

RESEARCH ARTICLE

# Identifying Change Points in a Covariate Effect on Time-to-Event Analysis with Reduced Isotonic Regression

Yong Ma<sup>1,2\*</sup>, Yinglei Lai<sup>1,3</sup>, John M. Lachin<sup>1,2</sup>

1. The Biostatistics Center, George Washington University, Rockville, MD, United States of America, 2. Department of Epidemiology and Biostatistics, Milken Institute School of Public Health, George Washington University, Washington, DC, United States of America, 3. Department of Statistics, George Washington University, Washington, DC, United States of America

\*[yma@bsc.gwu.edu](mailto:yma@bsc.gwu.edu)

## Abstract

Isotonic regression is a useful tool to investigate the relationship between a quantitative covariate and a time-to-event outcome. The resulting non-parametric model is a monotonic step function of a covariate  $X$  and the steps can be viewed as change points in the underlying hazard function. However, when there are too many steps, over-fitting can occur and further reduction is desirable. We propose a reduced isotonic regression approach to allow combination of small neighboring steps that are not statistically significantly different. In this approach, a second stage, the reduction stage, is integrated into the usual monotonic step building algorithm by comparing the adjacent steps using appropriate statistical testing. This is achieved through a modified dynamic programming algorithm. We implemented the approach with the simple exponential distribution and then its extension, the Weibull distribution. Simulation studies are used to investigate the properties of the resulting isotonic functions. We apply this methodology to the Diabetes Control and Complication Trial (DCCT) data set to identify potential change points in the association between HbA1c and the risk of severe hypoglycemia.



CrossMark  
click for updates

## OPEN ACCESS

**Citation:** Ma Y, Lai Y, Lachin JM (2014) Identifying Change Points in a Covariate Effect on Time-to-Event Analysis with Reduced Isotonic Regression. PLoS ONE 9(12): e113948. doi:10.1371/journal.pone.0113948

**Editor:** Gang Han, Yale School of Public Health, United States of America

**Received:** July 8, 2014

**Accepted:** October 31, 2014

**Published:** December 4, 2014

**Copyright:** © 2014 Ma et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License](http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability:** The authors confirm that all data underlying the findings are fully available without restriction. Data are available at the following web location: <http://www2.bsc.gwu.edu/bsc/webpage.php?no=18>.

**Funding:** The authors have no support or funding to report.

**Competing Interests:** The authors have declared that no competing interests exist.

## Introduction

In clinical practice, disease diagnosis and subsequent treatment are often guided by a strict threshold (i.e. change point) of a biomarker. For example, fasting plasma glucose (FPG) at 126 mg/dl is the cutoff to diagnose type II diabetes, and more intensive treatment is used when FPG reaches 140 mg/dl. Such change points are often identified through a large scale health study where disease risk

increases substantially when a biomarker level exceeds a change point. Because identifying change points is data driven, more recent research data would mandate the update of the change points. In the case of diabetes diagnosis, the diagnostic threshold was at FPG  $\geq 140$  mg/dl before 1997. However, in 1997, increased cardiovascular and micro-vascular disease risk at lower values prompted the American Diabetes Association to recommend lowering the diagnostic threshold to 126 mg/dl. Changes like this have huge effects on medical practice, especially the initiation of a treatment, hence a systematic approach to identify change points in a covariate is well worth the effort.

Ancukiewicz et al. [1] have established an isotonic regression method to model the relationship between a quantitative covariate and clinical events. The covariate is assumed to be discrete with multiple levels so that the model provides an estimate of the outcome at every discrete value of the covariate. The resulting model is a step function where each new step can be viewed as a change point. They used their method to identify a change point in the association of CD4 count with HIV risk and the method worked well. However, in situations where the data is dense, that is, there is a large number of subjects with the outcome event and support over many discrete levels of the covariate, the model can also include many mini-steps and further combination of some mini-steps is desirable. Schell and Singh (1997) [2] proposed the idea of ‘reduced isotonic regression’ in which a backward elimination procedure is used after the usual isotonic regression model is built. Salanti and Ulm (2005) [3] also proposed a two-step procedure to estimate threshold limit values with binary outcomes. In their approach, the second stage in the algorithm is a sequence of Fisher tests for the adjacent  $2 \times 2$  tables to accomplish a reduced model. Very recently, Han et al. (2013) [4] proposed to use a reduced piecewise exponential approach to improve the modeling of survival time. They also used a two step procedure in which all insignificant change-points are eliminated after first implementing an order restriction on the failure rate. A flaw in the two stage approach is that the resulting model may not give the global maximum likelihood. Thus, we propose to employ a global optimization approach, examining all potential combinations of isotonic models with the constraint that the adjacent steps are significantly different and then identify the one with the maximum likelihood. We implemented this approach with a modified dynamic programming algorithm proposed by Lai [5]. This approach was chosen over the popular *pool adjacent violators algorithm* (PAVA) because the later cannot guarantee a global optimization solution when the extra testing is required. Lai and Albert [6] described using the approach in a linear mixed effects model, here we apply the approach in a parametric time-to-event data analysis.

