Impact of Delayed Event Time on Cox and Logistic Regression Models and Its Application to GWAS

By

Rebecca Irlmeier

Thesis

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

in

Biostatistics

December 12, 2020

Nashville, Tennessee

Approved:

Qingxia (Cindy) Chen, Ph.D.

Jake Hughey, Ph.D.

Contents

List of Tables

List of Figures

1 Introduction

1.1 Genome-wide and Phenome-wide Association Studies

Genome-wide association studies (GWAS) rose to popularity about 15 years ago and have become "the traditional approach for the discovery of genetic variations contributing to a multitude of complex human traits and diseases" [24]. GWAS is used to determine the genetic markers, usually single-nucleotide polymorphisms (SNPs), that contribute to a particular phenotype or disease of interest within a population of unrelated individuals [1]. To conduct GWAS, DNA is obtained from patients with and without a disease of interest, and each person's genome is scanned for millions of variants to identify SNPs of interest. One way to apply GWAS is to determine significant associations with phecodes, which are derived from billing codes from the International Classification of Diseases to represent a phenotype of interest, in the electronic health record (EHR) [5]. Here, GWAS generally assumes that all genetic variants being studied are equally likely to be associated with the phecode of interest, to maximize the ability of discovering unknown associations [24]. The SNPs that are found significantly more often in people with the disease than those without the disease are considered to be associated with the phenotype of interest [1]. The use of large cohorts and the evolution of GWAS to have the ability to assess millions of SNPs have led to the discovery of many unique significant genotype-phenotype associations [16].

1.2 Cox Regression and Logistic Regression

Survival analysis is a term used to describe the statistical methods utilized when analyzing time-to-event data. The outcome variable is the time until the occurrence of an event of interest, which may be death, relapse, recurrence, among others. The survival time ranges from the time origin to the occurrence of the event or the date of last contact (called censoring) [3]. A widely-used method in survival analysis is Cox (proportional hazards) regression, which is a semi-parametric survival model [3]. The effect size for Cox regression

is the hazard ratio, which is the estimate of the ratio of the hazard rate in the treatment group versus that of the control group. A major assumption of Cox regression is that of proportional hazards, where each hazard ratio is assumed to be constant over time [3].

Logistic regression is used when the outcome variable is dichotomous. It is a generalized linear model that uses a logistic function, logit, to link the probability of the binary outcome to the linear predictor function [14]. The effect size for logistic regression is the odds ratio, which is the estimate of the ratio of the odds of the event in the treatment group versus the odds of the event in the control group.

Generally, Cox regression is used with survival data and logistic regression is used with binary data. However, though using Cox regression would allow for more information to be incorporated, logistic regression can still be used in survival data if the time-to-event information is ignored. This may occur since logistic regression is more widely understood and less computationally expensive than Cox regression in analysis [22].

1.3 Real-World Motivation

Traditionally, genomic studies have used logistic regression models to analyze the genetic data linked to EHR data, but this method does not consider the longitudinal nature of EHR observations. Cases are typically defined as individuals who experienced the event of interest at any timepoint in their record, without taking into account the time at which the event occurred. To incorporate this, in addition to logistic regression models that completely ignore the event time [23], [12], [7], logistic regression models that adjust for the time-to-event have been employed [26], [11], [21], [17], as well as logistic regression models that adjust for EHR length [9]. The use of Cox regression, which can account for both the right censoring and left truncation that occurs in EHR data, has also been explored [9]. Previous work has shown that Cox regression is advantageous over logistic regression in genomic studies using the EHR, in which it was found that Cox regression increased the power to detect genotype-phenotype associations [9]. However, though GWAS of SNPs often include time-to-event

data, logistic regression is regularly used instead of Cox regression in analysis since it is less computationally expensive, despite some recent efforts to speed up the analysis for GWAS [22], [15], [2].

Another resistance of using the Cox model in EHR-based analysis is the concern of recorded time accuracy. The longitudinal nature of EHR data is useful in that it provides information regarding disease development and progression due to repeated clinical visits [13]. Individuals enter the healthcare system at various ages (left truncation) and may leave the system before they have an event (right censoring). This time-to-event information can be utilized in certain modeling techniques. However, due to the structure of EHR data, the time-to-event that is used in Cox regression may not always be accurate. In GWAS, an individual is considered a case if they have evidence of a phecode at some point in their record, and the time-to-event is the age at which they first show this phecode. If an individual has large gaps in their record, the age at which they first show the phecode on their record could potentially be older than the age at which they actually developed the phenotype. We refer to the age difference between when an individual actually develops the phecode and when the phecode shows up on the record as the delayed event time in the time-to-event information in the EHR. As Cox models use the time-to-event information directly, there may be concern on their validity in the presence of delayed event time, especially when compared to logistic regression models. On the other hand, it is known that the Score test for a simple Cox regression model with one binary exposure is equivalent to the log-rank test, a nonparametric rank-based approach, and hence, robust to the independent delayed event time on the observed time [25].

In this paper, we sought to determine the impact of delayed event time on the performance of Cox regression and logistic regression models in simulations and for identifying genotype-phenotype associations in genetic data linked to EHR data. We explore when the delayed event time is independent and when it depends on a confounder, non-confounder, and the exposure of interest. We showed that while logistic regression does not model the time-to-

event directly, various logistic regression models used in GWAS were more sensitive to the delayed event time scenarios than Cox regression. We begin in Section 2 by describing the motivation and methods used in the simulations and genomic study application. Section 3 discusses the simulation study, while Section 4 reviews the GWAS application. We end with a discussion in Section 5 and conclude in Section 6.

2 Methods

2.1 Modeling Schemes

We first define the Cox model and three commonly used logistic regression models used in GWAS studies. The models are fit with an exposure variable, z, and two types of covariates $\mathbf{x_1}$ and $\mathbf{x_2}$, where $\mathbf{x_1}$ is a $p \times 1$ vector of confounders for the exposure and $\mathbf{x_2}$ is a $q \times 1$ vector of covariates that is associated with the outcome but not with the exposure. Both simulations with and without left truncation, T_{lt}, are conducted. The observed time, T_{obs}, is the minimum of the event time, T_e, and the right censoring time, T_c, for each observation. E is the event indicator and is defined as $E = I(T_e < T_c)$. One Cox regression model and three logistic regression models used in the GWAS literature in the presence of right censoring are considered.

1. Cox proportional hazards regression model (Cox):

$$h(T_{obs}|z, \mathbf{x_1}, \mathbf{x_2}) = h_0(t)exp\{\beta_1 z + \boldsymbol{\beta}_2'\mathbf{x_1} + \boldsymbol{\beta}_3'\mathbf{x_2}\} \tag{1}$$

2. Logistic regression model (adjusting for time difference) (LRM_{obs}):

$$logit[P(E = 1|z, \mathbf{x_1}, \mathbf{x_2}, T_d)] = \beta_0 + \beta_1 z + \boldsymbol{\beta}_2'\mathbf{x_1} + \boldsymbol{\beta}_3'\mathbf{x_2} + \beta_4 f(T_d) \tag{2}$$

where $T_d = T_{obs}$.

3. Logistic regression model (without adjusting for time) (LRM_u):

$$logit[P(E=1|z,\mathbf{x_1},\mathbf{x_2})] = \beta_0 + \beta_1 z + \boldsymbol{\beta}_2'\mathbf{x_1} + \boldsymbol{\beta}_3'\mathbf{x_2} \tag{3}$$

4. Logistic regression model (adjusting for record length) (LRM_{rl}):

$$logit[P(E=1|z,\mathbf{x_1},\mathbf{x_2},T_{rl},T_c)] = \beta_0 + \beta_1 z + \boldsymbol{\beta}_2'\mathbf{x_1} + \boldsymbol{\beta}_3'\mathbf{x_2} + \beta_4 f(T_{rl}) \tag{4}$$

where $T_{rl} = T_c$ is the EHR length. Note that Model (4) is usually not considered as an alternative of the Cox model as, unlike EHR-based application, T_c is not observable for $E = 1$ in time-to-event applications. When $E = 0$, $T_d = T_c$ and hence Model (2) has the same expression as Model (4).

With the existence of left truncation, the Cox model adapting to left truncation is readily available [10]. T_d in LRM_{obs} became $T_d = T_{obs} - T_{lt}$ and LRM_u remained the same. In LRM_{rl}, $T_{rl} = T_c - T_{lt}$, so Model (4) became:

4. Logistic regression model (adjusting for record length) (LRM_{rl}):

$$logit[P(E=1|z,\mathbf{x_1},\mathbf{x_2},T_{rl},T_c)] = \beta_0 + \beta_1 z + \boldsymbol{\beta}_2'\mathbf{x_1} + \boldsymbol{\beta}_3'\mathbf{x_2} + \beta_4 T_{rl} + \beta_5 f(T_c) \tag{4}$$

In all models, β_1, the coefficient of the exposure, is the parameter of interest, and the unknown function $f(\cdot)$ is modeled using a cubic smoothing spline with three degrees of freedom.

2.2 Delayed Event Time Scenarios

2.2.1 Delayed Diagnosis

To better understand the motivation, consider the following example: suppose the event of interest is being diagnosed with a certain disease (phecode), and there are two individuals

who develop the disease at the same time. Depending on certain characteristics of the patients, such as their financial standing or insurance status, the patients are diagnosed at different times after developing the disease. A patient who does not have insurance may likely put off going to the doctor until it is necessary and be diagnosed later, while a patient with insurance may go to the doctor right away. The time difference between when a patient develops the tumor and is diagnosed with the disease (or the phecode shows up on their record) is the delayed event time, ϵ, which is being simulated in the models. Only positive delayed event time is considered; for example, if a patient develops the disease at age 40, the delayed event time can only occur after age 40 until diagnosis. Different delayed event time scenarios are considered, and specific examples of these scenarios are given in Section 2.3.

Before the delayed event time, ϵ, is incorporated, the true event time and true censoring time are denoted as T_e and T_c, respectively. The true observed time is thus $T_{obs} = min(T_e, T_c)$ and the event indicator is $E = I(T_e < T_c)$. In this simulation, the delayed event time is added to the event time only, and the observed time with delayed event time is the minimum of the true event time plus delayed event time and the true censoring time: $\widetilde{T}_{obs} = min(T_e + \epsilon, T_c)$. This leads to an event indicator of $\widetilde{E} = I(T_e + \epsilon < T_c)$. Due to the nature of this simulation, the delayed event time that is added to T_e can lead to three different cases that relate $\widetilde{T}_{obs}$ with T_{obs}, in which $\widetilde{E}$ does not always equal E. These cases are explained in Appendix A, but it should be noted that the magnitude of the proportion of misclassified events will change the relative performance of the models. In addition, if left truncation is present, there are occurrences of the simulated event time being less than the simulated left truncation time. In the research to evaluate the Cox model with left truncation, these occurrences are usually removed from the simulated dataset as they are considered as not meeting the criteria or not at risk [8], [20]. However, to mimic the application to the EHR, an observation in this situation is considered a control since they do not have the event of interest during their record, which is from left truncation time to right censoring time.

2.2.2 Baseline Shifted

Another type of delayed event time occurs when the baseline time is shifted by a fixed delayed event time, ϵ. For example, consider that we are interested in the time from cancer diagnosis to cancer mortality. If the diagnosis time is delayed such as in Section 2.2.1, both the times of cancer related death (T_e) and the last record of the patient (T_c) from diagnosis are reduced by the same delayed event time. Compared to delayed diagnosis, baseline shifted is less common in practice. As the example that motivates this scenario does not have a left truncation design, only censoring without truncation is considered.

In baseline shifted, the delayed event time is subtracted from both T_e and T_c to obtain the observed time with the delayed event time: $\bar{T}_{obs} = min(T_e - \epsilon, T_c - \epsilon)$. This leads to an event indicator of $\bar{E} = I(T_e - \epsilon < T_c - \epsilon)$, so $\bar{E} = E$ and $\bar{T}_{obs} = T_{obs} - \epsilon$ for every observation. Thus, the observations do not partition into different delayed event cases as described in Appendix A for delayed diagnosis.

2.3 Distribution of Delayed Event Time

Five delayed event time scenarios are examined in this study. We consider when there is no delayed event time, which can occur if a patient is diagnosed with a disease as soon as it develops (or the phecode shows up on the EHR). If the phecode of interest is an acute disease requiring an emergency visit, the diagnosis time is most likely accurate. We consider delayed event time independent of the exposure or covariates, which is caused by any factor of the patient that is not related to the exposure and other covariates, such as a delayed clinic visit due to scheduling. In addition, we consider when delayed event time is associated with the exposure directly, which can occur if the delay is related to a particular SNP that is being studied or a drug of interest in a clinical trial. Another delayed event time scenario is when the delay is associated with a confounder of exposure. If the delayed event time is caused by a disease being easier to diagnose in one sex over the other since it is more common in that sex, and sex is a confounder of the exposure of interest, the confounding

scenario occurs. Last, we consider when the delayed event time is associated with a covariate that is independent of the exposure. For example, someone with a lower income may take longer to go to the doctor and be diagnosed, but income is not associated with a SNP or drug of interest.

3 Simulation Study

3.1 Data-Generation Process

We simulated data for the delayed diagnosis scenario motivated in Section 2.2.1 and the baseline shifted scenario motivated in Section 2.2.2. Specifically, two covariates x_1 and x_2 were independently generated from Bernoulli with $p = 0.3$ and $\mathrm{N}(0.5, 0.4)$, respectively. The exposure, z, was simulated from a Bernoulli distribution with $p = [1 + exp(1.25 - x_1)]^{-1}$, i.e., x_1 is a confounder for z.

Different distributions for the event time and censoring time were considered. We first simulated the event time from Model (1) with baseline hazard generated from either exponential(0.001) or log-normal$(6.5, 1)$. The former model belongs to the accelerated failure time model while the latter does not. The regression coefficients for x_1 and x_2 were log(2) and the coefficient for z was varied to examine the type I error rate and power. The censoring time was simulated from $Unif(a_1, a_2)$, where a_1 and a_2 were specified to obtain different numbers of observations in each delayed event case as explained in Appendix A. We also simulated censoring time from a multivariable Cox regression model with baseline hazard generated from exponential(0.002), where the parametric component included x_1 and x_2 for non-informative censoring. The regression coefficients for x_1 and x_2 were log(2). We conducted the delayed diagnosis simulation both with and without left truncation. When left truncation was present, it was simulated from $Unif(50, 150)$. The mean event rate varied in the simulations depending on the delayed event case, the coefficient for z, and the censoring distribution.

In the simulation study, we considered sample size $n = 500$ and fit the four models as described in Section 2.1. To evaluate the type I error and power of these models, we conducted 5000 simulations, where the regression coefficient for z was rejected if the p-value was less than 0.05. We evaluated the type I error when the coefficient for z was simulated to be zero, and evaluated the power when the coefficient for z was simulated to be log(1.1), log(1.15), log(1.25), log(1.5) and log(2).

3.1.1 Delayed Event Time Scenarios

We simulated five delayed event time scenarios which added delayed event time, ϵ, to T_e. When there was no delayed event time, the value of ϵ was equal to zero. Independent delayed event time was simulated from $Unif(b_1, b_2)$. When the delayed event time was associated with the exposure, z, it was simulated from $Unif(c_1, c_2)$ and $Unif(c_2, c_3)$ for subjects exposed and not exposed, respectively. The same distributions were used when the delayed event time was associated with the confounder, but for subjects with $x_1 = 1$ and $x_1 = 0$, respectively. Delayed event time that was associated with the covariate, x_2, was simulated from log-normal$(d \times x_2, 1)$. The parameters b_1, b_2, c_1, c_2, c_3, and d were varied to obtain different numbers of observations in each delayed event case as explained in Appendix A and explore different magnitudes of delayed event time.

3.2 Simulation Results

We used a series of simulations to compare the Cox regression and logistic regression models under different delayed event time scenarios to mimic the application in the EHR data. Since the effect sizes of the two methods are not equivalent (i.e., hazard ratios and odds ratios), the performance of the four models was compared in terms of type I error and power in the presence of delayed event time. We also evaluated the bias of the estimation for exposure for the Cox model only.

3.2.1 Simulation 1 - Delayed Diagnosis Results

The results of Simulation 1 (with left truncation) based on the five different delayed event time scenarios, when the event time is simulated from a Cox model with baseline hazard from an exponential distribution and the censoring time is simulated from a uniform distribution, are shown in Figure 1 and Figure 2. In Figure 1, in all of the delayed event time scenarios, except for when ϵ is associated with z, Model 1 (Cox) performs either the same or better than two of the logistic regression models. When the coefficient for z is zero, the type I error rate is near the nominal rate and the power increases as the effect size of the exposure increases. Models 3 (LRM_u) and 4 (LRM_{rl}) perform well and similarly. However, Model 2 (LRM_{obs}), performs substantially worse in terms of power than the other three models. This is because in Model 2 (LRM_{obs}), the effect of z leaks through T_d when it has a non-null effect.

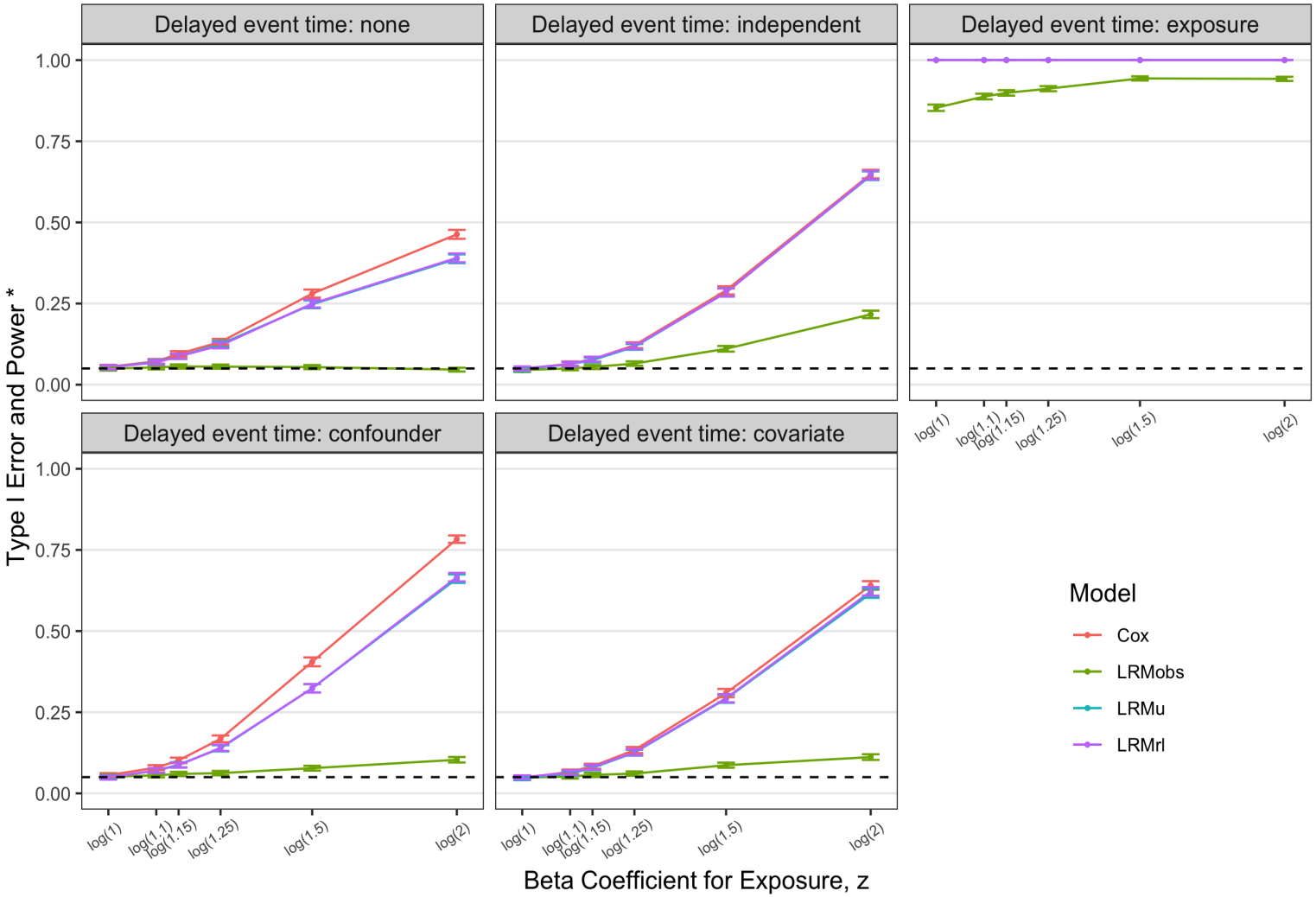


Figure 1: Results from Simulation 1 when the event time was generated from a Cox model with baseline hazard from an exponential distribution, the censoring time was generated from a uniform distribution, and there was left truncation. The parameters led to a large number of observations with a misclassified event status (detailed in Appendix A).

* Type I error evaluated at log(1). Power evaluated at log(1.1), log(1.15), log(1.25), log(1.5), log(2).

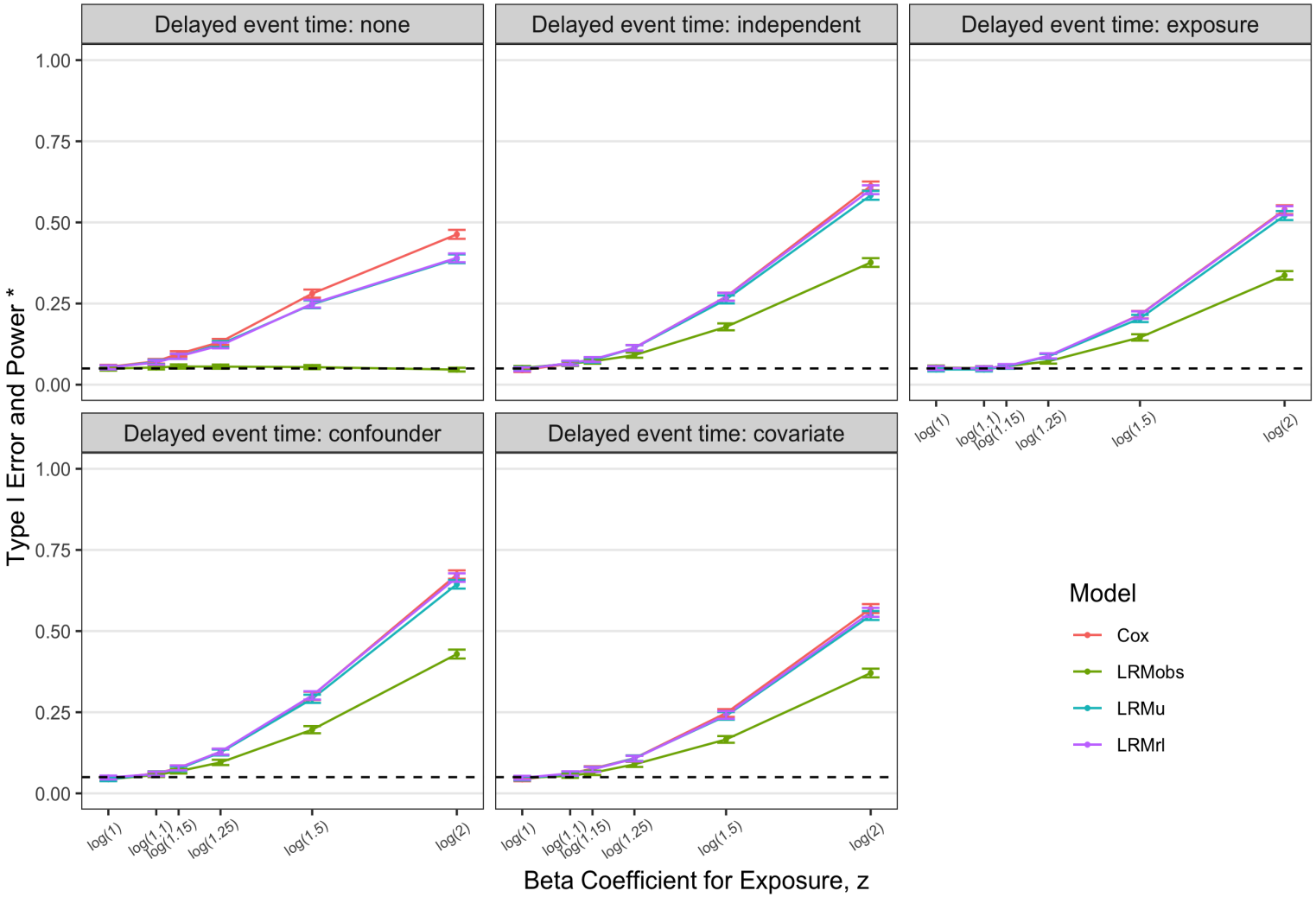


Figure 2: Results from Simulation 1 when the event time was generated from a Cox model with baseline hazard from an exponential distribution, the censoring time was generated from a uniform distribution, and there was left truncation. The parameters led to a small number of observations with a misclassified event status (detailed in Appendix A).

* Type I error evaluated at log(1). Power evaluated at log(1.1), log(1.15), log(1.25), log(1.5), log(2).

The only difference in the data-generation for Figure 1 and Figure 2 is the magnitude of the delayed event time, ϵ, and the censoring distribution to vary the proportion of subjects with a misclassified event status (see Appendix C, Figure 7a: Delayed Event Case 2 and Figure 7b: Delayed Event Case 2). The misclassification occurs when the delayed event time causes an observation who is originally a case to become a control. In Figure 1, the only delayed event time scenario in which none of the models have an acceptable performance is when ϵ depends on z. This scenario is almost impossible in a GWAS study, but it is likely for other EHR-based applications, such as drug repurposing [27]. When the proportion of misclassified events is high, all of the models are invalid, so a new method is needed with additional data collected to model the delayed event time in this scenario. However, when the proportion of misclassified subjects is small, the type I error of the models when the delayed event time depends on the exposure is controlled, as can be seen in Figure 2.

The corresponding figures for when the event-time is generated from a Cox model with baseline hazard from a log-normal distribution are in Appendix C, Figure 8 and Figure 9. These results are consistent to those previously described, with the exception that the type I error is slightly inflated in Figure 9 when the delayed event time depends on z.

When the censoring distribution is modified to be simulated from a Cox model that depends on x_1 and x_2 (Figure 3, Appendix C: Figure 10), Model 1 (Cox) always performs the best in terms of power, followed by Model 4 (LRM_{rl}). The difference in power between these two models is larger than when the censoring distribution is independent of the covariates. Again, when the delayed event time is associated with z and there is a high proportion of subjects with misclassified events, all four models are invalid.

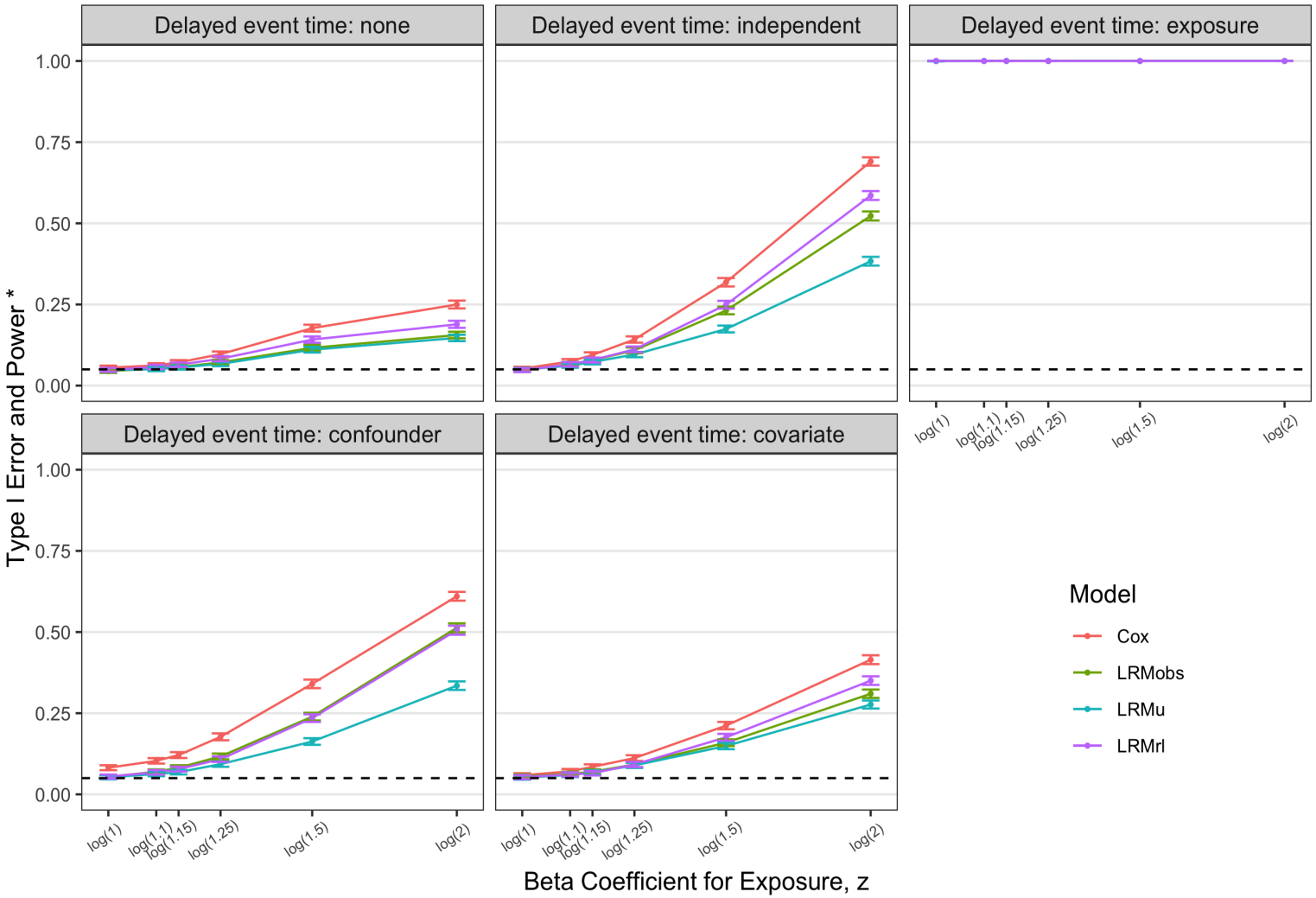


Figure 3: Results from Simulation 1 when the event time was generated from a Cox model with baseline hazard from an exponential distribution, the censoring time was generated from a Cox model with baseline hazard from an exponential distribution that depended on x, and there was left truncation. The parameters led to a large number of observations with a misclassified event status (detailed in Appendix A).

* Type I error evaluated at log(1). Power evaluated at log(1.1), log(1.15), log(1.25), log(1.5), log(2).

These results hold when there is no left truncation (Appendix C, Figures 13-18), with the exception that the type I error is always inflated when there is a small number of subjects with misclassified events.

We also evaluated the bias of the regression coefficient estimate for exposure, z, from Model 1 (Cox) in the different delayed event time scenarios and combinations of event time and censoring time distributions. When there is a small proportion of subjects with a misclassified event status, as in Figure 1, the bias ranged from -0.4104 to 0.2444. When the proportion is large, the bias for when the delayed event time depends on z increases slightly, widening the range from -0.4673 to 2.5374. However, the bias only has a magnitude greater than 0.5 when the delayed event time depends on the exposure, which we already stated requires a new method with additional data used to model the delayed event time. Including the observations who have a simulated event time earlier than their simulated left truncation time in the analysis as controls (since their time-to-event information would not be known in application) slightly increases the magnitude of the bias for z, compared to when we did the same simulations while removing these observations from the analysis (results not shown).

3.2.2 Simulation 2 - Baseline Shifted Results

The results for Simulation 2 are relatively consistent with those from Simulation 1 (see Appendix C, Figures 21-26). Again, in all the delayed event time scenarios except for when ϵ depends on z, Model 1 (Cox) performs the same or better than the logistic regression models in terms of statistical power, usually followed by Model 4 (LRM_{rl}). Model 2 (LRM_{obs}) generally performs the worst, though it is about the same as Model 3 (LRM_u) when the censoring distribution depends on the covariates. The models are invalid when the delayed event time is associated with z, with inflated type I error.

4 Genomic Study Application

4.1 Genomic Study Application Data-Generation Process

To determine the impact of delayed event time on Cox and logistic regression models in a real-data application, we conducted GWAS in the genetic data linked to EHR data [6]. We selected ten phenotypes in which to compare the ability of Cox and logistic regression models to detect known genotype-phenotype associations in the presence of simulated delayed event time, which are listed in Appendix B, Table 1. These phenotypes were chosen before the analysis was performed. Cases for each phenotype were defined as individuals who had the phecode in the EHR on two distinct dates, and controls as those who did not have the phecode in the EHR. Left truncation, T_{lt}, was present in the EHR and corresponded to the age at the first visit in the healthcare system. The observed age, T_{obs}, which was the event age for cases, T_e, and the right censoring age for controls, T_c, corresponded to the age on the second date of receiving the phecode (cases) or the age at the last visit (controls).

Since we aimed to understand the impact of delayed event time and the robustness of the models in the empirical data, we assumed the event age in the EHR data was the "true" event age for each patient who was a case (i.e., there was no delayed event time in the EHR). We simulated delayed event time, and it was added to the event time only, corresponding to Simulation 1 in which $\widetilde{T}_{obs} = min(T_e + \epsilon, T_c)$. Due to the structure of the EHR data, since only patients who had the phecode of consideration on two distinct dates had an age for the event time, the delayed event time was only added to the cases. Thus, a case could become a control in the presence of delayed event time if $T_e + \epsilon > T_c$, where T_c corresponded to their last ever visit. A control remained a control.

In the genomic application, we considered the four models described in Section 2.1. For all four models, the linear component included genotype and the first four components of genetic ancestry. The model either included a term for biological sex or the data were restricted to females or males only depending on the phenotype. Model 1 (Cox) used the

counting process formulation with left truncation and the observed age. Model 2 (LRM_{obs}) included additional terms for the age difference (as a cubic spline with three degrees of freedom), which was the difference between the observed age and the left truncation age, $T_d = T_{obs} - T_{lt}$. Model 3 (LRM_u) included no additional terms concerning age. Model 4 (LRM_{rl}) included additional terms for age at the last visit (as a cubic spline with three degrees of freedom) and the record length, which was the difference in age between the first ever and last ever visits.

4.1.1 Delayed Event Time Scenarios

We considered four delayed event time scenarios to add to the event age for each phenotype. We considered delayed event time that depended on significant SNPs. For a particular phecode, all the significant SNPs at the $P \leq 5 \times 10^{-8}$ significance level were selected. The number of significant SNPs ranged from 1 to 298 among the ten phecodes used. The coding for the SNP was the allele count. If a patient had at least one of the alleles, the delayed event time was simulated from $Unif(min = 0, max = 0.5)$. If the patient had none of the alleles, the delayed event time was simulated from $Unif(min = 0.5, max = 1)$. The scale of age was years, so values of delayed event time equal to 0.5 and 1 corresponded to 6 months and 1 year, respectively. We also considered delayed event time that depended on non-significant SNPs. For each phecode, the same number of SNPs that were significant at the $P \leq 5 \times 10^{-8}$ significance level were randomly sampled from the non-significant SNPs. The delayed event time was simulated in the same way as for the significant SNPs. We considered delayed event time that depended on sex, which was only used in phecodes that were associated with both females and males. In this case, it was simulated from $Unif(min = 0, max = 0.5)$ for females and $Unif(min = 0.5, max = 1)$ for males. Last, we simulated independent delayed event time from $Unif(min = 0, max = 1)$ for all patients.

4.2 Genomic Study Application Results

To study the robustness of Cox and logistic regression models in the presence of delayed event time, we compared the four models with every delayed event time scenario using genetic data linked to the EHR. A cohort of 49,792 individuals of European ancestry was used, and ten phenotypes were defined from the EHR. For each model and delayed event time combination, GWAS was run on 795,850 common SNPs. The Manhattan plots for the ten phenotypes are shown in Appendix C, Figures 29-38. Model 1 (Cox) generally detected the most significant SNPs, followed by Model 4 (LRM_{rl}), especially for common phenotypes.

