BMJ

BMJ 2013;347:f7118 doi: 10.1136/bmj.f7118 (Published 29 November 2013)

ENDGAMES

STATISTICAL QUESTION

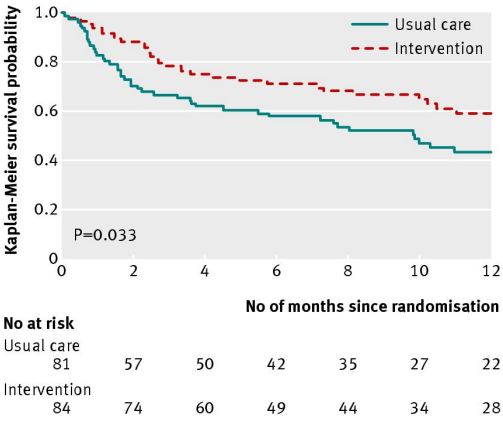
Kaplan-Meier survival curves: interpretation and communication of risk

Philip Sedgwick *reader in medical statistics and medical education*, Katherine Joekes *senior lecturer in clinical communication*

Centre for Medical and Healthcare Education, St George's, University of London, London, UK

Researchers investigated whether specialist nurse intervention reduced morbidity and mortality in patients with chronic heart failure. A randomised controlled trial study design was used. The intervention consisted of specialist nurse home visits in addition to routine care. The aim of the intervention was to educate patients about heart failure and its treatment. The control treatment consisted of routine care alone, with patients managed as usual by the admitting physician and, subsequently, general practitioner. Participants were 165 patients admitted to an acute medical admissions unit with heart failure as a result of left ventricular systolic dysfunction. The intervention started before discharge and continued for up to one year.[1]

The primary endpoint was a composite of death from all causes or first readmission to hospital with worsening heart failure. The Kaplan-Meier survival curves for the length of time after randomisation until occurrence of the primary endpoint were presented for the intervention and control treatment groups (figure). There was a significant difference in survival times between the treatment groups (log rank test P=0.033). The Kaplan-Meier survival probability estimates at 12 months were about 0.59 for intervention and 0.43 for control.



Kaplan-Meier survival curves for length of time after randomisation until occurrence of the primary endpoint (death from any cause or hospital readmission for heart failure) for the intervention and control treatment groups. The numbers of patients still at risk of experiencing the primary endpoint are indicated

The researchers concluded that specially trained nurses can improve the morbidity and mortality of patients admitted to hospital with chronic heart failure.

Which of the following statements, if any, can be inferred?

a) About 43% of the control group had not experienced the primary endpoint by the end of the 12 month follow-up

b) For the intervention group, the probability of experiencing the primary endpoint some time after 12 months was about 0.59

c) For any patient, the time taken to experience the primary endpoint after starting treatment would have been longer if he or she had received the intervention rather than the control treatment

p.sedgwick@sgul.ac.uk

Answers

Statement *b* can be inferred, whereas *a* and *c* cannot.

The trial investigated whether specialist nurse intervention, when compared with routine care, reduced morbidity and mortality in patients with chronic heart failure admitted to an acute medical admissions unit. For each patient the length of time after starting treatment until the occurrence of the primary endpoint (death from all causes or first readmission to hospital with worsening heart failure) was recorded. These times are referred to as "time to event" or "survival" data. Survival data have been described in a previous question.[2] The term "survival data" is perhaps misleading because the endpoint does not have to be death or an adverse event. The endpoint might be positive—such as recovery from an operation.

The Kaplan-Meier survival curves for the length of time after randomisation until occurrence of the primary endpoint for the intervention and control treatment groups are shown (figure). The numbers of patients still at risk of experiencing the primary endpoint are indicated—that is, the numbers of patients who have not died from any causes or not been readmitted to hospital with worsening heart failure after randomisation.

