About Time: When Times Are Missing
Management of Missing Data in Discrete-Time Survival Analyses

Nils L.M. van de Ven
Utrecht University
N.l.m.vandeven@students.uu.nl

ABSTRACT

Survival analysis is a method of analysis used to study event occurrence. Missing periods in discrete-time survival analyses are problematic, since whether an event occurs determines whether the subject is followed up upon. Seven strategies that can be used when missingness occurs (*case deletion, deletion upon missing, single imputation, multiple imputation, remembrance, the Non-Event-Strategy and the Event-Strategy*) are evaluated using four criteria: effect size bias, standard error bias, power and coverage rate of confidence intervals. Single imputation, multiple imputation and the Non-Event Strategy show good results. Single imputation performs slightly better, yet the Non-Event Strategy is easier to implement.

Keywords

Simulation study, discrete-time survival analysis, event occurrence, missing data, single imputation, multiple Imputation

INTRODUCTION

Event occurrence is studied in many different experimental settings, both in psychological and in medical research. In such studies, the research question is often whether and when such an event occurs. An example of a study where event occurrence is analyzed is a study of relapse after treatment for drug abuse, where researchers are comparing a new type of treatment to an older treatment and want to know if, and when, relapse occurs. In such a study, calculating relative risks or odds ratios does not consider how much time goes by before subjects start using drugs again. On the other hand, it is impossible to calculate a mean time to event and use an Analysis of Variance (ANOVA) to compare the mean times to event of treatments, since not all participants will return to drug abuse during the study, or return to drug abuse at all. Survival analysis, also known as Event History Analysis, takes this into consideration. It analyzes simultaneously both whether and when events occur[1]. It does so by calculating a hazard probability, (the probability of event occurrence during a time interval), and a survival probability (the probability at any given moment in time that the event has not yet occurred to the average participant).

SRC 2017, November 30, 2017, The Netherlands.

Using survival analysis, a variety of research questions can be answered. Singer and Willett coined the mnemonic 'the whether and when test' [1]: if a research question mentions the word 'whether' or 'when', then a survival analysis is probably useful. Examples of studies using survival analysis are numerous, including a study of relapse after treatment of alcoholism [2] and a study of the recurrence of child maltreatment [3].

Event occurrence can be measured both in continuous time and in discrete time [1]. Although measuring on a continuous scale is often ideal, this is generally difficult to accomplish in the social and behavioral sciences[4]. Instead, events are often measured in time intervals. This can lead to some loss of information. The decision to discretize data should therefore be made carefully. However, previous research has shown that discretizing data leads to little loss of power and small parameter and standard error bias [4]. When measuring in discrete time, a discrete-time hazard model is used[1]. In discrete time survival analysis event occurrence is measured in time intervals. For each interval, it is recorded whether the event occurred. Subjects are only followed up upon until event occurrence. After the event occurs, the subject is removed from the dataset for the remainder of the study.

Missing data is a common and difficult problem[5]. Especially in survival analysis a missing period is troublesome, since the observations are done on the condition that each subject in the sample has not yet experienced the event. A missing time period can therefore not be ignored, since it is then unknown whether the subject should still be under observation. Three types of missing data exist[6,7]: data *missing completely at random* (MCAR), where the missingness is independent of other variables, *missing at random* (MAR), in which case the missingness is independent of event occurrence, but dependent on some other variable in the model, and *missing not at random* (MNAR), in which case the missingness depends on event occurrence.

Although the problem of a missing period can often partially be solved by the researcher putting in extra effort to reach his subjects, it is unlikely missing observations can be avoided altogether. Multiple strategies to manage missing data exist. However, no formal study analyzing strategies in managing missing data in a discrete time survival analysis has yet been done. This thesis aims to study how, in survival analyses, missing periods are best managed.

METHOD

Modeling

Discrete time survival analysis describes event occurrence using two functions: the hazard function and the survivor function.

Table 1. *Possible values for the different parameters*

Parameter	Possible Values (meaning of values)		
ω	0.25 (few event occurrences)	0.5 (medium event occurrences)	0.75 (many event occurrences)
β	0.25 (low effect size for experimental group)	0.5 (medium effect size)	1 (high effect size)
τ	0.5 (event occurrence concentrated towards beginning of the study)	1 (event occurrence constant throughout the study)	2 (event occurrence concentrated towards end of study)
n	100 (small group size)	200 (medium group size)	500 (large group size)

The hazard is the probability of event occurrence during each time period. Since survival analysis only follows participants until the event occurs, the hazard probability is a conditional probability: it is the risk that an event occurs, given that it has not already occurred. The set of hazard probabilities as a function of time is called the hazard function. This function can be used to identify hazardous periods and to find out whether the probability of event occurrence changes over time[1].

