ORIGINAL ARTICLE

Visualizing the quantile survival time difference curve

Harald Heinzl PhD, Professor ⓘ | Martina Mittlboeck PhD, Professor ⓘ

Section for Clinical Biometrics, Center for Medical Statistics, Informatics, and Intelligent Systems, Medical University of Vienna, Vienna, Austria

Correspondence
Martina Mittlboeck, CeMSIIS, Medical University of Vienna, Spitalgasse 23, A-1090 Vienna, Austria.
Email: martina.mittlboeck@muv.ac.at

Abstract

The difference between the pth quantiles of 2 survival functions can be used to compare patients' survival between 2 therapies. Setting $p = 0.5$ yields the median survival time difference. Varying p between 0 and 1 defines the quantile survival time difference curve which can be straightforwardly estimated by the horizontal differences between 2 Kaplan-Meier curves. The estimate's variability can be visualized by adding either a bundle of resampled bootstrap step functions or, alternatively, approximate bootstrap confidence bands. The user-friendly SAS software macro %kmdiff enables the straightforward application of this exploratory graphical approach. The macro is described, and its application is exemplified with breast cancer data. The advantages and limitations of the approach are discussed.

KEYWORDS

bootstrap samples, confidence bands, exploratory data analysis, Kaplan-Meier curves, SAS macro

1 | INTRODUCTION

As consistent estimator of the survival function, the Kaplan-Meier curve[1] is the most commonly used graphical tool in survival analysis. It is extensively used to visually compare censored survival time curves between groups of patients distinguished by different therapies, biomarker categories, or demographic features. Estimates for survival probabilities at specific times (eg, 3-year survival probability) and specific quantiles of the survival time distribution (eg, median survival time) can be easily obtained.

The diagonal upper-left-to-lower-right nature of the plot, however, hampers the visual assessment of the difference between the pth quantiles of 2 Kaplan-Meier curves. To overcome this problem, the quantile survival time difference curve is defined and straightforwardly estimated by calculating the horizontal differences between the corresponding Kaplan-Meier curves. Two bootstrap-based approaches are suggested to illustrate the estimate's variability; alternatively, the traditional normal approximation method or a smoothed empirical likelihood approach could have been considered as well.[2]

The quantile survival time difference curve is not to be confused with the survival probability difference curve[3-6] and approaches related to it.[7-10] Besides that, the term "quantile difference" itself is not unambiguously defined. It could also refer to the difference

between the pth and qth quantiles of a single survival curve[11,12], eg, the interquartile range for $p = 0.75$ and $q = 0.25$.

In Section 2, the technical details of the approach are presented. The SAS macro %kmdiff is described and illustrated in Section 3. A brief discussion is given in Section 4. The SAS macro code was generated using Version 9.4 (for Windows) of the SAS software (copyright © 2002-2012 SAS Institute Inc.). SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

2 | METHODS

Assume that from a large patient population with survival function $S(t)$, right-censored survival data have been observed for a sample of n patients, where $x_1...x_n$ denote observed survival times and $a_1...a_n$ denote their corresponding censoring indicators, respectively. The Kaplan-Meier estimator of the survival function is defined as

$$\widehat{S}(t) = \prod_{t_i \le t} \left(1 - \frac{d_i}{n_i}\right), \qquad t_1 \le t \le t_K, \qquad (1)$$

where $0 \le t_1 < t_2 < ... < t_K$ are $K > 0$ different observed failure times, d_i is the number of failures, and n_i is the size of the risk set at t_i, $i = 1, ..., K$. If

no censored values have been observed before t_1, then $n_1 = n$. $\widehat{S}(t)$ is right continuous, $\widehat{S}(t) = 1$ for $0 \leq t < t_1$ and $0 \leq \widehat{S}(t_K) \leq \widehat{S}(t) \leq 1$ for $0 \leq t \leq t_K$. Note that $\widehat{S}(t_K)$ equals 0 only if all censored observations occur before t_K. If patients are still alive after t_K, then either $\widehat{S}(t)$ can be set equal to $\widehat{S}(t_K)$ from t_K to the largest censoring time or $\widehat{S}(t)$ can be considered not defined for $t > t_K$. In the current manuscript, the latter definition is applied. Consequently, in the case of $K = 0$, that is no observed failure times, $\widehat{S}(t)$ equals 1 for $t = 0$ and is not defined for $t > 0$.

