






ARTICLE

Simulation Modeling to Extend Clinical Trials of Adjuvant Chemotherapy Guided by a 21-Gene Expression Assay in Early Breast Cancer

Jinani Jayasekera , Joseph A. Sparano, Robert Gray , Claudine Isaacs , Allison Kurian , Suzanne O'Neill, Clyde B. Schechter, Jeanne Mandelblatt 

See the Notes section for the full list of authors' affiliations.

Correspondence to: Jinani Jayasekera, PhD, Georgetown-Lombardi Comprehensive Cancer Center, 3300 Whitehaven St, NW, Ste 4100, Washington, DC 20007 (e-mail: vcj74@georgetown.edu).

Abstract

Purpose: The Trial Assigning Individualized Options for Treatment (TAILORx) found chemotherapy could be omitted in many women with hormone receptor–positive, *HER2*-negative, node-negative breast cancer and 21-gene recurrence scores (RS) 11–25, but left unanswered questions. We used simulation modeling to fill these gaps.

Methods: We simulated women eligible for TAILORx using joint distributions of patient and tumor characteristics and RS from TAILORx data; treatment effects by RS from other trials; and competing mortality from the Surveillance, Epidemiology, and End Results program database. The model simulations replicated TAILORx design, and then tested treatment effects on 9-year distant recurrence-free survival (DRFS) in 14 new scenarios: eight subgroups defined by age (≤ 50 and > 50 years) and 21-gene RS (11–25/16–25/16–20/21–25); six different RS cut points among women ages 18–75 years (16–25, 16–20, 21–25, 26–30, 26–100); and 20-year follow-up. Mean hazard ratios SD, and DRFS rates are reported from 1000 simulations.

Results: The simulation results closely replicated TAILORx findings, with 75% of simulated trials showing noninferiority for chemotherapy omission. There was a mean DRFS hazard ratio of 1.79 (0.94) for endocrine vs chemoendocrine therapy among women ages 50 years and younger with RS 16–25; the DRFS rates were 91.6% (0.04) for endocrine and 94.8% (0.01) for chemoendocrine therapy. When treatment was randomly assigned among women ages 18–75 years with RS 26–30, the mean DRFS hazard ratio for endocrine vs chemoendocrine therapy was 1.60 (0.83). The conclusions were unchanged at 20-year follow-up.

Conclusions: Our results confirmed a small benefit in chemotherapy among women aged 50 years and younger with RS 16–25. Simulation modeling is useful to extend clinical trials, indicate how uncertainty might affect results, and power decision tools to support broader practice discussions.

In the era of personalized medicine, clinical trials remain the gold standard of evidence to inform oncology practice and guidelines (1). Evidence is considered especially robust when several trials have consistent conclusions across different settings and eligible populations. However, it is becoming less feasible to conduct multiple trials, especially ones that are large enough to allow analyses for broad, clinically relevant subgroups, evaluation of different eligibility cut points, and inclusion of long-term follow-up. Consequently, it can be difficult for

oncologists to translate trial evidence in to treatment discussions with many of their individual patients.

The Trial Assigning Individualized Options for Treatment (TAILORx) was a large personalized trial that used 21-gene expression profile results to stratify treatment assignment. The results indicated that invasive disease-free survival rates and distant recurrence rates were equivalent with endocrine vs chemoendocrine therapy among women with hormone-receptor positive (HR)+/ human epidermal growth factor

Received: May 17, 2019; Revised: July 22, 2019; Accepted: August 7, 2019

© The Author(s) 2019. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com.

receptor 2 (HER2)-negative, node-negative breast cancer and 21-gene recurrence scores (RS) of 11–25 (2). These results have the potential to affect treatment for a large number of women because the majority (70%) of women diagnosed with breast cancer in the United States each year have these tumor types (3, 4).

However, there were several issues remaining after TAILORx that may make it difficult for clinicians to translate this evidence into practice. First, TAILORx data indicated a benefit from chemotherapy in women ages 50 years and younger who had tumors with RS 16–25 (2), but these were retrospective, unplanned analyses (5). Second, earlier validation studies used midrange RS cut points of 18–30, whereas TAILORx cut points for testing chemotherapy effects were RS 11–25 (6,7), which potentially created uncertainty about optimal cut points for determining chemotherapy benefits, especially for women with scores close to the demarcations. Third, chemotherapy was definitely rather than randomly assigned for women with RS of 26 or more. Fourth, TAILORx did not provide data for clinicians caring for women who are interested in having information about outcomes beyond 9 years before making treatment decisions. Finally, TAILORx was one trial, and the results have not yet been replicated.

In such situations, simulation modeling provides an excellent virtual laboratory to replicate trials, confirm planned and unplanned trial analyses, and extend results to examine new scenarios that were not possible to test in the original trials (1, 8,9). Simulation modeling permits evaluation of large numbers of virtual individuals with all possible combinations of characteristics that might be seen in practice, synthesis of multiple evidence sources, and sampling across the full range of effects to quantify uncertainty in estimations (10). Consistency between results of a single trial, such as TAILORx, and model simulation results could provide clinicians and their patients with greater confidence when making decisions based on results.

