

Statistical Primer: Basics of survival analysis for the cardiothoracic surgeon

Daniel J.F.M. Thuijs¹, Graeme L. Hickey², Ruben L.J. Osnabrugge¹

¹ Department of Cardiothoracic Surgery, Erasmus University Medical Centre, Rotterdam, the Netherlands

² Department of Biostatistics, Institute of Translational Medicine, University of Liverpool, UK

Address for correspondence:

R.L.J. Osnabrugge
Department of Cardiothoracic Surgery
Erasmus University Medical Centre
's-Gravendijkwal 230
P.O. Box 2040, 3000 CA
Rotterdam, the Netherlands
Tel: 0031 6 11 068 955
Fax: 0031 1 00 703 399
Email: r.osnabrugge@erasmusmc.nl

Short title: How to survive survival analysis?

Meeting presentation: this was presented at the European Association for Cardio-Thoracic Surgery Annual Meeting, Vienna, Austria, 7-11 October 2017.

Word count (full article): 2204

Number of figures/tables: 2 figures

Abstract

Survival analysis incorporates various statistical methods specific to data on time until an event of interest. Whilst the event is often death, giving rise to the phrase ‘survival analysis’, the event might also be, for example, a reoperation. As such, it is sometimes referred to as *time-to-event analysis*.

5 Censoring sets survival analysis apart from other analyses: at the end of the follow-up period, not all subjects have experienced the event of interest, **and some subjects may drop out of the study prior to completion**. Survival data for a group of subjects is usually visualised by the Kaplan-Meier **estimator**, representing the probability of a subject remaining free of the event during follow-up.

10 There are several methods to compare survival between study groups, for example treatment arms, including the log-rank test and the Cox proportional hazards model. The log-rank test is an unadjusted nonparametric method, whereas the Cox proportional hazards model allows comparison while adjusting for multiple covariates. A principal assumption of the Cox proportional hazards model is that the relative hazard stays constant over time—so-called proportionality. Specific methods exist for comparison of survival to the general population. This article describes the fundamental concepts
15 every cardiothoracic surgeon should be aware of when analysing survival data and are illustrated with a clinical example.

Summary word count: 211

Key words: survival analysis, time-to-event analysis, statistics, Kaplan-Meier, Cox proportional
20 hazards.

Introduction

Survival analysis is of great importance for many medical specialties, including cardiothoracic surgery. It helps clinicians to provide valuable prognostic information; for example, on the best
25 treatment option for patients with coronary artery disease (CAD) when deciding between Percutaneous Coronary Intervention (PCI) or Coronary Artery Bypass Grafting (CABG).

A myriad of statistical methods is available when examining survival data. Those most commonly utilised in cardiothoracic research include the Kaplan-Meier estimator, log-rank test, and the Cox proportional hazards regression model [1-3]. To-date, the original papers by Kaplan and
30 Meier and Cox are ranked the 11th and 24th most cited papers of all time in the scientific literature, demonstrating the reach of the survival analysis methodology [4-6]. It is important for cardiothoracic surgeons to understand the fundamentals of these methods in order to adequately appraise and conduct survival analyses. Here, distinguished features of survival data are presented and the statistical methods most frequently used are discussed. To illustrate, this article includes a clinical example and
35 describes the reporting standards for publishing survival analyses in the *European Journal of Cardio-Thoracic Surgery* (EJCTS) and *Interactive Cardiovascular and Thoracic Surgery* (ICVTS).

Important aspects

Censoring. Censoring means that the time to the event of interest cannot be determined. This
40 commonly occurs when a subject drops out of the study or does not experience the event before the study follow-up period is completed [7, 8]. The aforementioned examples are known as *right-censoring*. Censoring can also be classified as a *left- and interval-censoring*. For example, if the event of interest was myocardial infarction (MI) after CABG, and the subject did not have a MI at 1-year follow-up, but did have a silent MI (pathological Q-waves on the ECG) at 2-year follow-up without
45 the exact date of the event, then the event time is interval censored, as we do not exactly know when the MI occurred, except sometime between the two clinic visits.

Defining the event of interest. It is important to clearly describe the event of interest. The conventional endpoint in survival analysis is death. Unlike in situations where logistic regression

might be used, this is not death within a fixed interval; rather, the outcome is the *time to the event*.

50 However, survival analysis can also be applied to other events; for example, major adverse cardiac and cerebrovascular events (MACCE), which is a combined endpoint of all-cause death, stroke, myocardial infarction (MI) and repeat revascularisation. This does not change anything in the analysis or graphical representation, but outcomes are then described as disease-free survival or cumulative event rate. However, when analysing non-terminal events, researchers should be aware of the concept
55 of competing risks [9].

