

How to Deal with Interval-Censored Data Practically while Assessing the Progression-Free Survival: A Step-by-Step Guide Using SAS and R Software

Audrey Emmanuelle Dugué¹, Marina Pulido², Sylvie Chabaud³, Lisa Belin⁴, and Jocelyn Gal⁵

Abstract

We describe how to estimate progression-free survival while dealing with interval-censored data in the setting of clinical trials in oncology. Three procedures with SAS and R statistical software are described: one allowing for a nonparametric maximum likelihood estimation of the survival curve using the EM-ICM (Expectation and Maximization-Iterative Convex Minorant) algorithm as described by Wellner and Zhan in 1997; a sensitivity analysis procedure in which the progression

time is assigned (i) at the midpoint, (ii) at the upper limit (reflecting the standard analysis when the progression time is assigned at the first radiologic exam showing progressive disease), or (iii) at the lower limit of the censoring interval; and finally, two multiple imputations are described considering a uniform or the nonparametric maximum likelihood estimation (NPMLE) distribution. *Clin Cancer Res*; 22(23): 5629–35. ©2016 AACR.

Introduction

In oncology, progression-free survival (PFS) is often used as the primary endpoint in clinical trials (1), as using the response rate to evaluate targeted therapies has been strongly criticized. PFS appears as a relevant alternative, although it does not represent a direct clinical benefit, such as overall survival. Although this endpoint was recommended by a consortium of American experts (2), it remains controversial. Fleming and colleagues showed that this endpoint was interesting to document efficacy of a treatment, but many biases could occur (3). One of them is related to the imaging schedule. In fact, for patients with solid tumors, the disease progression is based on the radiologic RECIST (Response Evaluation Criteria in Solid Tumors Version 1.1) criteria. Thus, imaging visits follows a previously planned schedule, but the precise date of progression is then unknown but occurs between two assessment dates. It might be thought that the solution is to suggest a more compact imaging schedule, but that is not in the interest of patients and implies too much financial expenditure (4). Moreover, in the context of data

sharing or retrospective analysis of "real-life" data, we have to deal with interval censoring while assessing PFS over time.

Since the 1950s, several methods have emerged to deal with interval-censored data, which may be parametric or not, as described precisely by Kim (5). Beyond those methods, multiple imputation procedures may also be considered, because the issue of interval censoring might be defined as a missing data problem. Any fixed-point imputation method, such as the usual PFS estimation, does not take the whole uncertainty into account in its variance estimate, and multiple imputation methods may, therefore, be very useful to estimate the PFS without bias and its variance with all the variability due to the interval censoring.

Some of these methods are implemented in statistical software (6): the Expectation–Maximization (EM) algorithm (7), the Iterative Convex Minorant (ICM) algorithm (8), and the EM-ICM algorithm (9). Although tutorials for them exist, some may be obsolete or not only for medical research purposes (10, 11).

Although these methods are now widely implemented and recommended in oncology by the European Medicines Agency and the U.S. Food and Drug Administration as part of a sensitivity analysis (12–14), only a few clinical trials have assessed PFS with methods that take interval-censored data into account (15–19).

To make these methods as accessible as possible, we provide a step-by-step guide from data formatting to output interpretation of interval-censoring methods using SAS and R software. To do so, each step was conducted in the database of a real phase II clinical trial assessing the efficacy of a first-line therapeutic regimen for patients with metastatic colorectal cancer.

Preamble

Dataset context

As an example of interval-censored data in oncology, 58 patients with unresectable, metastatic colorectal cancer in a

¹Department of Clinical Research, François Baclesse Cancer Center, Caen, France. ²Clinical and Epidemiological Research Unit, Bergonié Cancer Institute, Bordeaux, France. ³Biostatistics and Therapeutics Evaluation Unit, Léon Bérard Cancer Center, Lyon, France. ⁴Department of Biostatistics, Curie Cancer Institute, Paris, France. ⁵Epidemiology and Biostatistics Unit, Antoine Lacassagne Cancer Center, Nice, France.

Note: Supplementary data for this article are available at Clinical Cancer Research Online (<http://clincancerres.aacrjournals.org/>).

