Biometrika (2016), **103**, 3, *pp.* 742–745
Printed in Great Britain

doi: 10.1093/biomet/asw026
Advance Access publication 25 July 2016

On the win-ratio statistic in clinical trials with multiple types of event

By D. OAKES

Department of Biostatistics and Computational Biology, University of Rochester, Rochester, New York 14642, U.S.A.

david_oakes@urmc.rochester.edu

Summary

Pocock et al. (2012), following Finkelstein & Schoenfeld (1999), has popularized the win ratio for analysis of controlled clinical trials with multiple types of outcome event. The approach uses pairwise comparisons between patients in the treatment and control groups using a primary outcome, say the time to death, with ties broken using a secondary outcome, say the time to hospitalization. In general the observed pairwise preferences and the weight they attach to the component rankings will depend on the distribution of potential follow-up time. We present expressions for the win and loss probabilities for general bivariate survival models when follow-up of all patients is limited to a specified time horizon. In the special case of a bivariate Lehmann model we show that the win ratio does not depend on this horizon. We show how the win ratio may be estimated nonparametrically or from a parametric model. Extensions to events of three or more types are described. Application of the method of marginal estimation due to Wei et al. (1989) to this problem is described.

Some key words: Composite endpoint; Hougaard model; Semicompeting risks; Survival analysis.

1. Introduction

Outcome measures in controlled cardiovascular clinical trials often include the time from randomization to death or to the first occurrence of a nonfatal event such as hospitalization. How best to combine these two measures into a single primary outcome is the source of some controversy. The composite outcome, consisting of the time to the first event of either category, allows the comparison of treated and control groups to be made with standard univariate survival methods. However, this approach ignores deaths among patients who have experienced a nonfatal event. The win ratio, introduced by (Pocock et al., 2012) following an approach suggested by Finkelstein & Schoenfeld (1999) for the joint analysis of survival and longitudinal outcome data, is an appealing alternative method, prioritizing mortality. The method uses pairwise comparisons between each patient in the treated group and each patient in the control group. The two patients are ranked first on time to death, if it occurs, and if this gives no clear winner, on the time to nonfatal event. Pocock defined the win ratio as the ratio of the total number of preferences thus established for the treated group to the number for the control group. We use its reciprocal, the loss ratio, termed here $\hat{\theta}$, because this corresponds to the hazard ratio, with values exceeding unity indicating excess risk in the treated group. Two-sample U-statistic theory provides expressions for the asymptotic variance of the win-ratio statistic that account for the dependence of comparisons involving the same individual. Luo et al. (2015) derive an alternative standard error estimate using counting process methods.

As indicated in Luo et al. (2015) the parameter, θ say, estimated by $\hat{\theta}$, will depend on the distribution of potential follow-up times in the trial, and so may not apply to other trials or populations. In the extreme case, if all patients are followed until death, nonfatal events will be irrelevant. To clarify this it is helpful to define a horizon c of interest, and to consider the loss ratio θ_c that would be obtained if a population of patients were all followed up to time c. This parameter can be estimated from data subject to progressive

censorship up to time c, rather as, for example, a five-year survival rate can be estimated by life-table methods from data including follow-up times censored prior to five years.

2. Formulation for a single outcome

We first review methods for a single outcome, say mortality, with times from randomization to death denoted by T_j with $j = 0, 1$ for the control and treated groups respectively, and corresponding distribution and survivor functions $F_j(\cdot)$ and $\bar{F}_j(\cdot)$, assumed continuous. With complete observations, the loss ratio is

$$\theta = \frac{\mathrm{pr}(T_0 > T_1)}{\mathrm{pr}(T_1 > T_0)} = \frac{\int \bar{F}_0(t)\,\mathrm{d}F_1(t)}{\int \bar{F}_1(t)\,\mathrm{d}F_0(t)}.$$

With potential censoring of both T_j at time c, it is

$$\theta_c = \frac{\int_0^c \bar{F}_0(t)\,\mathrm{d}F_1(t)}{\int_0^c \bar{F}_1(t)\,\mathrm{d}F_0(t)}. \tag{1}$$

Even if the potential censoring times, c_0 and c_1 say, differ, (1) still holds, with $c = \min(c_0, c_1)$, since T_0 is known to be less than T_1 based on the observed data if and only if $T_0 < \min(c_0, c_1, T_1)$. Here and later, partial information, such as when $c_0 < T_0 < T_1 < c_1$, is ignored.