Follow-up period. It is essential to use a clinically appropriate follow-up period when analysing survival data of two different treatment-groups. For instance, a comparison of PCI versus CABG will be biased towards PCI if the follow-up period is too short to capture all relevant events of both treatment strategies.

60 Follow-up data is difficult to collect: once the patient is discharged from the hospital, it creates additional work to track their status. Consequently, it is essential to describe the method of collecting follow-up data [10]. It is important to recognise that in models such as the Cox regression model, the number of events is of most importance rather than the number of subjects; hence an appropriate follow-up time must be considered. Moreover, the follow-up time distribution should be appropriately
65 summarised, including the **median** follow-up time (with clarification of which method was used [11]), key dates (e.g. of final follow-up), the number of subjects lost to follow-up, and the number of events observed [8, 12]. **When plotting Kaplan-Meier survival estimates (see below), another useful summary is the number of patients still at risk at a set of discrete time points (Figure 1B).**

70 **Visualising survival data**

Life-table. A life-table is the most basic form of survival analysis and is instrumental in explaining how censoring is handled [13]. To construct a life-table, the follow-up time is split into discrete time periods. For each time period, the numbers at risk are calculated by accounting for the subjects who died and were censored. These calculations result in an estimated survival probability for
75 each time period [14].

Kaplan-Meier estimator. The standard graphical plot of the cumulative survival probability during a defined follow-up period is the Kaplan-Meier plot, which is based on the non-parametric product-limit estimator [1]. The product-limit estimator can effectively be considered as the generalisation of the life table approach to each unique continuous event time. Steps in the Kaplan-Meier plot occur at each unique event time. All the necessary and essential information has to be clearly displayed to the reader. Example Kaplan-Meier plots that either do or do not achieve acceptable reporting standards for the EJCTS and ICVTS are presented in Figure 1 [12]. Kaplan-Meier plots may also be presented as downward sloping (i.e. cumulative survival; Figure 1) or upward sloping (i.e. cumulative event rate; Figure 2). The latter has been advocated by experts in biostatistics [15]. Moreover, always be aware of a potential break in the vertical axis. In addition, to avoid misinterpretation of the graphs, authors might consider halting the Kaplan-Meier estimates when only a small proportion (e.g. 10%) of patients are at risk [15]. This is because the eye is drawn to the right hand-side of the Kaplan-Meier plot, which is where the uncertainty is greatest due to only a small remaining number of patients at risk. However, this restriction is only for the graph; all data and events should be retained for analysis.

Comparing survival data

The log-rank test and the Cox proportional hazards model are the two most commonly used tests for comparing differences in survival data between study groups. It is important to note that the factor on which comparison is to be made is recorded at baseline, i.e. before the time origin. It is not possible to 'reach' forward in time. A classic example involves heart transplantation. In the Stanford Heart Transplant Study, patients who were eligible for transplantation were followed until death or censorship [16]. The objective of the study was to assess whether patients who received a transplant live longer than those not receiving a transplant. Comparing the survival distributions between those who received a transplant at some point and those who did not is, however, not appropriate. Patients who received a transplant must have lived until the point of transplantation, and therefore must have contributed to survival time in the non-transplant group [17]. In short, transplantation status is not

known at the time of eligibility; hence, no conditioning on this variable should be made in advance. In such situations, one might consider i) including a time-dependent covariate [18]; ii) using a landmark analysis [19]; or iii) not including measurements which occurred after the time origin. However, the choice of approach would depend on the application.

Log-rank test. The log-rank test is a standard nonparametric method for unadjusted univariable comparison of survival data between two (or more) study groups [2]. The log-rank test assumes that censoring is unrelated to prognosis and compares the complete survival distributions. The null hypothesis is that there are no differences in survival distributions between study groups. When the resulting P -value from the log-rank test is small, the null hypothesis should be rejected.

Notwithstanding the ubiquity of the log-rank test, several variations have been proposed based on the inclusion of weights. These include the Wilcoxon, Tarone-Ware, Peto, and Fleming-Harrington tests [17].