Corresponding Author: Audrey Emmanuelle Dugué, François Baclesse Cancer Centre, Avenue du Général Harris, Caen 14000, France. Phone: 332-3145-5002; Fax: 332-3145-5158; E-mail: audrey.dugue@baclesse.unicancer.fr

doi: 10.1158/1078-0432.CCR-16-1017

©2016 American Association for Cancer Research.

nonrandomized phase II trial were reanalyzed. They received biweekly a 90-minute intravenous infusion of bevacizumab (5 mg/kg), followed by a 90-minute intravenous infusion of irinotecan (180 mg/m²), followed by a simplified LV5FU2 regimen [leucovorin (400 mg/m²) and bolus fluorouracil (400 mg/m²) on day 1 and a 46-hour infusion of fluorouracil (2,400 mg/m²)] (20).

Data formatting

PFS was defined as the time from study entry to the occurrence of the "event," namely progression or death. To compute any type of survival analyses, data have to be properly formatted with regard to the following:

- Complete observations dataset: For PFS, this means that all patients have progressed or died, and the only variable needed is the time elapsed until the occurrence of the event. This is really rare in clinical studies.
- Right-censored dataset: This is when the follow-up was not long enough to allow all events to occur during the study. In such cases, two variables are needed in the dataset: a status variable to know whether the event occurred (status = 1) or not (status = 0) and a time variable describing the time elapsed until the last news (if status = 0) or until the occurrence of the event (if status = 1).
- Interval-censored dataset: This means that the exact moment at which the event occurred is unknown, although there has been a time interval during which the event occurred. This is the focus of this article. In such cases, two variables are needed in the dataset: T1 and T2. The time period between T1 and T2 is the interval during which the event occurred. Data have to be formatted as follows:
 - For patients who died without progression (uncensored data):
 - T1 = T2 = survival time
 - For patients who have not progressed or died during the study (right-censored data):
 - T1 = time elapsed to last visit
 - T2 = infinite time (missing data for SAS or R)
 - For patients who have progressed (interval-censored data):
 - T1 = time elapsed to last visit showing no progression
 - T2 = time elapsed to first visit showing disease progression

If only a few or no patients are interval censored in the dataset, it is pointless to use methods that deal with interval-censored data.

Statistical analysis

In this article, we estimate PFS over time using three different procedures to compare advanced and more standard methods. The pros and cons of these procedures are presented in Table 1.

- "Interval-censored procedure": To assess PFS precisely over time, we first computed a nonparametric maximum likelihood estimation (NPMLE) of the survival curve using the EM-ICM algorithm (9).
- "Sensitivity analysis procedure": For the sake of simplicity, we also suggested a method proposed by Panageas and

colleagues using standard PFS estimations (Kaplan–Meier estimator) that assumes the progression time is (i) the middle of the interval, (ii) the upper limit of the interval (which corresponds to the usual procedure while considering the date of progression as being the date of the first radiographic scan showing progression), and (iii) the lower limit of the interval (21).

- "Multiple imputations": In parallel, we decided to present two multiple imputation methods to estimate PFS and confidence interval (CI), as the issue of interval censoring may be regarded as a missing data problem. The first one consisted of **imputed data according to a uniform distribution** in the censoring interval, and the second one consisted of **imputed data according to the NPMLE distribution**. Those methods might be less time-consuming than the interval-censored procedure presented above, but they are not straightforward when implemented in SAS and R. Thus, we have decided not to present them in details, but to provide only the results to compare them with the two other methods we proposed.

For the sensitivity procedure, CIs of PFS estimation were computed on the basis of the log hazard, namely $\log(-\log(\text{PFS}))$.

Statistical analyses were done using the SAS software 9.3 (Copyright 2002–2010 SAS Institute Inc.) and the R-3.0.1 software (Copyright 2014 The R Foundation for Statistical Computing) with interval, survival, splines, perm, lcens, and MLEcens packages (22–28).

The dataset used for this article is available in Supplementary Material S1. SAS and R input and output are presented in Supplementary Materials S2 and S3. The source code of multiple imputations methods is available in Supplementary Material S4.

Data Visualization and Description

After importing the data, it is necessary to first describe them in terms of number of patients with complete, right-censored, and interval-censored data for PFS.