Based on the results found in the simulations and Hughey et al [9], we calculated the true positive and true negative rates (TPRs and TNRs) of detecting associations for the models with each delayed event time scenario, using the Cox regression model with no delayed event time as the gold standard. Thus, the SNPs found to be significant at either the $P \leq 5 \times 10^{-8}$ or $P \leq 1 \times 10^{-5}$ significance level by Model 1 (Cox) with no delayed event time are considered the "true" associations at the respective significance level. The average TPRs and TNRs from all ten phecodes and corresponding 95% confidence intervals are reported in Appendix B, Table 25. The average TNRs are very high for all the model and delayed event time combinations due to the relatively small number of significant SNPs compared to the 795,850 SNPs that were analyzed in the GWAS. The average TPRs for each model and delayed event time combination can be visualized in Figure 4. Model 1 (Cox) and Model 4 (LRM_{rl}) have the highest true positive rates, even in the presence of delayed event time. The individual TPRs and TNRs for the phecodes are provided in Appendix B (Tables 3-22).

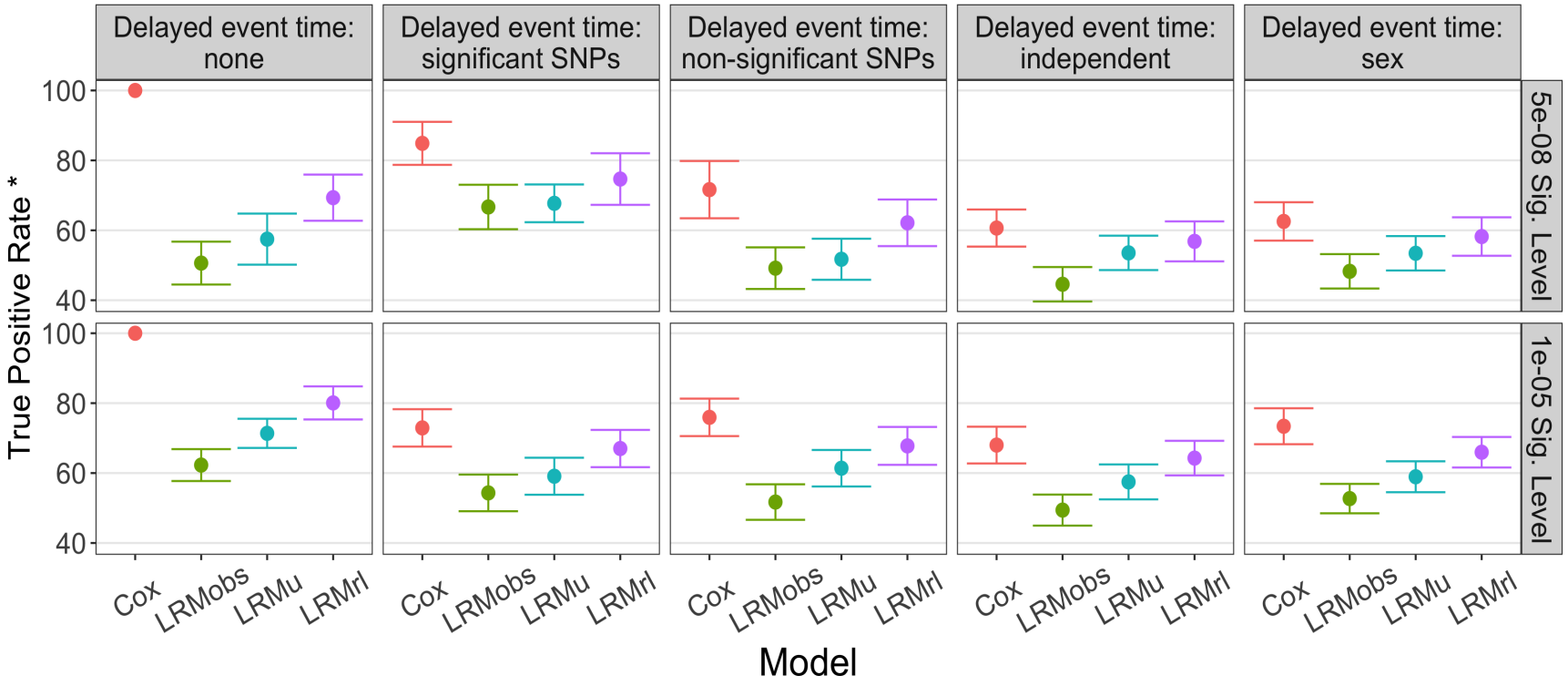


Figure 4: Average true positive rates for detecting significant SNPs from all ten phecodes for each model and delayed event time combination, using Model 1 (*Cox*) with no delayed event time as the gold standard. This application corresponds to the delayed diagnosis set-up.

* Based on Model 1 (*Cox*) - no delayed event time

We also plotted the p-values of Model 1 (Cox) with no delayed event time against the p-values of the remaining model and delayed event time combinations in Figure 5. The gray points indicate true positive or true negative SNPs, while the colored points represent false positive and false negative SNPs. The ideal performance of a model would be to have as few false positives (red points) and false negatives (blue points) as possible. In addition, the true negative and true positive SNPs (gray points) should follow closely along the 45° line. Model 1 (Cox) and Model 4 (LRM_{rl}) have the fewest false positive and false negative points, even in the presence of delayed event time. The true positive/true negative points follow most closely to the 45° line for Model 1 (Cox) compared to the logistic regression models, within each respective delayed event time scenario. The corresponding figures for the individual phecodes are given in Appendix C (Figures 39-48).

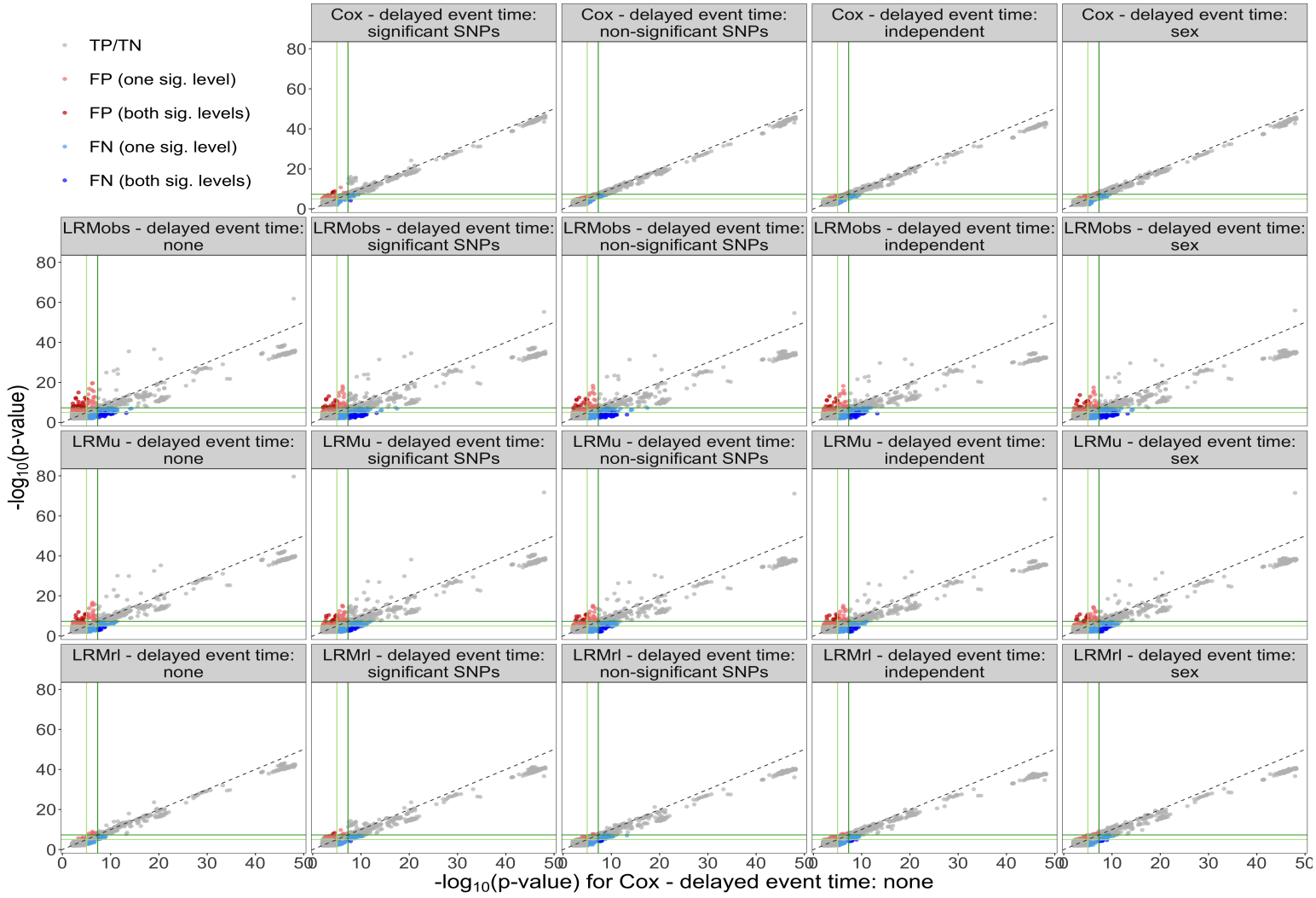


Figure 5: False positive and false negative SNPs for each model and delayed event time combination, using Model 1 (Cox) with no delayed event time as the gold standard, for all ten phecodes. Dark green lines correspond to $P \leq 5 \times 10^{-8}$ and light green lines correspond to $P \leq 1 \times 10^{-5}$.

We also used the GWAS results from each model/delayed event time combination for the ten phenotypes to determine each method's ability of detecting known associations from the NHGRI-EBI GWAS Catalog [4]. The results are shown in Figure 6, where each graph shows the four models for a particular delayed event time scenario. It can be seen that Model 1 (Cox) has the highest relative sensitivity compared to the other models across a range of p-value cutoffs, even with delayed event time. Model 4 (LRM_{rl}) generally seems to perform better than Models 2 (LRM_{obs}) and 3 (LRM_u) in detecting known associations.

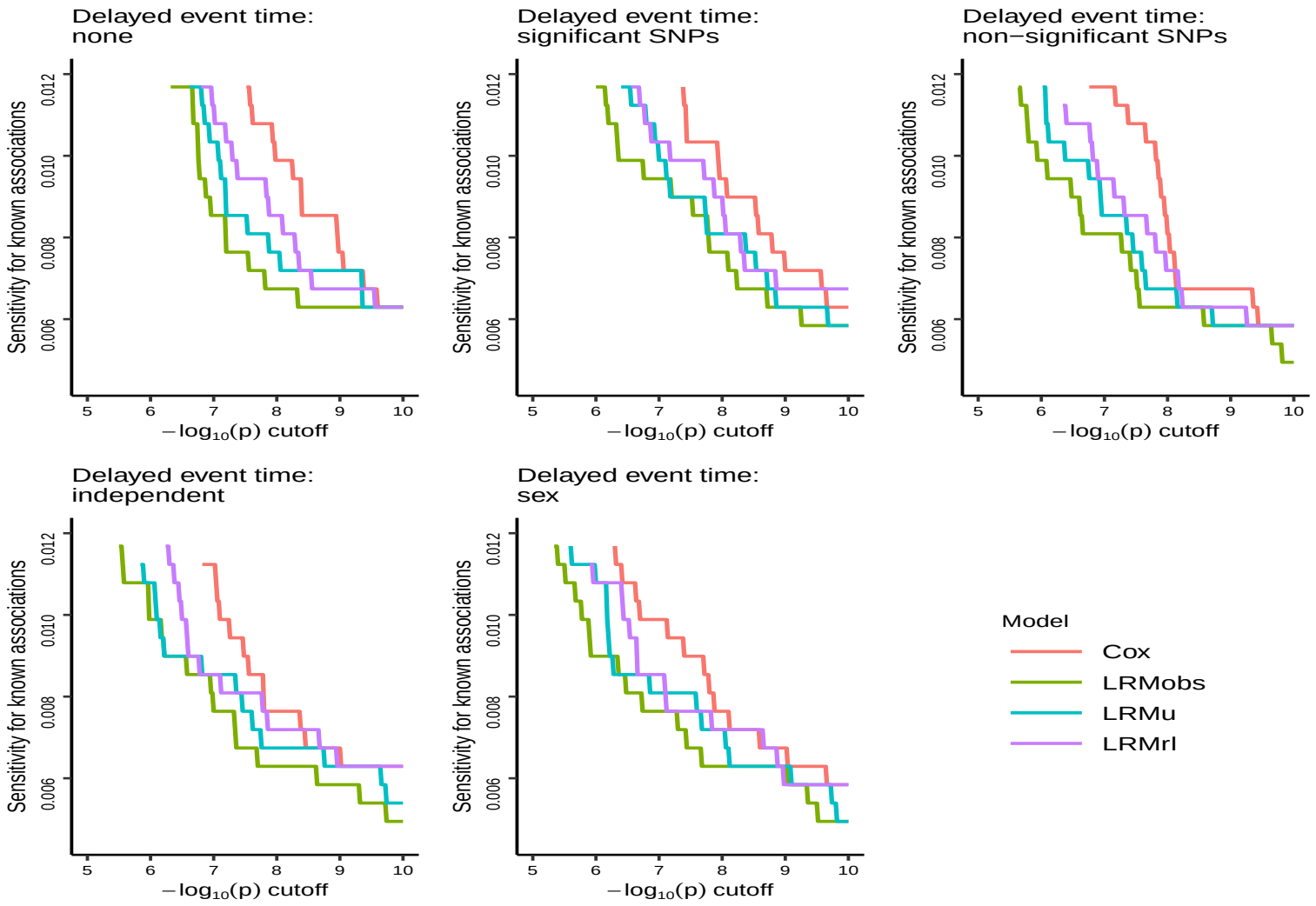


Figure 6: Sensitivity of each model and delayed event time combination for detecting known genotype-phenotype associations.

5 Discussion

Although rare in our motivating study, we considered when the censoring distribution was simulated from a Cox model that depended on both the covariates and exposure for comparison (see Figures 11, 12, 19, 20, 27, 28 in Appendix C). In this situation, informative censoring was observed. When the coefficient for z in the censoring distribution was $log(2)$, and thus there was moderate correlation between the exposure and censoring distribution, the bias in the coefficient for z was minimum. As the coefficient of z in the censoring distribution increased, thus leading to more severe informative censoring, the bias in the coefficient of z increased, even in the absence of delayed event time (see Appendix B, Table 2). At the existence of informative censoring, many methods were developed to extend the Cox model under different scenarios including, [18], [19], [27], [28], among many others, which is outside the scope of this paper and will be investigated in the future. Though not ideal, the Cox regression model without adjusting for informative censoring still outperformed the logistic regression models. Finally, the performance of LRM_obs and $LRMu$ deteriorated in this scenario, with uncontrolled type I error rate and decreasing power with increased effect size (see Appendix C, Figures ,12, 20, 28).

In the simulation study, we assumed that observations who had a simulated event time less than their left truncation time were a control, since the time-to-event information would not be known in application. Compared to otherwise identical analyses where these observations were removed in the simulations, the bias in the beta coefficient for exposure, z, increased slightly in magnitude when these patients were kept and treated as controls. This extends to the EHR application, where if a patient had the phenotype of interest before entry into a healthcare site, it would not be shown on the record. Due to our definition of a case, which was having the phecode of interest on two distinct dates, patients who had the event of interest before their first age in the record were considered controls, unless they had a recurrence during their record. A limitation of this study is the use of a single-site EHR, which restricted us to only consider patients as cases if they showed the phecode twice after

entering the record. If the first distinct date of showing the phecode on the record occurred at the first age in the record, this could be indicative of a patient who actually developed the phecode before entering the single-site EHR. This limitation could be alleviated if multi-site EHRs were combined.

There are limitations with the use of both Model 1 (Cox) with no delayed event time and the GWAS Catalog as the gold standards in the GWAS application. We made the assumption that the associations found to be significant by Model 1 (Cox) with no delayed event time were the truth based on previous work [9] and the results of the simulation study. These associations were used to calculate the true positive and true negative rates of the other model/delayed event time combinations, which could be misleading if some of the significant associations are incorrect. In addition, the use of the GWAS Catalog as the gold standard to determine the sensitivity of the Cox models is limiting, since most of the known genotype-phenotype associations were found by logistic or linear regression. Thus, it does not apply directly to associations found by Cox regression. All of the methods showed low sensitivity due to being underpowered for detecting the associations. However, it is promising that both the simulations and the GWAS application indicated that Cox regression has the best performance in detecting genotype-phenotype associations, even with these limitations.

Lastly, we did not determine the exact magnitude of delayed event time that would be acceptable in the EHR in order for the Cox model to continue to outperform the logistic regression models, as our main goal was to explore the impact of delayed event time on the performance of the models in general. However, in the simulations, we varied the parameters when simulating the delayed event time to obtain different numbers of observations with a misclassified event status, which led to different ranges of delayed event time magnitude. For example, when there was a small number of misclassified events and the delayed event time depended on the confounder, we set $c_1 = 20$ and $c_3 = 60$ days. To increase the proportion of misclassified events, we set $c_1 = 60$ and $c_3 = 1400$ days. Increasing the magnitude of the delayed event time caused all the methods to be invalid when the delayed event time

depended on the exposure, as explained in Section 3.2.1. However, for the other delayed event time scenarios, even when the magnitude of the delayed event time was large, the Cox regression model performed either the same or better as the logistic regression models in terms of statistical power, and the type I error rate was controlled. This gives some insight into the impact of the magnitude of delayed event time on the performance of the models.

6 Conclusion

Based on the use of both simulations and empirical data, we found that while logistic regression does not model the time-to-event directly, various logistic regression models used in the literature were more sensitive to delayed event time than Cox regression. The simulations showed that Cox regression had similar or modest improvement in statistical power over logistic regression at controlled type I error. These results were supported by the empirical data, where the Cox models steadily had the highest sensitivity to detect known genotype-phenotype associations under all scenarios of delayed event time. In the presence of delayed event time scenarios that might exist in EHRs, Cox regression outperformed the logistic regression models commonly used in genomic studies. Among the three logistic regression models, the logistic regression model that adjusts for record length, Model 4 (LRM_{rl}), is the preferred modeling scheme to use.

As stated in the Introduction, previous work has already shown the advantages of Cox regression over logistic regression in many scenarios [22], [26], including for use in genomic studies that utilize the EHR [9]. Our primary focus in this study was to determine if Cox regression still outperformed logistic regression when the time-to-event information in the EHR is incorrect, which we found to be true. This indicates that Cox regression is the most robust modeling scheme to delayed event time. Thus, even if time-to-event information is inaccurate, Cox regression may improve our ability to determine the significant genetic constitutes for a variety of diseases.

References

[1] Genome-wide association studies fact sheet. https://www.genome.gov/about-genomics/fact-sheets/Genome-Wide-Association-Studies-Fact-Sheet, Aug 2020.

[2] Wenjian Bi, Lars G Fritsche, Bhramar Mukherjee, Sehee Kim, and Seunggeun Lee. A fast and accurate method for genome-wide time-to-event data analysis and its application to uk biobank. *The American Journal of Human Genetics*, 107(2):222–233, 2020.

[3] A Brembilla, A Olland, M Puyraveau, G Massard, F Mauny, and PE Falcoz. Use of the cox regression analysis in thoracic surgical research. *Journal of thoracic disease*, 10:3891–3896, 2018.

[4] A Buniello, JAL MacArthur, M Cerezo, LW Harris, J Hayhurst, C Malangone, and et al. The nhgri-ebi gwas catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res*, 47:D1005–D1012, 2019.

[5] JC Denny, L Bastarache, and DM Roden. Phenome-wide association studies as a tool to advance precision medicine. *Annual review of genomics and human genetics*, 17:353–373, 2016.

[6] JC Denny, SL Van Driest, W-Q Wei, and DM Roden. The influence of big (clinical) data and genomics on precision medicine and drug development. *Clinical pharmacology and therapeutics*, 103:409–418, 2018.

[7] D Harold, R Abraham, P Hollingworth, R Sims, and et al. Genome-wide association study identifies variants at clu and picalm associated with alzheimer's disease. *Nature genetics*, 41(10):1088–1093, 2009.

[8] Penelope P Howards, Irva Hertz-Picciotto, and Charles Poole. Conditions for bias from differential left truncation. *American journal of epidemiology*, 165(4):444–452, 2007.

[9] J Hughey, S Rhoades, D Fu, L Bastarache, JC Denny, and Q Chen. Cox regression increases power to detect genotype-phenotype associations in genomic studies using the electronic health record. *BMC Genomics*, 20, 12 2019.

[10] JP Klein and ML Moeschberger. Survival analysis: Techniques for censored and truncated data. 2003.

[11] A Miyashita, A Koike, G Jun, and et al. Sorl1 is genetically associated with late-onset alzheimer's disease in japanese, koreans and caucasians. *PloS one*, 8(4):e58618, 2013.

[12] N Mullins, TB Bigdeli, and et al. Gwas of suicide attempt in psychiatric disorders and association with major depression polygenic risk scores. *JCO clinical cancer informatics*, 176(8):651–660, 2019.

[13] SA Pendergrass and DC Crawford. Using electronic health records to generate phenotypes for research. current protocols in human genetics. *Current Protocols in Human Genetics*, 100:e80, 2019.

[14] CYJ Peng, KL Lee, and Ingersoll GM. An introduction to logistic regression analysis and reporting. *The Journal of Educational Research*, 96:3–14, 2002.

[15] Abbas A Rizvi, Ezgi Karaesmen, Martin Morgan, Leah Preus, Junke Wang, Michael Sovic, Theresa Hahn, and Lara E Sucheston-Campbell. gwasurvivr: an r package for genome-wide survival analysis. *Bioinformatics*, 35(11):1968–1970, 2019.

[16] JR Robinson, JC Denny, DM Roden, and SL Van Driest. Genome-wide and phenome-wide approaches to understand variable drug actions in electronic health records. *Clinical and translational science*, 11:112–122, 2018.

[17] A Rogaeva, Y Meng, JH Lee, Y Gu, and et al. The neuronal sortilin-related receptor sorl1 is genetically associated with alzheimer disease. *Nature genetics*, 39(2):168–177, 2007.

[18] Andrea Rotnitzky and James M Robins. Semiparametric regression estimation in the presence of dependent censoring. *Biometrika*, 82(4):805–820, 1995.

[19] Douglas E Schaubel and Guanghui Wei. Double inverse-weighted estimation of cumulative treatment effects under nonproportional hazards and dependent censoring. *Biometrics*, 67(1):29–38, 2011.

[20] Enrique F Schisterman, Stephen R Cole, Aijun Ye, and Robert W Platt. Accuracy loss due to selection bias in cohort studies with left truncation. *Paediatric and perinatal epidemiology*, 27(5):491–502, 2013.

[21] J Simón-Sánchez, JJ van Hilten, B van de Warrenburg, B Post, and et al. Genome-wide association study confirms extant pd risk loci among the dutch. *European journal of human genetics*, 19(6):655–661, 2011.

[22] JR Staley, E Jones, S Kaptoge, AS Butterworth, Sweeting MJ, AM Wood, and et al. A comparison of cox and logistic regression for use in genome-wide association studies of cohort and case-cohort design. *European Journal of Human Genetics*, 25:854–862, 2017.

[23] H Syed, AL Jorgensen, and AP Morris. Evaluation of methodology for the analysis of 'time-to-event' data in pharmacogenomic genome-wide association studies. *Pharmacogenomics*, 17(8):907–915, 2016.

[24] Hamzah Syed. *Design, evaluation and application of methodology and software for time-to-event outcomes in pharmacogenetic genome-wide association studies.* PhD thesis, University of Liverpool, 2018.

[25] TM Therneau and PM Grambsch. Modeling survival data: Extending the cox model. 2000.

[26] JB van der Net, ACJW Janssens, MJC Eijkemans, JJP Kastelein, Sijbrands EJG, and EW Steyerberg. Cox proportional hazards models have more statistical power than logistic regression models in cross-sectional genetic association studies. *European Journal of Human Genetics*, 16:1111–1116, 2008.

[27] Margaret C Wu and Raymond J Carroll. Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*, pages 175–188, 1988.

[28] Yue Zhao, Amy H Herring, Haibo Zhou, Mirza W Ali, and Gary G Koch. A multiple imputation method for sensitivity analyses of time-to-event data with possibly informative censoring. *Journal of biopharmaceutical statistics*, 24(2):229–253, 2014.

7 Appendix A: Delayed Event Time Scenarios

7.1 Simulation Notation

The following are considered without left truncation:

- T_e is the true event time
- $T_c = T_{rl}$ is the true censoring time and the record length
- $T_{obs} = min(T_e, T_c) = T_d$ is the true observed time and time difference
- $E = I(T_e < T_c)$ is the event indicator
- ϵ is the delayed event time

7.1.1 Simulation 1 - Delayed Diagnosis

- $\widetilde{T}_d = \widetilde{T}_{obs} = min(T_e + \epsilon, T_c)$ is the observed time with delayed event time
- $\widetilde{E} = I(T_e + \epsilon < T_c)$ is the event indicator with delayed event time

There are three cases in which subjects can be partitioned once delayed event time is added to their true event time:

Case 1. $\widetilde{E} = 0$ and $T_c < T_e$:

$\widetilde{T}_d = T_c = min(T_e, T_c) = T_d$

$\widetilde{E} = E = 0$

Case 2. $\widetilde{E} = 0$ and $T_e < T_c < T_e + \epsilon$:

$\widetilde{T}_d = T_c < T_e + \epsilon$

$T_d = T_e$

$\Rightarrow \widetilde{T}_d \neq T_d \Rightarrow \widetilde{T}_d - T_d < \epsilon$

$\widetilde{E} = 0$ and $E = 1 \Longrightarrow$ delayed event time leads to a misclassfied event status

To estimate the proportion of observations in Case 2:

a) If T_c is uniformly distributed:

$$\begin{aligned} P(\text{Case 2}) &= P(T_e < T_c < T_e + \epsilon) \\ &= P(T_e < T_c | T_c < T_e + \epsilon) P(T_c < T_e + \epsilon) \\ &= [1 - P(T_c < T_e | T_c < T_e + \epsilon)] P(T_c < T_e + \epsilon) \\ &= [1 - P(T_c < T_e | T_c < T_e + \epsilon)] P(\widetilde{E} = 0) \end{aligned}$$

where $P(\widetilde{E} = 0)$ is the censoring rate in the data

$$\begin{aligned} &= \left[1 - \frac{T_e}{T_e + \epsilon}\right] P(\widetilde{E} = 0) \\ &= \left[\frac{\epsilon}{T_e + \epsilon}\right] P(\widetilde{E} = 0) \\ &= \left[\frac{R_e}{1 + R_e}\right] P(\widetilde{E} = 0) \end{aligned}$$

where $R_e = \dfrac{\epsilon}{T_e}$ is the relative delayed event time to the true event time

b) If T_c has density function $g(t)$, which is estimable in our application:

$$P(\text{Case 2}) = -\int_0^T \int_{t-\epsilon}^t g(\mu) d\mu d\widetilde{S}(t)$$

where $\widetilde{S}(t)$ is the survival function for $T_e^* = T_e + \epsilon$

Case 3. $\widetilde{E} = 1$ and $T_e + \epsilon < T_c$:

$\widetilde{T}_d = T_e + \epsilon$

$T_d = T_e$

$\Rightarrow \widetilde{T}_d = T_d + \epsilon$

$\widetilde{E} = E = 1$

7.1.2 Simulation 2 - Baseline Shifted

- $\bar{T}_d = min(T_e - \epsilon, T_c - \epsilon)$ is the observed time with delayed event time
- $\bar{E} = I(T_e - \epsilon < T_c - \epsilon)$ is the event indicator with delayed event time
- $\bar{E} = E$
- $\bar{T}_d = T_d - \epsilon$

7.2 Simulation 1 - Delayed Diagnosis

7.2.1 *Cox*

The likelihood function is:

$$\ell_i(\widetilde{T}_{d_i}, \widetilde{E_i}) = f(\widetilde{T}_{d_i})^{\widetilde{E}_i} \times S(\widetilde{T}_{d_i})^{1-\widetilde{E}_i}$$

Case 1. $= S(T_{d_i})^{1-E_i} = \ell_i(T_{d_i}, E_i)$

Case 2. $= S(\widetilde{T}_{d_i}) \neq f(T_{d_i}) = \ell_i(T_{d_i}, E_i)$

Case 3.

$$\begin{aligned}
&= f(T_{d_i} + \epsilon) = exp\{logf(T_{d_i} + \epsilon)\} \\
&= exp\left\{logf(T_{d_i}) + \frac{f'(T^*_{d_i})}{f(T^*_{d_i})}\epsilon\right\} \\
&\text{where } T^*_{d_i} \in [T_{d_i}, T_{d_i} + \epsilon] \\
&= f(T_{d_i})exp\left\{\frac{f'(T^*_{d_i})}{f(T^*_{d_i})}\epsilon\right\} \\
&= \ell_i(T_{d_i}, E_i)exp\left\{\frac{f'(T^*_{d_i})}{f(T^*_{d_i})}\epsilon\right\}
\end{aligned}$$

The log-likelihood function of $(\widetilde{T}_{d_i}, \widetilde{E_i})$ is:

Case 1. $= logS(T_{d_i}) = log\ell_i(T_{d_i}, E_i)$

Case 2. $= logS(\widetilde{T}_{d_i}) \neq logf(T_{d_i}) = log\ell_i(T_{d_i}, E_i)$

Case 3. $= log\ell_i(T_{d_i}, E_i) + \frac{f'(T^*_{d_i})}{f(T^*_{d_i})}\epsilon$

7.2.2 LRM_{obs}

Assuming we model $f(\widetilde{T}_d)$ linearly:

$$
\begin{aligned}
&logit\left[P(\widetilde{E}=1|Z,\mathbf{X},\widetilde{T}_d)\right] \\
&= \beta_0 + \beta_1 Z + \boldsymbol{\beta}_2'\mathbf{X} + \beta_3\widetilde{T}_d \\
&= \begin{cases} \beta_0 + \beta_1 Z + \boldsymbol{\beta}_2'\mathbf{X} + \beta_3 T_d & \text{if case 1} \\ \beta_0 + \beta_1 Z + \boldsymbol{\beta}_2'\mathbf{X} + \beta_3 T_c & \text{if case 2} \\ \beta_0 + \beta_1 Z + \boldsymbol{\beta}_2'\mathbf{X} + \beta_3 T_d + \beta_3\epsilon & \text{if case 3} \end{cases} \\
&\Rightarrow \begin{cases} = logit\left[P(E=1|Z,\mathbf{X},\widetilde{T}_d)\right] & \text{if case 1} \\ \neq logit\left[P(E=1|Z,\mathbf{X},\widetilde{T}_d)\right] & \text{if case 2} \\ = logit\left[P(E=1|Z,\mathbf{X},\widetilde{T}_d)\right] & \text{if case 3} \end{cases} \\
&\Rightarrow \begin{cases} \text{if case 1} \Rightarrow \begin{cases} = \beta_0 + \beta_1 Z + \boldsymbol{\beta}_2'\mathbf{X} + \beta_3 T_d & \text{if no delayed event time} \\ = \beta_0 + \beta_1 Z + \boldsymbol{\beta}_2'\mathbf{X} + \beta_3 T_d & \text{if delayed event time independent} \\ = \beta_0 + \beta_1 Z + \boldsymbol{\beta}_2'\mathbf{X} + \beta_3 T_d & \text{if delayed event time depends on Z} \\ = \beta_0 + \beta_1 Z + \boldsymbol{\beta}_2'\mathbf{X} + \beta_3 T_d & \text{if delayed event time depends on } \mathbf{X} \end{cases} \\ \text{if case 2} \Rightarrow \neq logit\left[P(E=1|Z,\mathbf{X},\widetilde{T}_d)\right] \\ \text{if case 3} \Rightarrow \begin{cases} = \beta_0 + \beta_1 Z + \boldsymbol{\beta}_2'\mathbf{X} + \beta_3 T_d & \text{if no delayed event time} \\ = (\beta_0 + \beta_3\epsilon) + \beta_1 Z + \boldsymbol{\beta}_2'\mathbf{X} + \beta_3 T_d & \text{if delayed event time independent} \\ = \beta_0 + (\beta_1 Z + \beta_3 g[Z]) + \boldsymbol{\beta}_2'\mathbf{X} + \beta_3 T_d & \text{if delayed event time depends on Z} \\ = \beta_0 + \beta_1 Z + (\boldsymbol{\beta}_2'\mathbf{X} + \beta_3 g[\mathbf{X}]) + \beta_3 T_d & \text{if delayed event time depends on } \mathbf{X} \end{cases} \end{cases}
\end{aligned}
$$

where $g[Z]$ is the delayed event time as a function of Z and $g[\mathbf{X}]$ is the delayed event time as a function of $\mathbf{X}$. Then we have:

$$
\begin{aligned}
&logit\left[P(\widetilde{E}=1|Z,\mathbf{X},\widetilde{T}_d)\right] \\
&\Rightarrow \begin{cases}
\text{if case 1} \Rightarrow \begin{cases}
= logit\left[P(E=1|Z,\mathbf{X},T_d)\right] & \text{if no delayed event time} \\
= logit\left[P(E=1|Z,\mathbf{X},T_d)\right] & \text{if delayed event time independent} \\
= logit\left[P(E=1|Z,\mathbf{X},T_d)\right] & \text{if delayed event time depends on Z} \\
= logit\left[P(E=1|Z,\mathbf{X},T_d)\right] & \text{if delayed event time depends on } \mathbf{X}
\end{cases} \\
\text{if case 2} \Rightarrow \neq logit\left[P(E=1|Z,\mathbf{X},T_d)\right] \\
\text{if case 3} \Rightarrow \begin{cases}
= logit\left[P(E=1|Z,\mathbf{X},T_d)\right] & \text{if no delayed event time} \\
\neq logit\left[P(E=1|Z,\mathbf{X},T_d)\right] & \text{if delayed event time independent} \\
\neq logit\left[P(E=1|Z,\mathbf{X},T_d)\right] & \text{if delayed event time depends on Z} \\
\neq logit\left[P(E=1|Z,\mathbf{X},T_d)\right] & \text{if delayed event time depends on } \mathbf{X}
\end{cases}
\end{cases}
\end{aligned}
$$

7.2.3 LRM_u

$$
\begin{aligned}
&logit\left[P(\widetilde{E}=1|Z,\mathbf{X})\right] \\
&= \beta_0 + \beta_1 Z + \boldsymbol{\beta}_2' \mathbf{X} \\
&\Rightarrow \begin{cases}
= logit\left[P(E=1|Z,\mathbf{X})\right] & \text{if case 1} \\
\neq logit\left[P(E=1|Z,\mathbf{X})\right] & \text{if case 2} \\
= logit\left[P(E=1|Z,\mathbf{X})\right] & \text{if case 3}
\end{cases}
\end{aligned}
$$

7.2.4 LRM_{rl}

Assuming we model $f(\widetilde{T}_{rl})$ linearly:

$$
\begin{aligned}
&logit\left[P(\widetilde{E}=1|Z,\mathbf{X},T_{rl})\right] \\
&= \beta_0 + \beta_1 Z + \boldsymbol{\beta}_2'\mathbf{X} + \beta_3 T_{rl} \\
&\Rightarrow \begin{cases} = logit\left[P(E=1|Z,\mathbf{X},T_{rl})\right] & \text{if case 1} \\ \neq logit\left[P(E=1|Z,\mathbf{X},T_{rl})\right] & \text{if case 2} \\ = logit\left[P(E=1|Z,\mathbf{X},T_{rl})\right] & \text{if case 3} \end{cases}
\end{aligned}
$$

