Treatment of Cancer randomized phase III trial 10853—A study by the EORTC Breast Cancer Cooperative Group and EORTC Radiotherapy Group. J Clin Oncol 24:3381-3387, 2006

5. Goldstein LJ, Gray R, Badve S, et al: Prognostic utility of the 21-gene assay in hormone receptor-positive operable breast cancer compared with classical clinicopathologic features. J Clin Oncol 26:4063-4071, 2008

6. Rockhill B, Spiegelman D, Byrne C, et al: Validation of the Gail et al model of breast cancer risk prediction and implications for chemoprevention. J Natl Cancer Inst 93:358-366, 2001

DOI: 10.1200/JCO.2012.43.9067; published online ahead of print at www.jco.org on July 23, 2012

Reply to K.J. Van Zee et al

We thank Van Zee and Patil[1] for their correspondence concerning our recently published article.[2] We agree with their observation that dividing our population into eight groups created small sample sizes, which may have affected our evaluation of the Memorial Sloan-Kettering Cancer Center nomogram.[3] To address this, we replotted calibration curves using quartiles, as suggested by Van Zee and Patil. This resulted in more than 120 patients in each group. We still found imperfect calibration, particularly in patients with the highest predicted risk, for whom the nomogram overestimated the observed risk on the 5- and 10-year calibration plots (Fig 1). In addition, we found that the accuracy of the nomogram seemed to decrease over time, given that two groups on the 10-year plot did not approach the reference line (Fig 1B). In our data set, 206 of 794 patients (25.9%) had at least 10 years of follow-up, whereas in the original Memorial Sloan-Kettering cohort, 17% of the patients were observed for at least 10 years. We would agree with Van Zee and Patil that calibration is more important to clinicians and patients; therefore, the finding of poor calibration in the highest-risk quartile is concerning and may limit the applicability of the nomogram when counseling these patients.

The Harrell C-index is used to quantify concordance for nomograms. A C-index of 0.70 to 0.80 is considered acceptable, and a C-index of 0.80 to 0.90 is considered excellent.[4] In their initial publication,[3] the Memorial Sloan-Kettering group reported a C-index of 0.70. This nomogram has not been validated previously by an internal or external data set. In our effort to validate their nomogram, we found a C-index of 0.63, which cannot be considered optimal on the basis of the definitions mentioned here. The authors[1] point out several models that are often used in clinical practice despite C-index scores in the same range as that of their nomogram. We are not suggesting that the nomogram should not be used in practice; whether or not to use it should be decided by individual practitioners. The nomogram-predicted risk of ipsilateral breast tumor recurrence would be just one variable used in counseling patients.

We agree that large multicenter trials have shown that use of radiation therapy (RT) is an important predictor of recurrence.[5,6] In our center, the majority of patients are treated with radiation unless they have small, low-grade tumors with wide margins. Of the 734 patients who were included in our evaluation of the nomogram, only four (0.5%) had positive margins, and 86 (11.7%) had close (< 2 mm) margins. Of the 205 patients (27.9%) who were treated without radiation, only one had positive margins and 24 had close margins. The authors[1]