In SAS

```
DATA WORK.MyData;
  set WORK.MyData;
  if T2>T1 then censoring='interval';
  if T2=. AND T1 NE. then censoring =
    'right';
  if T2=T1 then censoring='none';
run;
PROC FREQ DATA=WORK.MyData;
  tables censoring;
run;
```

In R

```
MyData<- within(MyData, censoring
<-ifelse(is.na(T2) & !is.na(T1),
'right',
  ifelse(T2!=T1, 'interval',
    ifelse(T2==T1, 'none', NA)))
table(MyData$censoring)
```

Table 1. Pros and cons of the three procedures

	Pros	Cons
NPMLE PFS	Least biased analysis Reflects all the uncertainty in variance estimation	Requires knowing the last radiologic visit without progression Time-consuming The most difficult to interpret
PFS sensitivity analysis	Reflects medical care Easy method Balanced bias	Requires knowing the last radiologic visit Returns three results, so not so easy to interpret Is sensitive to visits calendar, so is hardly comparable with literature results not using the same visits calendar
MI PFS	Balanced bias Easy to interpret	Requires knowing the last radiologic visit without progression Not straightforward when implemented in software

NOTE: (1) NPMLE PFS: taking interval-censored data into account by NPMLE; (2) PFS sensitivity analysis: defining (i) the midpoint, (ii) the upper limit (standard analysis), and (iii) the lower limit of censoring interval as progression time; (3) MI PFS: using uniform or NPMLE distribution. Abbreviations: MI, multiple imputation; NPMLE, nonparametric maximum likelihood estimation.

Interpretation

The dataset is composed of three variables: the subject identifier, T1, and T2. Our dataset contains 53 (91%) patients with interval-censored data, 3 (5%) patients with right-censored data, and 2 (3%) patients with complete data for PFS estimation. In such a case, it is recommended to estimate PFS with methods dealing with interval-censored data, as described below.

Interval-Censored Procedure

In SAS

PFS estimation. If version 13.1 of SAS/STAT in SAS 9.4 TS Level 1M1 is installed, the ICLIFETEST procedure may be used with the help of the well-described SAS support (29). Otherwise, you may use the EMICM macro, which may be downloaded from its help page (<http://support.sas.com/kb/24/980.html>). The macro has to be first compiled in SAS before running it with the data imported:

```
%EMICM(data=WORK.MyData, left=T1,
right=T2,
method=EMICM, out=intervalPFS,
options=plot, mRS=100);
```

Outputs should appear in the Results Viewer tab, and a new table "Intervalpfs" should be created in the work library, containing survival probabilities and SEs for each time interval defined with "Lower" and "Upper" columns.

PFS comparison. The ICLIFETEST procedure provides a TEST statement to compare several groups with a weighted generalized log-rank test. Otherwise, the ICSTEST macro allows for comparison of survival curves using generalized log-rank tests as well (30, 31). The SAS support of ICLIFETEST procedure well described the difference between both (29).

In R

Nonparametric estimation and comparison of interval-censored survival data are available in an add-on package called "interval."

PFS estimation.

```
intervalPFS<- icfit(Surv(T1, T2, type
="interval2") ~ 1, data = MyData,
conf.int=TRUE, control =
icfitControl(B=100, seed=1234))
```

The CI is estimated by bootstrapping, which is a resampling method that can be very time-consuming. By default, the number

of bootstrap samples is set at 200. To speed up the procedure, we reduced this number to 100 samples with the `icfitControl` function. Moreover, as bootstrapping relies on random sampling, the random seed is set to an arbitrary value so that the estimation of the CI does not change between two calls of the `icfit` function with the same entry values.

As the "intervalPFS" object contains only the event probability and not the cumulative survival probability, to see and plot the interval times with associated PFS estimation, call:

```
intPFS<- (1-cumsum(intervalPFS$pf))
intPFS<-data.frame(time.interval=
names(intPFS), PFS=intPFS, row.
names=NULL)
intPFS
plot(intervalPFS)
```

PFS comparison. The package interval also provides a function to compare several groups, called "icest" (22). Three different types of scores are available: the most commonly used log-rank scores of Sun (32), the scores associated with the grouped proportional hazards model of Finkelstein (33), and the generalized Wilcoxon-Mann-Whitney scores. The other options for scores only allow the permutation methods and follow cases where the error under the grouped continuous model is either normally distributed or distributed by some other distribution (34).

Sensitivity Analysis Procedure

Two new variables (a status and a time variable) are computed to conduct this analysis: Right-censored patients are labeled as censored (status = 0), with the time of censoring being T1; deceased patients are labeled as not censored (status = 1), with the time of the event at the time of death; and progressive patients are labeled as not censored (status = 1), with the time of the event at (i) the midpoint between T1 and T2, (ii) the lower limit of the interval, or (iii) the upper limit of the interval for the three separate analyses.