We used an established simulation model (1,11) to replicate and extend TAILORx results to different age and RS subgroups, RS cut points, and time horizons. The results are intended to demonstrate the potential contribution of simulation modeling to replicate and support translation of trial results in to routine oncology practice.

Methods

The study was approved by the Georgetown University Institutional Review Board and was considered as exempt research based on use of deidentified data.

Model Overview

We adapted an extant breast cancer model (Model GE) (1,11) developed in the National Cancer Institute–funded Cancer Intervention and Surveillance Modeling Network. The model randomly assigned women to endocrine vs chemoendocrine therapy and replicated TAILORx and tested treatment effects in 14 new scenarios: eight subgroups defined by age (≤ 50 and > 50 years) and 21-gene RS (11–25); six different RS cut points among women ages 18–75 (16–25, 16–20, 21–25, 26–30, 26–100); and 20-years of follow-up for women ages 18–75 years and RS 11–25, 16–25, 16–20, 21–25, 26–30, and 26 or more.

For each scenario, we first generated virtual samples of simulated women who would have been eligible for TAILORx (see below), each with a specific combination of age, tumor size, grade, hormonal status (estrogen receptor [ER] and/or

progesterone receptor [PR]), RS, and surgery type (mastectomy vs lumpectomy) from all possible joint combinations of these characteristics. Virtual women were then randomly assigned to endocrine vs chemoendocrine therapy and followed from random assignment to death or end of follow-up (9 or 20 years). Each virtual woman could remain event free until end of follow-up, experience a distant recurrence, die of breast cancer, or other causes (Figure 1), conditional on their treatment arm, age, RS, and tumor characteristics.

The primary endpoint for each simulated scenario was based on the secondary TAILORx endpoint, distant recurrence-free survival (DRFS) at 9 years, defined as time from recruitment to date of distant recurrence, or death with distant recurrence, if death was the first manifestation of distant recurrence. This corresponds to the Standardized Definitions for Efficacy End Points in Adjuvant Breast Cancer Trials definition of distant recurrence-free interval (12). We also simulated age-specific other-cause mortality for women ages 18–75 years.

Target Population and Sample Size

We included virtual women who were eligible for the TAILORx trial, including women ages 18 to 75 years with hormone receptor-positive, HER2-negative invasive, node-negative breast cancer with tumor size 1.1–5.0 cm (or 5 mm–1.0 cm and intermediate or poor histologic grade; “early breast cancer”), who had undergone lumpectomy (with radiotherapy) or mastectomy. The sample size of virtual women simulated was based on the TAILORx specifications for detection of relative differences in the effects of endocrine vs chemoendocrine therapy on DRFS, assuming a null hypothesis of no difference (2). The DRFS hazard ratio (HR) for the noninferiority margin was set at 1.61; noninferiority was based on whether the 95% confidence interval (CI) contained 1.0 (no difference), or the entire confidence interval less than 1.61 (2).

Model Inputs

The input parameters and sources are summarized in [Supplementary Table 1](#) (available online). We first simulated the characteristics of women (age, grade, tumor size, ER and PR status, RS, surgery type) eligible for each virtual scenario based on the joint distribution of characteristics derived from the individual-level deidentified TAILORx dataset provided by the Eastern Cooperative Oncology Group–American College of Radiology Imaging Network Cancer Research Group (personal communication, Robert Gray, 2019).

We then simulated events and time-to-events using data independent of TAILORx. Distant recurrence (and breast cancer death) events and time-to-events for endocrine vs chemoendocrine therapy conditional on patient and tumor attributes were simulated using competing risk-survival models fitted to an individual-level linked National Surgical Adjuvant Breast and Bowel Project (NSABP)–Genomic Health dataset provided by Genomic Health, Inc, Redwood City, CA (personal communication, Steve Shak, 2018) (7,13–15). The patient and tumor attributes included in the model were age, grade, tumor size, 21-gene RS, and interaction terms between age, RS, and tumor size. The subhazard ratios for treatment as well as patient and tumor attributes were calculated using Fine-Gray methods (16). Other-cause mortality was used as the competing risk. The time-to-event distributions were modeled semiparametrically by estimating proportional-hazards cumulative incidence functions (CIF). We allowed the CIF to be semiparametrically dependent

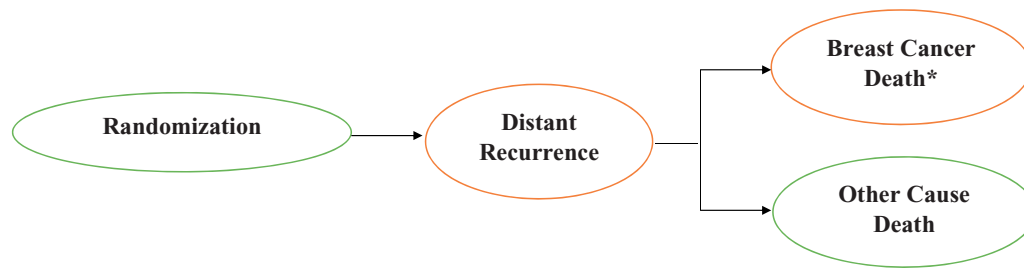


Figure 1. Progression of breast cancer following randomization as a state-transition process. *Breast cancer death without distant recurrence was considered the first manifestation of distant recurrence.

on various attributes of the patient and the tumor, including RS. That is, a baseline CIF was identified for a designated combination of reference values of the predictive patient and tumor attributes, and the CIF for other values of those attributes were derived by the application of subhazard ratios.

