OA Epidemiology
Open Access

Jóźwiak K, Moerbeek M. Optimal designs for trials with discrete-time survival endpoints. OA Epidemiology 2013 Dec 30;1(3):22.

This is a provisional PDF file.

Licensee OA Publishing London 2013. Creative Commons Attribution License (CC-BY)

Competing interests: none declared. Conflict of interests: none declared.
All authors contributed to conception and design, manuscript preparation, read and approved the final manuscript.
All authors abide by the Association for Medical Ethics (AME) ethical rules of disclosure.

Section: Systematic Reviews

Optimal designs for trials with discrete-time survival endpoints

K Jóźwiak, M Moerbeek

Utrecht University, The Netherlands

Acknowledgements:

This research was funded by a VIDI grant from the Netherlands Organization for Scientific Research (NWO) number 452-08-004.

Correspondence to:

K. Jóźwiak. Department of Epidemiology and Biostatistics. The Netherlands Cancer Institute. PO Box 90203. 1006 BE Amsterdam. The Netherlands. Phone: +31-20-5129074. K.jozwiak@nki.nl

Abstract

Randomized controlled trials require a lot of labor, time, and money. To not waste these efforts and to get scientifically useful results, they need to be carefully planned beforehand. There has been a lack of knowledge on planning designs of longitudinal experiments where events are observed in discrete time, but recently some relevant articles about that topic have been published. In experiments of that type subjects are observed in time periods and a design is a combination of the total number of subjects, the number of time periods they are observed for, and the allocation of subjects to the control and experimental groups. A good design is the combination of these factors that guarantees a sufficient level of statistical power at minimal costs. The optimal design methodology is introduced on a basic level and an example shows that designs based on standard calculations are not necessarily the best choice.

Key words: discrete-time longitudinal data; survival analysis; experimental studies; optimal design; statistical power

Introduction

In many fields of science researchers study a particular event in peoples' life as they are interested in the timing and/or causation of event occurrence, such as recovery from anorexia nervosa [1]. A method to identify and investigate such events, is to follow a group of subjects across time in a longitudinal study until they either experience the event or drop out from the study due to other reasons.

In the planning phase of a study it is important to select an appropriate metric of time, that is, a meaningful scale of time in which the event can occur. Time can be measured continuously or discretely. Time is measured continuously when the precise timing of event occurrence is known, and it is measured discretely when the event is measured at preselected points in time or in time intervals. Measuring discretely rather than continuously implies the exact time of event occurrence is unknown and so it results in a loss of information. The choice for measuring time discretely rather than continuously should therefore be justified well [2]. In retrospective studies time is often measured discretely since subjects may not be able to remember the exact time of event occurrence. In prospective studies the budget may be limited and endpoints are assessed periodically. In addition, measuring too frequently may cause participation burden and increased rates of attrition and may also raise ethical objections, especially when taking a measurement puts a high burden on the participants. In epidemiology discrete-time survival analysis is used in studies on the onset of alcohol and tobacco use and in studies on disease progress and relapse.

The metric of time is an important choice since it has an influence on the sample size to achieve a prespecified power level. Here we focus on the power to detect treatment effects in randomized controlled trials that compare one experimental condition to a control [3,4]. To conduct longitudinal trials where people are followed over time, a lot of labor, time, and money is needed. Therefore, trials need to be carefully planned before they are actually conducted, and the best way is to calculate the optimal design in the planning phase of a trial. The optimal design is the one among all designs that performs best with respect to some optimality criterion, for instance the power of the test on treatment effect. Power is related to

sample size and sufficient sample size ensures an effect of treatment is detected with large probability.

To plan a trial, researchers need to consider many aspects, for instance, how many participants should be recruited, how many of them should be randomized to the experimental condition and for how long they should be followed up. Guidelines on sample sizes and optimal designs for trials with continuous-time survival data have been studied in the past [5-9,10-12] and are implemented in software packages such as nQuery Advisor 7.0 [13], PASS [14] or Study Size [15]. These guidelines cannot be used for trials with discrete-time survival data because another metric of time is used. When subjects are measured discretely an important question is for how many time intervals they should be measured. This question is irrelevant for trials with continuous-time survival data because subjects are followed continuously rather than measured at some preselected points in time. In addition to that, the derivation of the optimal design and sample sizes calculations should be based on the model that is used for analysis once the data are collected. The Cox model is commonly used for continuous-time survival data [16] while discrete-time survival data are commonly analyzed by means of a generalized linear model [2], so this gives another reason why sample size calculations for trials with continuous-time survival data cannot be used for trials with survival endpoints in discrete time.

