

Effect of Time Measurement Error  
on Assessing Treatments with Time Dependent Effect

Yanshu Guo

A thesis  
submitted in partial fulfillment of the  
requirements for the degree of

Master of Science

University of Washington

2017

Committee:

Susanne May

Lianne Sheppard

Program Authorized to Offer Degree:

Biostatistics

©Copyright 2017

Yanshu Guo

University of Washington

**Abstract**

Effect of Time Measurement Error  
on Assessing Treatments with Time Dependent Effect

Yanshu Guo

Chair of the Supervisory Committee:

Professor Susanne May

Biostatistics

In the emergency medicine setting, it is often difficult to accurately record the time of injury without measurement error. If the treatment effect varies depending on the time from injury to treatment initiation, this type of measurement error may affect our analysis of the treatment effect. A Phase III trial for Tranexamic acid (TXA) in trauma patients with significant hemorrhage, CRASH-2, suggested a significant association between the time from injury to treatment initiation and treatment effect in that the benefit of TXA treatment reduces as the time to treatment initiation gets longer. TXA is a drug used to prevent fibrinolysis and reduce surgical blood loss for patients with major trauma, with traumatic brain injury (TBI) or after surgery. In this work, I study the effect measurement error might have on a combined analysis of two ongoing clinical trials of TXA in TBI. One of these trials is a Phase II clinical trial conducted by the Resuscitation Outcomes Consortium (ROC) in the United States and Canada. It is a

double blind, randomized placebo controlled clinical trial evaluating the efficacy of two dosing regimens of TXA in patients with TBI in a pre-hospital setting. The trial plans to enroll about 1,000 patients. The second trial, CRASH-3, is a Phase III double blind, randomized, placebo controlled multi-national trial of TXA in patients with TBI. This trial plans to enroll about 13,000 patients.

For the purpose of this study, I define a time dependent effect as a treatment effect that varies as times from injury to treatment initiation vary. Any observed treatment effect may be smaller or larger than the true effect due to time measurement error and the time to treatment initiation may be underestimated or overestimated. For treatments with potential time dependent effect like TXA, I investigate in this study whether and how much their estimated treatment effect may be affected by measurement error using simulations. I use logistic regression to fit models of interest with and without measurement error, with and without interaction term and compare power, (for some model comparisons) bias and Type I error rates. I first investigated the effect of measurement error under the setting resembling the ROC TXA trial, and then under other more general settings. Finally, I also investigated the potential effect of measurement error for the dataset combining the simulated CRASH-3 data and simulated ROC TXA data using meta-analysis.

The results were consistent with our expectation that the measurement error could reduce power for detecting treatment and interaction effect and increase estimation bias for treatment effect at time zero and interaction term and this impact of measurement error was only associated with the strength of absolute measurement error. However, if our assumptions are appropriate, an average 0.5 hour absolute

measurement error in the ROC TXA trial does not meaningfully impact our analysis results and an average 1 hour absolute measurement error in the meta-analysis (combining simulated CRASH-3 and ROC TXA data) can still generate a power above 80% to detect both treatment and interaction effect.

## Table of Contents

1. Introduction .....	8
2. Methods .....	12
2.1 Study design, population, and data generating models .....	13
2.1.1 Simulation method of the times from injury to treatment initiation .....	13
2.1.2 Randomization of treatment and simulation method of outcomes .....	15
2.1.3 Simulation for different scenarios .....	17
2.1.3.1 Simulation for ROC settings .....	17
2.1.3.2 Simulation for other settings .....	18
2.1.3.3 Simulation for IPD Meta-Analysis .....	19
2.2 Statistical analysis on ROC and other scenarios .....	21
2.2.1 Comparison among different analysis models .....	21
2.2.2 Preliminary study on degrees of freedom and power .....	23
2.2.2.1 Simulation for preliminary study .....	24
2.2.2.2 Analysis for preliminary study .....	26
2.3 IPD Meta-Analysis on ROC and CRASH-3 .....	27
2.3.1 IPD Meta-Analysis for continuous initiation time .....	27
2.3.2 IPD Meta-Analysis for crude initiation time .....	29
3. Results.....	32
3.1 Relationship between degrees of freedom and power .....	32
3.2 Differences among four models in simulated CRASH-2 setting .....	36
3.3 Differences among four models in other scenarios.....	37
3.3.1 Effects of proportion of underestimation on power and bias .....	37
3.3.2 Effects of distribution of true initiation time from injury on power and bias...39	
3.3.3 Effects of distribution of time measurement error on power and bias .....	41

3.3.4 Effects of the strength of treatment effect at time zero on power and bias .....	45
3.3.5 Effects of the strength of time dependent effect on power and bias .....	47
3.4 Effects of time measurement error on power and bias in IPD Meta-Analysis .....	49
4. Discussion and conclusion.....	52
References.....	62
Appendix .....	65

## 1. Introduction

A properly designed clinical trial is a gold standard for assessing whether an intervention has the postulated effect. Given that most of the biological mechanisms and knowledge of diseases are not fully understood, conducting an appropriately controlled randomized clinical trial is the most definitive method for Phase III studies to compare a new intervention against the current best standard treatment.

Tranexamic acid (TXA) is a drug used to inhibit fibrinolysis and reduce surgical blood loss for patients with traumatic brain injury (TBI) by reducing bleeding and the need for blood transfusion.<sup>1</sup> Two Phase III trials for TXA conducted in many countries, CRASH-2<sup>2</sup> and CRASH-3<sup>3</sup>, were designed to show the potential treatment effect of TXA on patients with traumatic injury by reducing bleeding and the need for blood transfusion. CRASH-2 focused on patients with significant haemorrhage, whereas CRASH-3 focuses on patients with TBI. While CRASH-3 is still enrolling patients, the completed CRASH-2 suggested a significant association between the time from injury to treatment initiation and treatment effect on the risk of death due to bleeding (test for interaction  $p < 0.0001$ )<sup>4</sup>. Receiving TXA treatment within 1 hour from injury significantly reduced the risk of death due to bleeding in hospital within four weeks of injury compared with the placebo group (RR 0.68, 95% CI 0.57–0.82;  $p < 0.0001$ ). TXA treatment given between 1 and 3 hours also reduced the risk of death due to bleeding in hospital within four weeks of injury (RR 0.79, 95% CI 0.64–0.97;  $p = 0.03$ ), but receiving TXA treatment after 3 hours from injury significantly increased the risk of death due to bleeding in hospital within four weeks of injury (RR 1.44, 95% CI 1.12–1.84;  $p = 0.004$ ). One possible explanation for the increased risk is that patients in the late phase of trauma can develop thrombotic disseminated intravascular coagulation and anti-fibrinolytics could

be contraindicated in this period.<sup>5,6</sup> Although disseminated intravascular coagulation is characterized by fibrin formation and coagulation, the rapid consumption of coagulation proteins can lead to their exhaustion, resulting in uncontrolled bleeding. The need to avoid giving an anti-fibrinolytic in this late phase was why the ROC trial restricted trial inclusion to patients who were within 2 hours of injury. In addition, the CRASH-3 trial changed its enrollment criteria from 1~8 hours to 1~3 hours (time from injury to enrollment) in September 2016.<sup>3</sup>

I call such a change of treatment effect depending on time from injury to treatment initiation a time dependent effect. Different from the concept of time dependent effect in other contexts like survival analysis, time dependent effect in this thesis only refers to the variation of treatment effect based on the times from injury to treatment initiation. The time dependent effect of TXA observed in CRASH-2 was an important aspect considered in the design of a TXA trial conducted by the Resuscitation Outcomes Consortium (ROC), a Phase II clinical trial for TXA in the United States and Canada. It was the reason for limiting enrollment to patients whose treatment could be initiated within 2 hours of estimated time of injury where treatment is initiated by emergency medicine providers in the field. The ROC study is an ongoing Phase II clinical trial for traumatic brain injury. To determine the efficacy of two dosing regimens of TXA initiated in the pre-hospital setting, the three arms of this study receive 1 gram Tranexamic Acid (TXA) prior to hospital arrival followed by a 1 gram TXA infusion over 8 hours after hospital arrival, 2 grams TXA prior to hospital arrival and 0.9% Sodium Chloride over 8 hours and 0.9% Sodium Chloride prior to hospital arrival and 0.9% Sodium Chloride over 8 hours injectable respectively. The primary endpoint is the dichotomized Glasgow Outcome Scale Extended (GOS-E) score at 6 months post injury.<sup>7</sup>

The GOS-E score is a global scale with 8 permissible grades that rates patients' survival status.<sup>8</sup> In the ROC TXA trial patients are assigned to categories of severe and moderate disability or good recovery based on phone interviews. Patients with GOS-E > 4 are considered as having favorable neurologic outcomes indicating good recovery.

A challenge with the CRASH-2, CRASH-3 and ROC TXA studies and the ultimate real-world application of this treatment (TXA) is that in many scenarios the estimated time from injury to treatment initiation may be measured with error. For example, witnesses or patients themselves may overestimate or underestimate the time when they experience or witness traffic accidents or other events that result in traumatic injury. Therefore, if there is time measurement error, it is important to investigate whether and how the analysis of time dependent treatment effects will be affected.

I anticipate that the observed times from injury to treatment initiation for patients with traumatic brain injury are a mixture of underestimation and overestimation and that the ratio of underestimation to overestimation varies in different scenarios. If TXA has a time dependent effect as observed in CRASH-2 and most of the times from injury to treatment initiation are underestimated, the observed treatment effect of TXA may be smaller than the true effect due to the combination of the measurement error and the existence of the time dependent effect. A beneficial treatment effect could also be missed if the observation time period includes both time periods with benefit as well as the time periods with harmful effects, because these effects could cancel each other out over time. This suggests that the observed treatment effect may also be influenced to different extents when features such as the sample size, the ratio of underestimation to

overestimation, the treatment effect at initiating time or other characteristics of this setting change. The main goal of this thesis is to investigate how the power of detecting a time dependent treatment effect of TXA will be influenced when I consider that time from injury to treatment initiation is measured with error. If there is a change in power of detecting the treatment effect or bias of the effect estimate, this thesis will also investigate the direction of change in different scenarios.

There are many previous studies conducted on measurement error<sup>9, 10</sup>, but the investigation of models including effect modification is limited. In practice, there are many treatments applied in an emergency and the true time from injury to treatment initiation typically cannot be recorded exactly. If the treatments have the time dependent effect as I defined, it might be critical to consider the uncertainty of measured time. This thesis will explore the effect of measurement error on estimating time dependent treatment effect in clinical trials with a focus on the ROC TXA and CRASH-3 trials.

## 2. Methods

Owing to the fact that the error of the time from injury to treatment initiation is typically immeasurable in real life, I generated the data by simulating true time from injury to treatment initiation using a truncated normal distribution<sup>11, 12</sup> and by simulating measurement error using a different truncated normal distribution similar to what is anticipated for the ROC TXA study. Then I simulated the treatment randomization and determined a zero/one outcome for each simulated record using a logistic model that included a time dependent effect. The outcome could represent neurologically intact survival at six months (as is the primary outcome for the ROC TXA study, in which the indicator for Glasgow Outcome Score Extended (GOS-E) score is less than or equal to 4) or death in hospital within 28 days (as is the primary outcome for the CRASH-3 trial). Based on this simulated dataset, I conducted analysis of different fitted models that included the true time from injury to treatment initiation or the time with measurement error in order to investigate how the power and estimation of these models would be affected by the measurement error. I assumed that there is no time dependent effect for the placebo group. For different fitted model scenarios, I estimated the power to detect the treatment effect, the power to detect the time dependent effect, the Type I error rate of falsely identifying a time effect for the placebo group and the bias of selected estimated coefficients. Finally, I simulated a ROC TXA and a CRASH-3 dataset using a similar strategy and performed a one-stage individual patient data (IPD) meta-analysis on the pooled dataset. All the datasets in this study were simulated 1,000 times. Further details regarding the simulated data and the fitted models are provided below.

## ***2.1 Study design, population, and data generating models***

The distributions and parameters for the simulation were chosen based on the information that was available for the ROC TXA (Appendix A), CRASH-3<sup>13</sup> and CRASH-2<sup>4</sup> trials. The parameters were then varied to investigate the effect of measurement error in different scenarios as I summarize below in Table 1.

### *2.1.1 Simulation method of the times from injury to treatment initiation*

I simulated data according to the background and characteristics of the CRASH-2, CRASH-3 and TXA studies. The distribution of the simulated times from injury to treatment was motivated by summary statistics reported for CRASH-2.<sup>4</sup> I will call it “true time” to distinguish it from the estimated time with measurement error. Even though the original true time from injury to treatment initiation at the individual level are impossible to ascertain, a dataset containing all times from 911 call to treatment initiation was available for all the subjects in ROC TXA trial and could provide some information about the shape of the distribution (Appendix A). The observed times from 911 call to treatment initiation were distributed asymmetrically with a peak at 0.5 hour. I suspect that in most cities in the United States and Canada, witnesses of the traumatic event and/or emergency medical providers are more likely to underestimate this time period. I furthermore assumed that an ambulance typically needs at least 15 minutes to be informed of and get to the site of injury and start the treatment. I chose to limit the maximum time from injury to treatment initiation to 12 hours. As a result, I used truncated normal distributions<sup>11,12</sup> to simulate the (true) time from injury to treatment initiation to be between 15 minutes and 12 hours. Even though the truncated normal distribution is symmetric around its (original) mean, asymmetric truncation and unequal probabilities of underestimation and overestimation could make the

distribution of observed time from injury to treatment initiation asymmetric. We obtained data from the ROC TXA trial that represented times from 911 call to treatment initiation. The distribution of these data was skewed to the right and it seemed reasonable that it could arise from a log normal distribution (see Appendix Figure 1). I nevertheless used truncated normal distributions for most of the simulated data settings and performed sensitivity analysis (presented in Appendix A and B) to investigate the potential consequences of this choice. The true time was distributed from 15 minutes to 12 hours, but after adding or subtracting the measurement error, only the time within the enrollment time frame 0~2 hours was selected as observed time. To make sure enough observed times within enrollment time frame could be generated, I used R function `rtruncnorm` to simulate a sample of  $1.5n$  subjects within enrollment time frame as the times from injury to treatment initiation, where  $n$  is the total number of subjects to be randomized. The truncated normal distribution has same shape as normal distribution but its domain is restricted to a certain range of values. When I simulated the distribution of true time from injury to treatment initiation, I also considered the time from injury to 911 call and the underestimation error which were both unobservable in real ROC TXA data. Therefore, the distribution of simulated true time from injury to treatment initiation had larger mean and standard deviation compared with the distribution of times from 911 call to treatment initiation in the ROC TXA trial (Appendix A),

Then I simulated the same number of time measurement errors using a different truncated normal distribution. These represent the difference between the true time from injury to treatment initiation time and the estimated (observed) time from injury to treatment initiation. This time measurement error could either be positive

corresponding to underestimation or negative corresponding to overestimation. I simulated positive absolute errors and then determined the sign of each measurement error with a desired percent  $p$  of over and underestimation (Appendix A Figure 1B showed 90% of underestimation). The simulated positive absolute errors were randomly assigned as negatives with the percent  $p$  and others were kept as positive with the percent  $(1-p)$ . The parameters of both distributions (of true time from injury to treatment initiation and of the absolute measurement error) were varied to investigate the impact such changes had on the bias and power to detect treatment effect (see Section 2.2.3 and Table 2).

To achieve varying degrees of underestimation and overestimation, for all the records, I added the (mixture of positive and negative) measurement error from the time from injury to obtain an estimated time distribution. The percent of underestimation by subtracting the measurement error was  $p$ . Then, a sample of size  $n$  within a prespecified maximum enrollment time frame (based on the estimated time from injury to treatment initiation) was used to represent the study population. The study population was selected only based on the estimated time and regardless of the true time value. A subject having estimated time within the enrollment time frame could be selected even if its true time is larger than the upper bound of the enrollment time frame. I used the enrollment time frame of 0 ~ 2 hours for ROC study, and used a combination of 0~3 hours and 0~8 hours for CRASH-3 in later IPD Meta-Analysis.

### *2.1.2 Randomization of treatment and simulation method of outcomes*

In the ongoing ROC TXA trial, there are two treatment arms with different dosing regimens: one loads 2 g TXA prior to hospital arrival (and placebo in the hospital) and

the other one loads 1 g TXA prior to hospital arrival followed by infusion of 1 g over 8 h after hospital arrival, but in this study on measurement error, I combined these two treatment arms and compared them to the placebo arm. Therefore, I randomized a total of 1,000 subjects with a 2:1 ratio to either treatment arm or placebo arm. I used the R function `rbinom` for generating random number under binomial (1, probability=2/3) to generate the treatment assignment for each subject.

The primary outcome is simulated as a zero/one indicator variable. Using the simulated treatment arm (`trt`), the simulated true time (`time.true`) and the parameters ( $\beta_0$ ,  $\beta_1$ ,  $\beta_2$  and  $\gamma$ ) motivated by CRASH-2 results, I calculated the log odds of probability  $q$  of having an unfavorable outcome according to equation 1 provided below. In this equation,  $\beta_0$  is the log odds of unfavorable outcome for the placebo group,  $\beta_1$  is the log odds ratio of unfavorable recovery comparing treatment group to placebo group at time zero and  $\beta_2$  is the log odds ratio of unfavorable recovery comparing patients in placebo group with time differing by 1 hour in the placebo group. The coefficient for the interaction term between treatment and the times from injury to treatment initiation,  $\gamma$  is how the log odds ratio comparing treatment group to placebo group changes by every unit of time (hour). I assume that  $\beta_2$  is 0, which means the log odds of placebo group is the same across all times from injury to treatment initiation. Of note, the interpretation of  $\beta_1$  is a theoretical construct, because it would represent a case where the treatment was initiated at exactly the same time when the patients got injured, which in real emergency medical settings is impossible.

$$\text{logit}(q) = \beta_0 + \beta_1 \text{trt} + \beta_2 \text{time.true} + \gamma \text{trt} \times \text{time.true} \quad \text{equation 1}$$

$\text{trt} = 1$ , treatment  
 $0$ , placebo

$\text{time.true}$  = true time from injury to treatment initiation time for both groups

Then I assigned a value for the outcome of 1 or 0 for each subject based on a binomial distribution with a probability  $q$  calculated according to equation 1 (above).

Because I assume that the log odds of the outcome for the placebo group is constant over all values of time from injury to treatment initiation,  $\beta_2$  was set to be 0. The values of  $\beta_0$ ,  $\beta_1$  and  $\gamma$  were motivated by results from previous studies and then varied in different scenarios as I describe below in Section 2.1.3 and Table 1.

### *2.1.3 Simulation for different scenarios*

In this thesis, I first investigated the effect of time measurement error on settings similar to the ROC TXA trial. Then I investigate the effect of measurement error under more general settings. Finally, I investigated the effect of measurement error on the dataset that combined simulated ROC TXA data and simulated CRASH-3 data using IPD meta-analysis.

#### *2.1.3.1 Simulation for settings similar to the ROC TXA trial*

First, I analyzed the effects of time measurement error in a simulated scenarios resembling the ROC TXA trial by fitting four logistic regression models. These are described below as Models 1-4. Based on the mean and standard deviation of the times from injury to treatment initiation in CRASH-2, I assumed a truncated normal distribution with a mean of 1.23 hours and a standard deviation of 0.514 hour (After truncation). I also assumed that time measurement error was truncated normally distributed from 0 to 0.5 hour with a mean of 0.163 hour and a standard deviation of 0.105 hour, the proportion of underestimated was 90%, the probability of outcome for placebo group was 0.7 (log odds  $\beta_0 = \ln\left(\frac{p_{placebo}}{1-p_{placebo}}\right) = 0.85$ ), the probability of outcome

for treatment group at time 0 was 0.56 (log odds ratio  $\beta_1 =$

$\ln\left(\frac{p_{treatment}}{1-p_{treatment}}\right) - \ln\left(\frac{p_{placebo}}{1-p_{placebo}}\right) = -0.6$ ) and the treatment switched to harmful at 3h

(interaction  $\gamma = -\frac{\beta_1}{crossing\ point(h)} = 0.2$ ) after injury. These parameters were the basis I

assumed for ROC TXA trial (Table1 Column 1).

### *2.1.3.2 Simulation for other settings*

Based on the basic setting similar to ROC TXA trial (Table1 Column 1), I varied the setting regarding a certain aspect while keep the other aspects same as the basic setting (Table 1 Column 2, 3, 4, 5, 6). There were five sets of scenarios with varied: (1) ratio of underestimation to overestimation; (2) distribution of the true initiation times from injury; (3) distribution of time measurement error; (4) magnitude of treatment effect at time zero; (5) magnitude of interaction between time from injury to treatment initiation and treatment effect.

**Table 1: How simulation parameters changed to investigate the effect of measurement error in different scenarios**

	1	2	3	4	5	6	7
Results	3.2 Table3	3.3.1 Figure2	3.3.2 Figure3	3.3.3 Figure4a	3.3.3 Figure4b	3.3.4 Figure5	3.3.5 Figure6
Maintain 80% power	Treatment	Treatment	Treatment	Treatment	Interaction	Treatment	Treatment
Sample size	1000	1400, 1400, 1400, 1400, 1400	900, 1300, 1400, 1400, 1400	1400, 1400, 1400, 1400, 1400	1700, 1600, 1500, 1400, 1300	15000, 3800, 2000, 1200, 900	2900, 2100, 1400, 1400, 1400
Log odds for placebo group; $\beta_0$ (eq1)	0.85	0.85	0.85	0.85	0.85	0.85	0.85
Time effect (log odds ratio) every hour; $\beta_2$ (eq1)	0	0	0	0	0	0	0
Proportion of Underestimation	90%	10%, 30%, 50%, 70%, 90%	90%	90%	90%	90%	90%
True time from injury to treatment initiation distribution; range=0.25~12; (mean; SD of original normal)	1.5; 1.5	1.5; 1.5	(0.5; 0.3) (1; 1) (1.5; 1.5) (2; 1.5) (2.5; 1.5)	1.5; 1.5	1.5; 1.5	1.5; 1.5	1.5; 1.5
True time from injury to treatment initiation distribution; range=0.25~12; (mean; SD after truncation)	1.23; 0.514	1.23; 0.514	(0.525; 0.238) (1.14; 0.509) (1.23; 0.514) (1.28; 0.524) (1.37; 0.521)	1.23; 0.514	1.23; 0.514	1.23; 0.514	1.23; 0.514
Absolute time measurement error distribution (range; mean; SD of original normal)	0~0.5; 0.1; 0.15	0~0.5; 0.1; 0.15	0~0.5; 0.1; 0.15	(0~0.5; 0.1; 0.15) (0~0.7; 0.2; 0.2) (0~1; 0.3; 0.2) (0~1.3; 0.4; 0.3) (0~1.5; 0.5; 0.4)	(0~0.5; 0.1; 0.15) (0~0.7; 0.2; 0.2) (0~1; 0.3; 0.2) (0~1.3; 0.4; 0.3) (0~1.5; 0.5; 0.4)	0~0.5; 0.1; 0.15	0~0.5; 0.1; 0.15
Absolute time measurement error truncated distribution (mean; SD after truncation)	0.163; 0.105	0.163; 0.105	0.163; 0.105	(0.163; 0.105) (0.257; 0.154) (0.331; 0.174) (0.450; 0.252) (0.564; 0.318)	(0.163; 0.105) (0.257; 0.154) (0.331; 0.174) (0.450; 0.252) (0.564; 0.318)	0.163; 0.105	0.163; 0.105
Treatment effect (log odds ratio) at time 0; $\beta_1$ (eq1)	-0.6	-0.6	-0.6	-0.6	-1.2	-0.15, -0.3, -0.45, -0.6, -0.75	-0.6
Interaction effect between time and treatment; $\gamma$ (eq1)	0.2	0.2	0.2	0.2	0.6	0.05, 0.1, 0.15, 0.2, 0.25	0.6, 0.3, 0.2, 0.15, 0.12
crossing time point	3h	3h	3h	3h	2h	3h	1h, 2h, 3h, 4h, 5h

*2.1.3.3 Simulation for IPD Meta-Analysis*

Finally, for the data of individual patient data (IPD) meta-analysis, I assumed a truncated normal distribution with a mean of 1.5 hours and a standard deviation of 1.5 hours before truncation for true time from injury to treatment initiation, the proportion of underestimation of such times was 90%, the probability of unfavorable neurologic outcome for placebo group was 0.7 (log odds  $\beta_0 = 0.85$ ), the probability of outcome for treatment group was 0.56 (log odds ratio  $\beta_1 = -0.6$ ) and the treatment switched to harmful at 3h (interaction  $\gamma = 0.2$ ) after injury. I investigated the effect of measurement error when I varied the original normal distribution of measurement error with (range; mean; sd): (0~0.5; 0.1; 0.15), (0~1; 0.3; 0.2), (0~1.3; 0.4; 0.3), (0~1.5; 0.5; 0.4), (0~2; 1; 0.5); The (mean; sd) after truncation were: (0.163, 0.106), (0.323, 0.175), (0.440, 0.249), (0.557, 0.314), (0.936, 0.426).

I used these parameters to simulate the datasets for both CRASH-3 with a sample size of 10,000 and ROC TXA with a sample size of 1,000. However, since CRASH-3 is a two-arm trial where subjects are randomized with a 1:1 ratio to either treatment arm or placebo arm, a binomial distribution (1, probability=1/2) was used generate treatment assignment for each subject. Another critical difference between the CRASH-3 and ROC TXA trial is the time frame of enrollment. The CRASH-3 trial started in September 2011 and restricted trial enrolment to patients who were within 8 hours of injury. However, after the CRASH-2 trial suggested a potential time dependent effect of TXA, CRASH-3 changed its protocol to limit enrolment to patients who were within 3 hours of injury starting in September 2016<sup>14</sup>. According to the protocol of CRASH-3, approximately total of 10,000 patients will be recruited and randomized to maintain 90% power and two sided alpha of 0.1 and detect a 15% relative reduction (from 20% to 17%) in all-cause mortality<sup>15</sup>.

Based on that information, the times from injury to treatment initiation and measurement error of 10,000 patients were simulated. By the end of July 2017, 9,800 patients had been randomized in this trial<sup>16</sup>. Assuming a constant enrollment rate, I simulated 7,143 patients receiving the TXA treatment within 8 hours after injury and 2,857 patients receiving the TXA treatment within 3h after injury. All other factors, including the distribution of the times from injury to treatment initiation, the distribution of measurement error, the underestimation proportion, the coefficients of main effects and interactions, were simulated identically to the ROC TXA dataset. These values were simulated based on previous CRASH-2 trial, to mimic the real situation. Finally, the dataset used for meta-analysis consisted of the simulated CRASH-3 data and the simulated ROC TXA data.

## ***2.2 Statistical analysis on ROC and other scenarios***

Using the simulated datasets resembling the ROC TXA trial, I examined the potential effects of measurement error on the (1) power to detect treatment effect and the interaction between treatment and initiation time; (2) bias of the estimated treatment effect, time dependent effect and their interaction, under the assumption of a true time dependent effect of TXA.

### ***2.2.1 Comparison among different analysis models***

As described above, I simulated the true time from injury to treatment initiation, estimated times, treatment assignment and outcome for 1,000 subjects. Then I fit 4 logistic regression models with the treatment and the time from injury to treatment initiation of all subjects to investigate how treatment time measurement error affects

bias and power of detecting treatment effect in different scenarios when different models were used.

Model 1: The correct model fit with true time from injury to treatment.

This analysis model is identical to the model I used for data generation. I estimated the main effect of treatment ( $\beta_1$ ), the main effect of the time from injury to treatment initiation ( $\beta_2$ ) and the interaction effect ( $\gamma$ ) by fitting the logistic regression with treatment and the true time from injury to treatment initiation and their interaction.

$$\text{logit}(q) = \beta_0 + \beta_1 \text{trt} + \beta_2 \text{time.true} + \gamma \text{trt} \times \text{time.true} \quad \text{Model 1 (M1)}$$

Model 2: The correct model fit with the estimated time which includes measurement error.

I still fit the correct logistic regression model in that it is identical to the model I used to simulate the datasets, but with estimated time from injury to treatment initiation that underestimates or overestimates the true time. Then I estimated the main effect of treatment ( $\beta_1$ ), the main effect of time from injury to treatment initiation ( $\beta_2$ ) and the interaction effect ( $\gamma$ ), examining whether or how much the estimates would change from Model 1 by considering the measurement error for time since injury.

$$\text{logit}(q) = \beta_0 + \beta_1 \text{trt} + \beta_2 \text{time.obs} + \gamma \text{trt} \times \text{time.obs} \quad \text{Model 2 (M2)}$$

Model 3: A model with estimated time but without treatment time interaction.

This model estimated the main effect of treatment ( $\beta_1$ ) and the main effect of time from injury to treatment initiation ( $\beta_2$ ), without the assumption that the treatment effect was modified by the initiation time.

$$\text{logit}(q) = \beta_0 + \beta_1 \text{trt} + \beta_2 \text{time.obs} \quad \text{Model 3 (M3)}$$

Model 4: A model only considering for treatment effect.