Midpoint of the interval

In SAS.

```
DATA WORK.MyData;
set WORK.MyData;
if T2=. then do;
status=0;
midpoint=T1; end;
```

Dugué et al.

```

if T2 NE . then do;
  status=1;
  midpoint=(T2+T1)/2; end;
run;

```

Now the LIFETEST procedure may be used to estimate the "midpoint" PFS over time:

```

PROC LIFETEST DATA=WORK.MyData
  PLOTS=s OUTSURV=WORK.midpointPFS;
  TIME midpoint*status(0);
run;

```

Outputs of the PFS should appear in the Results Viewer tab and a new table "Midpointpfs" should be created in the work library, containing survival probabilities and a 95% CI for each "event" time defined in the "Midpoint" column.

In R.

```

MyData<- within(MyData, status<-
  ifelse(is.na(T2), 0, 1))
MyData<- within(MyData, midpoint<-
  ifelse(is.na(T2), T1, (T2+T1)/2))

```

Now the "survfit" function of the "survival" package may be used to estimate the "midpoint" PFS over time:

```

midpointPFS <-survfit(Surv(midpoint,
  status)~1, data = MyData, conf.type=
  "log-log")

```

Outputs of this "midpoint" PFS are now saved in the "midpointPFS" object. The estimation is summarized with:

```
midpointPFS
```

Then the PFS estimation may be plotted over time by calling:

```

summary(midpointPFS)
plot(midpointPFS)

```

Upper limit of the interval

With this procedure considering progression at the date of first appearance, the status and event variables are the same as those already computed in the "midpoint procedure," but the time variable is equal to T2 in the event of interval censoring:

In SAS.

```

DATA WORK.MyData;
  set WORK.MyData;
  upper=midpoint;
  if T2 NE . AND T1 NE . then upper=T2;
run;

```

The LIFETEST procedure may then be used to estimate the upper PFS over time, as described in the midpoint procedure section, while replacing the "midpoint" variable by the "upper" variable.

In R.

```

MyData <- within(MyData, upper <-
  ifelse(!is.na(T2) & !is.na(T1),
    T2, midpoint))

```

The "survfit" function of the "survival" package may then be used to estimate the upper PFS over time, as described in the midpoint procedure section, while replacing the "midpoint" variable by the "upper" variable.

Lower limit of the interval

As detailed above, the time variable is now equal to T1 in the event of interval censoring.

Output interpretation

Interval-censored procedure

The shape of PFS assessed by an, NPMLE taking into account interval-censored data is similar to a Kaplan–Meier curve except for gray rectangles. In fact, as the exact PFS is unknown (no unique solution) in these periods of time, the rectangles display all the possibilities that PFS may be.

For each time interval, output tables returned by SAS and R display PFS probability. The main difference between them only relies on the display: **In R output, there are less time intervals than in SAS output because consecutive time intervals with the same PFS probabilities are pooled together.**

To estimate median PFS, the point where the survival probability reaches 0.5 must be established, namely during the time interval of 9.66 to 9.79 months. As the PFS is between 0.5329 and 0.3380 in this time interval, it is then simple to estimate that the median PFS is approximately reached at 9.68 months (Table 2).

95% CI of median PFS.

In SAS. It is not easy to estimate a CI of the median PFS with the EMICM macro, but SEs of survival estimations are available for each time interval.

In R. To find the 95% CI of this median, the same must be done in both columns of the 95% CI of survival over time (lower and upper bound).

- For the lower bound, one may call:

```

lower<-with(intervalPFS$CI,{i<-which(
  lower==min(lower[lower>=0.5])));
i<-c(i,tail(i,1)+1); data.frame
  (time=round(time[i],2), lower=
  lower[i]))

```

Table 2. Median PFS estimation with 95% CI obtained with the three procedures

	Median	95% CI
NPMLE PFS	9.68	(5.65–11.07)
PFS sensitivity analysis		
Midpoint	9.02	(6.79–10.35)
Upper limit (standard analysis)	10.22	(8.31–11.53)
Lower limit	7.62	(5.75–9.23)
MI PFS		
Uniform distribution	8.98	(6.82–10.05)
NPMLE distribution	9.01	(7.07–9.80)

NOTE: (1) NPMLE PFS: taking interval-censored data into account by NPMLE; (2) PFS sensitivity analysis: separately defining (i) the midpoint, (ii) the upper limit (standard analysis), and (iii) the lower limit of censoring interval as progression time; (3) MI PFS: using uniform or NPMLE distribution.

