

Statistical Theory and Related Fields

ISSN: 2475-4269 (Print) 2475-4277 (Online) Journal homepage: www.tandfonline.com/journals/tstf20

Assessing association between biomarkers and clinical outcomes–moving beyond *p*-values

Ryan Sun , Daniel Paydarfar , Lu Tian , Brian Claggett , Dae Hyun Kim , Scott D. Solomon , Mitchell A. Psotka & Lee-Jen Wei

To cite this article: Ryan Sun , Daniel Paydarfar , Lu Tian , Brian Claggett , Dae Hyun Kim , Scott D. Solomon , Mitchell A. Psotka & Lee-Jen Wei (17 Nov 2026): Assessing association between biomarkers and clinical outcomes–moving beyond *p*-values, Statistical Theory and Related Fields, DOI: 10.1080/24754269.2026.2763239

To link to this article: https://doi.org/10.1080/24754269.2026.2763239

© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

Published online: 17 Nov 2026.

Submit your article to this journal

Article views: 40

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=tstf20

STATISTICAL THEORY AND RELATED FIELDS
https://doi.org/10.1080/24754269.2026.2763239

Taylor & Francis
by informa

OPEN ACCESS Check for updates

Assessing association between biomarkers and clinical outcomes–moving beyond *p*-values

Ryan Sun[a,†], Daniel Paydarfar[b,†], Lu Tian[c], Brian Claggett[d], Dae Hyun Kim[e], Scott D. Solomon[d], Mitchell A. Psotka[f] and Lee-Jen Wei[b]

[a]Department of Biostatistics, University of Texas MD Anderson Cancer Center, Houston, TX, USA; [b]Department of Biostatistics, Harvard University, Boston, MA, USA; [c]Department of Biomedical Data Science, Stanford University, Stanford, CA, USA; [d]Division of Cardiovascular Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; [e]Hinda and Arthur Marcus Institute for Aging Research, Hebrew SeniorLife, Harvard Medical School, Boston, MA, USA; [f]Inova Schar Heart and Vascular, Falls Church, VA, USA

ABSTRACT

Understanding the relationship between biomarkers and clinical outcomes is essential in medical research. Although regression models are used to quantify these associations, translating statistical findings into clinical practice remains challenging. Standard practice often reduces assessment to a *p*-value threshold, labelling a biomarker an 'independent' predictor if it passes. This oversimplification obscures nuanced interactions: different modelling strategies can yield divergent interpretations, and results may misrepresent clinical relevance. Large studies may highlight statistically significant but clinically trivial biomarkers, while smaller studies may miss important predictors due to limited power. Moreover, regression coefficients – especially from nonlinear models – often fail to convey the magnitude of association in a clinically intuitive way. Using trial data, our paper examines these limitations and reframes association analysis as a step towards addressing clinical questions: When, and to what extent, does a biomarker improve outcome prediction? We propose that an incremental predictability measure provides a more clinically relevant and interpretable way to quantify biomarker–outcome associations. Our findings underscore the need for analytic approaches that move beyond statistical significance towards measures that reflect clinical value. By emphasizing predictive performance, we aim to improve the translation of statistical associations into actionable insights and support more accurate identification of biomarkers that matter for patient outcomes.

ARTICLE HISTORY
Received 30 April 2026
Revised 20 October 2026
Accepted 9 November 2026

KEYWORDS
Association; baseline covariates; *C*-statistic; Cox model; mean absolute difference; prediction; survival time

1. Introduction

Association assessments are routinely conducted using regression modelling to analyze complex relationships between variables. These analyses aim to explore how clinical outcomes vary in response to changes in factors such as a specific biomarker, either alone or in conjunction with other patient characteristics. Typically, researchers construct a working regression

CONTACT Lee-Jen Wei wei@hsph.harvard.edu Department of Biostatistics, Harvard University, 655 Huntington Avenue, Boston, MA 02115, USA

[†]Co-first author.

© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

model linking the clinical outcome, or its related population measure, to the biomarker of interest and/or other baseline variables. With standard regression analysis, the strength of the association is conventionally based on the resulting regression coefficient estimate and its p-value. If the nominal p-value falls below 0.05, the biomarker is often labelled an 'independent' predictor and is claimed to be associated with the outcome in the medical literature, frequently without much consideration of the coefficient's magnitude.

While regression methodologies are widely used to explore associations, challenges remain in translating these findings into clinical practice. The relationship between a biomarker and an outcome can be quite nuanced. In fact, it is often unclear how to quantify association with a single population measure, yet assessments frequently rely on overly simplified models that lack adequate biological and clinical context. Different modelling strategies can yield divergent clinical and statistical interpretations. Accurately quantifying the strength of an association therefore depends on selecting a model capable of capturing the complex interplay of variables. Moreover, translating associations into clinical utility based solely on p-values poses considerable challenges. In a large study, this approach may lead to the selection of numerous biomarkers that are mostly clinically unrelated to outcomes when using the conventional 0.05 threshold to claim an association. Conversely, in a moderately sized study, the identification of clinically relevant biomarkers may be hindered by insufficient statistical power.

Although the regression coefficient estimate from a fitted model is often used to quantify the strength of association, its interpretation is not straightforward, particularly with nonlinear working models like logistic or Cox regression in the presence of other factors. Furthermore, its interpretation is heavily influenced by the chosen model, a nuance well understood in quantitative science (Lin & Wei, 1989; White, 1981), but less so in the medical literature, as evidenced by recent publications in medical journals utilizing this sub-optimal approach to assess association (Bernhard et al., 2024; Cheng et al., 2023; He et al., 2023; Loi et al., 2023; Marfella et al., 2024; Neumann et al., 2024).

Traditionally, association analysis serves as an intermediate step towards broader clinical research objectives, such as deciphering causal disease progression pathways or predicting outcomes using biomarkers. To uncover truly clinically and statistically meaningful associations, it would be necessary to move beyond conventional statistical practices and consider more relevant clinical implications.

In this article, we approach the assessment of biomarker utility from a different perspective. Rather than solely focussing on a biomarker's association with survival, we investigate whether a specific biomarker or patient characteristic at baseline should be included in a model for predicting the future survival of patients in the presence of other influencing factors. To achieve this, we consider quantifying the clinically interpretable incremental value that incorporating such a biomarker adds to the model's predictive capability. Using data from a cardiovascular clinical trial, we illustrate the challenges associated with conventional methods and present a robust solution centered on predictive analytics for quantifying the association in a more clinically meaningful way.

2. Illustrative example for issues of conventional association analysis

For illustration, we utilize data from the Beta-Blocker Evaluation of Survival Trial (BEST) (Eichhorn et al., 2001). This clinical study aimed to determine whether the beta-blocker bucindolol conferred benefits to patients with advanced chronic heart failure. A total of 2708