Because the NSABP trials were conducted several decades earlier, we used data from TAILORx and the Oxford Overview (17,18) to adjust treatment effects on the CIFs to reflect therapy in TAILORx. Specifically, we adjusted hormonal effects for tamoxifen to effects seen with tamoxifen and aromatase inhibitors (18) and chemotherapy effects from CMF regimens to effects seen with CMF, anthracycline, and taxane-based regimens (17). Finally, to estimate competing other-cause mortality, we used a left-truncated survival model estimated from the Surveillance, Epidemiology, and End Results–Genomic Health dataset for women matching trial eligibility (19,20).

Statistical Analysis

TAILORx and each of the 14 scenarios was modeled separately and replicated 1000 times to quantify the uncertainty related to sampling variability for any given parameter value (1). Each of the 1000 trial replicates was randomly assigned its own set of treatment effects sampling from the “prior” distribution of the subhazard ratios derived from the competing risk-survival models derived from the NSABP trials as described above. This empiric Bayesian analysis approach captures uncertainty in all predictors’ effects on outcomes and sampling variation.

The hazard ratio of endocrine vs chemoendocrine therapy and its two-sided 95% confidence interval were determined for each of the 1000 simulations using Cox proportional hazards regressions. Kaplan-Meier curves for DRFS by treatment were used to evaluate the proportional hazard assumption, where parallel curves indicated that the treatment variable satisfies the proportional hazard assumption. DRFS by treatment group for each of the 1000 simulations was also found from Kaplan-Meier curves.

The results of the 1000 simulations were summarized using the following metrics for women for receiving endocrine vs chemoendocrine treatment: 1) a percentage of trial replicates (of the 1000) that showed noninferiority in the omission of chemotherapy as defined in the TAILORx protocol; 2) the mean hazard ratio comparing endocrine vs chemoendocrine therapy for distant recurrence across the 1000 simulations, and the SD of the mean; 3) the mean and SD of the DRFS rate for each treatment group at years 9 and 20; and 4) caterpillar plots for every 10th trial from an ordered list of 1000 trial replicates to assess the distribution of hazard ratios and corresponding 95% confidence intervals. Caterpillar plots (21) were used to illustrate the range of results that might be observed if TAILORx were

repeated, given the uncertainty built into the modeling of input parameters and treatment effects.

All analyses were conducted using STATA version 15.0 (StataCorp. 2017. Stata Statistical Software: Release 15. College Station, TX: StataCorp LP).

Results

Within each of the 14 simulated scenarios, virtual women had a similar distribution of characteristics (age, grade, tumor size, ER and PR status, surgery) by treatment arm as was seen among women enrolled in TAILORx (Table 1; Supplementary Table 2, available online).

Replication of TAILORx (DRFS in Women Ages 18–75 Years and RS 11–25)

In the actual TAILORx trial, at 9 years the hazards of DRFS were noninferior (HR = 1.10; 95% CI = 0.85 to 1.41; $P = .48$), and had similar DRFS rates (94.5% vs 95.0%) among women receiving endocrine vs chemoendocrine therapy, respectively. In the simulated trials, at 9 years, the mean DRFS hazard ratio (SD) for endocrine therapy vs chemoendocrine therapy was 1.15 (0.42) (Table 2). The mean 9-year DRFS rates were 94.0% (0.02) vs 94.4% (0.01) (absolute difference = 0.4%) for endocrine vs chemoendocrine therapy. The distribution of hazard ratios from a sample of 100 trials (of 1000 trial simulations) is shown in Figure 2. Of the 1000 simulated trials, 75% found noninferiority in the omission of chemotherapy on DRFS.

DRFS in Women Ages 18–50 Years and RS 11–25

Among women ages 18–50 years with RS 11–25, the mean DRFS HR (SD) for endocrine vs chemoendocrine therapy was 1.55 (0.74) (Table 2). The mean 9-year DRFS rates across the 1000 simulations for endocrine vs chemoendocrine therapy were 94.4% (0.02) vs 95.9% (0.01) (absolute difference = 1.5%). Out of the 1000 trial simulations, 59% found noninferiority in the omission of chemotherapy on distant recurrence (ie, a statistically significant chemotherapy benefit was found in only approximately 40% of the trial replications) (Supplementary Figure 1, available online).

DRFS in Women Ages 51–75 Years and RS 11–25

The mean DRFS HR (SD) at 9 years for endocrine vs chemoendocrine therapy for older women was 1.17 (0.37) (Table 2), there was no difference in absolute event rates, and 73% of trials found noninferiority in the omission of chemotherapy on DRFS (Supplementary Figure 2, available online).