Since most clinical trials consider the models only include main treatment effect term, Model 4 was also studied in this thesis to be compared with the models included time and interaction term. This model estimated only the main effect of treatment ( $\beta_1$ )

$$\text{logit}(q) = \beta_0 + \beta_1 \text{trt} \quad \text{Model 4 (M4)}$$

To evaluate and compare these four models, I first calculated the power to detect treatment effect with likelihood ratio tests comparing four models and their corresponding reduced models. The reduced model for M1 was fit with true initiation time from injury:  $\text{logit}(q) = \beta_0 + \beta_2 \text{time.true}$ . The reduced models for M2 and M3 were fit by including only the estimated initiation time:  $\text{logit}(q) = \beta_0 + \beta_2 \text{time.obs}$ . The reduced model for M4 was fitted by only including the intercept:  $\text{logit}(q) = \beta_0$ . The likelihood ratio tests for M1 and M2 were two degrees of freedom tests for whether both, the main treatment effect  $\beta_1$  and interaction term  $\gamma$ , equal to zero at significance level of 0.05. The likelihood ratio tests for M3 and M4 were one degree of freedom tests for only the main treatment effect  $\beta_1$  equal to zero at significance level of 0.05.

Because I assumed for the simulated data that the log odds of outcomes for placebo group does not change over time (from injury to treatment initiation), I performed Wald tests for M1, M2 and M3 to calculate the Type I error rate at significance level of 0.05. M4 did not include a time effect, so a Type I error rate for  $\beta_2$  was not applicable.

Given a simulated time dependent effect of TXA, I performed Wald test for the interaction between treatment and initiation time  $\gamma$  to calculate the power of detecting

time dependent effect at significance level of 0.05 in M1 and M2. M3 and M4 did not include an interaction term, so estimates of power to detect an interaction term do not apply to M3 or M4. To compare the estimation bias, I calculated the difference between the estimated coefficients and true coefficients used for simulation for M1 and M2.

### *2.2.2 Preliminary study on degrees of freedom and power*

During my exploratory study for the effect of measurement error in the simulations that reflected the ROC TXA scenario, I found that the powers of M3 and M4 were higher than that of M1, which was contrary to my original hypothesis that the correct model without measurement error should have largest power to detect the treatment effect. Then I speculated that the larger degrees of freedom of M1 made it less likely to reject the null hypothesis compared to the one degree of freedom tests. To understand in what range the power gained from less degrees of freedom could exceed the power lost from incorrect modeling and make these models more comparable in these ranges, I performed a preliminary exploratory analysis about the relationship between the degrees of freedom of different models and their power to detect the treatment effect.

#### *2.2.2.1 Simulation for preliminary study*

Since I was only interested in the effect of degrees of freedom on the power in this preliminary study, I used the same simulation methods as described in Section 2.1.1 and Section 2.1.2 to generate the data for 1000 subjects. I assumed that the probability of outcome for placebo group was 0.7 (log odds  $\beta_0 = 0.85$ ), the probability of outcome for treatment group was 0.56 at time zero (log odds ratio  $\beta_1 = -0.6$ ) and the treatment switched to harmful if initiated more than 1 hours (interaction  $\gamma = 0.6$ ) after injury. These parameters were set as constants for this preliminary study.

Then I varied the strength of interaction between treatment and treatment initiation time from injury and the distribution of that time according to the parameters given in Table 2. To vary the distribution of time from injury to treatment initiation, I increased the standard deviation and kept the mean at the upper bound of enrollment time frame. Then the mean of distribution of the time from injury to treatment initiation was at the lower bound of enrollment time frame and the standard deviation decreased. When the distributions of time from injury to treatment initiation were varied, enrollment time frame was limited as 0~2 hours and the effect of TXA in the treatment group would switch from beneficial to harmful at 1h after injury compared with placebo group.

When I varied the strength of interaction, I kept the treatment effect at time zero (log odds ratio  $\beta_1$ ) as -0.6, enrollment time frame was limited as 0~2 and the distribution of time from injury to treatment initiation with mean of 0 and standard deviation of 2 was kept same as well. I set the crossing time point at 1 hour first. That means compared to the placebo group, the effect of TXA in the treatment group would switch from beneficial to harmful at 1h after injury. Then I increase the crossing time point to 5 hours by one hour. This time point will be referred to as crossing time point in this thesis. When the treatment effect at time zero remains constant, the crossing time point depends on the strength of interaction between treatment and initiation time; smaller crossing time point indicates a stronger interaction.

Therefore, in two sets of scenarios above, the subjects changed between unevenly distributed and evenly distributed about the crossing time point by either shifting the

left skewed distribution of initiation time gradually to right skewed or changing the crossing time point over a constant initiation time distribution.

**Table 2: How distribution of time from injury to treatment initiation and strength of interaction changed to investigate the relationship between power and degrees of freedom**

	1	2
Results	3.1 Figure 1.1	3.1 Figure1.2
Distribution of true time from injury to treatment initiation time before truncation (mean; SD)	(2; 0.2), (2; 0.5), (2; 0.7), (2; 1), (2; 1.4), (2; 2), (0; 2), (0; 1.4), (0; 1), (0; 0.7), (0; 0.5) and (0; 0.2)	(0; 2)
Crossing time point of two treatment groups (hour)	1	0.5, 0.7, 0.85, 1, 1.5, 2, 2.2, 2.4, 2.6, 3 and 4
Enrollment time frame (hour)	0~2	0~2
Sample size	700, 1100, 1300, 1500, 1600, 1800, 1700, 1500, 1400, 1300, 1100, 700	1000, 1000, 1000, 1000, 1000, 1000, 1000, 1000, 1000, 1000, 1000, 1000
Log odds for placebo group; $\beta_0$ (eq1)	0.85	0.85
Time effect (log odds ratio) every hour; $\beta_2$ (eq1)	0	0
Treatment effect (log odds ratio) at time 0; $\beta_1$ (eq1)	-0.6	-0.6
Interaction effect between time and treatment; $\gamma$ (eq1)	0.6	1.2, 0.857, 0.706, 0.6, 0.4, 0.3, 0.273, 0.25, 0.231, 0.2 and 0.14

### 2.2.2.2 Analysis for preliminary study

I used same time dataset but didn't consider for measurement error. The full model (Model 1) and a reduced model without interaction term (Model 5) were both fitted with true time. Model 5 contains only the treatment main effect and time main effect. Compared with M1, M5 does not consider the interaction between treatment effect and initiation time.

$$\text{logit}(q) = \beta_0 + \beta_1 \text{trt} + \beta_2 \text{time.true} + \gamma \text{trt} \times \text{time.true} \quad \text{Model 1 (M1)}$$

$$\text{logit}(q) = \beta_0 + \beta_1 \text{trt} + \beta_2 \text{time.true} \quad \text{Model 5 (M5)}$$

The power of detecting treatment effect of test using Model 1 was calculated by the likelihood ratio test comparing Model 1 and a reduced model:  $\text{logit}(q) = \beta_0 +$

$\beta_2$  *time.true*. It was a two degrees of freedom test for both the main treatment effect  $\beta_1$  and interaction term  $\gamma$  equal to zero at significance level of 0.05.

The power of detecting treatment effect of test using Model 5 was calculated by the likelihood ratio test comparing Model 5 and a reduced model:  $\text{logit}(q) = \beta_0 + \beta_2 \text{time.true}$ . It was a one degrees of freedom test for only the main treatment effect  $\beta_1$  equals to zero at significance level of 0.05.

To visualize how the power to detect treatment effect of M1 and M5 would be different with each other under different settings, I maintained the power of test using Model 1 at around 80% by adjusting the sample size and used the same sample size as Model 1 in each scenario for Model 5. Two curves for the power of tests using two models under different settings of interaction and time distribution were plotted on same graph. (See Section 3.1)

## **2.3 IPD Meta-Analysis on ROC and CRASH-3**

A critical part of this thesis is to investigate how the power of detecting a time dependent treatment effect could be impacted by the presence of time measurement error when combining the data from CRASH-3 and the ROC TXA trial.

### *2.3.1 IPD Meta-Analysis for continuous initiation time*

Compared with classic meta-analysis that only based on summary statistics, individual patient data (IPD) meta-analysis utilizes more detailed information and can potentially reduce the publication bias of the classic literature based meta-analysis<sup>17, 18</sup>. IPD meta-analysis can also solve some scientific questions that cannot be solved by classic meta-analysis, but they both assume that the selected studies should be representative.<sup>19</sup>

One-stage model and two-stage model are two methods of performing meta-analysis on IPD. The one-stage method fits a single model for all the data and there were indicator variables for each of the studies included in the model. The two-stage method analyzes every study separately, and then computes the aggregated coefficients and p-values based on the estimated coefficients of each study and the weights of studies<sup>20</sup>. In this thesis, I used the one-stage method to combine the ROC TXA and CRASH-3 simulated datasets. This choice was pre-specified because sometimes one-stage and two-stage methods lead to different conclusions<sup>21</sup>. A previous study showed that one-stage method was more reliable when only few studies were available<sup>21</sup>. In the analysis of the simulated data, I gathered all the data from both studies and fit a single logistic regression model with a fixed effect for study on the combined data using Model 6 or 7.

Model 6 (M6):

$$\text{logit}(q) = \beta_0 + \beta_1 \text{trt} + \beta_2 \text{time.true} + \gamma \text{trt} \times \text{time.true} + \alpha \times I(\text{study} = \text{ROC})$$

Model 7 (M7):

$$\text{logit}(q) = \beta_0 + \beta_1 \text{trt} + \beta_2 \text{time.obs} + \gamma \text{trt} \times \text{time.obs} + \alpha \times I(\text{study} = \text{ROC})$$

M6 was the true model I used for simulation and included a fixed effect to adjust for the variance between two studies. M7 had same model structure as M6, but it was fitted with the estimated treatment initiations times that include measurement error.

For both models, to calculate the power of detecting a treatment effect, likelihood ratio tests comparing each of them and their corresponding reduced model were performed:

$\text{logit}(q) = \beta_0 + \beta_2 \text{time.true}$  for M6 and  $\text{logit}(q) = \beta_0 + \beta_2 \text{time.obs}$  for M7. They are

two degrees of freedom tests for both the main treatment effect  $\beta_1$  and interaction term  $\gamma$  equal to zero at significance level of 0.05.

Since there was a simulated time dependent effect of TXA, I performed Wald test for the interaction term  $\gamma$  to calculate the power of tests using M6 and M7 to detect interaction.

### 2.3.2 IPD Meta-Analysis for categorical initiation time

In CRASH-2 trial, the time from injury to treatment initiation was categorical variable and the heterogeneity in treatment effects across subgroups: 0~1 hour, 1~3 hours and >3 hours was assessed with chi-squared tests<sup>4</sup>. To make our study more comparable, I categorized the continuous variable, time from injury to treatment initiation, to 4 subgroups: 0~1 hour, 1~2 hours, 2~3 hours and >3 hour. Since the enrollment time frame of ROC TXA trial was 0~2 hours and that of CRASH-3 was combination of 0~3 hours and 0~8 hours, the continuous initiation time was also categorized at 2 hours. The logistic models M6 and M7 were fit with a categorized, thus more crude measure of time from injury to treatment initiation. They were written as M8 and M9 below.

Model 8 (M8):

$$\begin{aligned} \text{logit}(q) = & \beta_0 + \beta_1 \text{trt} + \beta_2 \times I(\text{time.true in 1~2 hours}) \\ & + \beta_3 \times I(\text{time.true in 2~3 hours}) + \beta_4 \times I(\text{time.true} > 3 \text{ hours}) \\ & + \gamma_1 \text{trt} \times I(\text{time.true in 1~2 hours}) + \gamma_2 \text{trt} \times I(\text{time.true in 2~3 hours}) \\ & + \gamma_3 \text{trt} \times I(\text{time.true} > 3 \text{ hours}) + \alpha \times I(\text{study} = \text{ROC}) \end{aligned}$$

Model 9 (M9):

$$\begin{aligned} \text{logit}(q) = & \beta_0 + \beta_1 \text{trt} + \beta_2 \times I(\text{time. obs in 1~2 hours}) + \beta_3 \times I(\text{time. obs in 2~3 hours}) \\ & + \beta_4 \times I(\text{time. obs} > 3 \text{ hours}) + \gamma_1 \text{trt} \times I(\text{time. obs in 1~2 hours}) \\ & + \gamma_2 \text{trt} \times I(\text{time. obs in 2~3 hours}) + \gamma_3 \text{trt} \times I(\text{time. obs} \\ & > 3 \text{ hours}) + \alpha \times I(\text{study} = \text{ROC}) \end{aligned}$$

I performed four degrees of freedom likelihood ratio tests comparing each of them to their corresponding reduced model:  $\text{logit}(q) = \beta_0 + \beta_2 \times I(\text{time. true in 1~2 hours}) + \beta_3 \times I(\text{time. true in 2~3 hours}) + \beta_4 \times I(\text{time. true} > 3 \text{ hours}) + \alpha \times I(\text{study} = \text{ROC})$  for M8 and

$\text{logit}(q) = \beta_0 + \beta_2 \times I(\text{time. obs in 1~2 hours}) + \beta_3 \times I(\text{time. obs in 2~3 hours}) + \beta_4 \times I(\text{time. obs} > 3 \text{ hours}) + \alpha \times I(\text{study} = \text{ROC})$  for M9. They tested for both the main treatment effect  $\beta_1$  and all interaction terms  $\gamma_1 \gamma_2 \gamma_3$  equal to zero at significance level of 0.05.

Then I performed likelihood ratio tests for power to detect interaction, comparing M8 M9 and their corresponding reduced model were performed:  $\text{logit}(q) = \beta_0 + \beta_1 \text{trt} + \beta_2 \times I(\text{time. true in 1~2 hours}) + \beta_3 \times I(\text{time. true in 2~3 hours}) + \beta_4 \times I(\text{time. true} > 3 \text{ hours}) + \alpha \times I(\text{study} = \text{ROC})$  for M8 and

$\text{logit}(q) = \beta_0 + \beta_1 \text{trt} + \beta_2 \times I(\text{time. obs in 1~2 hours}) + \beta_3 \times I(\text{time. obs in 2~3 hours}) + \beta_4 \times I(\text{time. obs} > 3 \text{ hours}) + \alpha \times I(\text{study} = \text{ROC})$  for M9. They are three degrees of freedom tests for all interaction terms  $\gamma_1 \gamma_2 \gamma_3$  equal to zero at significance level of 0.05.

Both models M8 and M9 assumed constant treatment effect over the time frame of each subgroups, but allowed a flexible structure that each subgroup could have different treatment effect compared with the reference subgroup 0~1 hour. This analysis

investigated the effect of measurement error when only crude data were available and made this study more comparable to the previous study on TXA, CRASH-2.

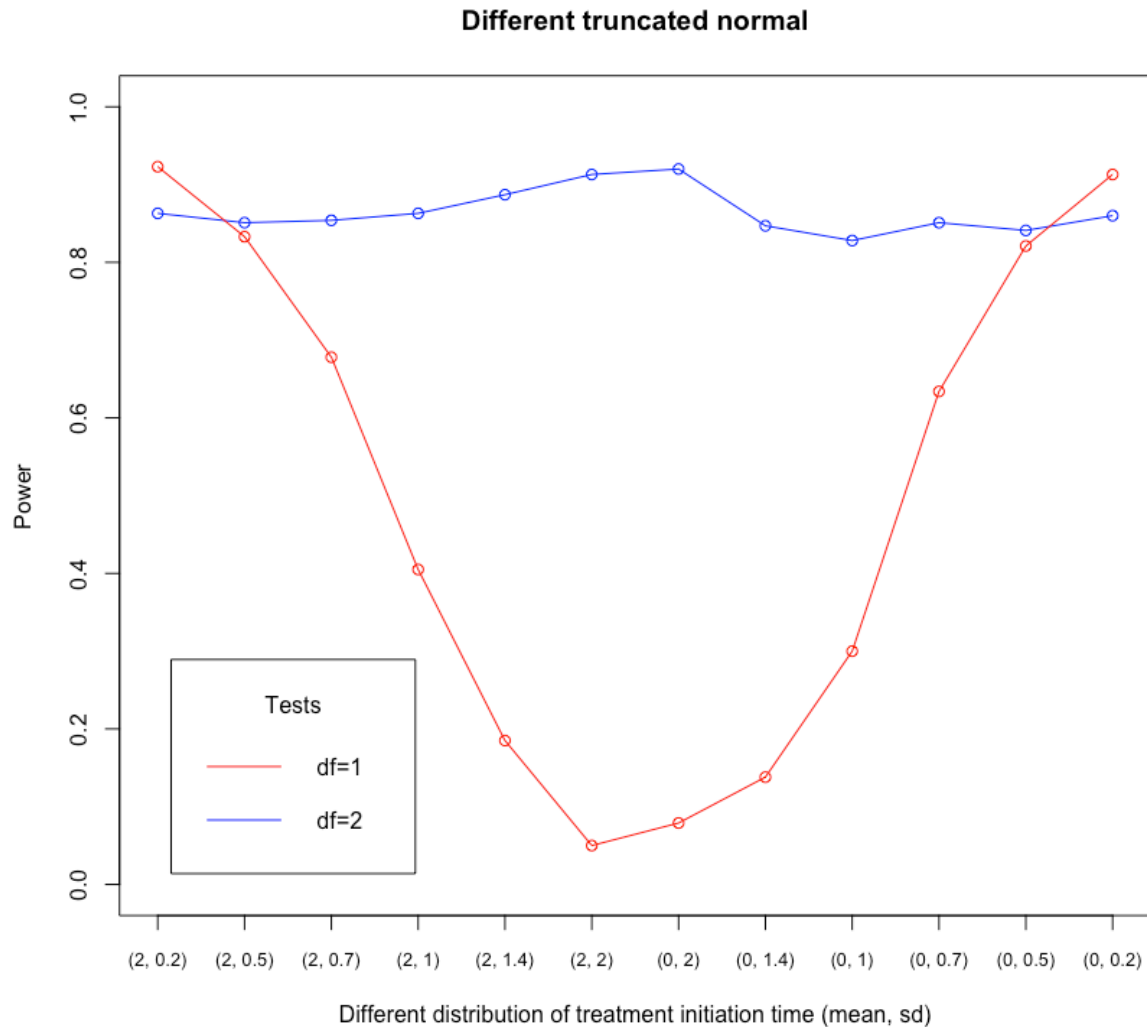
### **3. Results**

The results section is organized as follows. First, I present results on the relationship between the degrees of freedom and power of two likelihood ratio tests. Second, I present the potential effect of measurement error on a setting resembling the ROC TXA trial and third, how the choice of simulation parameters might influence this effect of measurement error. Finally, the results of meta-analysis showed the effect of measurement error on the dataset that pooled simulated ROC TXA data with simulated CRASH-3 data.

#### ***3.1 Relationship between degrees of freedom and power in the absence of time measurement error***

To analyze the relationship between the degrees of freedom of models and their power to detect the treatment effect, a full model that included an interaction term in addition to two main effects (M1) was compared with a reduced model that only included two main effects (M5; see Section 2.2.2.2). The mean of truncated normal before truncation was chosen at the upper bound of enrollment time frame (2 hours). I kept the mean constant but increased the variance from 0.2 to 2, making this highly skewed distribution flatter. Then I kept the mean constant at lower bound (0 hour) and decreased the variance from 2 to 0.2, making this distribution more skewed again. The first 6 distributions with mean at upper bound and increasing variance were symmetric to the other 6 distributions with mean at lower bound and decreasing variance. Therefore, the skewness of distribution of the initiation times shifts from the left to the right with less diminished skewness in the middle of the x-axis.

**Figure 1.1: Change in power when the distributions of time from injury to treatment initiation varied. The points represent the truncated normal distributions with original (mean, sd) before truncation: (2, 0.2); (2, 0.5); (2, 0.7); (2, 1); (2, 1.4); (2, 2); (0, 2); (0, 1.4); (0, 1); (0, 0.7); (0, 0.5); (0, 0.2), indicating the skew of initiation time distribution shift from left to right.**



The two points on the rightmost side and leftmost side represented the power to detect a treatment effect when times from injury to treatment initiation come from more highly skewed distributions. (Figure 1.1) At these two points, even though the data were simulated with the full model, the one degree of freedom likelihood ratio test for the (incorrect) reduced model had larger power compared with the two degree of freedom likelihood ratio test for the true model. As the standard deviation increased, the truncated normal distributions were less skewed and the power of the test using the

reduced model decreased. Two points in the middle of the x-axis presented the power when the truncated normal distribution with large standard deviation is close to a Uniform distribution. The power of the test using the reduced model was close to 0.05 in this setting. Both curves in Figure 1.1 were roughly symmetric, which was consistent with the symmetric distributions of time from injury to treatment initiation. The points at two ends indicated that when the distributions of time from injury to treatment initiation were highly skewed, the one degree of freedom likelihood ratio test for a reduced model gave larger power compared with the two degree of freedom likelihood ratio test for the true model. To maintain the power of two degrees of freedom test using M1 to detect treatment effect around 0.8, the sample size was increased when the distribution was flatter and decreased when the distribution was more skewed (Table 2).

**Figure 1.2: Change in power over crossing time point. The x-axis is the time point where the treatment effect changed from beneficial to harmful, indicating decreasing interaction strength when the treatment effect at time zero is constant. The power was calculated at 0.5h, 0.7h, 0.85h, 1h, 1.5h, 2h, 2.2h, 2.4h, 2.6h, 3h and 4h.**

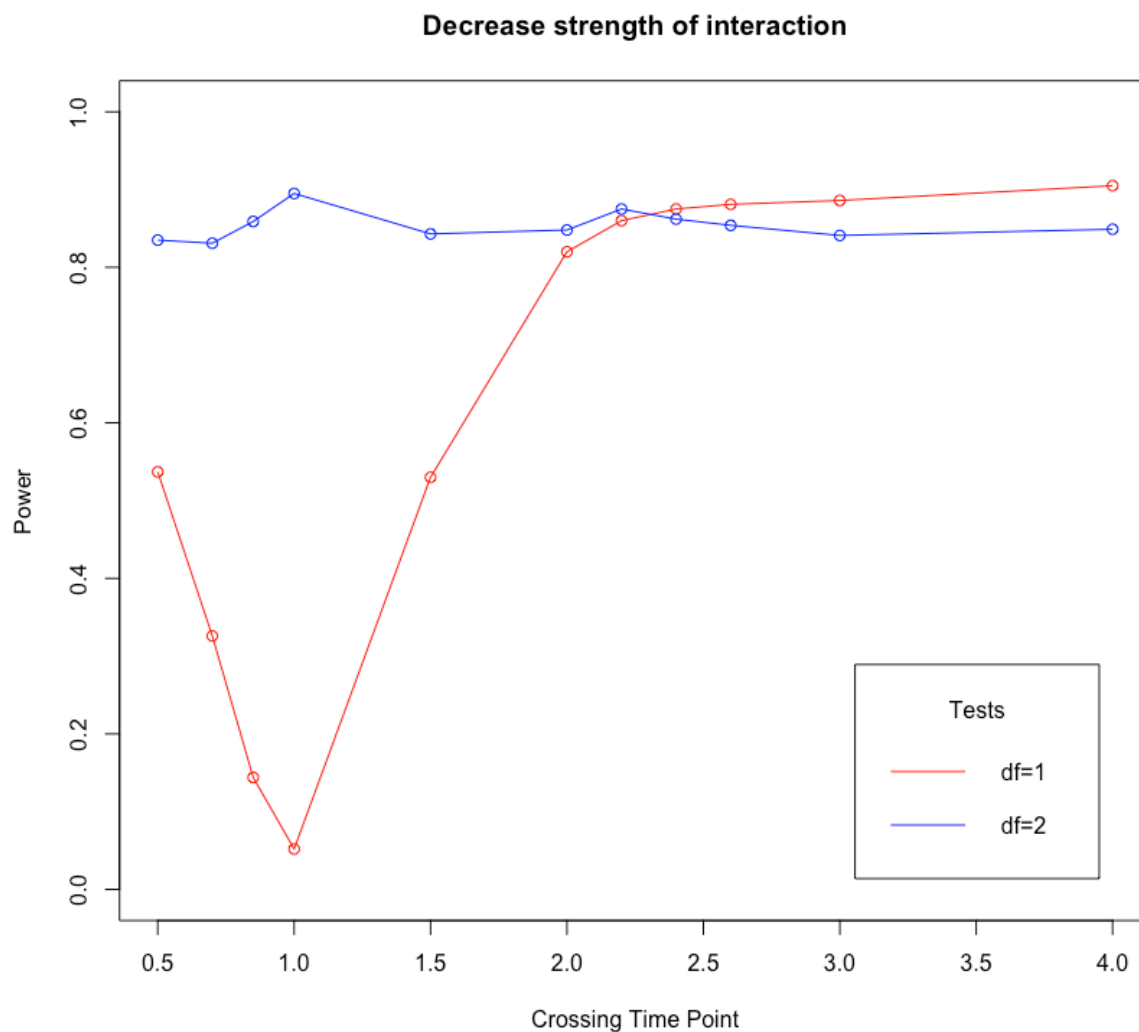


Figure 1.2 showed the power of one degree of freedom test dramatically decreased as the crossing time point got larger and approached 0.05 at 1 hour. Then the power kept increasing as the crossing time point increased from 1 hour and exceed the power of two degrees of freedom test at round 2.3 hours. The enrollment time frame was kept as 0 ~ 2 hour for this set of scenarios, which meant the treatment did not show harmful effect on any subject (in the simulated studies) if the crossing time point was larger than

2 hours. The power of the test using reduced model kept increasing when the interaction strength became either stronger or weaker compared with that at 1h. The points at 2.4h indicated that when strength of interaction decreased to a certain extent, the one degree of freedom likelihood ratio test for a reduced model had larger power compared to the two degree of freedom likelihood ratio test for the true model.

### 3.2 Differences among 4 models in ROC setting resembling the ROC TXA trial

To investigate the effects of time measurement error on the coefficient estimation, power to detect treatment effect, power to detect interaction and type I error in the setting resembling the ROC TXA trial, I Simulated the data for 1000 subject based on the parameters I assumed in Table 1 column 1.

**Table 3: Power, Type I error and bias of 4 models in an assumed “ROC setting”.**

	<b>M1:</b> $\text{logit}(q) = \beta_0 + \beta_1 \text{trt} + \beta_2 \text{time.true} + \gamma \text{trt} \times \text{time.true}$	<b>M2:</b> $\text{logit}(q) = \beta_0 + \beta_1 \text{trt} + \beta_2 \text{time.obs} + \gamma \text{trt} \times \text{time.obs}$	<b>M3:</b> $\text{logit}(q) = \beta_0 + \beta_1 \text{trt} + \beta_2 \text{time}$	<b>M4:</b> $\text{logit}(q) = \beta_0 + \beta_1 \text{trt}$
Power for Trt	0.668	0.656	0.736	0.739
Type I for $\beta_2$	0.05	0.054	0.172	NA
Power for $\gamma$	0.102	0.093	NA	NA
Bias for $\beta_0$	-0.00731	-0.00243	NA	NA
Bias for $\beta_1$	0.00388	0.0349	NA	NA
Bias for $\beta_2$	0.0112	0.00800	NA	NA
Bias for $\gamma$	-0.00928	-0.0143	NA	NA

\* NA: These results were not applicable, because interaction  $\gamma$  was not contained in M3 and M4, and time main effect  $\beta_2$  was not contained in M4.

\*  $\beta_0 = 0.85, \beta_1 = -0.6, \beta_2 = 0, \gamma = 0.2$

Given a sample size of 1000, the likelihood ratio test using correct model with true time from injury to treatment initiation (M1) had a power of 0.668 to detect the treatment effect, which was slightly larger compared with a power of 0.656 using model M2, but both tests using M1 and M2 had slightly lower power than the one degree of freedom likelihood ratio test using M3 and M4. The Type I error, the rate of falsely rejecting that the coefficient for the time main effect was zero, were around 0.05 for both likelihood ratio tests using M1 and M2 while the type I error for the test using M3 was much

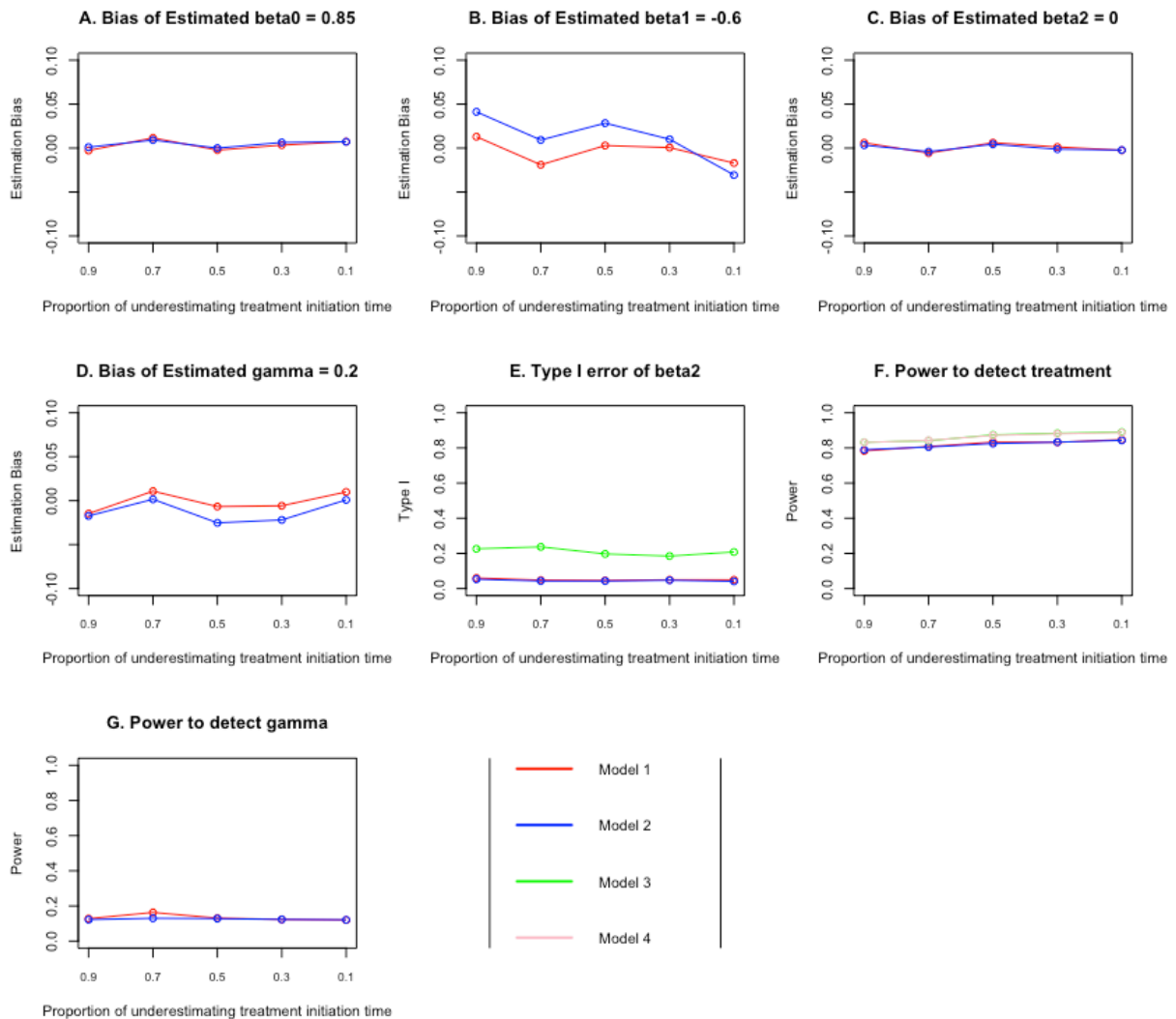
higher. The test using M1 had slightly higher power to detect the interaction effect than the test using M2, but their powers were both small around 0.1. For all the coefficients, estimation bias was small for both models, M1 and M2. In this scenario, all the results are consistent with our expectation.

