs=0.025)} \]

\[ RI_{mu} \leftarrow \text{apply(RI, 2, mean)} \]

\[ RI_{hi} \leftarrow \text{apply(RI, 2, quantile, probs=0.975)} \]

plot(post$times, RI_{hi}, type='l', lty=2, log='y',
    ylim=c(min(RI_{lo}, 1/RI_{hi}), max(1/RI_{lo}, RI_{hi})),
    xlab='t', ylab='RI(t, x)',
    sub='insulin(ins270=1) vs. no insulin(ins270=0)',
    main='Relative intensity of hospital admissions for diabetics')

lines(post$times, RI_{mu})

lines(post$times, RI_{lo}, lty=2)

lines(post$times, rep(1, K), col='darkgray')

## RI for insulin therapy seems fairly constant with time

mean(RI_{mu})

### Acknowledgments

Funding for this project was provided by the Advancing Healthier Wisconsin Research and Education Program under award 9520277. This publication was supported by the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant Number
UL1TR001436. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. *Conflict of Interest*: None.

**Disclaimer**

The data from the motivating example have been supplied by the Medical College of Wisconsin Clinical Research Data Warehouse. The interpretation and reporting of these data are the responsibility of the author(s) and in no way should be seen as an official policy of, or interpretation by, the Medical College of Wisconsin.
Table 1. Inclusion and Exclusion Criteria

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Patients</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary diabetes diagnosis in 2006-2007 (ICD-9 diagnoses: 250.x0, 250.x2)</td>
<td>6000</td>
<td>estimated total</td>
</tr>
<tr>
<td>Age 21 to 90</td>
<td>1495</td>
<td>random sample</td>
</tr>
<tr>
<td>Race White or Black</td>
<td>1471</td>
<td></td>
</tr>
<tr>
<td>ZIP code either 530xx or 532xx</td>
<td>1341</td>
<td></td>
</tr>
<tr>
<td>E&amp;M visit in 2006-2007</td>
<td>927</td>
<td></td>
</tr>
<tr>
<td>E&amp;M visit in 2012-2013</td>
<td>791</td>
<td></td>
</tr>
<tr>
<td>or deceased with E&amp;M visit within 1 year of death</td>
<td>520</td>
<td></td>
</tr>
<tr>
<td>Excluding secondary diabetes (ICD-9 diagnosis: 249.xx)</td>
<td>488</td>
<td></td>
</tr>
<tr>
<td>and gestational diabetes (ICD-9 diagnoses: 648.8, V12.21)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Table 2. Vital signs and laboratory values

<table>
<thead>
<tr>
<th>Vital signs</th>
<th>Missingness</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height (in.)</td>
<td>9.84%</td>
<td>1</td>
</tr>
<tr>
<td>Diastolic pressure (mmHg)</td>
<td>23.76%</td>
<td>2</td>
</tr>
<tr>
<td>Systolic pressure (mmHg)</td>
<td>23.76%</td>
<td>2</td>
</tr>
<tr>
<td>Pulse (/min.)</td>
<td>24.40%</td>
<td>4</td>
</tr>
<tr>
<td>Weight (oz.)</td>
<td>32.30%</td>
<td>5</td>
</tr>
<tr>
<td>Temperature (F)</td>
<td>43.41%</td>
<td>6</td>
</tr>
<tr>
<td>Respirations (/min.)</td>
<td>45.48%</td>
<td>11</td>
</tr>
<tr>
<td>Laboratory diagnostic tests</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creatinine (mg/dL)</td>
<td>44.58%</td>
<td>7</td>
</tr>
<tr>
<td>Calcium (mg/dL)</td>
<td>45.25%</td>
<td>8</td>
</tr>
<tr>
<td>Potassium (mmol/L)</td>
<td>45.28%</td>
<td>9</td>
</tr>
<tr>
<td>Blood urea nitrogen (mg/dL)</td>
<td>45.44%</td>
<td>10</td>
</tr>
<tr>
<td>Chloride (mmol/L)</td>
<td>45.61%</td>
<td>12</td>
</tr>
<tr>
<td>Sodium (mmol/L)</td>
<td>45.72%</td>
<td>13</td>
</tr>
<tr>
<td>Glycohemoglobin A1c (%)</td>
<td>46.67%</td>
<td>14</td>
</tr>
<tr>
<td>Bicarbonate (mmol/L)</td>
<td>47.72%</td>
<td>15</td>
</tr>
</tbody>
</table>
### Table 3. Codes for conditions (* combined with RxRisk, see Table 5)