Table 1. Comparison of patient and tumor characteristics by treatment arm in the actual and simulated trial assigning individualized options for treatment (TAILORx)

Characteristics	Actual TAILORx trial		Simulated TAILORx trial*	
	Endocrine therapy (n = 3399)	Chemoendocrine therapy (n = 3312)	(RS 11–25)	
			Endocrine therapy (n = 3357)	Chemoendocrine therapy (n = 3354)
Median age at diagnosis (range), y	55 (23–75)	55 (25–75)	56 (23–75)	56 (23–75)
Age ≤50 y, No. (%)	1139 (34)	1077 (33)	1059 (32)	1058 (32)
Mean tumor size (SD), cm	1.7 (0.81)	1.7 (0.77)	1.7 (0.80)	1.7 (0.80)
Median tumor size (range), cm	1.5 (1.2–2.0)	1.5 (1.2–2.0)	1.7 (1.2–2.3)	1.7 (1.2–2.3)
Tumor grade, %				
Low	29	29	33	33
Intermediate	57	57	58	58
High	14	14	9	9
Recurrence score, %				
11–15	35	35	32	32
16–20	40	40	40	40
21–25	24	24	28	28
Primary surgery, No. (%)				
Mastectomy	28	28	28	28
Lumpectomy	72	72	72	72
Hormonal status				
ER+/PR+	90	90	85	85
Other	10	10	15	15

*Numbers reported for simulated TAILORx represents averages across 1000 trial replicates. ER = estrogen receptor; PR = progesterone receptor; RS = recurrence score; TAILORx = Trial Assigning Individualized Options for Treatment.

DRFS in Women Ages 18–75 Years and RS 16–25

The mean DRFS HR (SD) for endocrine therapy vs chemoendocrine therapy was 1.71 (0.82) among women ages 18–75 years with RS 16–25 with 9 years of follow-up, with a corresponding mean 9-year DRFS rate of 91.0% (0.03) for endocrine and 94.1% (0.01) for chemoendocrine therapy (absolute difference = 3.1%) (Table 2). At 20 years, the mean DRFS rate with endocrine therapy was 88.3% (0.07) and 90.6% (0.06) with chemoendocrine therapy.

DRFS in Women Ages 18–50 Years and RS 16–25

At 9 years, the mean DRFS HR (SD) for endocrine therapy vs chemoendocrine therapy was 1.79 (0.94), with a small absolute increase in DRFS rates (Table 2); 43% of the trials found noninferiority in the omission of chemotherapy (Figure 2). The distribution of hazard ratios from a sample of 100 trials (of 1000 trial simulations) is shown in Figure 3. The mean DRFS rates were 91.6% (0.04) for endocrine and 94.8% (0.01) for chemoendocrine therapy (Table 2).

DRFS in Women Ages 51–75 Years and RS 16–25

The mean DRFS HR (SD) for endocrine vs chemoendocrine therapy was 1.18 (0.73) with no difference in DRFS rates (Table 2).

DRFS in RS 16–20 and RS 21–25

We also evaluated the potential differences in endocrine therapy vs chemoendocrine therapy effects in RS 16–20 and RS 21–25 categories (Table 2). In women 50 years or younger, chemotherapy was associated with a lower rate of distant recurrence than endocrine therapy when the scores were 16–20 (3.2% at 9 years) or 21–25 (3.2% at 9 years). The proportion of trials showing noninferiority

was lower in RS 16–20 (35%) and RS 21–25 (26%) groups compared with the original TAILORx study with RS 11–25 (75%).

DRFS in Women Ages 18–75 Years and RS 26–30

Under this scenario, simulated women randomly assigned to treatment had a relatively higher proportion of high-grade tumors (20%) (Supplementary Table 2, available online) compared with those in the RS 11–25 group (9%) (Table 1). The mean DRFS HR (SD) for endocrine vs chemoendocrine therapy was 1.60 (0.83) (Table 2). The mean absolute difference in this group was 2.3% and did not change at 20 years (Table 2). Approximately 62% of the trials found noninferiority in omission of chemotherapy (Supplementary Figure 3, available online).

DRFS in Women Ages 18–75 Years and RS 0–25 vs 26–100

The proportion of simulated women with high-grade tumors in the RS 0–25 and 26–100 groups were 8% and 36%, respectively (Supplementary Table 2, available online). The mean DRFS HR (SD) for endocrine therapy vs chemoendocrine therapy was 1.14 (0.39) in the 0–25 group, with an absolute difference of 0.6% at 9 years and 0.7% at 20 years. In the RS 26–100 group, hazard ratio was 2.94 (2.03) (Table 2), with an absolute difference of 3.4% at 9 years and 0.8% at 20 years; only 20% of the trials found noninferiority in the omission of chemotherapy.