The hazard function is often written as a function of a predictor. The hazard function is written as a logit function, which is the natural logarithm of the hazard odds. Such a discrete time hazard model is written as[1]:

$$\begin{aligned}\text{logit } h(t_j) &= \log\left(\frac{probability}{1-probability}\right) \\ &= [\alpha_1 D_1 \\ &+ \alpha_2 D_2 + \ldots + \alpha_j D_j] \\ &+ \beta_1 X_1\end{aligned} \quad (1)$$

where D_1, …, D_j are dummies for the different time periods and α_1, …, α_j represent the logit hazard of each time period, while β_1 represents the effect size for predictor X_1 on a logit scale. Note that a positive β results in a larger hazard, thus more participants will experience the event.

The survival function is, unlike the hazard function which analyzes individual time periods, an accumulation of event occurrence over time. As a function of time it represents the proportion of participants that did not yet experience event occurrence. At the beginning of the study, no events have yet occurred, and thus the survival probability is 1. From there on the function can only decrease or remain constant, but never increase.

Data generation

To evaluate strategies that manage missing data, a Monte Carlo Simulation Study was performed. The strategies were tested in 81 scenarios, based on the parameters ω (proportion of control group that experiences the event), τ (concentration of event occurrence), β (effect size of the experimental group on a logit scale) and n (total number of subjects). Table 1 shows the different values of the parameters. Each scenario contained twelve time periods during which the event could occur. In this simulation study only one predictor (control group vs. treatment condition) was used.

The datasets were created using a simulation program written specifically for this study. The simulation was performed using R [8] and the package MASS[9].

Simulation

For each of the scenarios a total of 2000 datasets were generated. In each iteration, data was removed at random. For each subject one period was removed. Each period had an equal probability to become missing. In some cases, the missing period comes after event occurrence. Since in survival analysis subjects are only followed up upon until the event occurs, this effectively means that these subjects have no missing period. Therefore, the subjects that experience the event later in the study have a higher risk of having a missing period. The missing data are therefore Missing At Random (MAR).

Strategies

Seven strategies to manage the missing data were evaluated:

1. *Case deletion (CD).* This strategy deletes all cases (subjects) with missing data.
2. *Deletion upon missing (DUM).* Subjects are deleted from the dataset once a missing period occurs, but the data up until the missing period is kept in the analysis.
3. *Single imputation (SI).* The missing valuable is replaced by a random but plausible value[10]
4. *Multiple imputation (MI).* Multiple datasets are created. In each dataset, the random but plausible valuable is imputed. The datasets are then analyzed and combined. Multiple imputation is considered the golden standard in handling missing data[7].
5. *Remembrance (R).* The subject is asked during the next time period whether he/she remembers whether the event occurred during the previous, missing period. The subject has a chance to answer incorrectly. In this study, the subject has a forty percent probability of being correct.
6. *Non-Event-Strategy (NES).* Each missing period is considered as if the event did not occur that period.
7. *Event-Strategy (ES).* Each missing period is considered as if the event did occur that period.

All seven strategies were tested on the same datasets.

Criteria for evaluation

The results of each strategy are compared to the original dataset without missings and evaluated using four criteria: relative effect size bias, relative standard error, power, and coverage rate for the confidence intervals of treatment effect. The first criterion calculates the percentage relative bias of effect size:

$$Relative\ effect\ size\ = 100\% * \frac{\hat{\beta}-\beta}{\beta}. \quad (2)$$

This equation calculates the difference between the effect size estimate $\hat{\beta}$ and the true effect size β as a percentage. A value of zero percent is considered ideal. A negative

percentage indicates an underestimation of the effect size, increasing the risk of a type II error (failure to reject a false null hypothesis (H_0), a 'false negative'), whereas a positive percentage indicates an overestimation of the effect size and therefore increases the risk of a type I error (rejection of a true H_0, a 'false positive'). The estimated effect size is the average estimation of 2000 generated datasets.

The second criterion estimates the percentage relative bias of standard error (SE) using the equation:

$$Relative\ SE\ = 100\% * \frac{\overline{\widehat{SE}\left(\hat{\beta}\right)} - SD(\hat{\beta})}{SD(\hat{\beta})}. \qquad (3)$$

The standard error is the average estimated standard error of 2000 replications. The difference between the standard error of the estimated effect size $\hat{\beta}$ and the standard deviation of the estimates of $\hat{\beta}$ is calculated and then expressed as a percentage of the standard deviation of the estimates of $\hat{\beta}$. Overestimation of the standard error may lead to not detecting significant effects, while an underestimation may lead to overstatement of significant effects[11].