The discrete-time survival approach was proposed by Cox [16] in the early seventies, discussed by Kalbfleisch and Prentice [17] and has been very well presented by Singer and Willett [2,18,19] in the early nineties. Recommendations for the design of trials with discrete-time survival endpoints have only been published recently, such as power analysis for simple randomized trials [20] and cluster randomized trials [21], and the optimal number of time periods and

allocation ratios [22-24]. These papers appeared in statistical journals and may not have drawn the attention of many epidemiologists while we believe the findings in these studies are also of importance to the audience of *Open Access Epidemiology*. The purpose of this contribution is to summarize the optimal design methodology by Jóźwiak and Moerbeek on a basic level. We do not include statistical equations or mathematical derivations but explain the rationale behind the optimal designs on a basic level. Our PODSE[25] software can be used to actually calculate the optimal design of a trial without any knowledge of the underlying matrix algebra. An example shows that the optimal design does not necessarily require the control and experimental group to be of equal size. Also, it shows that a design with fewer time points than initially planned may be more efficient in terms of estimating and testing the treatment effect.

Optimal Design

In the current approach it is assumed that subjects are followed over multiple time periods that have equal and fixed length. Subjects may be followed over a prespecified maximum number of time periods or they may be followed for a shorter amount of time. For instance, researchers want to conduct a study that lasts at most *12* months so they can observe subjects for *1*, *2*, ..., or *12* months. Measurements are taken at the beginning of the study and at the end of each time period but the particular event may occur at any time between the measurement points. So, time is measured discretely but the underlying process is continuous. Moreover, participants are randomized to one of two treatment groups, a control and an experimental group, and these groups do not need to be equally sized at the beginning of the study.

In each time period the event status is recorded for all subjects observed in that particular period. So, the data includes multiple records for each subject until and including the time interval the subject either experiences the event or gets censored. For such data the conditional hazard probability of event occurrence in each time period can be modeled by applying a generalized linear model (GLM), such as a logistic regression model [2]. To calculate an optimal design, it is necessary to know the values of the hazard probabilities in the control group.

The parameters of the statistical model and the asymptotic covariance matrix are estimated by iteratively re-weighted least squares [26]. The last entry in the covariance matrix is the variance of the treatment effect estimator and its value depends on the survival probabilities in both treatment groups, the total number of recruited subjects in a trial (N), the number of time periods (p), the proportion of subjects recruited in the experimental group (π_E) , and the attrition rate that is the proportion of subjects who leave the study between any two adjacent measurement occasions due to unforeseen reasons. A combination of the three quantities, (N, p, π_E) , is called a design and should be defined in the planning phase of a study. In some cases one of the quantities is fixed beforehand and then the other two need to be calculated. For example, researchers may have limited time for conducting a trial or there is a limited number of participants that can enter a trial. However, when none of the quantities is fixed or limited, researchers need to decide about a good combination of all three quantities.

The variance of the treatment effect estimator is used to calculate the statistical power for the test on treatment effect and it is common knowledge that the power is maximized for minimized variance. Since the main interest of an experiment is to find an effect of the treatment with sufficient probability, the minimal variance of the treatment effect is used as

optimality criterion. However, different designs can give the same statistical power and therefore, a cost constraint is also considered while choosing a good design. The cost of the trial can be calculated in four different ways and for each way, a separate cost function has been defined. The first cost function is used when all subjects are followed till the end of the trial even if they have already experienced the target event. It can be used, for instance, when researchers are interested in measuring not only the target event but also some other outcomes. The second cost function is used when all subjects are meant to be followed till the end of the trial but some of them leave the study due to unforeseen reasons, like moving out of town. So, this cost function takes attrition into account. The third cost function is used when subjects leave the study after experiencing the event and attrition is absent, while the fourth cost function is used when subjects leave the study after experiencing the event and some attrition may be present. Each cost function is a function of the number of recruited subjects, the number of subjects observed in each time period and two cost ratios. The first cost ratio compares the costs of recruitment in both treatment groups (f_1) and the second cost ratio relates the costs of recruiting one subject in a control group to the costs of an additional repeated measurement of one subject (f_2). The optimal design is the combination of the total number of recruited subjects, the number of time periods, and the proportion of subjects recruited in the experimental group that gives a sufficient power level at minimal cost or maximizes the power for a given budget. If the optimal design cannot be used in the trial at hand or if researchers do not want to conduct it for whatever reason, an alternative design can be found using the relative efficiency measure. This measure compares the optimal design to any alternative design and shows how often the alternative design should be replicated to be as efficient as the optimal design. When relative efficiencies are larger than *0.8* or *0.9*, the alternative design is highly efficient.

Using the presented approach only locally optimal designs are found because they are calculated for given values of the survival probabilities, the size of the effect treatment, the maximum number of time periods, and the cost ratios. The formulas for calculating such designs are derived using matrix algebra, but they are complicated and cannot be easily used by researchers planning a trial. To help finding the optimal design, the user-friendly interface computer program PODSE [25] has been created. In this program locally optimal designs are found, relative efficiencies of alternative designs are plotted, the statistical power is calculated for a given budget and the cost of the study is obtained for a given desired power level. In such a way, the user can evaluate few alternative designs for a certain trial and decide which one is the best. To find the optimal design the user needs to specify the underlying survival function and the size of the treatment effect. Also, the cost function and costs ratios f_1 and f_2 need to be specified. The PODSE software is a standalone executable program that can be freely downloaded from the following internet site: http://tinyurl.com/discrete-survival. A more extensive introduction to this program, including the matrix algebra that is used within the program, can be found in another paper of ours[25].