Discussion

This study illustrates the potential utility of simulation modeling to replicate and extend single, large clinical trials of personalized breast cancer therapy. Using data independent of

Table 2. Simulated trial results for hazard ratios and 95% confidence intervals associated with distant recurrence-free survival at 9 years and 20 years for endocrine vs chemoendocrine therapy by 21-gene expression recurrence score (RS) and age

Category	Age group, y	Sample size, mean*	Mean hazard ratio (SD)†	Trials showing noninferiority‡ in omission of chemo, %	9-year distant recurrence-free survival rates		20-year distant recurrence-free survival rates	
					Endocrine therapy, mean (SD), %	Chemoendocrine therapy, mean (SD), %	Endocrine therapy, mean (SD), %	Chemoendocrine therapy, mean (SD), %
RS 11–25	18–75	6711	1.15 (0.42)	75	94.0 (0.02)	94.4 (0.01)	89.7 (0.07)	90.7 (0.07)
	≤50	2449	1.55 (0.74)	59	94.4 (0.02)	95.9 (0.01)	NA	NA
	>50	3970	1.17 (0.37)	73	95.0 (0.02)	95.1 (0.01)	NA	NA
RS 16–25	18–75	2601	1.71 (0.82)	45	91.0 (0.03)	94.1 (0.01)	88.3 (0.07)	90.6 (0.05)
	≤50	2155	1.79 (0.94)	43	91.6 (0.04)	94.8 (0.01)	NA	NA
	>50	3970	1.18 (0.73)	70	92.1 (0.04)	92.8 (0.01)	NA	NA
RS 16–20	18–75	4437	1.72 (0.88)	34	90.8 (0.04)	94.2 (0.01)	85.6 (0.10)	88.0 (0.08)
	≤50	3771	1.73 (0.88)	35	91.7 (0.04)	94.9 (0.01)	NA	NA
	>50	4866	1.15 (0.70)	70	92.3 (0.04)	92.9 (0.01)	NA	NA
RS 21–25	18–75	2473	1.76 (0.93)	30	90.7 (0.04)	94.2 (0.01)	85.5 (0.10)	88.1 (0.08)
	≤50	2155	1.77 (0.97)	26	91.7 (0.04)	94.9 (0.01)	NA	NA
	>50	2694	1.17 (0.73)	65	92.3 (0.04)	92.9 (0.01)	NA	NA
RS 26–30	18–75	1372	1.60 (0.83)	62	91.6 (0.03)	93.9 (0.01)	89.7 (0.06)	91.0 (0.05)
RS 0–100	18–75	4310	1.14 (0.39)	75	94.2 (0.02)	94.8 (0.01)	89.8 (0.06)	90.5 (0.07)
	>25	1078	2.94 (2.03)	20	93.1 (0.01)	96.5 (0.02)	88.5 (0.08)	89.3 (0.08)

*Based on the original TAILORx protocol. NA = data inadequate to model long-term outcomes beyond 9 years in these subgroups; TAILORx = Trial Assigning Individualized Options for Treatment.

†Mean (SD) across 1000 trial replications.

‡Noninferiority was interpreted based on whether the confidence interval on the hazard ratio comparing endocrine vs chemoendocrine therapy contained the noninferiority margin (1.61) or no difference (1.00) according to the TAILORx protocol. The column provides the percentage of trials (out of 1000 simulations) that found noninferiority in the omission of chemotherapy on distant recurrence-free survival under each scenario.

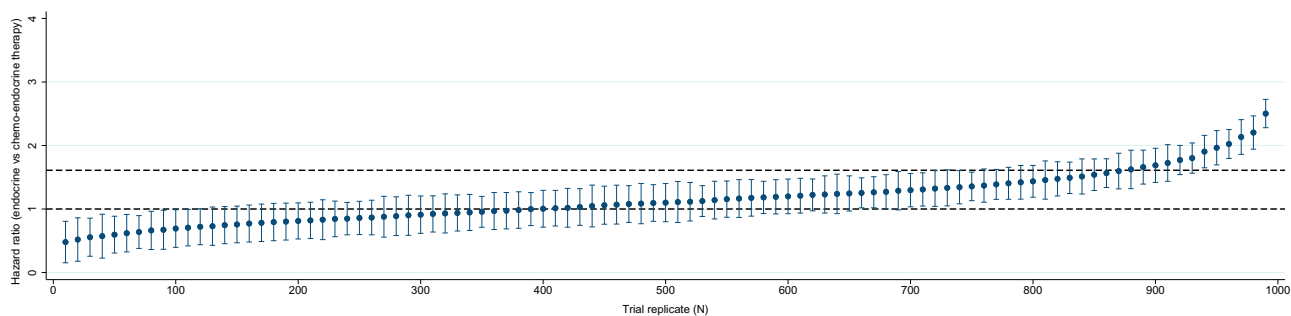


Figure 2. Distribution of hazard ratios and 95% confidence intervals for distant recurrence-free survival comparing endocrine therapy vs chemoendocrine therapy in 100 (out of 1000) replications of the Trial Assigning Individualized Options for Treatment (TAILORx). Horizontal dashed lines show noninferiority margin as defined in TAILORx. Noninferiority was interpreted based on whether the confidence interval on the hazard ratio comparing endocrine vs chemoendocrine therapy contained the noninferiority margin (1.61) or no difference (1.00). The hazard ratio for each trial replicate was derived from a Cox proportional hazards model.