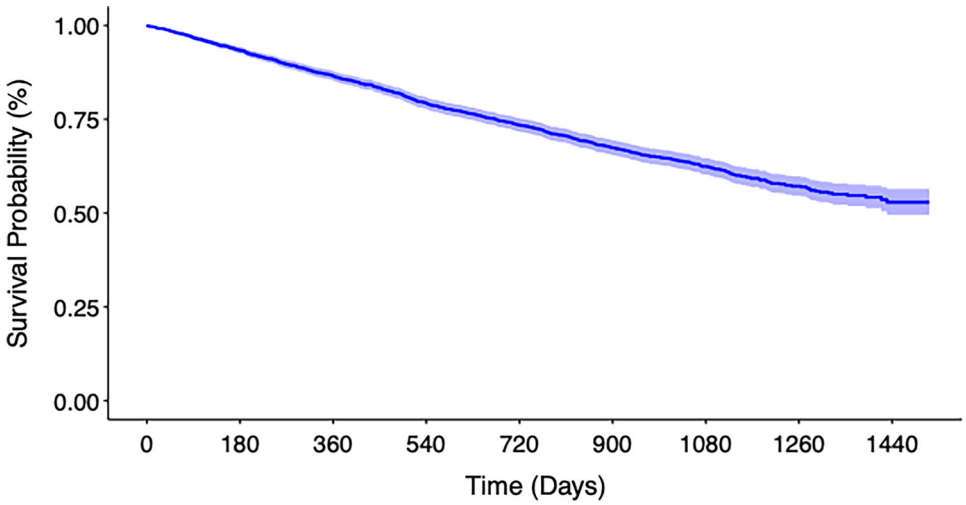


Figure 1. Kaplan-Meier curve for overall death of all subjects from BEST trial.

patients were enrolled, with overall survival time as the primary endpoint. Figure 1 shows the Kaplan-Meier curve for overall survival across both treatment groups, indicating an approximate 4-year survival rate of 53%. Due to censoring, data beyond this 4-year (1460-day) follow-up period were limited. Our analysis in this paper focuses on observations collected within this time frame.

From the 15 patient baseline factors in the US NIH-sponsored study database, we selected nine covariates generally considered relevant to survival outcome for illustrative purposes (Column 1 in Table 1(A)). We then fitted a standard Cox linear additive regression model to the survival data using these nine covariates in their original, unstandardized units (Cox, 1972). Table 1(A) (Columns 2–3) presents the resulting regression coefficient estimates and their corresponding statistics.

Consider, for instance, the effect of blood pressure, which achieved high statistical significance with a p-value less than 0.0001. In conventional analysis, this would typically lead to blood pressure being identified as an independent predictor of patient survival. Furthermore, the regression coefficient would be used to quantify the association's magnitude. For blood pressure, this coefficient is -0.01. One might then claim that for any fixed set of the other eight covariates, a one mmHg increase in blood pressure corresponds to a 0.01 decrease in the logarithm of the hazard. This interpretation is inherently challenging to grasp, given that hazard is not a probability metric like risk and lacks a benchmark value for easy comprehension and clinical interpretation. Moreover, the correlation between blood pressure and other covariates implies that fixing a patient's other eight covariate values would necessarily constrain blood pressure within a specific range, precluding unconstrained variation. Consequently, the traditional interpretation of the regression coefficient as a direct measure of the degree of association becomes tenuous. This issue is exacerbated when the model incorporates nonlinear terms, such as a quadratic component for blood pressure, rendering it even more difficult, if not impossible, to encapsulate the level of association via multiple regression coefficients from the linear and quadratic terms.

Table 1. Assessing strength of association between patients' baseline characteristics and survival.

(A) The individual covariate effects and the impact on predictability for survival time					
			Deleting a single covariate**		
Baseline characteristic/covariate*	Coefficient (95% CI)	*p*-value	*C*-statistic	MAD*** (days)	ΔMAD****
Rx	−0.14 (−0.27, −0.003)	0.04	0.68	380	2
Age	−0.003 (−0.010, 0.0052)	0.51	0.68	378	0
LVEF	−0.03 (−0.040, −0.021)	1.00×10^{-9}	0.67	381	3
eGFR	−0.02 (−0.021, −0.014)	3.13×10^{-19}	0.65	397	19
BPS	−0.01 (−0.015, −0.0069)	2.37×10^{-7}	0.67	379	1
Class	0.41 (0.21, 0.62)	6.39×10^{-5}	0.67	382	4
Diabetes	0.26 (0.12, 0.40)	2.33×10^{-4}	0.68	381	4
ISCH	0.33 (0.18, 0.49)	1.80×10^{-5}	0.67	377	-1
Afib	0.19 (−0.010, 0.39)	0.06	0.68	379	1
(B) No model (Kaplan-Meier)					
C-statistic = 0.50; MAD = 429 (days)					
(C) Model with nine covariates					
C-statistic = 0.68; MAD = 378 (days)					