In a nutshell, the algorithm examines all observed covariate ( $X$ ) values, from the smallest ( $x_1$ ) to the largest ( $x_n$ ), one at a time. At each  $X$  value, the algorithm will partition the values smaller or equal to  $X$  and identify an optimal step function satisfying the following three criteria: the function is isotonic, the distributions between two adjacent steps are significantly different, and the optimal step function has the maximum likelihood among all possible step functions that meet

the first two criteria. In the process of finding the optimal partition, all the other partitions that satisfy the first two conditions are recorded and saved for future use. This unique feature reduces the computing time from the  $O(2^{n-1})$  in a naive try to  $O(n^3)$ , assuming that there are  $n$  possible  $X$  values. The detailed description and related mathematical proofs about the modified dynamic programming algorithm was published elsewhere [5].

Large scale clinical trials like the Diabetes Control and Complication Trial (DCCT) [7] and the UK Prospective Diabetes Study (UKPDS) [8] demonstrated that improved glycemic control, represented by HbA1c (approximately a function of the 12 week average of glucose), reduces microvascular complications. However, they also showed that a lower glucose level is associated with elevated risk of severe hypoglycemia. It is therefore critical to identify the change points in the association between HbA1c and hypoglycemia to help establish a glycemic target which is low enough to minimize microvascular risk and yet not so low as to increase the risk of severe hypoglycemia. We apply this methodology to the DCCT data set to identify such change points.

## Methods

As in a parametric regression approach for time-to-event data, the null hypothesis here is that the covariate of interest  $X$  is not associated with survival time. The alternative hypothesis is that there exists at least one  $X$  value where the survival function changes significantly after reaching this value. If there are more than one change point, the change in survival function is monotonic. Without loss of generality, we only present the monotonically increasing scenario. Assuming that  $g(X)$  is a parameter in the survival time distribution, the hypothesis testing can be described as the following

$$\begin{aligned} H_0 : g(X_i) = \lambda \text{ is a constant} \\ H_1 : g(X_i) = \lambda_i, \lambda_1 \leq \lambda_2 \leq \dots \leq \dots \leq \lambda_m, \lambda_k \neq \lambda_{k+1} \text{ for some } k \text{ in } \{1, 2, \dots, m-1\} \end{aligned} \tag{1}$$

To establish a reduced isotonic regression as proposed, we need to specify the underlining survival function first. We start with the simple exponential distribution with a constant hazard in terms of time and then extend the results to the more robust Weibull distribution.

### Survival Time with an Exponential Distribution

When the event times follow an exponential distribution with constant hazard rate  $\lambda$ , the survival function can be expressed as

$$S(t) = e^{-\lambda t} \tag{2}$$

where  $\lambda$  can be expressed as an isotonic step function of the covariate  $X$ . The goal

of the algorithm is to determine whether each  $X$  value can be combined with its neighbors so that the final step function meets the three criteria described previously:  $\lambda$  being monotonically increasing, adjacent steps being statistically significantly different and having the overall maximum likelihood. The algorithm starts at the smallest covariate value ( $x_1$ ) and moves on until the final optimization is achieved.

Assume that  $\lambda_i$  is the hazard associated with the  $i$ th level of  $X$  ( $x_i$ ) where there are  $m_i$  observations. The time-to-event data for the  $j$ th participant in this group is represented as  $(\delta_{ij}, t_{ij})$ , where  $\delta_{ij}$  is the censoring indicator ( $\delta_{ij}=1$  indicating an event or  $\delta_{ij}=0$  indicating right-censoring) and  $t_{ij}$  is the survival time.

The log likelihood for all  $m_i$  observations can be expressed as

$$l(\lambda_i) = \sum_{j=1}^{m_i} \delta_{ij} \log(\lambda_i) - \sum_{j=1}^{m_i} \lambda_i t_{ij} \tag{3}$$

and  $\lambda_i$  can be estimated as

$$\hat{\lambda}_i = \frac{\sum_{j=1}^{m_i} \delta_{ij}}{\sum_{j=1}^{m_i} t_{ij}} \tag{4}$$

Now that the  $\lambda_i$  is estimated, we want to compare this  $\hat{\lambda}_i$  with the next one  $\hat{\lambda}_{i+1}$  so that the estimates are monotonically increasing and significantly different. We use Cox's  $F$ -test for the statistical testing since it is the most powerful test for comparing two exponential distributions [9] [10]. Assuming

$$\bar{t}_i = \frac{\sum_{j=1}^{m_i} t_{ij}}{\sum_{j=1}^{m_i} \delta_{ij}} \tag{5}$$

The ratio of the two  $F = \bar{t}_i / \bar{t}_{i+1}$  follows a  $F$ -distribution with  $u = 2 \sum_{j=1}^{m_i} \delta_{ij}$  and  $v = 2 \sum_{j=1}^{m_{i+1}} \delta_{(i+1)j}$  degrees of freedom. The test reject  $H_0$  if  $\bar{t}_i / \bar{t}_{i+1} > F_{u,v,\alpha}$  where the nominal  $\alpha$  is pre-specified. When either  $\hat{\lambda}_i \geq \hat{\lambda}_{i+1}$  or the  $F$  test is not significant at  $\alpha$ , the two  $X$  steps are combined and treated as a single step and the procedure continues. There are usually multiple partitions that will satisfy both the isotonic and significance criteria, among them, the one with the largest likelihood function is chosen as the optimization. The detailed algorithm used in the optimization is presented in the Supporting Information section.