### **3.3 Differences among 4 models in other scenarios**

#### *3.3.1 Effects of proportion of underestimating time from injury to treatment initiation on power and bias*

When I varied the proportion of underestimating time from injury to treatment initiation, I kept other simulation parameters constant (Table 1 Column 2) and set the sample size to 1400 to maintain the power to detect treatment effect in M1 around 80%. I decreased the proportion of underestimating time from injury to treatment initiation time from 90% to 10% by 20%.

**Figure 2: Estimation bias for all four coefficients in M1 and M2, Type I error for time effect (zero) in M1 M2 and M3 and power to detect treatment effect (all models) and interaction effect in M1 and M2 when the proportion of underestimating time from injury to treatment initiation is varied and 80% power detecting treatment effect in M1 was maintained. The x-axis is the proportion of underestimation of the treatment time.**



In this setting, the tests using M1 and M2 had similar power to detect treatment effect, but their power was slightly lower compared with the one degree of freedom tests using M3 and M4 (Figure 2F). There was a slight increasing trend for the power to detect treatment effect. As the proportion of underestimation decreased, the distributions for both simulated true time and observed time shifted towards zero and the distributions were more skewed. As the results of preliminary study showed, smaller sample size was

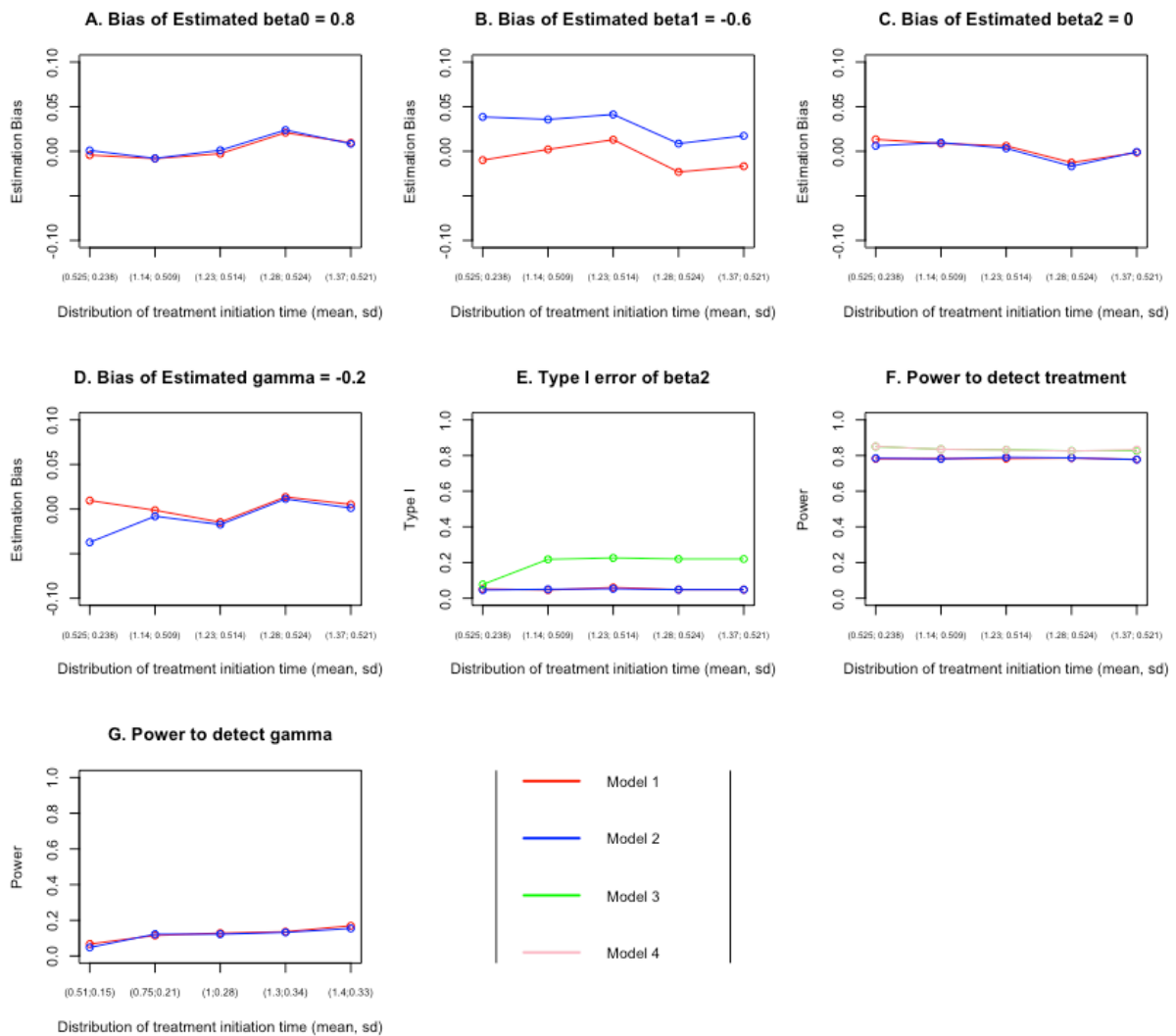
required to maintain the power to detect treatment effect when the distribution was more skewed, so the power increased here when I didn't change the sample size.

For Type I error, the rate of false significant time main effect, tests using M1 and M2 were around 0.05 as we expected but test using M3 was much larger around 0.2 (Figure 2E). Tests using M1 and M2 had similar power to detect the interaction effect when the proportion of underestimated time from injury to treatment initiation were varied, but the powers were very small around 0.1 given this sample size and strength of interaction (Figure 2G). For all estimated coefficients,  $\beta_0$   $\beta_1$   $\beta_2$  and  $\gamma$ , both M1 and M2 had small bias (Figure 2A 2B 2C 2D) and we could see M2 always underestimated  $\gamma$  compared with M1. When the percentage of underestimation varied, there was no obvious trend for the power to detect interaction, Type I error or bias.

### *3.3.2 Effects of distribution of true initiation time from injury on power and bias*

When I varied the distribution of true initiation time from injury, I kept other simulation parameters constant and manually adjusted the sample size (Table 1 Column 3) to maintain the power to detect treatment effect in M1 around 80%. I increased both the mean and standard deviation of the distribution of true initiation time from injury, shifting the distribution towards the crossing time point (3h).

**Figure 3: Estimation bias for all four coefficients in M1 and M2, Type I error for time effect (zero) in M1 M2 and M3 and power to detect treatment effect (all models) and interaction effect in M1 and M2 when the distribution of time from injury to treatment initiation varied and 80% power detecting treatment effect in M1 was maintained. The true times from injury to treatment initiation are distributed within the range of 15 minutes to 12 hours. The x-axis is the actual mean and standard deviation of truncated normal distribution after truncation. The scenarios with (mean, sd)= (0.525; 0.238), (1.14; 0.509), (1.23; 0.514), (1.28; 0.524) and (1.37; 0.521) are examined.**



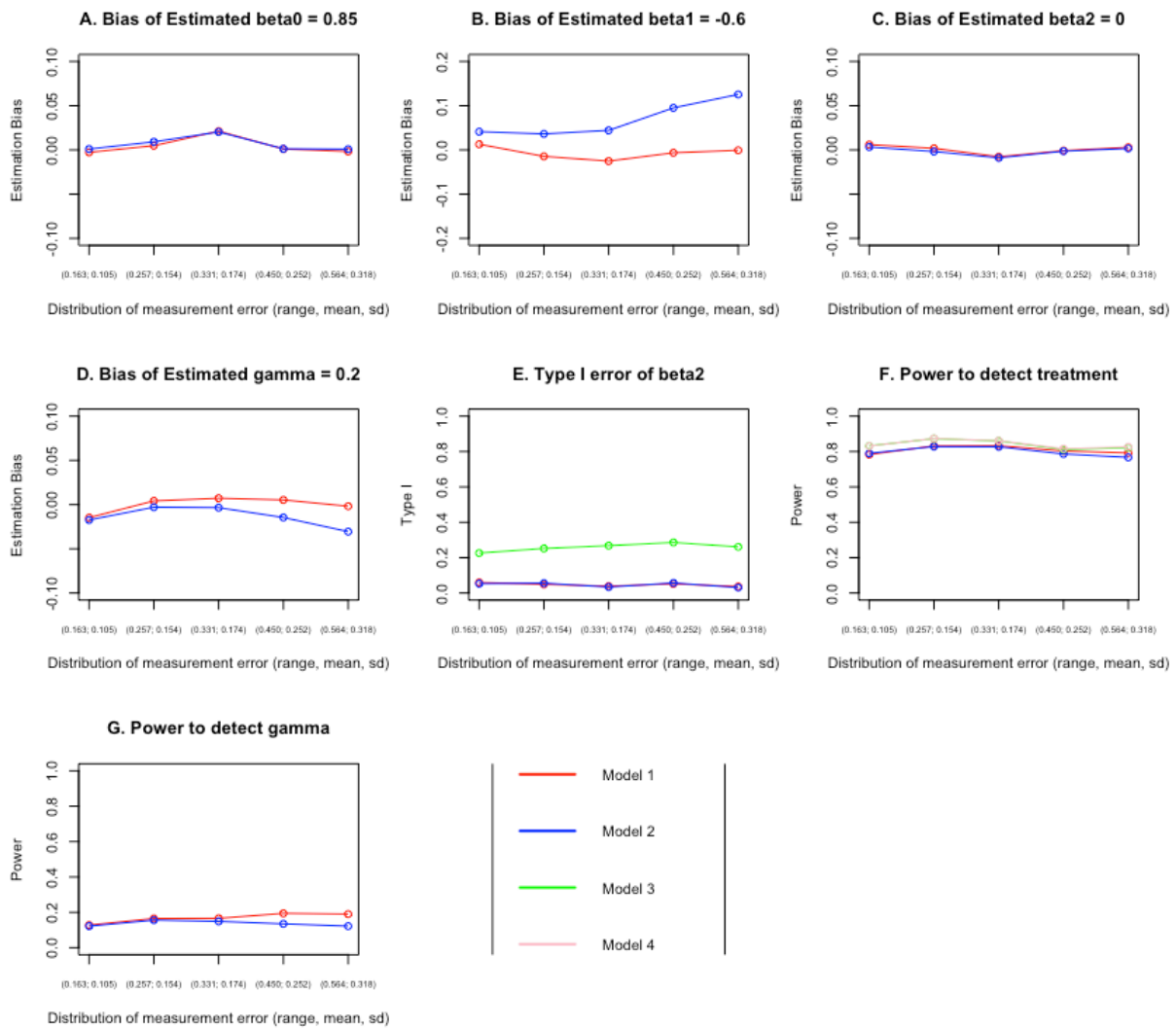
In the presence of measurement error, the test using M2 had similar power to detect treatment effect as the test using M1, but their power were smaller than the one degree of freedom tests using M3 M4 (Figure 3F). The tests using M1 and M2 to detect time main effect both had Type I error around 0.05 as I expected (Figure 3E). The test using

M2 had similar power to detect interaction as the test using M1, but their powers were both small at 0.05 when the mean was 0.5 and standard deviation was 0.3, and increased to around 0.15 when measurement error distribution had larger mean and variance (Figure 3G). For all estimated coefficients  $\beta_0$   $\beta_1$   $\beta_2$  and  $\gamma$ , both M1 and M2 had small bias (Figure 2A 2B 2C 2D) and we could see M2 always overestimated  $\beta_1$  and underestimated  $\gamma$  compared with M1.

### *3.3.3 Effects of distribution of time measurement error on power and bias*

When I varied the distribution of time measurement error, I kept other simulation parameters constant and manually adjusted the sample size (Table 1 Column 4) to maintain the power to detect treatment effect in M1 around 80%. I increased both the mean and standard deviation of the distribution of time measurement error.

**Figure 4a: Estimation bias for all four coefficients in M1 and M2, Type I error for time effect (zero) in M1 M2 and M3 and power to detect treatment effect (all models) and interaction effect in M1 and M2 when the distribution of time measurement error varied and 80% power detecting treatment effect in M1 was maintained. The x-axis is the actual mean and standard deviation of truncated normal distribution after truncation. The scenarios with (mean, sd)= (0.163; 0.105), (0.257; 0.154), (0.331; 0.174), (0.450; 0.252) and (0.564; 0.318) are examined.**

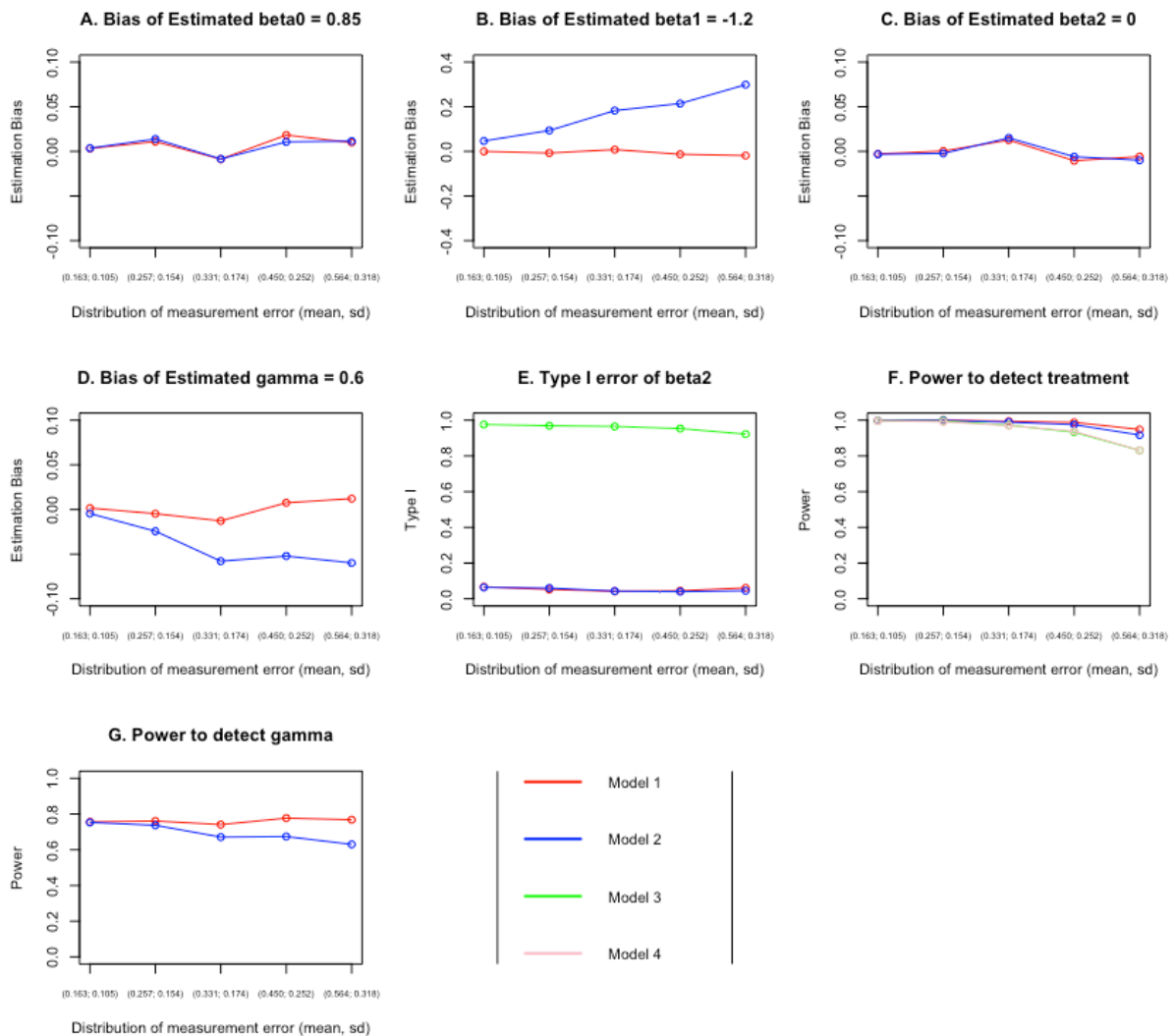


For the first three scenarios (measurement error distribution: mean  $\leq 0.33$  and  $sd \leq 0.17$ ), the tests using M1 and M2 had similar power to detect treatment effect and interaction. The powers of tests using M3 and M4 to detect treatment effect were slightly higher. The power of test using M2 to detect treatment effect was slightly lower than that using M1 when the mean of measurement error was 0.564 and standard error was 0.318 (Figure 4a F). The difference between tests using M1 and M2 in power to detect

interaction also became larger when mean of measurement error was increased to 0.45 and standard error was increased to 0.252 (Figure 4a G). The tests using M1 and M2 both had Type I error around 0.05 (Figure 4a E). For estimated  $\beta_0$  and  $\beta_2$ , both M1 and M2 had consistently low bias (Figure 4a A and 4a C). For estimated treatment effect at time zero  $\beta_1$ , M1 had a constantly low bias while the bias of M2 increased (Figure 4a B) and we could see M2 always overestimated  $\beta_1$  and underestimated  $\gamma$  compared with M1. As the mean and standard deviation of measurement error increased, there is decreasing trends of power to detect treatment effect and interaction for M2 M3 M4 and an increasing trend of treatment effect estimation bias for M2.

Even though Figure 4a E showed larger difference in power to detect interaction term when the measurement error increased, the tests using M1 and M2 both only had power around 0.2. To investigate the effect of measurement error on power to detect interaction, I kept other simulation parameters constant and manually adjusted the sample size (Table 1 Column 5) to maintain the power to detect interaction in M1 around 80%. I increased both the mean and standard deviation of the distribution of time measurement error.

**Figure 4b: Estimation bias for all four coefficients in M1 and M2, Type I error for time effect (zero) in M1 M2 and M3 and power to detect treatment effect (all models) and interaction effect in M1 and M2 when the distribution of time measurement error varied and 80% power detecting interaction in M1 was maintained. The x-axis is the actual mean and standard deviation of truncated normal distribution after truncation. The scenarios with (mean, sd)= (0.163; 0.105), (0.257; 0.154), (0.331; 0.174), (0.450; 0.252) and (0.564; 0.318) are examined.**



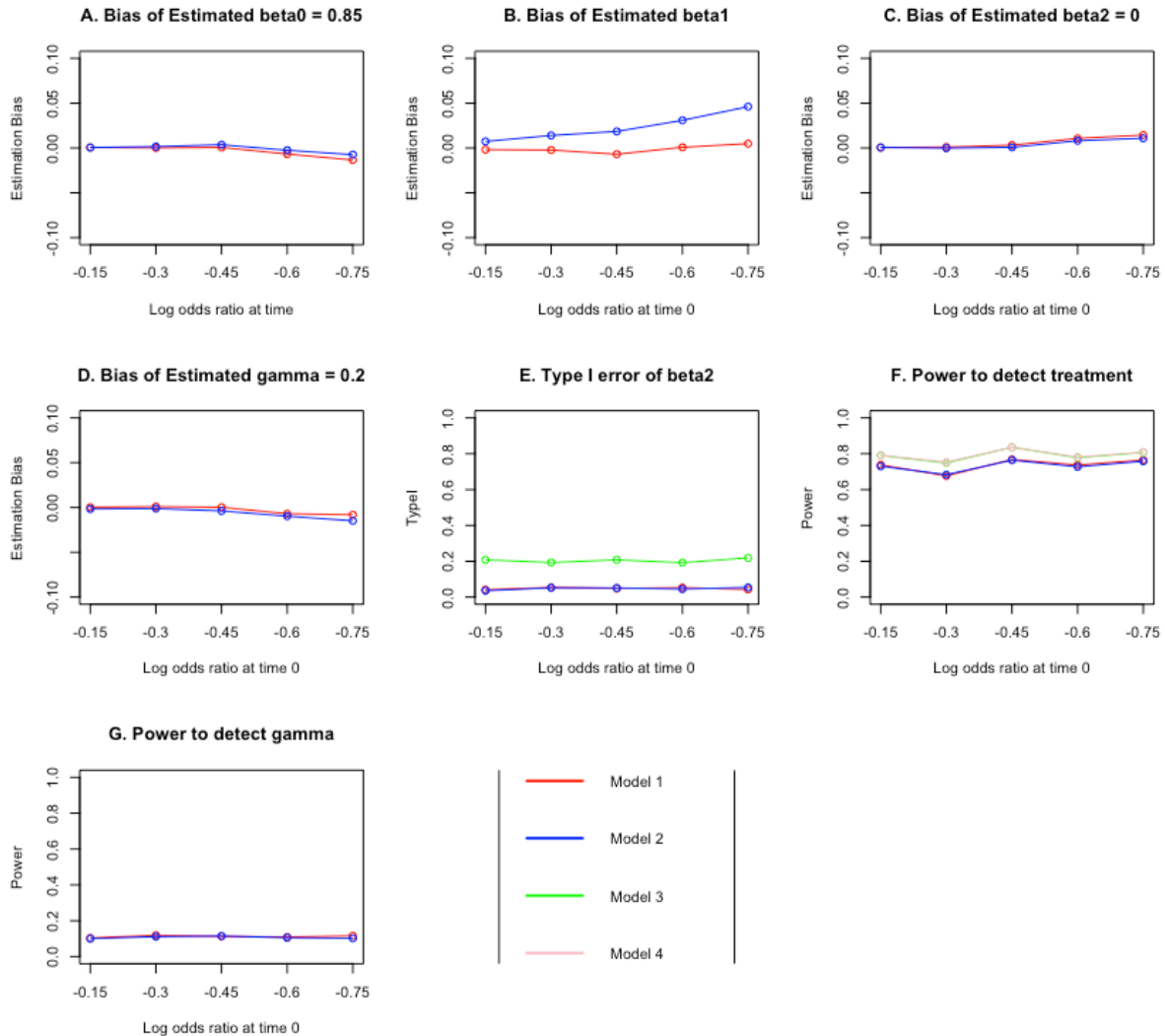
The power of test using M1 for detecting interaction was maintained around 80%. For both detecting treatment effect and detecting interaction, the tests using M1 and M2 had similar high power when the mean and standard deviation of measurement error were small. Then the power of test using M2 to detect interaction started decreasing when the mean was larger than 0.163 and standard deviation was larger than 0.105(Figure 4b G). Since the proportion of underestimation was kept at 90% and

distribution of true initiation time from injury was less skewed as the measurement error increased, so the power of test using M1 to detect treatment effect also slightly decreased when the sample size was adjusted to maintain the power to detect interaction term (Figure 4b F). The power of test using M2 for detecting treatment effect started decreasing when the mean was larger than 0.45 and standard deviation was larger than 0.252 (Figure 4b F). For Type I error, the rate of false significant time main effect, the tests using M1 and M2 were both around 0.05 under all scenarios (Figure 4b E). For estimated  $\beta_0$  and  $\beta_2$ , both M1 and M2 had constantly low bias (Figure 4b A and 4b C). For estimated treatment effect at time zero  $\beta_1$  and interaction  $\gamma$ , M1 had a consistently low bias while the bias of M2 steadily increased and we could see M2 always overestimated  $\beta_1$  and underestimated  $\gamma$  compared with M1 (Figure 4b B and 4b D). As the mean and standard deviation of measurement error increased and the power of test using M1 for detecting interaction was maintained at a high level, the test using M2 had decreasing trends of power to detect treatment effect and interaction and increasing trends of treatment effect estimation bias and interaction estimation bias.

#### *3.3.4 Effects of the strength of treatment effect at time 0 on power and bias*

When I varied the strength of treatment effect at time 0, I kept other simulation parameters constant and manually adjusted the sample size (Table 1 Column 6) to maintain the power to detect treatment effect in M1 around 80%. I decreased the log odds ratio at time 0 from -0.15 to -0.75 by 0.15, indicating an increase of treatment effect at time 0.

**Figure 5: Estimation bias for all four coefficients in M1 and M2, Type I error for time effect (zero) in M1 M2 and M3 and power to detect treatment effect (all models) and interaction effect in M1 and M2 when the strength of treatment effect at time 0 increased and 80% power detecting treatment effect in M1 was maintained. The x-axis is the log odds ratio at time 0. Five scenarios with log odds ratio=-0.15, -0.3, -0.45, -0.6, -0.75 were examined.**



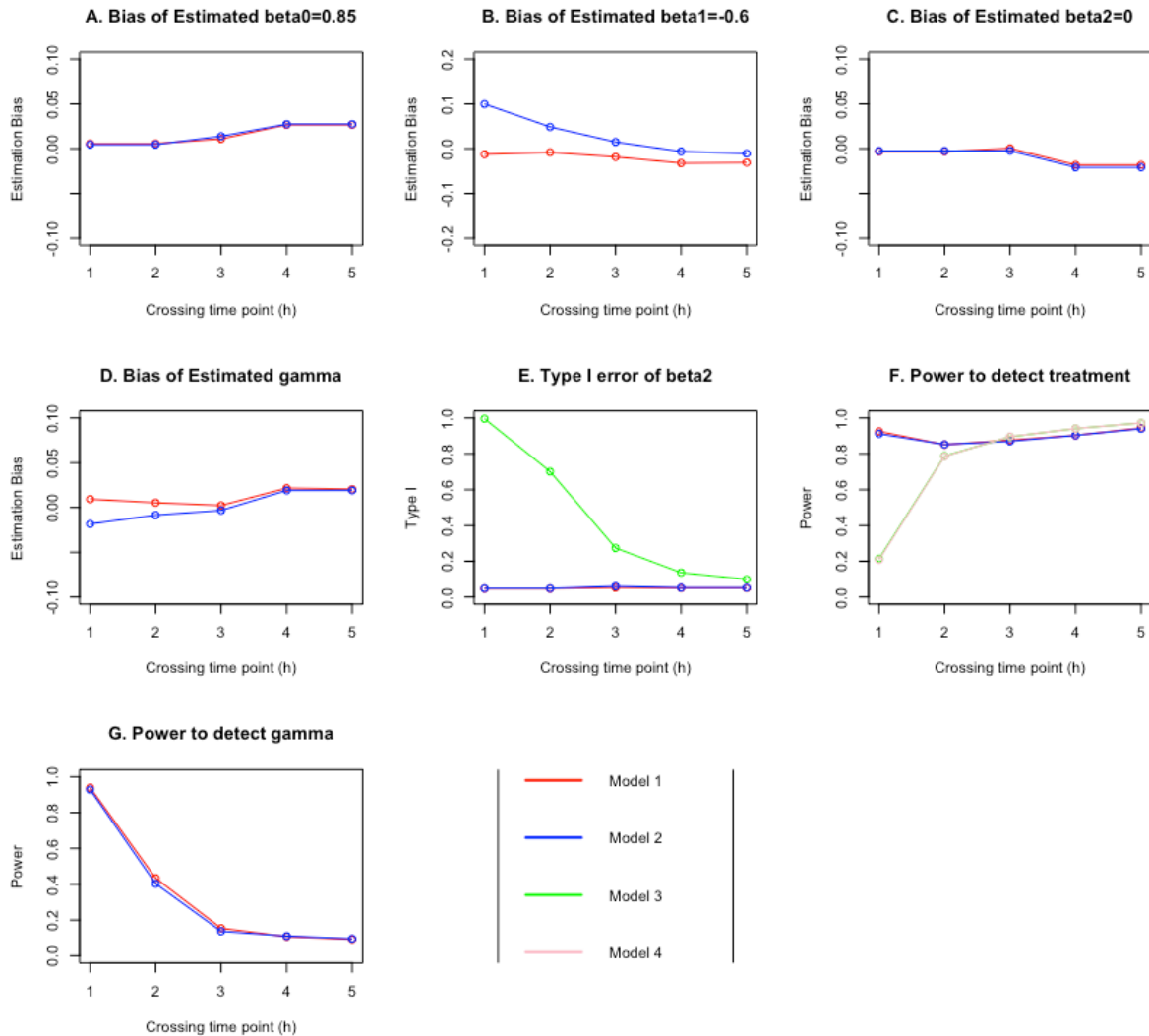
In the presence of measurement error, test using M2 had similar power to detect treatment effect as test using M1, but their powers were slightly smaller than one degree of freedom tests using models M3 M4 (Figure 5E). Both tests using M1 and M2 had Type I error around 0.05 under all scenarios and the test using M3 had a relatively larger Type I error around 0.2 (Figure 5F). Test using M2 had similar power to detect interaction as test using M1, but they were both small at around 0.15 (Figure 5G). For all estimated coefficients  $\beta_0$   $\beta_1$   $\beta_2$  and  $\gamma$ , both M1 and M2 had small bias and we could see

M2 always overestimated  $\beta_1$  compared with M1 (Figure 5A 5B 5C 5D). As strength of treatment effect at time 0 increased, the estimation bias of M2 for  $\beta_1$  also increased.