An application example: anorexia nervosa recovery

In modern society many people live with serious psychiatric disorders. One of them is anorexia nervosa that is characterized, among others, by extreme food restriction, an abnormally low body weight and intense fear of weight gain. All people suffering from anorexia should get medical and/or psychiatric treatment and there are many studies on how to help patients with this disorder. One study discussed by Herzog *et al*. [1] investigates recovery from anorexia nervosa. In this research a group of patients were hospitalized and

admitted to an inpatient treatment program that consisted of individual psychotherapy with behavioral and psychodynamic elements, group psychotherapy and consulting a social worker. However, after the treatment patients exhibit a low recovery rate.

Let us assume there is a new treatment for anorexia that increases recovery rates. This experimental treatment can be compared with the treatment presented by Herzog *et al.*[1] in a randomized controlled trial. The survival and hazard curves illustrated in their article can be used as survival and hazard curves for the control group. Moreover, let us assume that the experimental treatment has an effect of size *0.5* on the logit scale. The survival and hazard probabilities in both groups are shown in Figure 1. The rate of recovery at eleven years is *67%* and *83%* for the control and experimental condition, respectively. The chance of recovery is highest after 2 years and smallest after 9 years of admission to the program, and it is higher for people receiving the experimental condition.

Suppose patients do not leave the study after experiencing a recovery and are further observed to monitor their health. Since it is a longitudinal study, most likely there will be patients who leave the study due to other, unforeseen reasons and suppose in each year the attrition rate is 1%. To investigate the design of such an experiment we use the PODSE [25] program and present our results in Table 1. The paper by Herzog *et al.* [1] does not provide any information about the costs of recruiting and measuring participants, so we calculate optimal designs for four plausible combinations of the cost ratios: (f_1, f_2) = (5, 2), (5, 5), (10, 2), (10, 10) with fixing the value of the cost of recruiting one patient in the control group to 1 and using the second cost function. Both cost ratios are larger than 1, which implies the costs for a subject in the experimental condition are higher than those in the control and the costs to

include a subject are higher than the costs to take a measurement. We believe such cost ratios are plausible for the experiment at hand.

Table 1 lists the optimal number of time periods *p* and the proportion of subjects who are allocated to the experiment condition π_E for each pair (f_1, f_2). We observe the allocation proportion π_E decreases when the cost ratio f_1 increases. This result is intuitively sound since it implies fewer subjects are allocated to the experimental group when the costs per subject in the experimental group increase relative to those in the control group. For instance, 39% of the subjects are randomized to the experimental group when $f_1 = 5$, but 35% are randomized to this condition when this cost ratio increases to $f_1 = 10$ (while fixing f_2 to 2). Furthermore, we observe the optimal number of time periods *p* increases when at least one cost ratio increases. This shows that the optimal number of time periods decreases when it becomes more expensive to take a measurement (i.e. when the cost ratio f_2 decreases). This result is also intuitively sound since it makes sense to take fewer measurements and to sample more subjects when it becomes more expensive to take a measurement.

Table 1 also gives the required number of subjects *N* and related costs to achieve power levels of 0.8 or 0.9. As one can expect, more subjects are needed to achieve a higher power level and this results in higher trial costs. It should be mentioned that the costs do not only depend on the number of subjects *N* but also on the allocation ratio π_E and number of time periods *p*. When (f_1, f_2) = (5, 2) a total of 268 subjects are needed and the costs are equal to 1598. When (f_1, f_2) = (5, 5) slightly fewer subjects are needed but the costs decrease much since the proportion of subjects assigned to the experimental condition decreases and the number of time periods increases.

If researchers prefer an alternative design instead of the optimal one, the costs of the trial will be higher and the total number of patients will be smaller. Figure 2 shows relative efficiencies for alternative designs when $(f_1, f_2) = (5, 2)$. The optimal number of time periods is 6 and the optimal proportion of allocation to the experimental group is 0.39, but a trial with 6 time periods and equally sized groups has high efficiency. The costs of this trial slightly increase to 1663 and the number of patients slightly decreases to 260. However, if the follow-up time is prolonged then these two quantities change even more. For example, if two equally sized groups are followed up for 8 years then 232 patients can be recruited and the costs are equal to 1699, but if they are followed up for 10 years then 218 patients can be recruited and the costs are 1795.