### Survival Time with a Weibull Distribution

The Weibull distribution is an extension of the exponential distribution and its shape parameter,  $\gamma$ , determines the shape of the distribution of survival times. The Weibull survival function can be expressed as

$$S(t) = e^{-(\lambda t)^\gamma} \tag{6}$$

It is well known that, if a random variable  $T$  follows a Weibull distribution with parameters  $(\gamma, \lambda)$ , then  $T^\gamma$  will follow an exponential distribution with parameter  $\lambda^\gamma$ . Therefore, for a given  $\gamma$ , a simple power transformation of the survival times yields an exponential distribution. With this feature we can obtain estimates under a Weibull assumption by employing the algorithm already developed for the exponential distribution with power-transformed data. We assume that  $\lambda$  is a step function of  $X$ , and  $\gamma$  is a constant which will be estimated together with  $\lambda$ .

The log likelihood function for the observations with covariate value  $x_i$  under the Weibull assumption is

$$l(\lambda_i, \gamma) = \sum_{j=1}^{m_i} [\delta_{ij}(\gamma \log(\lambda_i) + \log(\gamma) + (\gamma - 1)\log t_{ij}) - (\lambda_i t_{ij})^\gamma] \tag{7}$$

and the likelihood for all data is

$$l(\lambda, \gamma) = \sum_{i=1}^n l(\lambda_i, \gamma) = \sum_{i=1}^n \sum_{j=1}^{m_i} [\delta_{ij}(\gamma \log(\lambda_i) + \log(\gamma) + (\gamma - 1)\log t_{ij}) - (\lambda_i t_{ij})^\gamma] \tag{8}$$

We use the following iterative steps to estimate  $\lambda$  and  $\gamma$ .

Step 1: Estimate  $\gamma$  by assuming that all observations are independently identically distributed (*i.i.d.*) from the same Weibull distribution with parameters  $(\lambda, \gamma)$ , i.e.,  $\lambda$  is the same for all  $x_i$ .

Step 2: With  $\gamma$  estimated, we transform the survival time  $t_{ij}$  to  $t_{ij}^\gamma$  and use the algorithm developed in the exponential case to estimate  $\lambda_i$

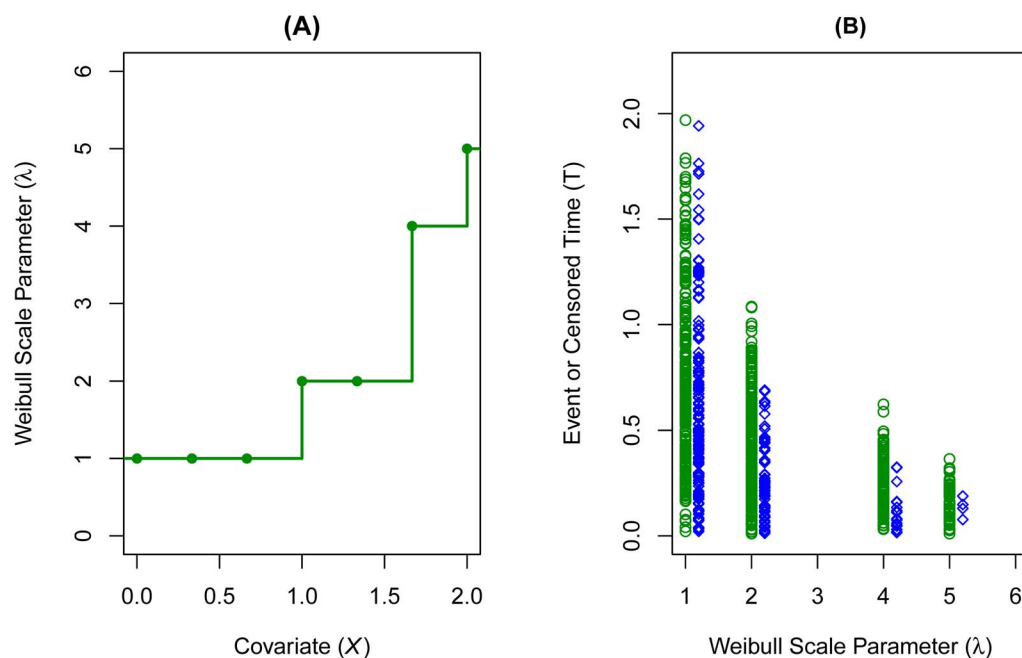
$$\hat{\lambda}_i^\gamma = \frac{\sum_{j=1}^{m_i} \delta_{ij}}{\sum_{j=1}^{m_i} t_{ij}^\gamma} \tag{9}$$

