Visualizing the quantile survival time difference curve

Harald Heinzl PhD, Professor | Martina Mittlboeck PhD, Professor

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

The \( p \)th 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 \( \hat{S}(t_{k-1}) \leq p \leq 1 \) and \( K > 0 \), the estimator for the \( p \)th quantile survival time can be defined as

\[
\hat{Q}(p) = \min \{ t | \hat{S}(t) \leq p \}.
\]

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 \( y_1, \ldots, y_m \) and \( b_1, \ldots, b_m \), respectively, and \( 0 \leq u_1 < u_2 < \ldots < u_l \) are the \( L \) different observed failure times. By analogy with formulae (1) and (2), a corresponding estimator \( \hat{Q}(p) \) for the \( p \)th quantile survival time \( Q(p) \) is obtained, \( \hat{S}(u_1) \leq p \leq 1 \). Note that the prime in the notations refers to the second population and its corresponding sample.

Let

\[
p_0 = \max \{ \hat{S}(t_{k}), \hat{S}(u_1) \}.
\]

Now, an estimator for the quantile survival time difference curve \( D(p) = Q(p) - \hat{Q}(p) \) can be defined as

\[
\bar{D}(p) = \hat{Q}(p) - Q(p)
\]

for \( p^0 \leq p \leq 1 \). If \( p^0 = 0 \), then \( \bar{D}(0) = t_{1} - u_1 \).

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

In a next step, the variability of the estimator \( \bar{D}(p) \) 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 \( \{x_1, a_1\}, \{x_2, a_2\}, \ldots, \{x_n, a_n\} \) and a sample of size \( m \) with replacement from \( \{y_1, b_1\}, \{y_2, b_2\}, \ldots, \{y_m, b_m\} \). These are the first bootstrap samples. \( \{y_1^*, b_1^*, \ldots, y_m^*, b_m^*\} \) from where the bootstrapped quantile survival time difference curve can be computed from the 2 bootstrapped quantile survival time curves, \( \bar{D}^*(p) = \hat{Q}^*(p) - \hat{Q}^{-*}(p) \), \( p^* \leq p \leq 1 \). Here, \( p^* \) 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, \( \bar{D}^*(p) \), \( p^* \leq p \leq 1 \), \( i = 1 \ldots \text{bundle} \). Depicting them together with \( \bar{D}(p) \) enables a visual assessment of the variation of \( \bar{D}(p) \). 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 \( \bar{D}(p) \), 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, \( \bar{D}^*(p) \), \( 0 \leq p^* \leq p \leq 1 \), \( i = 1 \ldots \text{...bootstrap} \). Here, the value of boot should depend on the chosen confidence level 100(1 - \( a \))%. As rule of thumb, we recommend boot \( \geq 100/a \), 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 \( \bar{D}^1(p), \bar{D}^2(p), \ldots, \bar{D}^\text{bootstrap}(p) \), where \( p^\text{max} \leq p \leq 1 \) and \( p^\text{max} = \max (p^1, p^2, \ldots, p^\text{bootstrap}) \). Note that \( p^\text{max} \) 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 \([p^\text{min}, p^\text{max}]\).

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

Now, for a given survival probability \( p \) with \( p^{1-\alpha} \leq p < p^\text{max} \) and all \( i \) with \( p^i \leq p \), \( \bar{D}^i(p) \) 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 \( \bar{D}^i(p) \). For all \( i \) with \( p^i > p \), \( \bar{D}^i(p) \) 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://cecsius. meduniwien.ac.at/en/kb/science/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
December 1989; 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 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

| 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-alph % 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” |

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

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)
3-10 all the refined approaches based thereon. but not as replacement of the survival probability difference curve and its sample variability. The curve is intended as useful complement to visually assess the estimated quantile survival time difference curve 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;144:95-101.
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).
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.

How to cite this article: Heinzl H, Mittlboeck M. Visualizing the quantile survival time difference curve. J Eval Clin Pract. 2018;24:708–712. https://doi.org/10.1111/jep.12948