The pth quantile survival time $Q(p)$ corresponds to the labelling of the survival function axis of the Kaplan-Meier plot and is equivalent to the common $(1 - p)$th quantile of a distribution function. For $\widehat{S}(t_K) \leq p \leq 1$ and $K > 0$, the estimator for the pth quantile survival time can be defined as

$$\widehat{Q}(p) = \min\left(t_j | \widehat{S}(t_j) \leq p\right). \tag{2}$$

Now assume that from a second large patient population (independent from the first one) with survival function $S'(t)$, right-censored survival data have been observed as well for a sample of m patients. Their observed survival times and censoring indicators are denoted by $y_1...y_m$ and $b_1...b_m$, respectively, and $0 \leq u_1 < u_2 < ... < u_L$ are the L different observed failure times. By analogy with formulae (1) and (2), a corresponding estimator $\widehat{Q'}(p)$ for the pth quantile survival time $Q'(p)$ is obtained, $\widehat{S'}(u_L) \leq p \leq 1$. Note that the prime in the notations refers to the second population and its corresponding sample. Let

$$p^0 = \max\left(\widehat{S}(t_K), \widehat{S'}(u_L)\right). \tag{3}$$

Now, an estimator for the quantile survival time difference curve $D(p) = Q(p) - Q'(p)$ can be defined as

$$\widehat{D}(p) = \widehat{Q}(p) - \widehat{Q'}(p) \tag{4}$$

for $p^0 \leq p \leq 1$. If $p^0 = 0$, then $\widehat{D}(0) = t_K - u_L$.

It seems obvious now to plot $\widehat{D}(p)$ against p. However, for the sake of comparability with the Kaplan Meier plot, p has to be plotted on the vertical axis against $\widehat{D}(p)$ on the horizontal axis, respectively.

In a next step, the variability of the estimator $\widehat{D}(p)$ will be visualized. For this purpose, 2 bootstrap solutions are suggested, both are based on Efron's classical bootstrap for censored data.[13]

2.1 | Bootstrap bundle

Draw a sample of size n with replacement from $(x_1, a_1), (x_2, a_2)...(x_n, a_n)$ and a sample of size m with replacement from $(y_1, b_1), (y_2, b_2)...(y_m, b_m)$. These are the first bootstrap samples, $(x_1^{*1}, a_1^{*1}), (x_2^{*1}, a_2^{*1})...(x_n^{*1}, a_n^{*1})$ and $(y_1^{*1}, b_1^{*1}), (y_2^{*1}, b_2^{*1})...(y_m^{*1}, b_m^{*1})$ from where the bootstrapped quantile survival time difference curve can be computed from the 2 bootstrapped quantile survival time curves, $\widehat{D}^{*1}(p) = \widehat{Q}^{*1}(p) - \widehat{Q'}^{*1}(p)$, $p^{*1} \leq p \leq 1$. Here, p^{*1} is defined in analogy to p^0 of formula (3).

Repeating the drawing of the bootstrap samples *bundle* times eventually yields *bundle* bootstrapped quantile difference curves, $\widehat{D}^{*i}(p)$, $p^{*i} \leq p \leq 1$, $i = 1...bundle$. Depicting them together with $\widehat{D}(p)$ enables a visual assessment of the variation of $\widehat{D}(p)$. In doing so, a subdued colour like light grey should be used for the bootstrapped quantile

difference curves, and a vibrant colour like green should be used for $\widehat{D}(p)$, respectively. Setting *bundle* to a value between 40 and 200 seems reasonable when using the bootstrap bundle approach.