Abbreviations: MI, multiple imputation; NPMLE, nonparametric maximum likelihood estimation.

The probability at 5.65 months falls from 0.53 to 0.48. So, in Fig. 1, the horizontal line representing a PFS of 0.5 (definition of median survival) cuts the lower line of the CI at this time.

- For the upper bound, the same code may be called while replacing lower by upper: The probability at 10.18 months is 0.55, and at 13.54 months it is 0.37. So, the probability 0.5 is reached within this time interval, at 11.07 months.

PFS estimation at a precise time. One may also need to estimate PFS at a precise time. To do so, the interval to which this time belongs must be found. For example, at 12 months, which belongs to the time interval of 10.18 to 13.54 months, as seen in the R output, or to 11.79 to 12.65 months, as seen in the SAS output, the associated survival probability is 0.248. At 10 months, which is not within any interval, the associated survival probability is between 0.338 and 0.248.

Sensitivity analysis procedure

In this procedure, PFS is assessed using the standard Kaplan-Meier estimator. The median PFS is thus directly estimated by SAS and R (Table 2). To estimate PFS at a precise time, the nearest time that is lower than or equal to the target time must

be found in the output table. For example, at 12 months, midpoint PFS is the same as at 11.385 months, namely 0.259 [95% CI (0.155–0.375)].

As described above for the midpoint PFS procedure, the estimation of the median PFS while considering the upper limit of the censoring interval is 10.215 months (95% CI, 8.310–11.530; Table 2). At 12 months, PFS is 0.310 (95% CI, 0.197–0.430). The estimation of the median PFS while considering the lower limit of the censoring interval is 7.62 months, (95% CI, 5.75–9.23; Table 2). At 12 months, the PFS is 0.207 (95% CI, 0.114–0.318).

Multiple imputation procedures

Both multiple imputation procedures gave close results (Table 2), so only the multiple imputation using the NPMLE distribution is presented in Fig. 1. We observed that the CI obtained with the multiple imputation procedure was narrower than the one obtained with the bootstrap resampling of the EM-ICM algorithm. In fact, as detailed by Hsu and colleagues, even with the multiple imputation based on the NPMLE, not all the uncertainty is taken into account, because those imputations are conditional on the data distribution and may not be viewed as a perfect multiple imputation procedure (35). Hsu and colleagues also propose adding a bootstrap stage before the imputation stage, but that makes it more time-consuming, thus losing this advantage.