The Cox proportional hazards model. The Cox proportional hazards model is widely used for the adjusted comparison of survival data [3]. Like other types of regression models, adjustment can be univariable or multivariable. In the latter case, it is therefore possible to assess the independent association of a specific variable of interest (e.g. treatment arm) on survival whilst adjusting for other confounders. The term *hazard* refers to the risk of an event occurring per unit of time, conditional on the subject having survived up to that time. The model is formulated in terms of the hazard function, $h_i(t)$, as follows

$$\log h_i(t) = \log h_0(t) + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi},$$

where $h_0(t)$ is the baseline hazard (a function of time), which by virtue of the partial likelihood estimation method developed by Cox, we can leave unspecified, hence the name **semiparametric model**. The measured baseline values for subject i are $x_{1i}, x_{2i}, \dots, x_{pi}$ and associated with coefficients $\beta_1, \beta_2, \dots, \beta_p$, respectively. The coefficients are estimated, but usually reported as exponentiated values, i.e. $\exp(\beta_j)$, called hazard ratios (HR), which like other regression models, are reported together with confidence intervals and P -values, allowing for inference. If a HR is significantly >1 ,

then we can conclude that an increase in that risk factor **corresponds** to an increase in the event hazard,
130 which in turn decreases the length of survival.

Assumptions in the Cox proportional hazards model. In the Cox proportional hazards (PH) model, the HR is assumed to be constant over time [7, 20]. Methods for verifying this important assumption include i) smoothed scaled Schoenfeld residual plots, ii) complementary Kaplan-Meier plots, iii) the Grambsch-Therneau test based on the Schoenfeld residuals, or iv) including time-
135 dependent coefficients into the regression model [21]. **In the case of crossing survival estimates**, which is not uncommon **in comparisons between interventional and surgical procedures**, the analysis should be modified to compensate for **non-PH**. If the variable that violates the proportionality assumption is categorical and not of inferential interest, then the simplest approach is to use it to stratify the baseline hazard. Alternative **options include: dividing** the follow-up time into intervals for which the hazard
140 ratios are known to be proportional [21]; **fitting** an extended Cox model [21, 22]; **fitting a** parametric survival model **that does not rely on the PH assumption; or interpret the HR as an average HR. More recently, it has been proposed to compute the restricted mean survival time instead [23].**

Comparison with the general population. Occasionally, an interest exists to compare the survival distribution of a study cohort to that of a general population. The standardized mortality ratio
145 describes the number of observed deaths (O) in relation to the number of expected deaths (E), retrieved from actuarial tables. To allow an appropriate interpretation of this O/E-ratio, it needs to be matched with respect to age, sex, ethnicity and socio-economic aspects. An O/E-ratio of > 1 indicates the observed mortality is higher than the expected mortality. Testing for significant differences in survival between the study and general population **is challenging** [24], but it **can be** performed by a
150 one-sided log-rank test [25].

Example

To illustrate the concepts and methods described in this article, survival data from the “Synergy between Percutaneous Coronary Intervention with Taxus and Cardiac Surgery” (SYNTAX) trial are
155 presented. In this trial, patients with multi-vessel coronary artery disease were randomized to PCI with

paclitaxel-eluting stents or CABG [26]. The five-year outcomes in subgroups with ($n = 452$) or without ($n = 1348$) diabetes were compared for MACCE. Five-year clinical follow-up was performed by a clinic visit or telephone call in 86.5% of CABG patients and 94.5% of PCI patients. The follow-up was complete (clinical follow-up or death) in 88.0% (CABG) and 95.2% (PCI).

160 Event rates were estimated with Kaplan-Meier **estimators (Figure 2)**. The survival at 5-years in the diabetic PCI arm was 88.0% and 89.1% for the diabetic CABG arm (log-rank **test** $P=0.48$). An unadjusted HR of 1.81 (CI 95% [1.31 to 2.48], $P<0.001$) suggests that a diabetic patient undergoing PCI has a 1.81 times significantly increased hazard of developing MACCE compared to a diabetic patient undergoing CABG.

165

Survival analysis software

Most of the described survival analyses can be executed using straightforward packages in software programs, including R (R Development Core Team, Vienna, Austria) using the ‘survival’ package; SPSS (IBM Corp., Armonk, New York, USA) using the ‘Survival’ tab from the ‘Analyse’ menu; and
170 Stata (StataCorp., College Station, Texas, USA), e.g. using the ‘st’ and ‘stcox’ commands. For each subject, the data required to implement software analyses are i) a continuous time observation which measures the time until an event or when a subject drops out of the study (censoring) and ii) a dichotomous observation describing whether the subject experienced the event or not (usually coded 1 and 0, respectively). Additional columns for baseline data (e.g. treatment groups, gender, age, etc.) can
175 also be included for log-rank tests and Cox proportional hazards regression. As for regression **modelling in general**, development must consider which **(and how many)** covariates to include, the functional form **and** interaction terms.