### *3.3.5 Effects of the strength of time dependent effect on power and bias*

When I varied the strength of time dependent effect, I kept other simulation parameters constant and manually adjusted the sample size (Table 1 Column 7) to maintain the power to detect treatment effect in M1 around 80%. I increased the crossing time point from 1 hour to 5 hours by 1 hour, indicating a decrease of time dependent effect when the treatment effect at time 0 was constant.

**Figure 6: Estimation bias for all four coefficients in M1 and M2, Type I error for time effect (zero) in M1 M2 and M3 and power to detect treatment effect (all models) and interaction effect in M1 and M2 when the strength of interaction decreased and 80% power detecting treatment effect in M1 was maintained. The x-axis is the crossing time point of two treatment groups. The scenarios with t=1h, 2h, 3h, 4h and 5h are examined.**



As the crossing time point increased, tests using M1 and M2 always had similar power to detect treatment effect. The tests using M1 and M2 had much larger power to detect treatment effect than the tests using M3 and M4 when two treatment groups had same log odds at 1 hour (Figure 6F). As the crossing time got later than 2 hours, the power of tests using M3 M4 increased and their power exceeded that of tests using M1 and M2 between 2 hours and 3 hours. For Type I error, the rate of false significant time main

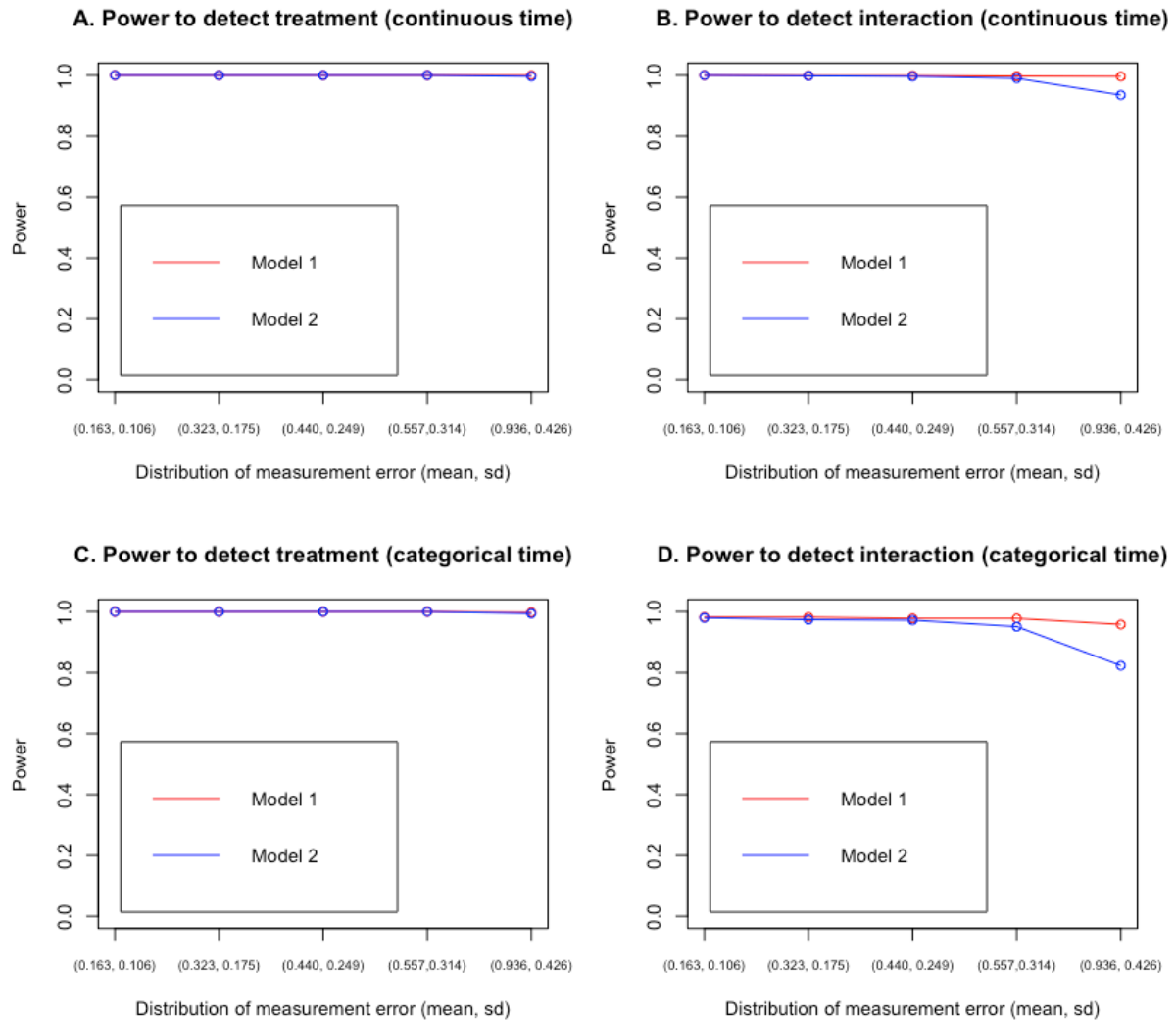
effect, tests using M1 and M2 were both around 0.05 under all scenarios. The test using M3 had a very large Type I error closed 1 at 1 hour and it kept decreasing as the crossing time point became larger (Figure 6F). As the strength of treatment effect decreased, the tests using M1 and M2 had similar power to detect interaction and they both declined rapidly (Figure 6G). For estimated coefficients  $\beta_0$ ,  $\beta_2$  and  $\gamma$ , both M1 and M2 had similar small bias (Figure 6A 6C 6D). M2 always overestimated  $\beta_1$  compared with M1, and the bias of M2 decreased as crossing time point decreased.

### ***3.4 Effects of time measurement error on power and bias in IPD Meta-Analysis***

I pooled a data resembling the ROC TXA trial with size of 1000 and a data resembling the CRASH-3 with size of 10000. The sample of 1000 was consisted of subjects with enrollment time frame 0~2 hours. The 7143 of 10000 subjects had enrollment time frame 0~8 hours and the other 2857 subjects had enrollment time frame 0~3 hours. I kept other simulation parameters constant and increased the both the mean and standard deviation of the distribution of time measurement error.

I fit the pooled data with M6, the true model I used for simulation and included a fixed effect to adjust for the variance between two studies, and M7 that having same model structure as M6 but fitted with the estimated time from injury to treatment initiation that includes measurement error. Then I fit both M6 and M7 with categorized time from injury to treatment initiation (M8 M9).

**Figure7: Power in IPD meta-analysis to detect treatment effect and interaction in M1 and M2 when the distribution of time measurement error varied. The x-axis is the actual mean and standard deviation of truncated normal distribution after truncation. The scenarios with (mean, sd)= (0.163, 0.106), (0.323, 0.175), (0.440, 0.249), (0.557, 0.314), (0.936, 0.426) are examined.**



Compared with the results from scenario resembling ROC TXA trial (section 3.2), both models fitted with simulated true time (M6 M8) and the model considered time measurement error (M7 M9) greatly improved the power of test to detect treatment effect and interaction closed to 1 by pooling the simulated ROC dataset with the simulated CRASH-3 dataset. As I increased the mean of measurement error to 1 hour, the time measurement error did not show influence on the power of detecting treatment effect, even in the cases where initiation time were categorized into 4

categories (Figure 6A 6C). When initiation times were measured as continuous variable, the time measurement error did not show influence on the power of detecting time treatment interaction until the mean of measurement error was 0.557 hour (33 minutes) (Figure 6B). When initiation times were categorized, the time measurement error didn't show influence on the power of detecting time treatment interaction until the mean of measurement error was 0.44 hour (26 minutes) (Figure 6D). In these five scenarios, tests using four models (M6, M7, M8 and M9) all had power above 80% to detect both treatment effect and interaction.

## 4. Discussion

The discussion section is organized as follows. First, I briefly introduce the approach I used in this thesis and provide some perspectives of the model. Second, I present my conclusion on the relationship between the degrees of freedom and power of two likelihood ratio tests. Third, I present my conclusion on the effect of measurement error on a setting resembling the ROC TXA trial and on other more general settings. Then, I present the conclusion on the meta-analysis for the simulated ROC TXA and simulated CRASH-3 data. Finally, I discussed the limitations of this study and the potential future directions.

In our study of time measurement error, its potential effects on the analysis of time dependent treatments like TXA were illustrated by performing simulations under different settings. The time dependent treatments as discussed in this thesis referred to trials where the treatment effect varied as the time from injury to treatment initiation varied. Using simulations, I fit logistic regression models of interest with interaction term and compare Type I error rates, power and estimation bias in the presence of measurement error. Models that did not include interaction effect (as often used in primary analyses of clinical trials) were also investigated and compared with full models that included interaction term in this study.

We used a specific model for many of the simulations. To make this model easier to understand on the probability scale, we provided different probabilities at different time points for both groups. Based on our assumptions for the ROC TXA trial, I assumed that the probability of unfavorable GOS-E score for placebo was 0.7 and constant over time. The probability of unfavorable GOS-E score for treatment group increased as the

time from injury to treatment initiation increased, and this increase of probability was linear on log odds scale. At time zero (which is only theoretically possible), the probability of unfavorable GOS-E score for treatment group was assumed to be 0.562. The minimum time at which patients could start to receive treatment we assumed was 15 minutes after injury, at this time the probability of unfavorable GOS-E score was assumed to be 0.574. At the maximum enrollment time of ROC TXA trial, 2 hours after injury, the probability was 0.657 for treatment group. Both of these probabilities represent a benefit of the treatment compared to the placebo group. At the crossing time point we assumed, 3 hours after injury, the probability was 0.70 for treatment group which is the same as that for placebo group. At the maximum enrollment time of CRASH-3, 8 hours after injury, the probability was 0.864 for treatment group and thus represents a harmful treatment effect compared to the placebo group (Appendix C).

The preliminary analysis on degrees of freedom and power in section 3.1 showed the situations where the one degree of freedom test in a model without an interaction term had larger power to detect a treatment effect than the two degrees of freedom test in a model that includes an interaction term when both models were fitted with simulated true time. If the time from injury to treatment initiation was evenly distributed over the enrollment time frame and the crossing time point was around the midpoint of enrollment time frame, including the treatment by time interaction term was important features for avoiding great power reduction. In this situation, because the number of subjects in the experimental group experienced harmful treatment effect was approximately equal to that experienced beneficial treatment effect, the treatment effect at initiation time and the time dependent effect would be missed if we assumed the treatment effect was constant over time (as in models M3 M4 and M5). When the

beneficial treatment effect and harmful treatment effect canceled each other out, the overall treatment effect should be zero. Therefore, the power of the one degree of freedom test to detect treatment effect was around 0.05 (Figure 1.1 and 1.2), which was consistent with my expectation because the power is actually the Type I error now.

When the subjects were not evenly distributed on the two sides of crossing time point, the power of the test for treatment effect in model 5

$(\text{logit}(q) = \beta_0 + \beta_1 \text{trt} + \beta_2 \text{time.true})$  increased. Because both M5 and  $\text{logit}(q) = \beta_0 + \beta_1 \text{trt} + \beta_2 \text{time.true} + \gamma \text{trt} \times \text{time.true}$  (M1) were compared with the same reduced model  $\text{logit}(q) = \beta_0 + \beta_2 \text{time.true}$ , the likelihood ratio test for M1 had one more degree of freedom than that for M5. The likelihood ratio test with less degree of freedom had smaller critical value, so the likelihood ratio test for M5 was more likely to reject the null hypothesis when M1 and M5 gave similar statistic values. For models tested by one degree of freedom likelihood ratio test like M5, the power gained from degree of freedom could sometimes exceed the power lost from incorrect modeling. Therefore, under the scenarios I simulated, the test using the model that only included the treatment main effect had slightly higher power to detect treatment effect than the correct model that included a time treatment interaction term, if the distribution of initiation time was highly skewed to the boundary of enrollment time frame.

My study on setting resembling ROC TXA trial showed that based on our assumptions for all the parameters, the ROC study with sample size of 1000 had power of 66.8% in the absence of measurement error. (Table 3) When the measurement error was considered in model  $\text{logit}(q) = \beta_0 + \beta_1 \text{trt} + \beta_2 \text{time.obs} + \gamma \text{trt} \times \text{time.obs}$  (M2) and true time from injury to treatment initiation could not be observed, the power to detect treatment effect was slightly attenuated compared with the model fitted with simulated

true time (M1). The Table 3 also showed that the reduced models  $\text{logit}(q) = \beta_0 + \beta_1 \text{trt} + \beta_2 \text{time. obs}$  (M3) and  $\text{logit}(q) = \beta_0 + \beta_1 \text{trt}$  (M4) had larger power to detect the treatment effect. As explained in the previous paragraph, even though M3 and M4 were not the model I used for generating data, power gained by less degree of freedom likelihood ratio tests exceeded the power lost from incorrect modeling. The simulation parameters for results in Table 3 were set to mimic the real setting of ROC TXA trial. Therefore, if our assumptions were appropriate (with average absolute measurement error of 10 minutes), initiation time measurement error or using incorrect model M3 M4 would not reduce the power to detect treatment effect; the measurement error would not greatly increase the estimation bias for all the coefficients in these scenarios.

The studies on M1 and M2 in more general settings (Section 3.3) revealed that in the scenarios I investigated, the increased mean and standard deviation of absolute measurement error introduced (1) larger estimation bias for treatment effect at time zero  $\beta_1$  and coefficient for the interaction between treatment and time main effect  $\gamma$ ; (2) reduced power to detect treatment effect and interaction. If my assumptions were appropriate, the effect of measurement error on power was only detectable when I increased the time dependent effect by increasing the strength of treatment effect at time 0 and decreasing the crossing time point from the setting similar to the ROC TXA trial (Figure 4b). When  $\beta_1$ , the log odds ratio of experimental group at time 0, was increased to -1.2 and the interaction was strengthened that crossing time was 2 hours, we could have sufficient power of 80% to detect interaction and see the impact of measurement error. In the setting resembling ROC TXA trial, if my assumptions were appropriate, even though the mean of absolute measurement error increased up to 1 hour, the impact of measurement error was relatively small (Figure 4a).

The studies on M3 and M4 in Results 3.3 revealed that in the scenarios I investigated, the impact of missing an interaction term on analyzing a time dependent treatment only associated with the distribution of initiation time and strength of time treatment interaction (Figure 6). This was consistent with my discussion on the preliminary analysis that using normal model that did not consider the interaction term for a time dependent treatment would greatly reduce the power to detect treatment effect when the time from injury to treatment initiation was evenly distributed over the enrollment time frame and the crossing time point was around the midpoint of enrollment time frame. Since the enrollment time frame was 0~2 hours and if a crossing time point of 3 hours is realistic for ROC TXA trial, that means given the treatment had interaction term, the power to detect treatment effect of test using model without considering interactions is expected to be larger than the power of test using model that includes the interaction term in real analysis.

Since obtaining accurate time is infeasible or extremely difficult in the emergency medical setting, we might expect a crucial effect of measurement on a time dependent treatment like ROC TXA. In most of cases, we can only obtain the accurate time between 911 call and treatment initiation, while the time periods between injury and 911 call are expected to be estimated with error. In some scenarios, the injury is occurred without witness in suburb and the estimated time might come with larger error. However, based on the results from general settings (Section 3.3), estimating time with large error or using incorrect model M3 M4 would not reduce the power to detect the treatment effect of the ROC TXA trial. Even though my simulations using average absolute error of 10 minutes were relatively conservative, my study on varying error distribution showed

that if my assumptions were appropriate, the effect of measurement was small even when the mean absolute measurement error increased to 1 hour. Because the enrolled patients could usually receive treatment within 1 hour after injury in the ROC TXA trial and we expected the mean absolute measurement error within 20 minutes, absolute measurement error of 1 hour should be considered as a rare case for real scenarios.

Our study on meta-analysis showed that due to large sample size, pooling simulated ROC data with simulated CRASH-3 data not only greatly improved the power to detect both treatment effect and interaction, but also made the power more robust to the measurement error. To perform the one stage IPD meta-analysis, I introduced an indicator variable for ROC study as a fixed effect to M1 and M2 (rewritten as M6 and M7 in Section 2.3.1) to adjust for the difference between two studies. I also fit M6 and M7 with categorized time value (rewritten as M8 and M9 in Section 2.3.1), making our meta-analysis more comparable with the CRASH-2 study and investigating the power when only crude times were available. Instead of a continuous time variable, in M8 and M9, an indicator variable was created for each of three subgroups: 1~2 hours, 2~3 hours and >3 hours; there were also three indicator variables for the interaction term between subgroup and treatment in both M8 and M9. Given sample size of 11000 in the presence of measurement error, the powers of tests using M7 and M9 to detect treatment effect were both improved to almost 1, even in the extreme case that mean absolute equaled to 1 hour (Figure 7A 7C). In this meta-analysis setting, the effect of measurement error on the power of detecting interaction was not obvious when measurement error was in the range we expected for real scenarios. Even in the extreme case that mean absolute measurement error equaled to 1 hour, the power in all the models to detect interaction were still maintain around 80%. Therefore, if we could

obtain the data of CRASH-3 and pool with the ROC TXA trial, we could get sufficient power to detect both treatment effect and interaction, even in the presence of measurement error or only categorized time data available.

According to the association I just found between the distribution of absolute measurement error, crossing time point, strength of treatment effect at time 0 and the potential effect of measurement error, our conclusions about the effect of measurement error in ROC trial and in meta-analysis highly depended on our assumptions for all the simulation parameters. In this study, one of the most important assumptions was that the results from CRASH-2 were representative of our expectation for ROC TXA and CRASH-3, but the objective of CRASH-2 was different from that of ROC TXA and CRASH-3. CRASH-2 investigated the treatment effect of TXA on haemorrhage, while ROC TXA and CRASH were for patients with traumatic brain injury. If that led to the change of treatment effect and interaction strength, our assumptions for simulation parameters might be inappropriate. On the other hand, CRASH-2 data were not accessible at individual patient level and we could only obtain the summary statistics<sup>4</sup>. The ROC TXA trial is not complete and the data from the ROC TXA trial only recorded the time from 911 call to the treatment initiation, so the information we obtained were very limited and the true distribution of time from injury to treatment initiation will never be available in real scenarios. We can see that if the actual crossing time point was in the enrollment time frame and the actual time from injury to treatment initiation distributed more symmetrically at the crossing time point, choosing an incorrect model would have strong effect on ROC study. Therefore, one limitation of this thesis is that the time from injury to treatment initiation and strength of time dependent effect may be inappropriate for simulating the settings of ROC and CRASH-2. That would make our

conclusion inappropriate. When more time data available in the future, we could use the results in this thesis to check if our conclusion on effect of measurement and choosing model still holds for the setting resembling ROC TXA trial.

Another potential limitation might be the assumptions for the distribution of time from injury to treatment initiation of CRASH-3. I used exactly same parameters to simulate the ROC and CRASH-3 for meta-analysis, except for sample size and enrollment time frame. However, different from ROC recruiting patients in Canada and United States, CRASH-3 is conducted in multiple countries around the world. Depending on infrastructure and geography this might lead to faster or slower emergency response times. In addition, the result of CT scan was one of the optional inclusion criteria of CRASH-3 and CT scan usually could only be conducted in hospital setting, while in the ROC TXA trial the emergency medical providers could start the treatment in the field. That meant that many patients enrolled in CRASH-3 might have larger time from injury to treatment initiation compared with the patients in ROC TXA who could have treatment initiated in prehospital setting. If patients in CRASH-3 typically get treatment sooner/later than patients in ROC trial, my simulations for the distribution of time from injury to treatment initiation and the distribution of absolute measurement error of CRASH-3 might not be representative of the studies. The effect of large absolute time measurement error on power of detecting treatment effect and interaction has been discussed above, so if the absolute measurement error of CRASH-3 is much larger than we simulated, the meta-analysis may not have sufficient power to detect treatment effect and interaction.

This study was simulated based on truncated normal distributions. On initial examination, the distribution of simulated estimated times from injury to treatment initiation seemed very different (more skewed and asymmetric) from the distribution of times from 911 call to treatment initiation seen in the ROC TXA trial (Appendix A Figure 1). Nevertheless, the times seen in the ROC TXA trial had a different starting point (911 call rather than time of injury). To make the data comparable I simulated times from injury to 911 call while also considering measurement error to obtain times from injury to treatment initiation for the ROC TXA trial. As a result, the distribution of these (ROC TXA data and simulation based) times from injury to treatment were much closer to the distribution of total time from injury to treatment initiation simulated based on truncated normal distributions (Appendix A Figure 2). Nevertheless, there are other potential options for the distribution of time from injury to treatment initiation or the measurement error, like log normal distribution. Based on a sensitivity analysis using log normal distribution (Appendix B), when I used log normal distribution to simulate the time from injury to treatment initiation instead of truncated normal distribution (Appendix B Figure 3), the effect of measurement error showed similar graphs and trends (Appendix B Figure 4). However, the bias in estimated the treatment effect at time zero and interaction term seemed slightly larger when using log normally distributed true time from injury to treatment initiation. We could investigate whether these results (bias in M1 and M2) were statistically significantly different and use other distributions to simulate time from injury to treatment initiation in the future work.

In addition, we only considered the general linear relationship with logit link between the outcome and treatment and time in this study. The interaction between treatment effect and time from injury to treatment initiation was simulated on logit scale and the

simulated data were fitted with only logistics models. If the real interactive relationship between treatment and initiation time was not on the logit scale, the models used in this study might not be able to detect treatment effect or time dependent effect even in the absence of measurement error. On the other hand, we could also investigate whether other regression models, like the relative risk regression used in CRASH-2<sup>4</sup>, could detect the treatment or interaction effect. In the future, we could study on other potential link functions to generalize our conclusion.

In conclusion, under the scenarios I simulated, the measurement error could reduce power for detecting treatment and interaction effect and increase estimation bias for  $\beta_1$  and  $\gamma$ ; and this impact of measurement error increased as the strength of absolute measurement error increased. If our assumptions are appropriate, an average 0.5 hour absolute measurement error in ROC setting does not impact our analysis results substantially and meta-analysis can still generate sufficient power above 80% to detect both treatment effect and interaction in the presence of an average 1 hour absolute measurement error. In addition, based on our assumptions for ROC TXA trial, using model without considering interactions would not lose power to detect treatment effect compared with test using correct model that includes the interaction term in real analysis.

## References:

1. Simmons, J., Sikorski, R.A., Pittet, J.F., Tranexamic Acid: From Trauma to Routine Perioperative Use. *Current opinion in anaesthesiology*. 28(2):191-200. doi:10.1097 (2015).
2. ClinicalTrials.gov. 2005 May – 2010 Feb. Identifier NCT00375258. Clinical Randomisation of an Antifibrinolytic in Significant Haemorrhage (CRASH-2). Available from:  
<https://clinicaltrials.gov/ct2/show/study/NCT00375258?term=NCT00375258&rank=1>
3. ClinicalTrials.gov. 2011 Sep - . Identifier NCT01402882. Clinical Randomisation of an Antifibrinolytic in Significant Head Injury (CRASH-3). Available from:  
<https://clinicaltrials.gov/ct2/show/NCT01402882?term=NCT01402882&recrs=a&rank=1>
4. Roberts, I., Shakur, H., Coats, T., Hunt, B., Balogun, E., Barnetson, L. et al. The CRASH-2 trial: a randomised controlled trial and economic evaluation of the effects of tranexamic acid on death, vascular occlusive events and transfusion requirement in bleeding trauma patients. *Health Technol Assess* 17(10) (2013).
5. Sawamura, A., Hayakawa, M., Gando, S., Kubota, N., Sugano, M., and WadaTet al. Disseminated intravascular coagulation with a fibrinolytic phenotype at an early phase of trauma predicts mortality. *Thromb Res* 124: 608- 13 (2009).
6. PrenticeCRBasis of antifibrinolytic therapy. *J Clin Pathol Suppl* 14: 35- 40 (1980).
7. ClinicalTrials.gov. 2015 May - . Identifier NCT01990768. Prehospital Tranexamic Acid Use for Traumatic Brain Injury (TXA). Available from:  
<https://clinicaltrials.gov/ct2/show/record/NCT01990768?term=NCT01990768&rank=1>

8. Wilson JTL, Pettigrew LEL, Teasdale GM. Structured interviews for the Glasgow Outcome Scale and the Extended Glasgow Outcome Scale: Guidelines for their use. *Journal of Neurotrauma* 15: 573-585 (1998).
9. Carroll R.J., Ruppert, D. & Stefanski, L.A. *Measurement Error in Nonlinear Models*. Chapman & Hall, London (1995).
10. Fuller, W.A. *Measurement Error Models*. Wiley, New York (1987).
11. <https://cran.r-project.org/web/packages/truncnorm/truncnorm.pdf>
12. Geweke, J., Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints. In *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, Ed. E. Keramidas and S. Kaufman, pp. 571-8. Fairfax Station, VA: Interface Foundation of North America.
13. Dewan Y, Komolafe EO, Mejía-Mantilla JH, Perel P, Roberts I, Shakur H CRASH-3 Collaborators. CRASH-3 - tranexamic acid for the treatment of significant traumatic brain injury: study protocol for an international randomized, double-blind, placebo-controlled trial. *Trials* 13:87 (2012).
14. [http://crash3.lshtm.ac.uk/files/6414/5994/7521/Why\\_reducing\\_time\\_to\\_treatment\\_is\\_critical\\_04Apr16\\_FINAL.pdf](http://crash3.lshtm.ac.uk/files/6414/5994/7521/Why_reducing_time_to_treatment_is_critical_04Apr16_FINAL.pdf)
15. [http://crash3.lshtm.ac.uk/files/6713/7362/3514/CRASH-3 Trial Protocol v1.0 1oct11 FINAL.pdf](http://crash3.lshtm.ac.uk/files/6713/7362/3514/CRASH-3_Trial_Protocol_v1.0_1oct11_FINAL.pdf)
16. <http://crash3.lshtm.ac.uk/>
17. Stewart, L. and Tierney, J., To IPD or Not to IPD? Advantages and Disadvantages of Systematic Reviews Using Individual Patient Data. *Evaluation And the Health Professions*, 25(1):7697, (2002)

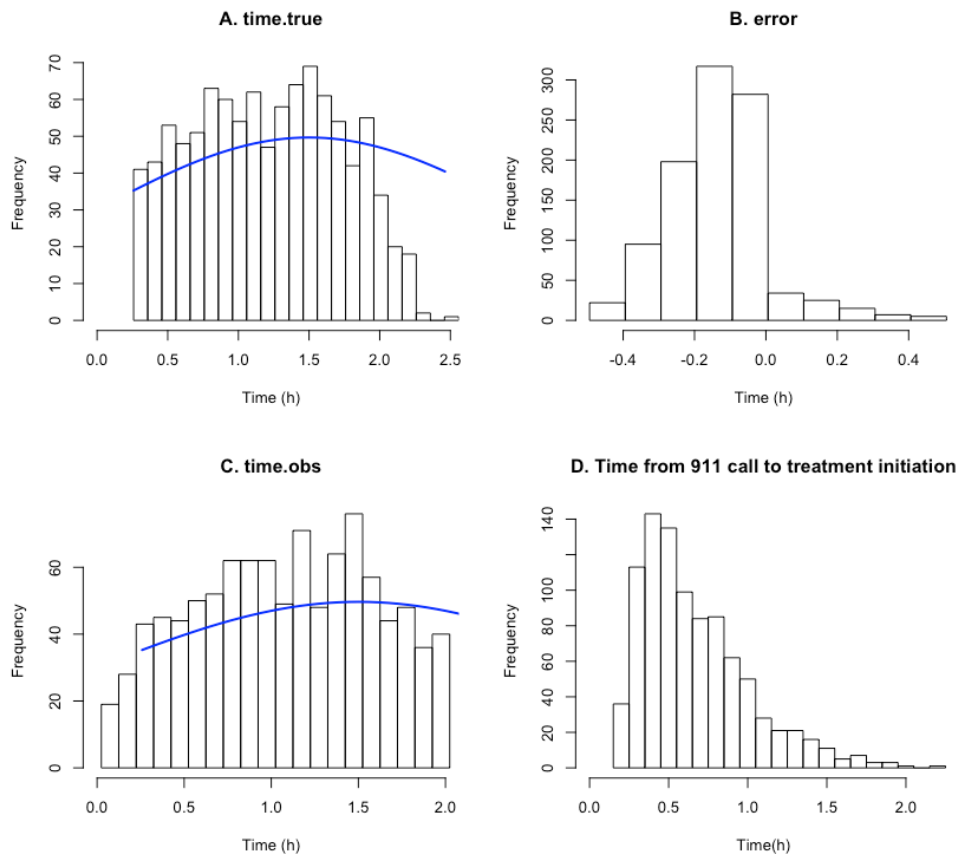
18. Dutton, M., Individual patient-level data meta-analysis: a comparison of methods for the diverse population collaboration data set. Florida State University, Ph.D. Dissertation (2011)
19. Stewart, L. and Parmar, M., Meta-analysis of the literature or of individual patient data: is there a difference? *The Lancet*, 2.5 (1993)
20. DerSimonian, R. and Laird, N., Meta-Analysis in Clinical Trials. *Controlled Clinical Trials*, 7:177{188, 1986. 1, 2.2.2, 2.3, 2.4, 2.4, 3.5.1, 3.5.3, 3.6, 4.1, 4.3.2, 8.1.1.
21. Debray, T.P., Moons, K.G., Abo-Zaid, G.M., Koffijberg, H. and Riley, R.D.. Individual participant data meta-analysis for a binary outcome: one-stage or two-stage? *PloS One* 8(4):e60650 (2013).