<table>
<thead>
<tr>
<th>Category</th>
<th>ICD-9/HCPCS Code(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute metabolic</td>
<td></td>
</tr>
<tr>
<td>Hyperinsulinism</td>
<td>962.3</td>
</tr>
<tr>
<td>Hypoglycemia</td>
<td>250.8x</td>
</tr>
<tr>
<td>Ketoacidosis</td>
<td>250.1x</td>
</tr>
<tr>
<td>Circulatory</td>
<td></td>
</tr>
<tr>
<td>Atrial fibrillation</td>
<td>427.31</td>
</tr>
<tr>
<td>Cardiomyopathy</td>
<td>425.4</td>
</tr>
<tr>
<td>Coronary artery disease</td>
<td>414.0x</td>
</tr>
<tr>
<td>Gangrene</td>
<td>785.4</td>
</tr>
<tr>
<td>Hypertension*</td>
<td>401.x</td>
</tr>
<tr>
<td>Eye</td>
<td></td>
</tr>
<tr>
<td>Blindness</td>
<td>369.xx</td>
</tr>
<tr>
<td>Retinopathy</td>
<td>362.0x</td>
</tr>
<tr>
<td>Kidney</td>
<td></td>
</tr>
<tr>
<td>Chronic kidney disease</td>
<td>585.x, E11.22</td>
</tr>
<tr>
<td>Dialysis</td>
<td>V45.1, V56.x, 90935:90937</td>
</tr>
<tr>
<td>Kidney failure</td>
<td>585.5, 585.6</td>
</tr>
<tr>
<td>Kidney transplant</td>
<td>V42.0, 50360, 50365</td>
</tr>
<tr>
<td>Nephropathy</td>
<td>583.81</td>
</tr>
<tr>
<td>Neurologic</td>
<td></td>
</tr>
<tr>
<td>Depression*</td>
<td>300.4, 311</td>
</tr>
<tr>
<td>Diabetic foot</td>
<td>713.5</td>
</tr>
<tr>
<td>Encephalopathy</td>
<td>348.30</td>
</tr>
<tr>
<td>Neuropathy</td>
<td>250.6x, 354.x, 355.x, 337.1, 353.5, 357.2, E11.40</td>
</tr>
<tr>
<td>Procedure/Surgery</td>
<td></td>
</tr>
<tr>
<td>Bariatric surgery</td>
<td>V45.86, 43775, 46344, 43846</td>
</tr>
<tr>
<td>CABG</td>
<td>V45.81, 33503:33505, 33510:33516, 4110F</td>
</tr>
<tr>
<td>PTCA</td>
<td>V45.82, 92982:92984, 92920, 92921</td>
</tr>
<tr>
<td>Other</td>
<td></td>
</tr>
<tr>
<td>Diabetic bone changes</td>
<td>731.8</td>
</tr>
<tr>
<td>Diabetic ulceration</td>
<td>707.1x, 707.8, 707.9</td>
</tr>
<tr>
<td>Medical nutrition therapy</td>
<td>97802, 97803</td>
</tr>
<tr>
<td>Sleep apnea</td>
<td>327.2x, 770.81, 770.82, 780.51, 780.53, 780.57, 786.03</td>
</tr>
</tbody>
</table>