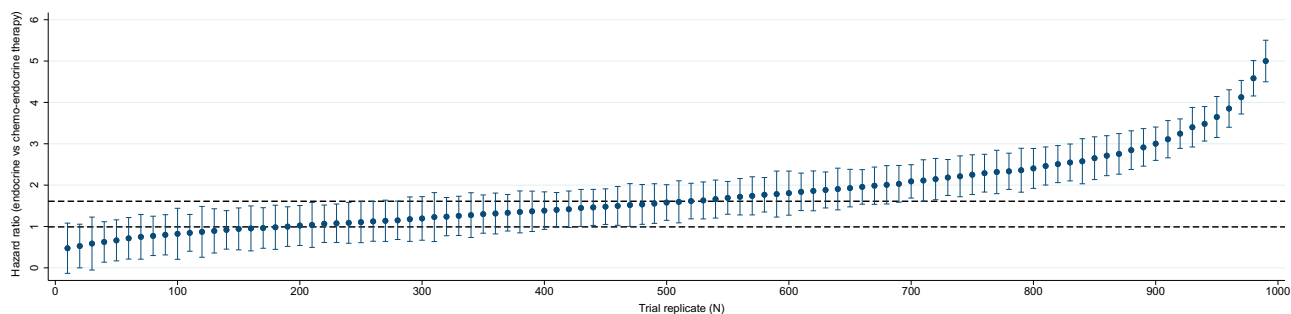


Figure 3. Distribution of hazard ratios and 95% confidence intervals for distant recurrence-free survival comparing endocrine therapy vs chemoendocrine therapy in 100 (out of 1000) replications of a clinical trial including only Trial Assigning Individualized Options for Treatment (TAILORx)-eligible younger women (≤ 50 years) with 21-gene recurrence scores ranging from 16 to 25. Horizontal dashed lines show noninferiority as defined in TAILORx. Noninferiority was interpreted based on whether the confidence interval on the hazard ratio comparing endocrine vs chemoendocrine therapy contained the noninferiority margin (1.61) or no difference (1.00). The hazard ratio for each trial replicate was derived from a Cox proportional hazards model.

TAILORx (1), model results confirmed that endocrine therapy was noninferior to chemoendocrine therapy for distant recurrence events among women ages 18–75 years with HR+/HER2-negative, node-negative breast cancers with RS 11–25. Further, this conclusion was seen in 75% of the 1000 trial replicates. The simulated trials also reproduced the TAILORx retrospective analysis results that chemoendocrine therapy was associated with a lower rate of distant recurrence events than endocrine therapy at 9 years among women age 50 years and younger with RS 16–25. The results were also similar across the RS 16–20 and RS 21–25 subgroups, with a lower proportion of trials finding noninferiority in the omission of chemotherapy compared with the TAILORx simulation. Extending TAILORx by randomly assigning chemotherapy among women with RS 26–100, most simulated trials showed a statistically significant benefit from chemoendocrine therapy, confirming the TAILORx choice to provide chemoendocrine therapy to all these participants. Simulation of 20 years of follow-up did not change overall conclusions. Variations among the 1000 trial replicates for each modeled scenario underscore expected differences when multiple trials are conducted given variability in populations and uncertainty in treatment effects.

Our results confirm that women ages 50 years and younger with RS 16–25 derive a relatively higher treatment benefit from adjuvant chemotherapy compared to older women (>50 years). These findings are consistent with results from a systematic review conducted by the Early Breast Cancer Trialists' collaborative group demonstrating that younger women derive a greater benefit from chemotherapy (5). This could be at least partly explained by an antiestrogenic effect associated with premature

menopause induced by chemotherapy in older premenopausal women (22). Therefore, our results add to the evidence that age should be factored into decisions about chemotherapy in women with early-stage, node-negative breast cancer. Further, a recent secondary analysis of TAILORx data examined whether a woman's recurrence risk based on classic clinical features such as tumor size and histologic grade adds prognostic information that is complementary to the 21-gene RS test (23). The results showed that high clinical risk features based on tumor size and grade can be used to identify a group of younger women (ages ≤ 50 years) with RS 16–25 who are likely to benefit from chemotherapy among TAILORx-eligible women. Accordingly, in our study, the observed chemotherapy effects in women ages 50 years and younger with RS 16–25 could also be explained by the higher benefit of chemotherapy derived by those simulated women with high clinical risk features.

The proportion of simulated trials that found noninferiority in the omission chemotherapy was lower in trials including only women with RS 26–30 (62%) and RS 26–100 (20%) compared with RS 11–25 (75%), indicating that a considerable proportion of women with scores greater than 25 could potentially benefit from chemotherapy. In addition to increasing distant recurrence event risk with increasing RS, these differences may also be partly explained by the relatively higher proportion of women with high-grade tumors belonging to the RS 26 or more group compared with those with RS 11–25. Consistent with previous findings, these results highlight the importance of integrating clinical-pathological features with RS to improve the accuracy of prognostic estimates (5,24).

The risk of distant recurrence events in early-stage breast cancer patients may remain up to 20 years (25). A recent meta-analysis showed that women with lymph node–negative disease could have an annual rate of distant recurrence at 1% from 5 to 20 years, resulting in a cumulative risk of distant recurrence of 13% (25). Our study confirms a cumulative increase in distant recurrence risk with increasing follow-up. However, the benefit of chemotherapy does not seem to increase at 20 years. Previous studies have shown that chemotherapy primarily prevents recurrences within 5 years of diagnosis (17). Even though more recurrences are expected with longer follow-up, the absolute benefits of chemotherapy remain small.

