A Note on Quantifying Follow-up in Studies of Failure Time

Michael Schemper, PhD, and Terry L. Smith, MS
Section of Clinical Biometrics, Department of Medical Computer Sciences, Vienna University, Vienna, Austria, (M.S.), and Department of Biomathematics, The University of Texas MD Anderson Cancer Center, Houston, Texas (T.L.S.)

In a recent review of survival analyses published in cancer journals, Altman et al [1] found that about half of the papers did not include any summary of follow-up time. Only 31% of those papers that did report median follow-up specified the method used to compute it.

Independently of Altman et al [1], we have surveyed articles in three medical journals—Journal of Clinical Oncology, Annals of Internal Medicine, and New England Journal of Medicine—within a 6-month period (January–June 1994) and identified 70 that used survival analysis. Among these, 47 (67%) included a statement regarding duration of follow-up, but 24 did not specify the method used to quantify follow-up (usually median follow-up). Among the 23 other articles, methods employed were (1) follow-up based only on censored times (14 articles); (2) specification of a minimum follow-up time (5 articles); (3) times from entry to death or last contact (1 article); (4) times from entry to end-of-study date (1 article); (5) other methods (2 articles). Median sample size in the surveyed articles was 236 (range 30–8331), and the median proportion of censoring was 60% (range 3–96%). This note shows that values of median follow-up may differ substantially depending on the method used.

Results of survival analysis apply to the time frame in which most of the individuals were observed. In particular, standard analytical methods for survival data, such as the log-rank test [2], the generalized Wilcoxon test [3], or the proportional hazards model [4], estimate average effects [5] for the observed response times and test those effects for significance. Thus the current reporting of follow-up is unsatisfactory.

The following methods have been used or suggested. We assume a medical study with staggered entry of all individuals between times \( T_1 \) and \( T_2 \), and analysis of the available data at a final end-of-study time, \( T_3 \). For each individual \( i \) (1 ≤ \( i \) ≤ \( n \)), we observe the time of entry into the study, \( t_{i0} \), and the final recorded date, \( t_{i2} \). If \( t_{i2} \) is the date of death, the status indicator, \( s_i \), assumes a value of 1. For

Address reprint requests to: Michael Schemper, PhD, Section of Clinical Biometrics, Department of Medical Computer Sciences, University of Vienna, Spitalg. 23, A-1090 Vienna, Austria.
Received June 12, 1995.

Controlled Clinical Trials 17:343-346 (1996)
© Elsevier Science Inc. 1996
655 Avenue of the Americas, New York, NY 10010
0197-2456/96/$15.00
PII S0197-2456(96)00075-X
individuals censored at time \( t_i \), either by loss to follow-up or because they are alive at end of study, \( s_i = 0 \).

Though follow-up may be fully described by an empirical distribution, for practical purposes citing some characteristics of it, e.g., the median and other percentiles, will suffice. The following distributions provide quantification of follow-up:

1. Observation time (T-OBS): \( T_{OBS} = t_{\tilde{u}} - t_i \)
2. Censoring times (T-CENS): \( T_{CENS} = t_{\tilde{u}} - t_i \) (only available if \( s_i = 0 \))
3. Time to end-of-study (T-END): \( T_{END} = T_3 - t_i \)
4. Known function time (KFT): \( KFT_i = t_{\tilde{u}} - t_i \) for \( s_i = 0 \) and \( T_3 - t_i \) for \( s_i = 1 \)
5. Kaplan–Meier estimate of potential follow-up (KM-PF)

KM-PF, also termed "reverse Kaplan–Meier" [1], is calculated in the same way as the Kaplan–Meier [6] estimate of the survival function, but with the meaning of the status indicator reversed. Thus death (\( s_i = 1 \)) censors the true but unknown observation time of an individual, and censoring (\( s_i = 0 \)) is an endpoint. The unobservable follow-up time of a deceased patient is interpreted as the follow-up time that potentially would have been obtained had that person not died.

6. Korn’s potential follow-up (KORN)

Quantiles for the Korn [7] distribution are obtained from a function that estimates the probability \( P \) to be under follow-up at time \( t' \) \( (t' > 0) \)

\[
P(t') = P(L > t'|E > t')\)P(E > t')
\]

where \( P(E > t') \) is the proportion of subjects with \( T_{END} > t', L \) is time until lost to follow-up, and \( P(L > t'|E > t') \) is computed from a reverse Kaplan–Meier estimate as in method 5, for subjects with \( T_{END} \geq t' \). In the following paragraph, we discuss these measures.

With increasing death hazards, method 1 increasingly underestimates potential follow-up of patients. Thus for two studies with identical follow-up of patients but with the first study having much poorer survival, T-OBS will take on much lower values for the first study than the second. Method 2 also systematically underestimates follow-up by only taking into account censored times: Longer individual follow-up times have a higher likelihood of being unavailable because of intermittent deaths than short individual follow-ups. Method 3, does not penalize for loss to follow-up and therefore can substantially overestimate actual follow-up. The rationale for method 4 as opposed to 1 is that the survival status of a dead individual cannot change, which is equivalent to having followed such an individual at least to the end-of-study time, \( T_3 \). This method, a hybrid of 1 and 3 tends to overestimate potential follow-up of patients with increasing loss to follow-up, as will be shown.