### References


### Table 4. Codes for Charlson diagnoses (* combined with RxRisk, see Table 5)

<table>
<thead>
<tr>
<th>Charlson Diagnoses</th>
<th>ICD-9 Code(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cerebrovascular disease</td>
<td>362.34, 430.xx:438.xx</td>
</tr>
<tr>
<td>Chronic pulmonary disease</td>
<td>416.8, 416.9, 490.x:496.x, 500.x:505.x, 506.4, 508.1, 508.8</td>
</tr>
<tr>
<td>Congestive heart failure</td>
<td>398.91, 402.01, 402.11, 402.91, 404.01, 404.03, 404.11, 404.13, 404.91, 404.93, 425.4x, 425.5x, 425.7x:425.9x, 428.xx</td>
</tr>
<tr>
<td>Dementia</td>
<td>290.xx, 294.1x, 331.2x</td>
</tr>
<tr>
<td>Diabetes with chronic complications</td>
<td>250.4x:250.6x</td>
</tr>
<tr>
<td>Hemiplegia or paraplegia</td>
<td>334.1, 342.x, 343.x, 344.0:344.6, 344.9</td>
</tr>
<tr>
<td>Malignancy*</td>
<td>140.x:172.x, 174.x:195.8, 200.x:208.x, 238.6</td>
</tr>
<tr>
<td>Metastatic solid tumor</td>
<td>196.xx:199.xx</td>
</tr>
<tr>
<td>Mild liver disease</td>
<td>070.22, 070.23, 070.32, 070.33, 070.44, 070.54, 070.6, 070.9, 570.x, 571.x, 573.3, 573.4, 573.8, 573.9, V42.7</td>
</tr>
<tr>
<td>Moderate/severe liver dis.</td>
<td>456.0:456.2, 572.2:572.8</td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td>410.xx, 412</td>
</tr>
<tr>
<td>Peptic ulcer disease*</td>
<td>531.x:534.x</td>
</tr>
<tr>
<td>Peripheral vascular disease*</td>
<td>093.0x, 437.3x, 440.xx, 441.xx, 443.1x:443.9x, 447.1x, 557.1x, 557.9x, V43.4</td>
</tr>
<tr>
<td>Renal disease*</td>
<td>403.01, 403.11, 403.91, 404.02, 404.03, 404.12, 404.13, 404.92, 404.93, 582.x, 583.0:583.7, 585.x, 586.x, 588.0, V42.0, V45.1, V56.x</td>
</tr>
<tr>
<td>Rheumatic disease*</td>
<td>446.5, 710.0:710.4, 714.0:714.2, 714.8, 725.x</td>
</tr>
</tbody>
</table>

### Table 5. Codes for adult RxRisk diagnoses (* combined with diagnosis codes, see Tables 3 and 4)

<table>
<thead>
<tr>
<th>Adult RxRisk Diagnoses</th>
<th>Medi-Span pharmaceutical class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anxiety and tension</td>
<td>57</td>
</tr>
<tr>
<td>Asthma</td>
<td>44</td>
</tr>
<tr>
<td>Cardiac disease</td>
<td>35</td>
</tr>
<tr>
<td>Peripheral vascular disease*</td>
<td>83</td>
</tr>
<tr>
<td>Depression*</td>
<td>58</td>
</tr>
<tr>
<td>Epilepsy</td>
<td>72</td>
</tr>
<tr>
<td>Gastric acid disorder*</td>
<td>49</td>
</tr>
<tr>
<td>Gout</td>
<td>68</td>
</tr>
<tr>
<td>Heart disease</td>
<td>32, 33, 34</td>
</tr>
<tr>
<td>Hyperlipidemia</td>
<td>39</td>
</tr>
<tr>
<td>Hypertension*</td>
<td>36</td>
</tr>
<tr>
<td>Malignancies*</td>
<td>21, 50</td>
</tr>
<tr>
<td>Parkinson’s disease</td>
<td>73</td>
</tr>
<tr>
<td>Psychosis</td>
<td>59</td>
</tr>
<tr>
<td>Renal disease*</td>
<td>82</td>
</tr>
<tr>
<td>Rheumatoid arthritis*</td>
<td>66</td>
</tr>
<tr>
<td>Thyroid disorder</td>
<td>28</td>
</tr>
<tr>
<td>Tuberculosis</td>
<td>9</td>
</tr>
</tbody>
</table>