2.2 | Bootstrap confidence bands

The generation of a confidence band also requires bootstrapped quantile difference curves, $\widehat{D}^{*i}(p)$, $0 \leq p^{*i} \leq p \leq 1$, $i = 1...boot$. Here, the value of *boot* should depend on the chosen confidence level $100(1 - \alpha)\%$. As rule of thumb, we recommend $boot \geq 100/\alpha$, that is $boot \geq 2000$ for a 2-sided 95% confidence band and $boot \geq 10000$ for a 2-sided 99% confidence band, respectively.

A confidence band for the quantile difference curve can be constructed from a series of pointwise confidence intervals for quantile differences. For a given survival probability p, the corresponding confidence interval is determined as the $(\alpha/2)$th and the $1 - (\alpha/2)$th quantile of the empirical distribution of $\widehat{D}^{*1}(p)$, $\widehat{D}^{*2}(p)...\widehat{D}^{*boot}(p)$, where $p^*_{max} \leq p \leq 1$ and $p^*_{max} = \max (p^{*1}, p^{*2}...p^{*boot})$. Note that p^*_{max} is the smallest probability for which the quantile difference can be computed for all bootstrap replications, and so, the confidence band is available for the interval $[p^*_{max}, 1]$.

The length of this interval can become rather short as it depends on the bootstrapped Kaplan-Meier curve (out of $2 \times boot$ such curves) that drops least. By applying a conservative approach, the confidence band can now be extended for $p^*_{1 - \alpha} \leq p < p^*_{max}$ as well, where $p^*_{1 - \alpha}$ is a specifically defined $(1 - \alpha)$ quantile of $p^{*1}, p^{*2}...p^{*boot}$. According to the usual definition of quantiles, the $(1 - \alpha)$ quantile is either one of the values $p^{*1}, p^{*2}...p^{*boot}$ or it can be any chosen value from an interval, say $[p^{*j}, p^{*k}]$. In the former case, set $p^*_{1-\alpha}$ equal to that value; in the latter case, set $p^*_{1 - \alpha} = p^{*k}$, respectively.

Now, for a given survival probability p with $p^*_{1 - \alpha} \leq p < p^*_{max}$ and all i with $p^{*i} \leq p$, $\widehat{D}^{*i}(p)$ is defined and a weight of $1/boot$ can be assigned; furthermore, M is set to an arbitrary value larger than the maximum of the absolute values of these $\widehat{D}^{*i}(p)$. For all i with $p^{*i} > p$, $\widehat{D}^{*i}(p)$ is undefined and will be replaced with 2 values, $-M$ and M, with weights $0.5/boot$. Hence, the sum of all weights will be 1 and the confidence interval can be determined as the $(\alpha/2)$th and the $1 - (\alpha/2)$th quantile of the now-weighted empirical distribution.

3 | RESULTS

3.1 | SAS macro

The SAS macro %kmdiff allows the straightforward application of the method to survival data sets and is freely available at https://cemsiis.meduniwien.ac.at/en/kb/science-research/software/statistical-software/.

The macro produces the standard Kaplan-Meier plot of SAS, and plots of the quantile differences with (i) a bootstrap bundle, (ii) bootstrap confidence bands, and (iii) both a bootstrap bundle and bootstrap confidence bands. The macro parameters are described in Table 1.