Note: * Rx: treatment (1 = bucindolol / 0 = placebo); Age: age (years); LVEF: left ventricular ejection fraction (%); eGFR: estimated glomerular filtration rate adjusted for body surface area (mL/min/1.73 m^2); BPS: blood pressure (mmHg); Class: New York Heart Association Class (1 = Class 4, 0 = Class 3); diabetes: (1 = yes, 0 = no); ISCH: previous ischemic heart failure (1 = yes, 0 = no); Afib: atrial fibrillation (1 = yes, 0 = no).
** Refitting the data with the Cox model with eight covariates after deleting this specific factor.
*** MAD: mean absolute difference.
**** The difference between MAD and 378 days, the MAD with nine covariates.

Notably, the treatment effect was statistically significant at the 0.05 level ($p = 0.04$), whereas age ($p = 0.51$) and Afib ($p = 0.06$) were not; the other six covariates had much smaller *p*-values (Table 1(A)) and would conventionally be claimed as independent predictors. As with blood pressure, it is unclear how to physically interpret the level of association through their respective regression coefficients.

3. Using model discriminability and predictability to quantify the strength of association

Conventionally, the modelling results from the nine covariates suggest that blood pressure is associated with survival. From a different angle, one might inquire about the incremental value of incorporating blood pressure into the model. To evaluate this, we can employ *C*-statistics to assess the model's discriminatory capacity (Uno et al., 2011). The *C*-statistic quantifies the concordance between observed survival times and the corresponding risk scores generated by Cox models. A patient's risk score is simply the sum of nine products of their covariate values and the corresponding regression coefficients from Table 1(A); a higher score indicates a shorter time to mortality. For instance, consider two patients with death times of 61 days and 893 days, and risk scores of 2.35 and −4.06, respectively. In this example, since $61 < 893$ and $-4.06 < 2.35$, the risk scores align with the observed event times,

reflecting concordance. By considering all pairs of patients in which the first patient experiences an earlier event than the second, the *C*-statistic, which ranges from 0.50 to 1, quantifies the proportion of concordant pairs among the 2708 patients in the BEST study. A higher *C*-statistic indicates a better-fitted model. Notably, Uno's *C*-statistic (Uno et al., 2011) is particularly suitable for models with censored survival time observations. For a model with no covariates, the *C*-statistic would be 0.50. Our model with nine covariates yields a *C*-statistic of 0.68 (Table 1(C)), representing a nontrivial improvement over 0.50 without using the baseline information.

To examine the incremental value of a specific factor, such as blood pressure, when added to the other eight covariates, we removed blood pressure from the full model and refit the Cox model using only the remaining eight. The *C*-statistic for the full nine-covariate model was 0.68. Interestingly, the model with eight covariates (excluding blood pressure) also yielded a *C*-statistic of 0.67. This demonstrates that the inclusion of blood pressure did not improve the model's discriminatory power beyond what the other eight covariates already provided. In essence, the eight-covariate model performs as well as the nine-covariate model in terms of discrimination. It is worth noting that a Cox model utilizing blood pressure as its sole covariate resulted in a *C*-statistic of 0.58.

Column 4 in Table 1(A) reports the *C*-statistic of the model obtained by deleting the corresponding covariate, with values ranging from 0.65 to 0.68. These values indicate that adding any single covariate to the model containing the other eight covariates yielded no significant incremental improvement in discrimination based on this metric.

While the *C*-statistic assesses a model's ability to differentiate between patients with earlier versus later events, it does not directly evaluate predictive accuracy. Rather than relying on the *C*-statistic, which is often marginally sensitive to differences in modelling (Wang et al., 2023), an alternative approach involves assessing models based on their predictive capabilities. One method for this is the mean absolute difference (MAD) between predicted and observed survival times, a measure that quantifies prediction precision (Wang et al., 2023). Specifically, for each patient, their covariates are entered into the fitted Cox model to obtain their predicted survival curve up to the 4-year follow-up period. The area under this curve represents the predicted survival time for that patient. For patients who died before four years, the observed restricted survival time is the death time; for patients known to survive beyond four years, it is 1460 days. For patients censored before four years, the restricted survival time is unknown, so these patients do not contribute a directly observed prediction error; instead, the observable errors are weighted by the inverse Kaplan-Meier estimate of the probability of remaining uncensored (Wang et al., 2023). The MAD is calculated as the resulting weighted average absolute difference, under the independent-censoring assumption.