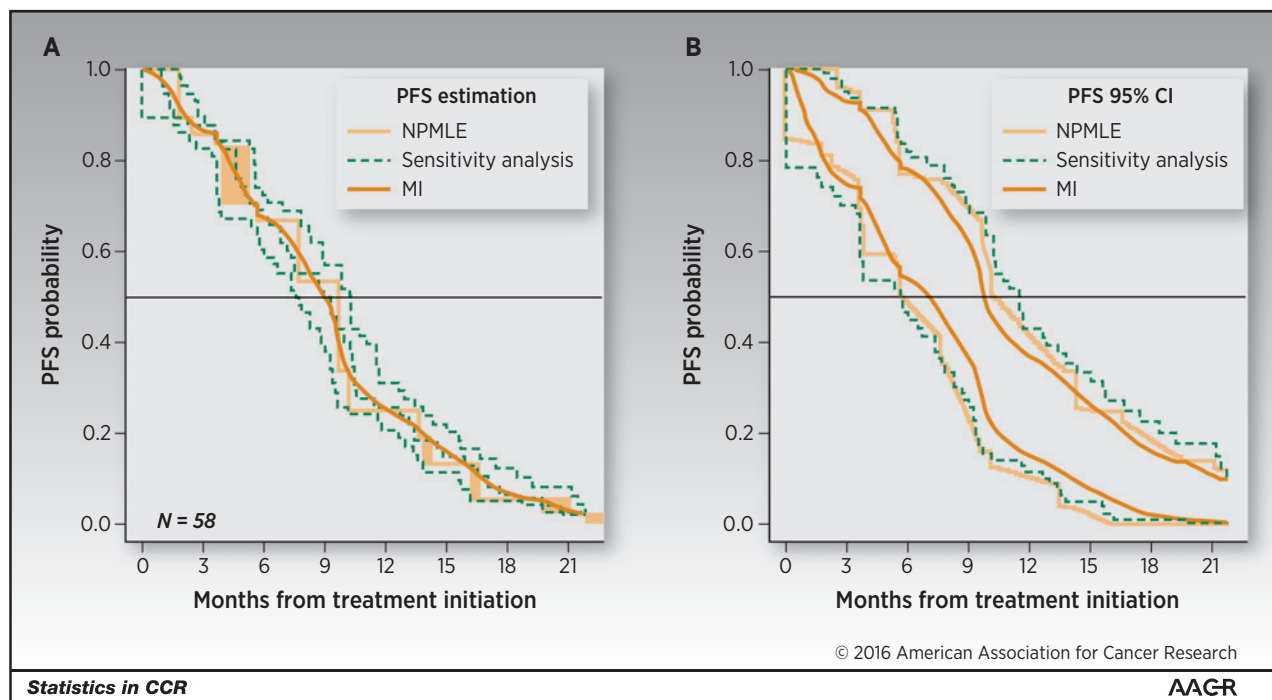


Figure 1.

PFS over time (A) and associated 95% CIs (B) obtained with the three procedures: (1) NPMLE PFS: taking interval-censored data into account; (2) PFS sensitivity analysis: separately defining (i) the midpoint, (ii) the upper limit (standard analysis), and (iii) the lower limit of censoring interval as progression time; and (3) MI PFS using the NPMLE distribution. B, For the PFS sensitivity analysis, the lower bound plotted is the lower bound of the 95% CI obtained while defining the lower limit of censoring interval as progression time, whereas the upper bound is the upper bound of the 95% CI obtained while defining the upper limit of censoring interval as progression time. MI, multiple imputation; NPMLE, nonparametric maximum likelihood estimation.

Discussion

In this article, we explain how to estimate PFS by using three different methods with interval-censored data using SAS and R statistical software and applied to a real dataset of patients with metastatic colorectal cancer enrolled in a phase II clinical trial.

As expected, the standard procedure (which is the procedure considering progression at the lower limit of the censoring interval) overestimates PFS in this example. In clinical trials with short time intervals between visits, this error may be negligible. The "midpoint procedure" is an easy-to-use method, but its scope of use is limited when used alone (36). We have, therefore, proposed a sensitivity analysis with the three systematic imputations: midpoint, upper limit, and lower limit of the interval, as recommended by Panageas and colleagues (21). The results obtained with the sensitivity analysis are similar to the results obtained with the other two procedures, however more complex.

The three procedures proposed here (the sensitivity analysis procedure, the interval-censored procedures, and the multiple imputation methods) require that the time of last visit before progression of each patient has to be recorded in the dataset, which is not always the case in the case report form. Other methods, such as the maximum bias, exist to analyze the worst-case scenario.

Beyond the wish to estimate PFS without bias, all the uncertainty in its variance estimate has to be taken into account. One solution is to add a bootstrap resampling stage, but it makes the procedures more time-consuming. Semiparametric maximum likelihood estimators or methods for grouped data (where all subjects follow the same assessment schedule) may be preferred in the case of large dataset (4, 37, 38).

Finally, caution is required when dealing with cancer types where progression is not only assessed by radiologic techniques. In settings such as gastrointestinal stromal tumor, progression may be established during a clinical examination, so the interval

to which the exact time of progression belongs may be unknown. However, it would certainly be shorter than with radiologic exams. Thus, we do not recommend the use of methods dealing with interval-censored data when these intervals are unknown or when the censoring mechanism is not independent of the event process.

As the dataset used here is available from the Supplementary Material, readers may practice before using such methods on their own data.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: A.E. Dugué

Development of methodology: A.E. Dugué

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): S. Chabaud

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): A.E. Dugué, J. Gal

Writing, review, and/or revision of the manuscript: A.E. Dugué, M. Pulido, S. Chabaud, L. Belin, J. Gal

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): A.E. Dugué

Acknowledgments

We thank Dr. Yves Becouarn from the Bergonié Cancer Institute for providing us the dataset of his clinical trial and Ray Cooke for his help in language and content editing of this manuscript. We thank the anonymous referees and the senior editor for their thoughtful comments, which have helped improve the presentation of this article.