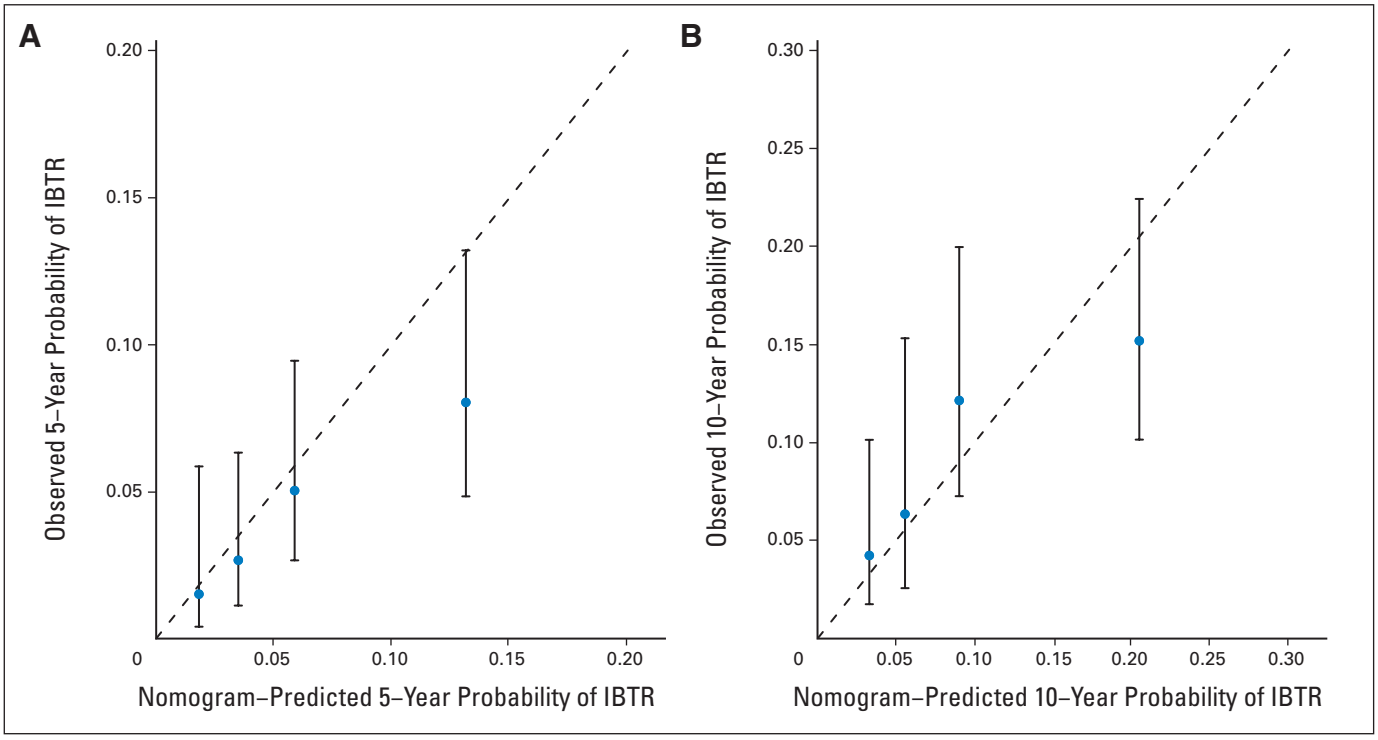

Fig 1. Calibration plots for (A) 5-year nomogram and (B) 10-year nomogram. Patients were grouped by quartiles of predicted risk. The *x*-axis is the nomogram-predicted probability of ipsilateral breast recurrence (IBTR). The *y*-axis is the observed probability of IBTR (Kaplan-Meier estimates). Broken line represents ideal nomogram; circles represent apparent predictive accuracy, which was calculated by plotting the mean Kaplan-Meier estimate for each quartile versus the mean nomogram-predicted probability for patients in each quartile.

ask whether there was something special about the patients with close or positive margins who did not receive RT. Patients who did not receive RT were older compared with patients who did receive RT (median age, 59 *v* 56 years; $P = .002$). Some of the patients in our cohort declined RT although it was recommended. Overall, more than 75% of the patients in our entire population received RT, and the recurrence rate in the entire population was low, which may explain why we did not find that RT was a predictor of ipsilateral breast tumor recurrence in our cohort.

Our definitions of time period of surgery and number of excisions differed from the definitions used by the Memorial Sloan-Kettering authors,[3] which we explained in the methods and results sections of our article.[2] Those variables with alternative definitions were only used in our Cox model. The 5- and 10-year predictive values were calculated through the Memorial Sloan-Kettering Web site and thus would not be affected by the definitions of variables. The nomogram was correctly applied for all of our patients.

Tumor size is an important variable for studying recurrence, and some studies have found an association between tumor size and recurrence of ductal carcinoma in situ.[7,8] Tumor size was not included in the Memorial Sloan-Kettering analysis. Perhaps tumor size was not always available in their cohort, which may be the reason why the number of excisions was used as a surrogate. There was no discussion in their article about the relationship between number of excisions and large tumor size.[3]

The effects of treatments and time period alone in the nomogram may not be sufficient to explain the outcomes. The nomogram will not completely account for differences in treatment trends over time as a result of the missing interaction term. The interaction term would provide different predictions for no RT before 1999 versus no RT after 1999 and RT before 1999 versus RT after 1999.