Step 3: We update  $\gamma$  with a MLE estimator by solving the following equation derived from (8)

$$\frac{\partial l}{\partial \gamma} = \sum_{i=1}^n \sum_{j=1}^{m_i} [\delta_{ij}(\log(\lambda_i t_{ij}) + \frac{1}{\gamma}) - (\lambda_i t_{ij})^\gamma \log(\lambda_i t_{ij})] = 0 \tag{10}$$

Step 4: Repeat steps 2 and 3 until the  $\gamma$  estimate converges, which is defined as change in  $\gamma$  is less than 0.1%.

Both the exponential and the Weibull algorithms have been implemented in the R statistical system and the codes can be found in the supplemental material.



**Figure 1. Underlying true model and time-to-event data.** (A) Weibull scale parameter  $\lambda$  is an isotonic function of  $X$ . (B) Event or censored time follows the Weibull distribution with a fixed shape parameter ( $\gamma = 2$ ) and scale parameter  $\lambda$  shown in (A). The green open circles represent event times and the blue open diamonds represent censored times.

doi:10.1371/journal.pone.0113948.g001

## Results

### An Example

We illustrate the algorithm using the following hypothetical example depicted in [Figure 1](#). We simulated time-to-event data that follows a Weibull distribution.  $X$  was the covariate of interest and had 7 distinct values  $x_i \in \{0, 1/3, 2/3, 1, 4/3, 5/3, 2\}$ , evenly spaced. A data set of 1000 observations was then generated by sampling a value of  $X$  from the set where the extremes each had probability  $1/12$ th and the other 5 values had probability  $1/6$ th. The corresponding hazard rates  $\lambda(x_i)$  were determined by the step function shown in [Figure 1](#) (A) and the shape parameter  $\gamma$  was set at 2.

To generate the event time  $T$ , we used the known fact that if  $T$  followed a Weibull distribution with parameters  $[\gamma, \lambda(x_i)]$  then  $T^\gamma$  would follow an exponential distribution with parameter  $\lambda(x_i)^\gamma$ . We randomly generated an event time  $T_0$  from the exponential distribution with rate  $\lambda(x_i)^\gamma$ . Thus, a back transformation of  $T = T_0^{\frac{1}{\gamma}}$  would create a random variable  $T$  following a Weibull distribution  $[\gamma, \lambda(x_i)]$ . For censoring, we used independently generated random numbers following uniform distributions in the interval between the minimum and the maximum of the event times as the censoring time  $C$ . The minimal of the event time and the censoring time  $\min(T, C)$  was used as the final survival time. The event indicator  $\delta$  was coded as  $\delta_{ij} = 1$  when  $T \leq C$  or 0 otherwise. [Figure 1](#) (B)

**Table 1.** Sample data following Weibull distributions with  $\gamma=2.0$  and  $\lambda$  as a step function of  $X$ .

True Parameters							
Steps	1			2		3	4
$x_i$	0	1/3	2/3	1	4/3	5/3	2
$n_i$	94	159	166	170	160	184	67
$\lambda_i$	1	1	1	2	2	4	5
Initial Estimates							
$\hat{\lambda}_i$	1.09	1.04	0.95	2.02	1.91	3.87	5.31
$se(\hat{\lambda}_i)$	0.07	0.05	0.05	0.09	0.08	0.15	0.33

$x_i$ : covariate with 7 distinct values;  $n_i$ : number of observations at each  $X$  value;  $\lambda_i$ : step function of  $X$  with 4 distinct values;  $\hat{\lambda}_i$ :  $\lambda$  estimated at each  $X$  value before the implementation of the reduced isotonic regression algorithm;  $se(\hat{\lambda}_i)$ : standard error of the  $\hat{\lambda}_i$ .

doi:10.1371/journal.pone.0113948.t001

displays survival times with open circles representing event times ( $\delta_{ij} = 1$ ) and open diamonds representing censored times ( $\delta_{ij} = 0$ ).