In the context of the BEST study data, for a model without any covariates, we predict any patient's survival using the restricted mean survival time (RMST), which was 1083 days up to the 4-year period. This mean survival time is calculated as the area under the Kaplan-Meier curve in Figure 1 up to 4 years. For this case, the MAD is 429 days (Table 1(B)). Now, considering the full model with nine covariates, the MAD is 378 days (Table 1(C)). This represents an incremental improvement of 51 days (429 minus 378) in prediction precision compared to the case without using baseline covariates.

We further quantify the incremental predictive value of each individual covariate using the MAD. Specifically, for each covariate, we sequentially remove it from the full model, refit the Cox model with the remaining eight covariates, and calculate the resulting MAD. These MAD values are reported in Column 5 of Table 1(A). Column 6 then shows the difference between these calculated MADs and the MAD of the full nine-covariate model (378 days).

The removal of blood pressure from the model resulted in a negligible change of only 1 day in MAD. In contrast, excluding estimated glomerular filtration rate (eGFR) led to a substantial reduction in prediction precision of 19 days. Utilizing this incremental predictability, as measured by the MAD difference, offers a more clinically and statistically interpretable approach than relying solely on regression coefficient estimates. For example, given that the total gain in prediction precision from the full nine-covariate model over the no-model case is 51 days, eGFR's contribution of 19 days represents approximately 37% of this overall improvement (19/51).

The above ratio of MAD reduction from a base model can serve as a valuable metric for evaluating a new biomarker's level of association with survival time. However, universally accepted threshold values for this ratio, quantifying the degree of association, remain to be established and require expert clinical input. Practical considerations, such as the cost or invasiveness of obtaining a biomarker, might necessitate a substantial incremental predictability to warrant its inclusion in future prognostic models. For instance, since eGFR is routinely recorded in clinical practice, its demonstrated predictive gain provides strong justification for its inclusion in survival prediction models. For predictive purposes, an important principle is to consider parsimonious candidate models that can be broadly applied. Thus, models with fewer biomarkers or other patient characteristics are generally preferred for clinical utility.

Metrics such as the *C*-statistic and MAD offer more relevant information regarding the level of association between a patient's characteristic and their survival time, surpassing the utility of *p*-values and regression coefficient estimates. Indeed, the American Statistical Association issued a formal statement cautioning against the misuse and misinterpretation of *p*-values in scientific decision-making (Wasserstein & Lazar, 2016). However, while the *C*-statistic primarily evaluates discriminatory power, the MAD provides a more intuitive approach to quantifying a biomarker's incremental value and the overall strength of a model's predictions. Unlike MAD, the *C*-statistic is often not sensitive enough to detect meaningful differences in predictive accuracy between models. As confirmed by our illustration using eGFR, models with nontrivial differences in their MAD values can still yield very similar *C*-statistics (Wang et al., 2023), underscoring the need for metrics that reflect prediction quality more directly. The computer code for implementing this MAD analysis is available at https://github.com/wx202/L1RMSTCOX.git.

4. Discussion

The concept of quantifying association is generally not well-defined. Specifically, when designing a study, a foremost consideration is how to identify the population quantity of interest, a point recently advocated strongly by ICH (International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use, 2019). Unfortunately, except in the simplest cases, such as a single binary covariate (e.g., a two-sample setting), this quantity for association remains unclear. While one might qualitatively claim that

two variables are associated, grasping or conceptualizing the level of that association is difficult.