7.3 Simulation 2 - Baseline Shifted

7.3.1 LRM_{obs}

Assuming we model $f(\bar{T}_d)$ linearly:

$$
\begin{aligned}
&logit\left[P(\bar{E}=1|Z,\mathbf{X},\bar{T}_d)\right] \\
&= \beta_0 + \beta_1 Z + \boldsymbol{\beta}_2'\mathbf{X} + \beta_3\bar{T}_d \\
&= \beta_0 + \beta_1 Z + \boldsymbol{\beta}_2'\mathbf{X} + \beta_3(T_d - \epsilon) \\
&= \beta_0 + \beta_1 Z + \boldsymbol{\beta}_2'\mathbf{X} + \beta_3 T_d - \beta_3\epsilon \\
&\Rightarrow \begin{cases}
= \beta_0 + \beta_1 Z + \boldsymbol{\beta}_2'\mathbf{X} + \beta_3 T_d & \text{if no delayed event time} \\
= (\beta_0 - \beta_3\epsilon) + \beta_1 Z + \boldsymbol{\beta}_2'\mathbf{X} + \beta_3 T_d & \text{if delayed event time independent} \\
= \beta_0 + (\beta_1 Z - \beta_3 g[Z]) + \boldsymbol{\beta}_2'\mathbf{X} + \beta_3 T_d & \text{if delayed event time depends on Z} \\
= \beta_0 + \beta_1 Z + (\boldsymbol{\beta}_2'\mathbf{X} - \beta_3 g[\mathbf{X}]) + \beta_3 T_d & \text{if delayed event time depends on } \mathbf{X}
\end{cases}
\end{aligned}
$$

where $g[Z]$ is the delayed event time as a function of Z and $g[\mathbf{X}]$ is the delayed event time as a function of $\mathbf{X}$. Then we have:

$$
\Rightarrow \begin{cases}
= logit\left[P(E=1|Z,\mathbf{X},T_d)\right] & \text{if no delayed event time} \\
\neq logit\left[P(E=1|Z,\mathbf{X},T_d)\right] & \text{if delayed event time independent} \\
\neq logit\left[P(E=1|Z,\mathbf{X},T_d)\right] & \text{if delayed event time depends on Z} \\
\neq logit\left[P(E=1|Z,\mathbf{X},T_d)\right] & \text{if delayed event time depends on } \mathbf{X}
\end{cases}
$$

7.3.2 LRM_u

$$
\begin{aligned}
&logit\left[P(\bar{E}=1|Z,\mathbf{X})\right] \\
&= \beta_0 + \beta_1 Z + \beta_2'\mathbf{X} \\
&= logit\left[P(E=1|Z,\mathbf{X})\right]
\end{aligned}
$$

7.3.3 LRM_{rl}

Assuming we model $f(\bar{T}_{rl})$ linearly, where $\bar{T}_{rl} = T_{rl} - \epsilon = T_c - \epsilon$:

$$
\begin{aligned}
&logit\left[P(\bar{E}=1|Z,\mathbf{X},T_{rl})\right] \\
&= \beta_0 + \beta_1 Z + \beta_2'\mathbf{X} + \beta_3(T_{rl}-\epsilon) \\
&\Rightarrow \begin{cases}
= \beta_0 + \beta_1 Z + \boldsymbol{\beta}_2'\mathbf{X} + \beta_3 T_{rl} & \text{if no delayed event time} \\
= (\beta_0 - \beta_3\epsilon) + \beta_1 Z + \boldsymbol{\beta}_2'\mathbf{X} + \beta_3 T_{rl} & \text{if delayed event time independent} \\
= \beta_0 + (\beta_1 Z - \beta_3 g[Z]) + \boldsymbol{\beta}_2'\mathbf{X} + \beta_3 T_{rl} & \text{if delayed event time depends on Z} \\
= \beta_0 + \beta_1 Z + (\boldsymbol{\beta}_2'\mathbf{X} - \beta_3 g[\mathbf{X}]) + \beta_3 T_{rl} & \text{if delayed event time depends on } \mathbf{X}
\end{cases}
\end{aligned}
$$

where $g[Z]$ is the delayed event time as a function of Z and $g[\mathbf{X}]$ is the delayed event time as a function of $\mathbf{X}$. Then we have:

$$
\Rightarrow \begin{cases}
= logit\left[P(E=1|Z,\mathbf{X},T_{rl})\right] & \text{if no delayed event time} \\
\neq logit\left[P(E=1|Z,\mathbf{X},T_{rl})\right] & \text{if delayed event time independent} \\
\neq logit\left[P(E=1|Z,\mathbf{X},T_{rl})\right] & \text{if delayed event time depends on Z} \\
\neq logit\left[P(E=1|Z,\mathbf{X},T_{rl})\right] & \text{if delayed event time depends on } \mathbf{X}
\end{cases}
$$

8 Appendix B: Additional Tables

Phenotype	Phecode	No Delayed Event Time Event Rate	Significant SNPs Event Rate	Non-significant SNPs Event Rate	Independent Event Rate	Sex Event Rate
Cancer of bronchus; lung	165.1	2.74%	1.95%	2.06%	1.78%	1.74%
Cancer of prostate *	185	6.52%	6.11%	5.80%	5.78%	-
Hypothyroidism	244	11.74%	10.94%	10.93%	10.64%	10.73%
Type 2 diabetes	250.2	14.33%	13.23%	13.23%	12.79%	12.73%
Vitamin D deficiency	261.4	6.72%	6.16%	6.11%	6.02%	6.12%
Hypercholesterolemia	272.11	9.99%	9.64%	9.75%	9.64%	9.62%
Insomnia	327.4	4.46%	4.00%	4.14%	4.11%	4.11%
Myocardial infarction	411.2	5.61%	4.54%	4.72%	4.66%	4.59%
Coronary atherosclerosis	411.4	16.83%	15.46%	15.51%	14.99%	14.89%
Atrial fibrillation	427.21	9.93%	8.65%	8.66%	8.3%	8.19%

Table 1: Phecodes used in the GWAS application. Includes information about the phenotype, phecode, and the event rate for each delayed event time scenario.

* Analysis performed for males only.

	Coefficient of z for T_c model		
Coefficient of z for T_e model	$ln(2)$	$ln(3)$	$ln(\exp\{2\})$
$ln(1)$	-0.020	-0.097	-6.733
$ln(1.1)$	-0.038 (-0.401)	-0.103 (-1.081)	-6.349 (-66.616)
$ln(1.15)$	-0.050 (-0.358)	-0.104 (-0.746)	-6.014 (-43.032)
$ln(1.25)$	-0.075 (-0.334)	-0.098 (-0.439)	-5.639 (-25.273)
$ln(1.5)$	-0.129 (-0.317)	-0.135 (-0.332)	-4.824 (-11.897)
$ln(2)$	-0.254 (0.367)	-0.239 (-0.345)	-3.910 (-5.641)

Table 2: Bias (relative bias) for the β coefficient of z from Model 1 (Cox) with no delayed event time when the correlation between the coefficient for z in the event time and the coefficient for z in the censoring time increases. This is related to informative censoring. These are the results from Simulation 1 when the event time was generated from a Cox model with baseline hazard from an exponential distribution, the censoring time was generated from a Cox model with baseline hazard from an exponential distribution that depended on x and z, and there was left truncation.

$P \leq 5 \times 10^{-8}$	# Significant SNPs	True Positive *	True Negative *
No delayed event time			
Cox	12	-	-
LRM_{obs}	0	0% (0/12)	100% (795838/795838)
LRM_{u}	4	33.33% (4/12)	100% (795838/795838)
LRM_{rl}	0	0% (0/12)	100% (795838/795838)
Delayed event time depends on significant SNPs			
Cox	19	100% (12/12)	99.9991% (795831/795838)
LRM_{obs}	15	100% (12/12)	99.9996% (795835/795838)
LRM_{u}	17	100% (12/12)	99.9994% (795833/795838)
LRM_{rl}	14	91.67% (11/12)	99.9996% (795835/795838)
Delayed event time depends on non-significant SNPs			
Cox	4	33.33% (4/12)	100% (795838/795838)
LRM_{obs}	0	0% (0/12)	100% (795838/795838)
LRM_{u}	1	8.33% (1/12)	100% (795838/795838)
LRM_{rl}	0	0% (0/12)	100% (795838/795838)
Delayed event time is independent			
Cox	0	0% (0/12)	100% (795838/795838)
LRM_{obs}	0	0% (0/12)	100% (795838/795838)
LRM_{u}	0	0% (0/12)	100% (795838/795838)
LRM_{rl}	0	0% (0/12)	100% (795838/795838)
Delayed event time depends on sex			
Cox	0	0% (0/12)	100% (795838/795838)
LRM_{obs}	0	0% (0/12)	100% (795838/795838)
LRM_{u}	0	0% (0/12)	100% (795838/795838)
LRM_{rl}	0	0% (0/12)	100% (795838/795838)

Table 3: Number of significant SNPs, true positive rates, and true negative rates for each model and delayed event time combination, using Model 1 (Cox) with no delayed event time as the gold standard, for cancer of bronchus; lung (phecode 165.1). The results are shown for the $P \leq 5 \times 10^{-8}$ significance level.

* Based on Model 1 (Cox) - no delayed event time

$P \leq 1 \times 10^{-5}$	# Significant SNPs	True Positive *	True Negative *
No delayed event time			
Cox	29	-	-
LRM_{obs}	22	68.97% (20/29)	99.9997% (795819/795821)
LRM_{u}	24	79.31% (23/29)	99.9999% (795820/795821)
LRM_{rl}	20	62.07% (18/29)	99.9997% (795819/795821)
Delayed event time depends on significant SNPs			
Cox	54	55.17% (16/29)	99.9952% (795783/795821)
LRM_{obs}	52	51.72% (15/29)	99.9954% (795784/795821)
LRM_{u}	51	51.72% (15/29)	99.9955% (795785/795821)
LRM_{rl}	46	48.28% (14/29)	99.996% (795789/795821)
Delayed event time depends on non-significant SNPs			
Cox	23	51.72% (15/29)	99.999% (795813/795821)
LRM_{obs}	19	41.38% (12/29)	99.9991% (795814/795821)
LRM_{u}	17	41.38% (12/29)	99.9994% (795816/795821)
LRM_{rl}	20	41.38% (12/29)	99.999% (795813/795821)
Delayed event time is independent			
Cox	20	48.28% (14/29)	99.9992% (795815/795821)
LRM_{obs}	13	13.79% (4/29)	99.9989% (795812/795821)
LRM_{u}	15	27.59% (8/29)	99.9991% (795814/795821)
LRM_{rl}	14	20.69% (6/29)	99.999% (795813/795821)
Delayed event time depends on sex			
Cox	21	44.83% (13/29)	99.999% (795813/795821)
LRM_{obs}	19	13.79% (4/29)	99.9981% (795806/795821)
LRM_{u}	17	17.24% (5/29)	99.9985% (795809/795821)
LRM_{rl}	14	13.79% (4/29)	99.9987% (795811/795821)

Table 4: Number of significant SNPs, true positive rates, and true negative rates for each model and delayed event time combination, using Model 1 (Cox) with no delayed event time as the gold standard, for cancer of bronchus; lung (phecode 165.1). The results are shown for both the $P \leq 1 \times 10^{-5}$ significance level.

* Based on Model 1 (Cox) - no delayed event time

$P \leq 5 \times 10^{-8}$	# Significant SNPs	True Positive *	True Negative *
No delayed event time			
Cox	8	-	-
LRM_{obs}	3	37.5% (3/8)	100% (795842/795842)
LRM_u	0	0% (0/8)	100% (795842/795842)
LRM_{rl}	5	62.5% (5/8)	100% (795842/795842)
Delayed event time depends on significant SNPs			
Cox	6	75% (6/8)	100% (795842/795842)
LRM_{obs}	3	37.5% (3/8)	100% (795842/795842)
LRM_u	0	0% (0/8)	100% (795842/795842)
LRM_{rl}	3	37.5% (3/8)	100% (795842/795842)
Delayed event time depends on non-significant SNPs			
Cox	5	62.5% (5/8)	100% (795842/795842)
LRM_{obs}	3	37.5% (3/8)	100% (795842/795842)
LRM_u	0	0% (0/8)	100% (795842/795842)
LRM_{rl}	4	50% (4/8)	100% (795842/795842)
Delayed event time is independent			
Cox	0	0% (0/8)	100% (795842/795842)
LRM_{obs}	0	0% (0/8)	100% (795842/795842)
LRM_u	0	0% (0/8)	100% (795842/795842)
LRM_{rl}	0	0% (0/8)	100% (795842/795842)

Table 5: Number of significant SNPs, true positive rates, and true negative rates for each model and delayed event time combination, using Model 1 (Cox) with no delayed event time as the gold standard, for cancer of prostate (phecode 185). The results are shown for the $P \leq 5 \times 10^{-8}$ significance level.

* Based on Model 1 (Cox) - no delayed event time

$P \leq 1 \times 10^{-5}$	# Significant SNPs	True Positive *	True Negative *
No delayed event time			
Cox	37	-	-
LRM_{obs}	32	51.35% (19/37)	99.9984% (795800/795813)
LRM_{u}	24	56.76% (21/37)	99.9996% (795810/795813)
LRM_{rl}	30	64.86% (24/37)	99.9992% (795807/795813)
Delayed event time depends on significant SNPs			
Cox	31	75.68% (28/37)	99.9996% (795810/795813)
LRM_{obs}	27	48.65% (18/37)	99.9989% (795804/795813)
LRM_{u}	24	51.35% (19/37)	99.9994% (795808/795813)
LRM_{rl}	28	64.86% (24/37)	99.9995% (795809/795813)
Delayed event time depends on non-significant SNPs			
Cox	34	67.57% (25/37)	99.9989% (795804/795813)
LRM_{obs}	32	40.54% (15/37)	99.9979% (795796/795813)
LRM_{u}	32	51.35% (19/37)	99.9984% (795800/795813)
LRM_{rl}	24	51.35% (19/37)	99.9994% (795808/795813)
Delayed event time is independent			
Cox	29	62.16% (23/37)	99.9992% (795807/795813)
LRM_{obs}	30	35.14% (13/37)	99.9979% (795796/795813)
LRM_{u}	28	43.24% (16/37)	99.9985% (795801/795813)
LRM_{rl}	23	48.65% (18/37)	99.9994% (795808/795813)

Table 6: Number of significant SNPs, true positive rates, and true negative rates for each model and delayed event time combination, using Model 1 (Cox) with no delayed event time as the gold standard, for cancer of prostate (phecode 185). The results are shown for the $P \leq 1 \times 10^{-5}$ significance level.

* Based on Model 1 (Cox)- no delayed event time

$P \leq 5 \times 10^{-8}$	# Significant SNPs	True Positive *	True Negative *
No delayed event time			
Cox	231	-	-
LRM_{obs}	106	44.59% (103/231)	99.9996% (795616/795619)
LRM_{u}	126	54.11% (125/231)	99.9999% (795618/795619)
LRM_{rl}	194	83.55% (193/231)	99.9999% (795618/795619)
Delayed event time depends on significant SNPs			
Cox	239	92.21% (213/231)	99.9967% (795593/795619)
LRM_{obs}	124	48.48% (112/231)	99.9985% (795607/795619)
LRM_{u}	145	57.14% (132/231)	99.9984% (795606/795619)
LRM_{rl}	207	84.42% (195/231)	99.9985% (795607/795619)
Delayed event time depends on non-significant SNPs			
Cox	233	94.37% (218/231)	99.9981% (795604/795619)
LRM_{obs}	116	48.92% (113/231)	99.9996% (795616/795619)
LRM_{u}	140	56.71% (131/231)	99.9989% (795610/795619)
LRM_{rl}	210	87.45% (202/231)	99.999% (795611/795619)
Delayed event time is independent			
Cox	225	88.74% (205/231)	99.9975% (795599/795619)
LRM_{obs}	133	52.81% (122/231)	99.9986% (795608/795619)
LRM_{u}	142	57.14% (132/231)	99.9987% (795609/795619)
LRM_{rl}	162	66.67% (154/231)	99.999% (795611/795619)
Delayed event time depends on sex			
Cox	268	95.67% (221/231)	99.9941% (795572/795619)
LRM_{obs}	142	53.68% (124/231)	99.9977% (795601/795619)
LRM_{u}	162	58.87% (136/231)	99.9967% (795593/795619)
LRM_{rl}	235	91.34% (211/231)	99.997% (795595/795619)

Table 7: Number of significant SNPs, true positive rates, and true negative rates for each model and delayed event time combination, using Model 1 (Cox) with no delayed event time as the gold standard, for hypothyroidism (phecode 244). The results are shown for the $P \leq 5 \times 10^{-8}$ significance level.

* Based on Model 1 (Cox) - no delayed event time

$P \leq 1 \times 10^{-5}$	# Significant SNPs	True Positive *	True Negative *
No delayed event time			
Cox	731	-	-
LRM_{obs}	434	56.22% (411/731)	99.9971% (795096/795119)
LRM_{u}	491	66.21% (484/731)	99.9991% (795112/795119)
LRM_{rl}	622	84.4% (617/731)	99.9994% (795114/795119)
Delayed event time depends on significant SNPs			
Cox	742	91.11% (666/731)	99.9904% (795043/795119)
LRM_{obs}	464	57.73% (422/731)	99.9947% (795077/795119)
LRM_{u}	540	64.98% (475/731)	99.9918% (795054/795119)
LRM_{rl}	644	79.62% (582/731)	99.9922% (795057/795119)
Delayed event time depends on non-significant SNPs			
Cox	753	92.48% (676/731)	99.9903% (795042/795119)
LRM_{obs}	446	55.4% (405/731)	99.9948% (795078/795119)
LRM_{u}	527	65.53% (479/731)	99.994% (795071/795119)
LRM_{rl}	637	79.62% (582/731)	99.9931% (795064/795119)
Delayed event time is independent			
Cox	731	88.92% (650/731)	99.9898% (795038/795119)
LRM_{obs}	488	60.33% (441/731)	99.9941% (795072/795119)
LRM_{u}	587	69.63% (509/731)	99.9902% (795041/795119)
LRM_{rl}	671	82.49% (603/731)	99.9914% (795051/795119)
Delayed event time depends on sex			
Cox	779	91.79% (671/731)	99.9864% (795011/795119)
LRM_{obs}	524	62.65% (458/731)	99.9917% (795053/795119)
LRM_{u}	617	71.82% (525/731)	99.9884% (795027/795119)
LRM_{rl}	725	86.87% (635/731)	99.9887% (795029/795119)

Table 8: Number of significant SNPs, true positive rates, and true negative rates for each model and delayed event time combination, using Model 1 (Cox) with no delayed event time as the gold standard, for hypothyroidism (phecode 244). The results are shown for the $P \leq 1 \times 10^{-5}$ significance level.

* Based on Model 1 (Cox) - no delayed event time

$P \leq 5 \times 10^{-8}$	# Significant SNPs	True Positive *	True Negative *
No delayed event time			
Cox	298	-	-
LRM_{obs}	153	48.66% (145/298)	99.999% (795544/795552)
LRM_{u}	196	61.74% (184/298)	99.9985% (795540/795552)
LRM_{rl}	268	88.93% (265/298)	99.9996% (795549/795552)
Delayed event time depends on significant SNPs			
Cox	201	66.11% (197/298)	99.9995% (795548/795552)
LRM_{obs}	129	41.61% (124/298)	99.9994% (795547/795552)
LRM_{u}	164	53.69% (160/298)	99.9995% (795548/795552)
LRM_{rl}	168	55.37% (165/298)	99.9996% (795549/795552)
Delayed event time depends on non-significant SNPs			
Cox	213	69.8% (208/298)	99.9994% (795547/795552)
LRM_{obs}	124	38.59% (115/298)	99.9989% (795543/795552)
LRM_{u}	165	53.36% (159/298)	99.9992% (795546/795552)
LRM_{rl}	206	67.45% (201/298)	99.9994% (795547/795552)
Delayed event time is independent			
Cox	224	71.81% (214/298)	99.9987% (795542/795552)
LRM_{obs}	148	47.32% (141/298)	99.9991% (795545/795552)
LRM_{u}	183	59.06% (176/298)	99.9991% (795545/795552)
LRM_{rl}	214	69.13% (206/298)	99.999% (795544/795552)
Delayed event time depends on sex			
Cox	258	80.54% (240/298)	99.9977% (795534/795552)
LRM_{obs}	193	62.42% (186/298)	99.9991% (795545/795552)
LRM_{u}	211	67.11% (200/298)	99.9986% (795541/795552)
LRM_{rl}	250	76.85% (229/298)	99.9974% (795531/795552)

Table 9: Number of significant SNPs, true positive rates, and true negative rates for each model and delayed event time combination, using Model 1 (Cox) with no delayed event time as the gold standard, for type 2 diabetes (phecode 250.2). The results are shown for the $P \leq 5 \times 10^{-8}$ significance level.

* Based on Model 1 (Cox)- no delayed event time

$P \leq 1 \times 10^{-5}$	# Significant SNPs	True Positive *	True Negative *
No delayed event time			
Cox	548	-	-
LRM_{obs}	483	76.28% (418/548)	99.9918% (795237/795302)
LRM_{u}	474	79.74% (437/548)	99.9953% (795265/795302)
LRM_{rl}	526	89.05% (488/548)	99.9952% (795264/795302)
Delayed event time depends on significant SNPs			
Cox	549	89.05% (488/548)	99.9923% (795241/795302)
LRM_{obs}	433	70.07% (384/548)	99.9938% (795253/795302)
LRM_{u}	454	78.28% (429/548)	99.9969% (795277/795302)
LRM_{rl}	535	85.4% (468/548)	99.9916% (795235/795302)
Delayed event time depends on non-significant SNPs			
Cox	578	92.34% (506/548)	99.9909% (795230/795302)
LRM_{obs}	465	68.43% (375/548)	99.9887% (795212/795302)
LRM_{u}	504	79.56% (436/548)	99.9914% (795234/795302)
LRM_{rl}	581	86.5% (474/548)	99.9865% (795195/795302)
Delayed event time is independent			
Cox	616	90.51% (496/548)	99.9849% (795182/795302)
LRM_{obs}	485	72.08% (395/548)	99.9887% (795212/795302)
LRM_{u}	534	79.93% (438/548)	99.9879% (795206/795302)
LRM_{rl}	606	86.68% (475/548)	99.9835% (795171/795302)
Delayed event time depends on sex			
Cox	603	88.69% (486/548)	99.9853% (795185/795302)
LRM_{obs}	490	69.34% (380/548)	99.9862% (795192/795302)
LRM_{u}	529	79.2% (434/548)	99.9881% (795207/795302)
LRM_{rl}	581	86.5% (474/548)	99.9865% (795195/795302)

Table 10: Number of significant SNPs, true positive rates, and true negative rates for each model and delayed event time combination, using Model 1 (Cox) with no delayed event time as the gold standard, for type 2 diabetes (phecode 250.2). The results are shown for the $P \leq 1 \times 10^{-5}$ significance level.

* Based on Model 1 (Cox) - no delayed event time

$P \leq 5 \times 10^{-8}$	# Significant SNPs	True Positive *	True Negative *
No delayed event time			
Cox	3	-	-
LRM_{obs}	3	100% (3/3)	100% (795847/795847)
LRM_{u}	3	100% (3/3)	100% (795847/795847)
LRM_{rl}	3	100% (3/3)	100% (795847/795847)
Delayed event time depends on significant SNPs			
Cox	5	66.67% (2/3)	99.9996% (795844/795847)
LRM_{obs}	5	66.67% (2/3)	99.9996% (795844/795847)
LRM_{u}	5	66.67% (2/3)	99.9996% (795844/795847)
LRM_{rl}	5	66.67% (2/3)	99.9996% (795844/795847)
Delayed event time depends on non-significant SNPs			
Cox	6	100% (3/3)	99.9996% (795844/795847)
LRM_{obs}	6	100% (3/3)	99.9996% (795844/795847)
LRM_{u}	6	100% (3/3)	99.9996% (795844/795847)
LRM_{rl}	6	100% (3/3)	99.9996% (795844/795847)
Delayed event time is independent			
Cox	6	100% (3/3)	99.9996% (795844/795847)
LRM_{obs}	7	100% (3/3)	99.9995% (795843/795847)
LRM_{u}	7	100% (3/3)	99.9995% (795843/795847)
LRM_{rl}	6	100% (3/3)	99.9996% (795844/795847)
Delayed event time depends on sex			
Cox	6	100% (3/3)	99.9996% (795844/795847)
LRM_{obs}	7	100% (3/3)	99.9995% (795843/795847)
LRM_{u}	7	100% (3/3)	99.9995% (795843/795847)
LRM_{rl}	6	100% (3/3)	99.9996% (795844/795847)

Table 11: Number of significant SNPs, true positive rates, and true negative rates for each model and delayed event time combination, using Model 1 (Cox) with no delayed event time as the gold standard, for vitamin D deficiency (phecode 261.4). The results are shown for the $P \leq 5 \times 10^{-8}$ significance level.

* Based on Model 1 (Cox) - no delayed event time

$P \leq 1 \times 10^{-5}$	# Significant SNPs	True Positive *	True Negative *
No delayed event time			
Cox	13	-	-
LRM_{obs}	16	92.31% (12/13)	99.9995% (795833/795837)
LRM_{u}	19	92.31% (12/13)	99.9991% (795830/795837)
LRM_{rl}	18	100% (13/13)	99.9994% (795832/795837)
Delayed event time depends on significant SNPs			
Cox	14	76.92% (10/13)	99.9995% (795833/795837)
LRM_{obs}	23	76.92% (10/13)	99.9984% (795824/795837)
LRM_{u}	22	76.92% (10/13)	99.9985% (795825/795837)
LRM_{rl}	18	84.62% (11/13)	99.9991% (795830/795837)
Delayed event time depends on non-significant SNPs			
Cox	17	84.62% (11/13)	99.9992% (795831/795837)
LRM_{obs}	21	84.62% (11/13)	99.9987% (795827/795837)
LRM_{u}	22	84.62% (11/13)	99.9986% (795826/795837)
LRM_{rl}	19	84.62% (11/13)	99.999% (795829/795837)
Delayed event time is independent			
Cox	20	84.62% (11/13)	99.9989% (795828/795837)
LRM_{obs}	21	84.62% (11/13)	99.9987% (795827/795837)
LRM_{u}	21	84.62% (11/13)	99.9987% (795827/795837)
LRM_{rl}	21	84.62% (11/13)	99.9987% (795827/795837)
Delayed event time depends on sex			
Cox	17	84.62% (11/13)	99.9992% (795831/795837)
LRM_{obs}	23	84.62% (11/13)	99.9985% (795825/795837)
LRM_{u}	22	84.62% (11/13)	99.9986% (795826/795837)
LRM_{rl}	21	92.31% (12/13)	99.9989% (795828/795837)

Table 12: Number of significant SNPs, true positive rates, and true negative rates for each model and delayed event time combination, using Model 1 (Cox) with no delayed event time as the gold standard, for vitamin D deficiency (phecode 261.4). The results are shown for the $P \leq 1 \times 10^{-5}$ significance level.

* Based on Model 1 (Cox) - no delayed event time

$P \leq 5 \times 10^{-8}$	# Significant SNPs	True Positive *	True Negative *
No delayed event time			
Cox	15	-	-
LRM_{obs}	16	80% (12/15)	99.9995% (795831/795835)
LRM_{u}	16	80% (12/15)	99.9995% (795831/795835)
LRM_{rl}	15	100% (15/15)	100% (795835/795835)
Delayed event time depends on significant SNPs			
Cox	11	73.33% (11/15)	100% (795835/795835)
LRM_{obs}	13	53.33% (8/15)	99.9994% (795830/795835)
LRM_{u}	14	73.33% (11/15)	99.9996% (795832/795835)
LRM_{rl}	11	73.33% (11/15)	100% (795835/795835)
Delayed event time depends on non-significant SNPs			
Cox	14	93.33% (14/15)	100% (795835/795835)
LRM_{obs}	16	73.33% (11/15)	99.9994% (795830/795835)
LRM_{u}	14	73.33% (11/15)	99.9996% (795832/795835)
LRM_{rl}	15	100% (15/15)	100% (795835/795835)
Delayed event time is independent			
Cox	13	86.67% (13/15)	100% (795835/795835)
LRM_{obs}	16	73.33% (11/15)	99.9994% (795830/795835)
LRM_{u}	14	73.33% (11/15)	99.9996% (795832/795835)
LRM_{rl}	15	100% (15/15)	100% (795835/795835)
Delayed event time depends on sex			
Cox	14	93.33% (14/15)	100% (795835/795835)
LRM_{obs}	16	73.33% (11/15)	99.9994% (795830/795835)
LRM_{u}	14	73.33% (11/15)	99.9996% (795832/795835)
LRM_{rl}	14	93.33% (14/15)	100% (795835/795835)

Table 13: Number of significant SNPs, true positive rates, and true negative rates for each model and delayed event time combination, using Model 1 (Cox) with no delayed event time as the gold standard, for hypercholesterolemia (phecode 272.11). The results are shown for the $P \leq 5 \times 10^{-8}$ significance level.

* Based on Model 1 (Cox) - no delayed event time

$P \leq 1 \times 10^{-5}$	# Significant SNPs	True Positive *	True Negative *
No delayed event time			
Cox	60	-	-
LRM_{obs}	34	41.67% (25/60)	99.9989% (795781/795790)
LRM_{u}	44	60% (36/60)	99.999% (795782/795790)
LRM_{rl}	52	85% (51/60)	99.9999% (795789/795790)
Delayed event time depends on significant SNPs			
Cox	53	83.33% (50/60)	99.9996% (795787/795790)
LRM_{obs}	43	45% (27/60)	99.998% (795774/795790)
LRM_{u}	41	53.33% (32/60)	99.9989% (795781/795790)
LRM_{rl}	53	80% (48/60)	99.9994% (795785/795790)
Delayed event time depends on non-significant SNPs			
Cox	58	93.33% (56/60)	99.9997% (795788/795790)
LRM_{obs}	37	40% (24/60)	99.9984% (795777/795790)
LRM_{u}	45	58.33% (35/60)	99.9987% (795780/795790)
LRM_{rl}	55	86.67% (52/60)	99.9996% (795787/795790)
Delayed event time is independent			
Cox	45	68.33% (41/60)	99.9995% (795786/795790)
LRM_{obs}	44	46.67% (28/60)	99.998% (795774/795790)
LRM_{u}	44	53.33% (32/60)	99.9985% (795778/795790)
LRM_{rl}	45	66.67% (40/60)	99.9994% (795785/795790)
Delayed event time depends on sex			
Cox	58	83.33% (50/60)	99.999% (795782/795790)
LRM_{obs}	47	50% (30/60)	99.9979% (795773/795790)
LRM_{u}	51	55% (33/60)	99.9977% (795772/795790)
LRM_{rl}	56	78.33% (47/60)	99.9989% (795781/795790)

Table 14: Number of significant SNPs, true positive rates, and true negative rates for each model and delayed event time combination, using Model 1 (Cox) with no delayed event time as the gold standard, for hypercholesterolemia (phecode 272.11). The results are shown for the $P \leq 1 \times 10^{-5}$ significance level.

* Based on Model 1 (Cox) - no delayed event time

$P \leq 5 \times 10^{-8}$	# Significant SNPs	True Positive *	True Negative *
No delayed event time			
Cox	1	-	-
LRM_{obs}	0	0% (0/1)	100% (795849/795849)
LRM_u	0	0% (0/1)	100% (795849/795849)
LRM_{rl}	0	0% (0/1)	100% (795849/795849)
Delayed event time depends on significant SNPs			
Cox	1	100% (1/1)	100% (795849/795849)
LRM_{obs}	1	100% (1/1)	100% (795849/795849)
LRM_u	1	100% (1/1)	100% (795849/795849)
LRM_{rl}	1	100% (1/1)	100% (795849/795849)
Delayed event time depends on non-significant SNPs			
Cox	0	0% (0/1)	100% (795849/795849)
LRM_{obs}	0	0% (0/1)	100% (795849/795849)
LRM_u	0	0% (0/1)	100% (795849/795849)
LRM_{rl}	0	0% (0/1)	100% (795849/795849)
Delayed event time is independent			
Cox	0	0% (0/1)	100% (795849/795849)
LRM_{obs}	0	0% (0/1)	100% (795849/795849)
LRM_u	0	0% (0/1)	100% (795849/795849)
LRM_{rl}	0	0% (0/1)	100% (795849/795849)
Delayed event time depends on sex			
Cox	0	0% (0/1)	100% (795849/795849)
LRM_{obs}	0	0% (0/1)	100% (795849/795849)
LRM_u	0	0% (0/1)	100% (795849/795849)
LRM_{rl}	0	0% (0/1)	100% (795849/795849)

Table 15: Number of significant SNPs, true positive rates, and true negative rates for each model and delayed event time combination, using Model 1 (Cox) with no delayed event time as the gold standard, for insomnia (phecode 327.4). The results are shown for the $P \leq 5 \times 10^{-8}$ significance level.

* Based on Model 1 (Cox) - no delayed event time

$P \leq 1 \times 10^{-5}$	# Significant SNPs	True Positive *	True Negative *
No delayed event time			
Cox	12	-	-
LRM_{obs}	14	66.67% (8/12)	99.9992% (795832/795838)
LRM_{u}	15	83.33% (10/12)	99.9994% (795833/795838)
LRM_{rl}	13	100% (12/12)	99.9999% (795837/795838)
Delayed event time depends on significant SNPs			
Cox	13	58.33% (7/12)	99.9992% (795832/795838)
LRM_{obs}	13	50% (6/12)	99.9991% (795831/795838)
LRM_{u}	13	58.33% (7/12)	99.9992% (795832/795838)
LRM_{rl}	14	58.33% (7/12)	99.9991% (795831/795838)
Delayed event time depends on non-significant SNPs			
Cox	10	66.67% (8/12)	99.9997% (795836/795838)
LRM_{obs}	12	50% (6/12)	99.9992% (795832/795838)
LRM_{u}	13	66.67% (8/12)	99.9994% (795833/795838)
LRM_{rl}	13	75% (9/12)	99.9995% (795834/795838)
Delayed event time is independent			
Cox	9	50% (6/12)	99.9996% (795835/795838)
LRM_{obs}	10	33.33% (4/12)	99.9992% (795832/795838)
LRM_{u}	13	58.33% (7/12)	99.9992% (795832/795838)
LRM_{rl}	15	83.33% (10/12)	99.9994% (795833/795838)
Delayed event time depends on sex			
Cox	14	75% (9/12)	99.9994% (795833/795838)
LRM_{obs}	13	58.33% (7/12)	99.9992% (795832/795838)
LRM_{u}	16	66.67% (8/12)	99.999% (795830/795838)
LRM_{rl}	14	75% (9/12)	99.9994% (795833/795838)

Table 16: Number of significant SNPs, true positive rates, and true negative rates for each model and delayed event time combination, using Model 1 (Cox) with no delayed event time as the gold standard, for insomnia (phecode 327.4). The results are shown for the $P \leq 1 \times 10^{-5}$ significance level.

* Based on Model 1 (Cox) - no delayed event time

$P \leq 5 \times 10^{-8}$	# Significant SNPs	True Positive *	True Negative *
No delayed event time			
Cox	4	-	-
LRM_{obs}	6	75% (3/4)	99.9996% (795843/795846)
LRM_{u}	6	100% (4/4)	99.9997% (795844/795846)
LRM_{rl}	3	75% (3/4)	100% (795846/795846)
Delayed event time depends on significant SNPs			
Cox	4	100% (4/4)	100% (795846/795846)
LRM_{obs}	6	100% (4/4)	99.9997% (795844/795846)
LRM_{u}	6	100% (4/4)	99.9997% (795844/795846)
LRM_{rl}	4	100% (4/4)	100% (795846/795846)
Delayed event time depends on non-significant SNPs			
Cox	3	75% (3/4)	100% (795846/795846)
LRM_{obs}	5	75% (3/4)	99.9997% (795844/795846)
LRM_{u}	6	100% (4/4)	99.9997% (795844/795846)
LRM_{rl}	3	75% (3/4)	100% (795846/795846)
Delayed event time is independent			
Cox	3	75% (3/4)	100% (795846/795846)
LRM_{obs}	4	50% (2/4)	99.9997% (795844/795846)
LRM_{u}	6	100% (4/4)	99.9997% (795844/795846)
LRM_{rl}	3	75% (3/4)	100% (795846/795846)
Delayed event time depends on sex			
Cox	1	25% (1/4)	100% (795846/795846)
LRM_{obs}	3	25% (1/4)	99.9997% (795844/795846)
LRM_{u}	4	50% (2/4)	99.9997% (795844/795846)
LRM_{rl}	1	25% (1/4)	100% (795846/795846)

Table 17: Number of significant SNPs, true positive rates, and true negative rates for each model and delayed event time combination, using Model 1 (Cox) with no delayed event time as the gold standard, for myocardial infarction (phecode 411.2). The results are shown for the $P \leq 5 \times 10^{-8}$ significance level.