Grant Support

This study was supported by François Badlesse Cancer Centre (Caen, France).

Received April 21, 2016; revised August 4, 2016; accepted August 12, 2016; published OnlineFirst September 23, 2016.

References

- Dhani N, Siu LL. Clinical trials and biomarker development with molecularly targeted agents and radiotherapy. *Cancer Metastasis Rev* 2008;27:339–49.
- Seymour L, Ivy SP, Sargent D, Spriggs D, Baker L, Rubinstein L, et al. The design of phase II clinical trials testing cancer therapeutics: consensus recommendations from the clinical trial design task force of the National Cancer Institute Investigational Drug Steering Committee. *Clin Cancer Res* 2010;16:1764–9.
- Fleming TR, Rothmann MD, Lu HL. Issues in using progression-free survival when evaluating oncology products. *J Clin Oncol* 2009;27:2874–80.
- Sun X, Li X, Chen C, Song Y. A review of statistical issues with progression-free survival as an interval-censored time-to-event endpoint. *J Biopharm Stat* 2013;23:986–1003.
- Kim J. Statistical inference of interval-censored failure time data. Seoul, Korea: The Korean Statistical Society; 2011.
- Zhang Z, Sun J. Interval censoring. *Stat Methods Med Res* 2010;19:53–70.
- Turnbull BW. The empirical distribution function with arbitrarily grouped, censored and truncated data. *J R Stat Soc Series B Stat Methodol* 1976;290–5.
- Groeneboom P, Wellner JA. Information bounds and nonparametric maximum likelihood estimation. 19 ed. Basel, Switzerland: Birkhäuser Basel; 1992.
- Wellner JA, Zhan Y. A hybrid algorithm for computation of the nonparametric maximum likelihood estimator from censored data. *J Am Stat Assoc* 1997;92:945–59.
- Gomez G, Calle ML, Oller R, Langohr K. Tutorial on methods for interval-censored data and their implementation in R. *Stat Model* 2009;9:259–97.
- Lindsey JC, Ryan LM. Tutorial in biostatistics methods for interval-censored data. *Stat Med* 1998;17:219–38.
- Stone AM, Bushnell W, Denne J, Sargent DJ, Amit O, Chen C, et al. Research outcomes and recommendations for the assessment of progression in cancer clinical trials from a PhRMA working group. *Eur J Cancer* 2011;47:1763–71.
- Committee for Medicinal Products for Human Use. Appendix 1 to the guideline on the evaluation of anticancer medicinal products in man (CHMP/EWP/205/95 REV. 3). Methodological considerations for using progression-free survival (PFS) as primary endpoint in confirmatory trials for registration. London, United Kingdom: European Medicines Agency; 2008.
- U.S. Food and Drug Administration. Guidance for industry: clinical trial endpoints for the approval of cancer drugs and biologics; 2007. Available from: <http://www.fda.gov/downloads/Drugs/.../Guidances/ucm071590.pdf>.
- Fernandez MC, Krailo MD, Gerbing RR, Matthay KK. A phase I dose escalation of combination chemotherapy with granulocyte-macrophage-colony stimulating factor in patients with neuroblastoma. *Cancer* 2000;88:2838–44.
- Krug LM, Kindler HL, Calvert H, Manegold C, Tsao AS, Fennell D, et al. Vorinostat in patients with advanced malignant pleural mesothelioma who have progressed on previous chemotherapy (VANTAGE-014): a phase 3, double-blind, randomised, placebo-controlled trial. *Lancet Oncol* 2015;16:447–56.
- Masterson TA, Serio AM, Mulhall JP, Vickers AJ, Eastham JA. Modified technique for neurovascular bundle preservation during radical