Discussion and Conclusion

It is important to study optimal designs for trials with discrete-time survival endpoints since in many experiments time cannot be measured continuously. Therefore, a large research project is currently investigating designs for such trials. The research summarized in this article is just a part of that project and more optimal design recommendations can already be found in the literature, for instance for trials where subjects are recruited at different points in time[27] and for trials with long-term survivors (i.e. subjects who will never experience the event)[28]. We also studied the effect of predictive covariates on statistical power levels[29]. More research is still being conducted and will be presented in the near future.

Until recently sample size formulae and optimal design recommendations were not available for trials with discrete-time survival endpoints. Findings for trials with continuous-time survival endpoints could not be used for reasons already discussed in the introduction. The

example shows that a trial with equal allocation to the experimental and control group is not necessarily the best choice. Also, using fewer time intervals than initially planned may increase efficiency. This illustrates the advantage of our optimal design methodology over standard approaches.

Of course some assumptions should be made before an optimal design can be derived. We did not take into account the possibility of recurrent events, such as multiple occurrences of relapse after periods of recovery in the anorexia example. Also, we ignored the possibility of competing events, such as multiple causes of death in a study on survival from some disease. The choice of the cost function should also be carefully justified as it has an effect on the design. We assumed subjects did not get censored after event occurrence and we allowed for the possibility of drop-out due to other reasons than event occurrence. Measuring after event occurrence implies increasing costs and taking drop-out into account means the optimal number of time intervals increases[22].

It should furthermore be mentioned that with the proposed methodology locally optimal designs are calculated. In other words, the optimal designs depend on the values of the model parameters, that is, the underlying survival function and treatment effect. In practice these values are often unknown and a prior guess could be based on expert knowledge or findings in the literature. An incorrect guess, however, may result in a loss of efficiency. Robust optimal designs, such as maximin or Bayesian optimal designs, may overcome this problem. With maximin designs a plausible range of all model parameters is specified and the maximin design is the one among all possible designs that maximizes the minimal relative efficiency over all plausible parameter values. Maximin designs have been proposed and studied in the context longitudinal studies beforehand[30]. With Bayesian optimal designs a prior distribution,

rather than a range, is used for all model parameters and the optimal design is found over this prior distribution.

References

1. Herzog W, Schellber D, Deter H-C. First recovery in anorexia nervosa patients in the long-term course: a discrete-time survival analysis. Journal of Consulting and Clinical Psychology. 1997; 65(1): 169-177.

2. Singer JD, Willett JB. Applied longitudinal data analysis. New York, NY: Oxford University Press; 2003.

3. Jadad AR, Enkin M. Randomized controlled trials: Questions, answers and musings. Malden: Blackwell; 2007.

4. Torgerson DJ, Torgerson CJ. Designing randomised trials in health, education and the social sciences: An introduction. Houndmills: Palgrave Macmillan; 2008.

5. Collet D. Modelling survival data in medical research. Boca Raton, FL: Chapman &Hall/CRC; 2003.

6. Machin D, Cheung YB, Parmar M. Survival analysis. A practical approach. West Sussex, England: Wiley; 2006.

7. Chow SC, Shao J, Wang H. Sample size calculations in clinical research. Boca Raton, FL: Chapman & Hall/CRC Biostatistics Series; 2008.

8. Hosmer DW, Lemeshow S, May S. Applied survival analysis. Regression modeling of time-to-event data. New York, NY: Wiley; 2008.

9. Julious SA. Sample sizes for clinical trials. Boca Raton, FL: Chapman & Hall/CRC; 2010.

10. Nam JM. Optimum sample sizes for the comparison of the control and treatment. Biometrics. 1973; 29: 101-108.

11. Gross AJ, Hunt HH, Cantor AB, Clark BC. Sample size determination in clinical trials with an emphasis on exponentially distributed responses. Biometrics. 1987; 43: 875-883.

12. McGree JM, Eccleston JA. Investigating design for survival models. Metrika. 2010; 72: 295-311.

13. Elashoff JD. nQuery Advisor version 7.0 users guide. Los Angeles, CA. http://www.statsolusa.com ; 2007.

14. Hintze DJL. Quick start manual. PASS Power Analysis and Sample Size System. NCSS, Kaysville, Utah; 2008.

15. CreoStat. Study Size 2.0. Software for analysis of sample size and power in study design. Retrieved from http://www.studysize.com/.

16. Cox DR. Regression models and life-tables. Journal of the Royal Statistical Society Series B (Methodological). 1972; 34(2): 187-220.

17. Kalbeisch JD, Prentice RL. The statistical analysis of failure time data. Hoboken, New Jersey: Wiley and Sons; 2002.

18. Singer JD, Willett JB. It's about time: using discrete-time survival analysis to study duration and the timing of events. Journal of Educational Statistics. 1993; 18(2): 155-195.

19. Willett JB, Singer JD. Investigating onset, cessation, relapse, and recovery: why you should, and how you can use discrete-time survival analysis to examine event occurrence. Journal of Consulting and Clinical Psychology. 1993; 61(6): 952-965.