* Based on Model 1 (Cox) - no delayed event time

$P \leq 1 \times 10^{-5}$	# Significant SNPs	True Positive *	True Negative *
No delayed event time			
Cox	19	-	-
LRM_{obs}	29	57.89% (11/19)	99.9977% (795813/795831)
LRM_u	24	68.42% (13/19)	99.9986% (795820/795831)
LRM_{rl}	17	78.95% (15/19)	99.9997% (795829/795831)
Delayed event time depends on significant SNPs			
Cox	17	36.84% (7/19)	99.9987% (795821/795831)
LRM_{obs}	23	42.11% (8/19)	99.9981% (795816/795831)
LRM_u	17	36.84% (7/19)	99.9987% (795821/795831)
LRM_{rl}	19	42.11% (8/19)	99.9986% (795820/795831)
Delayed event time depends on non-significant SNPs			
Cox	21	52.63% (10/19)	99.9986% (795820/795831)
LRM_{obs}	23	36.84% (7/19)	99.998% (795815/795831)
LRM_u	24	47.37% (9/19)	99.9981% (795816/795831)
LRM_{rl}	17	47.37% (9/19)	99.999% (795823/795831)
Delayed event time is independent			
Cox	18	31.58% (6/19)	99.9985% (795819/795831)
LRM_{obs}	26	36.84% (7/19)	99.9976% (795812/795831)
LRM_u	17	36.84% (7/19)	99.9987% (795821/795831)
LRM_{rl}	17	42.11% (8/19)	99.9989% (795822/795831)
Delayed event time depends on sex			
Cox	19	42.11% (8/19)	99.9986% (795820/795831)
LRM_{obs}	20	31.58% (6/19)	99.9982% (795817/795831)
LRM_u	19	36.84% (7/19)	99.9985% (795819/795831)
LRM_{rl}	15	36.84% (7/19)	99.999% (795823/795831)

Table 18: Number of significant SNPs, true positive rates, and true negative rates for each model and delayed event time combination, using Model 1 (Cox) with no delayed event time as the gold standard, for myocardial infarction (phecode 411.2). The results are shown for the $P \leq 1 \times 10^{-5}$ significance level.

* Based on Model 1 (Cox) - no delayed event time

$P \leq 5 \times 10^{-8}$	# Significant SNPs	True Positive *	True Negative *
No delayed event time			
Cox	181	-	-
LRM_{obs}	75	27.62% (50/181)	99.9969% (795644/795669)
LRM_{u}	117	51.93% (94/181)	99.9971% (795646/795669)
LRM_{rl}	164	88.95% (161/181)	99.9996% (795666/795669)
Delayed event time depends on significant SNPs			
Cox	152	81.77% (148/181)	99.9995% (795665/795669)
LRM_{obs}	68	27.07% (49/181)	99.9976% (795650/795669)
LRM_{u}	83	34.25% (62/181)	99.9974% (795648/795669)
LRM_{rl}	87	46.41% (84/181)	99.9996% (795666/795669)
Delayed event time depends on non-significant SNPs			
Cox	165	91.16% (165/181)	100% (795669/795669)
LRM_{obs}	67	27.07% (49/181)	99.9977% (795651/795669)
LRM_{u}	78	32.6% (59/181)	99.9976% (795650/795669)
LRM_{rl}	92	50.28% (91/181)	99.9999% (795668/795669)
Delayed event time is independent			
Cox	164	90.61% (164/181)	100% (795669/795669)
LRM_{obs}	73	30.94% (56/181)	99.9979% (795652/795669)
LRM_{u}	112	53.04% (96/181)	99.998% (795653/795669)
LRM_{rl}	122	66.3% (120/181)	99.9997% (795667/795669)
Delayed event time depends on sex			
Cox	139	76.24% (138/181)	99.9999% (795668/795669)
LRM_{obs}	70	28.73% (52/181)	99.9977% (795651/795669)
LRM_{u}	88	38.67% (70/181)	99.9977% (795651/795669)
LRM_{rl}	85	46.96% (85/181)	100% (795669/795669)

Table 19: Number of significant SNPs, true positive rates, and true negative rates for each model and delayed event time combination, using Model 1 (Cox) with no delayed event time as the gold standard, for coronary atherosclerosis (phecode 411.4). The results are shown for the $P \leq 5 \times 10^{-8}$ significance level.

* Based on Model 1 (Cox) - no delayed event time

$P \leq 1 \times 10^{-5}$	# Significant SNPs	True Positive *	True Negative *
No delayed event time			
Cox	423	-	-
LRM_{obs}	234	39.48% (167/423)	99.9916% (795360/795427)
LRM_u	275	51.77% (219/423)	99.993% (795371/795427)
LRM_{rl}	234	52.48% (222/423)	99.9985% (795415/795427)
Delayed event time depends on significant SNPs			
Cox	394	79.43% (336/423)	99.9927% (795369/795427)
LRM_{obs}	181	31.21% (132/423)	99.9938% (795378/795427)
LRM_u	246	48.23% (204/423)	99.9947% (795385/795427)
LRM_{rl}	220	50.12% (212/423)	99.999% (795419/795427)
Delayed event time depends on non-significant SNPs			
Cox	365	74.7% (316/423)	99.9938% (795378/795427)
LRM_{obs}	178	30.02% (127/423)	99.9936% (795376/795427)
LRM_u	248	47.04% (199/423)	99.9938% (795378/795427)
LRM_{rl}	214	48.94% (207/423)	99.9991% (795420/795427)
Delayed event time is independent			
Cox	343	72.34% (306/423)	99.9953% (795390/795427)
LRM_{obs}	220	40.9% (173/423)	99.9941% (795380/795427)
LRM_u	246	48.7% (206/423)	99.995% (795387/795427)
LRM_{rl}	216	49.65% (210/423)	99.9992% (795421/795427)
Delayed event time depends on sex			
Cox	325	69.03% (292/423)	99.9959% (795394/795427)
LRM_{obs}	195	33.57% (142/423)	99.9933% (795374/795427)
LRM_u	247	48.7% (206/423)	99.9948% (795386/795427)
LRM_{rl}	212	48.23% (204/423)	99.999% (795419/795427)

Table 20: Number of significant SNPs, true positive rates, and true negative rates for each model and delayed event time combination, using Model 1 (Cox) with no delayed event time as the gold standard, for coronary atherosclerosis (phecode 411.4). The results are shown for the $P \leq 1 \times 10^{-5}$ significance level.

* Based on Model 1 (Cox) - no delayed event time

$P \leq 5 \times 10^{-8}$	# Significant SNPs	True Positive *	True Negative *
No delayed event time			
Cox	126	-	-
LRM_{obs}	126	92.86% (117/126)	99.9989% (795715/795724)
LRM_{u}	123	93.65% (118/126)	99.9994% (795719/795724)
LRM_{rl}	123	94.44% (119/126)	99.9995% (795720/795724)
Delayed event time depends on significant SNPs			
Cox	118	93.65% (118/126)	100% (795724/795724)
LRM_{obs}	121	92.06% (116/126)	99.9994% (795719/795724)
LRM_{u}	121	92.06% (116/126)	99.9994% (795719/795724)
LRM_{rl}	116	91.27% (115/126)	99.9999% (795723/795724)
Delayed event time depends on non-significant SNPs			
Cox	122	96.83% (122/126)	100% (795724/795724)
LRM_{obs}	121	91.27% (115/126)	99.9992% (795718/795724)
LRM_{u}	121	92.86% (117/126)	99.9995% (795720/795724)
LRM_{rl}	116	91.27% (115/126)	99.9999% (795723/795724)
Delayed event time is independent			
Cox	118	93.65% (118/126)	100% (795724/795724)
LRM_{obs}	119	91.27% (115/126)	99.9995% (795720/795724)
LRM_{u}	121	92.86% (117/126)	99.9995% (795720/795724)
LRM_{rl}	115	91.27% (115/126)	100% (795724/795724)
Delayed event time depends on sex			
Cox	116	92.06% (116/126)	100% (795724/795724)
LRM_{obs}	119	91.27% (115/126)	99.9995% (795720/795724)
LRM_{u}	121	92.86% (117/126)	99.9995% (795720/795724)
LRM_{rl}	114	90.48% (114/126)	100% (795724/795724)

Table 21: Number of significant SNPs, true positive rates, and true negative rates for each model and delayed event time combination, using Model 1 (Cox) with no delayed event time as the gold standard, for atrial fibrillation (phecode 427.21). The results are shown for the $P \leq 5 \times 10^{-8}$ significance level.

* Based on Model 1 (Cox) - no delayed event time

$P \leq 1 \times 10^{-5}$	# Significant SNPs	True Positive *	True Negative *
No delayed event time			
Cox	186	-	-
LRM_{obs}	197	72.04% (134/186)	99.9921% (795601/795664)
LRM_{u}	194	75.81% (141/186)	99.9933% (795611/795664)
LRM_{rl}	186	83.87% (156/186)	99.9962% (795634/795664)
Delayed event time depends on significant SNPs			
Cox	175	83.33% (155/186)	99.9975% (795644/795664)
LRM_{obs}	177	69.89% (130/186)	99.9941% (795617/795664)
LRM_{u}	173	70.97% (132/186)	99.9948% (795623/795664)
LRM_{rl}	156	76.88% (143/186)	99.9984% (795651/795664)
Delayed event time depends on non-significant SNPs			
Cox	176	83.33% (155/186)	99.9974% (795643/795664)
LRM_{obs}	151	69.89% (130/186)	99.9974% (795643/795664)
LRM_{u}	149	72.04% (134/186)	99.9981% (795649/795664)
LRM_{rl}	158	76.34% (142/186)	99.998% (795648/795664)
Delayed event time is independent			
Cox	179	83.33% (155/186)	99.997% (795640/795664)
LRM_{obs}	157	70.43% (131/186)	99.9967% (795638/795664)
LRM_{u}	161	72.58% (135/186)	99.9967% (795638/795664)
LRM_{rl}	163	77.96% (145/186)	99.9977% (795646/795664)
Delayed event time depends on sex			
Cox	165	81.18% (151/186)	99.9982% (795650/795664)
LRM_{obs}	155	70.43% (131/186)	99.997% (795640/795664)
LRM_{u}	147	70.43% (131/186)	99.998% (795648/795664)
LRM_{rl}	154	75.81% (141/186)	99.9984% (795651/795664)

Table 22: Number of significant SNPs, true positive rates, and true negative rates for each model and delayed event time combination, using Model 1 (Cox) with no delayed event time as the gold standard, for atrial fibrillation (phecode 427.21). The results are shown for the $P \leq 1 \times 10^{-5}$ significance level.

* Based on Model 1 (Cox) - no delayed event time

$P \leq 5 \times 10^{-8}$	# Significant SNPs	True Positive *	True Negative *
No delayed event time			
Cox	879	-	-
LRM_{obs}	488	49.6% (436/879)	99.9993% (7957569/7957621)
LRM_{u}	591	61.89% (544/879)	99.9994% (7957574/7957621)
LRM_{rl}	775	86.92% (764/879)	99.9999% (7957610/7957621)
Delayed event time depends on significant SNPs			
Cox	756	81% (712/879)	99.9994% (7957577/7957621)
LRM_{obs}	485	49.03% (431/879)	99.9993% (7957567/7957621)
LRM_{u}	556	56.88% (500/879)	99.9993% (7957565/7957621)
LRM_{rl}	616	67.24% (591/879)	99.9997% (7957596/7957621)
Delayed event time depends on non-significant SNPs			
Cox	765	84.41% (742/879)	99.9997% (7957598/7957621)
LRM_{obs}	458	46.87% (412/879)	99.9994% (7957575/7957621)
LRM_{u}	531	55.18% (485/879)	99.9994% (7957575/7957621)
LRM_{rl}	652	72.13% (634/879)	99.9998% (7957603/7957621)
Delayed event time is independent			
Cox	753	81.91% (720/879)	99.9996% (7957588/7957621)
LRM_{obs}	500	51.19% (450/879)	99.9994% (7957571/7957621)
LRM_{u}	585	61.32% (539/879)	99.9994% (7957575/7957621)
LRM_{rl}	637	70.08% (616/879)	99.9997% (7957600/7957621)
Delayed event time depends on sex			
Cox	802	84.16% (733/871)	99.999% (7161710/7161779)
LRM_{obs}	550	56.49% (492/871)	99.9992% (7161721/7161779)
LRM_{u}	607	61.88% (539/871)	99.9991% (7161711/7161779)
LRM_{rl}	705	75.43% (657/871)	99.9993% (7161731/7161779)

Table 23: Number of significant SNPs, true positive rates, and true negative rates for each model and delayed event time combination, using Model 1 (Cox) with no delayed event time as the gold standard, for all ten phecodes. The results are shown for the $P \leq 5 \times 10^{-8}$ significance level.

* Based on Model 1 (Cox) - no delayed event time

$P \leq 1 \times 10^{-5}$	# Significant SNPs	True Positive *	True Negative *
No delayed event time			
Cox	2058	-	-
LRM_{obs}	1495	59.52% (1225/2058)	99.9966% (7956172/7956442)
LRM_{u}	1584	67.83% (1396/2058)	99.9976% (7956254/7956442)
LRM_{rl}	1718	78.52% (1616/2058)	99.9987% (7956340/7956442)
Delayed event time depends on significant SNPs			
Cox	2042	85.67% (1763/2058)	99.9965% (7956163/7956442)
LRM_{obs}	1436	55.98% (1152/2058)	99.9964% (7956158/7956442)
LRM_{u}	1581	64.63% (1330/2058)	99.9968% (7956191/7956442)
LRM_{rl}	1733	73.71% (1517/2058)	99.9973% (7956226/7956442)
Delayed event time depends on non-significant SNPs			
Cox	2035	86.39% (1778/2058)	99.9968% (7956185/7956442)
LRM_{obs}	1384	54.03% (1112/2058)	99.9966% (7956170/7956442)
LRM_{u}	1581	65.21% (1342/2058)	99.997% (7956203/7956442)
LRM_{rl}	1738	73.71% (1517/2058)	99.9972% (7956221/7956442)
Delayed event time is independent			
Cox	2010	82.99% (1708/2058)	99.9962% (7956140/7956442)
LRM_{obs}	1494	58.65% (1207/2058)	99.9964% (7956155/7956442)
LRM_{u}	1666	66.52% (1369/2058)	99.9963% (7956145/7956442)
LRM_{rl}	1791	74.15% (1526/2058)	99.9967% (7956177/7956442)
Delayed event time depends on sex			
Cox	2001	83.67% (1691/2021)	99.9957% (7160319/7160629)
LRM_{obs}	1486	57.84% (1169/2021)	99.9956% (7160312/7160629)
LRM_{u}	1665	67.29% (1360/2021)	99.9957% (7160324/7160629)
LRM_{rl}	1792	75.85% (1533/2021)	99.9964% (7160370/7160629)

Table 24: Number of significant SNPs, true positive rates, and true negative rates for each model and delayed event time combination, using Model 1 (Cox) with no delayed event time as the gold standard, for all ten phecodes. The results are shown for the $P \leq 1 \times 10^{-5}$ significance level.

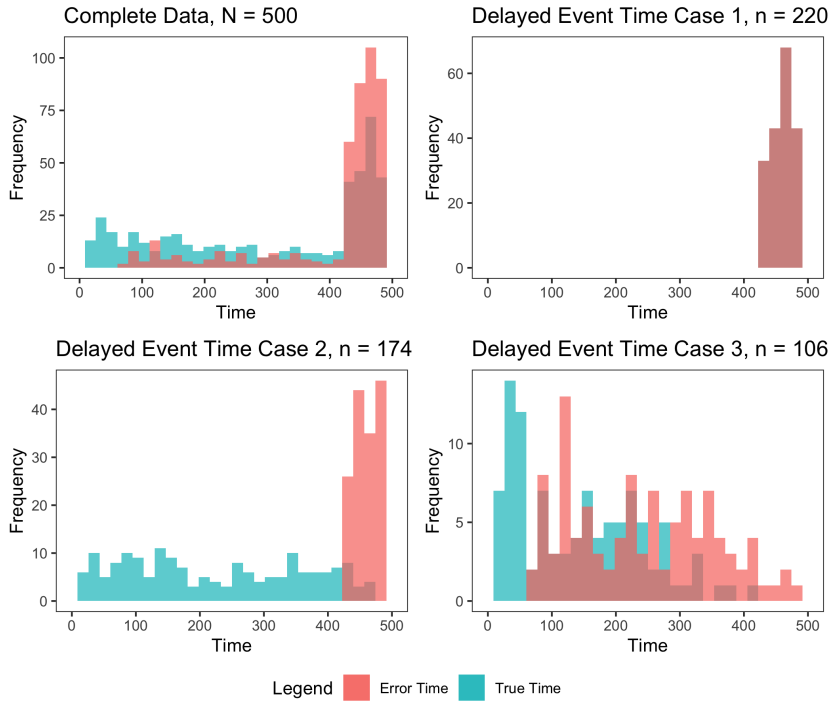
* Based on Model 1 (Cox) - no delayed event time

	$P \leq 5 \times 10^{-8}$		$P \leq 1 \times 10^{-5}$	
	TPR (95% CI) *	**TNR (95% CI) ***	**TPR (95% CI) ***	**TNR (95% CI) ***
No delayed event time				
Cox	Reference	Reference	Reference	Reference
LRM_{obs}	50.62% (44.49%, 56.75%)	99.99% (99.99%, 100.00%)	62.29% (57.72%, 66.85%)	99.99% (99.97%, 100.00%)
LRM_{u}	57.48% (50.17%, 64.78%)	99.99% (99.99%, 100.00%)	71.37% (67.20%, 75.54%)	99.99% (99.98%, 100.00%)
LRM_{rl}	69.34% (62.73%, 75.94%)	100.00% (99.99%, 100.00%)	80.07% (75.34%, 84.79%)	99.99% (99.98%, 100.00%)
Delayed event time depends on significant SNPs				
Cox	84.87% (78.72%, 91.03%)	99.99% (99.95%, 100.00%)	72.92% (67.57%, 78.27%)	99.99% (99.93%, 100.00%)
LRM_{obs}	66.67% (60.31%, 73.03%)	99.99% (99.96%, 100.00%)	54.33% (49.08%, 59.58%)	99.99% (99.93%, 100.00%)
LRM_{u}	67.72% (62.31%, 73.12%)	99.99% (99.96%, 100.00%)	59.10% (53.80%, 64.39%)	99.99% (99.93%, 100.00%)
LRM_{rl}	74.66% (67.29%, 82.04%)	100.00% (99.97%, 100.00%)	67.02% (61.69%, 72.35%)	99.99% (99.94%, 100.00%)
Delayed event time depends on non-significant SNPs				
Cox	71.63% (63.44%, 79.82%)	100.00% (99.98%, 100.00%)	75.94% (70.58%, 81.30%)	99.99% (99.96%, 100.00%)
LRM_{obs}	49.17% (43.21%, 55.12%)	99.99% (99.98%, 100.00%)	51.71% (46.63%, 56.79%)	99.99% (99.96%, 100.00%)
LRM_{u}	51.72% (45.85%, 57.59%)	99.99% (99.98%, 100.00%)	61.40% (56.17%, 66.61%)	99.99% (99.96%, 100.00%)
LRM_{rl}	62.14% (55.48%, 68.81%)	100.00% (99.98%, 100.00%)	67.78% (62.36%, 73.20%)	99.99% (99.96%, 100.00%)
Delayed event time is independent				
Cox	60.65% (55.34%, 65.96%)	100.00% (99.98%, 100.00%)	68.01% (62.75%, 73.26%)	99.99% (99.96%, 100.00%)
LRM_{obs}	44.57% (39.65%, 49.49%)	99.99% (99.97%, 100.00%)	49.41% (44.98%, 53.85%)	99.99% (99.96%, 100.00%)
LRM_{u}	53.54% (48.62%, 58.46%)	99.99% (99.97%, 100.00%)	57.48% (52.48%, 62.48%)	99.99% (99.96%, 100.00%)
LRM_{rl}	56.84% (51.12%, 62.55%)	100.00% (99.98%, 100.00%)	64.28% (59.34%, 69.23%)	99.99% (99.96%, 100.00%)
Delayed event time depends on sex				
Cox	62.54% (57.05%, 68.03%)	99.99% (99.97%, 100.00%)	73.40% (68.26%, 78.54%)	99.99% (99.96%, 100.00%)
LRM_{obs}	48.27% (43.35%, 53.19%)	99.99% (99.97%, 100.00%)	52.70% (48.49%, 56.91%)	99.99% (99.95%, 100.00%)
LRM_{u}	53.43% (48.51%, 58.34%)	99.99% (99.97%, 100.00%)	58.95% (54.52%, 63.37%)	99.99% (99.95%, 100.00%)
LRM_{rl}	58.22% (52.72%, 63.72%)	99.99% (99.98%, 100.00%)	65.96% (61.60%, 70.33%)	99.99% (99.96%, 100.00%)

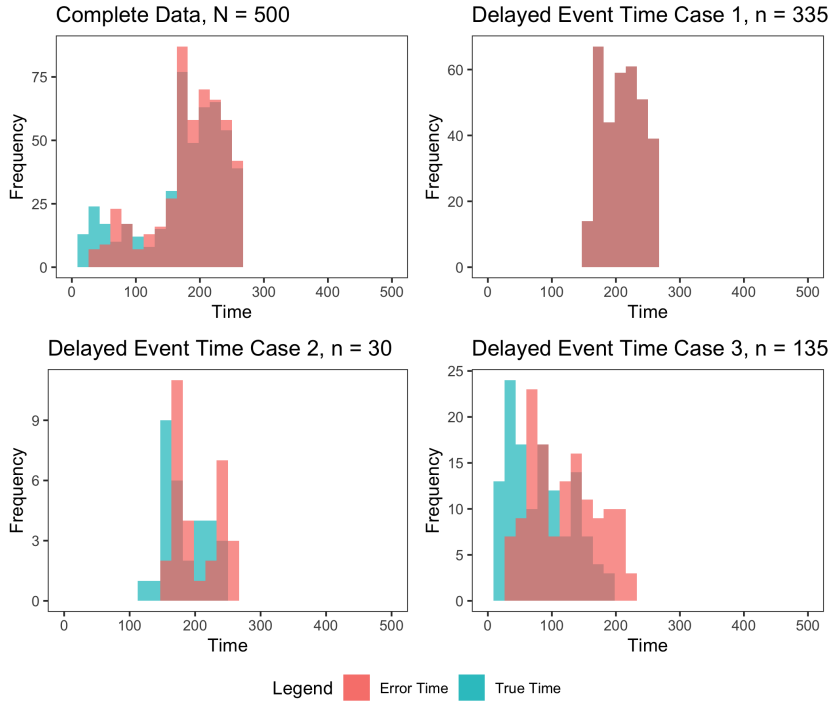
Table 25: Average true positive and true negative rates (95% confidence interval) for each model and delayed event time combination, using Model 1 (Cox) with no delayed event time as the gold standard. The average is calculated from ten phecodes, which are given in Appendix B, Table 1. The results are shown for both the $P \leq 5 \times 10^{-8}$ and $P \leq 1 \times 10^{-5}$ significance levels.

* Based on Model 1 (Cox) - no delayed event time

9 Appendix C: Additional Figures



(a) Parameters led to a large number of observations with a misclassified event status.



(b) Parameters led to a small number of observations with a misclassified event status.

Figure 7: Histograms of counts of observations in each delayed event time case in Simulation 1 when the event time was generated from a Cox model with baseline hazard from an exponential distribution. The delayed event time cases are explained in detail in Appendix A.

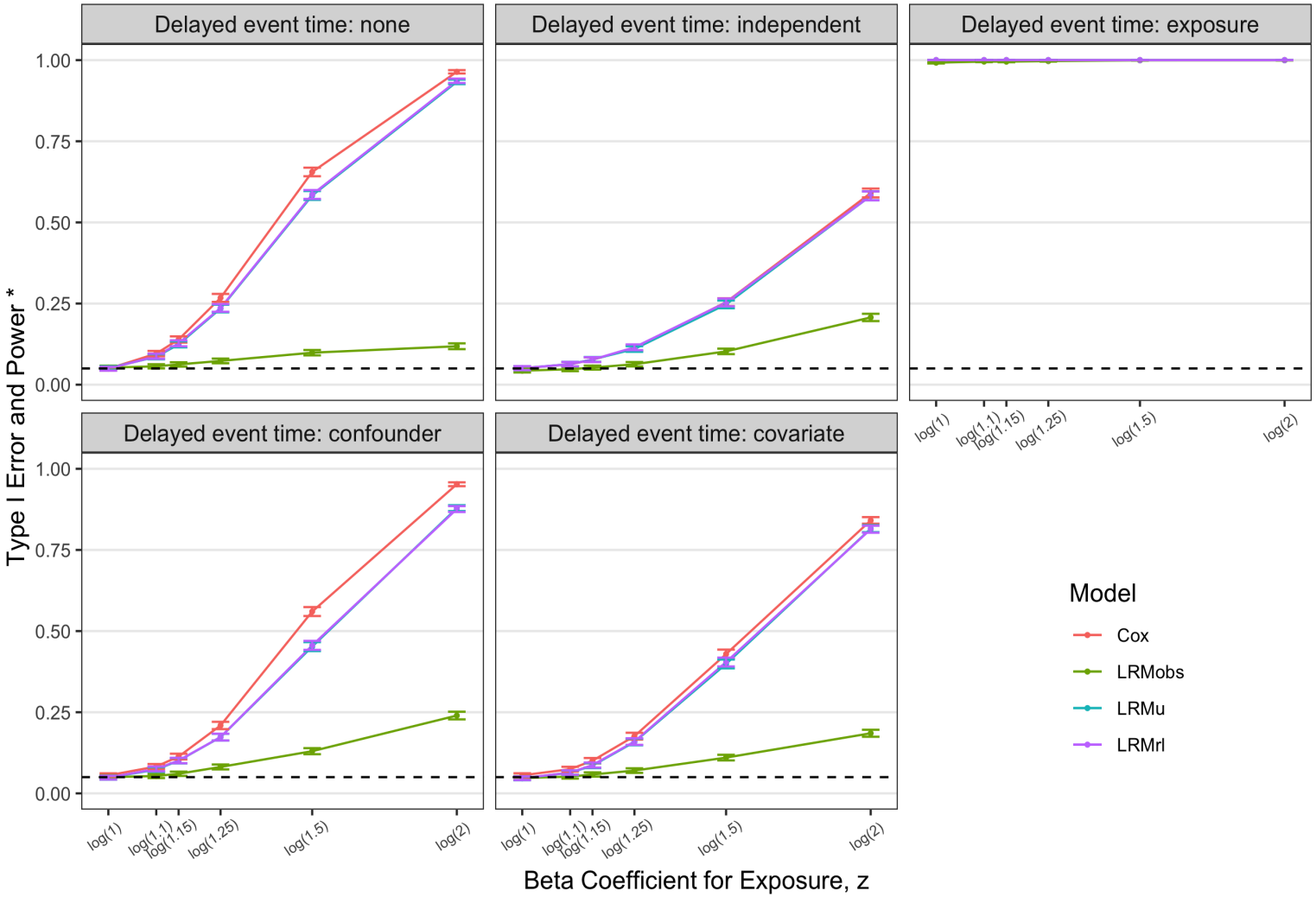


Figure 8: Results from Simulation 1 when the event time was generated from a Cox model with baseline hazard from a log-normal distribution, the censoring time was generated from a uniform distribution, and there was left truncation. The parameters led to a large number of observations with a misclassified event status (detailed in Appendix A).

* Type I error evaluated at log(1). Power evaluated at log(1.1), log(1.15), log(1.25), log(1.5), log(2).

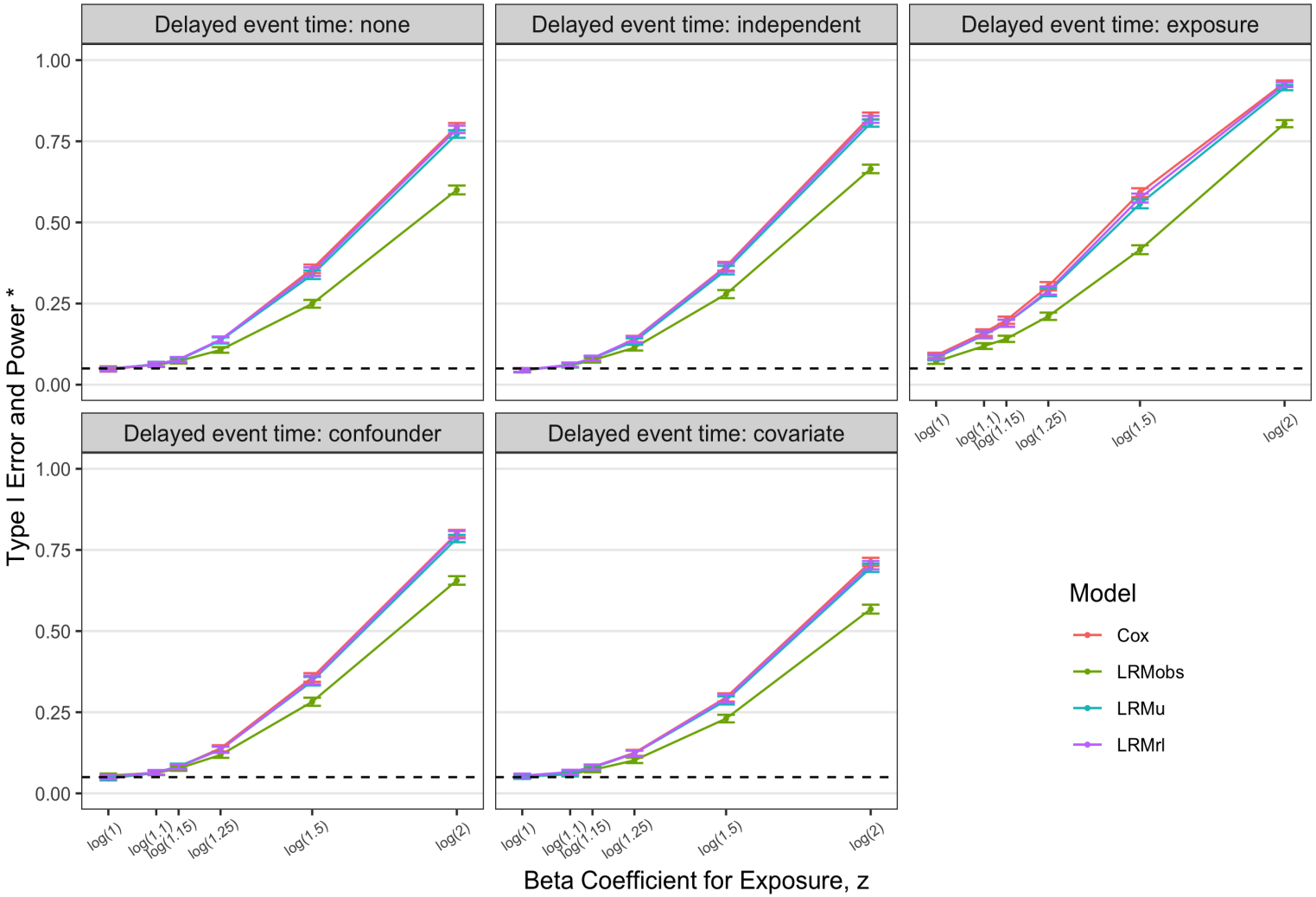


Figure 9: Results from Simulation 1 when the event time was generated from a Cox model with baseline hazard from a log-normal distribution, the censoring time was generated from a uniform distribution, and there was left truncation. The parameters led to a small number of observations with a misclassified event status (detailed in Appendix A).

* Type I error evaluated at log(1). Power evaluated at log(1.1), log(1.15), log(1.25), log(1.5), log(2).

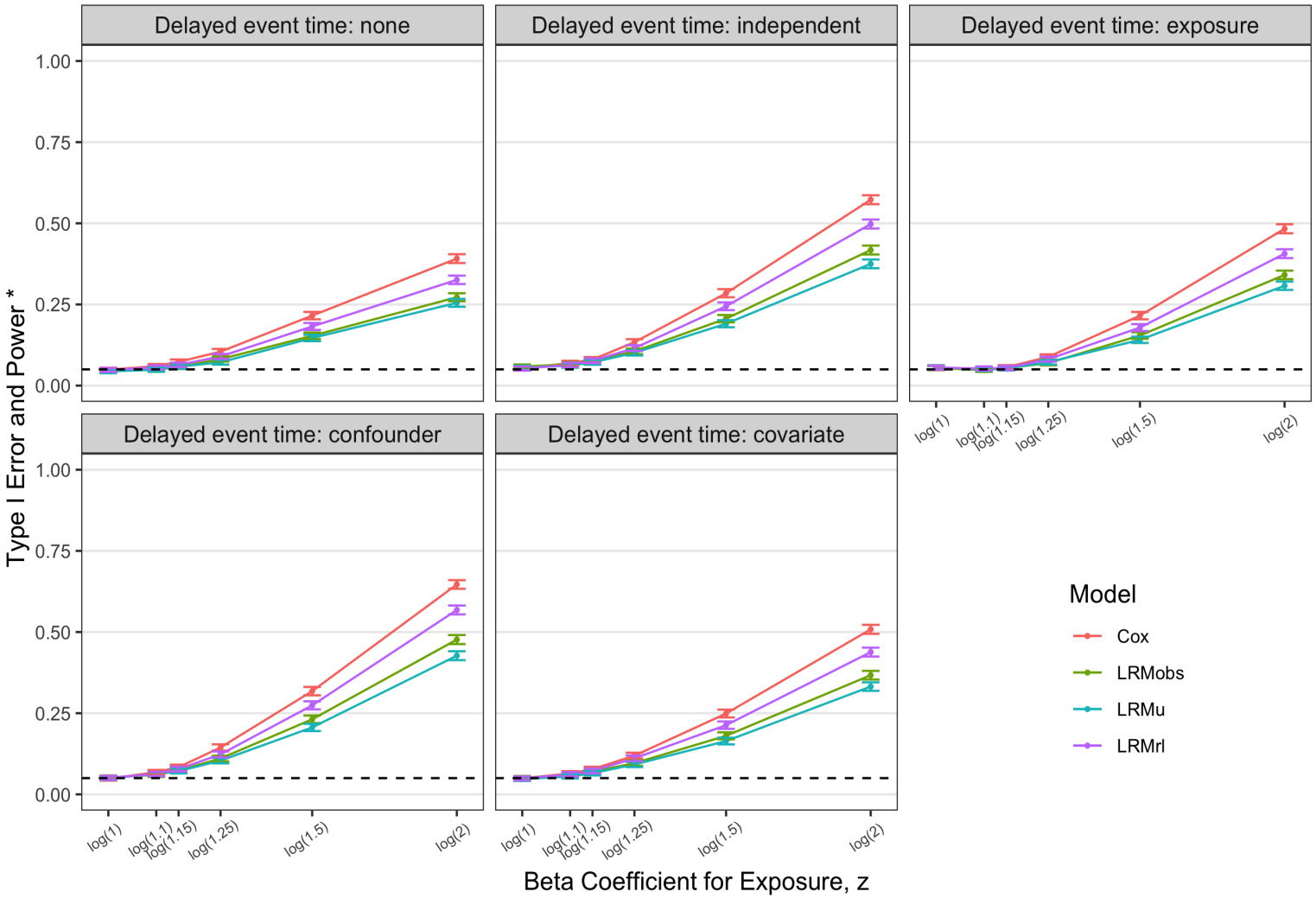


Figure 10: Results from Simulation 1 when the event time was generated from a Cox model with baseline hazard from an exponential distribution, the censoring time was generated from a Cox model with baseline hazard from an exponential distribution that depended on x, and there was left truncation. The parameters led to a small number of observations with a misclassified event status (detailed in Appendix A).