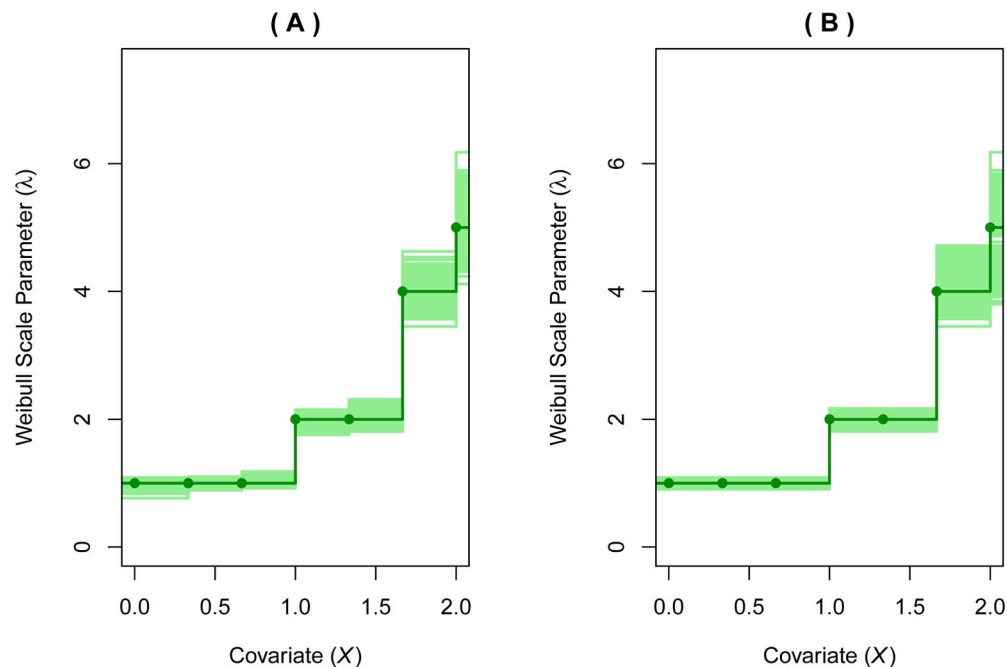
Table 1 shows the distinct values of  $X$ , number of subjects with each value of  $X$ , true hazard rates  $\lambda$  and the initial individually estimated hazard rates and their standard errors (with  $\gamma$  estimated at 2.02 in the final iteration of the algorithm). Before the constraint of monotonicity, the seven distinct estimates ( $\hat{\lambda}_i$ s) were close to the true values ( $\lambda_i$ s), however, no change point(s) could be determined because each  $\lambda$  is a distinct value.

We applied the algorithm to the data set to obtain a reduced isotonic regression model. The same example was repeated 1000 times, each time with a slightly different random data set and the results are shown in Figure 2. Panels (A) corresponds to models from the regular isotonic regression and panel (B) from the reduced isotonic regression with pre-specified testing significance at 0.0001. In panel (B), the bands of the estimates around the true  $\lambda$  values at 1 and 2 are much tighter, indicating improved model fit from incorporating the significance testing. The very small nominal  $\alpha$  was chosen for this example to demonstrate the effect of statistical testing. Such a stringent significance level could be too strict for real world data and shouldn't always be used.

### Small Sample Performance

A good modeling strategy is a strategy that still works when sample size is small. In the case of time-to-event data, due to censoring, the statistical information depends on the number of subjects experiencing the event, which is smaller than the number of participants in the study. Here we evaluate the performance of the reduced isotonic regression employing combinations of sample size and percent of censoring that yields 500, 200, 100 and 50 events.

We used the previously described example again. Adjacent steps were tested at nominal  $\alpha = 0.05$  and each scenario was repeated 1000 times. The results are shown in Table 2 and Table 3. Table 2 describes the frequency of various steps we



**Figure 2. Simulation results from 1000 repetitions.** (A) Regular isotonic regression without testing between steps. (B) Reduced isotonic regression with nominal  $\alpha = 0.0001$ . The dark green lines represent the underlying true model.

doi:10.1371/journal.pone.0113948.g002

identify in the 1000 repetitions. Table 3 summarizes the means and mean squared errors of the estimates of  $\hat{\gamma}$  and  $\hat{\lambda}_1, \dots, \hat{\lambda}_7$ . As the number of events decreases, the number of correctly identified steps (4) decreases and the model estimates are more likely to be biased with bigger variances. However even when the number of events is as small as 50, the models are able to capture the pattern of the underlining true model at about 25% of the times and only miss 1 step at 70% of the times. The parameter estimates are reasonably close to the true parameters.

### Model Diagnostics and Other Features

Cox-Snell [11] residuals can be applied to assess whether the model assumptions are accurate. If the model fits the data, and we plot Cox-Snell residual  $r_i$  against the negative log of the survival function of the residual  $-\log \hat{S}(r_i)$ , it should be a straight line with unit slope and zero intercept.

Although the nominal significance between the steps of the final model is pre-specified (herein at level  $\alpha$ ), such testing between any two steps does not provide an overall test of the significance of the covariate effect in the reduced isotonic regression model. Under certain conditions the likelihood ratio test of the covariate significance may follow a chi-square distribution. However, the degrees of freedom is unknown. We propose to use a permutational approach to obtain



**Table 2.** Number of steps identified with various event numbers and percent censored.

Event N	Sample N	% Censored mean(range)	Number of Steps				
			2	3	4	5	6
~500	1000	50.5(37.4–60.1)	0	86	785	126	3
	800	37.9(26.4–47.4)	0	126	745	122	7
	500	0	0	223	680	96	1
~200	400	50.7(34.3–62.5)	0	376	527	95	2
	320	37.5(24.1–49.1)	0	440	485	72	3
	200	0	0	509	429	59	3
~100	200	50.8(30.5–65.5)	0	550	384	62	4
	160	37.4(23.1–53.1)	0	577	372	49	2
	100	0	0	621	326	51	2
~50	100	50.5(31.0–70.0)	20	694	266	20	0
	80	37.4(20.0–60.0)	8	699	268	25	0
	50	0	19	737	229	18	0

The combination of "Sample N" and "%censore" is used to yield the targeted number of events in the "Event N" column, repeated 1000 times.

doi:10.1371/journal.pone.0113948.t002

the distribution of the likelihood ratio test under the null. This distribution will allow us to calculate the p-value of the covariate of interest.