20. Jóźwiak K, Moerbeek M. Power analysis for trials with discrete-time survival endpoints. Journal of Educational and Behavioral Statistics. 2012; 37(5): 630-654.

21. Moerbeek M. Sample size issues for cluster randomized trials with discrete-time survival endpoints. Methodology. 2012; 8(4): 146-158.

22. Jóźwiak K, Moerbeek M. Cost-effective designs for trials with discrete-time survival endpoints. Computational Statistics and Data Analysis. 2012; 56: 2086-2096.

23. Jóźwiak K, Moerbeek M. Optimal treatment allocation and study duration for trials with discrete-time survival endpoints. Journal of Statistical Planning and Inference. 2013; 143: 971-982.

24. Moerbeek M, Jóźwiak K. Optimal designs for event history analysis. Zeitschrift für Psychologie. 2013; 221(3): 160-173.

25. Jóźwiak K, Moerbeek M. PODSE: A computer program for optimal design of trials with discrete-time survival endpoints. Computer Methods and Programs in Biomedicine. 2013; 111: 115-127.

26. McCullagh P, Nelder JA. Generalized linear models. Boca Raton, FL: Chapman and Hall; 1983.

27. Jóźwiak K, Moerbeek M. Accrual by groups in trials with discrete-time survival endpoints. Clinical Trials. 2012; 10: 32-42.

28. Moerbeek M. Sufficient sample sizes for discrete-time survival analysis mixture models. Structural Equation Modeling. In press.

29. Safarkhani M, Moerbeek, M. Covariate Adjustment Strategy Increases Power in the Randomized Controlled Trial with Discrete-Time Survival Endpoints. Journal of Educational and Behavioral Statistics. 2013; 38(4): 355-380.

30. Ouwens M, Tan F, Berger M. Maximin D-optimal designs for longitudinal mixed effects models. Biometrics. 2002; 58(4): 735-741.