* Type I error evaluated at log(1). Power evaluated at log(1.1), log(1.15), log(1.25), log(1.5), log(2).

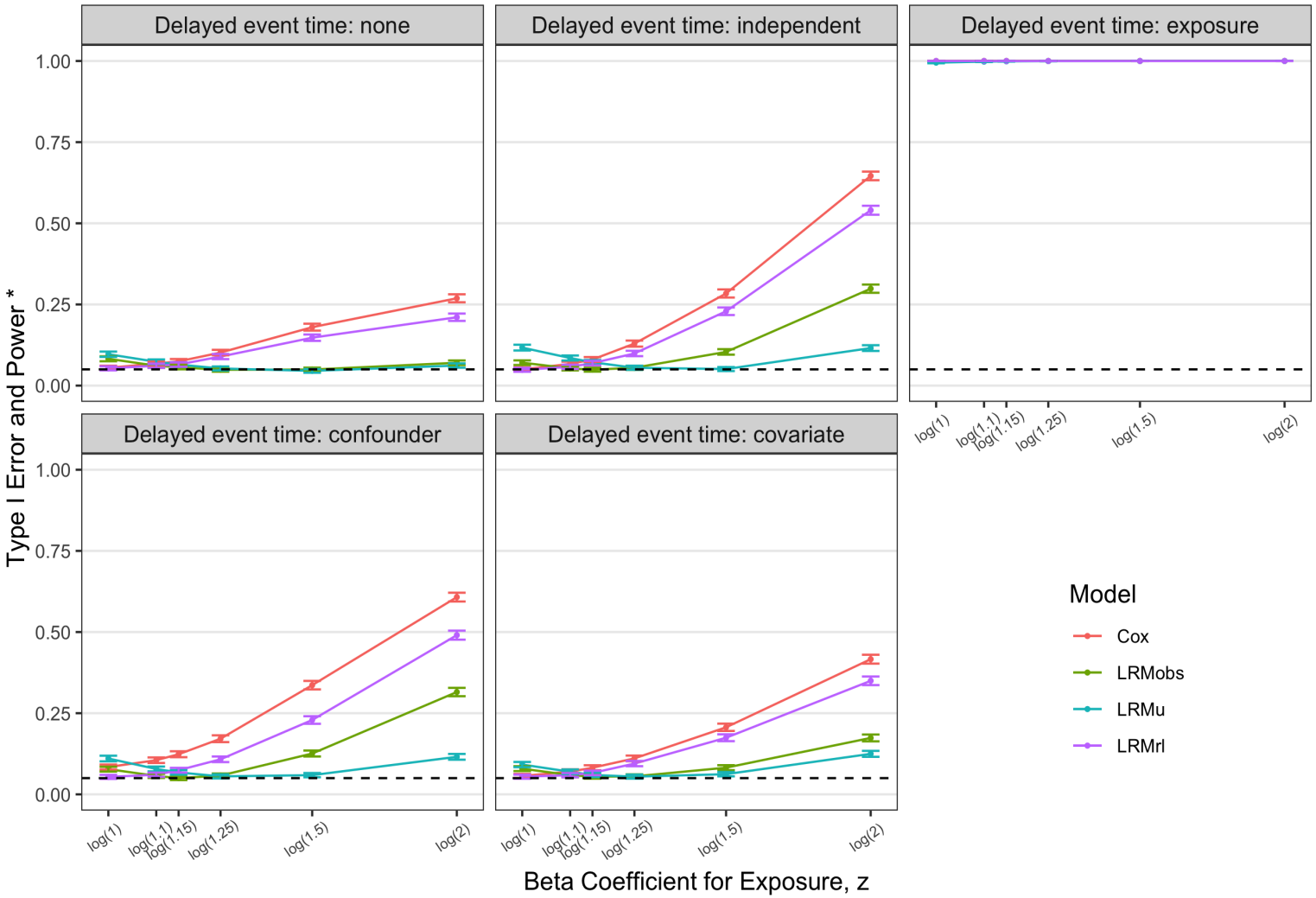


Figure 11: Results from Simulation 1 when the event time was generated from a Cox model with baseline hazard from an exponential distribution, the censoring time was generated from a Cox model with baseline hazard from an exponential distribution that depended on x and z, and there was left truncation. The parameters led to a large number of observations with a misclassified event status (detailed in Appendix A).

* Type I error evaluated at log(1). Power evaluated at log(1.1), log(1.15), log(1.25), log(1.5), log(2).

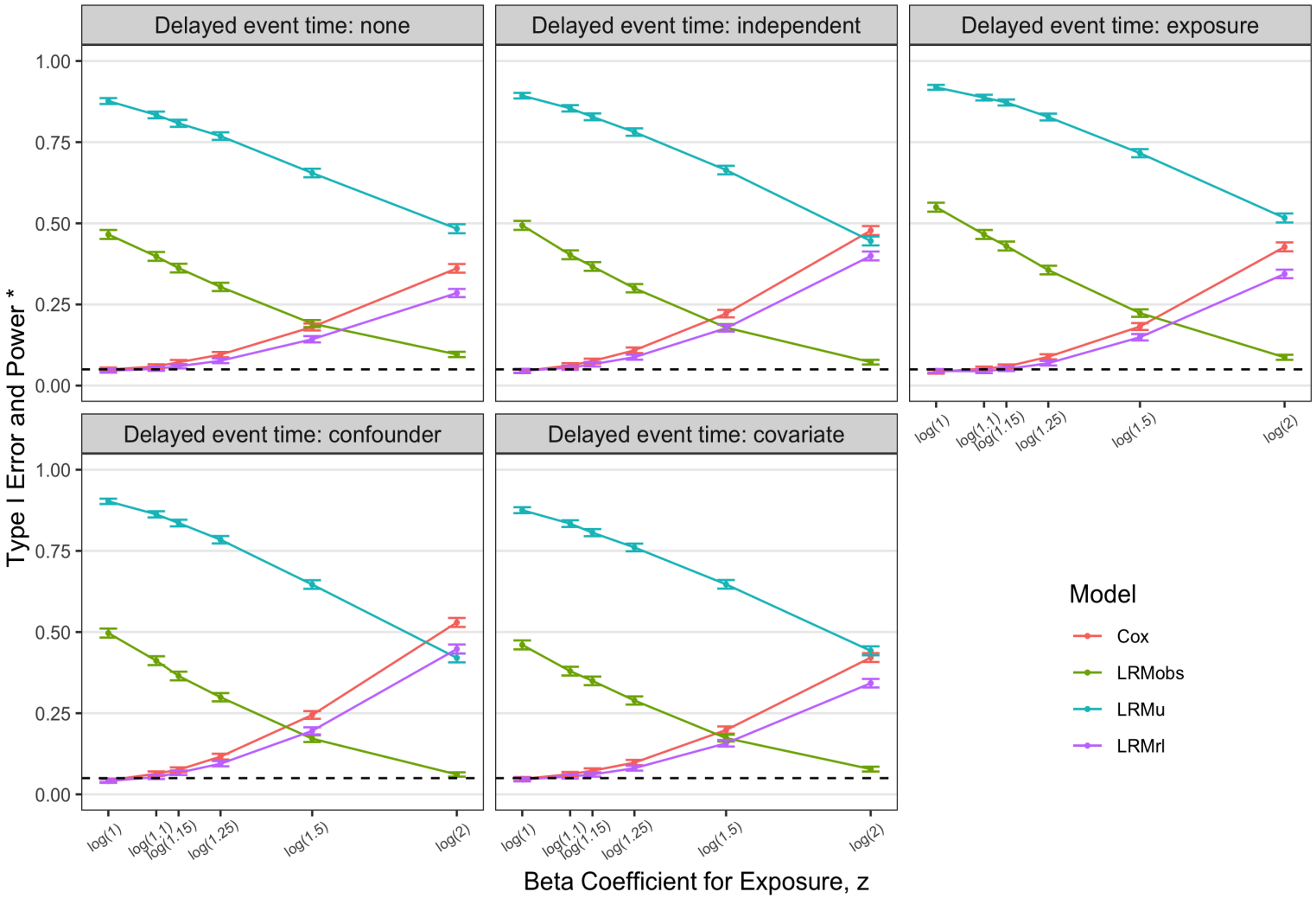


Figure 12: Results from Simulation 1 when the event time was generated from a Cox model with baseline hazard from an exponential distribution, the censoring time was generated from a Cox model with baseline hazard from an exponential distribution that depended on x and z, and there was left truncation. The parameters led to a small number of observations with a misclassified event status (detailed in Appendix A).

* Type I error evaluated at log(1). Power evaluated at log(1.1), log(1.15), log(1.25), log(1.5), log(2).

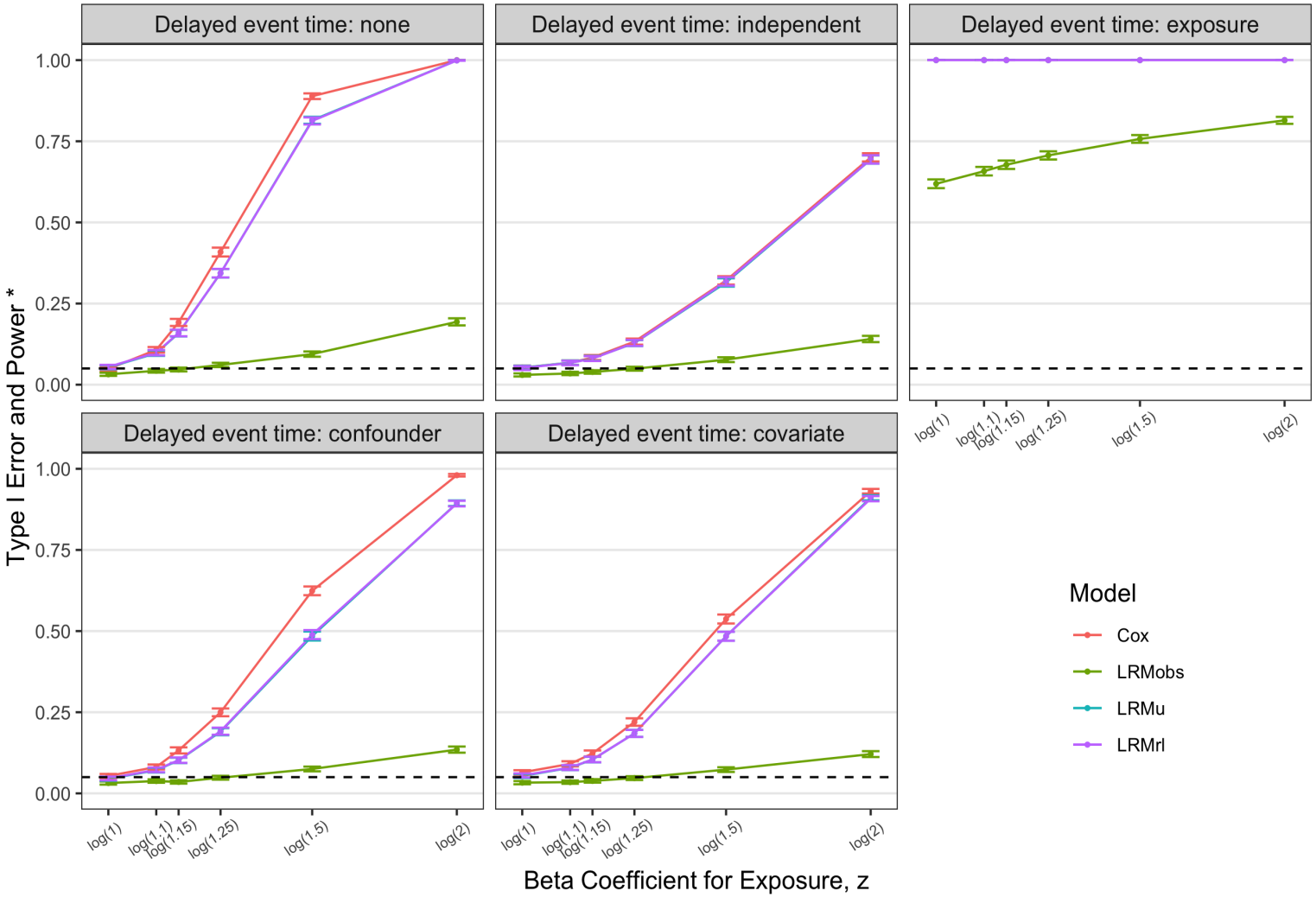


Figure 13: Results from Simulation 1 when the event time was generated from a Cox model with baseline hazard from an exponential distribution and the censoring time was generated from a uniform distribution. The parameters led to a large number of observations with a misclassified event status (detailed in Appendix A).

* Type I error evaluated at log(1). Power evaluated at log(1.1), log(1.15), log(1.25), log(1.5), log(2).

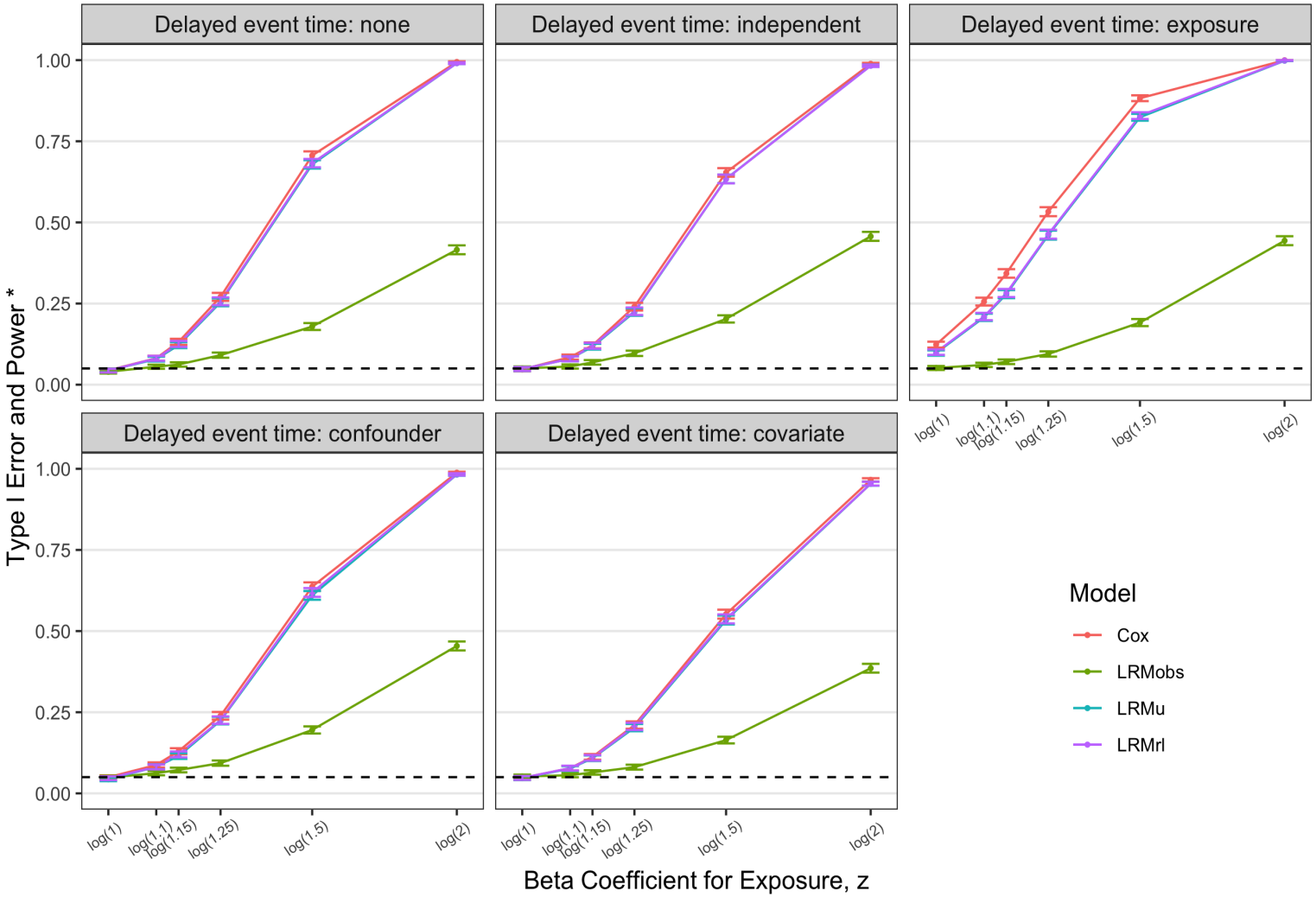


Figure 14: Results from Simulation 1 when the event time was generated from a Cox model with baseline hazard from an exponential distribution and the censoring time was generated from a uniform distribution. The parameters led to a small number of observations with a misclassified event status (detailed in Appendix A).

* Type I error evaluated at log(1). Power evaluated at log(1.1), log(1.15), log(1.25), log(1.5), log(2).

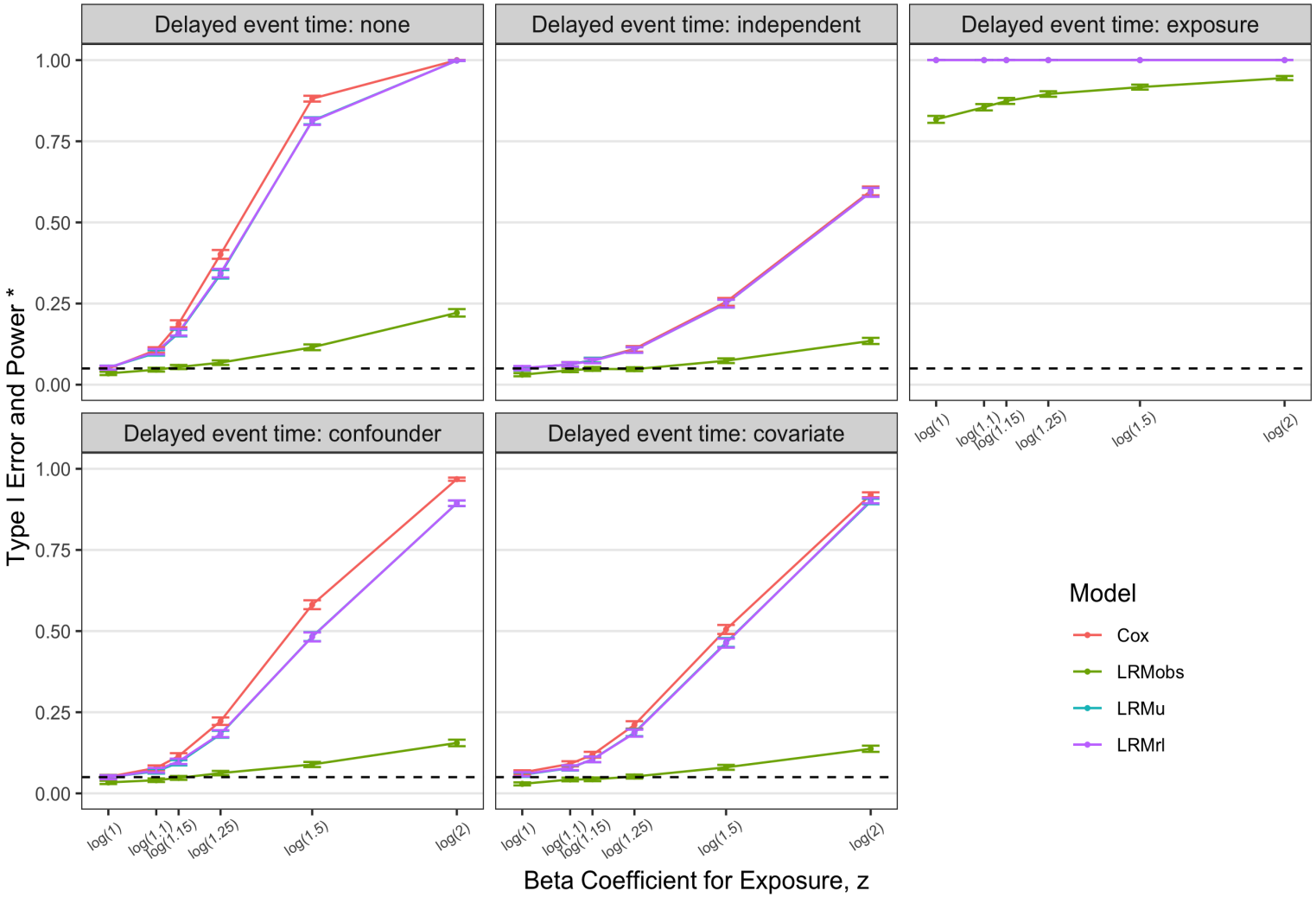


Figure 15: Results from Simulation 1 when the event time was generated from a Cox model with baseline hazard from a log-normal distribution and the censoring time was generated from a uniform distribution. The parameters led to a large number of observations with a misclassified event status (detailed in Appendix A).

* Type I error evaluated at log(1). Power evaluated at log(1.1), log(1.15), log(1.25), log(1.5), log(2).

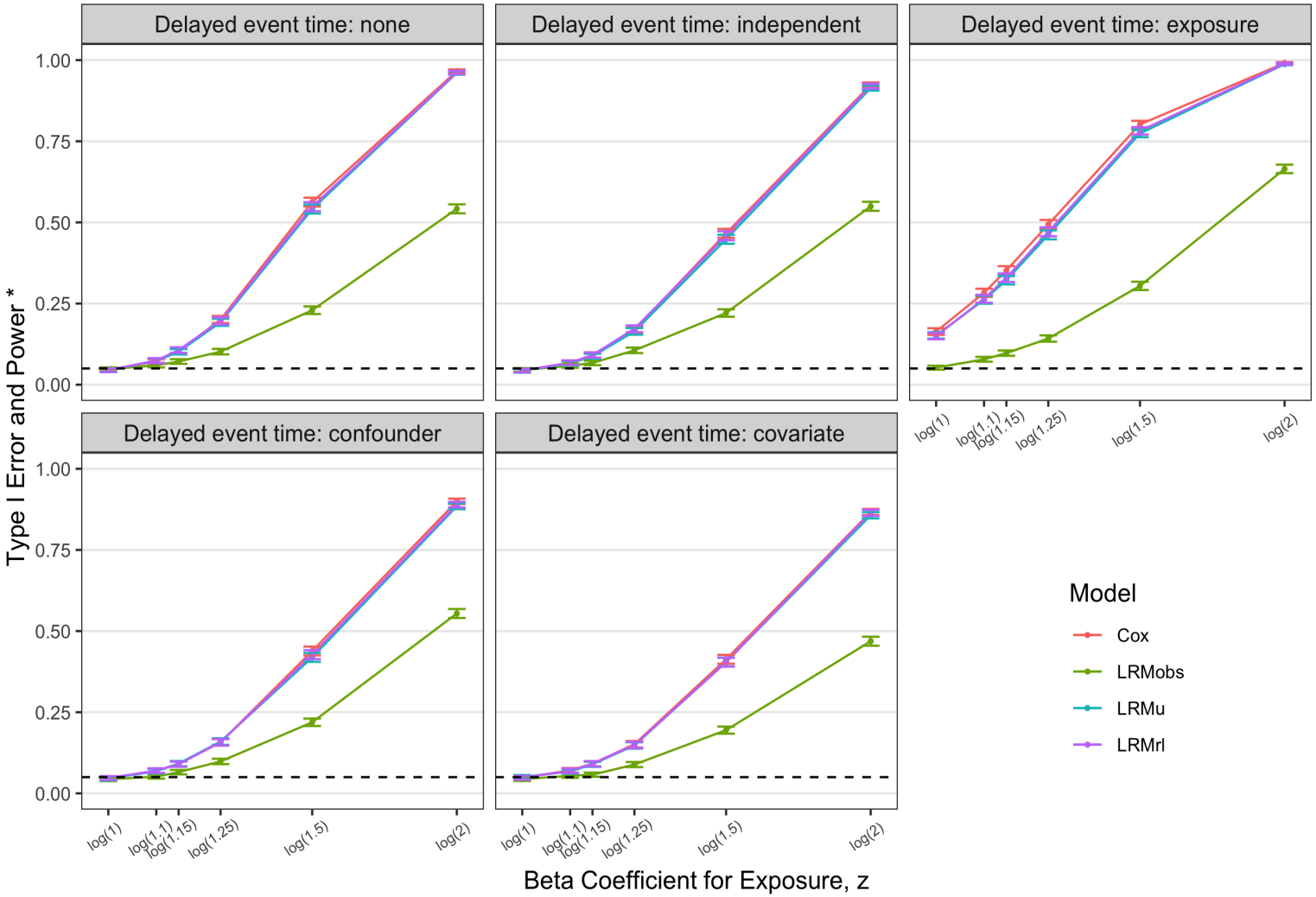


Figure 16: Results from Simulation 1 when the event time was generated from a Cox model with baseline hazard from a log-normal distribution and the censoring time was generated from a uniform distribution. The parameters led to a small number of observations with a misclassified event status (detailed in Appendix A).

* Type I error evaluated at log(1). Power evaluated at log(1.1), log(1.15), log(1.25), log(1.5), log(2).

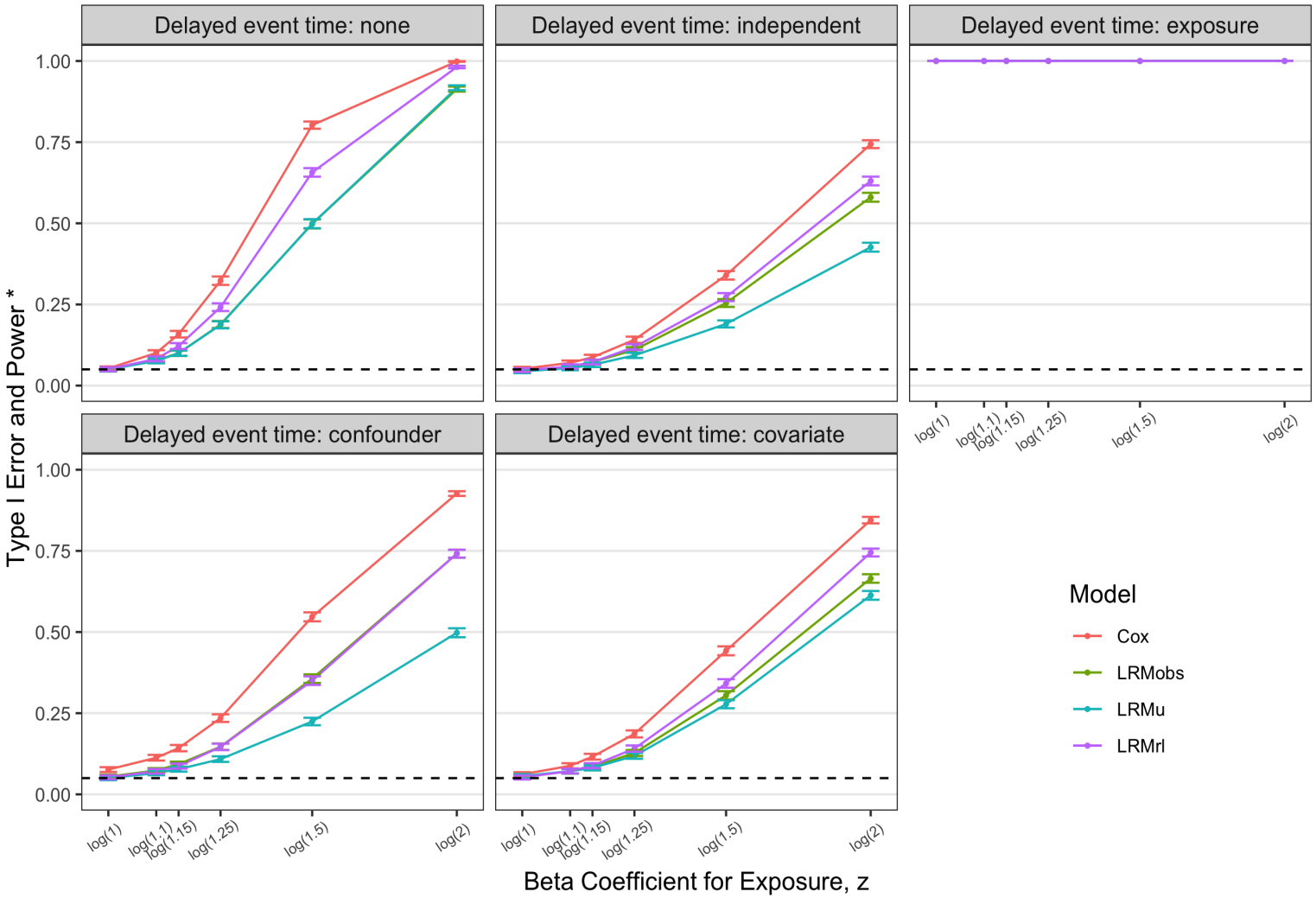


Figure 17: Results from Simulation 1 when the event time was generated from a Cox model with baseline hazard from an exponential distribution and the censoring time was generated from a Cox model with baseline hazard from an exponential distribution that depended on x. The parameters led to a large number of observations with a misclassified event status (detailed in Appendix A).

* Type I error evaluated at log(1). Power evaluated at log(1.1), log(1.15), log(1.25), log(1.5), log(2).

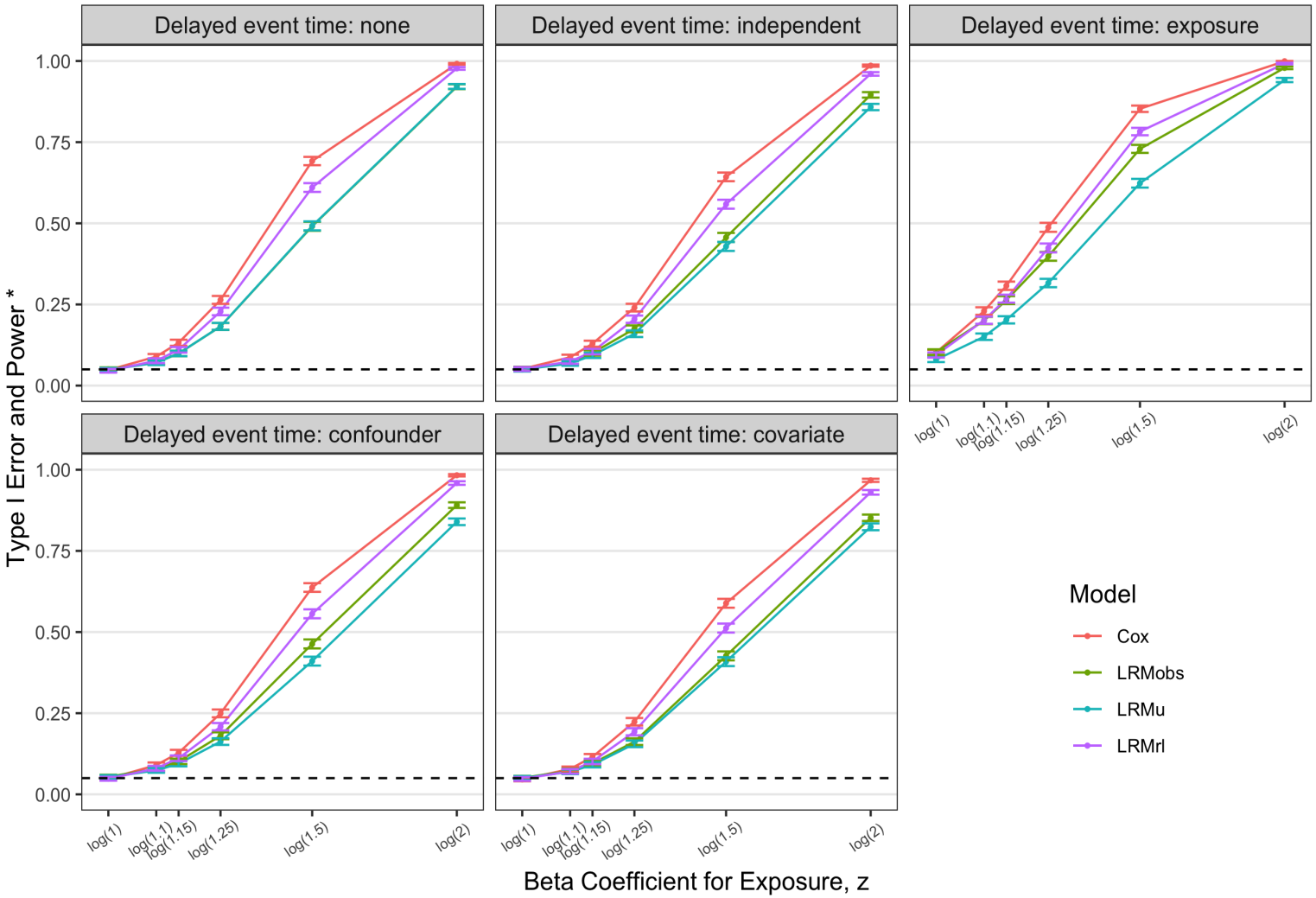


Figure 18: Results from Simulation 1 when the event time was generated from a Cox model with baseline hazard from an exponential distribution and the censoring time was generated from a Cox model with baseline hazard from an exponential distribution that depended on x. The parameters led to a small number of observations with a misclassified event status (detailed in Appendix A).

* Type I error evaluated at log(1). Power evaluated at log(1.1), log(1.15), log(1.25), log(1.5), log(2).

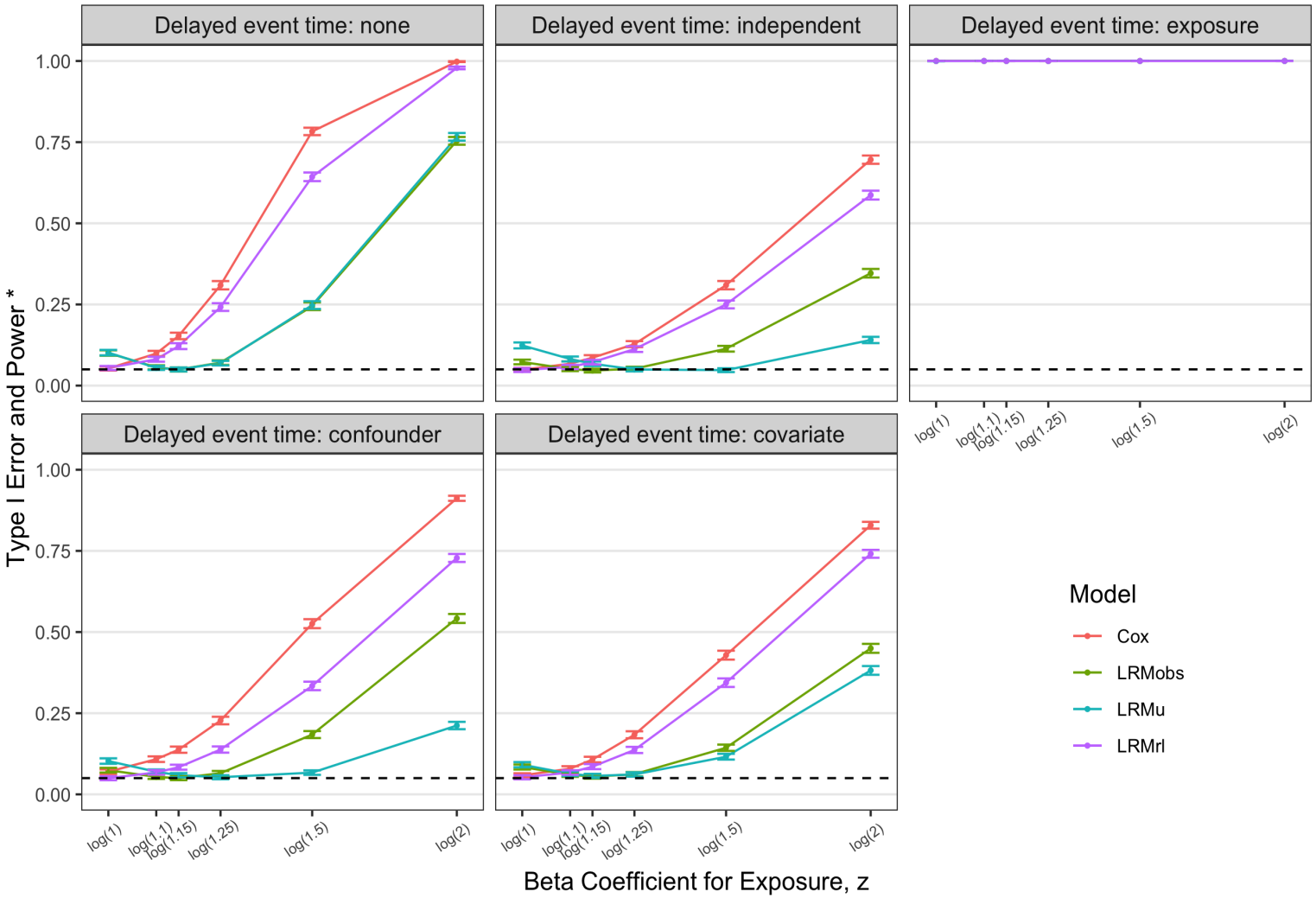


Figure 19: Results from Simulation 1 when the event time was generated from a Cox model with baseline hazard from an exponential distribution and the censoring time was generated from a Cox model with baseline hazard from an exponential distribution that depended on x and z. The parameters led to a large number of observations with a misclassified event status (detailed in Appendix A).