3.2 | Example

Patients with primary node positive breast cancer have been recruited by the German Breast Cancer Study Group (GBSG) from July 1984 to
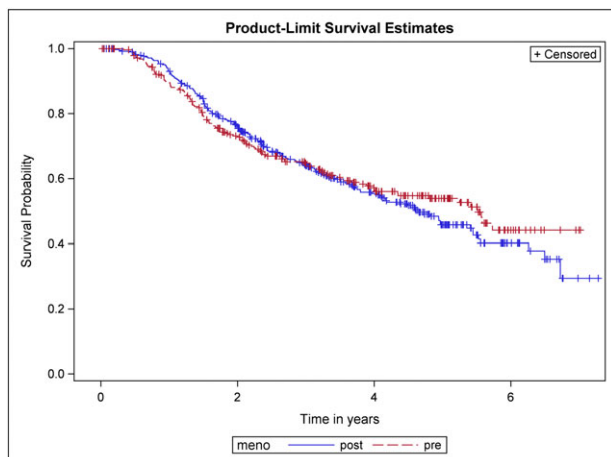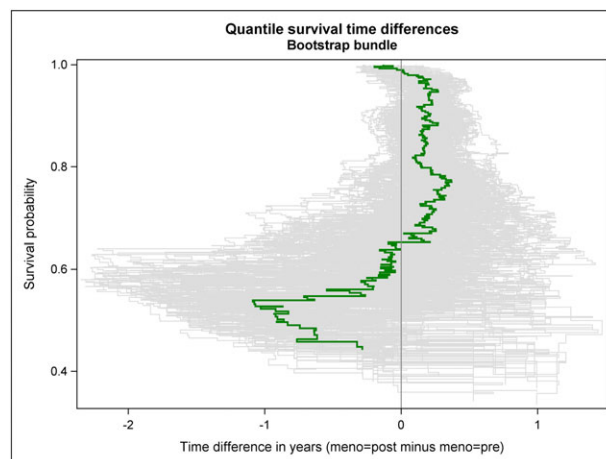
710 | WILEY—Journal of Evaluation in Clinical Practice
International Journal of Public Health Policy and Health Services Research

HEINZL AND MITTLBOECK

TABLE 1 Parameters of the SAS macro %kmdiff

Parameter	Description
data	Input SAS data set (required); its name must not start with two underscores, and it must not contain variables whose names start with two underscores (in particular, "__g", "__t", and "__status")
time	Survival time variable (required)
timeunit	Time units of the survival time variable; default is "years"
status	Survival status variable (required)
censval	Censoring status value(s) to be used in PROC LIFETEST; default is "0"; use the %str() function to specify more than 1 value
group	Covariate which distinguishes the two groups of interest (required)
gvalue1	Numerical value of first group of interest (required)
gvalue2	Numerical value of second group of interest (required)
grouplbl	Label of group variable (optional)
gvallbl1	Label of first group value (optional)
gvallbl2	Label of second group value (optional)
alpha	100-alpha % is the pointwise two-sided confidence level; default for alpha is "5"
boot	Number of bootstrap replications for computing confidence bands; default is "2000", minimum is "100"
bundle	Number of bootstrap replications shown in figures; default is "200"
seedval	Bootstrap random number seed; default is "0"

December 1989[14]; the GBSG data set can be either obtained from http://biostat.mc.vanderbilt.edu/wiki/Main/DataSets or from http://biom131.imbi.uni-freiburg.de/biom/Royston-Sauerbrei-book/. It will be used in the following for illustration purpose. The data set contains several clinical variables and the recurrence free survival (RFS) time of 686 female patients of whom 299 suffered an event (cancer recurrence or death).

In Figure 1, the Kaplan-Meier curves (product-limit survival estimates) for 290 premenopausal and 396 postmenopausal women are shown. The curves lie close to each other and intersect right before 6 months and then again right before 3 years; albeit the first intersection is hard to see. Before the first and after the second intersection, the premenopausal women show better RFS, in the time between the postmenopausal women are better off.

The estimated quantile RFS time difference curve is shown as green solid line in Figures 2–4; thereby, 3 distinct ways to visualize its variability have been applied. A bundle of 200 bootstrap

FIGURE 2 RFS probabilities are plotted against quantile RFS time differences between postmenopausal and premenopausal women with primary node positive breast cancer (green solid line). The variability of the estimated curve is illustrated by a bundle of 200 bootstrap replications (grey solid lines)

replications of the quantile RFS time difference curve are added as grey solid lines to Figure 2. A 95% confidence band is shown as green transparent area in Figure 3; it has been derived from 2000 bootstrap replications of the quantile RFS time difference curve. In Figure 4, key features of Figures 2 and 3 have been combined; however, the 95% confidence band is depicted with 2 black solid lines now. To allow independent replications of Figures 2–4, the seedval parameter (bootstrap random number seed) was set to a value larger than 0, concretely 44181.

The quantile RFS time differences curve vividly shows the extent of the horizontal gaps between the 2 Kaplan-Meier curves. In particular, the 2 intersections of the Kaplan-Meier curves can be easily distinguished in the quantile RFS time difference curve.