Conclusion

180 Survival analysis concentrates on the time until an event of interest. Specific methods are used to analyse and compare survival data, including Kaplan-Meier method, the log-rank-test and the Cox proportional hazards model. When fitting a Cox regression model, it is crucial to evaluate whether the

proportional hazards assumption is met. Here, only the fundamentals routinely used by cardiothoracic surgeons in their daily research are described, and this paper serves as a guide for publishing their analysis in EJCTS and ICVTS. Yet of course, advanced survival analysis methods are available. These include alternative regression models to the Cox model, inclusion of time-varying covariates, competing risks data, recurrent events data, frailty models (i.e. the inclusion of random effects), and the joint modelling of survival and longitudinal data [9, 20, 27-29].

190 **Funding statement**

This work was not supported by any external funding.

FIGURE LEGENDS

Figure 1. Two examples of Kaplan-Meier estimates in patients with advanced lung cancer.

195 Panel A: a figure that lacks vital information or does not adhere to the EJCTS and ICVTS reporting
guidelines. Panel B: a publication-ready figure. The **plot** in Panel A is not suitable for publishing in
EJCTS or ICVTS; for example, it i) lacks a legend clarifying which group is represented by each line,
ii) has no units on the horizontal and vertical axes, iii) uses time-intervals of 200 days that are difficult
to interpret, iv) does not display a table giving the number of patients at risk at different time points,
200 and v) does not indicate the censoring times. In contrast, the **plot** in Panel B demonstrates an
appropriate way of reporting. **In Panel B, dotted lines denote 95% confidence bands, and tick marks
denote censoring times. Reproduced with permission from Hickey *et al.* [12] and Oxford University
Press.**

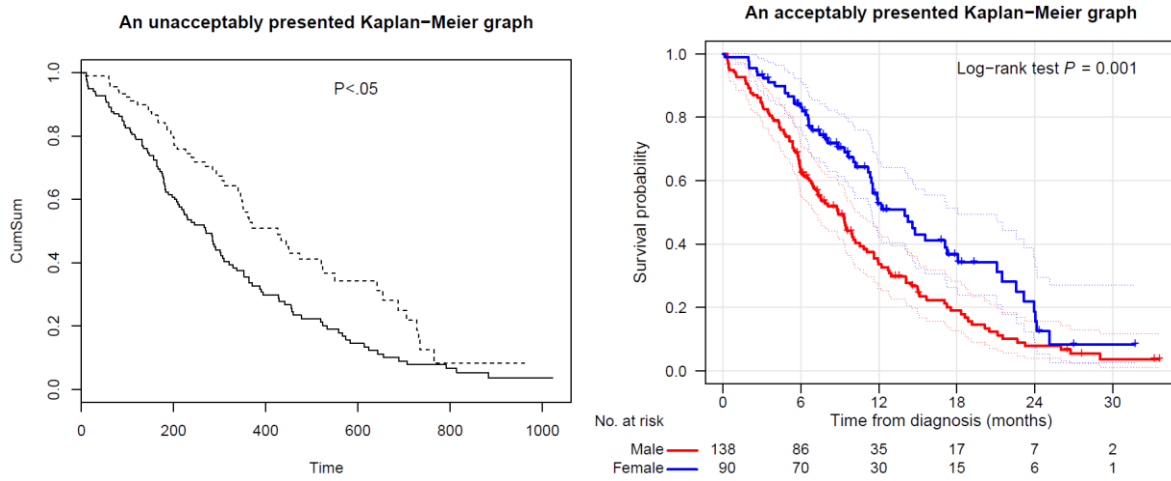
205 **Figure 2. Five-year outcomes of PCI versus CABG in diabetic patients in the SYNTAX trial.**

Kaplan–Meier estimates of major adverse cardiac or cerebrovascular events (MACCE) in diabetic
patients undergoing PCI versus CABG. Coronary artery bypass grafting (*dotted line*); PCI:
percutaneous coronary intervention (*solid line*). **Reproduced with permission** from Kappetein A.P. *et*
al. [26] **and** Oxford University Press.

