## Adding a new analytical procedure with clinical interpretation in the tool box of survival analysis

In a long-term comparative oncology trial, progression-free survival or overall survival time is often the study endpoint. The hazard ratio (HR) has been routinely utilized to quantify the between-group difference in survival analysis for the past five decades [1]. The validity of HR estimation procedure depends on a strong assumption of proportional hazards (PH), i.e. the ratio of the two hazard functions is constant over the entire study period. When the PH assumption is not met, the resulting estimate is not a simple average of HRs over time and is difficult to interpret clinically [2–6]. As indicated by Fei et al. [7], a standard goodness-of-fit test for the adequacy of the PH assumption generally is not informative. First, it has insufficient statistical power to detect model misspecification when the number of events of interest in the trial is small. Second, when the number of events is large, the same test may identify even a negligible misspecification. For most immunotherapy studies discussed in the article [7], the PH assumption appeared to be violated upon visual inspection. Moreover, even when the PH assumption is plausible, it is not clear that a statistically significant HR of, for example, 0.80 (immunotherapy versus control), could be translated to inform effective clinical decision making. When the underlying hazard for the control arm is low, a reduction of 20% may not represent

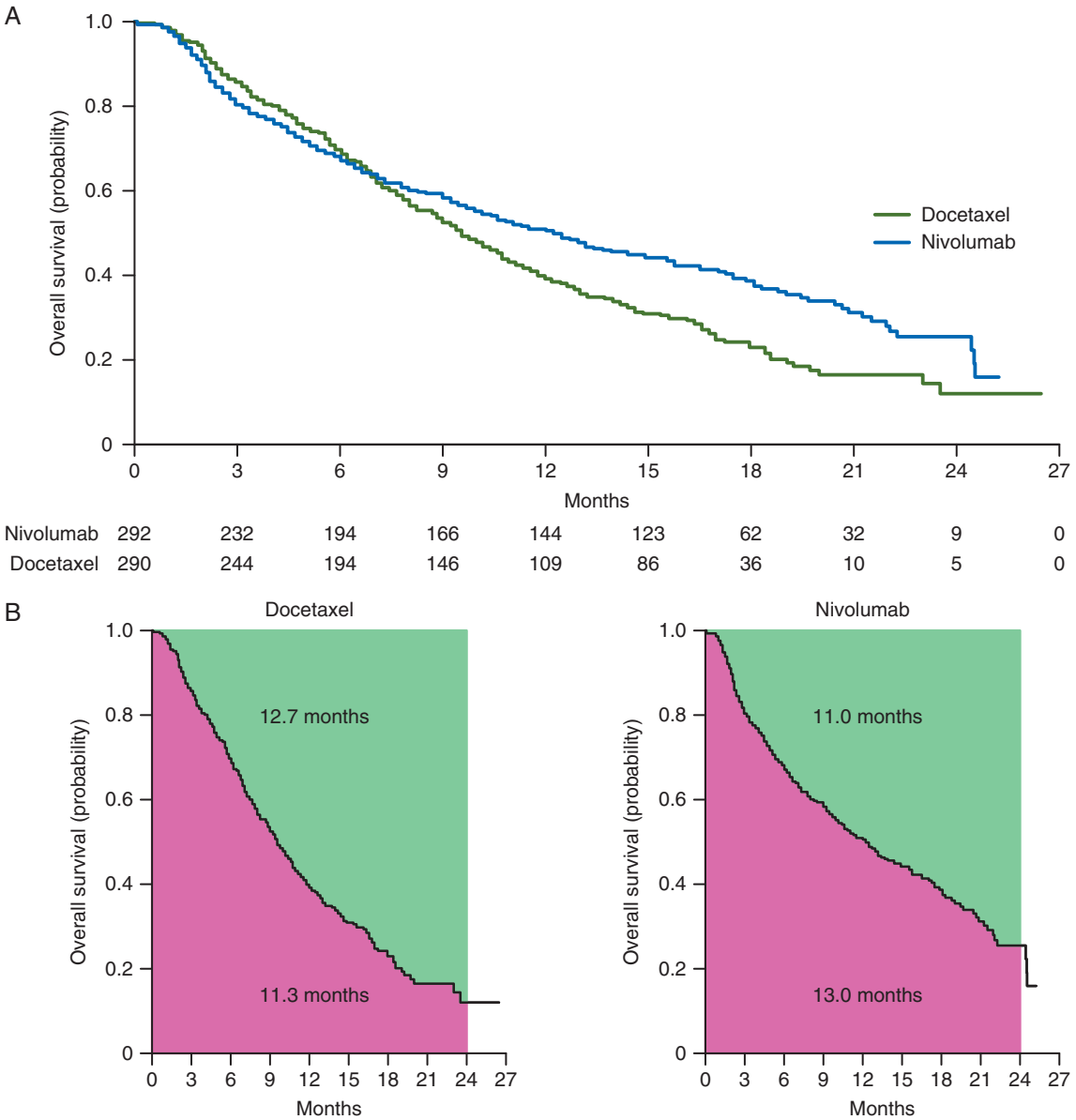


**Figure 1.** Estimated survival curves, restricted mean survival times (RMST) and restricted mean time lost (RMTL) based on reconstructed overall survival data for lung cancer study. (A) Kaplan–Meier curves for nivolumab (blue) and docetaxel (green). (B) RMST through 24 months (the area under the Kaplan–Meier curve) and RMTL through 24 months (the area above the Kaplan–Meier curve) for docetaxel (left) and nivolumab (right).

a clinically meaningful treatment effect. The treatment decision process in practice should not be based on a single contrast such as HR without a benchmark value from the control arm. These issues and concerns have been discussed extensively [2–6]. One may argue that the HR or the log-rank test is a valid procedure to reject a null hypothesis of no treatment effect even without the PH assumption. However, it is known that such a test may lack the power to detect a treatment effect when PH assumption is not valid. Coupled with HR, median survival time or the survival rate at a specific time point is often used to summarize the 'local aspect' of the survival profile for each group. The median survival time may not capture the long-term survival profile or may not be estimable due to limited follow-up time in the study. Good alternatives to HR are highly desirable.

