8-1-2018

Statistical Tests Used to Validate the American Joint Committee on Cancer Eighth Edition Prognostic Stage Compared With the Anatomic Stage in Breast Cancer.

Yilong Zhang

Xiaoxia Han
Henry Ford Health System, XHAN2@hfhs.org

Recommended Citation

Zhang Y, Han X. Statistical Tests Used to Validate the American Joint Committee on Cancer Eighth Edition Prognostic Stage Compared With the Anatomic Stage in Breast Cancer. JAMA Oncol. 2018 Aug 1;4(8):1137.

Role of the Funder/Sponsor: The National Cancer Institute had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

1. Sinnamon AJ, Neuwirth MG, Gimotty PA, et al. Association of first-in-class immune checkpoint inhibition and targeted therapy with survival in patients with stage IV melanoma. *JAMA Oncol*. 2018;4(1):126-128.

2. Cronin A, Tian L, Uno H. strmst2 and strmst2pw: new commands to compare survival curves using the restricted mean survival time. *Stata J*. 2016;16(3):702-716.

3. Hodi FS, O'Day SJ, McDermott DF, et al. Improved survival with ipilimumab in patients with metastatic melanoma. *N Engl J Med*. 2010;363(8):711-723.

4. Robert C, Thomas L, Bondarenko I, et al. Ipilimumab plus dacarbazine for previously untreated metastatic melanoma. *N Engl J Med*. 2011;364(26):2517-2526.

5. Chapman PB, Hauschild A, Robert C, et al; BRIM-3 Study Group. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *N Engl J Med*. 2011;364(26):2507-2516.

6. Maio M, Grob JJ, Aamdal S, et al. Five-year survival rates for treatment-naive patients with advanced melanoma who received ipilimumab plus dacarbazine in a phase III trial. *J Clin Oncol*. 2015;33(10):1191-1196.

Statistical Tests Used to Validate the American Joint Committee on Cancer Eighth Edition Prognostic Stage Compared With the Anatomic Stage in Breast Cancer

To the Editor Weiss et al[1] validated the American Joint Committee on Cancer Eighth Edition prognostic stage and compared it with the anatomic stage in breast cancer in 2 large cohorts. The authors used the Harrell C index to qualify the models' predictive performance based on prognostic stage and anatomic stage, respectively. The authors further determined the significance between the Harrell C index of the prognostic stage and anatomic stage using the R package compareC. In the MD Anderson cohort, the Harrell C indices for the prognostic stage and the anatomic stage are 0.8357 and 0.7370 ($P < .001$). In the California Cancer Registry, the Harrell C indices for the prognostic stage and the anatomic stage are 0.8426 and 0.8097 ($P < .001$).

With censored data, it is well known that the Harrell C index can overestimate the C index. Weiss et al[1] did not report the proportion of censored data for the 2 cohorts. Based on the Kaplan-Meier curves in the article, the 2 cohorts have approximately 75% subjects for whom no event was observed and who were censored at the end of the study, especially those with stage IA to IIB disease. Furthermore, to provide a valid inference, the method implemented in the R package compareC requires a strong condition that might not hold in practice.[2] If the condition does not hold, the compareC method can induce a serious bias and inflated type I error.[2] An alternative way is to use the inverse probability of censoring weighting estimator proposed by Uno et al[3] (R package SurvC1), but the bias may be nonnegligible if the censored proportion is high.[2] Another way is to assume a Cox proportional hazards (PH) model or proportional odds model and then apply the method proposed by Gonen and Heller[4] (R package CPE) or Zhang and Shao[5] (R package evacure) to estimate and compare the concordance indices.

The authors also report the Akaike information criterion (AIC) to compare model fits. For univariate analysis, the Harrell C index and the inverse probability of censoring weighted C statistics can be estimated directly without assuming a model. It is not clear why a model is required to estimate the C index and further compare the model by using the AIC. The Gonen and Heller estimator requires a Cox PH model, yet the goodness-of-fit test of the Cox PH model is more important than the AIC because the violation of the PH assumption can lead to a biased estimator.

Yilong Zhang, PhD
Xiaoxia Han, PhD

Author Affiliations: Merck Research Laboratories, Rahway, New Jersey (Zhang); Henry Ford Health System, Detroit, Michigan (Han).

Corresponding Author: Yilong Zhang, PhD, Merck Research Laboratories, 90 E Scott Ave, Rahway, NJ 07065 (yilong.zhang@merck.com).

Published Online: June 7, 2018. doi:10.1001/jamaoncol.2018.0884

Conflict of Interest Disclosures: None reported.

Editorial Note: This letter was shown to the corresponding author of the original article, who declined to reply on behalf of the authors.

1. Weiss A, Chavez-MacGregor M, Lichtensztajn DY, et al. Validation study of the American Joint Committee on Cancer Eighth Edition prognostic stage compared with the anatomic stage in breast cancer. *JAMA Oncol*. 2018;4(2):203-209.

2. Han X, Zhang Y, Shao Y. On comparing 2 correlated C indices with censored survival data. *Stat Med*. 2017;36(25):4041-4049.

3. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat Med*. 2011;30(10):1105-1117.

4. Gönen M, Heller G. Concordance probability and discriminatory power in proportional hazards regression. *Biometrika*. 2005;92(4):965-970.

5. Zhang Y, Shao Y. Concordance measure and discriminatory accuracy in transformation cure models. *Biostatistics*. 2018;19(1):14-26.

Estimating and Interpreting the Overall Survival Benefit of Checkpoint Inhibitors via Meta-analysis

To the Editor Lee et al[1] conducted an interesting meta-analysis to estimate the relative efficacy of checkpoint inhibitor vs docetaxel for treatment of advanced non–small cell lung carcinoma. The meta-analysis consists of 5 comparative clinical trials (CheckMate-017, CheckMate-057, Keynote-010, OAK, POPLAR) with the overall survival (OS) end point. For each study, the hazard ratio (HR) was used to quantify the treatment effect. A weighted average of 5 HRs was constructed as the pooled treatment effect from checkpoint inhibitors using the fixed-effects inverse-variance-weighted method. This resulted in a combined HR (checkpoint inhibitor vs docetaxel) of 0.69 (95% CI, 0.63-0.75).

There are a couple of issues regarding this meta-analysis. First, except for CheckMate-017, checkpoint inhibitors had delayed clinical OS benefit. That is, Kaplan-Meier curves for 2 treatment groups in each trial overlapped considerably for the early part of the study. Thus, the HR was not a constant over the entire study follow-up time. For this situation, it would be difficult to interpret individual HRs clinically and the HR estimate would not be an appropriate measure to quantify the OS benefit from checkpoint inhibitor use.[2,3] Second, even when the HR was constant over time for each study, one would not be able to identify a meaningful patient population for which the aforementioned pooled estimate of 0.69 could be interpreted as its HR unless those 5 underlying HRs are identical (an unlikely scenario).[4]

For a single study, a robust alternative summary for the between-treatment difference in OS could be the difference of 2 survival rates or restricted mean survival times (RMST) at a specific time point.[2,3] For example, for CheckMate-057, RMSTs for OS with 24-month follow-up were 13.0 and 11.3 months for