210

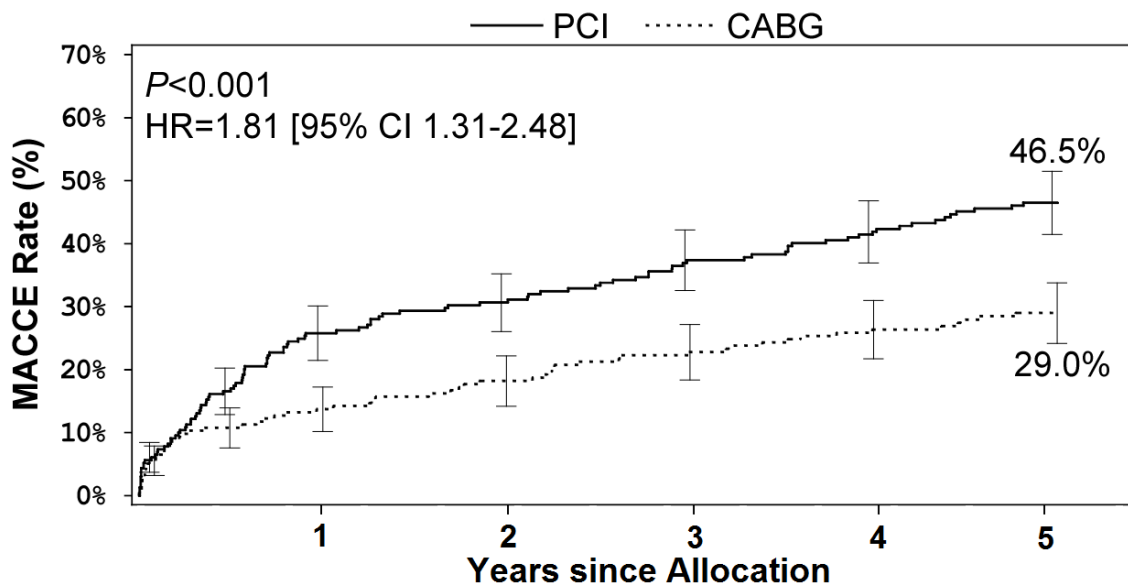
FIGURES

Figure 1.



215

Figure 2.



Numbers at risk:

| | t=0 | 1 | 2 | 3 | 4 | 5 |
|------|-----|-----|-----|-----|-----|-----|
| PCI | 231 | 211 | 176 | 169 | 156 | 140 |
| CABG | 221 | 193 | 178 | 175 | 163 | 151 |