Fei et al. [7] considered an alternative to HR to quantify the group difference based on the ratio of two restricted mean survival times (RMST) or restricted mean time lost (RMTL) [2–5, 8, 9]. As an illustration of the RMST, in Figure 1A, we present the Kaplan–Meier curves for the immunotherapy group and control group based on the reconstructed individual patient overall survival data from Borghaei et al. [10, 11], which is one of the studies utilized by Fei et al. [7]. Figure 1B depicts the RMST (or t-year mean survival time) and RMTL for each arm. The RMST, the area under the Kaplan–Meier curve, through 24 months is 13.0 for nivolumab versus 11.3 months for docetaxel. That is, on average, patients treated by nivolumab would survive 13.0 months out of 24-month follow-up. The corresponding RMTLs, the area above Kaplan–Meier curve, are 11.0 and 12.7 months, respectively. The ratio of RMSTs (nivolumab versus docetaxel) is 1.15 [95% confidence interval (CI) 1.03–1.29], and corresponding ratio of RMTLs is 0.87 (95% CI 0.77–0.97). The validity of these estimates and CI's requires no model assumptions. Moreover, there are absolute values from the control group with which to better interpret these ratios clinically.

Fei et al. [7] compared the ratio of RMSTs or RMTLs with the HR. However, these two ratios are not comparable summaries for the treatment effect since they estimate different population quantities. When the survival rates are low, the ratio of RMTLs is often numerically similar to HR since the survival time for each group can be approximated by an exponential distribution. On the other hand, the relative merit of these two ratios may be assessed via the statistical power for detecting the positive treatment effect. Empirically, the authors found that these two ratios as test statistics tend to have coherent results with respect to the statistical significance using type-I error rate of 0.05. This, coupled with other recent publications [12], ease concerns that the RMST-based tests might not be as powerful as the HR-based test when the PH assumption is plausible.

Regarding the time-window from zero to a time point $t$ to define the RMST or RMTL, Fei et al. [7] commented that this choice of '$t$' should be based on a clinical consideration at the design stage. As suggested by the authors, after the data were collected, various time-windows may be chosen empirically. For example, one may identify the last observed or censored survival time as the upper bound of the time-window for each group, then choose the minimum of these two values as time '$t$' for the RMSTs or RMTLs. For example, in the above example, we may choose 25.5 months instead of 24 months. The resulting ratios are identical to those with $t$ being 24 months. Note that the HR estimate may utilize less data than RMSTs. In the above example, the information that contributes to the HR estimation ends at $t = 23.5$ months, which is the minimum of the last observed event time in each of the two groups (Figure 1A). This phenomenon is not widely known in practice. Moreover, since the HR estimation procedure is event driven, the above '$t$' cannot be completely determined at the design stage.