## Appendix:

### *A. Justification for Truncated Normal*

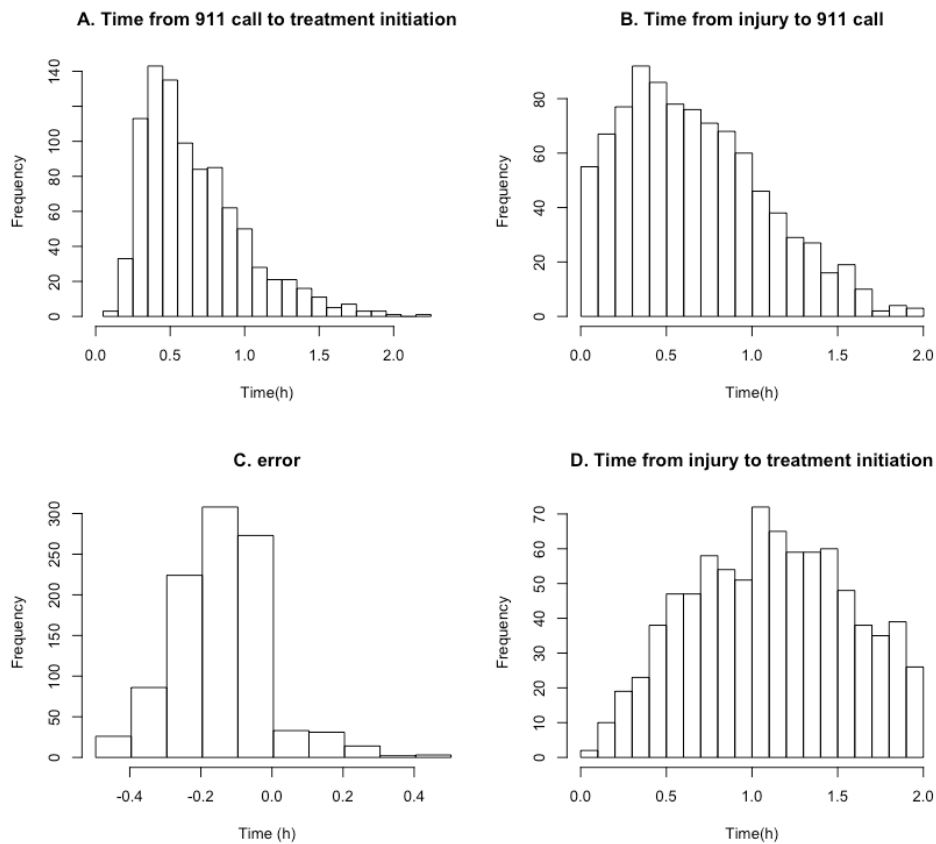
The simulations I performed in this thesis for true initiation time from injury were truncated normally distributed. Figure 1 shows the distributions of true time from injury to treatment initiation (A), measurement error (B) and estimated initiation time (C) under the setting resembling the ROC TXA trial.

**Figure 1. Distributions of truncated normally simulated true time from injury to treatment initiation, simulated measurement error, simulated estimated time and real time from 911 call to treatment initiation collected from ROC study**



The shape of simulated estimated time (Figure 1C) is not similar as the time from 911 call to treatment initiation actually observed in ROC TXA trial (Figure 1D), so I considered the time from injury to 911 call (Figure 2B) and combined it with real time from 911 call to treatment initiation (Figure 1D) and measurement error (Figure 1C), comparing the distribution of total time with the estimated time I simulated (Figure 1B).

**Figure 2. Distributions of real time from 911 call to treatment initiation collected from ROC study, simulated true time from injury to 911 call, simulated measurement error and time from injury to treatment initiation based on real ROC data.**



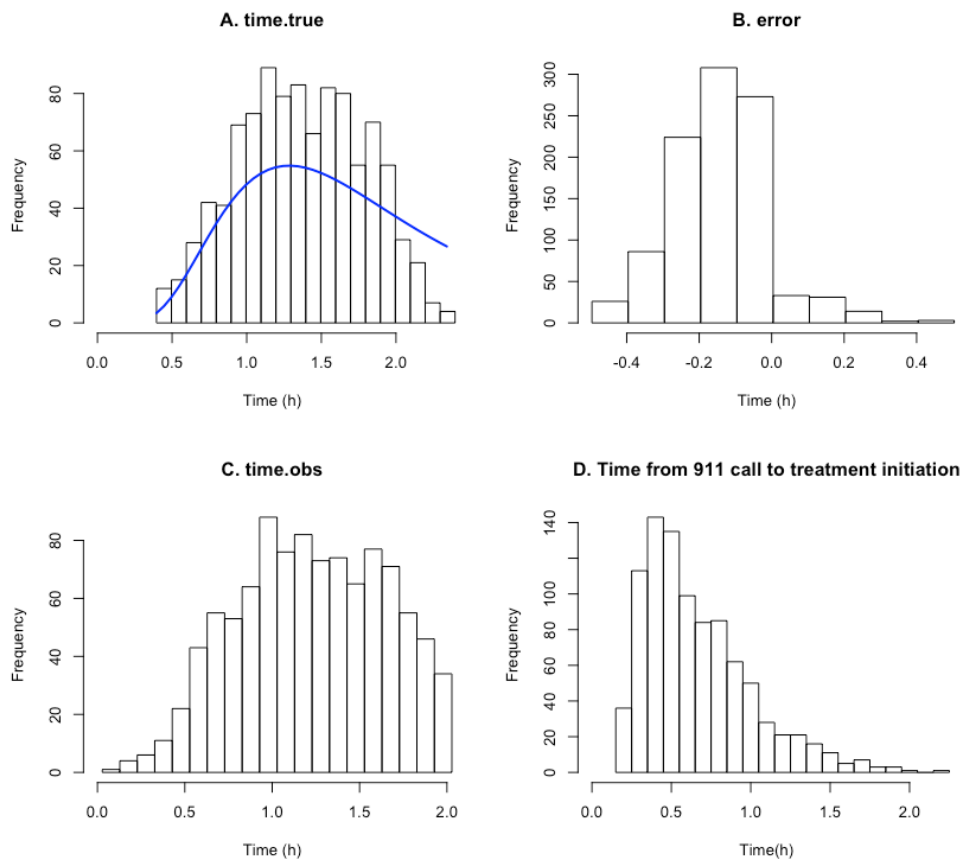
In Figure 2, the distribution of time from 911 call to treatment initiation actually observed in ROC TXA trial (A) and measurement error (D) are identical as Figure 1D and Figure 1D. I simulated the time from injury to 911 call (Figure 2B) with a truncated normal from 0 to 2, mean of 0.5 hour and standard deviation of 0.5 hour (mean and sd before truncation). Since this time period is always positive, I added this time to the time from 911 to treatment from real ROC data (Figure 2A). Since the time from 911 call to treatment is relatively accurate, then I added the measurement error (Figure 2C, same as Figure 1B) to the total time from injury to treatment. Figure 2D only shows the final time distribution within the time frame of 0~2 hours we can observe for the real ROC data. Now observed time I directly simulated with truncation normal distribution

(Figure 1C) is more closed to the observed time we expected from the real ROC data (Figure 2D).

*B. Sensitivity analysis using Log Normal distribution*

To investigate whether my results for the effect of measurement error would change if I used log normal distribution for simulating true initiation time from injury, I repeated the simulations and tests in Section 3.3.3a.

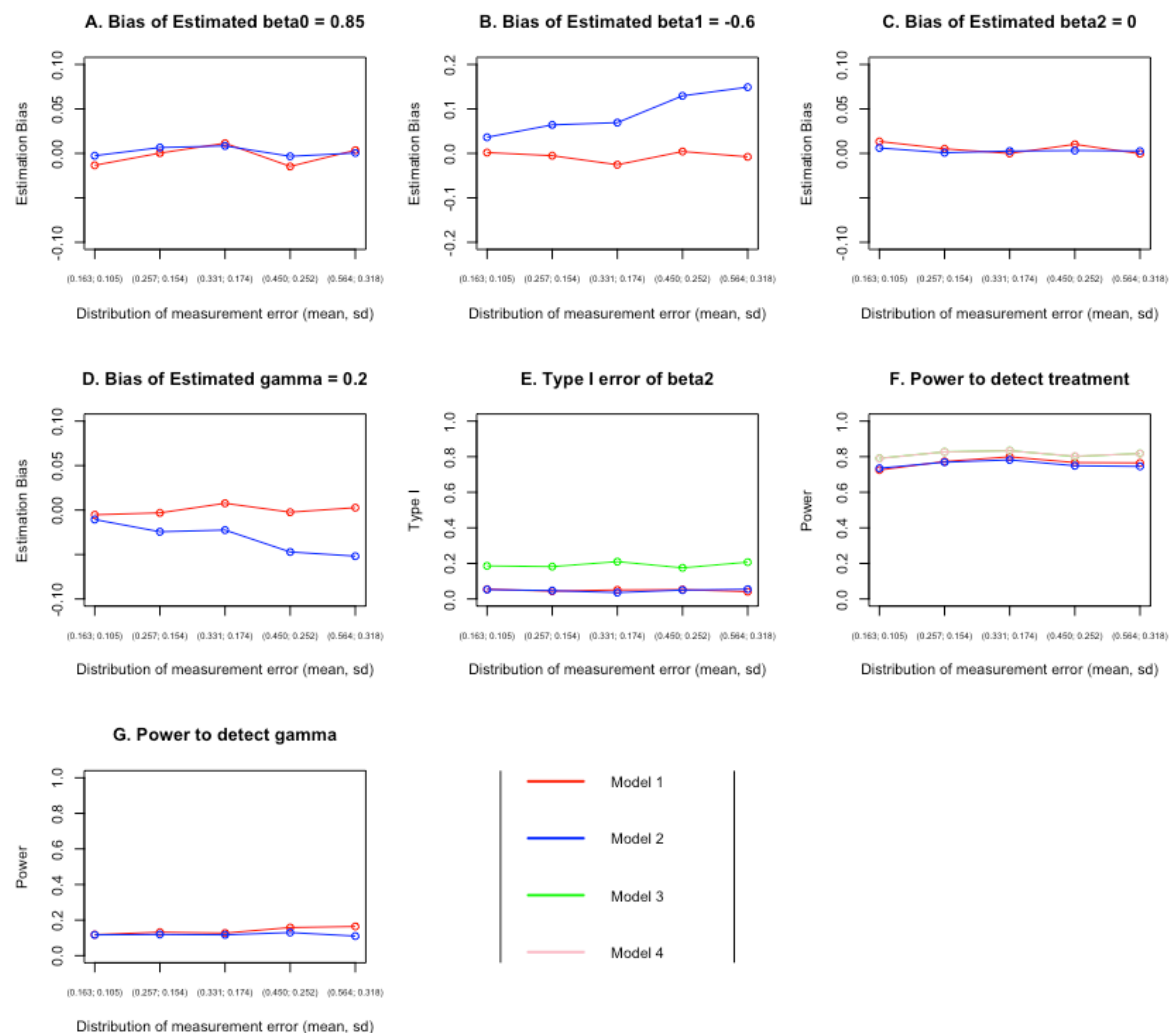
**Figure 3. Distributions of log normally simulated true time from injury to treatment initiation, simulated measurement error, simulated estimated time and real time from 911 call to treatment initiation collected from ROC study**



I simulated the true initiation time from injury with a truncated log normal distribution (still truncated at 15 minutes and 12 hours, mean and standard deviation before truncation were both 0.5 hour) (Figure 3A). To make it more comparable with the truncated normal I used above (Figure 1A), I kept the peak of distribution at similar

position (1~1.5 hours) as the truncated normal. I still used truncated normal for the distribution of absolute measurement error (Figure 3B). After adding the measurement error to the true initiation time from injury, I got the distribution of estimated time from injury to treatment initiation (Figure 3C).

**Figure 4: Estimation bias for all four coefficients in M1 and M2, Type I error for time effect (zero) in M1 M2 and M3 and power to detect treatment effect (all models) and interaction effect in M1 and M2 when the distribution of true initiation time from injury was log normal, the distribution of time measurement error varied and 80% power detecting treatment effect in M1 was maintained. The x-axis is the actual mean and standard deviation of truncated normal distribution after truncation. The scenarios with (mean, sd)= (0.163; 0.105), (0.257; 0.154), (0.331; 0.174), (0.450; 0.252) and (0.564; 0.318) are examined.**

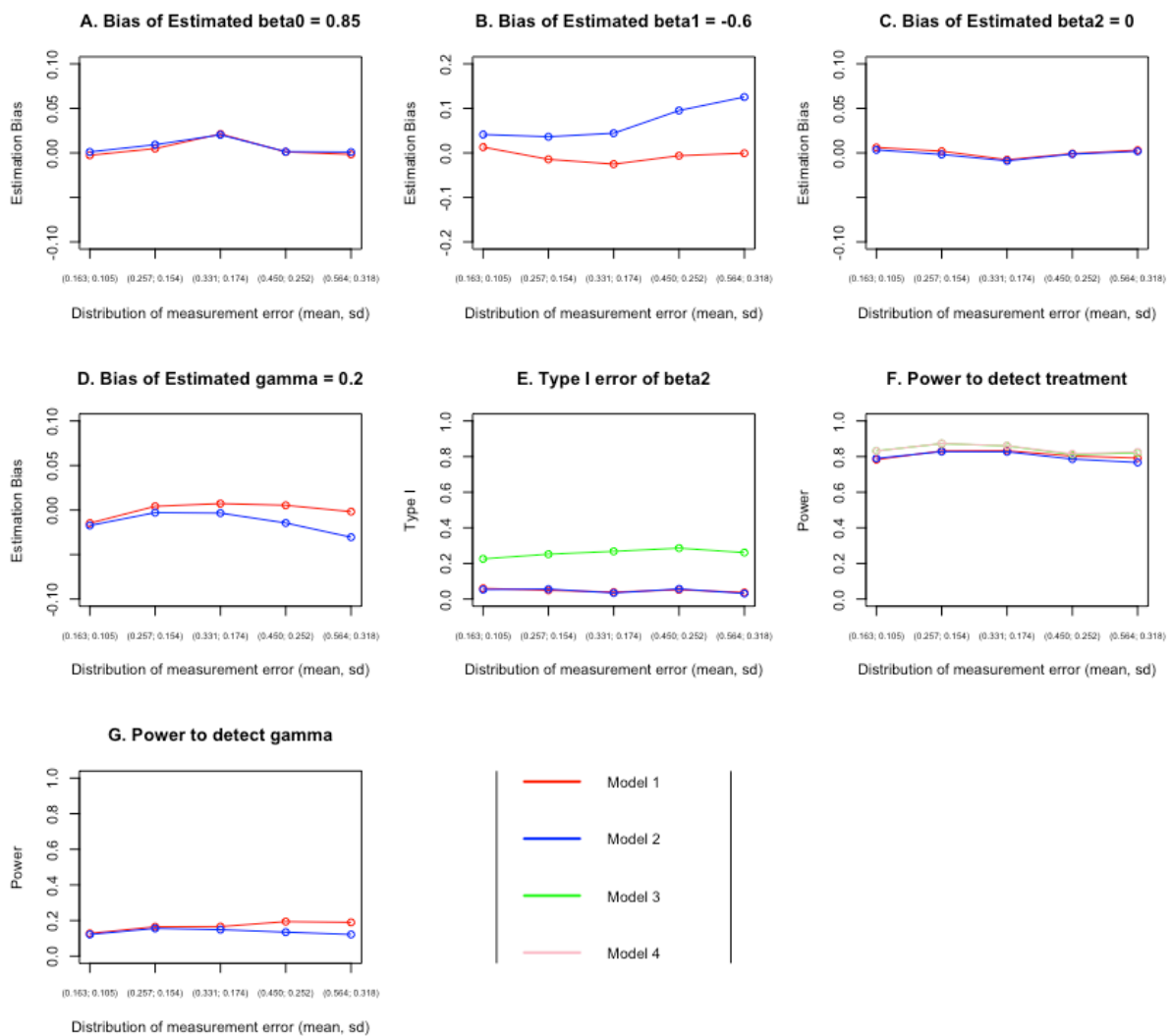


Then I repeated all the tests and comparisons performed in Section 3.3.3a (Figure 4).

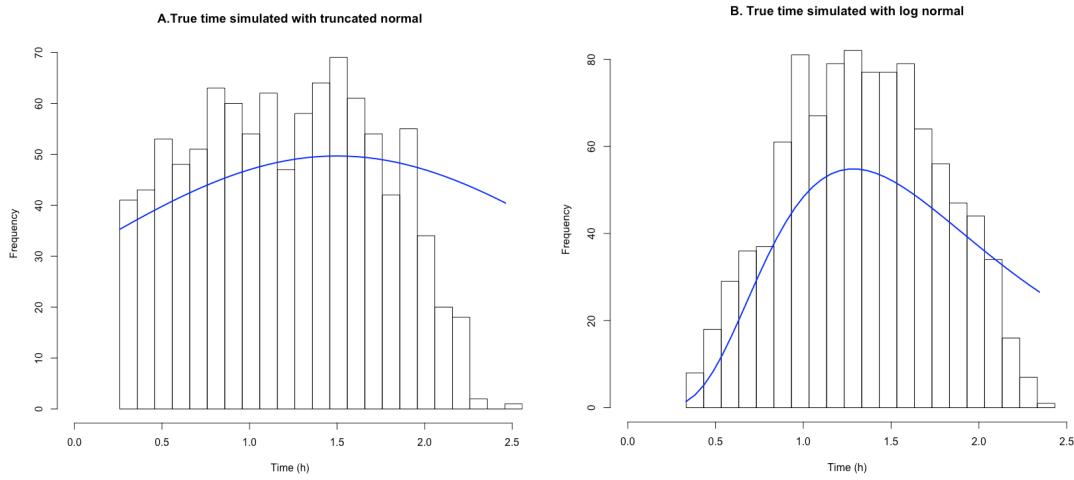
Compared the original results when truncated normal distribution was used for true

initiation time from injury (Figure 5), the results from log normal (Figure 4) showed very similar graphs and trends. However, the bias in estimated treatment effect at time zero  $\beta_1$  and interaction term  $\gamma$  seemed slightly larger when using log normally distributed true time from injury to treatment initiation. We did not investigate whether these results (bias in M1 and M2) were statistically significantly different.

**Figure 5: Estimation bias for all four coefficients in M1 and M2, Type I error for time effect (zero) in M1 M2 and M3 and power to detect treatment effect (all models) and interaction effect in M1 and M2 when the distribution of true initiation time from injury was truncated normal, the distribution of time measurement error varied and 80% power detecting treatment effect in M1 was maintained. The x-axis is the actual mean and standard deviation of truncated normal distribution after truncation. The scenarios with (mean, sd)= (0.163; 0.105), (0.257; 0.154), (0.331; 0.174), (0.450; 0.252) and (0.564; 0.318) are examined.**

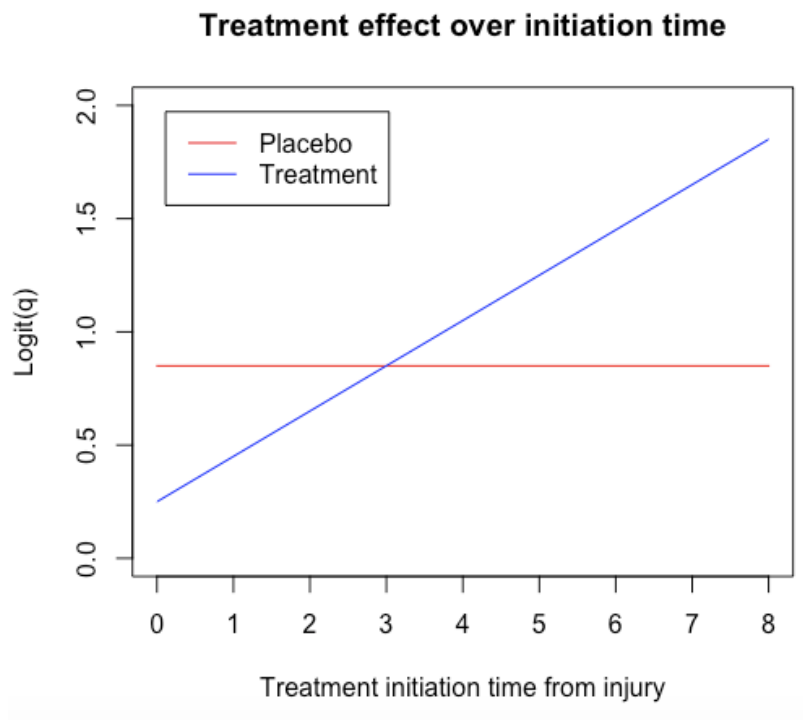


**Figure 5: Distributions of truncated normally simulated true time from injury to treatment initiation and distributions of log normally simulated true time from injury to treatment initiation.**

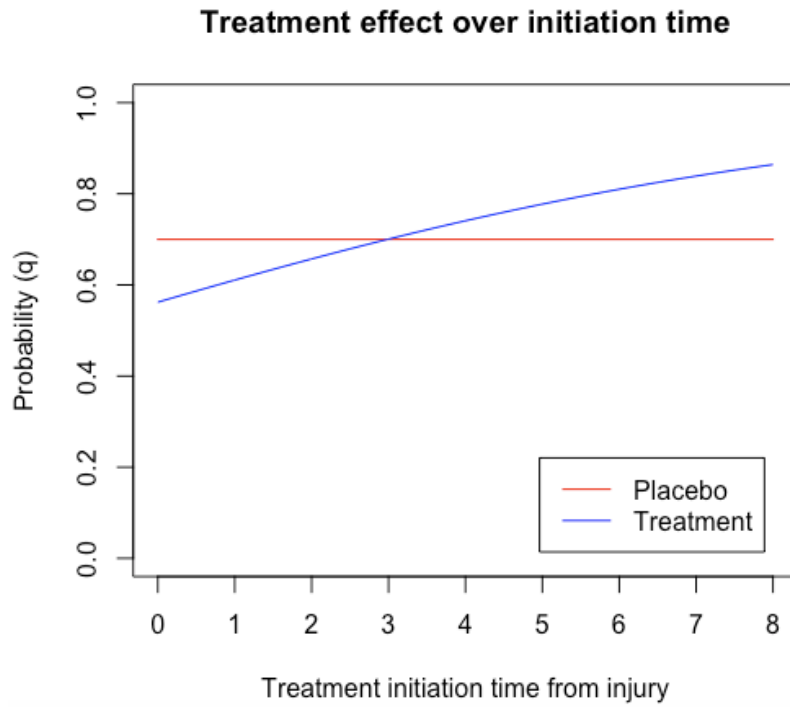


*C. The model of treatment and interaction effect for TXA*

**Figure 6: Log odds (logit(q)) of unfavorable outcome over time from injury to treatment initiation for both placebo and treatment groups**



**Figure 7: Probability (q) of unfavorable outcome over treatment initiation time for both placebo and treatment groups.**



**Table 1. Log odds (logit(q))probability (q) of unfavorable outcome at different time points for both placebo and treatment groups.**

Time point	Log Odds		Probability	
	Placebo	Treatment	Placebo	Treatment
0 hour	0.85	0.25	0.70	0.562
15 minutes	0.85	0.30	0.70	0.574
2 hours	0.85	0.65	0.70	0.657
3 hours	0.85	0.85	0.70	0.70
8 hours	0.85	1.85	0.70	0.864

D. All the models in this study

**Table 2. All the models**

M1	$\text{logit}(q) = \beta_0 + \beta_1 \text{trt} + \beta_2 \text{time.true} + \gamma \text{trt} \times \text{time.true}$
M2	$\text{logit}(q) = \beta_0 + \beta_1 \text{trt} + \beta_2 \text{time.obs} + \gamma \text{trt} \times \text{time.obs}$
M3	$\text{logit}(q) = \beta_0 + \beta_1 \text{trt} + \beta_2 \text{time.obs}$
M4	$\text{logit}(q) = \beta_0 + \beta_1 \text{trt}$
M5	$\text{logit}(q) = \beta_0 + \beta_1 \text{trt} + \beta_2 \text{time.true}$
M6	$\text{logit}(q) = \beta_0 + \beta_1 \text{trt} + \beta_2 \text{time.true} + \gamma \text{trt} \times \text{time.true} + \alpha \times I(\text{study} = \text{ROC})$
M7	$\text{logit}(q) = \beta_0 + \beta_1 \text{trt} + \beta_2 \text{time.true} + \gamma \text{trt} \times \text{time.obs} + \alpha \times I(\text{study} = \text{ROC})$
M8	$\begin{aligned} \text{logit}(q) = & \beta_0 + \beta_1 \text{trt} + \beta_2 \times I(\text{time.true in } 1\sim 2 \text{ hours}) + \beta_3 \times I(\text{time.true in } 2\sim 3 \text{ hours}) \\ & + \beta_4 \times I(\text{time.true} > 3 \text{ hours}) + \gamma_1 \text{trt} \times I(\text{time.true in } 1\sim 2 \text{ hours}) \\ & + \gamma_2 \text{trt} \times I(\text{time.true in } 2\sim 3 \text{ hours}) + \gamma_3 \text{trt} \times I(\text{time.true} \\ & > 3 \text{ hours}) + \alpha \times I(\text{study} = \text{ROC}) \end{aligned}$
M9	$\begin{aligned} \text{logit}(q) = & \beta_0 + \beta_1 \text{trt} + \beta_2 \times I(\text{time.obs in } 1\sim 2 \text{ hours}) + \beta_3 \times I(\text{time.obs in } 2\sim 3 \text{ hours}) \\ & + \beta_4 \times I(\text{time.obs} > 3 \text{ hours}) + \gamma_1 \text{trt} \times I(\text{time.obs in } 1\sim 2 \text{ hours}) \\ & + \gamma_2 \text{trt} \times I(\text{time.obs in } 2\sim 3 \text{ hours}) + \gamma_3 \text{trt} \times I(\text{time.obs} \\ & > 3 \text{ hours}) + \alpha \times I(\text{study} = \text{ROC}) \end{aligned}$

## R Code:

```
library(truncnorm)
library(lmtest)
getTime <- function(n, min, max, time.min, time.max, time.m, time.sd, error.min, error.max,
                    error.m, error.sd, ratio) {
  true <- rtruncnorm(n+5000, a=time.min, b=time.max, mean = time.m, sd = time.sd)
  error <- rtruncnorm(n+5000, a=error.min, b=error.max, mean = error.m, sd = error.sd)
  obs <- rep(0, n+5000)
  for (i in 1:(n+5000)){
    p <- rbinom(1, 1, ratio)
    if (p == 1) {
      error[i] <- -error[i]
    }
    else {
      error[i] <- error[i]
    }
    obs[i] <- true[i] + error[i]
  }
  t <- data.frame(true, error, obs)
  colnames(t) <- c("time.true", "error", "time.obs")
  t <- subset(t, time.obs > min & time.obs < max)
  t <- t[sample(nrow(t), n),]

  par(mfrow=c(2,2))
  h <- hist(t[,1], breaks = seq(min(t[,1]), max(t[,1])+0.1, 0.1), xlim = c(0,max(t[,1])),
           xlab = "Time (h)", main = paste("A.", colnames(t)[1]))
  xfit<-seq(min(t[,1]),max(t[,1]),length=40)
  yfit<-dtruncnorm(xfit,a=min(t[,1]), b=max(t[,1]), mean=time.m,sd=time.m)
  yfit <- yfit*diff(h$mids[1:2])*length(t[,1])
  lines(xfit, yfit, col="blue", lwd=2)
  hist(t[,2], breaks = seq(min(t[,2]), max(t[,2])+0.1, 0.1), xlim = c(min(t[,2]),max(t[,2])),
       xlab = "Time (h)", main = paste("B.", colnames(t)[2]))
  h <- hist(t[,3], breaks = seq(min(t[,3]), max(t[,3])+0.1, 0.1), xlim = c(0,max(t[,3])),
           xlab = "Time (h)", main = paste("C.", colnames(t)[3]))
  xfit<-seq(min(t[,1]),max(t[,1]),length=40)
  yfit<-dtruncnorm(xfit,a= min(t[,1]), b= max(t[,1]), mean=time.m,sd=time.sd)
  yfit <- yfit*diff(h$mids[1:2])*length(t[,1])
  lines(xfit, yfit, col="blue", lwd=2)
```

```

hist(callToTrt, breaks = seq(min(callToTrt), max(callToTrt)+0.1, 0.1), xlim = c(0,max(callToTrt)), xlab =
"Time(h)",
  main = "D. Time from 911 call to treatment initiation")
return(t)
}