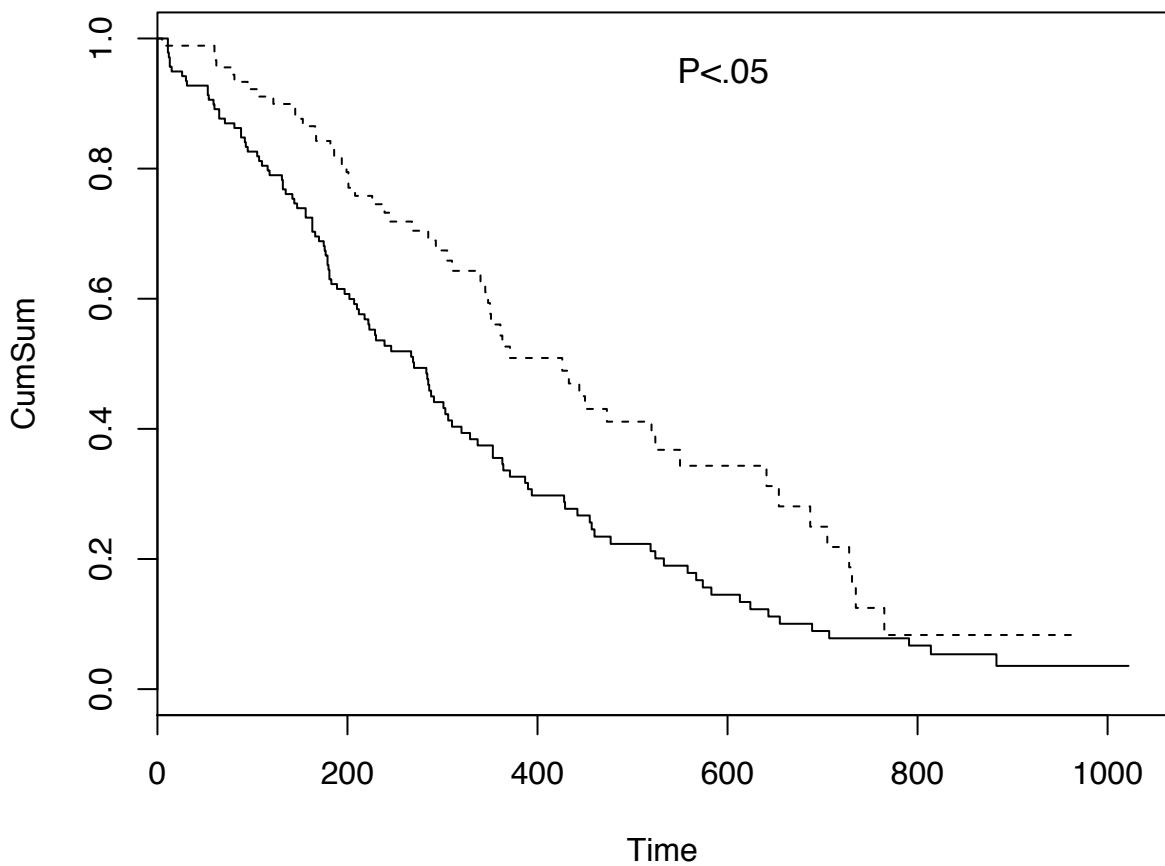
REFERENCES

- 220 [1] Kaplan E.L. MP. *Nonparametric Estimation from Incomplete Observations*. Journal of the American Statistical Association 1958;**53**:457-81.
- [2] Bland JM, Altman DG. *The logrank test*. BMJ 2004;**328**:1073.
- [3] Cox DR. *A note on the graphical analysis of survival data*. Biometrika 1979;**66**:188-90.
- 225 [4] Van Noorden R, Maher B, Nuzzo R. *The top 100 papers*. Nature 2014;**514**:550-3.
- [5] Kaplan E.L. MP. *Nonparametric estimation from incomplete observations*. J Am Stat Assoc 1958;**53**:457 - 81.
- [6] Cox DR. *Regression Models and life-tables*. Journal of the Royal Statistical Society 1972;**34**:187 - 220.
- 230 [7] Ahmed FE, Vos PW, Holbert D. *Modeling survival in colon cancer: a methodological review*. Mol Cancer 2007;**6**:15.
- [8] Clark TG, Bradburn MJ, Love SB, Altman DG. *Survival analysis part I: basic concepts and first analyses*. Br J Cancer 2003;**89**:232-8.
- [9] Grunkemeier GL, Jin R, Eijkemans MJ, Takkenberg JJ. *Actual and actuarial probabilities of competing risks: apples and lemons*. Ann Thorac Surg 2007;**83**:1586-92.
- 235 [10] Akins CW, Miller DC, Turina MI, Kouchoukos NT, Blackstone EH, Grunkemeier GL *et al*. *Guidelines for reporting mortality and morbidity after cardiac valve interventions*. J Thorac Cardiovasc Surg 2008;**135**:732-8.
- [11] Schemper M, Smith TL. *A note on quantifying follow-up in studies of failure time*. Control Clin Trials 1996;**17**:343-6.
- 240 [12] Hickey GL, Dunning J, Seifert B, Sodeck G, Carr MJ, Burger HU *et al*. *Statistical and data reporting guidelines for the European Journal of Cardio-Thoracic Surgery and the Interactive CardioVascular and Thoracic Surgery*. Eur J Cardiothorac Surg 2015;**48**:180-93.
- [13] Cutler SJ, Ederer F. *Maximum utilization of the life table method in analyzing survival*. J Chronic Dis 1958;**8**:699-712.
- 245 [14] Berkson J, Gage RP. *Calculation of survival rates for cancer*. Proc Staff Meet Mayo Clin 1950;**25**:270-86.
- [15] Pocock SJ, Clayton TC, Altman DG. *Survival plots of time-to-event outcomes in clinical trials: good practice and pitfalls*. Lancet 2002;**359**:1686-9.
- 250 [16] Crowley J, Hu M. *Covariance Analysis of Heart Transplant Survival Data*. Journal of the American Statistical Association 1977;**72**:27-36.
- [17] Kleinbaum DG, Klein M. *Survival Analysis*. 3rd edn New York: Springer, 2012.
- [18] Oakes D, Peterson DR. *Survival methods: additional topics*. Circulation 2008;**117**:2949-55.
- 255 [19] Dafni U. *Landmark analysis at the 25-year landmark point*. Circ Cardiovasc Qual Outcomes 2011;**4**:363-71.
- [20] Bradburn MJ, Clark TG, Love SB, Altman DG. *Survival analysis part II: multivariate data analysis--an introduction to concepts and methods*. Br J Cancer 2003;**89**:431-6.
- [21] Bradburn MJ, Clark TG, Love SB, Altman DG. *Survival analysis Part III: multivariate data analysis -- choosing a model and assessing its adequacy and fit*. Br J Cancer 260 2003;**89**:605-11.
- [22] Tableman M KJ. *Survival analysis using S*. Chapman and Hall/CRC, 2003.
- [23] Royston P, Parmar MK. *Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome*. BMC Med Res Methodol 2013;**13**:152.
- 265