Rearranging (1) gives $\int_0^c \bar{F}_0(t)\,\mathrm{d}F_1(t) = \theta_c \int_0^c \bar{F}_1(t)\,\mathrm{d}F_0(t)$ and, easily, θ_c is c-free if and only if $\bar{F}_1(t) = \bar{F}_0(t)^\theta$, so that $\bar{F}_1$ and $\bar{F}_0$ satisfy the proportional hazards model with hazard ratio θ. Of course, if this model is assumed, then the preferred estimate of θ uses Cox's partial likelihood, whereas the win-ratio estimate is based on the Wilcoxon statistic, which is inefficient under the proportional hazards model. To estimate θ_c from progressively censored data without assuming proportional hazards, simply replace F_0 and F_1 by their Kaplan–Meier estimates $\hat{F}_0$ and $\hat{F}_1$, giving (Efron, 1967)

$$\hat{\theta}_c = \int_0^c \{1 - \hat{F}_0(t)\}\,\mathrm{d}\hat{F}_1(t) \Big/ \int_0^c \{1 - \hat{F}_1(t)\}\,\mathrm{d}\hat{F}_0(t).$$

3. Formulation of the model with two types of event

Suppose now that (T, X) denote times to death and first nonfatal event, respectively. We write (T, X) rather than (X, T) because T precedes X in priority, though not in time. By convention if either event does not occur the corresponding time is set to infinity. Consider a typical pair of individuals from the control and treated groups with data (T_0, X_0) and (T_1, X_1), respectively, and subject to censoring at times c_0 and c_1. This is a semicompeting risks problem in that T_j censors X_j. With $c = \min(c_0, c_1)$, the treated individual wins if $\{T_0 < \min(c, T_1)\} \cup [\{X_0 < \min(X_1, c)\} \cap \{T_1 \not< \min(c, T_0)\}]$ and loses if $\{T_1 < \min(c, T_0)\} \cup [\{X_1 < \min(X_0, c)\} \cap \{T_0 \not< \min(c, T_1)\}]$. The pair is indeterminate if $c < \min(T_0, T_1, X_0, X_1)$ and these three possibilities are almost surely mutually exclusive and exhaustive. To model the win and loss probabilities for comparing two individuals with potential censoring times c_0 and c_1 with $c = \min(c_0, c_1)$ we specify, for $j = 0, 1$, the marginal densities $f_j(t)$ of T_j, with survivor functions $\bar{F}_j(t)$, and the improper conditional densities $g_j(x \mid c)$ of X_j given $T_j > c$, with survivor functions $\bar{G}_j(x \mid c) = 1 - \int_0^x g_j(u|c)\,\mathrm{d}u$. The win and loss probabilites are respectively

$$\int_0^c f_0(t)\bar{F}_1(t)\,\mathrm{d}t + \bar{F}_1(c)\bar{F}_0(c) \int_0^c g_0(x \mid c)\bar{G}_1(x \mid c)\,\mathrm{d}x, \tag{2}$$

$$\int_0^c f_1(t)\bar{F}_0(t)\,\mathrm{d}t + \bar{F}_0(c)\bar{F}_1(c) \int_0^c g_1(x \mid c)\bar{G}_0(x \mid c)\,\mathrm{d}x, \tag{3}$$

and that of a tie is $\bar{F}_0(c)\bar{F}_1(c)\bar{G}_0(c \mid c)\bar{G}_1(c \mid c)$. Then θ_c is the ratio of (3) to (2). This answers the question: what is θ?

4. BIVARIATE LEHMANN MODELS AND c-FREE θ

For clear interpretability of the loss or win ratio we would like proportional hazards models to hold for $F_1(t)$, $F_0(t)$ and for the pair $G_1(x|c)$, $G_0(x|c)$ for each value of c, all with the same value of θ. There is a simple characterization of such models:

THEOREM 1. *The marginal survivor functions $\bar{F}_0(t)$, $\bar{F}_1(t)$ and the conditional survivor functions $\bar{G}_0(x \mid t)$, $\bar{G}_1(x \mid t)$ both satisfy proportional hazards models, in t, and in x for given t respectively, with the same θ if and only if the corresponding joint survivor functions $\bar{H}_j(t, x) = \mathrm{pr}(T_j > t, X_j > x)$ satisfy*

$$\bar{H}_1(t, x) = \bar{H}_0(t, x)^{\theta}. \tag{4}$$

Proof. The proof is immediate, since $\bar{H}_j(t, x) = \bar{F}_j(t)\bar{G}_j(x \mid t)$. □

Balkema & Resnick (1977) showed that the survivor function $\bar{H}(t, x)$ is such that $\bar{H}(t, x)^{\theta}$ is a survivor function for all $\theta > 0$ if and only if the continuation odds ratio

$$\phi(t, x) = \bar{H}(t, x)\frac{\partial^2 \bar{H}(t, x)}{\partial x \partial t} \Big/ \frac{\partial \bar{H}(t, x)}{\partial x}\frac{\partial \bar{H}(t, x)}{\partial t} \geqslant 1,$$

a type of positive dependence. See also Yang & Ying (2001). This condition is satisfied by all bivariate frailty models (Oakes, 1989), and the model takes a simple form if the frailty distribution is infinitely divisible. The model of Hougaard (1986) with

$$\bar{H}(t, x) = \exp(-[\{-\log \bar{F}(t)\}^{1/\alpha} + \{-\log \bar{G}(x)\}^{1/\alpha}]^{\alpha}) \quad (0 < \alpha < 1; 0 < x < t),$$

obtained from a positive stable frailty distribution, is particularly convenient, as it allows the strength of dependence, parameterized by α, to be specified independently of the proportionality factor θ.