**Bleich, Justin, Kapelner, Adam, George, Edward I. and Jensen, Shane T.** (2014).
Table 6. Hospital admissions and deaths

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deceased</td>
<td>488</td>
</tr>
<tr>
<td>Yes</td>
<td>30 (6.1)</td>
</tr>
<tr>
<td>No</td>
<td>458 (93.9)</td>
</tr>
<tr>
<td>Hospital Admissions</td>
<td>488</td>
</tr>
<tr>
<td>Mean, SD</td>
<td>1.1 2.2</td>
</tr>
<tr>
<td>Min, Max</td>
<td>0 16</td>
</tr>
<tr>
<td>1</td>
<td>79 (16.2)</td>
</tr>
<tr>
<td>2</td>
<td>35 (7.2)</td>
</tr>
<tr>
<td>3</td>
<td>15 (3.1)</td>
</tr>
<tr>
<td>4</td>
<td>16 (3.3)</td>
</tr>
<tr>
<td>5</td>
<td>13 (2.7)</td>
</tr>
<tr>
<td>6</td>
<td>8 (1.6)</td>
</tr>
<tr>
<td>7+</td>
<td>14 (2.9)</td>
</tr>
</tbody>
</table>


**Dorie, V, Chipman, HA and McCulloch, RE.** (2016). dbarts: discrete Bayesian Additive
Fig. 1. As the number of previous hospital admissions, $N_i(t-)$, increases; the cumulative intensity function of a new hospital admission increases respectively. The increasing numbers on the far right ($1, \ldots, 6$) are the corresponding profiles from Table 8.

Regression Trees sampler. [https://cran.r-project.org/web/packages/dbarts](https://cran.r-project.org/web/packages/dbarts) accessed on 11/01/2016.

Fig. 2. Relative intensities and 95% credible intervals for previous admission profiles 3 to 1 (black) and 6 to 1 (gray) (see the profiles in Table 8).


FISHER, ES, BYNUM, JP AND SKINNER, JS. (2009). Slowing the growth of health care costs:
Fig. 3. Lower levels of serum calcium substantially increase the cumulative intensity for new hospital admissions. The decreasing numbers on the far right are the corresponding serum calcium levels.

lessons from regional variation. *NEJM* 360, 849–52.

Fig. 4. Intensities plotted at all sixteen settings, i.e., \((x_1, x_2, x_3, x_4)\) are all binary. However, there are only fourteen distinct \(\beta\)'s, therefore, only fourteen curves. These intensities are proportional.


GHOSH, D AND LIN, DY. (2000). Nonparametric analysis of recurrent events and death. *Bio-
Fig. 5. Intensities plotted at all sixteen settings, i.e., \((x_1, x_2, x_3, x_4)\) are all binary. However, there are only fourteen distinct \(\beta\)'s, therefore, only fourteen curves. These intensities are nonproportional. In the early (later) months, \(x_4 = 0\) produce the upper (lower) seven curves (black); and in the early (later) months, \(x_4 = 1\) produce the lower (upper) seven curves (gray).

*metrics* 56, 554–62.

GHOSH, D and LIN, DY. (2003). Semiparametric analysis of recurrent events data in the
Fig. 6. In the proportional setting, Cox generally has higher $R^2$ between the estimated and actual cumulative intensity across 400 simulated data sets.


Fig. 7. In the proportional setting, Cox generally has smaller bias estimating the cumulative intensity across 400 simulated data sets. Horizontal axis markings indicate the time (in months) at which the cumulative intensity was estimated followed by the estimation method: B for BART (blue) vs. C for Cox (red).

Fig. 8. In the proportional setting, Cox generally has smaller root mean square error (RMSE) estimating the cumulative intensity across 400 simulated data sets. Horizontal axis markings indicate the time (in months) at which the cumulative intensity was estimated followed by the estimation method: B for BART (blue) vs. C for Cox (red).

Fig. 9. In the proportional setting, Cox has better 95% interval coverage than BART at 12 months, and BART has better coverage at 24, 36 and 48 months in estimating the cumulative intensity across 400 simulated data sets. Horizontal axis markings indicate the time (in months) at which the cumulative intensity was estimated followed by the estimation method: B for BART (blue) vs. C for Cox (red).