* Type I error evaluated at log(1). Power evaluated at log(1.1), log(1.15), log(1.25), log(1.5), log(2).

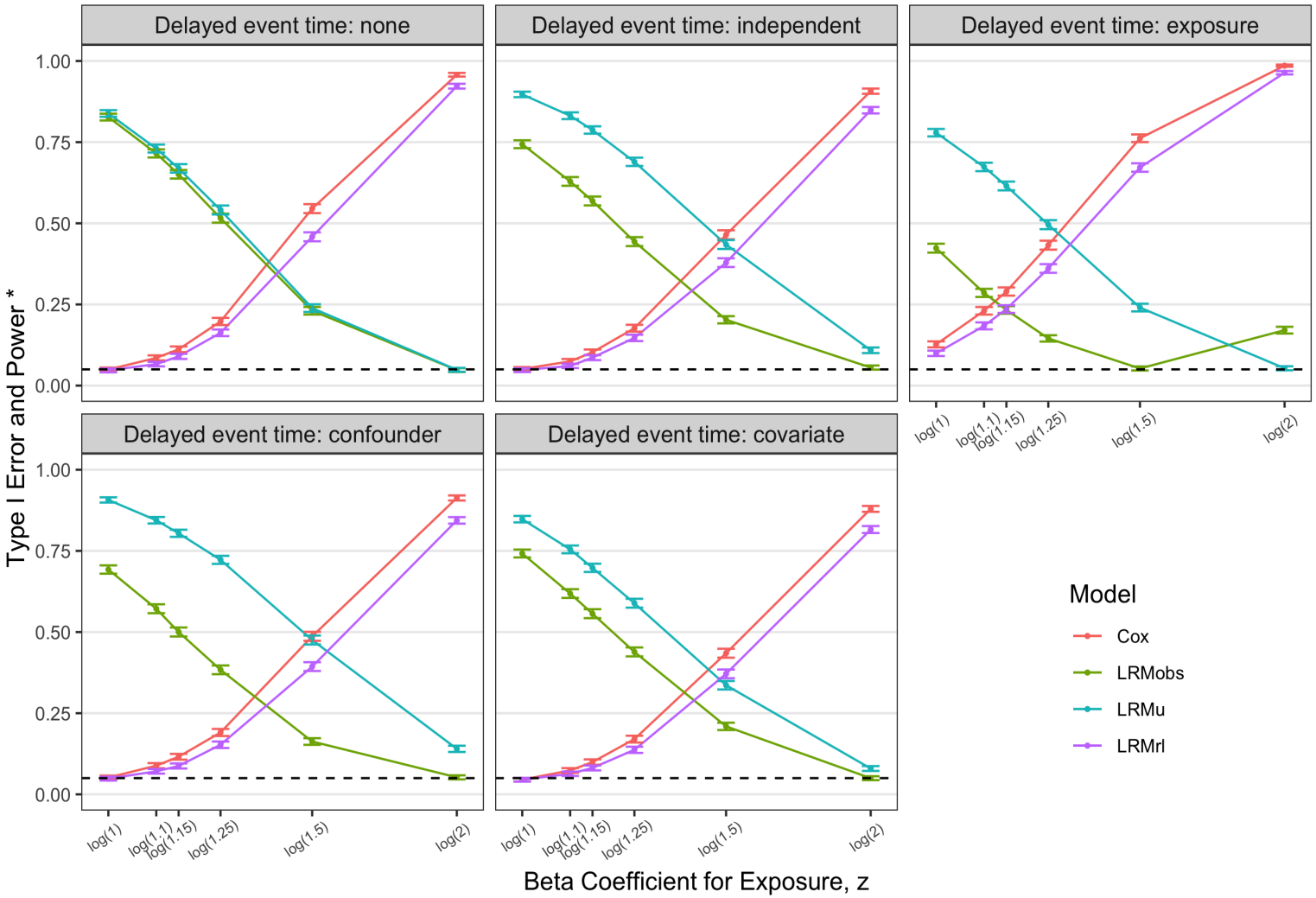


Figure 20: Results from Simulation 1 when the event time was generated from a Cox model with baseline hazard from an exponential distribution and the censoring time was generated from a Cox model with baseline hazard from an exponential distribution that depended on x and z. The parameters led to a small number of observations with a misclassified event status (detailed in Appendix A).

* Type I error evaluated at log(1). Power evaluated at log(1.1), log(1.15), log(1.25), log(1.5), log(2).

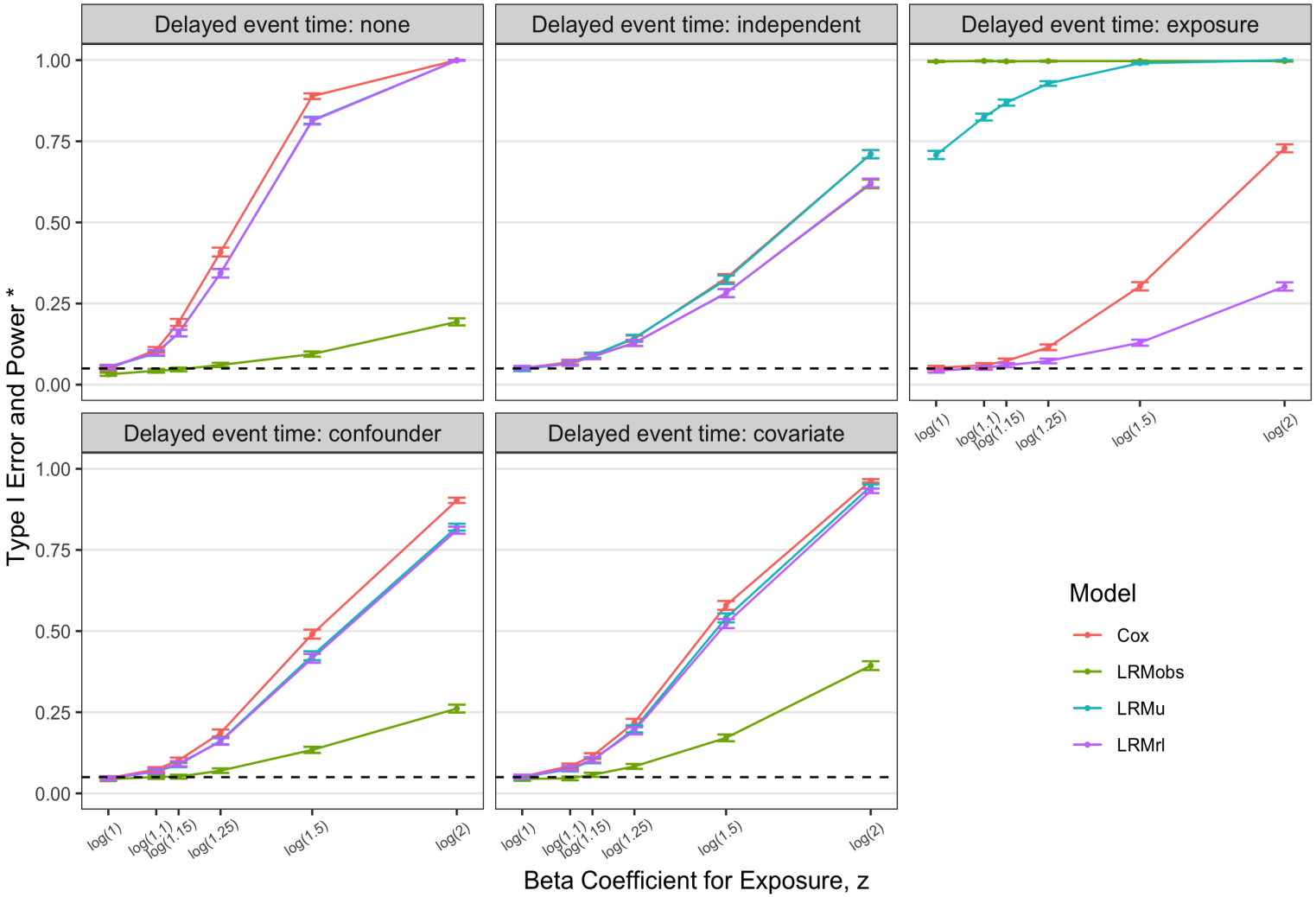


Figure 21: Results from Simulation 2 when the event time was generated from a Cox model with baseline hazard from an exponential distribution and the censoring time was generated from a uniform distribution. The parameters are the same as those in Figure 13.

* Type I error evaluated at log(1). Power evaluated at log(1.1), log(1.15), log(1.25), log(1.5), log(2).

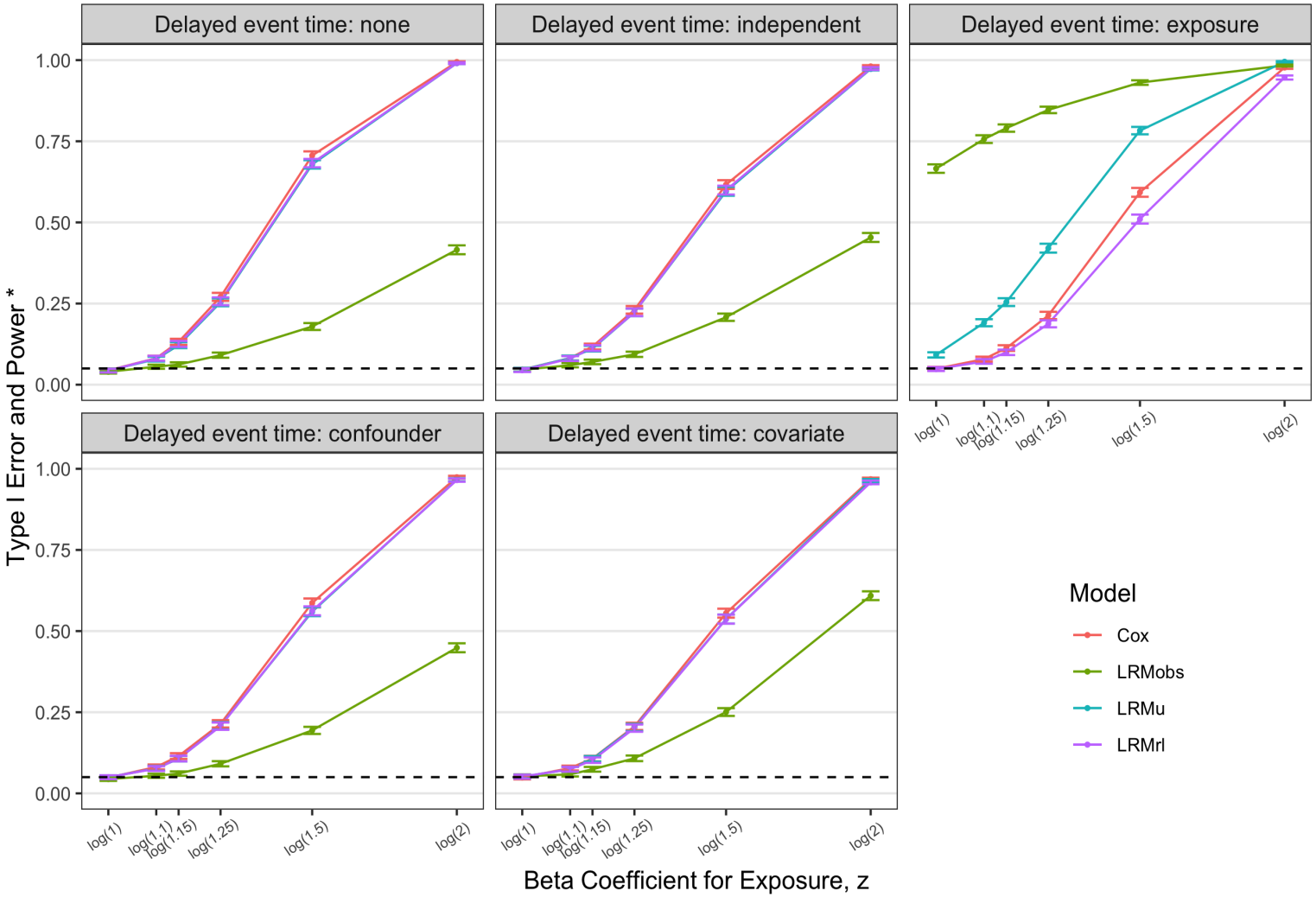


Figure 22: Results from Simulation 2 when the event time was generated from a Cox model with baseline hazard from an exponential distribution and the censoring time was generated from a uniform distribution. The parameters are the same as those in Figure 14.

* Type I error evaluated at log(1). Power evaluated at log(1.1), log(1.15), log(1.25), log(1.5), log(2).

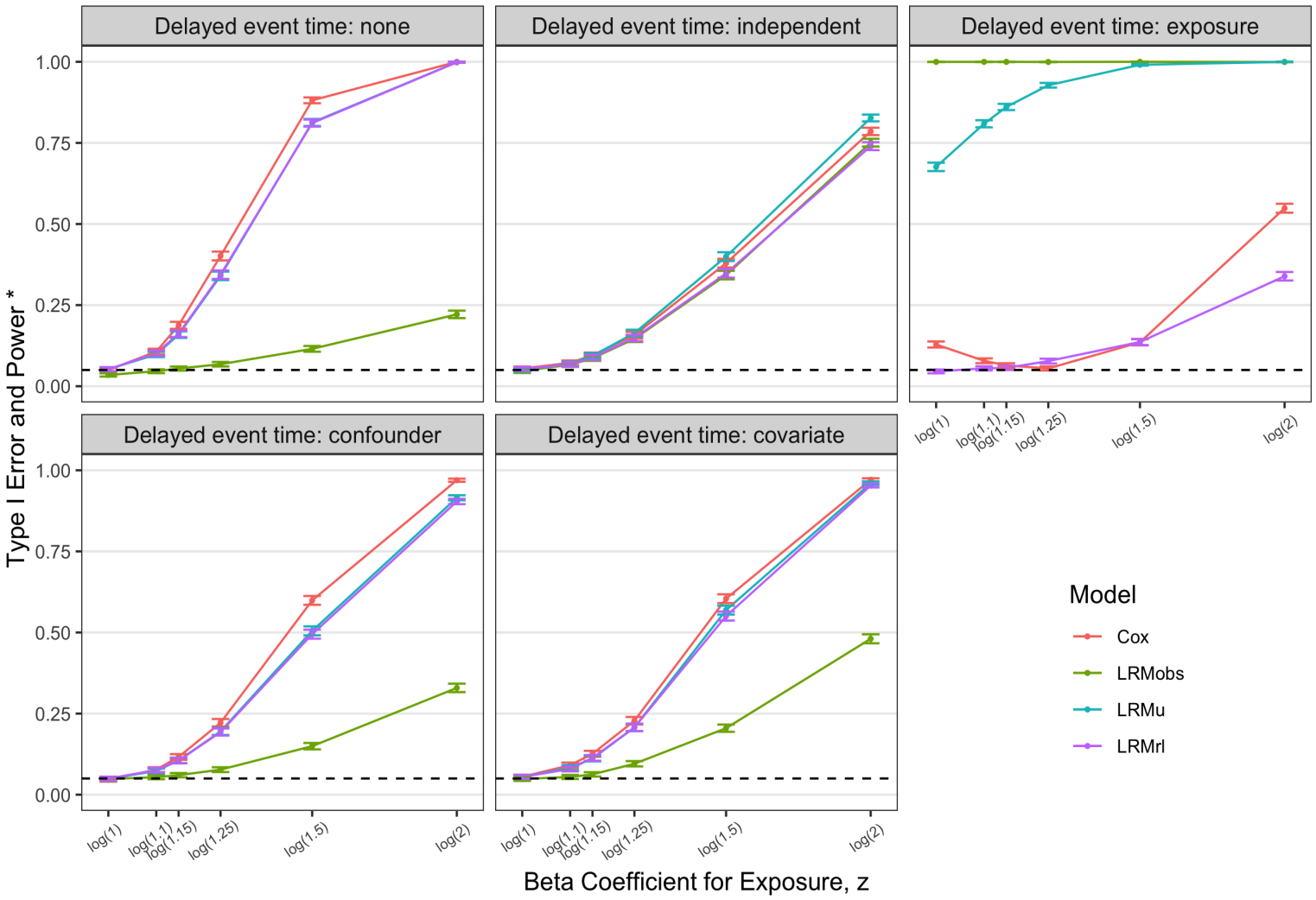


Figure 23: Results from Simulation 2 when the event time was generated from a Cox model with baseline hazard from a log-normal distribution and the censoring time was generated from a uniform distribution. The parameters are the same as those in Figure 15.

* Type I error evaluated at log(1). Power evaluated at log(1.1), log(1.15), log(1.25), log(1.5), log(2).

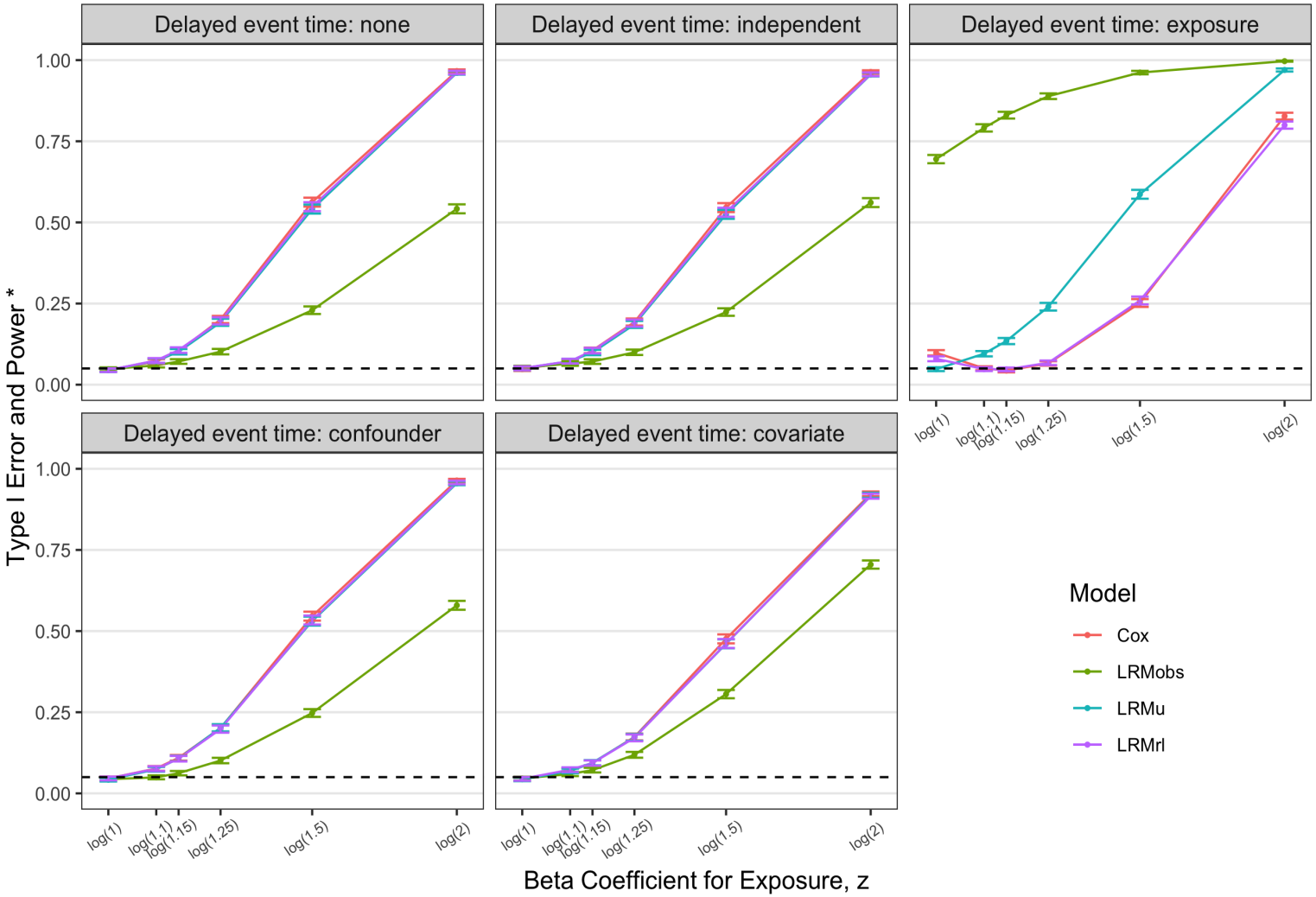


Figure 24: Results from Simulation 2 when the event time was generated from a Cox model with baseline hazard from a log-normal distribution and the censoring time was generated from a uniform distribution. The parameters are the same as those in Figure 16.

* Type I error evaluated at log(1). Power evaluated at log(1.1), log(1.15), log(1.25), log(1.5), log(2).

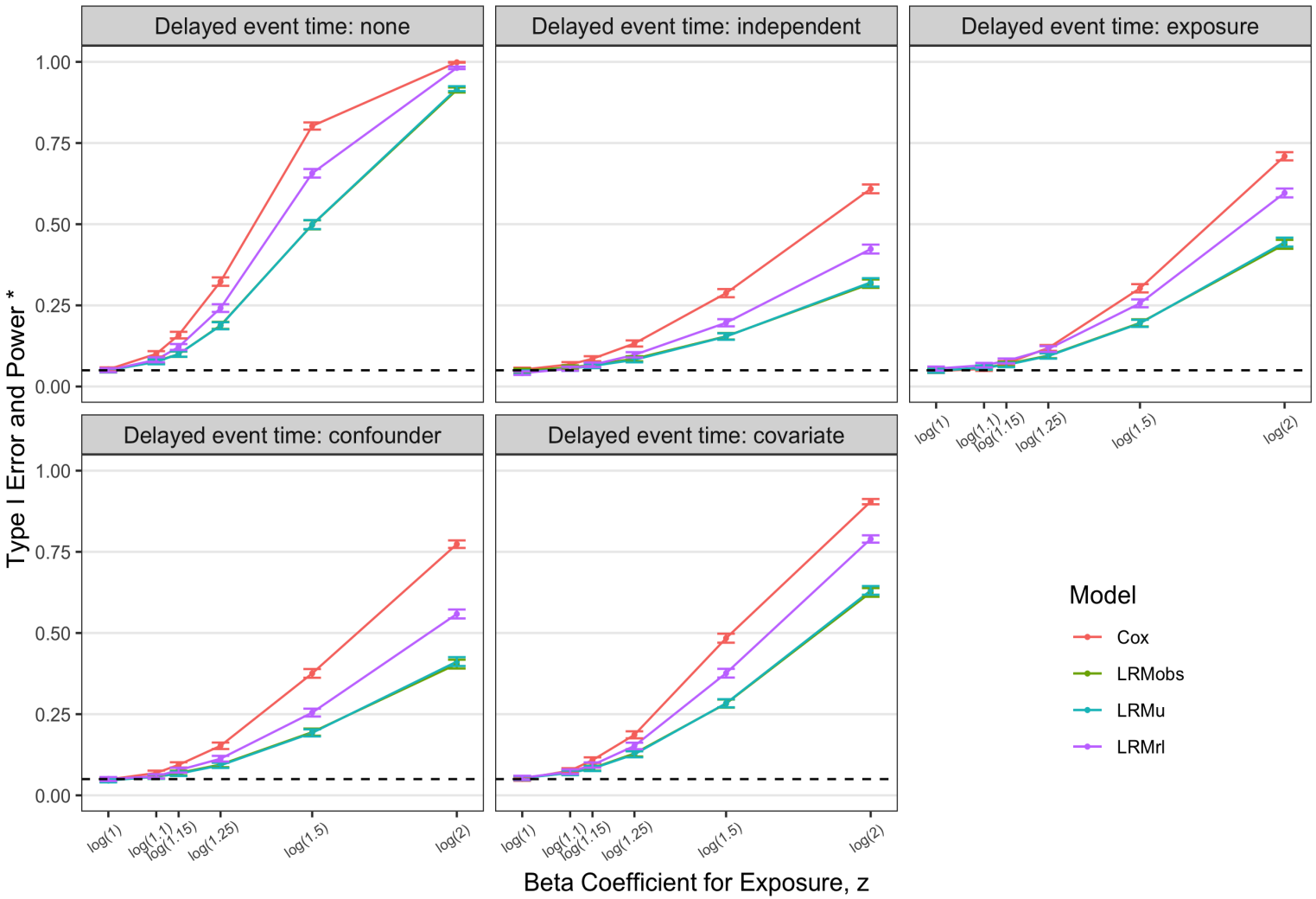


Figure 25: Results from Simulation 2 when the event time was generated from a Cox model with baseline hazard from an exponential distribution and the censoring time was generated from a Cox model with baseline hazard from an exponential distribution that depended on x. The parameters are the same as those in Figure 17.

* Type I error evaluated at log(1). Power evaluated at log(1.1), log(1.15), log(1.25), log(1.5), log(2).

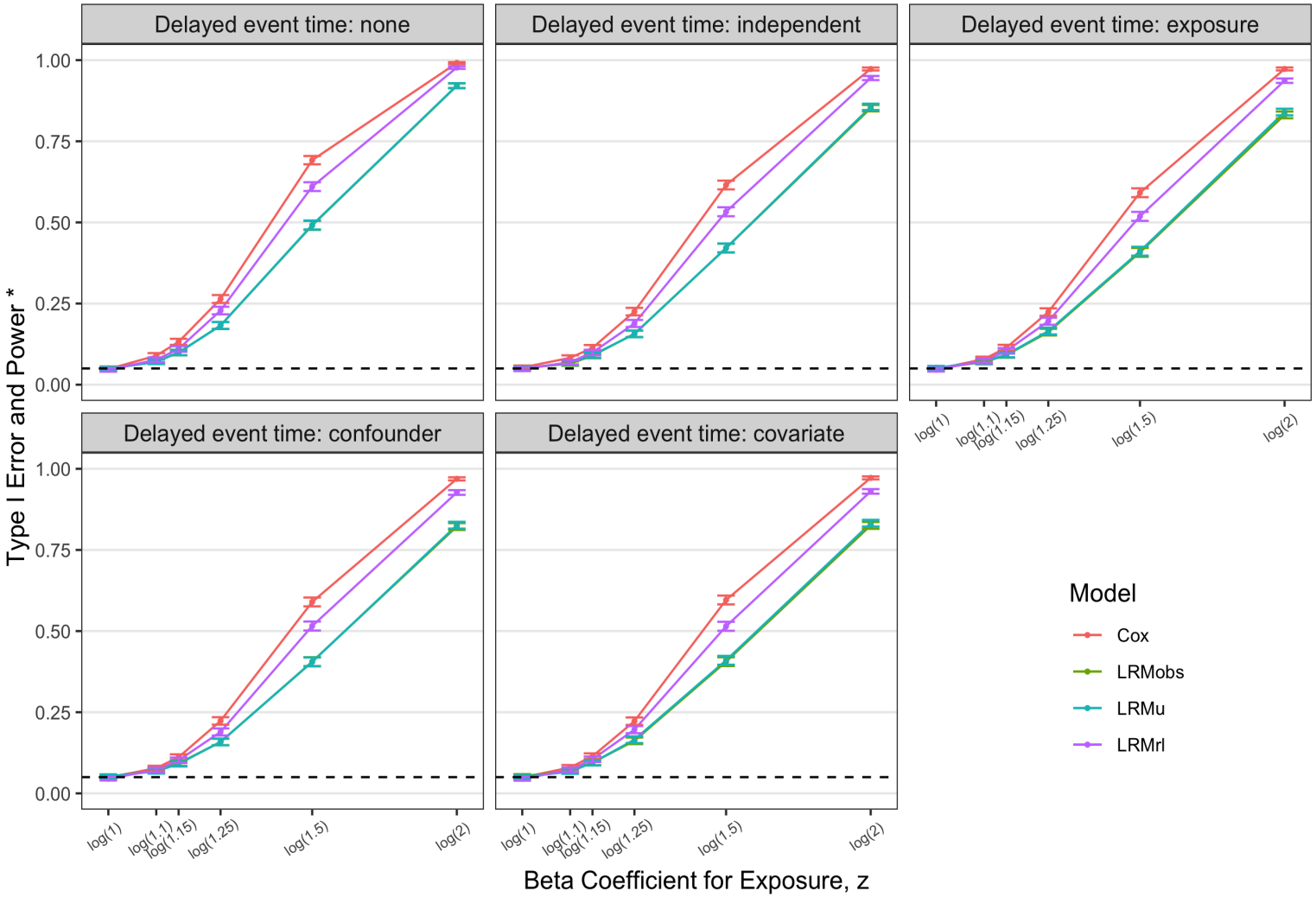


Figure 26: Results from Simulation 2 when the event time was generated from a Cox model with baseline hazard from an exponential distribution and the censoring time was generated from a Cox model with baseline hazard from an exponential distribution that depended on x. The parameters are the same as those in Figure 18.

* Type I error evaluated at log(1). Power evaluated at log(1.1), log(1.15), log(1.25), log(1.5), log(2).

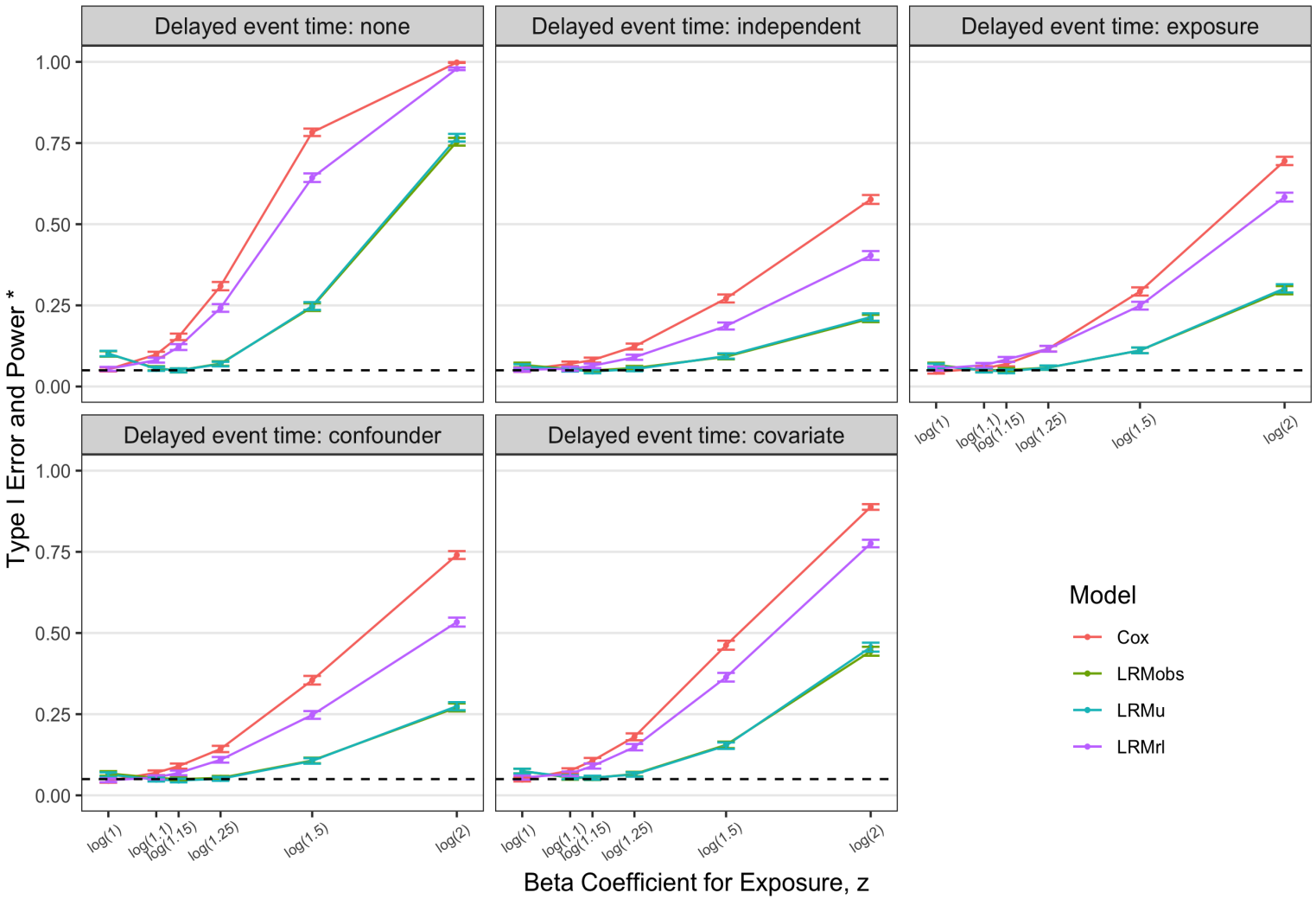


Figure 27: Results from Simulation 2 when the event time was generated from a Cox model with baseline hazard from an exponential distribution and the censoring time was generated from a Cox model with baseline hazard from an exponential distribution that depended on x and z. The parameters are the same as those in Figure 19.

* Type I error evaluated at log(1). Power evaluated at log(1.1), log(1.15), log(1.25), log(1.5), log(2).