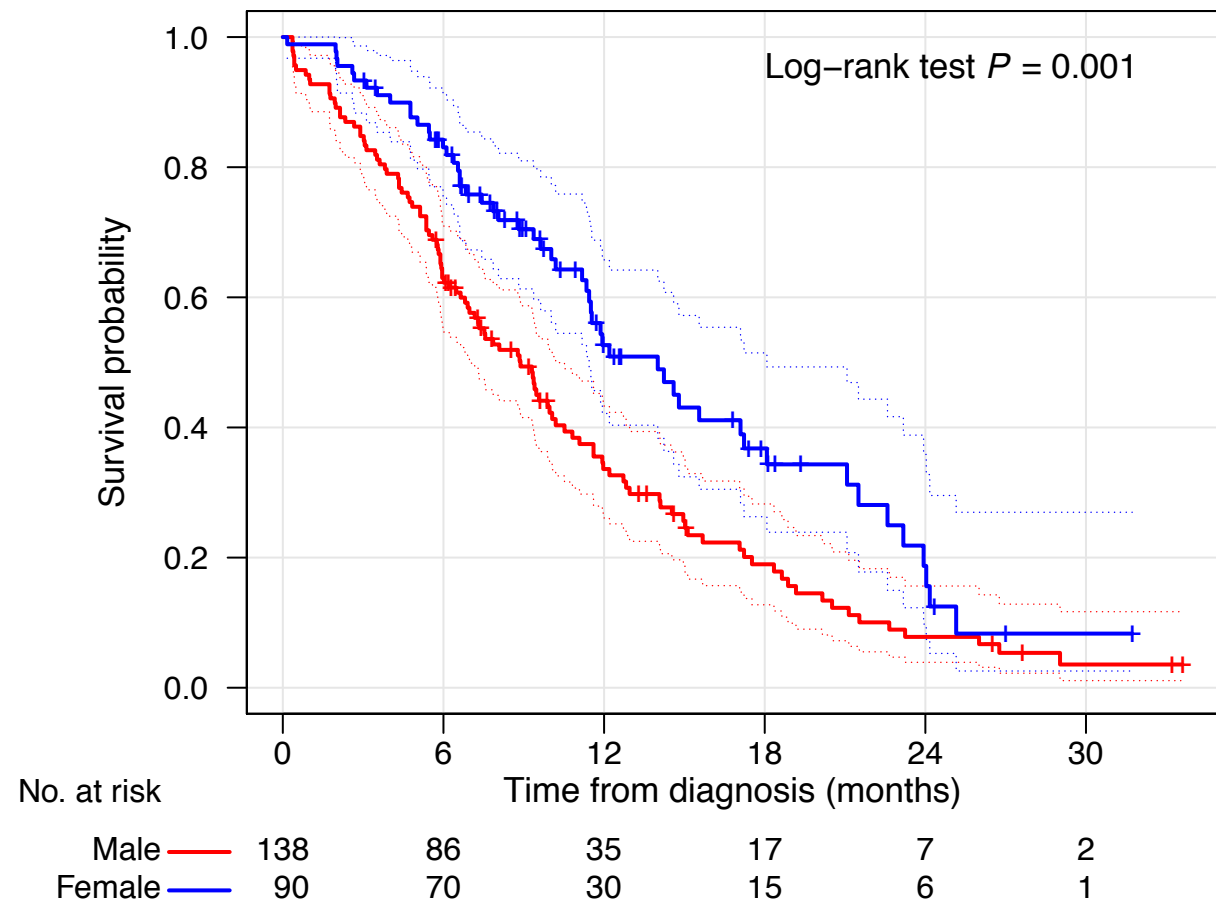
- [24] Osnabrugge RL, Head SJ, Kappetein AP. *Methodologic issues regarding background mortality in observational studies*. J Thorac Cardiovasc Surg 2011;**142**:1289-90; author reply 90.
- 270 [25] Finkelstein DM, Muzikansky A, Schoenfeld DA. *Comparing Survival of a Sample to That of a Standard Population*. JNCI: Journal of the National Cancer Institute 2003;**95**:1434-39.
- 275 [26] Kappetein AP, Head SJ, Morice MC, Banning AP, Serruys PW, Mohr FW *et al*. *Treatment of complex coronary artery disease in patients with diabetes: 5-year results comparing outcomes of bypass surgery and percutaneous coronary intervention in the SYNTAX trial*. Eur J Cardiothorac Surg 2013;**43**:1006-13.
- [27] Amorim LD, Cai J. *Modelling recurrent events: a tutorial for analysis in epidemiology*. Int J Epidemiol 2015;**44**:324-33.
- [28] Clark TG, Bradburn MJ, Love SB, Altman DG. *Survival analysis part IV: further concepts and methods in survival analysis*. Br J Cancer 2003;**89**:781-6.
- 280 [29] Andrinopoulou ER, Rizopoulos D, Jin R, Bogers AJ, Lesaffre E, Takkenberg JJ. *An introduction to mixed models and joint modeling: analysis of valve function over time*. Ann Thorac Surg 2012;**93**:1765-72.