It is still difficult to understand the theoretical properties of the parameters estimated from the reduced isotonic regression algorithm. To circumvent the problem, we use the distribution free bootstrap approach [12] to calculate the confidence intervals of the parameter estimates. A bootstrap sample with replacement is created from the original data set and model parameters are generated. This is repeated multiple times and a distribution of the parameter estimates is created. The 95% confidence intervals for the parameter estimates is therefore constructed.

### Application

The Diabetes Control and Complication Trial (DCCT) was a clinical trial aimed at comparing intensive treatment, i.e., at least 3 insulin injections a day, to the traditional treatment, once or twice a day for Type 1 diabetes mellitus (T1DM) patients. Although the intensive treatment significantly delayed the onset and slowed the progression of retinopathy, neuropathy and nephropathy, there is a two-to-three fold increase in episodes of severe hypoglycemia (low blood sugar) that could lead to coma (unconsciousness) and/or siezures [7]. Here we use the methods developed in the previous sections to explore the relationship between HbA1c and severe hypoglycemia for the 711 participants in the intensive treatment group. The DCCT hypoglycemia data is described by Lachin [13] and can be obtained from the following web site: <http://www2.bsc.gwu.edu/bsc/webpage.php?no=18>.

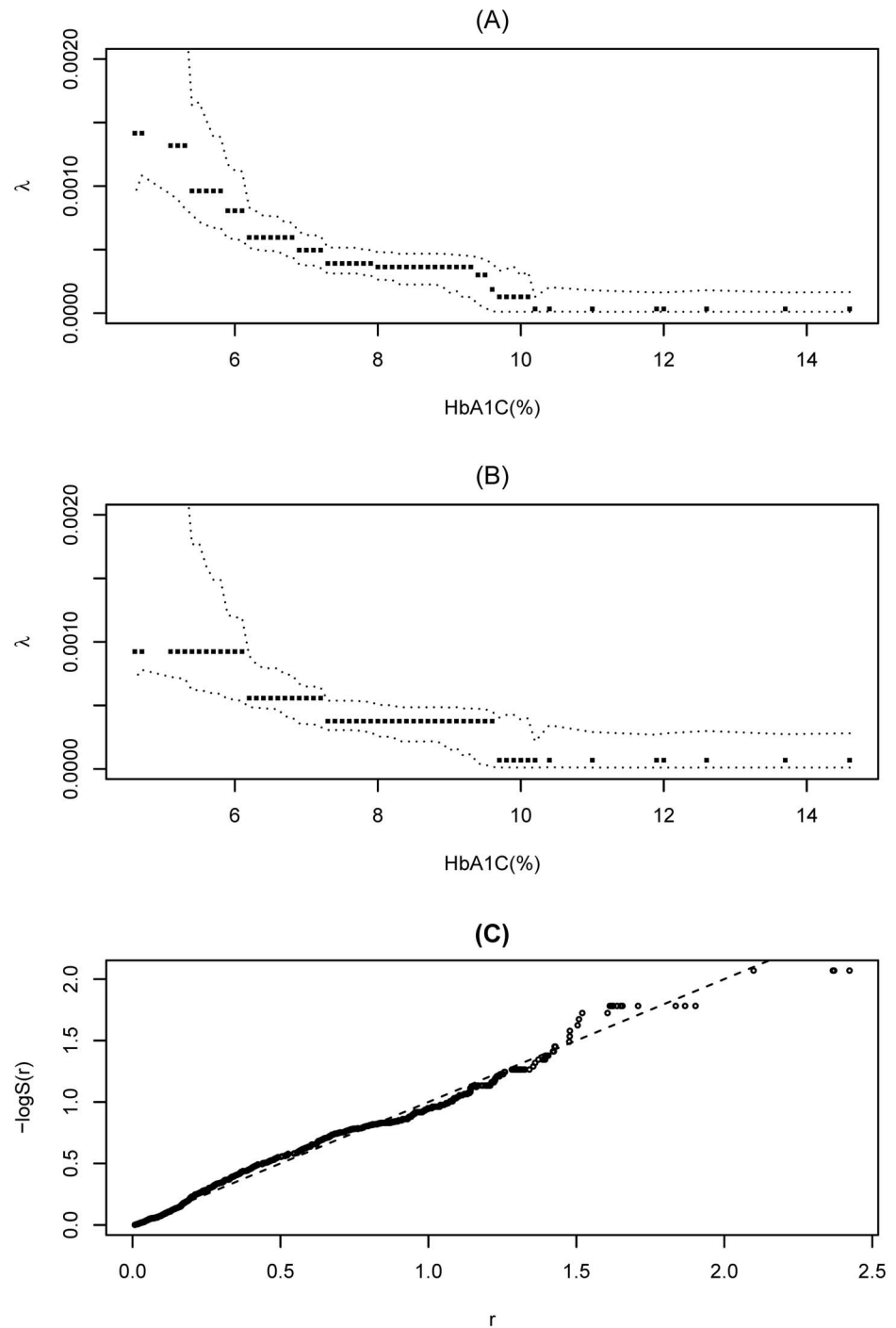
**Table 3.** Precision of parameter estimates with various event numbers and percent censored.