- prostatectomy: association between technique and recovery of erectile function. *BJU Int* 2008;101:1217–22.
18. Robert C, Dummer R, Gutzmer R, Lorigan P, Kim KB, Nyakas M, et al. Selumetinib plus dacarbazine versus placebo plus dacarbazine as first-line treatment for BRAF-mutant metastatic melanoma: a phase 2 double-blind randomised study. *Lancet Oncol* 2013;14:733–40.
 19. Sakamoto J, Teramukai S, Nakazato H, Ohashi Y. A re-analysis of a randomized clinical trial for gastric cancer using interval censoring. *Jpn J Clin Oncol* 1997;27:445–6.
 20. Becouarn Y, Cany L, Pulido M, Beyssac R, Texereau P, Le MV, et al. FOLFIRI(R) and bevacizumab in first-line treatment for colorectal cancer patients: safety, efficacy and genetic polymorphisms. *BMC Res Notes* 2014;7:260.
 21. Panageas KS, Ben-Porat L, Dickler MN, Chapman PB, Schrag D. When you look matters: the effect of assessment schedule on progression-free survival. *J Natl Cancer Inst* 2007;99:428–32.
 22. Fay MP, Shaw PA. Exact and asymptotic weighted logrank tests for interval censored data: the interval R package. *J Stat Softw* 2010;36:pii:i02.
 23. Therneau TM, Grambsch PM. Modeling survival data: extending the Cox model. New York, NY: Springer Verlag; 2000.
 24. Therneau T. A package for survival analysis in S, Version 2.38. Available from: <http://cran.r-project.org/web/packages/snow/index.html> 2015.
 25. Team RC. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2012.
 26. Gentleman R, Vandal A. Icnens: NPMLE for censored and truncated data. R package version 1.24. 0; 2012. Available from: <http://www.bioconductor.org/packages/release/bioc/html/icnens.html>.
 27. Groeneboom P, Jongbloed G, Wellner JA. The support reduction algorithm for computing non-parametric function estimates in mixture models. *Scand J Stat* 2008;35:385–99.
 28. Maathuis MH. Reduction algorithm for the NPMLE for the distribution function of bivariate interval-censored data. *J Comput Graph Stat* 2012;14:352–62.
 29. Guo C, So Y, Johnston G. Analyzing interval-censored data with the ICLIFETEST procedure. SAS Institute, Inc.; 2014. Available from: <https://support.sas.com/resources/papers/proceedings14/SAS279-2014.pdf>.
 30. Sun J, Zhao Q, Zhao X. Generalized log-rank tests for interval-censored failure time data. *Scand J Stat* 2005;32:49–57.
 31. Zhao Q, Sun J. Generalized log-rank test for mixed interval-censored failure time data. *Stat Med* 2004;23:1621–9.
 32. Sun J. A non-parametric test for interval-censored failure time data with application to AIDS studies. *Stat Med* 1996;15:1387–95.
 33. Finkelstein DM. A proportional hazards model for interval-censored failure time data. *Biometrics* 1986;42:845–54.
 34. Fay MP. Rank invariant tests for interval censored data under the grouped continuous model. *Biometrics* 1996;52:811–22.
 35. Hsu CH, Taylor JM, Murray S, Commenges D. Multiple imputation for interval censored data with auxiliary variables. *Stat Med* 2007;26:769–81.
 36. Sun X, Chen C. Comparison of Finkelstein's method with the conventional approach for interval-censored data analysis. *Stat Biopharm Res* 2010;2:97–108.
 37. Farrington CP. Interval censored survival data: a generalized linear modelling approach. *Stat Med* 1996;15:283–92.
 38. Goetghebuer E, Ryan L. Semiparametric regression analysis of interval-censored data. *Biometrics* 2000;56:1139–44.

Clinical Cancer Research

How to Deal with Interval-Censored Data Practically while Assessing the Progression-Free Survival: A Step-by-Step Guide Using SAS and R Software

Audrey Emmanuelle Dugué, Marina Pulido, Sylvie Chabaud, et al.

Clin Cancer Res 2016;22:5629-5635. Published OnlineFirst September 23, 2016.

Updated version Access the most recent version of this article at:
doi:[10.1158/1078-0432.CCR-16-1017](https://doi.org/10.1158/1078-0432.CCR-16-1017)

Supplementary Material Access the most recent supplemental material at:
<http://clincancerres.aacrjournals.org/content/suppl/2016/09/23/1078-0432.CCR-16-1017.DC1>

Cited articles This article cites 28 articles, 2 of which you can access for free at:
<http://clincancerres.aacrjournals.org/content/22/23/5629.full#ref-list-1>

E-mail alerts [Sign up to receive free email-alerts](#) related to this article or journal.

Reprints and Subscriptions To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions To request permission to re-use all or part of this article, use this link
<http://clincancerres.aacrjournals.org/content/22/23/5629>.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.