The average results across 1000 trial replicates for each scenario provided robust estimates of chemotherapy effects. However, we also observed that there was variation in these effects, with only a proportion of replicates in each scenario finding noninferiority in the omission of chemotherapy. This observation is related to several factors. First, there is expected uncertainty in the model-input parameters based on prior distributions of those parameters. Second, each trial randomly selects a set of eligible participants, so individual trials may include somewhat different populations, mirroring the reality of clinical practice. Third, there is variation in the predictive power of the 21-gene RS (7). Therefore, chemotherapy decisions in practice settings should integrate RS with traditional pathological and clinical measures to improve accuracy of prognostic outcomes for individual women (24).

Our analyses have several limitations. Historical trials (NSABP B14 and B20) provided limited sample size to derive input parameters evaluating treatment effects in smaller RS and age subgroups, contributing to uncertainty and variation in results. Another limitation is that we were unable to model the effects of specific chemotherapy regimens and durations of hormonal therapy (17,25,26). It is unclear whether the effects of adjuvant chemotherapy will vary based on type of chemotherapy within RS and age categories. Furthermore, we did not model the effects of treatment by RS based on mode of detection because screening mammography data were not available from any of our data sources (27–30). Another limitation is that patients' HER2 status was frequently missing in the Surveillance, Epidemiology, and End Results–Genomic Health dataset. Because HER2-positive tumors primarily have high RS, any misclassification of HER2 should not have affected our main findings confirming lack of chemotherapy benefits in patients with HER2-negative breast cancer and RS of 11–25 (31).

Overall, simulation modeling provides a powerful computational tool to synthesize evidence from various sources to evaluate treatment outcomes (and associated uncertainty) for different combinations of individual characteristics and an efficient laboratory to replicate and extend clinical trial results. Further, the output generated from simulation and modeling could be converted into a “calculation engine” to power decision tools (32) that could assist communication about treatment during clinical encounters.

Funding

This work was supported by a Lombardi Comprehensive Cancer Center American Cancer Society Young Investigator Award (ACS IRG 92-152-20) and the Cancer Prevention Research Fellowship sponsored by the American Society of Preventive Oncology and Breast Cancer Research Foundation (ASPO-17-001) to JJ. This work was supported in

part by the National Institutes of Health under National Cancer Institute Grant 1U01CA199218. The research was also supported in part by Grant R35 CA197289 to JM.

Notes

Affiliations of authors: Department of Oncology, Georgetown University Medical Center and Cancer Prevention and Control Program, Georgetown-Lombardi Comprehensive Cancer Center, Washington, DC (JJ, CI, SO, JM); Departments of Family and Social Medicine and Epidemiology and Population Health (CBS), and Department of Oncology at Montefiore Medical Center (JAS), Albert Einstein College of Medicine, Bronx, NY; Department of Biostatistics at Harvard University and Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA (RG); Departments of Medicine and of Health Research and Policy, Stanford University School of Medicine, Stanford University, Palo Alto, CA (AK).

CBS and JM contributed equally to this work.

JJ, SO, CBS, and JM have nothing to disclose.

JAS owns stock in Metastat; has served in an advisory role for Genentech/Roche, Novartis, AstraZeneca, Celgene, Lilly, Celldex, Pfizer, Prescient Therapeutics, Juno Therapeutics, and Merrimack; and has received research funding from Prescient Therapeutics, Deciphera, Genentech/Roche, Merck, Novartis, and Merrimack.

RG has received research funding from Abbott Molecular, Agios, Amgen, AstraZeneca, Bristol-Myers Squibb, Boehringer Ingelheim, Celgene, Genentech/Roche, Genomic Health, Genzyme, GlaxoSmithKline, ImClone Systems, Janssen-Ortho, Kanisa, Millennium, Nodality, Onyx, OSI Pharmaceuticals, Pfizer, Sanofi, Sequentia, Syndax, and Novartis.

CI has received research funding from Novartis (Inst), Pfizer (Inst), Genentech (Inst), Tesaro (Inst) and has served in a consulting or advisory role for Pfizer, Genentech, Novartis, AstraZeneca, Medivation, NanoString Technologies, Genentech, Celgene, Pfizer, AstraZeneca, and Caris Life Sciences. She receives patents, royalties, other intellectual property from UpToDate.

AK has received research funding to her institution from Myriad Genetics.

The study funders had no role in the design of the study; the collection, analysis, or interpretation of the data; the writing of the manuscript; or the decision to submit the manuscript for publication.

We thank Genomic Health, Inc, for provision of proprietary, deidentified, locked NSABP-Genomic Health data. We are also grateful to JAS and RG, and the Eastern Cooperative Oncology Group–American College of Radiology Imaging Network Cancer Research Group for sharing deidentified individual data for development of model parameters.

Earlier versions of this study were presented at the 2018 and 2019 Annual American Society for Clinical Oncology meetings, Chicago, Illinois, June 2–3, and the American Society of Preventive Oncology Annual Meeting in Tampa, Florida, March 12, 2019.