5. SOME COMMENTS ON ESTIMATION

Under (4), valid estimates of θ may be obtained from the Cox's partial likelihood applied to (i) times to death, ignoring nonfatal events, and (ii) time to first nonfatal event, with deaths as censorings. To see this, note that the risk sets for first nonfatal events at time t include only subjects who have survived, event-free, to time t and that their hazard functions at t are those of the conditional survivor functions $\bar{G}_j(t \mid t)$, which also have proportionality factor θ. Each partial likelihood yields a valid estimate of standard error. The two estimates of θ are not asymptotically independent, but may be combined by the method of Wei et al. (1989). While better estimates may exist, efficiency gains over this method are likely to be small. The model may be extended to allow different values of θ for deaths and first nonfatal events.

Under (4) the usual composite endpoint, the time to the earlier of death or nonfatal event, also follows a proportional hazards model. However the analysis of this endpoint will be less efficient than the combined analysis of Wei et al. (1989).

As for single outcome data, simple analogous estimates are available for θ_c in (2) and (3), without any assumptions on the form of the $H_j(t, x)$, and the data used to estimate θ_c may be subject to general progressive censorship, provided this permits consistent estimation of the relevant survivor functions over $(0, c)$. The analogous estimate of $\hat{G}_j(x \mid c)$ is constructed from all individuals whose time to death is known to exceed c. In a fully nonparametric approach any follow-up beyond time c is irrelevant to the estimation of θ_c.

6. EXTENSIONS TO THREE OR MORE TYPES OF EVENT

The win-ratio method extends easily to the comparison of pairs of individuals on the time to three or more prioritized outcomes, for example, severe, moderate or mild headache. Consider times to the first severe headache, breaking ties (i) using times to the first moderate headache, and (ii) using the time to the first mild headache. Headaches of severity equal to or lesser than one that has already occurred in the

same individual may be ignored as they do not affect any comparison. The calculation of the win and loss probabilities requires only:

(i) the marginal distribution of the time to the first severe headache;
(ii) the improper conditional distribution of the time to the first moderate headache given that no severe headache has occurred before time c, where $c = \min(c_0, c_1)$ as before;
(iii) the improper conditional distribution of the time to the first mild headache, given that no severe or moderate headache has occurred before time c.

Writing Y, T and X for the times to severe, moderate and mild headaches ($Y > T > X$), we can decompose the joint survivor functions as a product of conditional survivor functions, $\bar{H}_j(y, t, x) = \bar{E}_j(y)\bar{F}_j(t \mid y)\bar{G}_j(x \mid t, y)$ However the win and loss probabilities involve only $\bar{E}_j(y)$, $\bar{F}_j(t \mid c)$ and $\bar{G}_j(x \mid c, c)$, so will not fully determine the $\bar{H}_j(y, t, x)$. So $\bar{H}_1(y, t, x) = \bar{H}_0(y, t, x)^\theta$ is sufficient, but not necessary, for the proportional hazards model to hold for $\bar{E}_j$, $\bar{F}_j$ and $\bar{G}_j$ The approach of Wei et al. (1989) still applies under the proportionality hypothesis, again with events of higher priority censoring follow-up for events of lower priority. The existence of multivariate Lehmann models relates to the question of min- or max-infinite divisibility and is more complicated in the general multivariate than in the bivariate case; see for example Resnick (1987) Chapter 5.

7. Discussion

Similar questions of definition and comparability arise in combining analysis of mortality with that of longitudinal outcome data, the problem discussed by Finkelstein & Schoenfeld (1999). The present approach could be extended to this problem by using the longitudinal process to define a hierarchy of intermediate events, for example the first time the process crosses certain specified levels.

References

Balkema, A. A. & Resnick, J. (1977). Max-infinite divisibility. *J. Appl. Prob.* **14**, 309–19.

Efron, B. (1967). The two sample problem with censored data In *Proc. 5th Berkeley Symp. Math. Statist. Prob.* **4**, L. LeCam and J. Neyman, eds. Berkeley: University of California Press, pp. 831–53.

Finkelstein, D. M. & Schoenfeld, D. A. (1999). Combining mortality and longitudinal measures in clinical trials. *Statist. Med.* **18**, 1341–54.

Hougaard, P. (1986). A class of multivariate survival distributions. *Biometrika* **73**, 671–8.

Luo, X. L., Tian, H., Mohanty, S. & Tsai, W. Y. (2015). An alternative approach to confidence interval estimation for the win ratio statistic. *Biometrics* **71**, 139–45.

Oakes, D. (1989). Bivariate survival models induced by frailties. *J. Am. Statist. Assoc.* **84**, 487–93.

Pocock, S. J., Ariti, C. A., Collier, T. J. & Wang, D. (2012). The win ratio: A new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. *Eur Heart J.* **13**, 176–82.

Resnick, S. I. (1987). *Extreme Values, Regular Variation and Point Processes*. New York: Springer.

Wei, L. J., Lin, D. Y. & Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J. Am. Statist. Assoc.* **84**, 1065–73.

Yang, Y. & Ying, Z. (2001). Marginal proportional hazards models for multiple event-time data. *Biometrika* **88**, 581–6.

[*Received March* 2016. *Revised May* 2016]