Ideally, to quantify association at the population level, the chosen measure should be independent of model assumptions (McCaw et al., 2021). If this measure is contingent upon the model (e.g., relying on a proportional hazards assumption) and the assumed model subsequently fails to adequately fit the data, interpreting the association estimate becomes quite challenging. On the other hand, defining a quantity for the level of association at the population level without modelling is difficult, if not impossible. Consequently, there are no direct ways to estimate such a conceptually vague parameter with data.

While the incremental value of a new biomarker (e.g., blood pressure) for prediction is still subject to the selection of the base model (e.g., the Cox model with the other eight covariates), the prediction approach provides a valid method to assess the relative merit of different working models. This stands in stark contrast to quantifying the degree of association using a regression coefficient, where validity is heavily dependent on the correctness of the working model.

Lastly, using MAD for variable selection in regression analysis appears promising. Traditionally, selection procedures are mainly statistically oriented, such as using AIC, BIC, or p-values (Hastie et al., 2009), often lacking clinical input. For example, one could employ a stepwise backward variable selection approach starting with the nine-covariate model. At each step, a covariate with a small ΔMAD, such as age, would be deleted, the model refit, and the resulting precision loss (in MAD) evaluated. This iterative process would stop when the MAD shows a nontrivial increase compared to that of the original full model. Using data from BEST, we found that a parsimonious model with four covariates (left ventricular ejection fraction, eGFR, New York Heart Association Class, and diabetes status) yielded an MAD of 380 days, which is only slightly higher than the 378 days for the full model. The corresponding C-statistic is 0.66. This represents an important consideration when building a parsimonious prediction model for practical future use.

Beyond its direct utility in model comparison and variable selection for censored event-free time observations, our proposal in this article exemplifies a broader shift towards prediction-based approaches for quantifying biomarker contributions. The growing interest in assessing the practical utility of biomarker selection in survival analysis is further underscored by methodologies like SHAP (SHapley Additive exPlanations) values (Lundberg & Lee, 2017), which quantify individual covariate contributions. While a detailed discussion of SHAP values for our specific cases falls outside the scope of this article's primary theme, their development reinforces the far-reaching applicability of prediction precision-based metrics like MAD.

Author contributions

CRediT: **Ryan Sun:** Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Writing – review & editing; **Daniel Paydarfar:** Formal analysis, Methodology, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing; **Lu Tian:** Conceptualization, Investigation, Methodology, Project administration, Supervision, Validation, Writing – review & editing; **Brian Claggett:** Conceptualization, Methodology, Project administration, Validation, Writing – review & editing; **Dae Hyun Kim:** Validation, Writing – review & editing; **Scott D. Solomon:** Validation, Writing – review & editing; **Mitchell A. Psotka:** Validation, Writing – review & editing; **Lee-Jen Wei:** Conceptualization, Data curation, Investigation, Methodology, Project administration, Supervision, Validation, Writing – original draft, Writing – review & editing

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Bernhard, B., Ge, Y., Antiochos, P., Heydari, B., Islam, S., Sanchez Santiuste, N., Steel, K. E., Bingham, S., Mikolich, J. R., Arai, A. E., Bandettini, W. P., A. R. Patel, Shanbhag, S. M., Farzaneh-Far, A., Heitner, J. F., Shenoy, C., Leung, S. W., Gonzalez, J. A., Raman, S. V., … Kwong, R. Y. (2024). Association of adverse clinical outcomes with peri-infarct ischemia detected by stress cardiac magnetic imaging. *Journal of the American College of Cardiology*, *84*(5), 417–429.

Cheng, E., Shi, Q., Shields, A. F., Nixon, A. B., Shergill, A. P., Ma, C., Guthrie, K. A., Couture, F., Kuebler, P., Kumar, P., Tan, B., Krishnamurthi, S. S., Ng, K., E. M. O'Reilly, Brown, J. C., Philip, P. A., Caan, B. J., Feliciano, E. M. C., & J. A. Meyerhardt (2023). Association of inflammatory biomarkers with survival among patients with stage III colon cancer. *JAMA Oncology*, *9*(3), 404–413.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, *34*(2), 187–202.