It is important to make statistics more translational so that clinicians and patients can use them for decision making under the risk-cost-benefit consideration. The HR is not a readily translatable summary measure of the between-group difference. Alternative approaches that provide a robust and interpretable quantitative summary, such as difference or ratio of RMSTs or RMTLs, may be considered. We thank Fei et al. [7] for providing us with useful information regarding the relative merits between procedures using HR and RMST, and the editors for inviting comments on RMST as a new analytical procedure in the tool box of survival analysis.

H. Uno[1†], B. Claggett[2†], L. Tian[3], H. Fu[4], B. Huang[5], D. H. Kim[6,7‡] & L. J. Wei[8*‡]

[1]Division of Population Sciences, Department of Medical Oncology, Dana Farber Cancer Institute, Boston; [2]Division of Cardiovascular Medicine, Brigham and Women's Hospital, Boston; [3]Department of Health Research and Policy, Stanford University, Stanford; [4]Eli Lilly and Company, Indianapolis; [5]Pfizer Inc, Groton; [6]Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital, Boston; [7]Division of Gerontology, Department of Medicine, Beth Israel Deaconess Medical Center, Boston; [8]Department of Biostatistics, Harvard University, Boston, USA

(*E-mail: wei@hsph.harvard.edu)

[†]Both authors contributed equally as first authors

[‡]Both authors contributed equally as last authors.

## Funding

The work was partially supported by NIH/NHLBI (R01 HL089778), NIH/AHRQ (R00 HS022193), NIH/NIA (R21 AG049385) and NIH/NIA (K08 AG0511587).

## Disclosure

HF in an employee of Eli Lilly. BH is an employee of Pfizer. All remaining authors have declared no conflicts of interest.

## References

1. Cox DR. Regression models and life-tables. J R Stat Soc Series B Stat Methodol 1972; 34(2): 187–220.
2. Uno H, Claggett B, Tian L et al. Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. J Clin Oncol 2014; 32(22): 2380–2385.
3. Uno H, Wittes J, Fu H et al. Alternatives to hazard ratios for comparing the efficacy or safety of therapies in noninferiority studies. Ann Intern Med 2015; 163(2): 127–112.
4. A'Hern RP. Restricted mean survival time: an obligatory end point for time-to-event analysis in cancer trials? J Clin Oncol 2016; 34(28): 3474–3476.
5. Chappell R, Zhu X. Describing differences in survival curves. JAMA Oncol 2016; 2(7): 906–907.

6. Péron J, Roy P, Ozenne B et al. The net chance of a longer survival as a patient-oriented measure of treatment benefit in randomized clinical trials. JAMA Oncol 2016; 2(7): 901–905.
7. Fei L, Sheng Z, Qing W, Wenfeng L. Treatment effects measured by restricted mean survival time in trials of immune checkpoint inhibitors for cancer. Ann Oncol 2018; 29: doi.org/10.1093/annonc/mdy075.
8. 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.
9. Royston P, Parmar MK. Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. BMC Med Res Methodol 2013; 13(1): 152.
10. Borghaei H, Paz-Ares L, Horn DR et al. Nivolumab versus docetaxel in advanced nonsquamous non–small-cell lung cancer. N Engl J Med 2015; 373(17): 1627–1639.
11. Guyot P, Ades AE, Ouwens MJ, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. BMC Med Res Methodol 2012; 12(1): 9.
12. Tian L, Fu H, Ruberg SJ et al. Efficiency of two sample tests via the restricted mean survival time for analyzing event time observations. Biometrics 2017 September 12 [Epub ahead of print], doi: 10.1111/biom.12770.