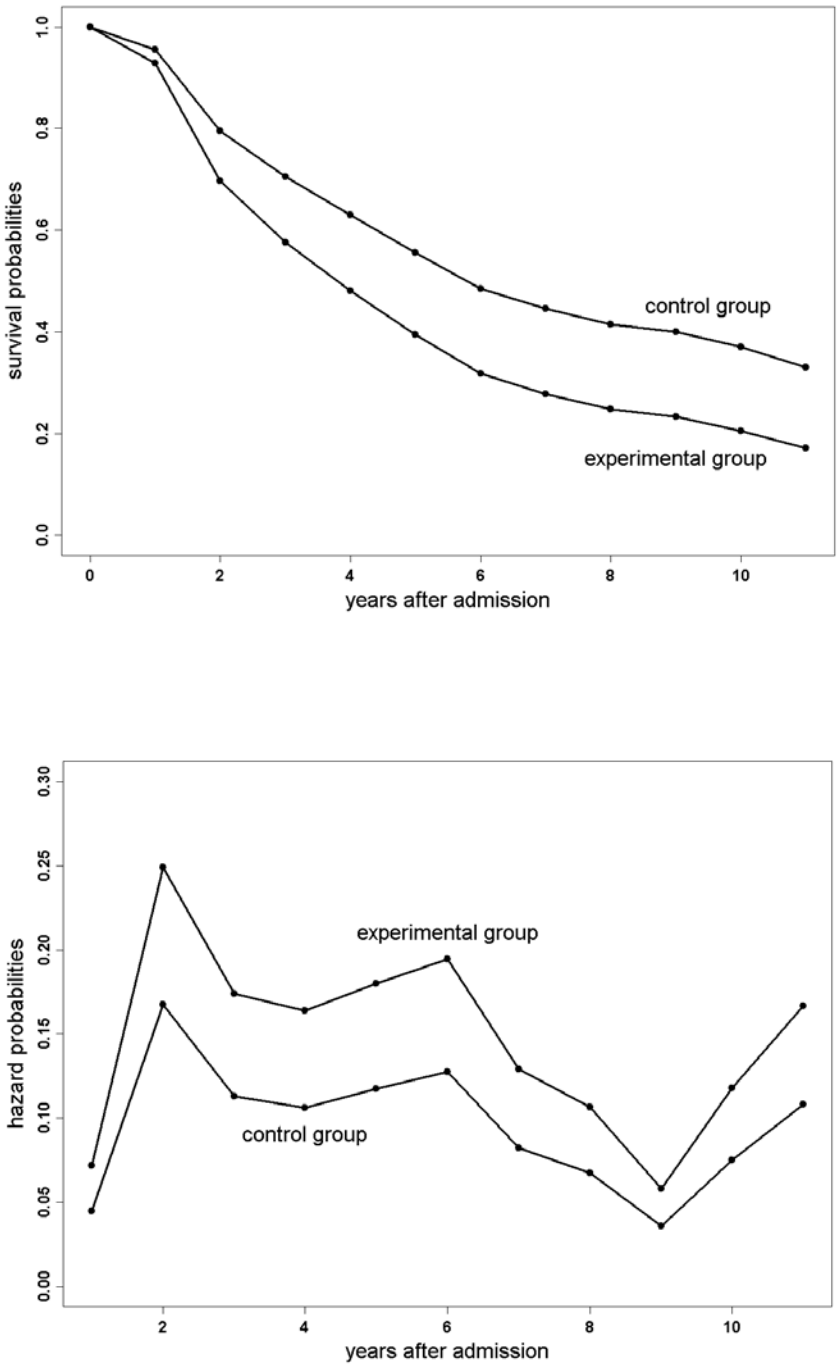

Figure 1. Survival and hazard probabilities of anorexia nervosa recovery in the control and experimental groups.

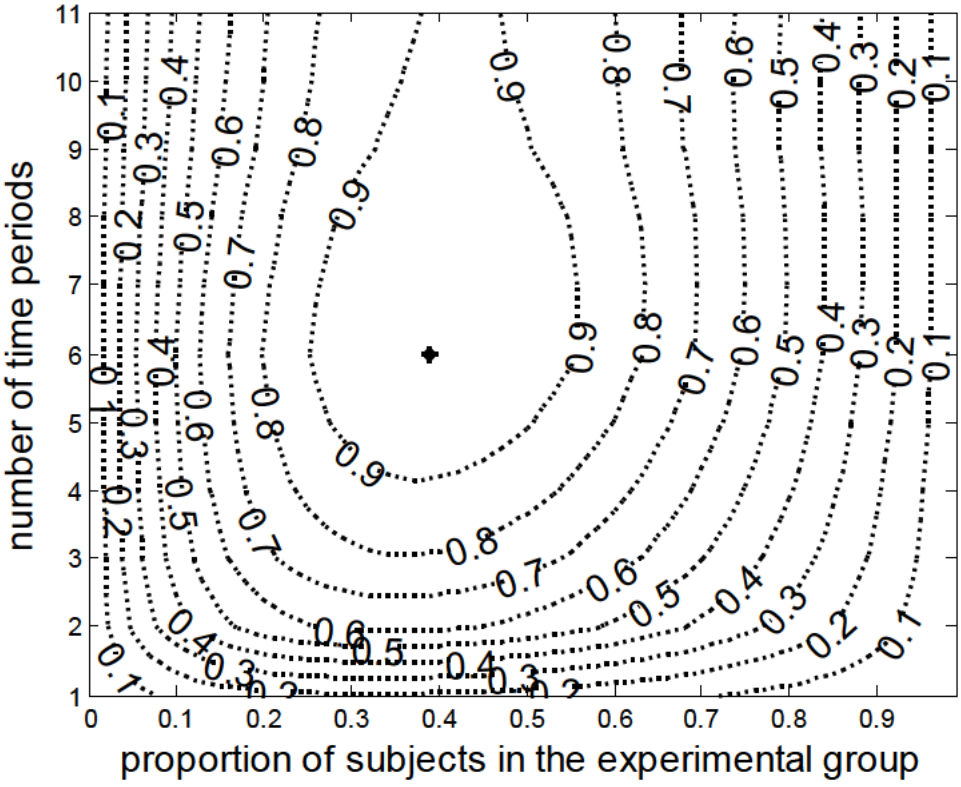

Figure 2. Relative efficiencies for anorexia nervosa recovery example with (f_1, f_2)=(*5,2*), where the dot represents the optimal design and each contour line is a collection of designs (p, π_E) that have the same relative efficiency.

Table 1. Optimal designs (N, p, π_E)=(total number of recruited patients, number of years, proportion of patients in the experimental group) with power of *0.8* or *0.9* and the costs of these designs for varying values of cost ratios f_1(costs of recruiting one patient in experimental group/costs of recruiting one patient in control group), f_2(costs of recruiting one patient in control group/costs of additional measurement per patient) for the anorexia nervosa example.

f_1	f_2	p	π_E	*power = 0.8*		*power = 0.9*	
				N	*cost*	*N*	*cost*
5	2	6	0.39	268	1598	358	2134
5	5	8	0.37	243	1023	325	1367
10	2	7	0.35	259	2078	348	2790
10	5	11	0.32	232	1425	310	1905

Table 1: Table 1.docx

Table 1. Optimal designs (N, p, π_E)=(total number of recruited patients, number of years, proportion of patients in the experimental group) with power of *0.8* or *0.9* and the costs of these designs for varying values of cost ratios f_1(costs of recruiting one patient in experimental group/costs of recruiting one patient in control group), f_2(costs of recruiting one patient in control group/costs of additional measurement per patient) for the anorexia nervosa example.