```

```

getTime2 <- function(n, min, max, time.min, time.max, time.m, time.sd, error.min, error.max, error.m,
  error.sd, ratio) {
true <- rlnorm(n+5000, mean = time.m, sd = time.sd)
true <- true[time.min < true & true<time.max]
error <- rtruncnorm(length(true), a=error.min, b=error.max, mean = error.m, sd = error.sd)

obs <- rep(0, length(true))
for (i in 1:(length(true))){
  p <- rbinom(1, 1, ratio)
  if (p == 1) {
    error[i] <- -error[i]
  }
  else {
    error[i] <- error[i]
  }
  obs[i] <- true[i] + error[i]
}
t <- data.frame(true, error, obs)
colnames(t) <- c("time.true", "error", "time.obs")
t <- subset(t, time.obs > min & time.obs < max)
t <- t[sample(nrow(t), n),]

```

```

par(mfrow=c(2,2))
h <- hist(t[,1], breaks = seq(min(t[,1]), max(t[,1])+0.1, 0.1), xlim = c(0,max(t[,1])),
  xlab = "Time (h)", main = paste("A.", colnames(t)[1]))
xfit<-seq(min(t[,1]),max(t[,1]),length=40)
yfit<-dlnorm(xfit, meanlog = time.m, sdlog = time.sd)
yfit <- yfit*diff(h$mids[1:2])*length(t[,1])
lines(xfit, yfit, col="blue", lwd=2)
hist(t[,2], breaks = seq(min(t[,2]), max(t[,2])+0.1, 0.1), xlim = c(min(t[,2]),max(t[,2])),
  xlab = "Time (h)", main = paste("B.", colnames(t)[2]))
hist(t[,3], breaks = seq(min(t[,3]), max(t[,3])+0.1, 0.1), xlim = c(0,max(t[,3])),
  xlab = "Time (h)", main = paste("C.", colnames(t)[3]))

```

```

hist(callToTrt, breaks = seq(min(callToTrt), max(callToTrt)+0.1, 0.1), xlim = c(0,max(callToTrt)), xlab =
"Time(h)",
  main = "D. Time from 911 call to treatment initiation")
return(t)
}

```

```

getData <- function(n, timeData, log.odd0, log.or1, logit.increase1) {
  a1 <- sample(1:n, size=round(n/3)) # randomly choose placebo group
  a2 <- setdiff(1:n, a1) # randomly choose treatment 1 group
  x <- cbind(rep(0, n), rep(0, n))

  for (i in a1) {x[i, 1] <- 1} # assign treatment to each Xi in placebo arm
  for (i in a2) {x[i, 2] <- 1} # assign treatment to each Xi in treatment arm
  colnames(x) <- c("plb", "trt")

  p <- rep(0,n)
  y <- rep(0,n) # assign 0 to all y

  for (i in 1:n) {
    z <- log.odd0 + log.or1*x[i, 2] + logit.increase1*timeData[i, 1]*x[i, 2]
    p[i] <- exp(z)/(1+exp(z))
    y[i] <- rbinom(1, 1, p[i]) # reassign 1 to y for placebo group with probability p0
  }
  mydata <- data.frame(y, p, x, timeData)
  return(mydata)
}

```

```

testModel <- function(n, simulation, timeData, log.odd0, log.or1, logit.increase1) {
  coef.true <- rep(0,4)
  coef.obs1 <- rep(0,4)
  coef.obs2 <- rep(0,3)
  coef.obs3 <- rep(0,2)

  typeItrt.true <- 0
  typeItrt.obs1 <- 0
  typeItrt.obs2 <- 0
  typeItrt.obs3 <- 0

  typeI.true <- 0

```

```

typeI.obs1 <- 0
typeI.obs2 <- 0

typeII.true <- 0
typeII.obs1 <- 0

for (i in 1:simulation){
  set.seed(i)
  mydata <- getData(n, timeData, log.odd0, log.or1, logit.increase1)

  fit.true <- glm(y~trt+time.true+trt:time.true, family = binomial, data = mydata)
  fit.trueRed <- glm(y~time.true, family = binomial, data = mydata)
  fit.obs1 <- glm(y~trt+time.obs+trt:time.obs, family = binomial, data = mydata)
  fit.obs1Red <- glm(y~time.obs, family = binomial, data = mydata)
  fit.obs2 <- glm(y~trt+time.obs, family = binomial, data = mydata)
  fit.obs2Red <- glm(y~time.obs, family = binomial, data = mydata)
  fit.obs3 <- glm(y~trt, family = binomial, data = mydata)
  fit.obs3Red <- glm(y~1, family = binomial, data = mydata)

  for (j in 1:4) {
    coef.true[j] <- coef.true[j]+fit.true$coefficients[j]
    coef.obs1[j] <- coef.obs1[j]+fit.obs1$coefficients[j]
  }

  coef.obs2[1] <- coef.obs2[1]+fit.obs2$coefficients[1]
  coef.obs2[2] <- coef.obs2[2]+fit.obs2$coefficients[2]
  coef.obs2[3] <- coef.obs2[3]+fit.obs2$coefficients[3]

  coef.obs3[1] <- coef.obs3[1]+fit.obs3$coefficients[1]
  coef.obs3[2] <- coef.obs3[2]+fit.obs3$coefficients[2]

  if(lrtest(fit.true,fit.trueRed)[2,5] > 0.05) { typeIItrt.true <- typeIItrt.true+1 }
  if(lrtest(fit.obs1,fit.obs1Red)[2,5] > 0.05) { typeIItrt.obs1 <- typeIItrt.obs1+1 }
  if(lrtest(fit.obs2,fit.obs2Red)[2,5] > 0.05) { typeIItrt.obs2 <- typeIItrt.obs2+1 }
  if(lrtest(fit.obs3,fit.obs3Red)[2,5] > 0.05) { typeIItrt.obs3 <- typeIItrt.obs3+1 }

  if(summary(fit.true)$coefficients[3,4] < 0.05) { typeI.true <- typeI.true+1 }
  if(summary(fit.obs1)$coefficients[3,4] < 0.05) { typeI.obs1 <- typeI.obs1+1 }
  if(summary(fit.obs2)$coefficients[3,4] < 0.05) { typeI.obs2 <- typeI.obs2+1 }
}

```

```

if(summary(fit.true)$coefficients[4,4] > 0.05) { typeIInt.true <- typeIInt.true+1 }
if(summary(fit.obs1)$coefficients[4,4] > 0.05) { typeIInt.obs1 <- typeIInt.obs1+1 }
}
return(list(coef.true = coef.true/simulation, coef.obs1 = coef.obs1/simulation,
           coef.obs2 = coef.obs2/simulation, coef.obs3 = coef.obs3/simulation,
           powerTrt.true = 1 - typeIInt.true/simulation, powerTrt.obs1 = 1 - typeIInt.obs1/simulation,
           powerTrt.obs2 = 1 - typeIInt.obs2/simulation, powerTrt.obs3 = 1 - typeIInt.obs3/simulation,
           typeI.true = typeI.true/simulation, typeI.obs1 = typeI.obs1/simulation,
           typeI.obs2 = typeI.obs2/simulation,
           powerInt.true = 1 - typeIInt.true/simulation, powerInt.obs1 = 1 - typeIInt.obs1/simulation))
}

#### In CRASH 2 setting
set.seed(1)
t <- getTime(1000, 0, 2, 0.25, 12, 1.5, 1.5, 0, 0.5, 0.1, 0.15, 0.9)
testModel(1000, 1000, t, 0.85, -0.6, 0.6/3)

# Ratio (0.9 --> 0.1)
set.seed(1)
t <- getTime(1400, 0, 2, 0.25, 12, 1.5, 1.5, 0, 0.5, 0.1, 0.15, 0.9)
r9 <- testModel(1400, 1000, t, 0.85, -0.6, 0.6/3)
set.seed(1)
t <- getTime(1400, 0, 2, 0.25, 12, 1.5, 1.5, 0, 0.5, 0.1, 0.15, 0.7)
r7 <- testModel(1400, 1000, t, 0.85, -0.6, 0.6/3)
set.seed(1)
t <- getTime(1400, 0, 2, 0.25, 12, 1.5, 1.5, 0, 0.5, 0.1, 0.15, 0.5)
r5 <- testModel(1400, 1000, t, 0.85, -0.6, 0.6/3)
set.seed(1)
t <- getTime(1400, 0, 2, 0.25, 12, 1.5, 1.5, 0, 0.5, 0.1, 0.15, 0.3)
r3 <- testModel(1400, 1000, t, 0.85, -0.6, 0.6/3)
set.seed(1)
t <- getTime(1400, 0, 2, 0.25, 12, 1.5, 1.5, 0, 0.5, 0.1, 0.15, 0.1)
r1 <- testModel(1400, 1000, t, 0.85, -0.6, 0.6/3)

par(mfrow=c(3,3))
b0.true <- c(r9$coef.true[1], r7$coef.true[1], r5$coef.true[1], r3$coef.true[1], r1$coef.true[1])
b0.obs1 <- c(r9$coef.obs1[1], r7$coef.obs1[1], r5$coef.obs1[1], r3$coef.obs1[1], r1$coef.obs1[1])
b0.obs2 <- c(r9$coef.obs2[1], r7$coef.obs2[1], r5$coef.obs2[1], r3$coef.obs2[1], r1$coef.obs2[1])
b0.obs3 <- c(r9$coef.obs3[1], r7$coef.obs3[1], r5$coef.obs3[1], r3$coef.obs3[1], r1$coef.obs3[1])
plot(c(1:5), b0.true-0.85, type = "o", col = "red", ylab = "Estimation Bias", xaxt = 'n',

```

```
ylim = c(-0.1, 0.1), xlab = "Proportion of underestimating treatment initiation time", main = "A. Bias of  
Estimated beta0 = 0.85")
```

```
lines(c(1:5), b0.obs1-0.85, type = "o", col = "blue")
```

```
axis(side=1,at=c(1:5),labels=c("0.9", "0.7", "0.5", "0.3", "0.1"), cex.axis = 0.8)
```

```
b1.true <- c(r9$coef.true[2], r7$coef.true[2], r5$coef.true[2], r3$coef.true[2], r1$coef.true[2])
```

```
b1.obs1 <- c(r9$coef.obs1[2], r7$coef.obs1[2], r5$coef.obs1[2], r3$coef.obs1[2], r1$coef.obs1[2])
```

```
b1.obs2 <- c(r9$coef.obs2[2], r7$coef.obs2[2], r5$coef.obs2[2], r3$coef.obs2[2], r1$coef.obs2[2])
```

```
b1.obs3 <- c(r9$coef.obs3[2], r7$coef.obs3[2], r5$coef.obs3[2], r3$coef.obs3[2], r1$coef.obs3[2])
```

```
plot(c(1:5),b1.true+0.6, type = "o",col = "red",ylab = "Estimation Bias", xaxt = 'n',
```

```
ylim = c(-0.1, 0.1), xlab = "Proportion of underestimating treatment initiation time",main = "B. Bias of  
Estimated beta1 = -0.6")
```

```
lines(c(1:5), b1.obs1+0.6, type = "o", col = "blue")
```

```
axis(side=1,at=c(1:5),labels=c("0.9", "0.7", "0.5", "0.3", "0.1"), cex.axis = 0.8)
```

```
b2.true <- c(r9$coef.true[3], r7$coef.true[3], r5$coef.true[3], r3$coef.true[3], r1$coef.true[3])
```

```
b2.obs1 <- c(r9$coef.obs1[3], r7$coef.obs1[3], r5$coef.obs1[3], r3$coef.obs1[3], r1$coef.obs1[3])
```

```
b2.obs2 <- c(r9$coef.obs2[3], r7$coef.obs2[3], r5$coef.obs2[3], r3$coef.obs2[3], r1$coef.obs2[3])
```

```
plot(c(1:5), b2.true, type = "o",col = "red", ylab = "Estimation Bias", xaxt = 'n',
```

```
ylim = c(-0.1, 0.1), xlab = "Proportion of underestimating treatment initiation time",main = "C. Bias of  
Estimated beta2 = 0")
```

```
lines(c(1:5), b2.obs1, type = "o", col = "blue")
```

```
axis(side=1,at=c(1:5),labels=c("0.9", "0.7", "0.5", "0.3", "0.1"), cex.axis = 0.8)
```

```
b3.true <- c(r9$coef.true[4], r7$coef.true[4], r5$coef.true[4], r3$coef.true[4], r1$coef.true[4])
```

```
b3.obs1 <- c(r9$coef.obs1[4], r7$coef.obs1[4], r5$coef.obs1[4], r3$coef.obs1[4], r1$coef.obs1[4])
```

```
plot(c(1:5), b3.true-0.2, type = "o",col = "red", ylab = "Estimation Bias", xaxt = 'n',
```

```
ylim = c(-0.1, 0.1), xlab = "Proportion of underestimating treatment initiation time",main = "D. Bias of  
Estimated gamma = 0.2")
```

```
lines(c(1:5), b3.obs1-0.2, type = "o", col = "blue")
```

```
axis(side=1,at=c(1:5),labels=c("0.9", "0.7", "0.5", "0.3", "0.1"), cex.axis = 0.8)
```

```
typeI.true <- c(r9$typeI.true, r7$typeI.true, r5$typeI.true, r3$typeI.true, r1$typeI.true)
```

```
typeI.obs1 <- c(r9$typeI.obs1, r7$typeI.obs1, r5$typeI.obs1, r3$typeI.obs1, r1$typeI.obs1)
```

```
typeI.obs2 <- c(r9$typeI.obs2, r7$typeI.obs2, r5$typeI.obs2, r3$typeI.obs2, r1$typeI.obs2)
```

```
plot(c(1:5), typeI.true, type = "o",col = "red", ylab = "Type I", ylim = c(0, 1), xaxt = 'n',
```

```
xlab = "Proportion of underestimating treatment initiation time", main = "E. Type I error of beta2")
```

```
lines(c(1:5), typeI.obs1, type = "o", col = "blue")
```

```

lines(c(1:5), type=l.obs2, type = "o", col = "green")
axis(side=1,at=c(1:5),labels=c("0.9", "0.7", "0.5", "0.3", "0.1"), cex.axis = 0.8)

powers.true <- c(r9$powerTrt.true, r7$powerTrt.true, r5$powerTrt.true, r3$powerTrt.true,
r1$powerTrt.true)
powers.obs1 <- c(r9$powerTrt.obs1, r7$powerTrt.obs1, r5$powerTrt.obs1, r3$powerTrt.obs1,
r1$powerTrt.obs1)
powers.obs2 <- c(r9$powerTrt.obs2, r7$powerTrt.obs2, r5$powerTrt.obs2, r3$powerTrt.obs2,
r1$powerTrt.obs2)
powers.obs3 <- c(r9$powerTrt.obs3, r7$powerTrt.obs3, r5$powerTrt.obs3, r3$powerTrt.obs3,
r1$powerTrt.obs3)
plot(c(1:5), powers.true, type = "o",col = "red", ylab = "Power", xaxt = 'n',
ylim = c(0, 1), xlab = "Proportion of underestimating treatment initiation time", main = "F. Power to
detect treatment")
lines(c(1:5), powers.obs1, type = "o", col = "blue")
lines(c(1:5), powers.obs2, type = "o", col = "green")
lines(c(1:5), powers.obs3, type = "o", col = "pink")
axis(side=1,at=c(1:5),labels=c("0.9", "0.7", "0.5", "0.3", "0.1"), cex.axis = 0.8)

powersInt.true <- c(r9$powerInt.true, r7$powerInt.true, r5$powerInt.true, r3$powerInt.true,
r1$powerInt.true)
powersInt.obs1 <- c(r9$powerInt.obs1, r7$powerInt.obs1, r5$powerInt.obs1, r3$powerInt.obs1,
r1$powerInt.obs1)
plot(c(1:5), powersInt.true, type = "o",col = "red", ylab = "Power", ylim = c(0, 1), xaxt = 'n',
xlab = "Proportion of underestimating treatment initiation time", main = "G. Power to detect gamma")
lines(c(1:5), powersInt.obs1, type = "o", col = "blue")
axis(side=1,at=c(1:5),labels=c("0.9", "0.7", "0.5", "0.3", "0.1"), cex.axis = 0.8)

plot(1, type="n", axes=FALSE, xlab="", ylab="")
legend(1, 1, legend = c("Model 1", "Model 2", "Model 3", "Model 4"), col=c("red", "blue", "green", "pink"),
lwd=2, cex=1, xjust=0.5, yjust=0.5)

# Distribution of initiation time
set.seed(1)
t <- getTime(800, 0, 2, 0.25, 12, 0.5, 0.3, 0, 0.5, 0.1, 0.15, 0.9)
i1 <- testModel(800, 1000, t, 0.85, -0.6, 0.6/3)
set.seed(1)
t <- getTime(1300, 0, 2, 0.25, 12, 1, 1, 0, 0.5, 0.1, 0.15, 0.9)
i2 <- testModel(1300, 1000, t, 0.85, -0.6, 0.6/3)
set.seed(1)

```

```

t <- getTime(1400, 0, 2, 0.25, 12, 1.5, 1.5, 0, 0.5, 0.1, 0.15, 0.9)
i3 <- testModel(1400, 1000, t, 0.85, -0.6, 0.6/3)
set.seed(1)
t <- getTime(1400, 0, 2, 0.25, 12, 2, 1.5, 0, 0.5, 0.1, 0.15, 0.9)
i4 <- testModel(1400, 1000, t, 0.85, -0.6, 0.6/3)
set.seed(1)
t <- getTime(1500, 0, 2, 0.25, 12, 2.5, 1.5, 0, 0.5, 0.1, 0.15, 0.9)
i5 <- testModel(1500, 1000, t, 0.85, -0.6, 0.6/3)

par(mfrow=c(3,3))
b0.true <- c(i1$coef.true[1], i2$coef.true[1], i3$coef.true[1], i4$coef.true[1], i5$coef.true[1])
b0.obs1 <- c(i1$coef.obs1[1], i2$coef.obs1[1], i3$coef.obs1[1], i4$coef.obs1[1], i5$coef.obs1[1])
b0.obs2 <- c(i1$coef.obs2[1], i2$coef.obs2[1], i3$coef.obs2[1], i4$coef.obs2[1], i5$coef.obs2[1])
b0.obs3 <- c(i1$coef.obs3[1], i2$coef.obs3[1], i3$coef.obs3[1], i4$coef.obs3[1], i5$coef.obs3[1])
plot(c(1:5), b0.true-0.85, type = "o", col = "red", xlab = "Distribution of treatment initiation time (mean,
sd)",
      ylab = "Estimation Bias", ylim = c(-0.1, 0.1), main = "A. Bias of Estimated beta0 = 0.8", xaxt = 'n')
lines(c(1:5), b0.obs1-0.85, type = "o", col = "blue")
axis(side=1, at=c(1:5), labels=c("(0.525; 0.238)", "(1.14; 0.509)", "(1.23; 0.514)", "(1.28; 0.524)", "(1.37;
0.521)"), cex.axis = 0.6)

b1.true <- c(i1$coef.true[2], i2$coef.true[2], i3$coef.true[2], i4$coef.true[2], i5$coef.true[2])
b1.obs1 <- c(i1$coef.obs1[2], i2$coef.obs1[2], i3$coef.obs1[2], i4$coef.obs1[2], i5$coef.obs1[2])
b1.obs2 <- c(i1$coef.obs2[2], i2$coef.obs2[2], i3$coef.obs2[2], i4$coef.obs2[2], i5$coef.obs2[2])
b1.obs3 <- c(i1$coef.obs3[2], i2$coef.obs3[2], i3$coef.obs3[2], i4$coef.obs3[2], i5$coef.obs3[2])
plot(c(1:5), b1.true+0.6, type = "o", col = "red", xlab = "Distribution of treatment initiation time (mean,
sd)",
      ylab = "Estimation Bias", ylim = c(-0.1, 0.1), main = "B. Bias of Estimated beta1 = -0.6", xaxt = 'n')
lines(c(1:5), b1.obs1+0.6, type = "o", col = "blue")
axis(side=1, at=c(1:5), labels=c("(0.525; 0.238)", "(1.14; 0.509)", "(1.23; 0.514)", "(1.28; 0.524)", "(1.37;
0.521)"), cex.axis = 0.6)

b2.true <- c(i1$coef.true[3], i2$coef.true[3], i3$coef.true[3], i4$coef.true[3], i5$coef.true[3])
b2.obs1 <- c(i1$coef.obs1[3], i2$coef.obs1[3], i3$coef.obs1[3], i4$coef.obs1[3], i5$coef.obs1[3])
b2.obs2 <- c(i1$coef.obs2[3], i2$coef.obs2[3], i3$coef.obs2[3], i4$coef.obs2[3], i5$coef.obs2[3])
plot(c(1:5), b2.true, type = "o", col = "red", xlab = "Distribution of treatment initiation time (mean, sd)",
      ylab = "Estimation Bias", ylim = c(-0.1, 0.1), main = "C. Bias of Estimated beta2 = 0", xaxt = 'n')
lines(c(1:5), b2.obs1, type = "o", col = "blue")
axis(side=1, at=c(1:5), labels=c("(0.525; 0.238)", "(1.14; 0.509)", "(1.23; 0.514)", "(1.28; 0.524)", "(1.37;
0.521)"), cex.axis = 0.6)

```

```

b3.true <- c(i1$coef.true[4], i2$coef.true[4], i3$coef.true[4], i4$coef.true[4], i5$coef.true[4])
b3.obs1 <- c(i1$coef.obs1[4], i2$coef.obs1[4], i3$coef.obs1[4], i4$coef.obs1[4], i5$coef.obs1[4])
plot(c(1:5), b3.true-0.6/3, type = "o", col = "red", xlab = "Distribution of treatment initiation time (mean,
sd)",
  ylab = "Estimation Bias", ylim = c(-0.1, 0.1), main = "D. Bias of Estimated gamma = -0.2", xaxt = 'n')
lines(c(1:5), b3.obs1-0.6/3, type = "o", col = "blue")
axis(side=1, at=c(1:5), labels=c("(0.525; 0.238)", "(1.14; 0.509)", "(1.23; 0.514)", "(1.28; 0.524)", "(1.37;
0.521)"), cex.axis = 0.6)

type1.true <- c(i1$type1.true, i2$type1.true, i3$type1.true, i4$type1.true, i4$type1.true)
type1.obs1 <- c(i1$type1.obs1, i2$type1.obs1, i3$type1.obs1, i4$type1.obs1, i4$type1.obs1)
type1.obs2 <- c(i1$type1.obs2, i2$type1.obs2, i3$type1.obs2, i4$type1.obs2, i4$type1.obs2)
plot(c(1:5), type1.true, type = "o", col = "red", xlab = "Distribution of treatment initiation time (mean, sd)",
  ylab = "Type I", ylim = c(0, 1), main = "E. Type I error of beta2", xaxt = 'n')
lines(c(1:5), type1.obs1, type = "o", col = "blue")
lines(c(1:5), type1.obs2, type = "o", col = "green")
axis(side=1, at=c(1:5), labels=c("(0.525; 0.238)", "(1.14; 0.509)", "(1.23; 0.514)", "(1.28; 0.524)", "(1.37;
0.521)"), cex.axis = 0.6)

powers.true <- c(i1$powerTrt.true, i2$powerTrt.true, i3$powerTrt.true, i4$powerTrt.true,
i5$powerTrt.true)
powers.obs1 <- c(i1$powerTrt.obs1, i2$powerTrt.obs1, i3$powerTrt.obs1, i4$powerTrt.obs1,
i5$powerTrt.obs1)
powers.obs2 <- c(i1$powerTrt.obs2, i2$powerTrt.obs2, i3$powerTrt.obs2, i4$powerTrt.obs2,
i5$powerTrt.obs2)
powers.obs3 <- c(i1$powerTrt.obs3, i2$powerTrt.obs3, i3$powerTrt.obs3, i4$powerTrt.obs3,
i5$powerTrt.obs3)
plot(c(1:5), powers.true, type = "o", col = "red", xlab = "Distribution of treatment initiation time (mean,
sd)",
  ylab = "Power", ylim = c(0, 1), main = "F. Power to detect treatment", xaxt = 'n')
lines(c(1:5), powers.obs1, type = "o", col = "blue")
lines(c(1:5), powers.obs2, type = "o", col = "green")
lines(c(1:5), powers.obs3, type = "o", col = "pink")
axis(side=1, at=c(1:5), labels=c("(0.525; 0.238)", "(1.14; 0.509)", "(1.23; 0.514)", "(1.28; 0.524)", "(1.37;
0.521)"), cex.axis = 0.6)

powersInt.true <- c(i1$powerInt.true, i2$powerInt.true, i3$powerInt.true, i4$powerInt.true,
i5$powerInt.true)

```

```

powersInt.obs1 <- c(i1$powerInt.obs1, i2$powerInt.obs1, i3$powerInt.obs1, i4$powerInt.obs1,
i5$powerInt.obs1)
plot(c(1:5), powersInt.true, type = "o", col = "red", xlab = "Distribution of treatment initiation time (mean,
sd)",
  ylab = "Power", ylim = c(0, 1), main = "G. Power to detect gamma", xaxt = 'n')
lines(c(1:5), powersInt.obs1, type = "o", col = "blue")
axis(side=1, at=c(1:5), labels=c("(0.51;0.15)", "(0.75;0.21)", "(1;0.28)", "(1.3;0.34)", "(1.4;0.33)"), cex.axis =
0.7)

plot(1, type="n", axes=FALSE, xlab="", ylab="")
legend(1, 1, legend = c("Model 1", "Model 2", "Model 3", "Model 4"), col=c("red", "blue", "green", "pink"),
  lwd=2, cex=1, xjust=0.5, yjust=0.5)

# Distribution of error (range)
set.seed(1)
t <- getTime(1400, 0, 2, 0.25, 12, 1.5, 1.5, 0, 0.5, 0.1, 0.15, 0.9)
e3 <- testModel(1400, 1000, t, 0.85, -0.6, 0.6/3)
set.seed(1)
t <- getTime(1500, 0, 2, 0.25, 12, 1.5, 1.5, 0, 0.7, 0.2, 0.2, 0.9)
e5 <- testModel(1500, 1000, t, 0.85, -0.6, 0.6/3)
set.seed(1)
t <- getTime(1600, 0, 2, 0.25, 12, 1.5, 1.5, 0, 1, 0.3, 0.2, 0.9)
e7 <- testModel(1600, 1000, t, 0.85, -0.6, 0.6/3)
set.seed(1)
t <- getTime(1700, 0, 2, 0.25, 12, 1.5, 1.5, 0, 1.3, 0.4, 0.3, 0.9)
e1 <- testModel(1700, 1000, t, 0.85, -0.6, 0.6/3)
set.seed(1)
t <- getTime(1800, 0, 2, 0.25, 12, 1.5, 1.5, 0, 1.5, 0.5, 0.4, 0.9)
e13 <- testModel(1800, 1000, t, 0.85, -0.6, 0.6/3)

par(mfrow=c(3,3))
b0.true <- c(e3$coef.true[1], e5$coef.true[1], e7$coef.true[1], e1$coef.true[1], e13$coef.true[1])
b0.obs1 <- c(e3$coef.obs1[1], e5$coef.obs1[1], e7$coef.obs1[1], e1$coef.obs1[1], e13$coef.obs1[1])
b0.obs2 <- c(e3$coef.obs2[1], e5$coef.obs2[1], e7$coef.obs2[1], e1$coef.obs2[1], e13$coef.obs2[1])
b0.obs3 <- c(e3$coef.obs3[1], e5$coef.obs3[1], e7$coef.obs3[1], e1$coef.obs3[1], e13$coef.obs3[1])
plot(c(1:5), b0.true-0.85, type = "o", col = "red", xlab = "Distribution of measurement error (mean, sd)",
  ylab = "Estimation Bias", ylim = c(-0.1, 0.1), main = "A. Bias of Estimated beta0 = 0.85", xaxt = 'n')
lines(c(1:5), b0.obs1-0.85, type = "o", col = "blue")
axis(side=1, at=c(1:5), labels=c("(0.163; 0.105)", "(0.257; 0.154)", "(0.331; 0.174)",
  "(0.450; 0.252)", "(0.564; 0.318)"), cex.axis = 0.6)

```

```

b1.true <- c(e3$coef.true[2], e5$coef.true[2], e7$coef.true[2], e1$coef.true[2], e13$coef.true[2])
b1.obs1 <- c(e3$coef.obs1[2], e5$coef.obs1[2], e7$coef.obs1[2], e1$coef.obs1[2], e13$coef.obs1[2])
b1.obs2 <- c(e3$coef.obs2[2], e5$coef.obs2[2], e7$coef.obs2[2], e1$coef.obs2[2], e13$coef.obs2[2])
b1.obs3 <- c(e3$coef.obs3[2], e5$coef.obs3[2], e7$coef.obs3[2], e1$coef.obs3[2], e13$coef.obs3[2])
plot(c(1:5), b1.true+0.6, type = "o", col = "red", xlab = "Distribution of measurement error (mean, sd)",
      ylab = "Estimation Bias", ylim = c(-0.2, 0.2), main = "B. Bias of Estimated beta1 = -0.6", xaxt = 'n')
lines(c(1:5), b1.obs1+0.6, type = "o", col = "blue")
axis(side=1,at=c(1:5),labels=c("(0.163; 0.105)", "(0.257; 0.154)", "(0.331; 0.174)",
                               "(0.450; 0.252)", "(0.564; 0.318)"), cex.axis = 0.6)