Power, the third criterion, is the probability that a statistical test finds an effect, given that this effect exists [12]. It measures the proportion of datasets that correctly reject the H_0 using $\alpha = 0.05$ (the probability of a type I error).

The fourth criterion is the coverage rate. Coverage is the proportion of generated datasets for which the 95% confidence interval contains the actual, true parameter value[10].

There is no consensus on what is 'acceptable' bias. Muthén and Muthén [11] suggest a 10% limit for parameter and standard error bias and a 5% limit for the standard error of the parameter for which power is being assessed. They also suggest that the coverage rate should be between 0.91 and 0.98. Schafer and Graham[10] believe that bias becomes problematic when it's absolute size is greater than one half of the estimates standard error, or when coverage is below 0.90. In this study, the Muthén and Muthén criteria are used.

RESULTS

The estimates of the effect size β and its standard error were very extreme in eighteen scenarios for strategies *CD* and *DUM*. These were all scenarios with a low event occurrence, event occurrence concentrated towards the end of the study and small group sizes. These estimates are likely based on a mathematical error due to one or more of the periods in these scenarios not having any events. Analysis of the individual iterations of these scenarios showed a small amount of iterations with extreme estimates. These iterations were then deleted for all seven strategies within the same scenario. This aids the comparison between strategies within one scenario.

Convergence was achieved for all datasets in all scenarios.

Relative Effect Size Bias

Strategy *DUM*, *SI*, *MI* and *NES* produce acceptable mean effect size biases when averaged over all scenarios, given the 10% limit proposed by Muthén and Muthén (2009). Of these strategies, strategy *NES* has the smallest mean bias (0.64%), followed by strategy *MI* (-0.86%), then *SI* (2.02%), then *DUM* (3.20%). Although on average strategies *DUM*, *SI*, *MI* and *NES* produce similar mean biases equal to 0%, strategy *MI* has a larger bias than the other three strategies in the scenarios where $\omega = 0.25$, $\tau = 2$ and $\beta = 0.25$, and where $\omega = 0.25$, $\tau = 2$ and $\beta = 0.5$ (Scenarios with little event occurrence, where event occurrence is concentrated towards the end of the study, and where there is a small effect size). Strategy *CD*, *R* and *ES* all show a large, negative biases, respectively -77.72%, -48.71% and -60.77%.

In each strategy, the bias decreases as event occurrence ω and effect size β increase. This effect is stronger for ω than for β. As event occurrence concentrates towards the end of the trial (τ increases), the bias increases as well.

Relative Standard Error Bias

Strategy *SI* and *NES* have an acceptable mean bias when averaged over all scenarios (under 5%[11]) for the standard error (0.77% and 3.00%, respectively) while strategy *MI* is close to this limit (6.67%). Strategy *CD* shows the worst result (74.27%), followed by strategy *DUM* (41.99%), *ES* (-19.57%) and *R* (-16.75%). Strategy *R* and *ES* both underestimate the standard error equally. Strategy *CD* overestimates the standard error by almost double the amount that strategy *DUM* does.

For each strategy, the bias comes closer to 0% for a larger ω (higher event occurrence) and a lower τ (concentration of event occurrence towards the beginning of the study). A larger β has only a marginal effect on decreasing the absolute value of the bias. There seems to be no interaction effect between ω, β and τ.

Power

Strategy *SI*, *NES* and *MI* each have a power almost indistinguishable from the power of the complete dataset). Strategy *CD*, *DUM*, *R* and *ES* have a considerably lower power.

Across the scenarios, the power of strategies *SI*, *MI* and *NES* remains about equal to the power of the complete dataset for different values of ω. Strategy *CD*, *DUM*, *R* and *SE* have a decreasing power as ω (event occurrence) increases. The parameters τ and β each have a positive main effect on the power of the different strategies: as these parameters increase (concentration of events towards the end of the study and increasing effect size), so does the power. This effect has an equal size for the benchmark and all strategies except *CD*. For strategy *DUM*, *R* and *ES*, when $\beta = 1$, the decrease in power that goes with an increase in ω, is much larger than the decrease in power when $\beta = 0.25$.

Coverage Rate

Using the coverage rate suggested by Muthén and Muthén (2009), strategies *DUM*, *SI*, *MI* and *NES* have overall acceptable coverage rates (0.95, 0.93, 0.96 and 0.95, respectively). Strategies *CD*, *R* and *ES* have low mean coverage rates (0.68, 0.61 and 0.50, respectively). Across all scenarios, strategy *DUM*, *SI*, *MI* and *NES* are hard to distinguish from each other.

As τ and β decrease, the coverage increases. As ω increases, the coverage rate also increases.