Event N	Sample N	$\gamma=2$	$x=$	0	1/3	2/3	1	4/3	5/3	2
			$\lambda=$	1	1	1	2	2	4	5
mean and mean squared error for the estimates of $\lambda$										
~500	1000	2.01		0.99	1.00	1.01	1.99	2.01	4.02	5.00
		0.0051		0.0043	0.0027	0.0030	0.0061	0.0065	0.040	0.13
	800	2.01		0.99	1.00	1.01	1.99	2.01	4.02	4.99
		0.0052		0.0037	0.0022	0.0025	0.0072	0.0073	0.044	0.16
	500	2.01		0.99	1.00	1.01	1.99	2.01	4.06	4.95
		0.0050		0.0021	0.0014	0.0017	0.0079	0.0085	0.072	0.22
~200	400	2.01		0.97	1.00	1.01	2.00	2.01	4.10	4.90
		0.014		0.013	0.0081	0.0093	0.015	0.016	0.11	0.39
	320	2.02		0.99	1.00	1.01	1.99	2.02	4.10	4.85
		0.013		0.0078	0.0053	0.0062	0.016	0.018	0.12	0.47
	200	2.03		0.99	1.00	1.01	2.00	2.02	4.14	4.86
		0.015		0.0047	0.0035	0.0055	0.018	0.022	0.18	0.59
~100	200	2.04		0.97	1.01	1.03	1.99	2.04	4.16	4.83
		0.030		0.023	0.016	0.022	0.039	0.043	0.22	0.66
	160	2.05		0.98	1.01	1.03	1.99	2.04	4.20	4.82
		0.030		0.019	0.012	0.016	0.034	0.042	0.27	0.65
	100	2.06		0.98	1.00	1.03	2.01	2.05	4.18	4.89
		0.032		0.0088	0.0070	0.014	0.041	0.055	0.32	1.28
~50	100	2.07		0.97	1.00	1.08	1.96	2.07	4.17	4.82
		0.069		0.043	0.044	0.085	0.11	0.13	0.35	1.36
	80	2.08		0.97	1.02	1.07	1.98	2.10	4.23	4.90
		0.072		0.032	0.027	0.057	0.11	0.14	0.53	1.65
	50	2.14		0.99	1.02	1.07	1.98	2.13	4.24	4.99
		0.091		0.021	0.020	0.050	0.11	0.24	0.79	2.60

For the mean and mean squared error columns, the first row is the mean and the second row is the mean squared error. Percent and range of censoring is the same as shown in [Table 2](#).

doi:10.1371/journal.pone.0113948.t003

The event is the first occurrence of severe hypoglycemia and the covariate of interest is the participant’s HbA1c at study entry. We applied the reduced isotonic regression with Weibull assumption and a nominal  $\alpha=0.05$  to the data and present the results in [Figure 3](#). Panel (A) is the regular isotonic regression without testing between steps and the resulting model has many small steps. When the testing between steps is added, a parsimonious model with only 3 change points (6.2, 7.3 and 9.6) is obtained. Model in panel (B) suggests that even though it is ideal to lower patients’ HbA1C level to as close to normal (5.6 or lower) as possible, we need to monitor the level closely when it crosses 9.6, 7.3, and 6.2 to avoid the occurrence of severe hypoglycemia. Overall significance of HbA1C is  $p=0.026$ , estimated from the permutation approach. The model estimate for the shape parameter is  $\hat{\gamma}=0.76$  ( $p=0.001$ ), suggesting that the hypoglycemic events



**Figure 3. Modeling HbA1c and risk of severe hypoglycemia.** (A) Regular isotonic regression without testing between steps. (B) Reduced isotonic regression with nominal  $\alpha=0.05$  (C) Cox-Snell residual plot of Model B. Dotted lines in (A) and (B) represent 95% Confidence Intervals of  $\lambda$ .

doi:10.1371/journal.pone.0113948.g003

tend to occur early in the implementation of the intensive therapy. The Cox-Snell residual plot (panel (C)) indicates that the Weibull assumption is valid.