The distinguishing feature of survival data is that usually some participants will not have experienced the primary endpoint before the end of the study. In the trial above, some participants in each group had not experienced the primary endpoint (death or readmission to hospital) during the 12 months of follow-up. For these patients, survival times were recorded as the total length of follow-up—that is, 12 months. Exactly when they experienced the primary endpoint is not known. All that can be inferred is that they experienced the primary outcome some time after 12 months. For that reason, their survival times are called "right censored" or simply "censored." Some participants may also have had censored observations because, for example, they withdrew from the study or were lost to follow-up before experiencing the primary endpoint. It can be inferred that there were censored observations in the trial above because the Kaplan-Meier survival curves for both treatment groups did not reach zero probability by the end of the 12 months of follow-up.

The Kaplan-Meier survival curves displayed do not indicate straightforward absolute probabilities—that is, they do not indicate the proportion of patients in each treatment group that had not experienced the primary endpoint during follow-up. This is because each treatment group had "censored" observations. The probabilities shown are called Kaplan-Meier survival probabilities and have a unique interpretation. The survival probabilities are conditional ones and indicate the probability of experiencing the primary endpoint beyond a certain length of follow-up.

The Kaplan-Meier survival probability estimates at 12 months were about 0.59 for the intervention group and 0.43 for the control group. As described above, the Kaplan-Meier survival curves displayed do not indicate straightforward absolute probabilities. Therefore, statement *a* cannot be inferred—about 43% of the control group had not experienced the primary endpoint by the end of the 12 months of follow-up. However, statement *b* can be inferred—the probability of reaching the primary endpoint some time after 12 months was about 0.59 for the intervention group.

Traditional hypothesis testing with a two sided alternative hypothesis, described in a previous endgame,[3] was used to compare treatment groups in survival times. The test was significant (P=0.033) and the null hypothesis of no difference in survival times in the population between the intervention and control treatments was rejected in favour of the alternative. Because the survival curve for the intervention group was above that for the control group (figure), it can be concluded that the intervention group had significantly longer survival times as a whole. Therefore, mortality and morbidity were lower in the intervention group than in the control group. Although it can be concluded that the survival times for the intervention group as a whole were significantly longer than those of the control group, statement *c* cannot be inferred. That is, it cannot be inferred that all patients would have taken longer to experience the primary endpoint if they had received the intervention rather than the control treatment. The morbidity and mortality for some patients may have been enhanced if they had received the control treatment rather than the intervention. Furthermore, it was not possible to predict how long a patient would survive and not be readmitted to hospital if he or she had received one or other of the treatments.

It can be difficult to communicate the efficacy of a new intervention to patients, policy makers, or healthcare professionals on the basis of Kaplan-Meier survival data. Kaplan-Meier survival times can be hard to understand, and this could present further uncertainty. For example, it might not be useful to tell people that the probability of experiencing the primary endpoint (death or readmission to hospital) some time after 12 months is about 0.59 for the intervention and 0.43 for standard care without any indication of the prognosis beyond that time. Furthermore, percentages can be challenging to patients, policy makers, and healthcare professionals. It may be helpful to use natural frequencies and round numbers and to personalise the message. For example, when comparing treatment options a clinician could explain to a patient, "If there are 100 people in your situation, 60 will survive and not have been readmitted for exacerbating heart failure within 12 months if they have specialist nurse intervention compared with 45 with standard care." The use of such natural frequencies makes the data more accessible. The composite primary outcome (death and hospital readmission) is complex and may lead to further confusion, and a more detailed exploration of the study results would be needed to understand the impact of the intervention fully.

The probability estimates of the primary endpoint (death or hospital readmission) are based on groups of patients and are difficult to apply to a single patient. When offering the intervention to a patient on the basis of such results, the uncertainty inherent in the situation and the fact that the new intervention cannot provide any guarantees must be explained. Although it has been shown that the intervention is significantly superior to the control, this may not be the case for all patients. Furthermore, it is not possible to predict how long patients would survive or not be readmitted to hospital if they received either the intervention or control treatment.

Competing interests: None declared.

1 Blue L, Lang E, McMurray JJV, Davie AP, McDonagh TA, Murdoch DR, et al. Randomised controlled trial of specialist nurse intervention in heart failure. *BMJ* 2001;323:715-8.
2 Sedgwick P. Survival (time to event) data II. *BMJ* 2010;341:c3665.
3 Sedgwick P. Statistical hypothesis testing. *BMJ* 2010;340:c2059.

Cite this as: *BMJ* 2013;347:f7118