```

```

b2.true <- c(e3$coef.true[3], e5$coef.true[3], e7$coef.true[3], e1$coef.true[3], e13$coef.true[3])
b2.obs1 <- c(e3$coef.obs1[3], e5$coef.obs1[3], e7$coef.obs1[3], e1$coef.obs1[3], e13$coef.obs1[3])
b2.obs2 <- c(e3$coef.obs2[3], e5$coef.obs2[3], e7$coef.obs2[3], e1$coef.obs2[3], e13$coef.obs2[3])
plot(c(1:5), b2.true, type = "o", col = "red", xlab = "Distribution of measurement error (mean, sd)",
      ylab = "Estimation Bias", ylim = c(-0.1, 0.1), main = "C. Bias of Estimated beta2 = 0", xaxt = 'n')
lines(c(1:5), b2.obs1, type = "o", col = "blue")
axis(side=1,at=c(1:5),labels=c("(0.163; 0.105)", "(0.257; 0.154)", "(0.331; 0.174)",
                               "(0.450; 0.252)", "(0.564; 0.318)"), cex.axis = 0.6)

```

```

b3.true <- c(e3$coef.true[4], e5$coef.true[4], e7$coef.true[4], e1$coef.true[4], e13$coef.true[4])
b3.obs1 <- c(e3$coef.obs1[4], e5$coef.obs1[4], e7$coef.obs1[4], e1$coef.obs1[4], e13$coef.obs1[4])
plot(c(1:5), b3.true-0.6/3, type = "o", col = "red", xlab = "Distribution of measurement error (mean, sd)",
      ylab = "Estimation Bias", ylim = c(-0.1, 0.1), main = "D. Bias of Estimated gamma = 0.2", xaxt = 'n')
lines(c(1:5), b3.obs1-0.6/3, type = "o", col = "blue")
axis(side=1,at=c(1:5),labels=c("(0.163; 0.105)", "(0.257; 0.154)", "(0.331; 0.174)",
                               "(0.450; 0.252)", "(0.564; 0.318)"), cex.axis = 0.6)

```

```

typeI.true <- c(e3$typeI.true, e5$typeI.true, e7$typeI.true, e1$typeI.true, e13$typeI.true)
typeI.obs1 <- c(e3$typeI.obs1, e5$typeI.obs1, e7$typeI.obs1, e1$typeI.obs1, e13$typeI.obs1)
typeI.obs2 <- c(e3$typeI.obs2, e5$typeI.obs2, e7$typeI.obs2, e1$typeI.obs2, e13$typeI.obs2)
plot(c(1:5), typeI.true, type = "o", col = "red", xlab = "Distribution of measurement error (mean, sd)",
      ylab = "Type I", ylim = c(0, 1), main = "E. Type I error of beta2", xaxt = 'n')
lines(c(1:5), typeI.obs1, type = "o", col = "blue")
lines(c(1:5), typeI.obs2, type = "o", col = "green")
axis(side=1,at=c(1:5),labels=c("(0.163; 0.105)", "(0.257; 0.154)", "(0.331; 0.174)",
                               "(0.450; 0.252)", "(0.564; 0.318)"), cex.axis = 0.6)

```

```

powers.true <- c(e3$powerTrt.true, e5$powerTrt.true, e7$powerTrt.true, e1$powerTrt.true,
e13$powerTrt.true)
powers.obs1 <- c(e3$powerTrt.obs1, e5$powerTrt.obs1, e7$powerTrt.obs1, e1$powerTrt.obs1,
e13$powerTrt.obs1)
powers.obs2 <- c(e3$powerTrt.obs2, e5$powerTrt.obs2, e7$powerTrt.obs2, e1$powerTrt.obs2,
e13$powerTrt.obs2)
powers.obs3 <- c(e3$powerTrt.obs3, e5$powerTrt.obs3, e7$powerTrt.obs3, e1$powerTrt.obs3,
e13$powerTrt.obs3)
plot(c(1:5), powers.true, type = "o", col = "red", xlab = "Distribution of measurement error (mean, sd)",
      ylab = "Power", ylim = c(0, 1), main = "F. Power to detect treatment", xaxt = 'n')
lines(c(1:5), powers.obs1, type = "o", col = "blue")
lines(c(1:5), powers.obs2, type = "o", col = "green")
lines(c(1:5), powers.obs3, type = "o", col = "pink")
axis(side=1, at=c(1:5), labels=c("(0.163; 0.105)", "(0.257; 0.154)", "(0.331; 0.174)",
                                "(0.450; 0.252)", "(0.564; 0.318)"), cex.axis = 0.6)

powersInt.true <- c(e3$powerInt.true, e5$powerInt.true, e7$powerInt.true, e1$powerInt.true,
e13$powerInt.true)
powersInt.obs1 <- c(e3$powerInt.obs1, e5$powerInt.obs1, e7$powerInt.obs1, e1$powerInt.obs1,
e13$powerInt.obs1)
plot(c(1:5), powersInt.true, type = "o", col = "red", xlab = "Distribution of measurement error (mean, sd)",
      ylab = "Power", ylim = c(0, 1), main = "G. Power to detect gamma", xaxt = 'n')
lines(c(1:5), powersInt.obs1, type = "o", col = "blue")
axis(side=1, at=c(1:5), labels=c("(0.163; 0.105)", "(0.257; 0.154)", "(0.331; 0.174)",
                                "(0.450; 0.252)", "(0.564; 0.318)"), cex.axis = 0.6)

plot(1, type="n", axes=FALSE, xlab="", ylab="")
legend(1, 1, legend = c("Model 1", "Model 2", "Model 3", "Model 4"), col=c("red", "blue", "green", "pink"),
      lwd=2, cex=1, xjust=0.5, yjust=0.5)

##### Test for lognormal #####
set.seed(1)
t <- getTime2(1400, 0, 2, 0.25, 12, 0.5, 0.5, 0, 0.5, 0.1, 0.15, 0.9)
el3 <- testModel(1400, 1000, t, 0.85, -0.6, 0.6/3)
set.seed(1)
t <- getTime2(1500, 0, 2, 0.25, 12, 0.5, 0.5, 0, 0.7, 0.2, 0.2, 0.9)
el5 <- testModel(1500, 1000, t, 0.85, -0.6, 0.6/3)
set.seed(1)
t <- getTime2(1600, 0, 2, 0.25, 12, 0.5, 0.5, 0, 1, 0.3, 0.2, 0.9)
el7 <- testModel(1600, 1000, t, 0.85, -0.6, 0.6/3)

```

```

set.seed(1)
t <- getTime2(1700, 0, 2, 0.25, 12, 0.5, 0.5, 0, 1.3, 0.4, 0.3, 0.9)
el1 <- testModel(1700, 1000, t, 0.85, -0.6, 0.6/3)
set.seed(1)
t <- getTime2(1800, 0, 2, 0.25, 12, 0.5, 0.5, 0, 1.5, 0.5, 0.4, 0.9)
el13 <- testModel(1800, 1000, t, 0.85, -0.6, 0.6/3)

par(mfrow=c(3,3))
b0.true <- c(el3$coef.true[1], el5$coef.true[1], el7$coef.true[1], el1$coef.true[1], el13$coef.true[1])
b0.obs1 <- c(el3$coef.obs1[1], el5$coef.obs1[1], el7$coef.obs1[1], el1$coef.obs1[1], el13$coef.obs1[1])
b0.obs2 <- c(el3$coef.obs2[1], el5$coef.obs2[1], el7$coef.obs2[1], el1$coef.obs2[1], el13$coef.obs2[1])
b0.obs3 <- c(el3$coef.obs3[1], el5$coef.obs3[1], el7$coef.obs3[1], el1$coef.obs3[1], el13$coef.obs3[1])
plot(c(1:5), b0.true-0.85, type = "o", col = "red", xlab = "Distribution of measurement error (mean, sd)",
      ylab = "Estimation Bias", ylim = c(-0.1, 0.1), main = "A. Bias of Estimated beta0 = 0.85", xaxt = 'n')
lines(c(1:5), b0.obs1-0.85, type = "o", col = "blue")
axis(side=1,at=c(1:5),labels=c("(0.163; 0.105)", "(0.257; 0.154)", "(0.331; 0.174)",
                               "(0.450; 0.252)", "(0.564; 0.318)"), cex.axis = 0.6)

b1.true <- c(el3$coef.true[2], el5$coef.true[2], el7$coef.true[2], el1$coef.true[2], el13$coef.true[2])
b1.obs1 <- c(el3$coef.obs1[2], el5$coef.obs1[2], el7$coef.obs1[2], el1$coef.obs1[2], el13$coef.obs1[2])
b1.obs2 <- c(el3$coef.obs2[2], el5$coef.obs2[2], el7$coef.obs2[2], el1$coef.obs2[2], el13$coef.obs2[2])
b1.obs3 <- c(el3$coef.obs3[2], el5$coef.obs3[2], el7$coef.obs3[2], el1$coef.obs3[2], el13$coef.obs3[2])
plot(c(1:5), b1.true+0.6, type = "o", col = "red", xlab = "Distribution of measurement error (mean, sd)",
      ylab = "Estimation Bias", ylim = c(-0.2, 0.2), main = "B. Bias of Estimated beta1 = -0.6", xaxt = 'n')
lines(c(1:5), b1.obs1+0.6, type = "o", col = "blue")
axis(side=1,at=c(1:5),labels=c("(0.163; 0.105)", "(0.257; 0.154)", "(0.331; 0.174)",
                               "(0.450; 0.252)", "(0.564; 0.318)"), cex.axis = 0.6)

b2.true <- c(el3$coef.true[3], el5$coef.true[3], el7$coef.true[3], el1$coef.true[3], el13$coef.true[3])
b2.obs1 <- c(el3$coef.obs1[3], el5$coef.obs1[3], el7$coef.obs1[3], el1$coef.obs1[3], el13$coef.obs1[3])
b2.obs2 <- c(el3$coef.obs2[3], el5$coef.obs2[3], el7$coef.obs2[3], el1$coef.obs2[3], el13$coef.obs2[3])
plot(c(1:5), b2.true, type = "o", col = "red", xlab = "Distribution of measurement error (mean, sd)",
      ylab = "Estimation Bias", ylim = c(-0.1, 0.1), main = "C. Bias of Estimated beta2 = 0", xaxt = 'n')
lines(c(1:5), b2.obs1, type = "o", col = "blue")
axis(side=1,at=c(1:5),labels=c("(0.163; 0.105)", "(0.257; 0.154)", "(0.331; 0.174)",
                               "(0.450; 0.252)", "(0.564; 0.318)"), cex.axis = 0.6)

b3.true <- c(el3$coef.true[4], el5$coef.true[4], el7$coef.true[4], el1$coef.true[4], el13$coef.true[4])
b3.obs1 <- c(el3$coef.obs1[4], el5$coef.obs1[4], el7$coef.obs1[4], el1$coef.obs1[4], el13$coef.obs1[4])

```

```

plot(c(1:5), b3.true-0.6/3, type = "o", col = "red", xlab = "Distribution of measurement error (mean, sd)",
      ylab = "Estimation Bias", ylim = c(-0.1, 0.1), main = "D. Bias of Estimated gamma = 0.2", xaxt = 'n')
lines(c(1:5), b3.obs1-0.6/3, type = "o", col = "blue")
axis(side=1,at=c(1:5),labels=c("(0.163; 0.105)", "(0.257; 0.154)", "(0.331; 0.174)",
                               "(0.450; 0.252)", "(0.564; 0.318)"), cex.axis = 0.6)

```

```

typeI.true <- c(el3$typeI.true, el5$typeI.true, el7$typeI.true, el1$typeI.true, el13$typeI.true)
typeI.obs1 <- c(el3$typeI.obs1, el5$typeI.obs1, el7$typeI.obs1, el1$typeI.obs1, el13$typeI.obs1)
typeI.obs2 <- c(el3$typeI.obs2, el5$typeI.obs2, el7$typeI.obs2, el1$typeI.obs2, el13$typeI.obs2)
plot(c(1:5), typeI.true, type = "o", col = "red", xlab = "Distribution of measurement error (mean, sd)",
      ylab = "Type I", ylim = c(0, 1), main = "E. Type I error of beta2", xaxt = 'n')
lines(c(1:5), typeI.obs1, type = "o", col = "blue")
lines(c(1:5), typeI.obs2, type = "o", col = "green")
axis(side=1,at=c(1:5),labels=c("(0.163; 0.105)", "(0.257; 0.154)", "(0.331; 0.174)",
                               "(0.450; 0.252)", "(0.564; 0.318)"), cex.axis = 0.6)

```

```

powers.true <- c(el3$powerTrt.true, el5$powerTrt.true, el7$powerTrt.true, el1$powerTrt.true,
                el13$powerTrt.true)
powers.obs1 <- c(el3$powerTrt.obs1, el5$powerTrt.obs1, el7$powerTrt.obs1, el1$powerTrt.obs1,
                el13$powerTrt.obs1)
powers.obs2 <- c(el3$powerTrt.obs2, el5$powerTrt.obs2, el7$powerTrt.obs2, el1$powerTrt.obs2,
                el13$powerTrt.obs2)
powers.obs3 <- c(el3$powerTrt.obs3, el5$powerTrt.obs3, el7$powerTrt.obs3, el1$powerTrt.obs3,
                el13$powerTrt.obs3)
plot(c(1:5), powers.true, type = "o", col = "red", xlab = "Distribution of measurement error (mean, sd)",
      ylab = "Power", ylim = c(0, 1), main = "F. Power to detect treatment", xaxt = 'n')
lines(c(1:5), powers.obs1, type = "o", col = "blue")
lines(c(1:5), powers.obs2, type = "o", col = "green")
lines(c(1:5), powers.obs3, type = "o", col = "pink")
axis(side=1,at=c(1:5),labels=c("(0.163; 0.105)", "(0.257; 0.154)", "(0.331; 0.174)",
                               "(0.450; 0.252)", "(0.564; 0.318)"), cex.axis = 0.6)

```

```

powersInt.true <- c(el3$powerInt.true, el5$powerInt.true, el7$powerInt.true, el1$powerInt.true,
                  el13$powerInt.true)
powersInt.obs1 <- c(el3$powerInt.obs1, el5$powerInt.obs1, el7$powerInt.obs1, el1$powerInt.obs1,
                  el13$powerInt.obs1)
plot(c(1:5), powersInt.true, type = "o", col = "red", xlab = "Distribution of measurement error (mean, sd)",
      ylab = "Power", ylim = c(0, 1), main = "G. Power to detect gamma", xaxt = 'n')
lines(c(1:5), powersInt.obs1, type = "o", col = "blue")
axis(side=1,at=c(1:5),labels=c("(0.163; 0.105)", "(0.257; 0.154)", "(0.331; 0.174)",
                               "(0.450; 0.252)", "(0.564; 0.318)"), cex.axis = 0.6)

```

```
"(0.450; 0.252)", "(0.564; 0.318)"), cex.axis = 0.6)
```

```
plot(1, type="n", axes=FALSE, xlab="", ylab="")  
legend(1, 1, legend = c("Model 1", "Model 2", "Model 3", "Model 4"), col=c("red", "blue", "green", "pink"),  
      lwd=2, cex=1, xjust=0.5, yjust=0.5)
```

```
# Test for interaction when distribution of error (range)
```

```
set.seed(1)
```

```
t <- getTime(1700, 0, 2, 0.25, 12, 1.5, 1.5, 0, 0.3, 0, 0.1, 0.9)
```

```
p3 <- testModel(1700, 1000, t, 0.85, -1.2, 1.2/2)
```

```
set.seed(1)
```

```
t <- getTime(1600, 0, 2, 0.25, 12, 1.5, 1.5, 0, 0.5, 0.1, 0.15, 0.9)
```

```
p5 <- testModel(1600, 1000, t, 0.85, -1.2, 1.2/2)
```

```
set.seed(1)
```

```
t <- getTime(1500, 0, 2, 0.25, 12, 1.5, 1.5, 0, 0.7, 0.2, 0.2, 0.9)
```

```
p7 <- testModel(1500, 1000, t, 0.85, -1.2, 1.2/2)
```

```
set.seed(1)
```

```
t <- getTime(1400, 0, 2, 0.25, 12, 1.5, 1.5, 0, 1, 0.3, 0.2, 0.9)
```

```
p1 <- testModel(1400, 1000, t, 0.85, -1.2, 1.2/2)
```

```
set.seed(1)
```

```
t <- getTime(1300, 0, 2, 0.25, 12, 1.5, 1.5, 0, 1.3, 0.4, 0.3, 0.9)
```

```
p13 <- testModel(1300, 1000, t, 0.85, -1.2, 1.2/2)
```

```
par(mfrow=c(3,3))
```

```
b0.true <- c(p3$coef.true[1], p5$coef.true[1], p7$coef.true[1], p1$coef.true[1], p13$coef.true[1])
```

```
b0.obs1 <- c(p3$coef.obs1[1], p5$coef.obs1[1], p7$coef.obs1[1], p1$coef.obs1[1], p13$coef.obs1[1])
```

```
b0.obs2 <- c(p3$coef.obs2[1], p5$coef.obs2[1], p7$coef.obs2[1], p1$coef.obs2[1], p13$coef.obs2[1])
```

```
b0.obs3 <- c(p3$coef.obs3[1], p5$coef.obs3[1], p7$coef.obs3[1], p1$coef.obs3[1], p13$coef.obs3[1])
```

```
plot(c(1:5), b0.true-0.85, type = "o", col = "red", xlab = "Distribution of measurement error (mean, sd)",
```

```
      ylab = "Estimation Bias", ylim = c(-0.1, 0.1), main = "A. Bias of Estimated beta0 = 0.85", xaxt = 'n')
```

```
lines(c(1:5), b0.obs1-0.85, type = "o", col = "blue")
```

```
axis(side=1, at=c(1:5), labels=c("(0.163; 0.105)", "(0.257; 0.154)", "(0.331; 0.174)",
```

```
      "(0.450; 0.252)", "(0.564; 0.318)"), cex.axis = 0.6)
```

```
b1.true <- c(p3$coef.true[2], p5$coef.true[2], p7$coef.true[2], p1$coef.true[2], p13$coef.true[2])
```

```
b1.obs1 <- c(p3$coef.obs1[2], p5$coef.obs1[2], p7$coef.obs1[2], p1$coef.obs1[2], p13$coef.obs1[2])
```

```
b1.obs2 <- c(p3$coef.obs2[2], p5$coef.obs2[2], p7$coef.obs2[2], p1$coef.obs2[2], p13$coef.obs2[2])
```

```
b1.obs3 <- c(p3$coef.obs3[2], p5$coef.obs3[2], p7$coef.obs3[2], p1$coef.obs3[2], p13$coef.obs3[2])
```

```
plot(c(1:5), b1.true+1.2, type = "o", col = "red", xlab = "Distribution of measurement error (mean, sd)",
```

```
      ylab = "Estimation Bias", ylim = c(-0.4, 0.4), main = "B. Bias of Estimated beta1 = -1.2", xaxt = 'n')
```

```

lines(c(1:5), b1.obs1+1.2, type = "o", col = "blue")
axis(side=1,at=c(1:5),labels=c("(0.163; 0.105)", "(0.257; 0.154)", "(0.331; 0.174)",
"(0.450; 0.252)", "(0.564; 0.318)"), cex.axis = 0.6)

```

```

b2.true <- c(p3$coef.true[3], p5$coef.true[3], p7$coef.true[3], p1$coef.true[3], p13$coef.true[3])
b2.obs1 <- c(p3$coef.obs1[3], p5$coef.obs1[3], p7$coef.obs1[3], p1$coef.obs1[3], p13$coef.obs1[3])
b2.obs2 <- c(p3$coef.obs2[3], p5$coef.obs2[3], p7$coef.obs2[3], p1$coef.obs2[3], p13$coef.obs2[3])
plot(c(1:5), b2.true, type = "o", col = "red", xlab = "Distribution of measurement error (mean, sd)",
ylab = "Estimation Bias", ylim = c(-0.1, 0.1), main = "C. Bias of Estimated beta2 = 0", xaxt = 'n')
lines(c(1:5), b2.obs1, type = "o", col = "blue")
axis(side=1,at=c(1:5),labels=c("(0.163; 0.105)", "(0.257; 0.154)", "(0.331; 0.174)",
"(0.450; 0.252)", "(0.564; 0.318)"), cex.axis = 0.6)

```

```

b3.true <- c(p3$coef.true[4], p5$coef.true[4], p7$coef.true[4], p1$coef.true[4], p13$coef.true[4])
b3.obs1 <- c(p3$coef.obs1[4], p5$coef.obs1[4], p7$coef.obs1[4], p1$coef.obs1[4], p13$coef.obs1[4])
plot(c(1:5), b3.true-0.6, type = "o", col = "red", xlab = "Distribution of measurement error (mean, sd)",
ylab = "Estimation Bias", ylim = c(-0.1, 0.1), main = "D. Bias of Estimated gamma = 0.6", xaxt = 'n')
lines(c(1:5), b3.obs1-0.6, type = "o", col = "blue")
axis(side=1,at=c(1:5),labels=c("(0.163; 0.105)", "(0.257; 0.154)", "(0.331; 0.174)",
"(0.450; 0.252)", "(0.564; 0.318)"), cex.axis = 0.6)

```

```

typeI.true <- c(p3$typeI.true, p5$typeI.true, p7$typeI.true, p1$typeI.true, p13$typeI.true)
typeI.obs1 <- c(p3$typeI.obs1, p5$typeI.obs1, p7$typeI.obs1, p1$typeI.obs1, p13$typeI.obs1)
typeI.obs2 <- c(p3$typeI.obs2, p5$typeI.obs2, p7$typeI.obs2, p1$typeI.obs2, p13$typeI.obs2)
plot(c(1:5), typeI.true, type = "o", col = "red", xlab = "Distribution of measurement error (mean, sd)",
ylab = "Type I", ylim = c(0, 1), main = "E. Type I error of beta2", xaxt = 'n')
lines(c(1:5), typeI.obs1, type = "o", col = "blue")
lines(c(1:5), typeI.obs2, type = "o", col = "green")
axis(side=1,at=c(1:5),labels=c("(0.163; 0.105)", "(0.257; 0.154)", "(0.331; 0.174)",
"(0.450; 0.252)", "(0.564; 0.318)"), cex.axis = 0.6)

```

```

powers.true <- c(p3$powerTrt.true, p5$powerTrt.true, p7$powerTrt.true, p1$powerTrt.true,
p13$powerTrt.true)
powers.obs1 <- c(p3$powerTrt.obs1, p5$powerTrt.obs1, p7$powerTrt.obs1, p1$powerTrt.obs1,
p13$powerTrt.obs1)
powers.obs2 <- c(p3$powerTrt.obs2, p5$powerTrt.obs2, p7$powerTrt.obs2, p1$powerTrt.obs2,
p13$powerTrt.obs2)

```

```

powers.obs3 <- c(p3$powerTrt.obs3, p5$powerTrt.obs3, p7$powerTrt.obs3, p1$powerTrt.obs3,
p13$powerTrt.obs3)
plot(c(1:5), powers.true, type = "o", col = "red", xlab = "Distribution of measurement error (mean, sd)",
      ylab = "Power", ylim = c(0, 1), main = "F. Power to detect treatment", xaxt = 'n')
lines(c(1:5), powers.obs1, type = "o", col = "blue")
lines(c(1:5), powers.obs2, type = "o", col = "green")
lines(c(1:5), powers.obs3, type = "o", col = "pink")
axis(side=1, at=c(1:5), labels=c("(0.163; 0.105)", "(0.257; 0.154)", "(0.331; 0.174)",
                                "(0.450; 0.252)", "(0.564; 0.318)"), cex.axis = 0.6)

powersInt.true <- c(p3$powerInt.true, p5$powerInt.true, p7$powerInt.true, p1$powerInt.true,
p13$powerInt.true)
powersInt.obs1 <- c(p3$powerInt.obs1, p5$powerInt.obs1, p7$powerInt.obs1, p1$powerInt.obs1,
p13$powerInt.obs1)
plot(c(1:5), powersInt.true, type = "o", col = "red", xlab = "Distribution of measurement error (mean, sd)",
      ylab = "Power", ylim = c(0, 1), main = "G. Power to detect gamma", xaxt = 'n')
lines(c(1:5), powersInt.obs1, type = "o", col = "blue")
axis(side=1, at=c(1:5), labels=c("(0.163; 0.105)", "(0.257; 0.154)", "(0.331; 0.174)",
                                "(0.450; 0.252)", "(0.564; 0.318)"), cex.axis = 0.6)

plot(1, type="n", axes=FALSE, xlab="", ylab="")
legend(1, 1, legend = c("Model 1", "Model 2", "Model 3", "Model 4"), col=c("red", "blue", "green",
"pink"), lwd=2, cex=1, xjust=0.5, yjust=0.5)

# Treatment effect at time 0
set.seed(1)
t <- getTime(20000, 0, 2, 0.25, 12, 1.5, 1.5, 0, 0.5, 0.1, 0.15, 0.9)
b1 <- testModel(20000, 1000, t, 0.85, -0.15, 0.15/3)
set.seed(1)
t <- getTime(4500, 0, 2, 0.25, 12, 1.5, 1.5, 0, 0.5, 0.1, 0.15, 0.9)
b2 <- testModel(4500, 1000, t, 0.85, -0.3, 0.3/3)
set.seed(1)
t <- getTime(2300, 0, 2, 0.25, 12, 1.5, 1.5, 0, 0.5, 0.1, 0.15, 0.9)
b3 <- testModel(2300, 1000, t, 0.85, -0.45, 0.45/3)
set.seed(1)
t <- getTime(1200, 0, 2, 0.25, 12, 1.5, 1.5, 0, 0.5, 0.1, 0.15, 0.9)
b4 <- testModel(1200, 1000, t, 0.85, -0.6, 0.6/3)
set.seed(1)
t <- getTime(800, 0, 2, 0.25, 12, 1.5, 1.5, 0, 0.5, 0.1, 0.15, 0.9)

```

```

b5 <- testModel(800, 1000, t, 0.85, -0.75, 0.75/3)

par(mfrow=c(3,3))
b0.true <- c(b1$coef.true[1], b2$coef.true[1], b3$coef.true[1], b4$coef.true[1], b5$coef.true[1])
b0.obs1 <- c(b1$coef.obs1[1], b2$coef.obs1[1], b3$coef.obs1[1], b4$coef.obs1[1], b5$coef.obs1[1])
b0.obs2 <- c(b1$coef.obs2[1], b2$coef.obs2[1], b3$coef.obs2[1], b4$coef.obs2[1], b5$coef.obs2[1])
b0.obs3 <- c(b1$coef.obs3[1], b2$coef.obs3[1], b3$coef.obs3[1], b4$coef.obs3[1], b5$coef.obs3[1])
plot(c(1, 2, 3, 4, 5), b0.true-0.85, type = "o", col = "red", xlab = "Log odds ratio at time",
      ylab = "Estimation Bias", ylim = c(-0.1, 0.1), main = "A. Bias of Estimated beta0 = 0.85", xaxt = 'n')
lines(c(1, 2, 3, 4, 5), b0.obs1-0.85, type = "o", col = "blue")
axis(side=1,at=c(1:5),labels=c("-0.15", "-0.3", "-0.45", "-0.6", "-0.75"))

b1.true <- c(b1$coef.true[2], b2$coef.true[2], b3$coef.true[2], b4$coef.true[2], b5$coef.true[2])
b1.obs1 <- c(b1$coef.obs1[2], b2$coef.obs1[2], b3$coef.obs1[2], b4$coef.obs1[2], b5$coef.obs1[2])
b1.obs2 <- c(b1$coef.obs2[2], b2$coef.obs2[2], b3$coef.obs2[2], b4$coef.obs2[2], b5$coef.obs2[2])
b1.obs3 <- c(b1$coef.obs3[2], b2$coef.obs3[2], b3$coef.obs3[2], b4$coef.obs3[2], b5$coef.obs3[2])
plot(c(1, 2, 3, 4, 5), b1.true+c(0.15, 0.3, 0.45, 0.6, 0.75), type = "o", col = "red", xlab = "Log odds ratio at
time 0",
      ylab = "Estimation Bias", ylim = c(-0.1, 0.1), main = "B. Bias of Estimated beta1", xaxt = 'n')
lines(c(1, 2, 3, 4, 5), b1.obs1+c(0.15, 0.3, 0.45, 0.6, 0.75), type = "o", col = "blue")
axis(side=1,at=c(1:5),labels=c("-0.15", "-0.3", "-0.45", "-0.6", "-0.75"))

b2.true <- c(b1$coef.true[3], b2$coef.true[3], b3$coef.true[3], b4$coef.true[3], b5$coef.true[3])
b2.obs1 <- c(b1$coef.obs1[3], b2$coef.obs1[3], b3$coef.obs1[3], b4$coef.obs1[3], b5$coef.obs1[3])
b2.obs2 <- c(b1$coef.obs2[3], b2$coef.obs2[3], b3$coef.obs2[3], b4$coef.obs2[3], b5$coef.obs2[3])
plot(c(1, 2, 3, 4, 5), b2.true, type = "o", col = "red", xlab = "Log odds ratio at time 0",
      ylab = "Estimation Bias", ylim = c(-0.1, 0.1), main = "C. Bias of Estimated beta2 = 0", xaxt = 'n')
lines(c(1, 2, 3, 4, 5), b2.obs1, type = "o", col = "blue")
axis(side=1,at=c(1:5),labels=c("-0.15", "-0.3", "-0.45", "-0.6", "-0.75"))

b3.true <- c(b1$coef.true[4], b2$coef.true[4], b3$coef.true[4], b4$coef.true[4], b5$coef.true[4])
b3.obs1 <- c(b1$coef.obs1[4], b2$coef.obs1[4], b3$coef.obs1[4], b4$coef.obs1[4], b5$coef.obs1[4])
plot(c(1, 2, 3, 4, 5), b3.true-c(0.15, 0.3, 0.45, 0.6, 0.75)/3, type = "o", col = "red", xlab = "Log odds ratio at
time 0",
      ylab = "Estimation Bias", ylim = c(-0.1, 0.1), main = "D. Bias of Estimated gamma = 0.2", xaxt = 'n')
lines(c(1, 2, 3, 4, 5), b3.obs1-c(0.15, 0.3, 0.45, 0.6, 0.75)/3, type = "o", col = "blue")
axis(side=1,at=c(1:5),labels=c("-0.15", "-0.3", "-0.45", "-0.6", "-0.75"))

typeI.true <- c(b1$typeI.true, b2$typeI.true, b3$typeI.true, b4$typeI.true, b5$typeI.true)
typeI.obs1 <- c(b1$typeI.obs1, b2$typeI.obs1, b3$typeI.obs1, b4$typeI.obs1, b5$typeI.obs1)