f_1	f_2	p	π_E	*power = 0.8*		*power = 0.9*	
				N	*cost*	*N*	*cost*
5	2	6	0.39	268	1598	358	2134
5	5	8	0.37	243	1023	325	1367
10	2	7	0.35	259	2078	348	2790
10	5	11	0.32	232	1425	310	1905

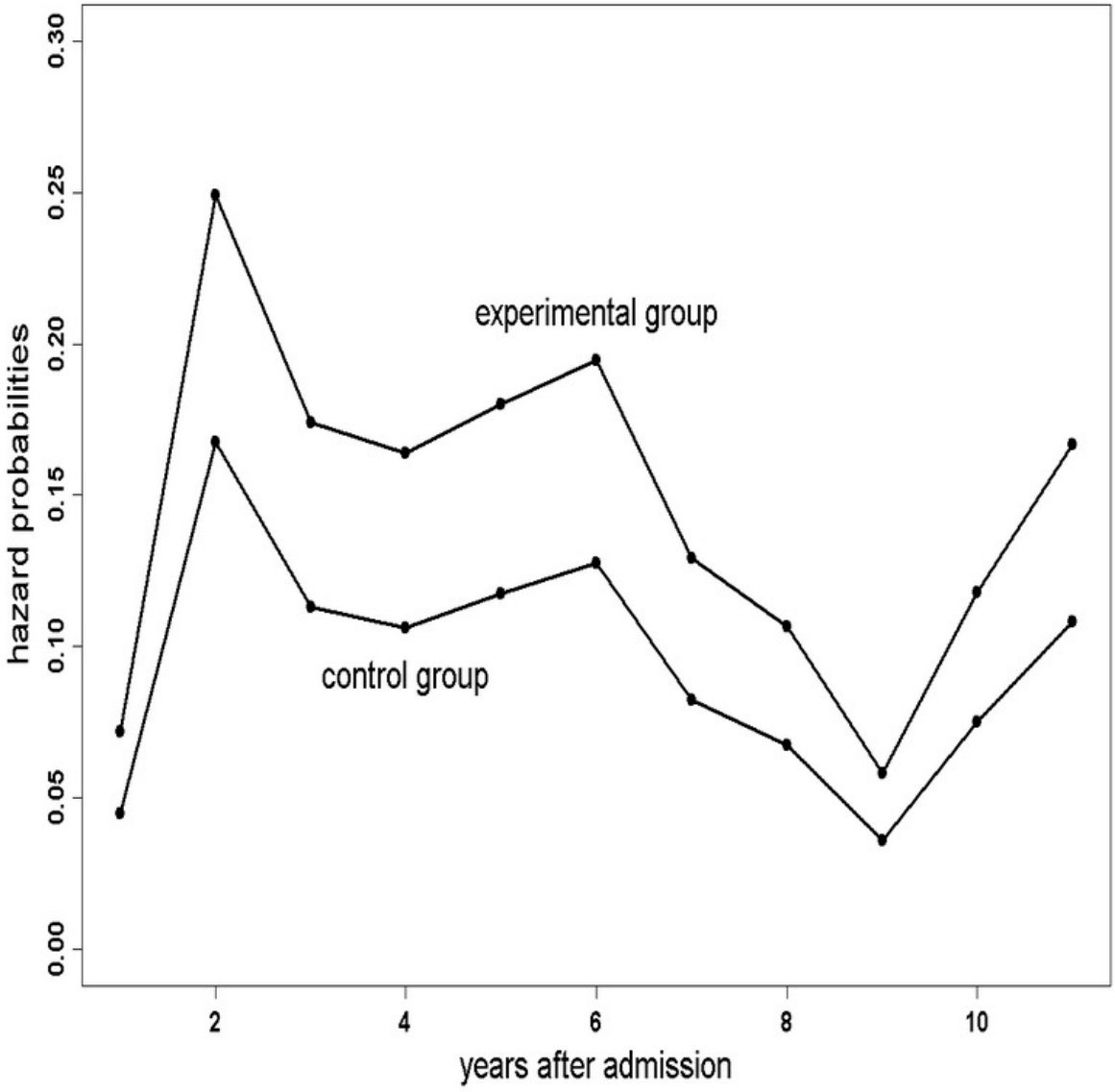

Figure 1: hazard_curves2.TIF

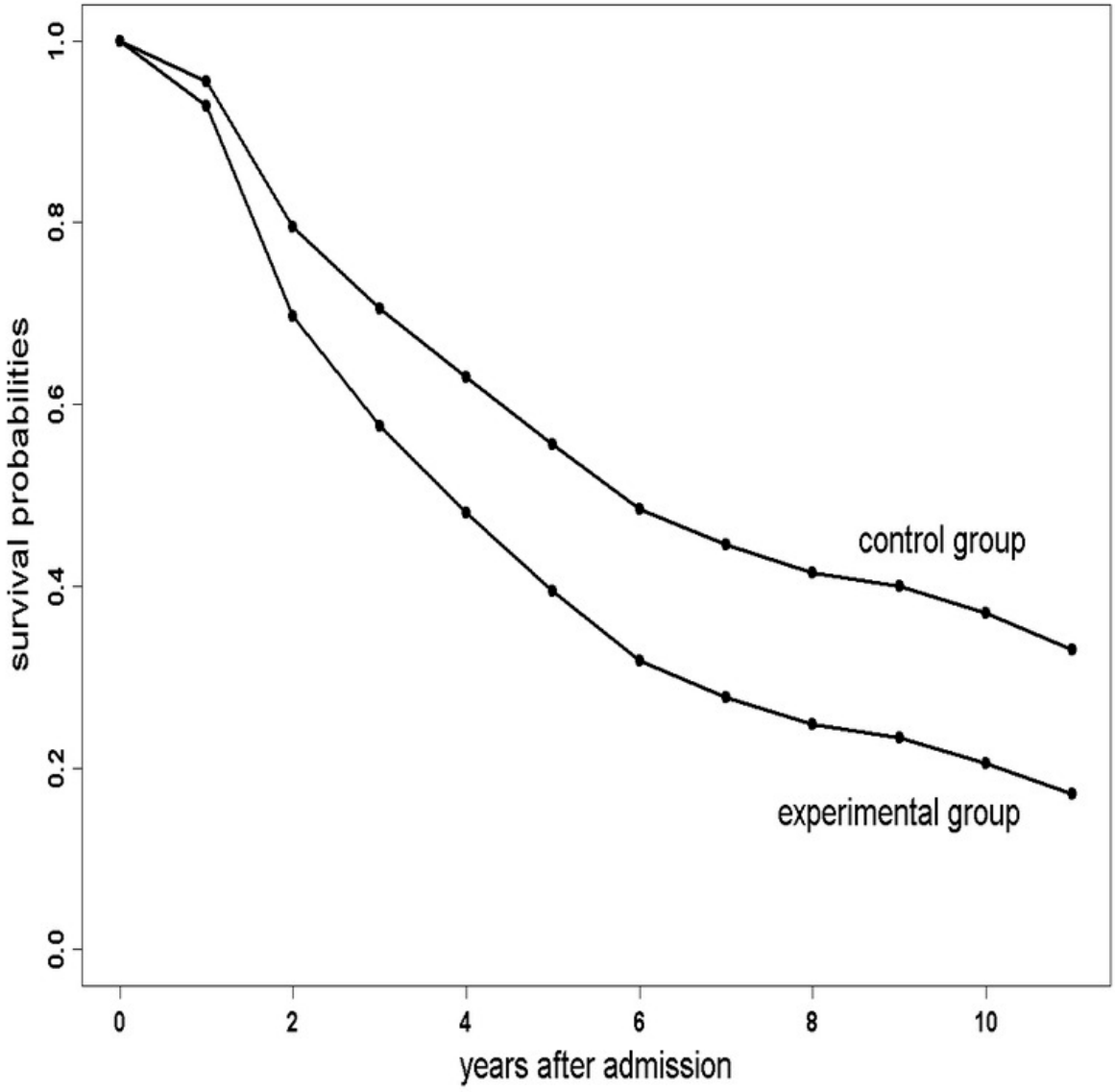

Figure 2: survival_curves2.TIF

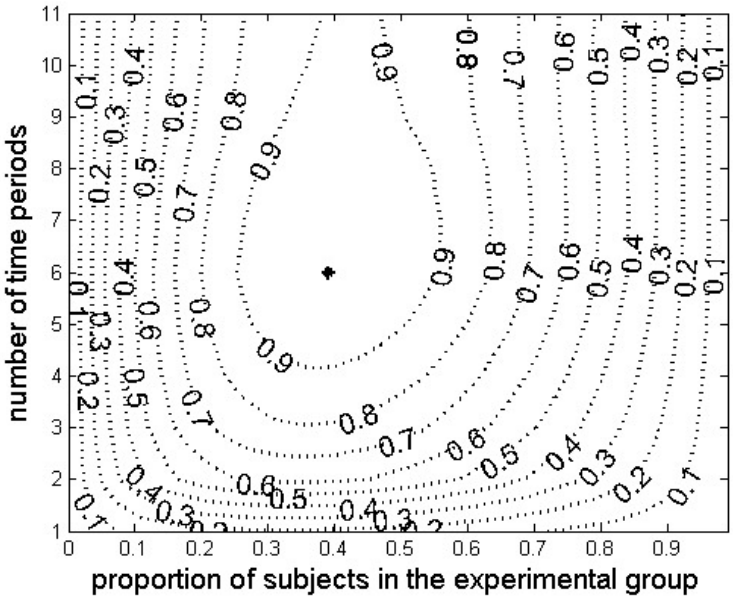

Figure 3: re_plot2.TIF