Both the bundle of bootstrap replications and the 95% confidence band reveal the increasing variability of the observed quantile difference with decreasing survival probability (ie, decreasing size of the risk set). Simple random fluctuations can provide a plausible explanation
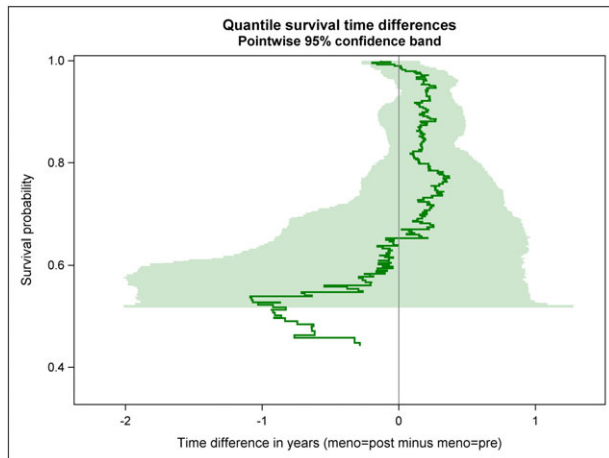

FIGURE 1 Kaplan-Meier curves showing recurrence-free survival (RFS) of postmenopausal and premenopausal women with primary node positive breast cancer

FIGURE 3 RFS probabilities are plotted against quantile RFS time differences between postmenopausal and premenopausal women with primary node positive breast cancer (green solid line). The variability of the estimated curve is illustrated by a bootstrap-based 95% confidence band (2000 replications, green transparent area)
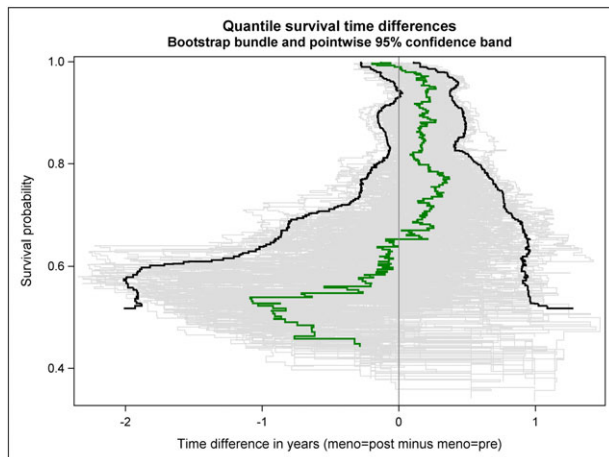

FIGURE 4 RFS probabilities are plotted against quantile RFS time differences between postmenopausal and premenopausal women with primary node positive breast cancer (green solid line). The variability of the estimated curve is illustrated by both a bundle of 200 bootstrap replications (grey solid lines) and a bootstrap-based 95% confidence band (2000 replications, black solid lines)

for the observed differences in quantile RFS times between premenopausal and postmenopausal women over nearly the whole observed probability range. The only exception is observed for survival probabilities around 94%.

4 | DISCUSSION

The SAS macro %kmdiff provides an easy-to-use computational tool to visually assess the estimated quantile survival time difference curve and its sample variability. The curve is intended as useful complement but not as replacement of the survival probability difference curve and all the refined approaches based thereon.[3-10]

The motivation to plot a quantile survival time difference curve may be threefold. Firstly, patients as well as health professionals may find time differences more intuitive and easier to interpret than

probabilities and probability differences. Secondly, the curve shows an overall picture unlike the isolated snippet provided by the commonly reported median survival time difference. And thirdly, it can become rather difficult to assess horizontal (and also vertical) differences between 2 Kaplan-Meier curves; the visual perception is often affected by the shortest (Euclidean) distance between the 2 curves.

The main purpose of the SAS macro %kmdiff is to support the exploration of (possibly time-dependent) group effects in the presence of right censoring. Using the macro for confirmatory purposes would require the prespecification of statistical hypotheses and the proper adjustment for any multiple testing; it should also be taken into account that the macro produces a pointwise confidence band.