Min Yi, Funda Meric-Bernstam, Henry M. Kuerer, Elizabeth A. Mittendorf, Isabelle Bedrosian, Anthony Lucci, Rosa F. Hwang, and Jaime R. Crow
The University of Texas MD Anderson Cancer Center, Houston, TX

Sheng Luo
School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX

Kelly K. Hunt
The University of Texas MD Anderson Cancer Center, Houston, TX

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The author(s) indicated no potential conflicts of interest.

REFERENCES

1. Van Zee KJ, Patil S: Validation of a nomogram for predicting risk of local recurrence for ductal carcinoma in situ. J Clin Oncol 30:3143-3144, 2012

2. Yi M, Meric-Bernstam F, Kuerer HM, et al: Evaluation of a breast cancer nomogram for predicting risk of ipsilateral breast tumor recurrences in patients with ductal carcinoma in situ after local excision. J Clin Oncol 30:600-607, 2012

3. Rudloff U, Jacks LM, Goldberg JI, et al: Nomogram for predicting the risk of local recurrence after breast-conserving surgery for ductal carcinoma in situ. J Clin Oncol 28:3762-3769, 2010

4. Hosmer D, Lemeshow S: Applied Logistic Regression (ed 2). New York, NY, Wiley, 2000, pp 162

5. EORTC Breast Cancer Cooperative Group, EORTC Radiotherapy Group, Bijker N, et al: Breast-conserving treatment with or without radiotherapy in ductal carcinoma-in-situ: Ten-year results of European Organisation for Research and Treatment of Cancer randomized phase III trial 10853—A study by the EORTC Breast Cancer Cooperative Group and EORTC Radiotherapy Group. J Clin Oncol 24:3381-3387, 2006

6. Fisher B, Dignam J, Wolmark N, et al: Lumpectomy and radiation therapy for the treatment of intraductal breast cancer: Findings from National Surgical Adjuvant Breast and Bowel Project B-17. J Clin Oncol 16:441-452, 1998

7. Silverstein MJ, Lagios MD: Choosing treatment for patients with ductal carcinoma in situ: Fine tuning the University of Southern California/Van Nuys Prognostic Index. J Natl Cancer Inst Monogr 2010:193-196, 2010

8. Silverstein MJ, Lagios MD, Craig PH, et al: A prognostic index for ductal carcinoma in situ of the breast. Cancer 77:2267-2274, 1996

DOI: 10.1200/JCO.2012.43.9406; published online ahead of print at www.jco.org on July 23, 2012

Design of Testis Cancer Trials: Who Knows Best?

To the Editor: In their editorial, Nichols et al[1] criticize our trial in intermediate prognosis germ cell cancer (GCC) for reasons of futility.[2] We disagree with this conclusion and would like to explain the reasons.

In the International Germ Cell Consensus Conference Classification,[3] metastatic GCC is divided into three groups, including an intermediate-risk group (25% of patients; 75% chance of cure with the standard regimen) and a poor-risk group (15% of patients; 50% chance of cure). In 1995, all genitourinary cancer collaborative groups in Europe agreed on two separate strategies that were designed to improve the management of GCC: one was to investigate high-dose chemotherapy in poor-risk disease, and the second was the addition of paclitaxel (T) to standard-dose bleomycin, etoposide, and cisplatin (BEP) in intermediate-risk disease.[4] The phase II/III trial objective to improve cure from 75% to 85% (10%) was founded on a dose-finding study of T-BEP.[5] The T-BEP versus BEP protocol was supported by national scientific boards as well as investigators from sites in the United States who participated in the independent data monitoring committee. Futility stopping rules at the end of the phase II trial were not met, and the independent data monitoring committee repeatedly concluded that there was scientific value in continuing the study. If the study had accrued as originally predicted, it would have been completed in 5 years; thus, it seemed feasible to conduct the study in parallel with the previously mentioned study,[6] which was enrolling patients with poor-prognosis disease and investigating the concept of dose-intensified chemotherapy. Unfortunately, accrual in both trials was hampered by difficulties with respect to trial conduct in Europe, with increasing administrative hurdles in clinical research that were accompanied by a lack of financial resources. Eventually, both trials had to be stopped prematurely as a result of slow accrual.

Despite the loss of power that resulted from the curtailed accrual (data on only 85 of 98 events were available), the analyses that were free of potential bias because of the uneven distribution