Eichhorn, E. J., Domanski, M. J., Krause-Steinrauf, H., Bristow, M. R., & Lavori, P. W. (Beta-Blocker Evaluation of Survival Trial Investigators, 2001). A trial of the beta-blocker bucindolol in patients with advanced chronic heart failure. *New England Journal of Medicine*, *344*(22), 1659–1667.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer.

He, J., Bundy, J. D., Geng, S., Tian, L., He, H., & Li, X. (2023). Social, behavioral, and metabolic risk factors and racial disparities in cardiovascular disease mortality in U.S. adults: An observational study. *Annals of Internal Medicine*, *176*(9), 1200–1208.

International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (2019). Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials E9(R1). https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf

Lin, D. Y., & Wei, L. J. (1989). The robust inference for the Cox proportional hazards model. *Journal of the American Statistical Association*, *84*(408), 1074–1078.

Loi, S., Salgado, R., Schmid, P., Cortes, J., Cescon, D. W., Winer, E. P., Toppmeyer, D. L., Rugo, H. S., De Laurentiis, M., Nanda, R., Iwata, H., Awada, A., Tan, A. R., Sun, Y., Karantza, V., Wang, A., Huang, L., Saadatpour, A., Cristescu, R., … Adams, S. (2023). Association between biomarkers and clinical outcomes of pembrolizumab monotherapy in patients with metastatic triple-negative breast cancer: KEYNOTE-086 exploratory analysis. *JCO Precision Oncology*, *7*, e2200317.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In U. von Luxburg, I. Guyon, S. Bengio, H. Wallach, & Rob Fergus (Eds.), *Advances in Neural Information Processing Systems* (Vol. 30, pp. 4765–4774). Curran Associates, Inc.

Marfella, R., Prattichizzo, F., Sardu, C., Fulgenzi, G., Graciotti, L., Spadoni, T., D'Onofrio, N., Scisciola, L., La Grotta, R., Frigé, C., Pellegrini, V., Municinò, M., Siniscalchi, M., Spinetti, F., Vigliotti, G., Vecchione, C., Carrizzo, A., Accarino, G., Squillante, A., … Paolisso, G. (2024). Microplastics and nanoplastics in atheromas and cardiovascular events. *New England Journal of Medicine*, *390*(10), 900–910.

McCaw, Z. R., Tian, L., Wei, J., Claggett, B. L., Bretz, F., Fitzmaurice, G., & Wei, L. J. (2021). Choosing clinically interpretable summary measures and robust analytic procedures for quantifying the treatment difference in comparative clinical studies. *Statistics in Medicine*, *40*(28), 6235–6242.

Neumann, J. T., Twerenbold, R., Weimann, J., Ballantyne, C. M., Benjamin, E. J., Costanzo, S., de Lemos, J. A., de Filippi, C. R., Castelnuovo, A. D., Donfrancesco, C., Dörr, M., Eggers, K. M., Engström, G., Felix, S. B., M. M. Ferrario, Gansevoort, R. T., Giampaoli, S., Giedraitis, V., Hedberg, P., … Ojeda, F. (2024). Prognostic value of cardiovascular biomarkers in the population. *JAMA*, *331*(22), 1898–1909.

Uno, H., Cai, T., Pencina, M. J., D'Agostino, R. B., & Wei, L. J. (2011). On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Statistics in Medicine*, *30*(10), 1105–1117.

Wang, X., Claggett, B. L., Tian, L., Malachias, M. V. B., Pfeffer, M. A., & Wei, L. J. (2023). Quantifying and interpreting the prediction accuracy of models for the time of a cardiovascular event–moving beyond C statistic: A review. *JAMA Cardiology*, *8*(3), 290–295.

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA statement on *p*-values: Context, process, and purpose. *The American Statistician*, *70*(2), 129–133.

White, H. (1981). Consequences and detection of misspecified nonlinear regression models. *Journal of the American Statistical Association*, *76*(374), 419–433.