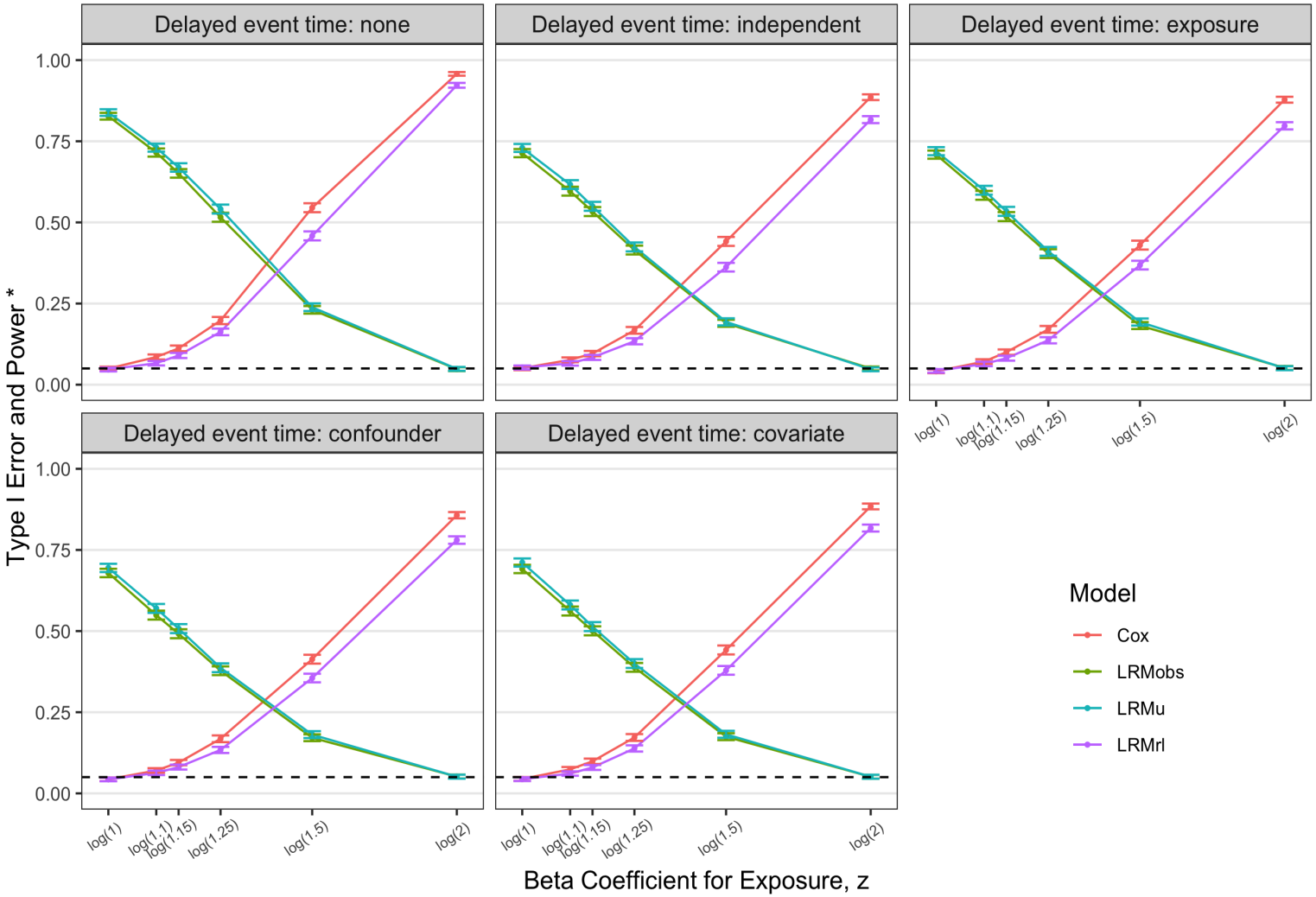


Figure 28: Results from Simulation 2 when the event time was generated from a Cox model with baseline hazard from an exponential distribution and the censoring time was generated from a Cox model with baseline hazard from an exponential distribution that depended on x and z. The parameters are the same as those in Figure 20.

* Type I error evaluated at log(1). Power evaluated at log(1.1), log(1.15), log(1.25), log(1.5), log(2).

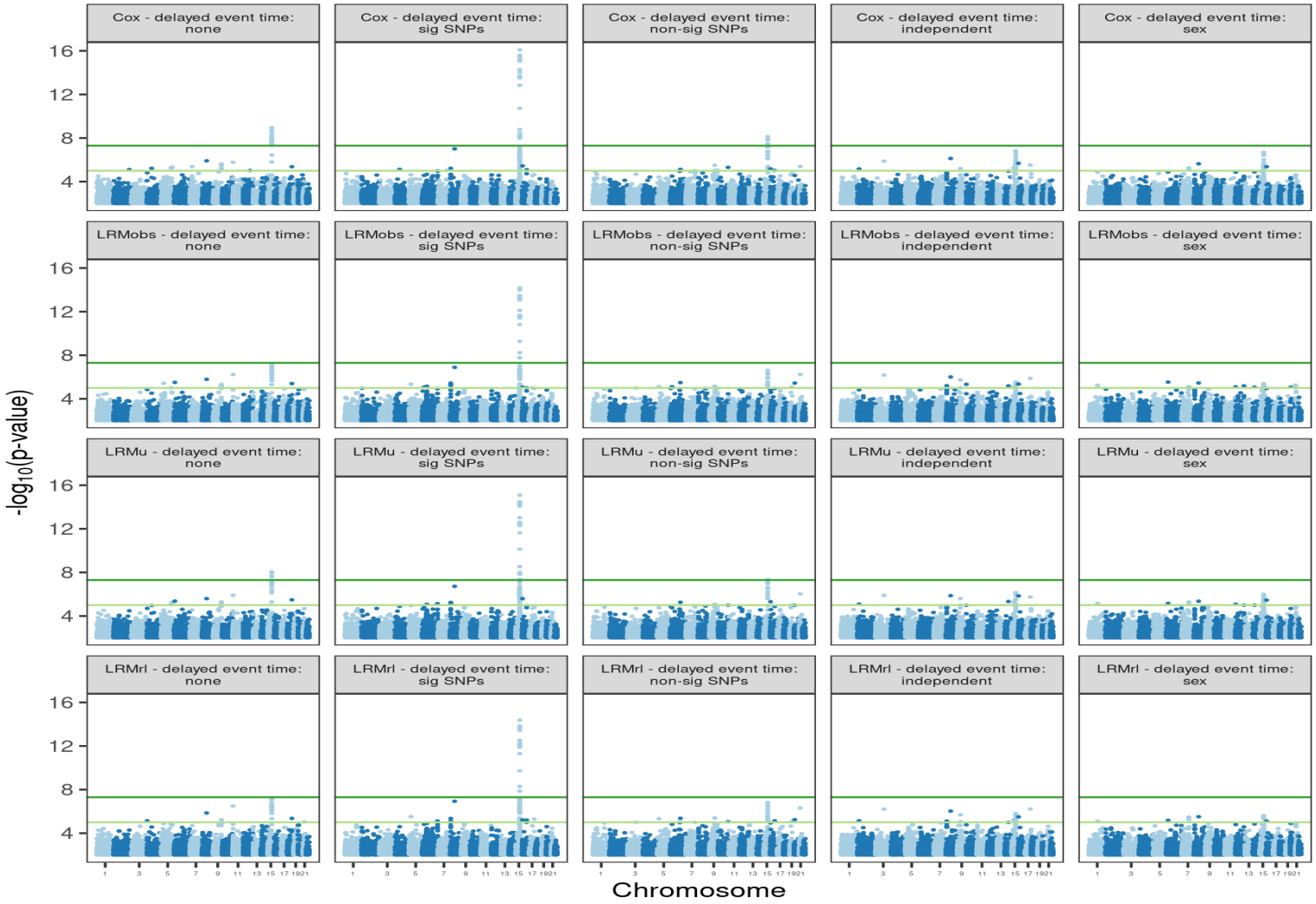


Figure 29: Manhattan plots of GWAS results for cancer of bronchus; lung (phecode 165.1) for each model and delayed event time combination. The dark green line corresponds to $P \leq 5 \times 10^{-8}$ and the light green line corresponds to $P \leq 1 \times 10^{-5}$.

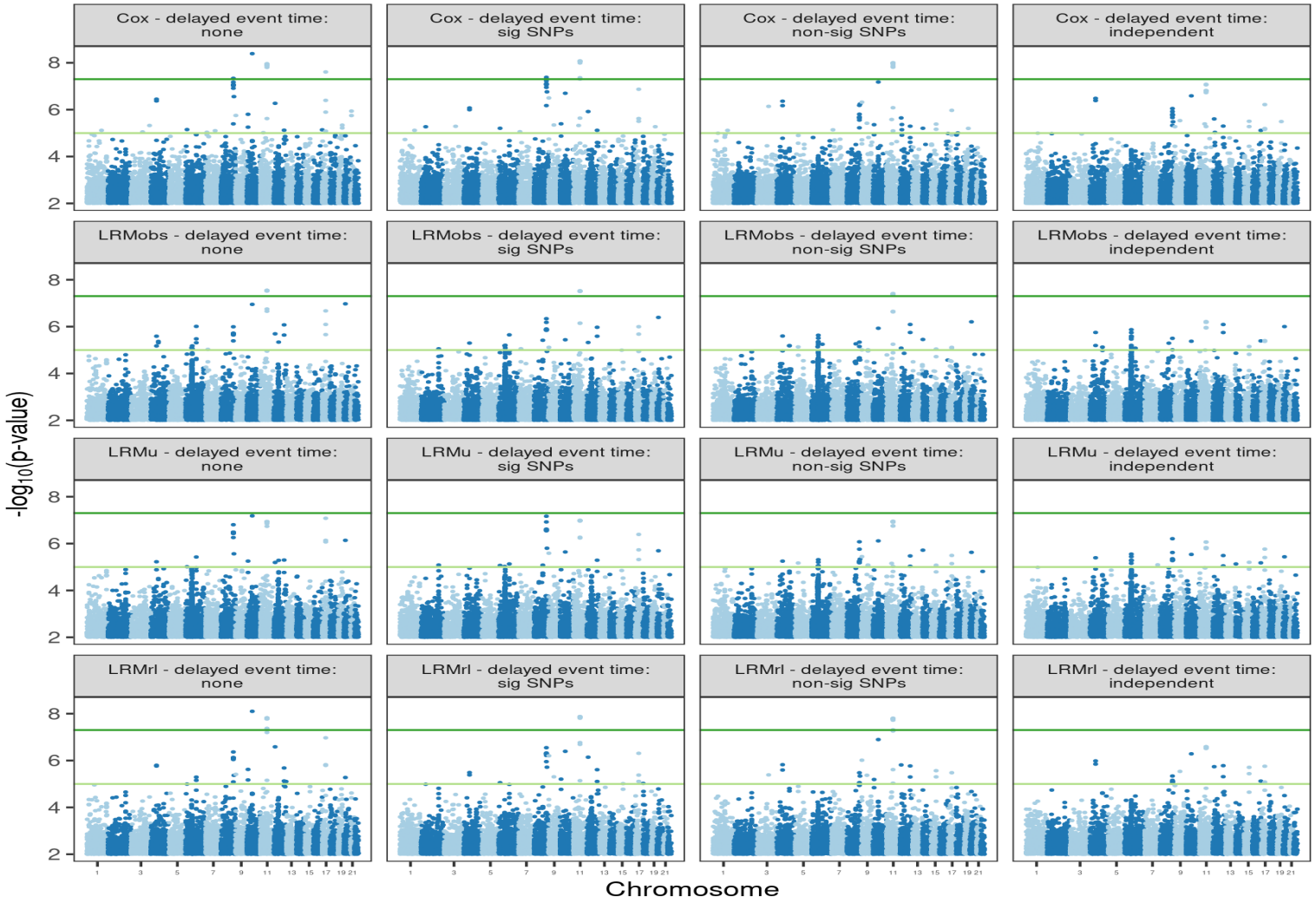


Figure 30: Manhattan plots of GWAS results for cancer of prostate (phecode 185) for each model and delayed event time combination. The dark green line corresponds to $P \leq 5 \times 10^{-8}$ and the light green line corresponds to $P \leq 1 \times 10^{-5}$.

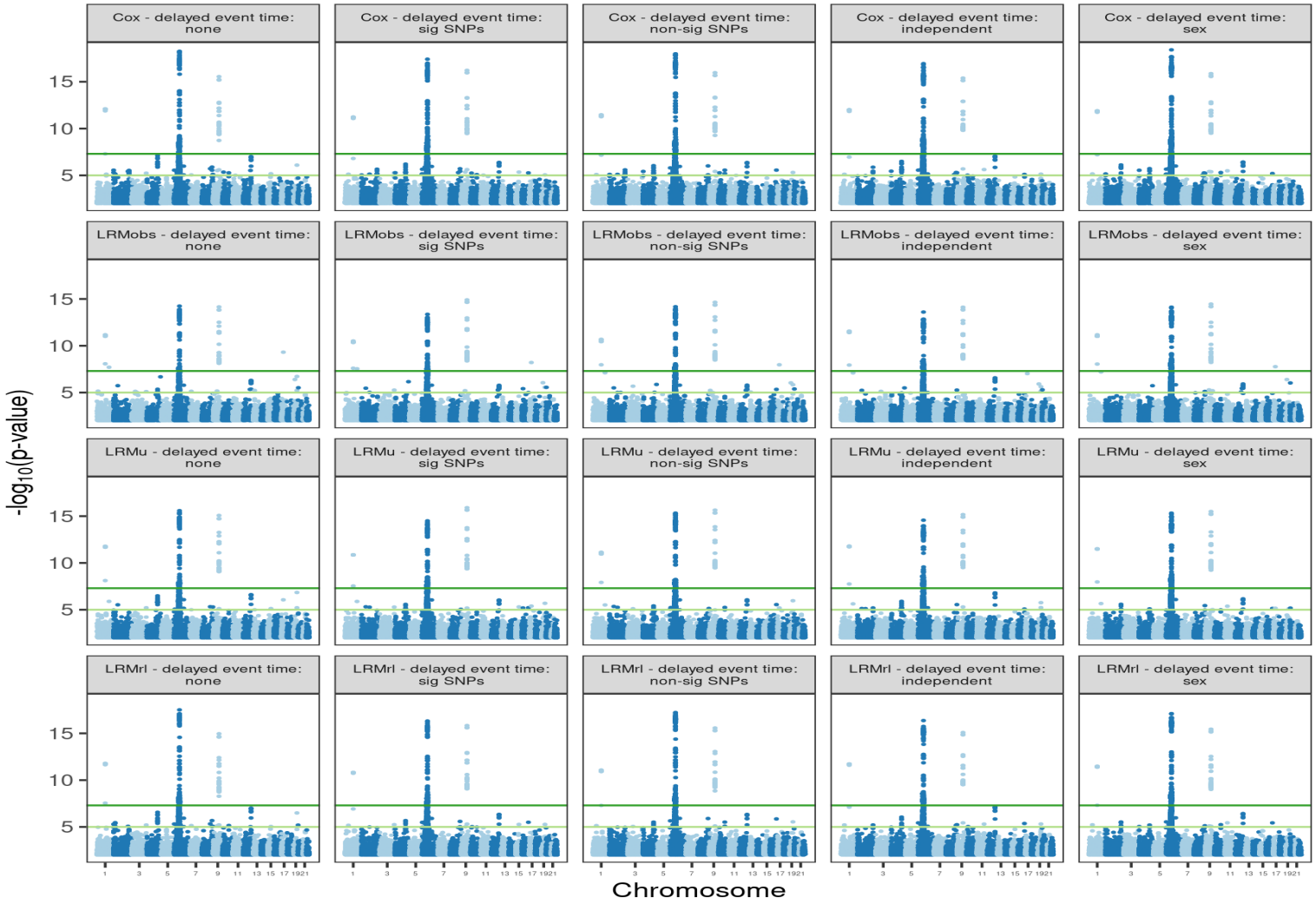


Figure 31: Manhattan plots of GWAS results for hypothyroidism (phecode 244) for each model and delayed event time combination. The dark green line corresponds to $P \leq 5 \times 10^{-8}$ and the light green line corresponds to $P \leq 1 \times 10^{-5}$.

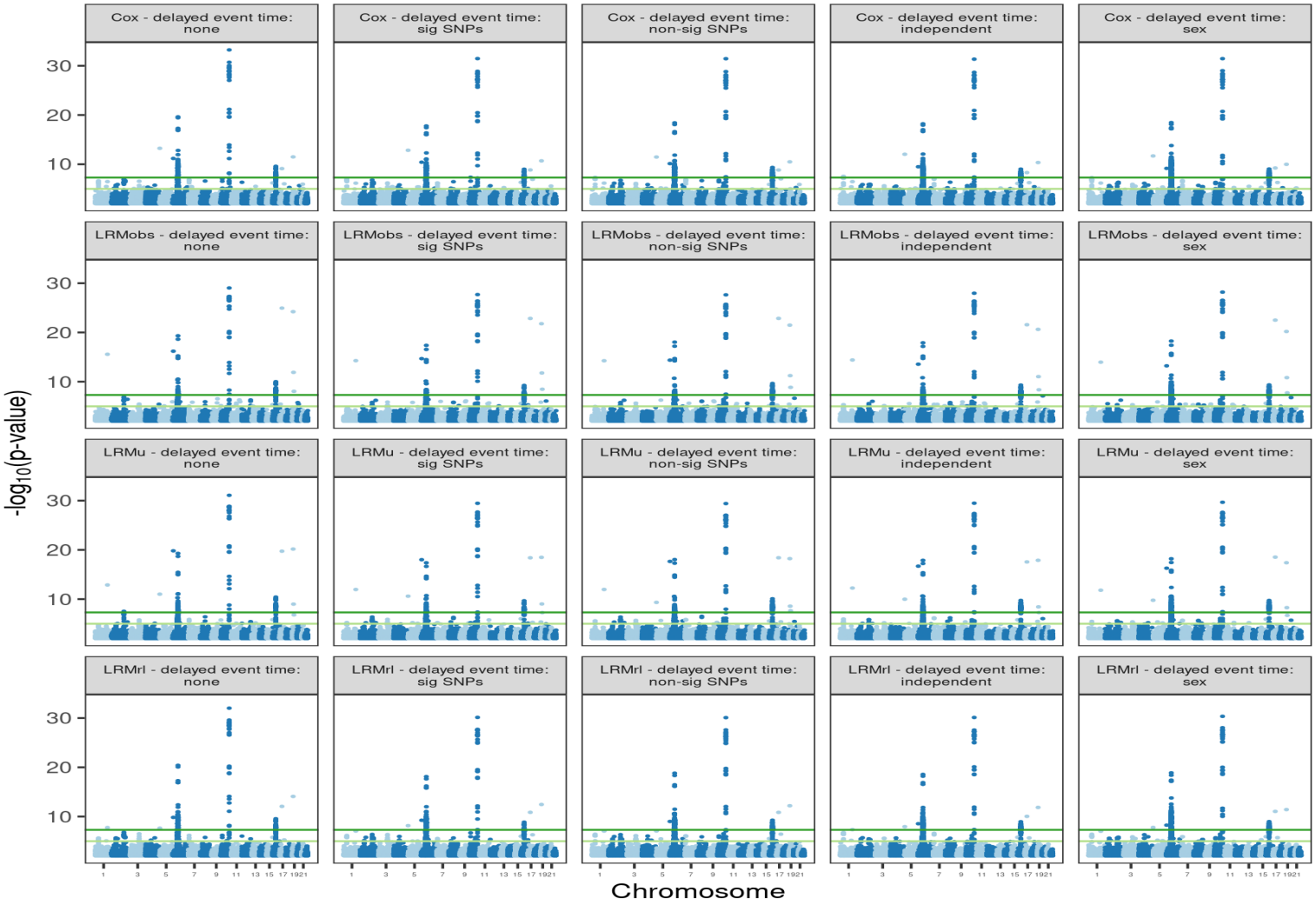


Figure 32: Manhattan plots of GWAS results for type 2 diabetes (phecode 250.2) for each model and delayed event time combination. The dark green line corresponds to $P \leq 5 \times 10^{-8}$ and the light green line corresponds to $P \leq 1 \times 10^{-5}$.

Figure 33: Manhattan plots of GWAS results for vitamin D deficiency (phecode 261.4) for each model and delayed event time combination. The dark green line corresponds to $P \leq 5 \times 10^{-8}$ and the light green line corresponds to $P \leq 1 \times 10^{-5}$.

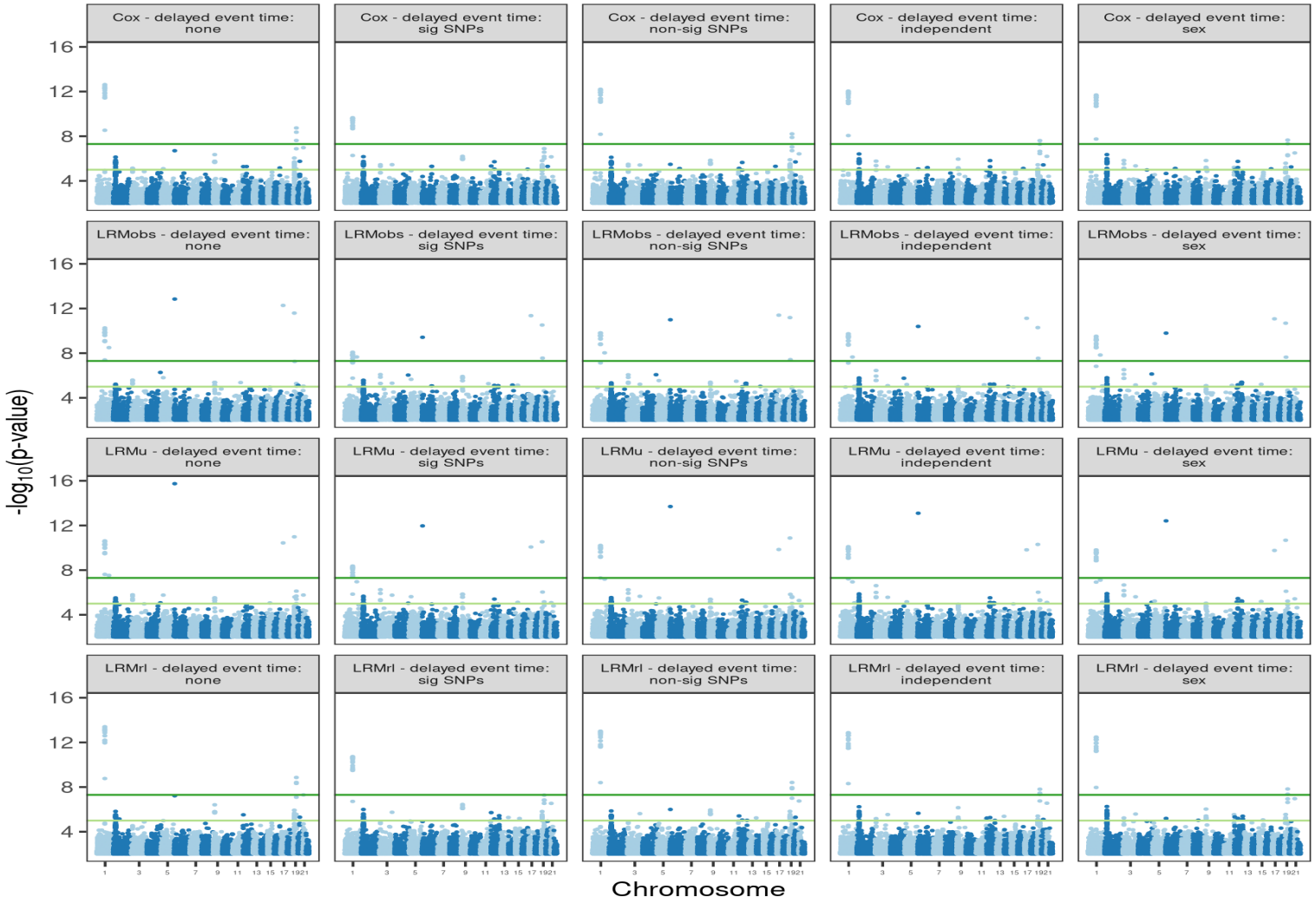


Figure 34: Manhattan plots of GWAS results for hypercholesterolemia (phecode 272.11) for each model and delayed event time combination. The dark green line corresponds to $P \leq 5 \times 10^{-8}$ and the light green line corresponds to $P \leq 1 \times 10^{-5}$.

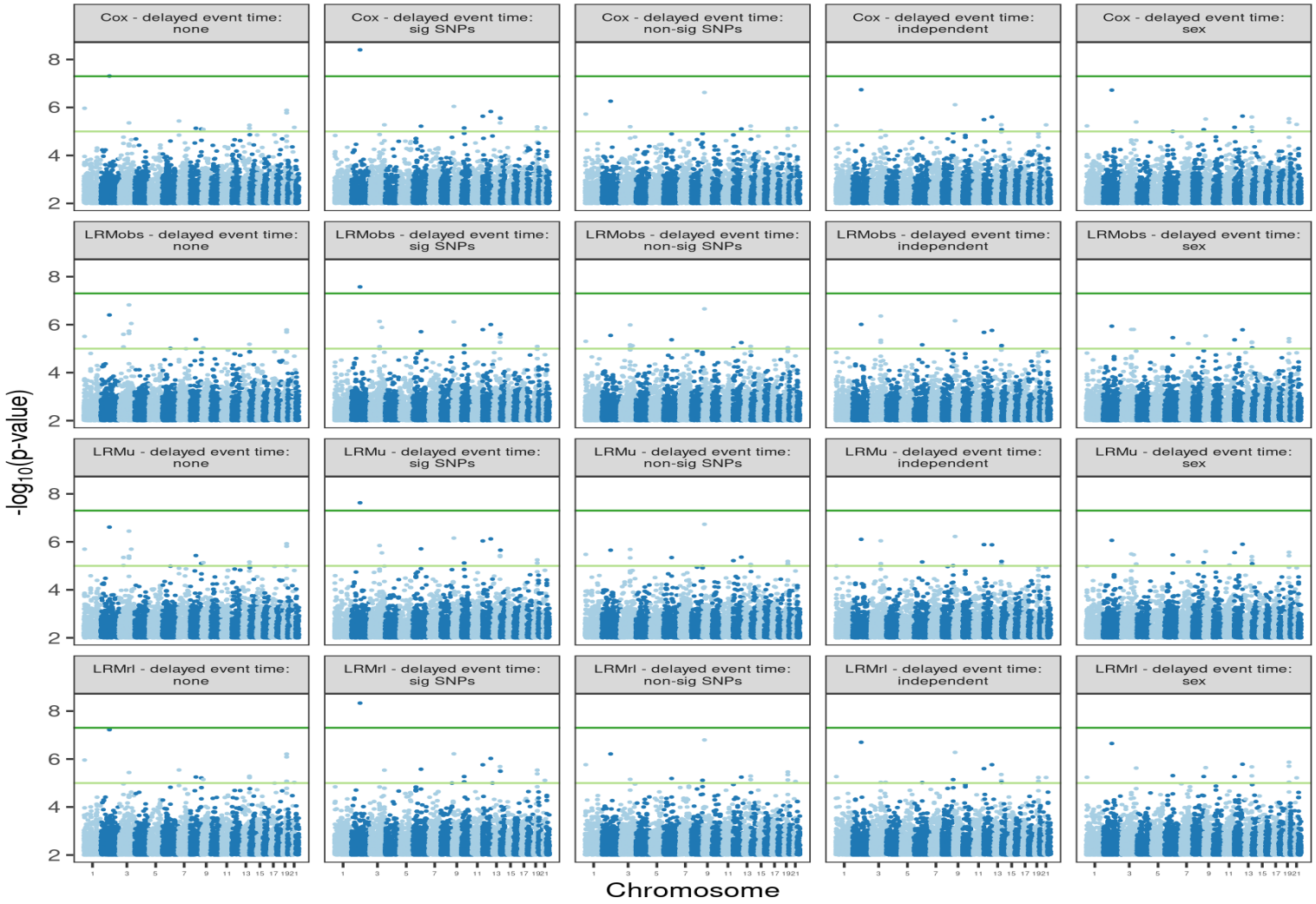


Figure 35: Manhattan plots of GWAS results for insomnia (phecode 327.4) for each model and delayed event time combination. The dark green line corresponds to $P \leq 5 \times 10^{-8}$ and the light green line corresponds to $P \leq 1 \times 10^{-5}$.

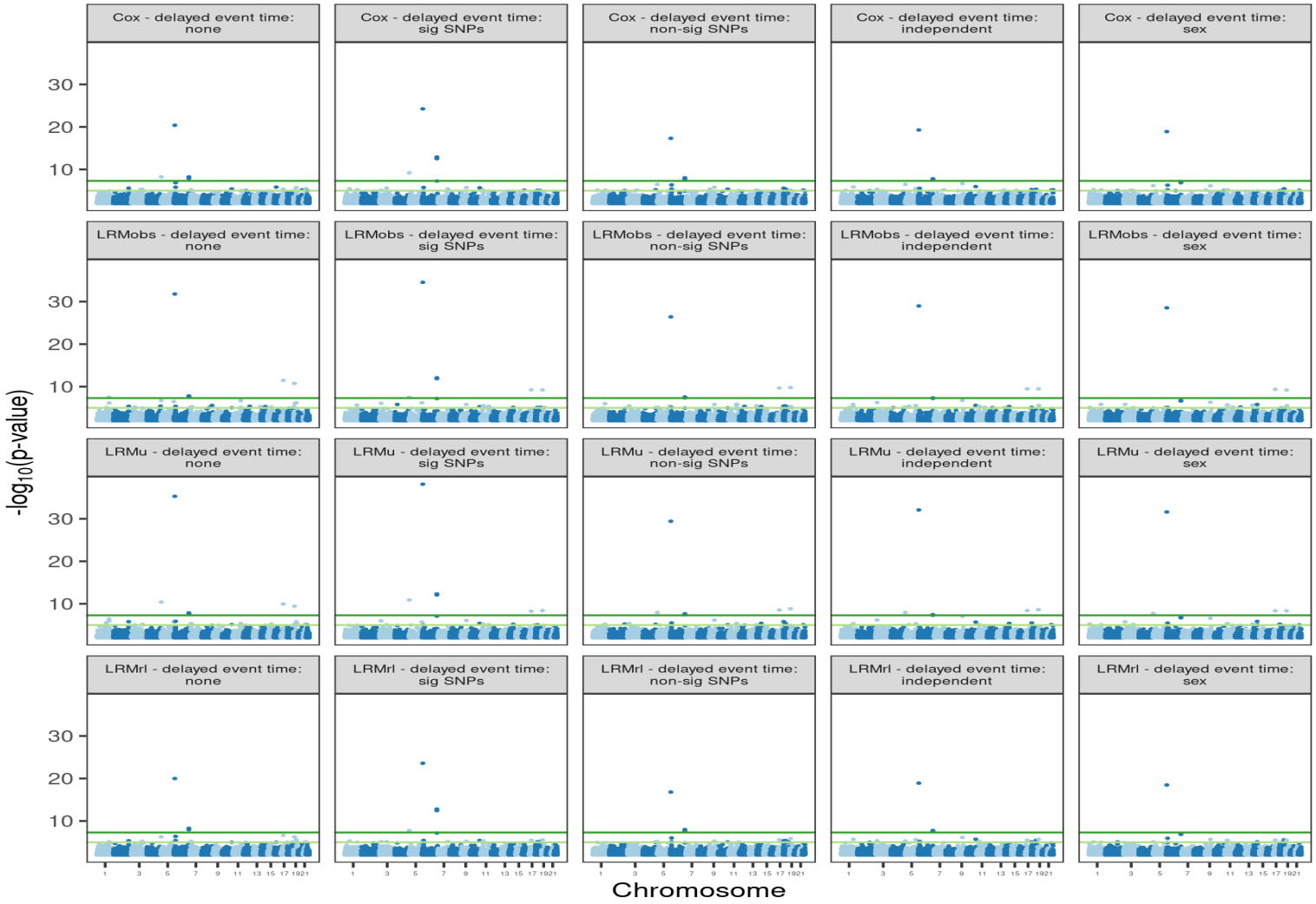


Figure 36: Manhattan plots of GWAS results for myocardial infarction (phecode 411.2) for each model and delayed event time combination. The dark green line corresponds to $P \leq 5 \times 10^{-8}$ and the light green line corresponds to $P \leq 1 \times 10^{-5}$.

Figure 37: Manhattan plots of GWAS results for coronary atherosclerosis (phecode 411.4) for each model and delayed event time combination. The dark green line corresponds to $P \leq 5 \times 10^{-8}$ and the light green line corresponds to $P \leq 1 \times 10^{-5}$.

Figure 38: Manhattan plots of GWAS results for atrial fibrillation (phecode 427.21) for each model and delayed event time combination. The dark green line corresponds to $P \leq 5 \times 10^{-8}$ and the light green line corresponds to $P \leq 1 \times 10^{-5}$.

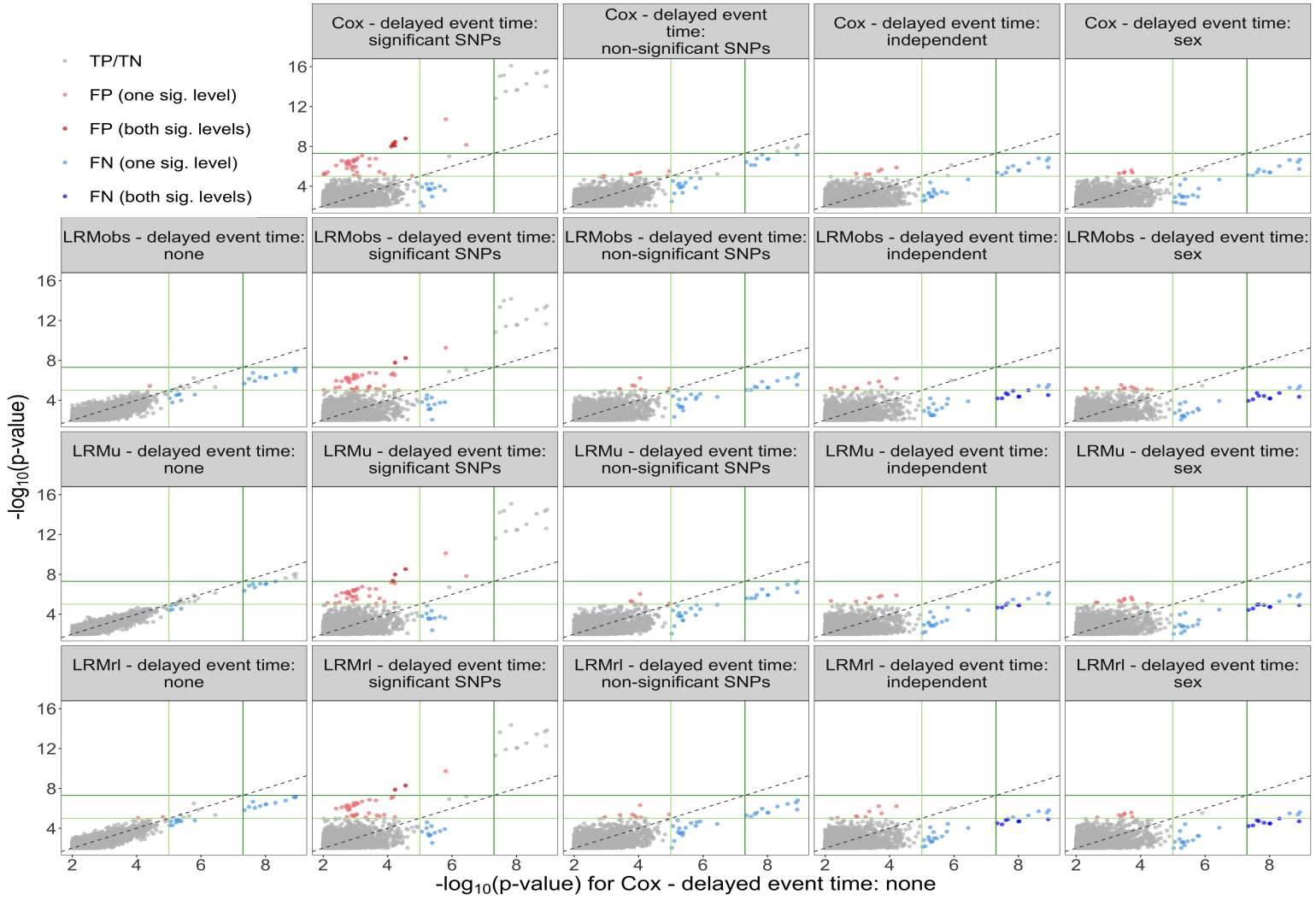


Figure 39: False positive and false negative SNPs for each model and delayed event time combination, using Model 1 (Cox) with no delayed event time as the gold standard, for cancer of bronchus; lung (phecode 165.1). Dark green lines correspond to $P \leq 5 \times 10^{-8}$ and light green lines correspond to $P \leq 1 \times 10^{-5}$.