doi:10.1093/annonc/mdy109
Published online 3 April 2018

## Circulating tumor DNA detection in hepatocellular carcinoma

Circulating tumor DNA (ctDNA) analysis has demonstrated excellent specificity and satisfactory sensitivity to detect somatic mutations across many tumor types and particularly in patients with metastatic disease [1, 2]. ctDNA detection rate and level primarily depend on tumor burden, proliferation and tumor type [3]. Its clinical use as a theragnostic biomarker for detecting *EGFR* mutation has been approved in stage IV non-small-cell lung cancer patients, in whom invasive tumor biopsy may be perilous because of limited accessibility and/or pre-existing organ dysfunction (e.g. emphysema). In that regard, biopsies of hepatocellular carcinoma (HCC) face the same challenges, with poor accessibility of some liver lesions and pre-existing cirrhosis and/or coagulopathy due to liver failure. In turn, HCC belongs to the few tumor types in which diagnosis can be established without an invasive biopsy, using combined radiological and biological (alpha-fetoprotein) criteria. Patients diagnosed with HCC have two main therapeutic options: locoregional treatments (i.e. surgery, embolization, radiofrequency or transplantation), prioritized in patients with non-metastatic HCC amenable to such treatments, and systemic treatments (e.g. sorafenib and regorafenib, two inhibitors of protein kinases with antiproliferative and antiangiogenic properties) proposed to patients with metastatic disease and/or unfit for locoregional treatments. Recent sequencing studies revealed potential therapeutic targets in 11 pathways in $\geq$5% of HCCs, mutations being found in *TERT* promoter (60%), WNT/β-catenin pathway (54%), PI3K-AKT-mTOR pathway (51%), TP53/cell cycle (49%) and mitogen-activated protein kinase pathway (MAPK, 43%) [4]. The presence of these somatic mutations cannot, however, be assessed if tumor tissue is not available from either biopsy or resection (which is often the case in late stage HCC patients)—opening a large window of opportunity for ctDNA analysis.

Prior studies have highlighted the increased level of cell-free circulating DNA (cfcDNA) in HCC patients [5, 6]. Jiang et al. also studied the size profiles of plasma DNA fragments in HCC patients: they showed that ctDNA fragments are shorter (peak at 166 bp) than normal cfcDNA, suggesting that most of ctDNA is derived from cancer cell apoptosis [7]. Three reports have been published on ctDNA detection in HCC patients with localized disease amenable to surgical resection. First, Liao et al. carried out ctDNA analysis by targeted next-generation sequencing (t-NGS) focusing on the three most frequently genomically altered genes (*TERT*, *CTNNB1* and *TP53*), and detected somatic mutations in only 20% ($n = 8/41$) of plasma samples (Table 1). Second, using droplet digital PCR (ddPCR) targeting four loci of the same three genes, Huang and colleagues detected at least one somatic mutation in cfcDNA in 56% ($n = 27/48$) of HCC patients before surgery. Third, Ono et al. used customized assays targeting tumor somatic rearrangements (previously characterized by whole genome sequencing of a tumor sample): 15% of patients ($n = 7/46$) had detectable ctDNA levels before surgery. It is important to note that these three studies focused on operable patients—i.e. patients with limited tumor burden—and targeted a limited number of genes.

In this issue of *Annals of Oncology*, Ng et al. have extended the patient population beyond operable HCC patients and the number of targeted genes, in order to answer whether ctDNA could be used as a liquid biopsy tool in HCC patients, by comparing liquid and tissue biopsies in 30 patients [8]. For this purpose, they designed a t-NGS panel targeting the most recurrent HCC mutations across 33 protein-coding genes, 2 long non-coding RNA genes, 4 promoters and 7 additional cancer gene mutation hotspots.

A clinically relevant analysis was then carried out that distinguished somatic mutations detected in the blood with high allelic frequency (mutations detected '*de novo*' without any knowledge on tumor tissue status) from that not called by algorithms (because of a lower allelic frequency) but observed manually in the plasma DNA following tumor tissue sequencing results. Using the '*de novo*' approach, Ng et al. detected at least one mutation in 27% ($n = 8/30$) of patients. While this rate might look disappointing at first sight, the authors clearly established that ctDNA detection was strongly correlated with the largest tumor diameter. Among cases with lesions >5 cm or with distant metastases, i.e. in patients primarily ineligible for surgical resection, the ctDNA detection rate rose to 86% ($n = 6/7$). Using the less stringent approach, based on a prior sequencing of the tumor tissue, the ctDNA detection sensitivity increased to 63% in the whole cohort ($n = 19/30$). This suggests that mutated DNA fragments are indeed detectable in the blood of most patients, even at an early stage, although mutations are hardly distinguishable by current algorithms from the background noise of sequencing. This highlights the potential interest of more sensitive NGS techniques, which may combine unique molecular identifiers and correction for sequencing error rates [9, 10]. Results obtained by Ng et al. have been confirmed by three other studies [11–13] (Table 1);