```

```

typeI.obs2 <- c(b1$typeI.obs2, b2$typeI.obs2, b3$typeI.obs2, b4$typeI.obs2, b5$typeI.obs2)
plot(c(1, 2, 3, 4, 5), typeI.true, type = "o", col = "red", xlab = "Log odds ratio at time 0",
      ylab = "TypeI", ylim = c(0, 1), main = "E. Type I error of beta2", xaxt = 'n')
lines(c(1, 2, 3, 4, 5), typeI.obs1, type = "o", col = "blue")
lines(c(1, 2, 3, 4, 5), typeI.obs2, type = "o", col = "green")
axis(side=1, at=c(1:5), labels=c("-0.15", "-0.3", "-0.45", "-0.6", "-0.75"))

powers.true <- c(b1$powerTrt.true, b2$powerTrt.true, b3$powerTrt.true, b4$powerTrt.true,
b5$powerTrt.true)
powers.obs1 <- c(b1$powerTrt.obs1, b2$powerTrt.obs1, b3$powerTrt.obs1, b4$powerTrt.obs1,
b5$powerTrt.obs1)
powers.obs2 <- c(b1$powerTrt.obs2, b2$powerTrt.obs2, b3$powerTrt.obs2, b4$powerTrt.obs2,
b5$powerTrt.obs2)
powers.obs3 <- c(b1$powerTrt.obs3, b2$powerTrt.obs3, b3$powerTrt.obs3, b4$powerTrt.obs3,
b5$powerTrt.obs3)
plot(c(1, 2, 3, 4, 5), powers.true, type = "o", col = "red", xlab = "Log odds ratio at time 0",
      ylab = "Power", ylim = c(0, 1), main = "F. Power to detect treatment", xaxt = 'n')
lines(c(1, 2, 3, 4, 5), powers.obs1, type = "o", col = "blue")
lines(c(1, 2, 3, 4, 5), powers.obs2, type = "o", col = "green")
lines(c(1, 2, 3, 4, 5), powers.obs3, type = "o", col = "pink")
axis(side=1, at=c(1:5), labels=c("-0.15", "-0.3", "-0.45", "-0.6", "-0.75"))

powersInt.true <- c(b1$powerInt.true, b2$powerInt.true, b3$powerInt.true, b4$powerInt.true,
b5$powerInt.true)
powersInt.obs1 <- c(b1$powerInt.obs1, b2$powerInt.obs1, b3$powerInt.obs1, b4$powerInt.obs1,
b5$powerInt.obs1)
plot(c(1, 2, 3, 4, 5), powersInt.true, type = "o", col = "red", xlab = "Log odds ratio at time 0",
      ylab = "Power", ylim = c(0, 1), main = "G. Power to detect gamma", xaxt = 'n')
lines(c(1, 2, 3, 4, 5), powersInt.obs1, type = "o", col = "blue")
axis(side=1, at=c(1:5), labels=c("-0.15", "-0.3", "-0.45", "-0.6", "-0.75"))

plot(1, type="n", axes=FALSE, xlab="", ylab="")
legend(1, 1, legend = c("Model 1", "Model 2", "Model 3", "Model 4"), col=c("red", "blue", "green", "pink"),
      lwd=2, cex=1, xjust=0.5, yjust=0.5)

# Interaction strength
set.seed(1)
t <- getTime(2700, 0, 2, 0.25, 12, 1.5, 1.5, 0, 0.5, 0.1, 0.15, 0.9)
s1 <- testModel(2700, 1000, t, 0.85, -0.6, 0.6)
set.seed(1)

```

```

t <- getTime(2700, 0, 2, 0.25, 12, 1.5, 1.5, 0, 0.5, 0.1, 0.15, 0.9)
s2 <- testModel(2700, 1000, t, 0.85, -0.6, 0.6/2)
set.seed(1)
t <- getTime(1600, 0, 2, 0.25, 12, 1.5, 1.5, 0, 0.5, 0.1, 0.15, 0.9)
s3 <- testModel(1600, 1000, t, 0.85, -0.6, 0.6/3)
set.seed(1)
t <- getTime(1400, 0, 2, 0.25, 12, 1.5, 1.5, 0, 0.5, 0.1, 0.15, 0.9)
s4 <- testModel(1400, 1000, t, 0.85, -0.6, 0.6/4)
set.seed(1)
t <- getTime(1400, 0, 2, 0.25, 12, 1.5, 1.5, 0, 0.5, 0.1, 0.15, 0.9)
s5 <- testModel(1400, 1000, t, 0.85, -0.6, 0.6/5)

par(mfrow=c(3,3))
b0.true <- c(s1$coef.true[1], s2$coef.true[1], s3$coef.true[1], s4$coef.true[1], s5$coef.true[1])
b0.obs1 <- c(s1$coef.obs1[1], s2$coef.obs1[1], s3$coef.obs1[1], s4$coef.obs1[1], s5$coef.obs1[1])
b0.obs2 <- c(s1$coef.obs2[1], s2$coef.obs2[1], s3$coef.obs2[1], s4$coef.obs2[1], s5$coef.obs2[1])
b0.obs3 <- c(s1$coef.obs3[1], s2$coef.obs3[1], s3$coef.obs3[1], s4$coef.obs3[1], s5$coef.obs3[1])
plot(c(1, 2, 3, 4, 5), b0.true-0.85, type = "o", col = "red", xlab = "Crossing time point (h)",
     ylab = "Estimation Bias", ylim = c(-0.1, 0.1), main = "A. Bias of Estimated beta0=0.85")
lines(c(1, 2, 3, 4, 5), b0.obs1-0.85, type = "o", col = "blue")

b1.true <- c(s1$coef.true[2], s2$coef.true[2], s3$coef.true[2], s4$coef.true[2], s5$coef.true[2])
b1.obs1 <- c(s1$coef.obs1[2], s2$coef.obs1[2], s3$coef.obs1[2], s4$coef.obs1[2], s5$coef.obs1[2])
b1.obs2 <- c(s1$coef.obs2[2], s2$coef.obs2[2], s3$coef.obs2[2], s4$coef.obs2[2], s5$coef.obs2[2])
b1.obs3 <- c(s1$coef.obs3[2], s2$coef.obs3[2], s3$coef.obs3[2], s4$coef.obs3[2], s5$coef.obs3[2])
plot(c(1, 2, 3, 4, 5), b1.true+0.6, type = "o", col = "red", xlab = "Crossing time point (h)",
     ylab = "Estimation Bias", ylim = c(-0.2, 0.2), main = "B. Bias of Estimated beta1=-0.6")
lines(c(1, 2, 3, 4, 5), b1.obs1+0.6, type = "o", col = "blue")

b2.true <- c(s1$coef.true[3], s2$coef.true[3], s3$coef.true[3], s4$coef.true[3], s5$coef.true[3])
b2.obs1 <- c(s1$coef.obs1[3], s2$coef.obs1[3], s3$coef.obs1[3], s4$coef.obs1[3], s5$coef.obs1[3])
b2.obs2 <- c(s1$coef.obs2[3], s2$coef.obs2[3], s3$coef.obs2[3], s4$coef.obs2[3], s5$coef.obs2[3])
plot(c(1, 2, 3, 4, 5), b2.true, type = "o", col = "red", xlab = "Crossing time point (h)",
     ylab = "Estimation Bias", ylim = c(-0.1, 0.1), main = "C. Bias of Estimated beta2=0")
lines(c(1, 2, 3, 4, 5), b2.obs1, type = "o", col = "blue")

b3.true <- c(s1$coef.true[4], s2$coef.true[4], s3$coef.true[4], s4$coef.true[4], s5$coef.true[4])

```

```

b3.obs1 <- c(s1$coef.obs1[4], s2$coef.obs1[4], s3$coef.obs1[4], s4$coef.obs1[4], s5$coef.obs1[4])
plot(c(1, 2, 3, 4, 5), b3.true-0.6/c(1:5), type = "o", col = "red", xlab = "Crossing time point (h)",
      ylab = "Estimation Bias", ylim = c(-0.1, 0.1), main = "D. Bias of Estimated gamma")
lines(c(1, 2, 3, 4, 5), b3.obs1-0.6/c(1:5), type = "o", col = "blue")

typeI.true <- c(s1$typeI.true, s2$typeI.true, s3$typeI.true, s4$typeI.true, s5$typeI.true)
typeI.obs1 <- c(s1$typeI.obs1, s2$typeI.obs1, s3$typeI.obs1, s4$typeI.obs1, s5$typeI.obs1)
typeI.obs2 <- c(s1$typeI.obs2, s2$typeI.obs2, s3$typeI.obs2, s4$typeI.obs2, s5$typeI.obs2)
plot(c(1, 2, 3, 4, 5), typeI.true, type = "o", col = "red", xlab = "Crossing time point (h)",
      ylab = "Type I", ylim = c(0, 1), main = "E. Type I error of beta2")
lines(c(1, 2, 3, 4, 5), typeI.obs1, type = "o", col = "blue")
lines(c(1, 2, 3, 4, 5), typeI.obs2, type = "o", col = "green")

powers.true <- c(s1$powerTrt.true, s2$powerTrt.true, s3$powerTrt.true, s4$powerTrt.true,
s5$powerTrt.true)
powers.obs1 <- c(s1$powerTrt.obs1, s2$powerTrt.obs1, s3$powerTrt.obs1, s4$powerTrt.obs1,
s5$powerTrt.obs1)
powers.obs2 <- c(s1$powerTrt.obs2, s2$powerTrt.obs2, s3$powerTrt.obs2, s4$powerTrt.obs2,
s5$powerTrt.obs2)
powers.obs3 <- c(s1$powerTrt.obs3, s2$powerTrt.obs3, s3$powerTrt.obs3, s4$powerTrt.obs3,
s5$powerTrt.obs3)
plot(c(1, 2, 3, 4, 5), powers.true, type = "o", col = "red", xlab = "Crossing time point (h)",
      ylab = "Power", ylim = c(0, 1), main = "F. Power to detect treatment")
lines(c(1, 2, 3, 4, 5), powers.obs1, type = "o", col = "blue")
lines(c(1, 2, 3, 4, 5), powers.obs2, type = "o", col = "green")
lines(c(1, 2, 3, 4, 5), powers.obs3, type = "o", col = "pink")

powersInt.true <- c(s1$powerInt.true, s2$powerInt.true, s3$powerInt.true, s4$powerInt.true,
s5$powerInt.true)
powersInt.obs1 <- c(s1$powerInt.obs1, s2$powerInt.obs1, s3$powerInt.obs1, s4$powerInt.obs1,
s5$powerInt.obs1)
plot(c(1, 2, 3, 4, 5), powersInt.true, type = "o", col = "red", xlab = "Crossing time point (h)",
      ylab = "Power", ylim = c(0, 1), main = "G. Power to detect gamma")
lines(c(1, 2, 3, 4, 5), powersInt.obs1, type = "o", col = "blue")

plot(1, type="n", axes=FALSE, xlab="", ylab="")
legend(1, 1, legend = c("Model 1", "Model 2", "Model 3", "Model 4"), col=c("red", "blue", "green", "pink"),
      lwd=2, cex=1, xjust=0.5, yjust=0.5)

```

```
#####
###
##### Meta-analysis#####
getDataCrash <- function(n, timeData, log.odd0, log.or1, logit.increase1) {
  a1 <- sample(1:n, size=(n/2)) # randomly choose placebo group
  a2 <- setdiff(1:n, a1) # randomly choose treatment 1 group
  x <- cbind(rep(0, n), rep(0, n))

  for (i in a1) {x[i, 1] <- 1} # assign treatment to each Xi
  for (i in a2) {x[i, 2] <- 1}
  colnames(x) <- c("plb", "trt")

  p <- rep(0,n)
  y <- rep(0,n) # assign 0 to all y

  for (i in 1:n) {
    z <- log.odd0 + log.or1*x[i, 2] + logit.increase1*timeData[i, 1]*x[i, 2]
    p[i] <- exp(z)/(1+exp(z))
    y[i] <- rbinom(1, 1, p[i]) # reassign 1 to y for placebo group with probability p0
  }
  mydata <- data.frame(y, p, x, timeData)
  return(mydata)
}

testPool <- function(n1, n2, n3, simulation, tTXA, tCRASH1, tCRASH2, log.odd0, log.or1, logit.increase1) {
  pwrtrt.true <- 0
  pwrtrt.obs1 <- 0

  pwrint.true <- 0
  pwrint.obs1 <- 0

  pwrtrtCat.true <- 0
  pwrtrtCat.obs1 <- 0

  pwrintCat.true <- 0
  pwrintCat.obs1 <- 0

  for (i in 1:simulation){
    set.seed(i)

```

```

dataTXA <- getData(n1, tTXA, log.odd0, log.or1, logit.increase1)
dataTXA$study <- 1
### Create CRASH data
dataCRASH1 <- getDataCrash(n2, tCRASH1, log.odd0, log.or1, logit.increase1)
dataCRASH1$study <- 0
dataCRASH2 <- getDataCrash(n3, tCRASH2, log.odd0, log.or1, logit.increase1)
dataCRASH2$study <- 0
###Create Pool data
mydata <- rbind(dataTXA, dataCRASH1, dataCRASH2)
### Fit regression for pooled data with fix effect
fit.true <- glm(y~trt+time.true+trt:time.true+study, family = binomial, data = mydata)
fit.trueRed <- glm(y~time.true+study, family = binomial, data = mydata)
fit.obs1 <- glm(y~trt+time.obs+trt:time.obs+study, family = binomial, data = mydata)
fit.obs1Red <- glm(y~time.obs+study, family = binomial, data = mydata)

if(lrtest(fit.true, fit.trueRed)[2,5] < 0.05) { pwrtrt.true <- pwrtrt.true+1 }
if(lrtest(fit.obs1, fit.obs1Red)[2,5] < 0.05) { pwrtrt.obs1 <- pwrtrt.obs1+1 }

if(summary(fit.true)$coefficients[5,4] < 0.05) { pwrprint.true <- pwrprint.true+1 }
if(summary(fit.obs1)$coefficients[5,4] < 0.05) { pwrprint.obs1 <- pwrprint.obs1+1 }
### Create categorical time variable
mydata <- data.frame(mydata, c(1:(n1+n2+n3)), c(1:(n1+n2+n3)))
colnames(mydata)[9:10] <- c("time.trueCat", "time.obsCat")
for (j in 1:nrow(mydata)) {
  if(0 <= mydata$time.true[j] & mydata$time.true[j] < 1) {mydata$time.trueCat[j] <- 0}
  else if(1 <= mydata$time.true[j] & mydata$time.true[j] < 2) {mydata$time.trueCat[j] <- 1}
  else if(2 <= mydata$time.true[j] & mydata$time.true[j] < 3) {mydata$time.trueCat[j] <- 2}
  else if(3 <= mydata$time.true[j]){mydata$time.trueCat[j] <- 3}
  if(0 <= mydata$time.obs[j] & mydata$time.obs[j] < 1) {mydata$time.obsCat[j] <- 0}
  else if(1 <= mydata$time.obs[j] & mydata$time.obs[j] < 2) {mydata$time.obsCat[j] <- 1}
  else if(2 <= mydata$time.obs[j] & mydata$time.obs[j] < 3) {mydata$time.obsCat[j] <- 2}
  else if(3 <= mydata$time.obs[j]){mydata$time.obsCat[j] <- 3}
}
### Fit regression for categorical time
fit.trueCat <- glm(y~trt+as.factor(time.trueCat)+trt:as.factor(time.trueCat)+study, family = binomial,
data = mydata)
fit.trueCatRed <- glm(y~as.factor(time.trueCat)+study, family = binomial, data = mydata)
fit.trueCatintRed <- glm(y~trt+as.factor(time.trueCat)+study, family = binomial, data = mydata)
fit.obsCat <- glm(y~trt+as.factor(time.obsCat)+trt:as.factor(time.obsCat)+study, family = binomial, data
= mydata)

```

```

fit.obsCatRed <- glm(y~as.factor(time.obsCat)+study, family = binomial, data = mydata)
fit.obsCatintRed <- glm(y~trt+as.factor(time.obsCat)+study, family = binomial, data = mydata)

if(lrtest(fit.trueCat, fit.trueCatRed)[2,5] < 0.05) { pwrtrtCat.true <- pwrtrtCat.true+1 }
if(lrtest(fit.obsCat, fit.obsCatRed)[2,5] < 0.05) { pwrtrtCat.obs1 <- pwrtrtCat.obs1+1 }

if(lrtest(fit.trueCat, fit.trueCatintRed)[2,5] < 0.05) { pwrintCat.true <- pwrintCat.true+1 }
if(lrtest(fit.obsCat, fit.obsCatintRed)[2,5] < 0.05) { pwrintCat.obs1 <- pwrintCat.obs1+1 }

}
return(list(pwrtrt.true = pwrtrt.true/simulation, pwrtrt.obs1 = pwrtrt.obs1/simulation,
           pwrint.true = pwrint.true/simulation, pwrint.obs1 = pwrint.obs1/simulation,
           pwrtrtCat.true = pwrtrtCat.true/simulation, pwrtrtCat.obs1 = pwrtrtCat.obs1/simulation,
           pwrintCat.true = pwrintCat.true/simulation, pwrintCat.obs1 = pwrintCat.obs1/simulation))
}

set.seed(1)
tTXA <- getTime(1000, 0, 2, 0.25, 12, 1.5, 1.5, 0, 0.5, 0.1, 0.15, 0.9)
tCRASH1 <- getTime(7143, 0, 8, 0.25, 12, 1.5, 1.5, 0, 0.5, 0.1, 0.15, 0.9)
tCRASH2 <- getTime(2857, 0, 3, 0.25, 12, 1.5, 1.5, 0, 0.5, 0.1, 0.15, 0.9)
Pool1 <- testPool(1000, 7143, 2857, 1000, tTXA, tCRASH1, tCRASH2, 0.85, -0.6, 0.6/3)

set.seed(1)
tTXA <- getTime(1000, 0, 2, 0.25, 12, 1.5, 1.5, 0, 1, 0.3, 0.2, 0.9)
tCRASH1 <- getTime(7143, 0, 8, 0.25, 12, 1.5, 1.5, 0, 1, 0.3, 0.2, 0.9)
tCRASH2 <- getTime(2857, 0, 3, 0.25, 12, 1.5, 1.5, 0, 1, 0.3, 0.2, 0.9)
Pool2 <- testPool(1000, 7143, 2857, 1000, tTXA, tCRASH1, tCRASH2, 0.85, -0.6, 0.6/3)

set.seed(1)
tTXA <- getTime(1000, 0, 2, 0.25, 12, 1.5, 1.5, 0, 1.3, 0.4, 0.3, 0.9)
tCRASH1 <- getTime(7143, 0, 8, 0.25, 12, 1.5, 1.5, 0, 1.3, 0.4, 0.3, 0.9)
tCRASH2 <- getTime(2857, 0, 3, 0.25, 12, 1.5, 1.5, 0, 1.3, 0.4, 0.3, 0.9)
Pool3 <- testPool(1000, 7143, 2857, 1000, tTXA, tCRASH1, tCRASH2, 0.85, -0.6, 0.6/3)

set.seed(1)
tTXA <- getTime(1000, 0, 2, 0.25, 12, 1.5, 1.5, 0, 1.5, 0.5, 0.4, 0.9)
tCRASH1 <- getTime(7143, 0, 8, 0.5, 12, 1.5, 1.5, 0, 1.5, 0.5, 0.4, 0.9)
tCRASH2 <- getTime(2857, 0, 3, 0.5, 12, 1.5, 1.5, 0, 1.5, 0.5, 0.4, 0.9)
Pool4 <- testPool(1000, 7143, 2857, 1000, tTXA, tCRASH1, tCRASH2, 0.85, -0.6, 0.6/3)

```

```

set.seed(1)
tTXA <- getTime(1000, 0, 2, 0.25, 12, 1.5, 1.5, 0, 2, 1, 0.5, 0.9)
tCRASH1 <- getTime(7143, 0, 8, 0.5, 12, 1.5, 1.5, 0, 2, 1, 0.5, 0.9)
tCRASH2 <- getTime(2857, 0, 3, 0.5, 12, 1.5, 1.5, 0, 2, 1, 0.5, 0.9)
Pool5 <- testPool(1000, 7143, 2857, 1000, tTXA, tCRASH1, tCRASH2, 0.85, -0.6, 0.6/3)

par(mfrow=c(2,2))
powersTrt.true <- c(Pool1$pwrtrt.true, Pool2$pwrtrt.true, Pool3$pwrtrt.true, Pool4$pwrtrt.true,
Pool5$pwrtrt.true)
powersTrt.obs1 <- c(Pool1$pwrtrt.obs1, Pool2$pwrtrt.obs1, Pool3$pwrtrt.obs1, Pool4$pwrtrt.obs1,
Pool5$pwrtrt.obs1)
plot(c(1:5), powersTrt.true, type = "o", col = "red", xlab = "Distribution of measurement error (mean, sd)",
ylab = "Power", ylim = c(0, 1), main = "A. Power to detect treatment (continuous time)", xaxt = 'n')
lines(c(1:5), powersTrt.obs1, type = "o", col = "blue")
axis(side=1, at=c(1:5), labels=c("(0.163, 0.106)", "(0.323, 0.175)", "(0.440, 0.249)", "(0.557, 0.314)", "(0.936,
0.426)"), cex.axis = 0.7)
legend("bottomleft", inset=.05, c("Model 1", "Model 2"), lty=c(1,1), col = c("red", "blue"))

powersInt.true <- c(Pool1$pwrInt.true, Pool2$pwrInt.true, Pool3$pwrInt.true, Pool4$pwrInt.true,
Pool5$pwrInt.true)
powersInt.obs1 <- c(Pool1$pwrInt.obs1, Pool2$pwrInt.obs1, Pool3$pwrInt.obs1, Pool4$pwrInt.obs1,
Pool5$pwrInt.obs1)
plot(c(1:5), powersInt.true, type = "o", col = "red", xlab = "Distribution of measurement error (mean, sd)",
ylab = "Power", ylim = c(0, 1), main = "B. Power to detect interaction (continuous time)", xaxt = 'n')
lines(c(1:5), powersInt.obs1, type = "o", col = "blue")
axis(side=1, at=c(1:5), labels=c("(0.163, 0.106)", "(0.323, 0.175)", "(0.440, 0.249)", "(0.557, 0.314)", "(0.936,
0.426)"), cex.axis = 0.7)
legend("bottomleft", inset=.05, c("Model 1", "Model 2"), lty=c(1,1), col = c("red", "blue"))

powersTrtCat.true <- c(Pool1$pwrtrtCat.true, Pool2$pwrtrtCat.true, Pool3$pwrtrtCat.true,
Pool4$pwrtrtCat.true, Pool5$pwrtrtCat.true)
powersTrtCat.obs1 <- c(Pool1$pwrtrtCat.obs1, Pool2$pwrtrtCat.obs1, Pool3$pwrtrtCat.obs1,
Pool4$pwrtrtCat.obs1, Pool5$pwrtrtCat.obs1)
plot(c(1:5), powersTrtCat.true, type = "o", col = "red", xlab = "Distribution of measurement error (mean,
sd)",
ylab = "Power", ylim = c(0, 1), main = "C. Power to detect treatment (categorical time)", xaxt = 'n')
lines(c(1:5), powersTrtCat.obs1, type = "o", col = "blue")
axis(side=1, at=c(1:5), labels=c("(0.163, 0.106)", "(0.323, 0.175)", "(0.440, 0.249)", "(0.557, 0.314)", "(0.936,
0.426)"), cex.axis = 0.7)
legend("bottomleft", inset=.05, c("Model 1", "Model 2"), lty=c(1,1), col = c("red", "blue"))

```

```

powersIntCat.true <- c(Pool1$pwrintCat.true, Pool2$pwrintCat.true, Pool3$pwrintCat.true,
Pool4$pwrintCat.true, Pool5$pwrintCat.true)
powersIntCat.obs1 <- c(Pool1$pwrintCat.obs1, Pool2$pwrintCat.obs1, Pool3$pwrintCat.obs1,
Pool4$pwrintCat.obs1, Pool5$pwrintCat.obs1)
plot(c(1:5), powersIntCat.true, type = "o", col = "red", xlab = "Distribution of measurement error (mean,
sd)",
  ylab = "Power", ylim = c(0, 1), main = "D. Power to detect interaction (categorical time)", xaxt = 'n')
lines(c(1:5), powersIntCat.obs1, type = "o", col = "blue")
axis(side=1, at=c(1:5), labels=c("(0.163, 0.106)", "(0.323, 0.175)", "(0.440, 0.249)", "(0.557, 0.314)", "(0.936,
0.426)"), cex.axis = 0.7)
legend("bottomleft", inset=.05, c("Model 1", "Model 2"), lty=c(1,1), col = c("red", "blue"))

```

### Real ROC TXA data

```

library(gdata)
realTXA <- read.csv("/Users/guoyanshu/Desktop/Thesis/txasample.csv", header = FALSE)
realTXA$V1 <- as.numeric(substr(realTXA$V1, 1, 2))*60+as.numeric(substr(realTXA$V1, 4, 5))
realTXA$V3 <- as.numeric(substr(realTXA$V3, 1, 2))*60+as.numeric(substr(realTXA$V3, 4, 5))
realTXA$callToTrt <- realTXA$V3 - realTXA$V1
realTXA$callToTrt <- realTXA$callToTrt/60
callToTrt <- realTXA$callToTrt[!is.na(realTXA$callToTrt)]
callToTrt <- callToTrt[0<callToTrt]
par(mfrow=c(2,2))
hist(callToTrt, breaks = seq(min(callToTrt)-0.1, max(callToTrt)+0.1, 0.1), xlab = "Time(h)",
  xlim = c(0, max(callToTrt)+0.1), main = "A. Time from 911 call to treatment initiation")
injToCall <- rtruncnorm(length(callToTrt), a=0, b=2, mean = 0.5, sd = 0.5)
hist(injToCall, breaks = seq(min(injToCall), max(injToCall)+0.1, 0.1), xlab = "Time(h)",
  xlim = c(0,2), main = "B. Time from injury to 911 call")
hist(t[,2], breaks = seq(min(t[,2]), max(t[,2])+0.1, 0.1), xlim = c(min(t[,2]),max(t[,2])),
  xlab = "Time (h)", main = paste("C.", colnames(t)[2]))
callTrt <- callToTrt+injToCall+t$error[1:length(callToTrt)]
hist(callTrt[0<callTrt&callTrt<2], breaks = seq(0, 2, 0.1), xlab = "Time(h)",
  xlim = c(0,2), main = "D. Time from injury to treatment initiation")
mean(realTXA$callToTrt[!is.na(realTXA$callToTrt)])

```

```

par(mfrow=c(1,1))
plot(c(0, 8), c(0.85, 0.85), type = "l", col = "red", ylab = "Logit(q)", xaxt = 'n',
  ylim = c(0, 2), xlab = "Treatment initiation time from injury", main = "Treatment effect over initiation
time")
lines(c(0, 3, 8), c(0.25, 0.85, 1.85), type = "l", col = "blue")

```

```

axis(side=1,at=c(0:8),labels=c(0:8))
legend("topleft", inset=.05, c("Placebo","Treatment"), lty=c(1,1), col = c("red", "blue"))

plot(c(0, 8), c(0.7, 0.7), type = "l",col = "red", ylab = "Probability (q)", xaxt = 'n',
      ylim = c(0, 1), xlab = "Treatment initiation time from injury",main = "Treatment effect over initiation
time")
p <- seq(0, 8, by = 0.01 )
lines(p, exp(0.25+0.2*p)/(1+exp(0.25+0.2*p)), type = "l", col = "blue")
axis(side=1,at=c(0:8),labels=c(0:8))
legend("bottomright", inset=.05, c("Placebo","Treatment"), lty=c(1,1), col = c("red", "blue"))

###Comparison between log normal and truncated normal
par(mfrow=c(1,1))
h <- hist(t[,1], breaks = seq(min(t[,1]), max(t[,1])+0.1, 0.1), xlim = c(0,max(t[,1])),
        xlab = "Time (h)", main = "A.True time simulated with truncated normal")
xfit<-seq(min(t[,1]),max(t[,1]),length=40)
yfit<-dtruncnorm(xfit,a=min(t[,1]), b=max(t[,1]), mean=1.5,sd=1.5)
yfit <- yfit*diff(h$mids[1:2])*length(t[,1])
lines(xfit, yfit, col="blue", lwd=2)

h <- hist(t[,1], breaks = seq(min(t[,1]), max(t[,1])+0.1, 0.1), xlim = c(0,2.5),
        xlab = "Time (h)", main = "B. True time simulated with log normal")
xfit<-seq(min(t[,1]),max(t[,1]),length=40)
yfit<-dlnorm(xfit, meanlog = 0.5, sdlog = 0.5)
yfit <- yfit*diff(h$mids[1:2])*length(t[,1])
lines(xfit, yfit, col="blue", lwd=2)

```