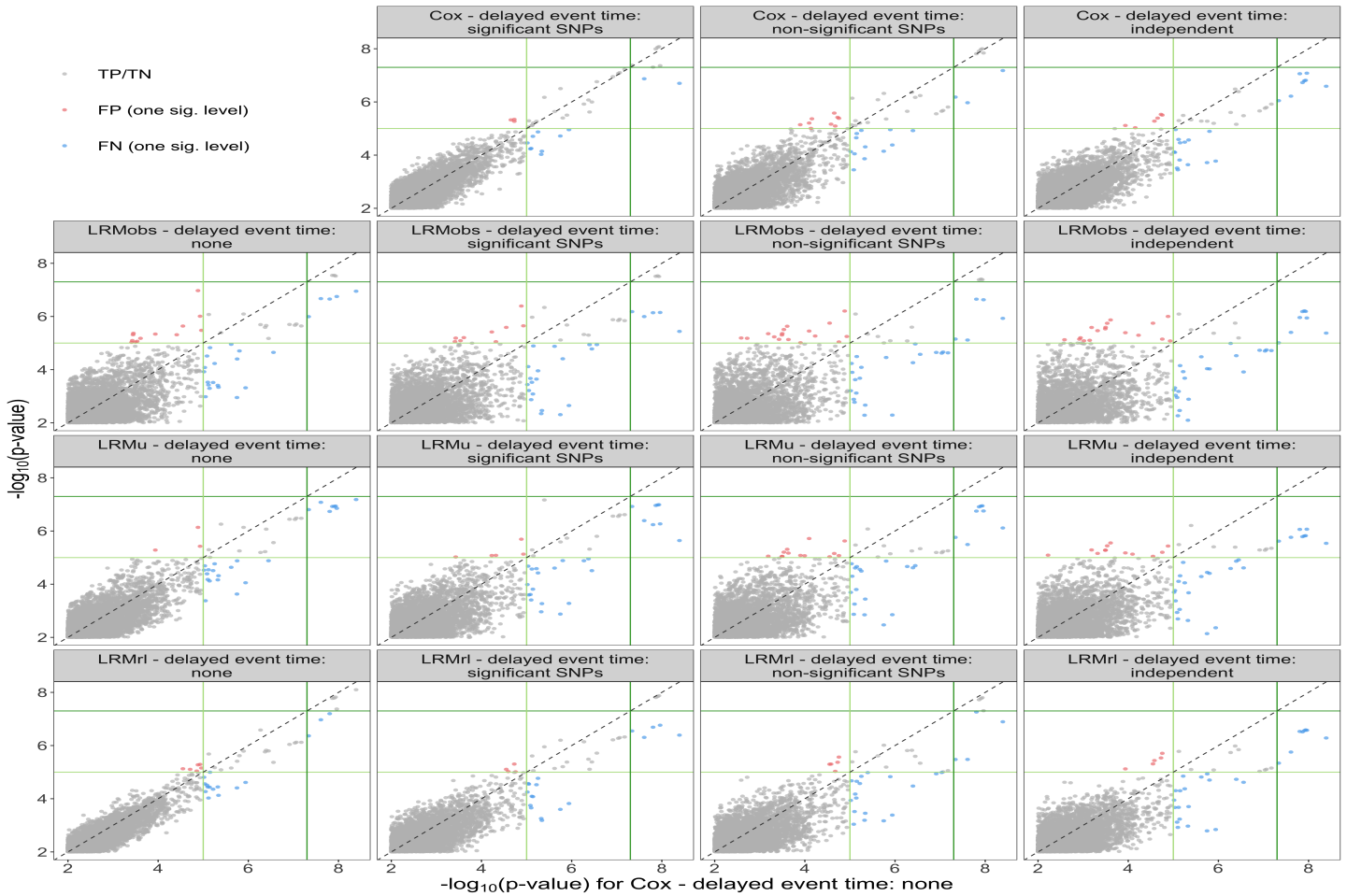


Figure 40: False positive and false negative SNPs for each model and delayed event time combination, using Model 1 (Cox) with no delayed event time as the gold standard, for cancer of prostate (phecode 185). Dark green lines correspond to $P \leq 5 \times 10^{-8}$ and light green lines correspond to $P \leq 1 \times 10^{-5}$.

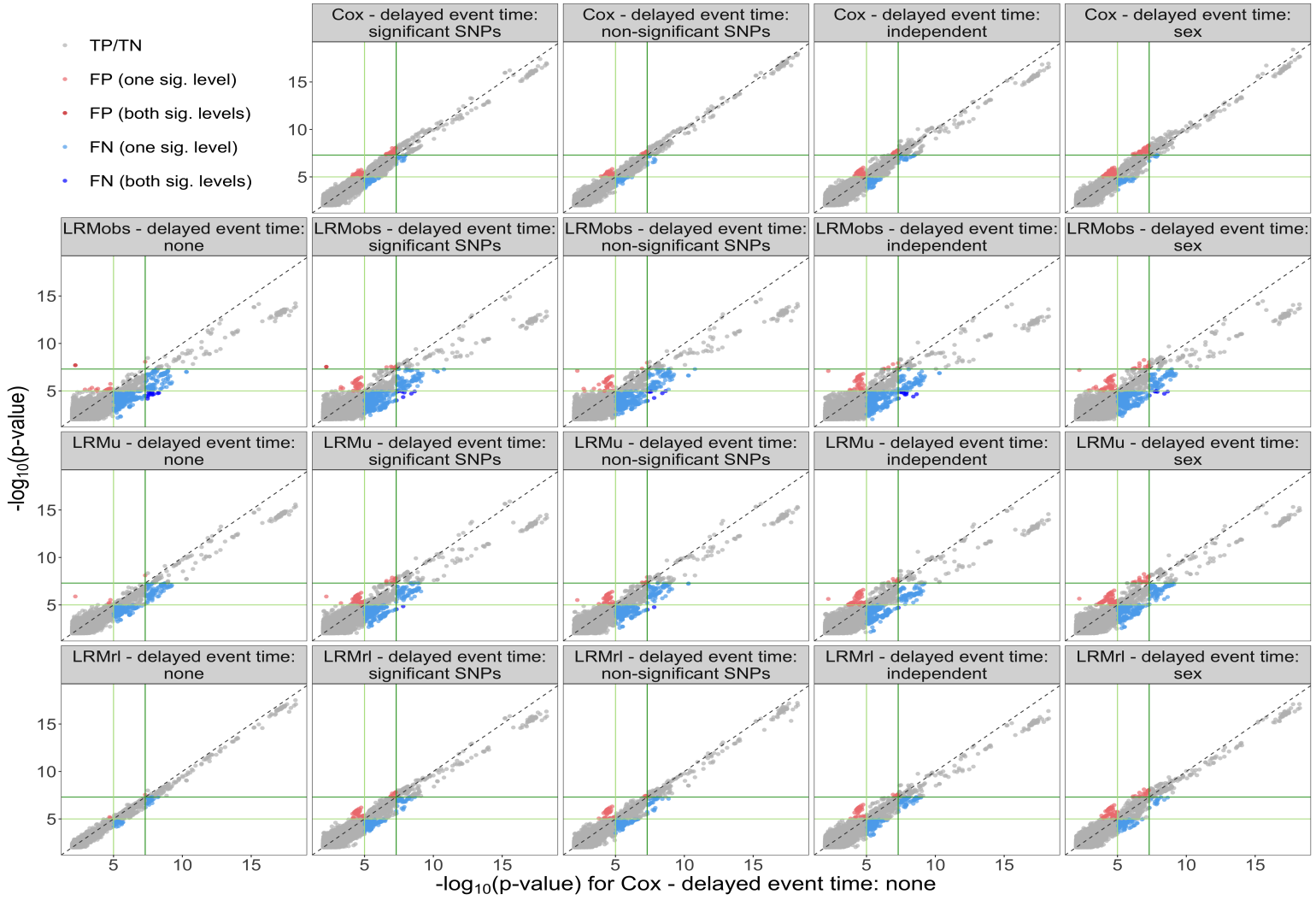


Figure 41: False positive and false negative SNPs for each model and delayed event time combination, using Model 1 (Cox) with no delayed event time as the gold standard, for hypothyroidism (phecode 244). Dark green lines correspond to $P \leq 5 \times 10^{-8}$ and light green lines correspond to $P \leq 1 \times 10^{-5}$.

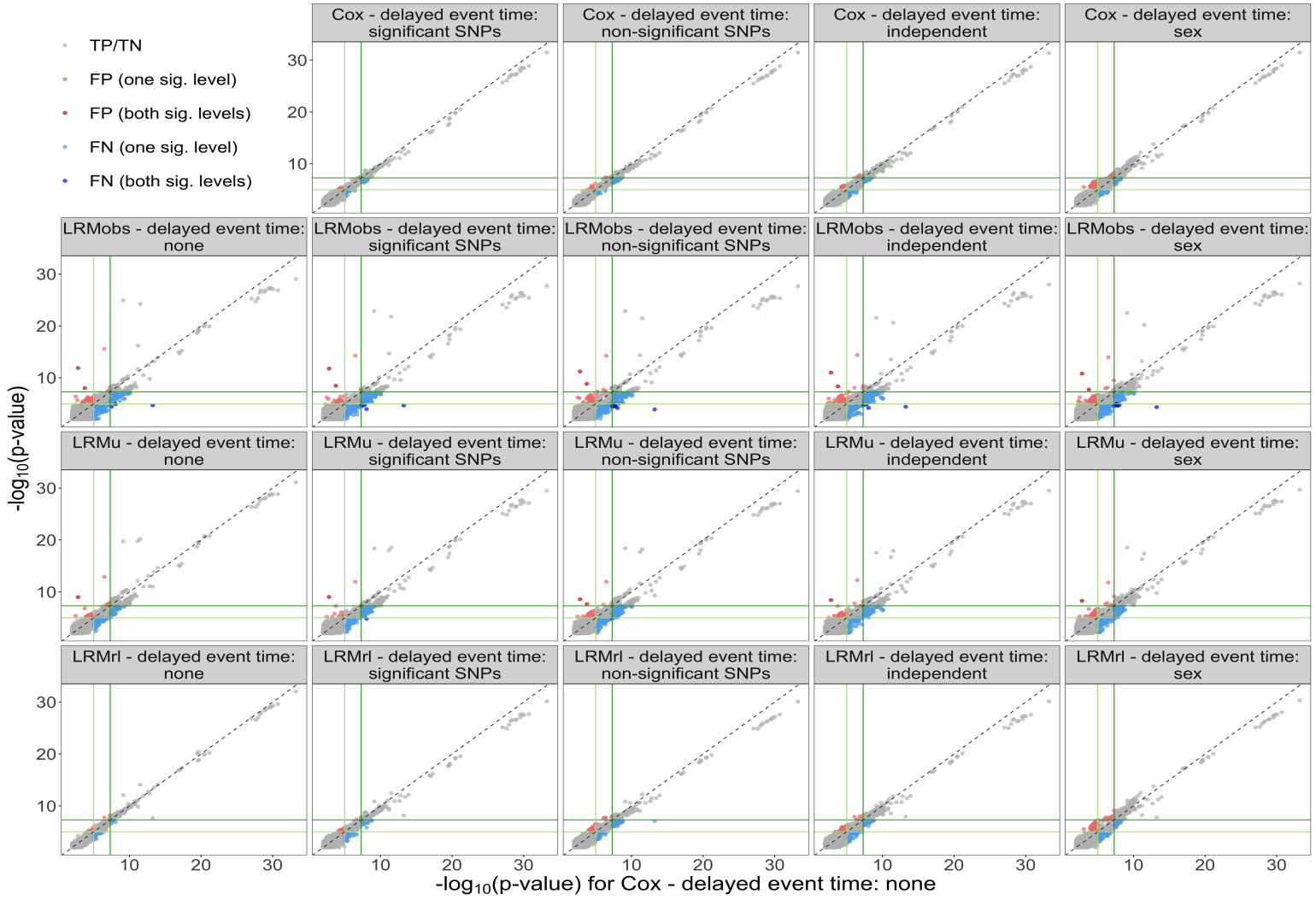


Figure 42: False positive and false negative SNPs for each model and delayed event time combination, using Model 1 (Cox) with no delayed event time as the gold standard, for type 2 diabetes (phecode 250.2). Dark green lines correspond to $P \leq 5 \times 10^{-8}$ and light green lines correspond to $P \leq 1 \times 10^{-5}$.

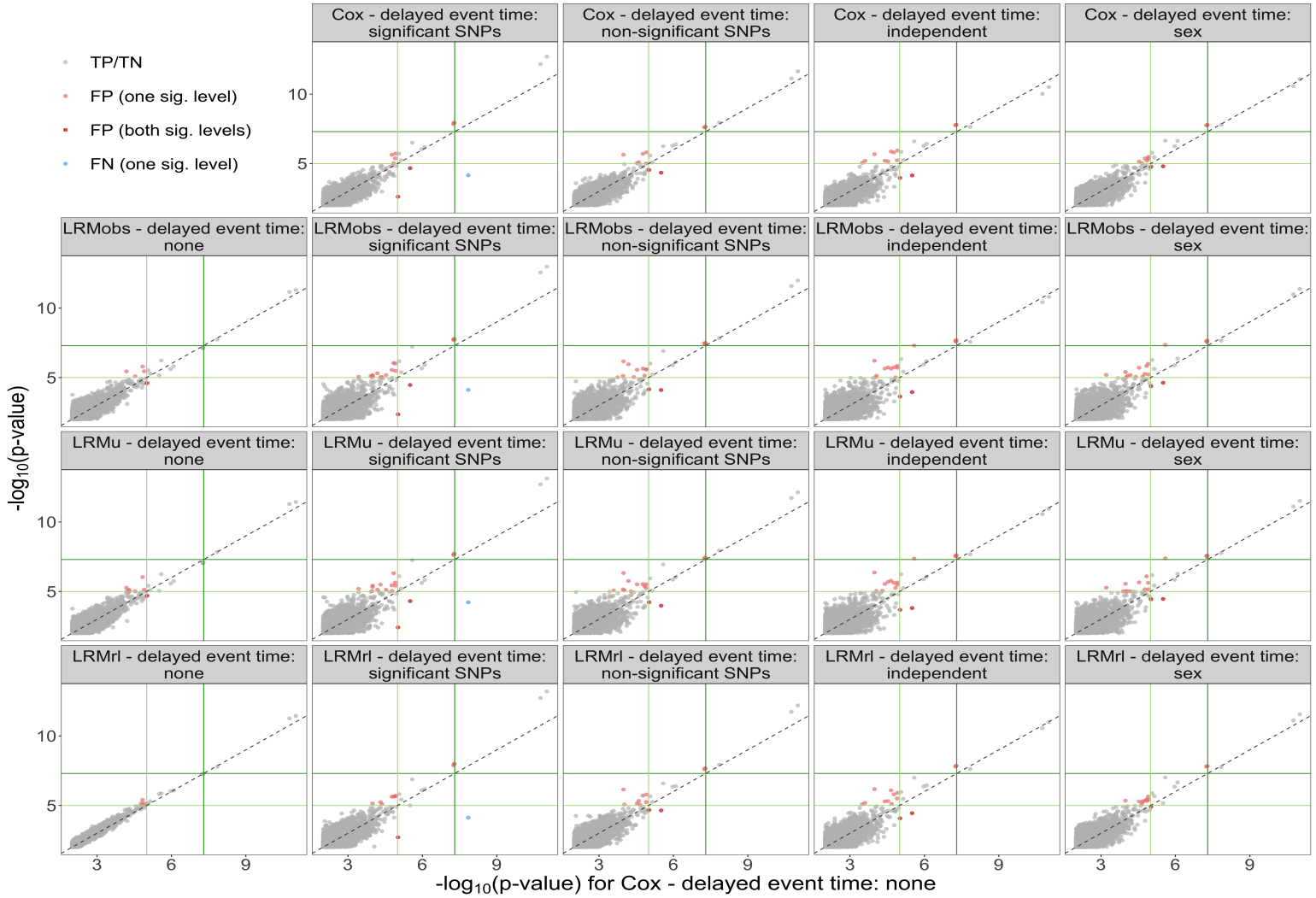


Figure 43: False negative SNPs for each model and delayed event time combination, using Model 1 (*Cox*) with no delayed event time as the gold standard, for vitamin D deficiency (phecode 261.4). Dark green lines correspond to $P \leq 5 \times 10^{-8}$ and light green lines correspond to $P \leq 1 \times 10^{-5}$.

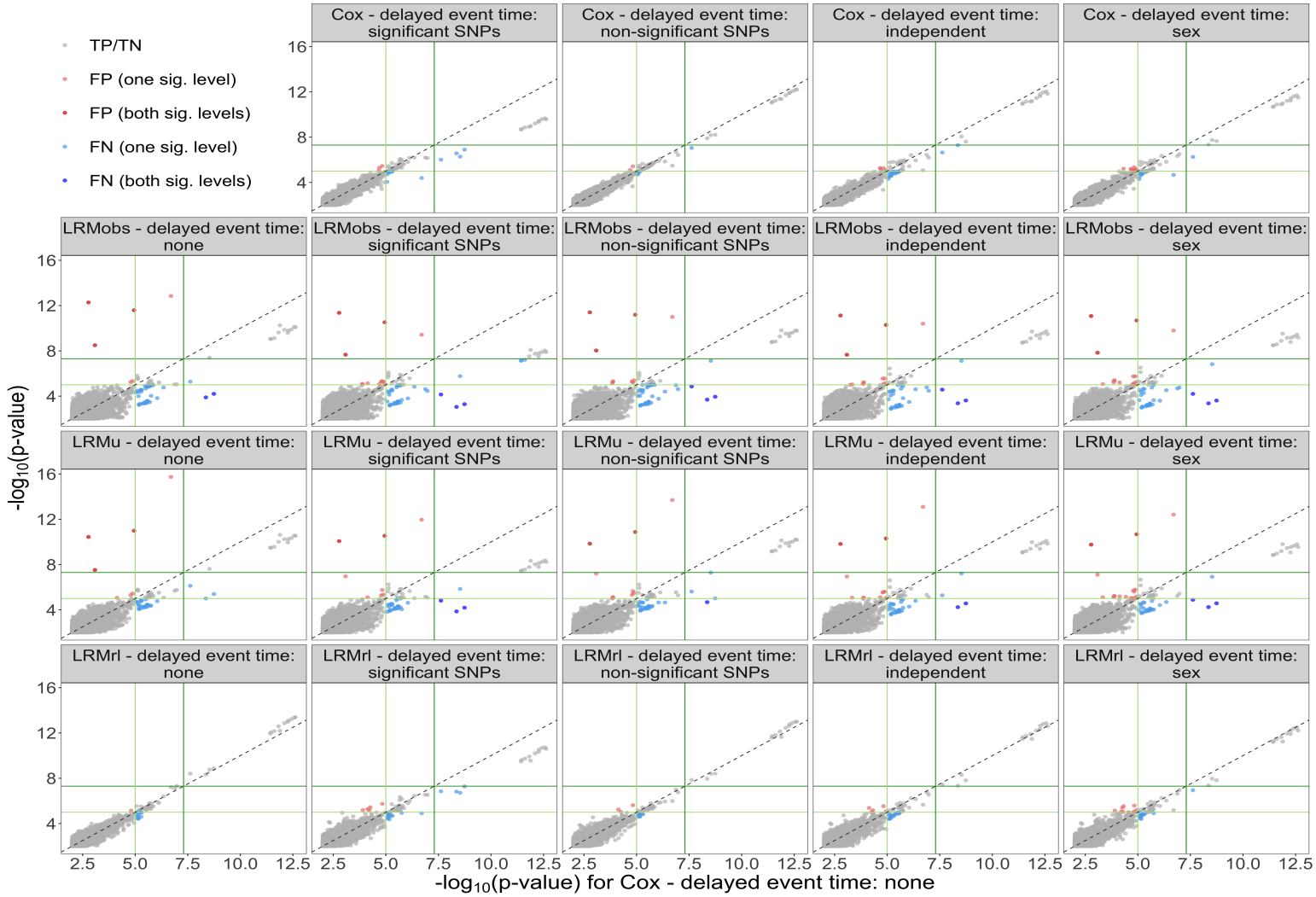


Figure 44: False negative SNPs for each model and delayed event time combination, using Model 1 (Cox) with no delayed event time as the gold standard, for hypercholesterolemia (phecode 272.11). Dark green lines correspond to $P \leq 5 \times 10^{-8}$ and light green lines correspond to $P \leq 1 \times 10^{-5}$.

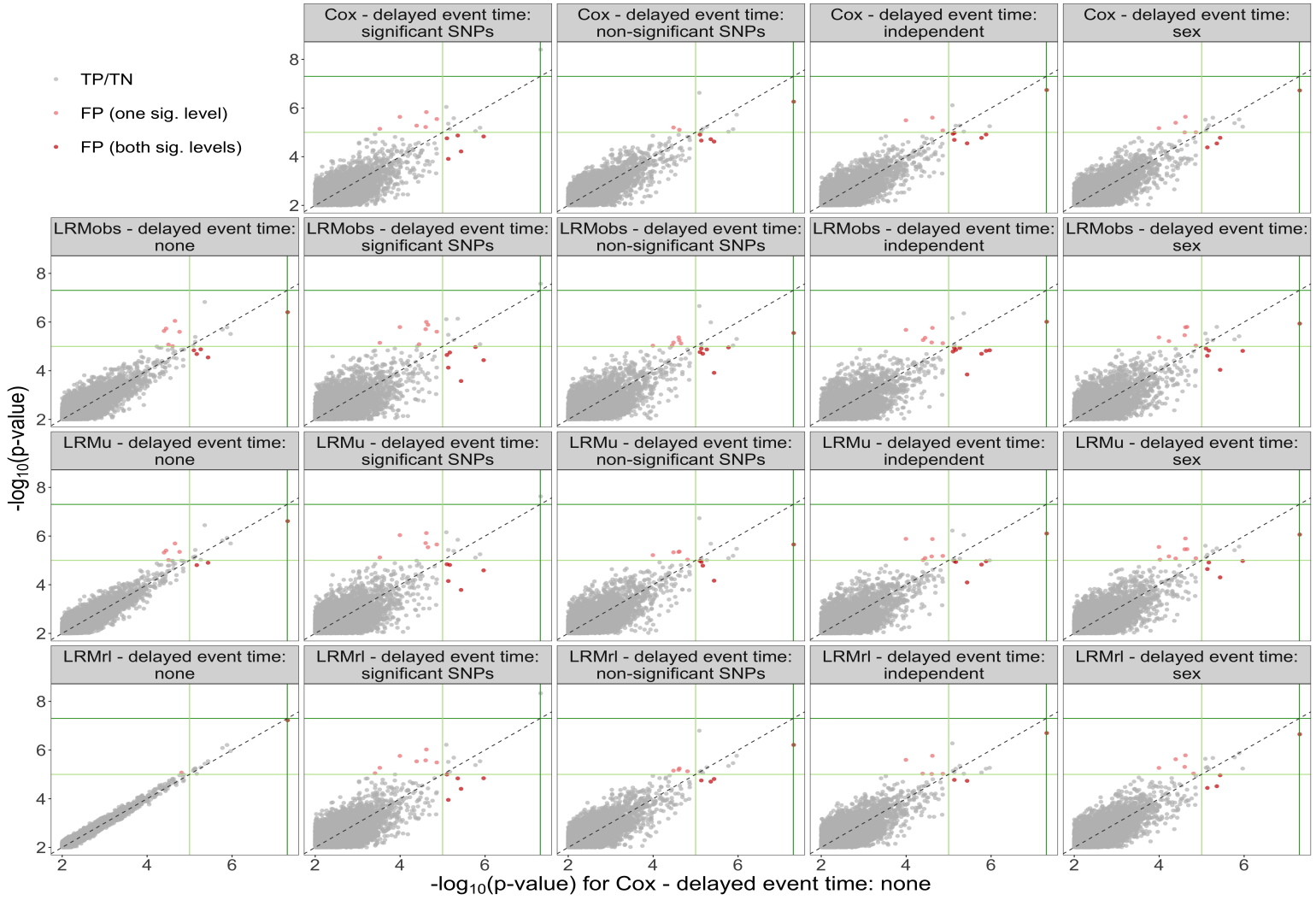


Figure 45: False positive and false negative SNPs for each model and delayed event time combination, using Model 1 (Cox) with no delayed event time as the gold standard, for insomnia (phecode 327.4). Dark green lines correspond to $P \leq 5 \times 10^{-8}$ and light green lines correspond to $P \leq 1 \times 10^{-5}$.

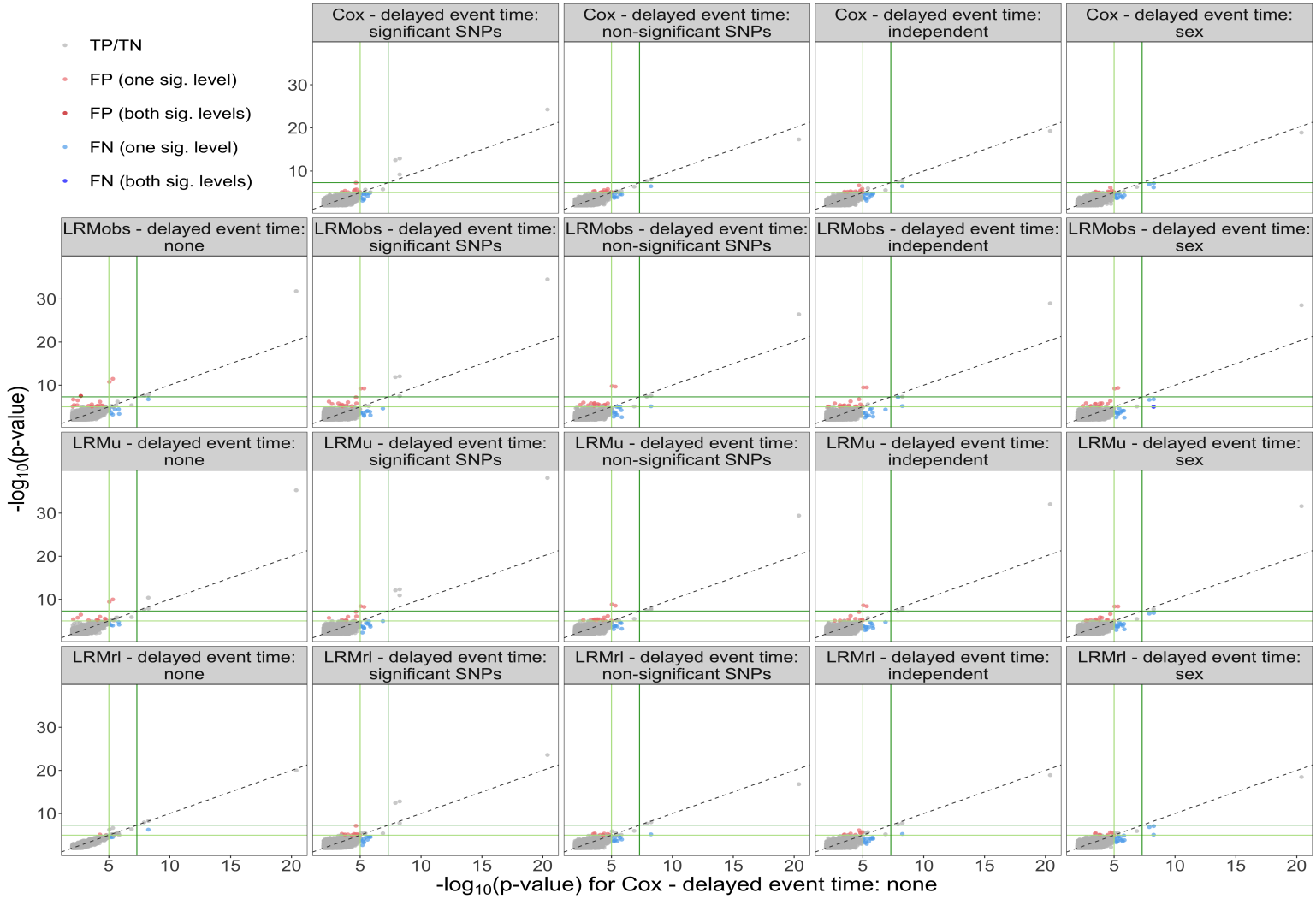


Figure 46: False positive and false negative SNPs for each model and delayed event time combination, using Model 1 (*Cox*) with no delayed event time as the gold standard, for myocardial infarction (phecode 411.2). Dark green lines correspond to $P \leq 5 \times 10^{-8}$ and light green lines correspond to $P \leq 1 \times 10^{-5}$.

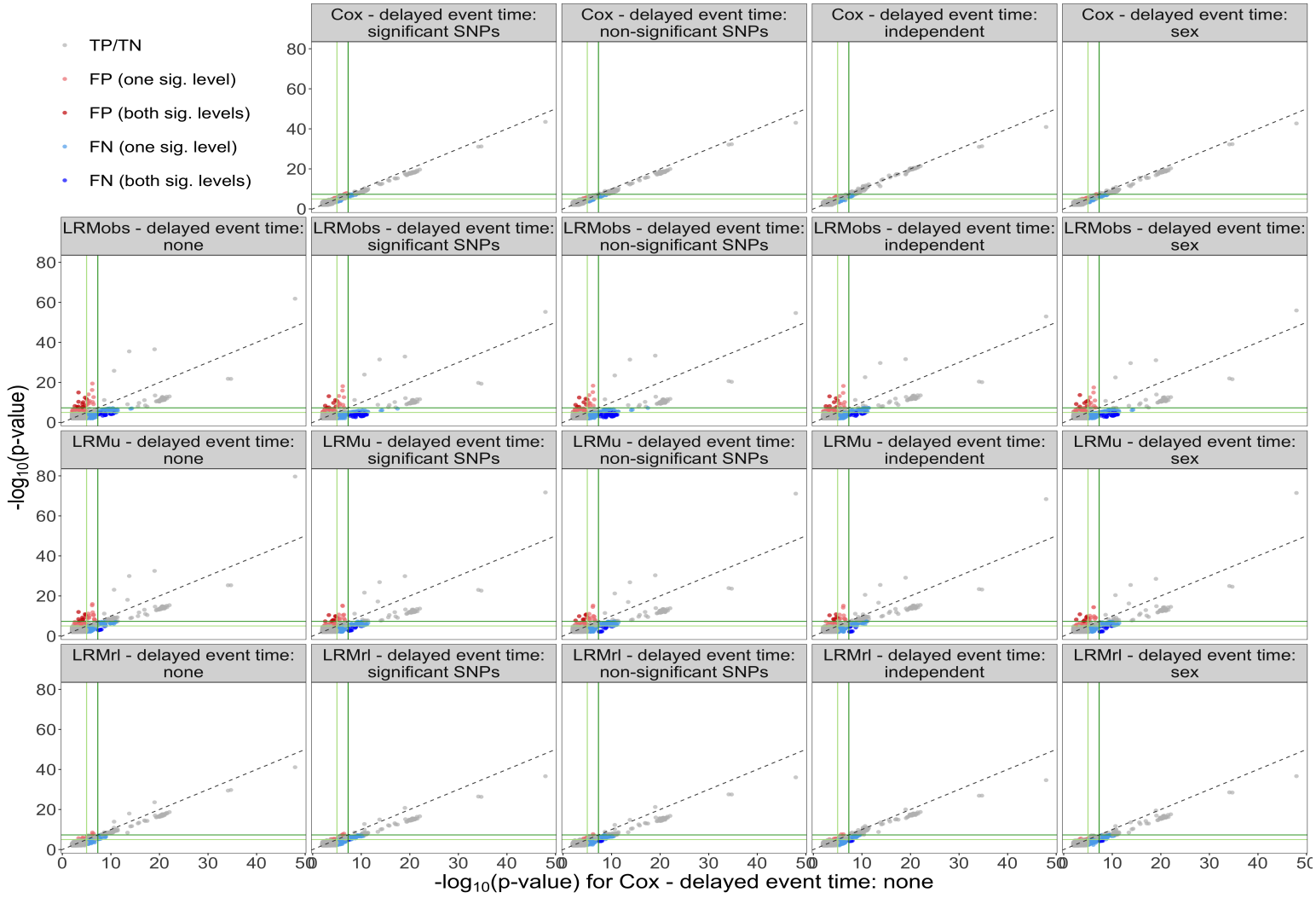


Figure 47: False positive and false negative SNPs for each model and delayed event time combination, using Model 1 (Cox) with no delayed event time as the gold standard, for coronary atherosclerosis (phecode 411.4). Dark green lines correspond to $P \leq 5 \times 10^{-8}$ and light green lines correspond to $P \leq 1 \times 10^{-5}$.

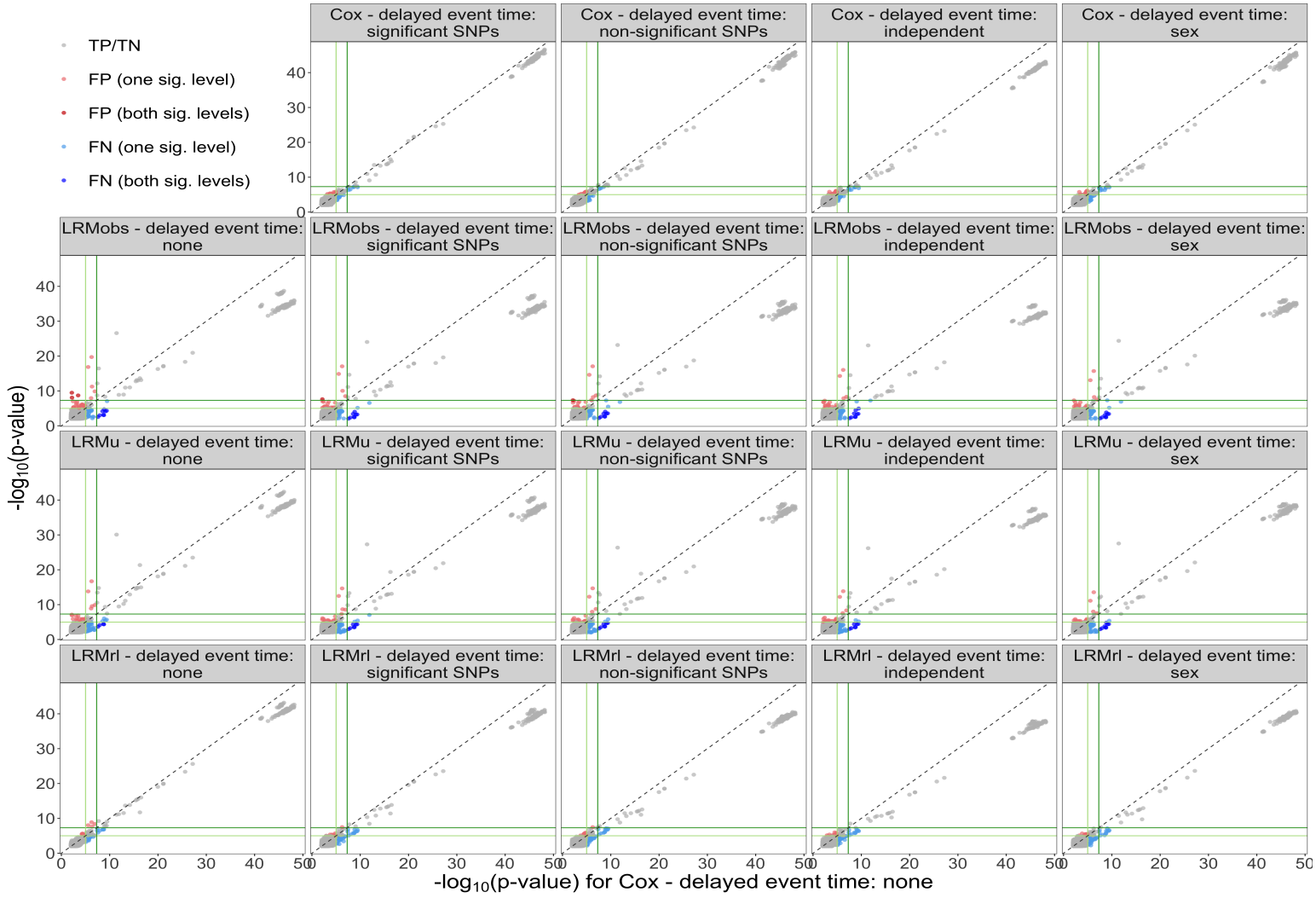


Figure 48: False positive and false negative SNPs for each model and delayed event time combination, using Model 1 (Cox) with no delayed event time as the gold standard, for atrial fibrillation (phecode 427.21). Dark green lines correspond to $P \leq 5 \times 10^{-8}$ and light green lines correspond to $P \leq 1 \times 10^{-5}$.