## Discussion

Herein we demonstrated how reduced isotonic regression can be implemented in parametric time-to-event data analysis with survival time following an exponential or Weibull distribution.

As shown in the simulation studies, adding statistical testing between steps can reduce number of steps falsely introduced by noise. Although in the simulation example we chose  $\gamma=2$  for clarity of presentation, we have examined the scenario when  $\gamma=0.5$  and obtained similar results [14]. A  $\gamma < 1$  might be more representative of medical data of chronic diseases and the fact that our approach worked well with the DCCT data ( $\gamma=0.76$ ) is reassuring. In the DCCT example, the regular isotonic regression model produced too many change points and were not directly useful for the clinical practice. Although we could identify a couple change points from the regular model by eyeballing the figure, it is not systematic and very subjective. By using the statistical testing we were able to build a parsimonious model with only a few change points. Obviously, the nominal isotonic testing level  $\alpha$  will influence the number of change points. As the nominal  $\alpha$  becomes smaller the number of change points decreases. As the methodology allows for user's choices of the nominal  $\alpha$ , in real world data analysis, we recommend to start with a big  $\alpha$  at 1.0, i.e., no testing done between steps, to obtain an exploratory check of the association between the covariate and the outcome. After that, a smaller nominal  $\alpha$  can be applied to obtain a more parsimonious model with fewer change points for practical use. A methodological approach such as those used in choosing the smoothing parameter value in non-parametric data analysis can be developed to choose a single best nominal  $\alpha$ , however, it is beyond the scope of this paper with both a caveat and possible extension of the method given.

In health research or epidemiological studies, we often want to evaluate whether a covariate of interest is associated with the outcome independently of the effects of other covariates. This is usually achieved by adding (or adjusting for) other covariates known to be associated with the outcome in the model. In this case, we can add the known covariates to the algorithm and solve for them simultaneously with the covariate of interest. Estimates of these covariates can be solved in the same way as the shape parameter  $\gamma$  in the Weibull case, i.e., held as constants while solving for the parameters related to  $X$ .

## Supporting Information

**Appendix S1. The modified dynamic programming algorithm.**

[doi:10.1371/journal.pone.0113948.S001](https://doi.org/10.1371/journal.pone.0113948.S001) (PDF)

## Appendix S2. R programs developed for the reduced isotonic regression in survival analysis.

[doi:10.1371/journal.pone.0113948.S00](https://doi.org/10.1371/journal.pone.0113948.S00) (TXT)

## Acknowledgments

The authors would like to thank Michael J Schell and Gang Han for their critical review and insightful comments which refined the article.

## Author Contributions

Conceived and designed the experiments: YM YL JML. Performed the experiments: YM. Analyzed the data: YM. Contributed reagents/materials/analysis tools: YM. Wrote the paper: YM YL JML. Designed the software used for the analysis: YM YL.

## References

1. **Ancukiewicz M, Finkelstein DM, Schoenfeld DA** (2003) Modeling the relationship between continuous covariates and clinical events using isotonic regression. *Stat Med* 22: 3151–3159.
2. **Michael J, Schell MJ, Bahadur S** (1997) The Reduced monotonic regression method. *J Am Stat Assoc* 92: 128–135.
3. **Salanti G, Kurt U** (2005) A non-parametric framework for estimating threshold limit values. *BMC Med Res Methodol* 5: 36.
4. **Han G, Schell MJ, Kim J** (2014) Improved survival modeling in cancer research using a reduced piecewise exponential approach. *Stat Med* 33(1): 59–73.
5. **Lai Y** (2011) On the adaptive partition approach to the detection of multiple change-points. *PLoS ONE* Volume 6(5).
6. **Lai Y, Albert PS** (2014) Identifying multiple change points in a linear mixed effects model. *Stat Med* 33(6): 1015–28.
7. **The Diabetes Control and Complications Trial Research Group** (1993) The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N Engl J Med* Sep 30 329(14): 977–86.
8. **UK Prospective Diabetes Study (UKPDS) Group** (1998) Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). *Lancet* 352(9131): 837–53.
9. **Gehan EA, Thomas DG** (1969) The Performance of some two-sample tests in small samples with and without censoring. *Biometrika* 56: 127–132.
10. **Lee ET, Desu MM, Gehan EA** (1975) A Monte-Carlo study of the power of some two-sample tests. *Biometrika* 62: 425–432.
11. **Cox DR, Snell EJ** (1968) A General definition of Residuals. *J R Stat Soc Series B* 30: 248–275.
12. **Efron B** (1979) “Bootstrap Methods: Another Look at the Jackknife”. *Ann Stat* 7(1): 1–26.
13. **Lachin JM** (2010) *Biostatistical Methods: The Assessment of Relative Risks*. Second Edition. John Wiley & Sons, 672 p.
14. **Ma Y** (2010) *Flexible Isotonic Regression in Survival Data Analysis*. PhD Dissertation, The George Washington University. Available: <http://pqdtopen.proquest.com/pqdtopen/doc/749787764.html?FMT=AI> Accessed August 2010.