While methods 1–4 are unsuitable for reasons noted, method 6 is not subject to these limitations. It accounts for different risks of loss to follow-up between early and later recruitment. This provision, which adds complexity in calculating the follow-up distribution, requires a specific computer program. This method was not cited in our survey; in the relevant years of the Science Citation Index [8] we found only five citations, suggesting that the Korn method is considered to require too much effort to obtain a description of follow-up. Furthermore, the Korn method (and also methods 3 and 4) requires specification of an end-of-study date \( T_3 \), which will often be subject to somewhat arbitrary choice when the status of each living
Table 1  Performance of Follow-up Measures with Simulated Survival Data

<table>
<thead>
<tr>
<th>Death Hazard Rate</th>
<th>Loss Hazard Rate</th>
<th>Resulting % Censored</th>
<th>Measures of Follow-up (Median)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>T-OBS</td>
</tr>
<tr>
<td>Increasing survival hazard</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.01</td>
<td>0</td>
<td>70.3</td>
<td>28.1</td>
</tr>
<tr>
<td>0.03</td>
<td>0</td>
<td>37.6</td>
<td>18.4</td>
</tr>
<tr>
<td>0.10</td>
<td>0</td>
<td>6.1</td>
<td>7.1</td>
</tr>
<tr>
<td>Increasing loss hazard</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.04</td>
<td>0.002</td>
<td>30.4</td>
<td>15.2</td>
</tr>
<tr>
<td>0.04</td>
<td>0.010</td>
<td>37.7</td>
<td>13.5</td>
</tr>
<tr>
<td>0.04</td>
<td>0.040</td>
<td>56</td>
<td>8.6</td>
</tr>
</tbody>
</table>

Method 5, KM-PF, which incorporates Korn's useful concept of potential follow-up, can be computed with readily available computer programs. Also, KM-PF accurately reflects changes in the quality of follow-up with increasing follow-up time of individuals; however, KM-PF requires the additional, possibly mild, assumption that the quality of follow-up does not depend on recruitment time.

To give an impression of how much the inappropriate methods 1-4 can give misleading results in practice, we provide results from a Monte Carlo study. For typical conditions likely to occur in practice, we sampled 5000 survival times from an exponential distribution with specified hazard rate. We then censored these survival times using the procedure by Gehan and Thomas [9] to model a clinical trial. For all conditions a fixed accrual period of 48 time units and constant accrual rate were assumed, plus an additional observation period of 12 time units. This amounts to T-END uniformly distributed in the interval [12,60] with a median of 36 time units. In order to investigate the effect of loss to follow-up, we assumed that time till loss to follow-up also followed an exponential distribution with specified hazard rate.

The results in Table 1 imply that T-CENS and in particular T-OBS substantially underestimate follow-up as compared with the more appropriate Korn method, whereas the reverse holds for T-END and KFT when there are losses to follow-up. Note that only KORN and KM-PF do not vary with increasing death hazard rate but only with increasing loss hazard rate.

Furthermore we compared KORN and KM-PF in a similar way under conditions of loss to follow-up depending on recruitment time. From Table 2 we conclude

Table 2  Comparison of the Korn and Kaplan-Meier Methods with Recruitment-Dependent Loss to Follow-up in Simulated Survival Data

<table>
<thead>
<tr>
<th>Death Hazard Rate</th>
<th>First Half of Recruitment Time</th>
<th>Second Half of Recruitment Time</th>
<th>Resulting % Censored</th>
<th>Measures of Follow-up (Median)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>KORN</td>
</tr>
<tr>
<td>0.04</td>
<td>0.02</td>
<td>0.02</td>
<td>45</td>
<td>22.4</td>
</tr>
<tr>
<td>0.04</td>
<td>0</td>
<td>0.04</td>
<td>38</td>
<td>33.2</td>
</tr>
<tr>
<td>0.04</td>
<td>0.04</td>
<td>0</td>
<td>45</td>
<td>22.5</td>
</tr>
</tbody>
</table>
that the more involved method by Korn and the much simpler KM-PF method hardly differ in their results even if the additional assumption of KM-PF is heavily violated.

For routine quantification of follow-up we therefore suggest use of the median and further quantiles of Kaplan–Meier estimated potential follow-up, which is easily obtained with standard software and can be meaningfully interpreted.

While follow-up of a clinical trial will usually be described by quantiles of a single overall distribution, occasionally the effect of covariates on follow-up might be of interest, such as when the comparability of the follow-up of treatment arms or of different contributing clinical centers is in doubt. In these or related situations, separate quantifications of follow-up may be illustrative and statistical tests [2,3] or even Cox’s model [4] could be applied in the usual way, but with the meaning of the status indicator reversed.

REFERENCES