Fig. 10. In the proportional setting, BART generally has longer 95% interval length estimating the cumulative intensity across 400 simulated data sets. Horizontal axis markings indicate the time (in months) at which the cumulative intensity was estimated followed by the estimation method: B for BART (blue) vs. C for Cox (red).

Fig. 11. In the nonproportional setting, BART generally has higher $R^2$ between the estimated and actual cumulative intensity across 400 simulated data sets.

Fig. 12. In the nonproportional setting, BART generally has smaller bias estimating the cumulative intensity across 400 simulated data sets. Horizontal axis markings indicate the time (in months) at which the cumulative intensity was estimated followed by the estimation method: B for BART (blue) vs. C for Cox (red).

Nonproportional setting: cumulative intensity RMSE

Fig. 13. In the nonproportional setting, BART generally has smaller root mean square error (RMSE) estimating the cumulative intensity across 400 simulated data sets. Horizontal axis markings indicate the time (in months) at which the cumulative intensity was estimated followed by the estimation method: B for BART (blue) vs. C for Cox (red).

Fig. 14. In the nonproportional setting, BART generally has better 95% interval coverage estimating the cumulative intensity across 400 simulated data sets. Horizontal axis markings indicate the time (in months) at which the cumulative intensity was estimated followed by the estimation method: B for BART (blue) vs. C for Cox (red).

Fig. 15. In the nonproportional setting, BART generally has longer 95% interval length estimating the cumulative intensity across 400 simulated data sets. Horizontal axis markings indicate the time (in months) at which the cumulative intensity was estimated followed by the estimation method: B for BART (blue) vs. C for Cox (red).

index.php/jss/article/view/v070104].

KAPLAN, MH AND FEINSTEIN, AR. (1974). The importance of classifying initial co-morbidity
Fig. 16. All splits in the tree involve ca (serum calcium), cci.pud (peptic ulcer disease), cci.pvd (peripheral vascular disease), ins270 (insulin therapy) or n (number of previous admissions, i.e., \(N_i(t-\)) except cci.tumor at the bottom involving a small group.


McCulloch, RE and Sparapani, RA. (2016). Bayesian Additive Regression Trees. [https:


REFERENCES


### Table 7. Patient characteristics and hospital admissions

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Beginning of Follow-up</th>
<th>End of Follow-up</th>
<th>Hospital Admissions</th>
<th>Unadjusted Relative Intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>488</td>
<td>488</td>
<td>525</td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>216 (44.3)</td>
<td>216 (44.3)</td>
<td>228 (43.4)</td>
<td>0.98 (0.69, 1.40)</td>
</tr>
<tr>
<td>F</td>
<td>272 (55.7)</td>
<td>272 (55.7)</td>
<td>297 (56.6)</td>
<td>1.00 (reference)</td>
</tr>
<tr>
<td>Race</td>
<td>488</td>
<td>488</td>
<td>525</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>174 (35.7)</td>
<td>174 (35.7)</td>
<td>265 (50.5)</td>
<td>1.85 (1.30, 2.63)</td>
</tr>
<tr>
<td>White</td>
<td>314 (64.3)</td>
<td>314 (64.3)</td>
<td>260 (49.5)</td>
<td>1.00 (reference)</td>
</tr>
<tr>
<td>Age</td>
<td>488</td>
<td>488</td>
<td>525</td>
<td></td>
</tr>
<tr>
<td>Mean, SD</td>
<td>56.0</td>
<td>60.9</td>
<td>60.3</td>
<td>15.7</td>
</tr>
<tr>
<td>Min, Max</td>
<td>21.0</td>
<td>26.0</td>
<td>24.0</td>
<td>92.0</td>
</tr>
<tr>
<td>21-44</td>
<td>104 (21.3)</td>
<td>66 (13.5)</td>
<td>87 (16.6)</td>
<td>1.00 (reference)</td>
</tr>
<tr>
<td>45-54</td>
<td>127 (26.0)</td>
<td>94 (19.3)</td>
<td>99 (18.9)</td>
<td>0.90 (0.50, 1.62)</td>
</tr>
<tr>
<td>55-64</td>
<td>116 (23.8)</td>
<td>137 (28.1)</td>
<td>115 (21.9)</td>
<td>0.85 (0.48, 1.53)</td>
</tr>
<tr>
<td>65-74</td>
<td>78 (16.0)</td>
<td>92 (18.9)</td>
<td>100 (19.0)</td>
<td>1.21 (0.65, 2.28)</td>
</tr>
<tr>
<td>75+</td>
<td>63 (12.9)</td>
<td>99 (20.3)</td>
<td>124 (23.6)</td>
<td>1.48 (0.83, 2.64)</td>
</tr>
<tr>
<td>ZIP3 area</td>
<td>488</td>
<td>488</td>
<td>525</td>
<td></td>
</tr>
<tr>
<td>532/urban</td>
<td>378 (77.5)</td>
<td>378 (77.5)</td>
<td>454 (86.5)</td>
<td>1.90 (1.19, 3.05)</td>
</tr>
<tr>
<td>530/suburb</td>
<td>110 (22.5)</td>
<td>110 (22.5)</td>
<td>71 (13.5)</td>
<td>1.00 (reference)</td>
</tr>
<tr>
<td>Insurance</td>
<td>488</td>
<td>488</td>
<td>525</td>
<td></td>
</tr>
<tr>
<td>Government &gt;65</td>
<td>141 (28.9)</td>
<td>191 (39.1)</td>
<td>224 (42.7)</td>
<td>1.00 (reference)</td>
</tr>
<tr>
<td>Government ≤65</td>
<td>134 (27.5)</td>
<td>138 (28.3)</td>
<td>208 (39.6)</td>
<td>1.08 (0.73, 1.59)</td>
</tr>
<tr>
<td>Commercial ≤65</td>
<td>192 (39.3)</td>
<td>143 (29.3)</td>
<td>71 (13.5)</td>
<td>0.31 (0.20, 0.48)</td>
</tr>
<tr>
<td>Other ≤65</td>
<td>21 (4.3)</td>
<td>16 (3.3)</td>
<td>22 (4.2)</td>
<td>0.77 (0.45, 1.32)</td>
</tr>
<tr>
<td>Insulin</td>
<td>488</td>
<td>488</td>
<td>525</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>191 (39.1)</td>
<td>206 (42.2)</td>
<td>391 (74.5)</td>
<td>4.56 (3.35, 6.21)</td>
</tr>
<tr>
<td>No</td>
<td>297 (60.9)</td>
<td>282 (57.8)</td>
<td>134 (25.5)</td>
<td>1.00 (reference)</td>
</tr>
<tr>
<td>Metformin</td>
<td>488</td>
<td>488</td>
<td>525</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>176 (36.1)</td>
<td>176 (36.1)</td>
<td>168 (32.0)</td>
<td>0.76 (0.53, 1.08)</td>
</tr>
<tr>
<td>No</td>
<td>312 (63.9)</td>
<td>312 (63.9)</td>
<td>357 (68.0)</td>
<td>1.00 (reference)</td>
</tr>
<tr>
<td>A1c(%)</td>
<td>389</td>
<td>416</td>
<td>488</td>
<td></td>
</tr>
<tr>
<td>Mean, SD</td>
<td>7.1</td>
<td>7.1</td>
<td>6.9</td>
<td>1.7</td>
</tr>
<tr>
<td>Min, Max</td>
<td>4.4</td>
<td>18.5</td>
<td>4.5</td>
<td>15.6</td>
</tr>
<tr>
<td>&lt; 9</td>
<td>355 (91.3)</td>
<td>366 (88.0)</td>
<td>437 (89.5)</td>
<td>1.00 (reference)</td>
</tr>
<tr>
<td>≥ 9</td>
<td>34 (8.7)</td>
<td>50 (12.0)</td>
<td>51 (10.5)</td>
<td>1.25 (0.83, 1.88)</td>
</tr>
<tr>
<td>Missing</td>
<td>99</td>
<td>72</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>Calcium(mg/dL)</td>
<td>415</td>
<td>435</td>
<td>520</td>
<td></td>
</tr>
<tr>
<td>Mean, SD</td>
<td>9.3</td>
<td>9.4</td>
<td>8.9</td>
<td>0.7</td>
</tr>
<tr>
<td>Min, Max</td>
<td>7.2</td>
<td>10.9</td>
<td>7.2</td>
<td>11.6</td>
</tr>
<tr>
<td>≤ 9.1</td>
<td>151 (36.4)</td>
<td>125 (28.7)</td>
<td>334 (64.2)</td>
<td>7.35 (5.67, 9.51)</td>
</tr>
<tr>
<td>&gt; 9.1</td>
<td>264 (64.6)</td>
<td>310 (71.3)</td>
<td>186 (35.8)</td>
<td>1.00 (reference)</td>
</tr>
<tr>
<td>Missing</td>
<td>73</td>
<td>53</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>PVD</td>
<td>488</td>
<td>488</td>
<td>525</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>172 (18.5)</td>
<td>272 (29.2)</td>
<td>488 (52.4)</td>
<td>16.16 (11.19, 23.34)</td>
</tr>
<tr>
<td>No</td>
<td>316 (55.5)</td>
<td>216 (38.0)</td>
<td>37 (6.5)</td>
<td>1.00 (reference)</td>
</tr>
<tr>
<td>PUD</td>
<td>488</td>
<td>488</td>
<td>525</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>239 (22.7)</td>
<td>333 (31.7)</td>
<td>480 (45.7)</td>
<td>7.49 (5.09, 11.01)</td>
</tr>
<tr>
<td>No</td>
<td>249 (55.5)</td>
<td>155 (34.5)</td>
<td>45 (10.0)</td>
<td>1.00 (reference)</td>
</tr>
</tbody>
</table>
Table 8. *Profiles for N_i(t−) incrementing only at the months indicated (see Figures 1 and 2)*