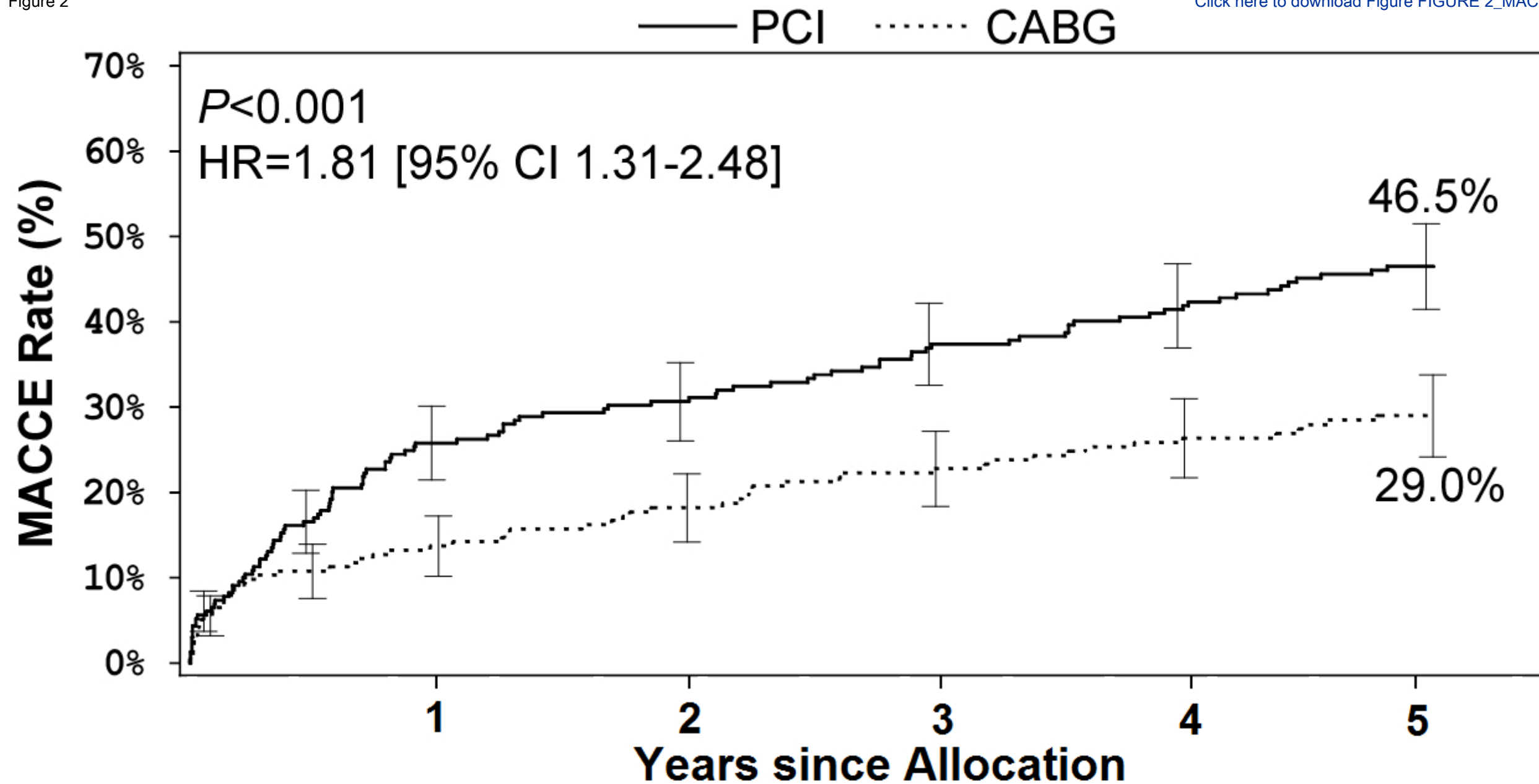
Figure 1

An unacceptably presented Kaplan–Meier graph



An acceptably presented Kaplan–Meier graph





Numbers at risk:

| | t=0 | | 1 | 2 | 3 | 4 | 5 |
|------|-----|-----|-----|-----|-----|-----|-----|
| PCI | 231 | 211 | 176 | 169 | 156 | 140 | 129 |
| CABG | 221 | 193 | 178 | 175 | 163 | 151 | 143 |