CONCLUSION

Single imputation (*SI*), multiple imputation (*MI*) and the Non-Event-Strategy (*NES*) perform equally well for each criterion of evaluation. Of these three, single imputation performed slightly better than the other strategies, and if the researcher has the statistical knowledge then it is certainly the best strategy to use. The Non-Event Strategy is the easiest strategy to implement however, since it requires no extended knowledge of statistics. It is both a fast and easy method. There are no specific scenarios where one strategy stands out.

The available literature on missing data in other scientific fields suggests that single and multiple imputation are good strategies. Multiple imputation is generally preferred over single imputation[7,10], but this was not found in this study. The Non-Event-Strategy produces acceptable results as well. This can be explained by the fact that since the probability that the event occurs during the specific missing time period is low, even if throughout the trial event occurrence is high. Only if event occurrence would be extremely high during each period, then this strategy would not work. Most subjects would then experience the event within very few periods however, and it can then be argued that the study should either be using smaller time periods or use a continuous time survival analysis. There is no literature available to either validate or invalidate this strategy, since it was designed specifically for this study.

The results also showed that as τ increases (meaning event occurrence concentrates towards the end of the study), the bias increases as well. This is logical, since in this study, each participant would have one missing period. Because survival analysis only observes people that have not yet experienced the event, only the missing periods before an event can produce bias. Missing periods following an event do not produce bias, since those periods are not analyzed. As event occurrence shifts towards the end of the study, more missing periods become actual missings and as such the bias increases. The results also show that as ω (proportion of participants in control group that experiences the event) increases, the bias decreases. The logic behind this follows the same line of reasoning as for τ: as ω increases, and thus more participants experience the event, the probability that a missing period comes after an event increases. As therefore the number of actual missing periods decreases, the bias decreases as well.

Although the fact that hazard can change over time was factored in, not all possible scenarios were considered. For instance, the hazard rate can be high in the middle of a trial, or have multiple peaks and troughs throughout the study. This was not tested for in these scenarios. However, since single and multiple imputation and the non-event strategy stood out from the other four strategies in all 81 scenarios, it is expected that this also holds in other scenarios. Another limitation was that each subject had a 40% probability of remembering event occurrence correctly. This is an arbitrary percentage, which will be different for various studies. As this probability increases, this strategy is expected to obtain better results. Future research could focus on at what probability this strategy starts to produce acceptable results. A final limitation could be that due to the way missingness was introduced, not all scenarios contained the same amount of missingness. Although larger biases were found in scenarios with more missings, these larger biases were found for all strategies. This did not seem to affect the hierarchy of strategies.

ROLE OF THE STUDENT

Nils van de Ven was an undergraduate student working under the supervision of Mirjam Moerbeek and Maryam Safarkhani. The topic was chosen by Nils from a list of available topics. The R script of the simulation was handed to the student. The different strategies were thought of in collaboration between the student and the supervisors. The analyzation of the results, formulation of the conclusion and the writing were done by the student.

REFERENCES

1. Singer JD, Willett JB. *Applied Longitudinal Data Analysis*. New York: Oxford University Press; 2003.
2. Kadden RM, Cooney NL, Getter H, Litt MD. Matching alcoholics to coping skills or interactional therapies: posttreatment results. *J Consult Clin Psychol*. 1991;59(4):598-601. doi:10.1037/0022-006X.57.6.698.
3. MacMillan HL, Thomas BH, Jamieson E, et al. Effectiveness of home visitation by public-health nurses in prevention of the recurrence of child physical abuse and neglect: A randomised controlled trial. *Lancet*. 2005;365(9473):1786-1793.
4. Moerbeek M, Schormans J. The Effect of Discretizing Survival Times in Randomized Controlled Trials. *Methodol Eur J Res Methods Behav Soc Sci*. 2015;1(2002):1-10.
5. van Buuren S. *Flexible Imputation of Missing Data*. Boca Raton: CRC Press; 2012.
6. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley; 1987.
7. Treiman DJ. *Quantative Data Analysis*. First Edit. San Francisco: Jossey-Bass; 2008.
8. R Development Core Team. R: A language and Environment for Statistical Computing. 2013. http://ww.r-project.rog.
9. Venables WN, Ripley BD. Modern Applied Statistics with S. 2002. http://www.stats.ox.ac.uk/pub/MASS4.
10. Schafer JL, Graham JW. Missing data: Our view of the state of the art. *Psychol Methods*. 2002;7(2):147-177.
11. Muthén LK, Muthén BO. How to Use a Monte Carlo Study to Decide on Sample Size and Determine How to Use a Monte Carlo Study to Decide on Sample Size and Determine Power. *Struct Equ Model A Multidiscip J*. 2009;9(4):599-620..
12. Field A. *Discovering Statistics Using IBM SPSS Statistics*. 4th editio. London: Sage; 2013.