<table>
<thead>
<tr>
<th>Profile</th>
<th>6</th>
<th>12</th>
<th>18</th>
<th>24</th>
<th>30</th>
<th>36</th>
<th>42</th>
<th>48</th>
<th>54</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
</tbody>
</table>

Table 9. *Relative intensity of hospital admission by serum calcium level*

<table>
<thead>
<tr>
<th>Ca (mg/dL)</th>
<th>RI</th>
<th>95% CI</th>
<th>Pr(RI = 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.4</td>
<td>2.61</td>
<td>(1.00, 7.43)</td>
<td>0.343</td>
</tr>
<tr>
<td>8.5</td>
<td>2.01</td>
<td>(1.00, 4.56)</td>
<td>0.474</td>
</tr>
<tr>
<td>8.6</td>
<td>1.62</td>
<td>(1.00, 3.78)</td>
<td>0.636</td>
</tr>
<tr>
<td>8.9</td>
<td>1.24</td>
<td>(1.00, 3.06)</td>
<td>0.828</td>
</tr>
<tr>
<td>9.5</td>
<td>1.00</td>
<td>reference</td>
<td>1.000</td>
</tr>
<tr>
<td>10.3</td>
<td>0.92</td>
<td>(0.30, 1.00)</td>
<td>0.854</td>
</tr>
</tbody>
</table>

Table 10. *Relative intensity of hospital admission by important binary variables*

<table>
<thead>
<tr>
<th>Variable</th>
<th>RI</th>
<th>95% CI</th>
<th>Pr(RI = 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insulin</td>
<td>2.17</td>
<td>(1.00, 3.34)</td>
<td>0.049</td>
</tr>
<tr>
<td>PVD</td>
<td>2.29</td>
<td>(1.00, 3.85)</td>
<td>0.104</td>
</tr>
<tr>
<td>PUD</td>
<td>1.66</td>
<td>(0.92, 3.36)</td>
<td>0.292</td>
</tr>
</tbody>
</table>