There is a potential limitation of bootstrap-based confidence intervals for quantile differences: in particular in very small samples and risk sets, the actual coverage probability may considerably deviate from the nominal one.[15,16] Naturally, this limitation also affects the bootstrap bundle approach; that is, the graphically shown bootstrap replications might only provide a distorted picture of the variability of the estimated quantile differences in very small samples and risk sets.

Confidence intervals could also be obtained by normal approximation or a smoothed empirical likelihood method.[2] However, the former will need moderate to large sample sizes to work properly, whereas the latter requires the selection of a kernel bandwidth by cross-validation which is computationally burdensome.[2] Besides that, due to the close relationship between the bootstrap and the empirical likelihood, it seems reasonable to assume that the empirical likelihood will face similar problems as the bootstrap in very small samples and risk sets.

The presented SAS macro %kmdiff only uses Base SAS and SAS/STAT procedures for statistical computations; graphical representations are based on the ODS Graphics procedure SGPLOT. Note that the user can easily modify the SGPLOT code or replace it with purpose-built SAS code to obtain specially tailored graphical output.

In conclusion, the SAS macro %kmdiff provides a useful exploratory tool for medical researchers as it brings in an additional dimension to the assessment and communication of group differences in patient survival.

ORCID

Harald Heinzl http://orcid.org/0000-0003-4711-2379
Martina Mittlboeck http://orcid.org/0000-0003-1097-4047

REFERENCES

1. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc*. 1958;53(282):457-481.

2. Yang H, Yau C, Zhao Y. Smoothed empirical likelihood inference for the difference of two quantiles with right censoring. *J Stat Plann Infer*. 2014;146:95-101.

3. Ahmed N, Subramanian S. Semiparametric simultaneous confidence bands for the difference of survival functions. *Lifetime Data Anal*. 2016;22(4):504-530.

4. Coory M, Lamb KE, Sorich M. Risk-difference curves can be used to communicate time-dependent effects of adjuvant therapies for early stage cancer. *J Clin Epidemiol*. 2014;67(9):966-972.

5. Zhang MJ, Fine J. Summarizing differences in cumulative incidence functions. *Stat Med*. 2008;27(24):4939-4949.

6. McKeague Ian W, Zhao Y. Comparing distribution functions via empirical likelihood. *Int J Biostatistics*. 2005;1(1).

712 | WILEY—Journal of Evaluation in Clinical Practice
International Journal of Public Health Policy and Health Services Research

HEINZL AND MITTLBOECK

7. Royston P, Parmar MK. The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Stat Med.* 2011;30(19):2409-2421.

8. Shen Y, Fleming TR. Weighted mean survival test statistics: A class of distance tests for censored survival data. *J R Stat Soc Series B Stat Meth.* 1997;59(1):269-280.

9. Zhao L, Claggett B, Tian L, et al. On the restricted mean survival time curve in survival analysis. *Biometrics.* 2016;72(1):215-221.

10. Zhao L, Tian L, Uno H, et al. Utilizing the integrated difference of two survival functions to quantify the treatment contrast for designing, monitoring, and analyzing a comparative clinical study. *Clin Trials* (London, England). 2012;9(5):570-577.

11. Shen J, He S. Empirical likelihood for the difference of quantiles under censorship. *Stat Papers.* 2007;48(3):437-457.

12. Xun L, Shao L, Zhou Y. Efficiency of estimators for quantile differences with left truncated and right censored data. *Stat Probab Lett.* 2017;121:29-36.

13. Efron B. Censored data and the bootstrap. *J Am Stat Assoc.* 1981;76(374):312-319.

14. Royston P, Sauerbrei W. *Multivariable model-building; A pragmatic approach to regression analysis based on fractional polynomials for modelling continuous variables.* Hoboken, NJ: Wiley; 2008.

15. Heinzl H, Mittlboeck M. Assessing a hypothesis test for the difference between two quantiles from independent populations. *Comm Stat Simulat Comput.* 2017;46(5):3540-3552.

16. Hutson AD. A distribution function estimator for the difference of order statistics from two independent samples. *Stat Papers.* 2009;50(1):203-208.