References

- Jayasekera J, Li Y, Schechter CB, et al. Simulation modeling of cancer clinical trials: application to omitting radiotherapy in low-risk breast cancer. *J Natl Cancer Inst*. 2018;110(12):1360–1369.
- Sparano JA, Gray RJ, Makower DF, et al. Adjuvant chemotherapy guided by a 21-gene expression assay in breast cancer. *N Engl J Med*. 2018;379(2):111–121.

3. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin*. 2018; 68(1):7–30.
4. Gradishar WJ, Anderson BO, Balassanian R, et al. NCCN Guidelines Insights breast cancer, version 1.2016. *J Natl Compr Canc Netw*. 2015;13(12): 1475–1485.
5. Dowsett M, Turner N. Estimating risk of recurrence for early breast cancer: integrating clinical and genomic risk. *J Clin Oncol*. 2019;37(9):689–692.
6. Sparano JA, Paik S. Development of the 21-gene assay and its application in clinical practice and clinical trials. *J Clin Oncol*. 2008;26(5):721–728.
7. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med*. 2004;351(27): 2817–2826.
8. de Koning HJ, Meza R, Plevritis SK, et al. Benefits and harms of computed tomography lung cancer screening strategies: a comparative modeling study for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2014;160(5): 311–320.
9. Mandelblatt JS, Stout NK, Schechter CB, et al. Collaborative modeling of the benefits and harms associated with different U.S. breast cancer screening strategies. *Ann Intern Med*. 2016;164(4):215–225.
10. Briggs A. Probabilistic analysis of cost-effectiveness models: statistical representation of parameter uncertainty. *Value Health*. 2005;8(1):1–2.
11. Schechter CB, Near AM, Jayasekera J, et al. Structure, function, and applications of the Georgetown-Einstein (GE) breast cancer simulation model. *Med Decis Making*. 2018;38(suppl 1):66s–77s.
12. Hudis CA, Barlow WE, Costantino JP, et al. Proposal for standardized definitions for efficacy end points in adjuvant breast cancer trials: the STEEP system. *J Clin Oncol*. 2007;25(15):2127–2132.
13. Fisher B, Costantino J, Redmond C, et al. A randomized clinical trial evaluating tamoxifen in the treatment of patients with node-negative breast cancer who have estrogen-receptor-positive tumors. *N Engl J Med*. 1989;320(8): 479–484.
14. Fisher B, Dignam J, Wolmark N, et al. Tamoxifen and chemotherapy for lymph node-negative, estrogen receptor-positive breast cancer. *J Natl Cancer Inst*. 1997;89(22):1673–1682.
15. Paik S, Tang G, Shak S, et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol*. 2006;24(23):3726–3734.
16. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc*. 1999;94(446):496–509.
17. Peto R, Davies C, Godwin J, et al. Comparisons between different polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100,000 women in 123 randomised trials. *Lancet*. 2012;379(9814): 432–444.
18. Dowsett M, Cuzick J, Ingle J, et al. Meta-analysis of breast cancer outcomes in adjuvant trials of aromatase inhibitors versus tamoxifen. *J Clin Oncol*. 2010; 28(3):509–518.
19. Petkov VI, Miller DP, Howlader N, et al. Breast-cancer-specific mortality in patients treated based on the 21-gene assay: a SEER population-based study. *NPJ Breast Cancer*. 2016;2(1):16017.
20. Cho H, Mariotto AB, Mann BS, et al. Assessing non-cancer-related health status of US cancer patients: other-cause survival and comorbidity prevalence. *Am J Epidemiol*. 2013;178(3):339–349.
21. Doran H, Bates D, Bliese P, et al. Estimating the multilevel Rasch model: with the lme4 package. *J Stat Softw* 2007;20(2):18.
22. Swain SM, Jeong JH, Geyer CE Jr, et al. Longer therapy, iatrogenic amenorrhea, and survival in early breast cancer. *N Engl J Med*. 2010;362(22): 2053–2065.
23. Sparano JA, Gray RJ, Ravdin PM, et al. Clinical and genomic risk to guide the use of adjuvant therapy for breast cancer. *N Engl J Med*. 2019;380(25): 2395–2405.
24. Tang G, Cuzick J, Costantino JP, et al. Risk of recurrence and chemotherapy benefit for patients with node-negative, estrogen receptor-positive breast cancer: recurrence score alone and integrated with pathologic and clinical factors. *J Clin Oncol*. 2011;29(33):4365–4372.
25. Pan H, Gray R, Braybrooke J, et al. 20-year risks of breast-cancer recurrence after stopping endocrine therapy at 5 years. *N Engl J Med*. 2017;377(19): 1836–1846.
26. Foldi J, O'Meara T, Marczyk M, et al. Defining risk of late recurrence in early-stage estrogen receptor-positive breast cancer: clinical versus molecular tools. *J Clin Oncol*. 2019;37(16):1365–1369.
27. Shen Y, Yang Y, Inoue LY, et al. Role of detection method in predicting breast cancer survival: analysis of randomized screening trials. *J Natl Cancer Inst*. 2005;97(16):1195–1203.
28. Joensuu H, Lehtimäki T, Holli K, et al. Risk for distant recurrence of breast cancer detected by mammography screening or other methods. *JAMA*. 2004; 292(9):1064–1073.
29. Wishart GC, Greenberg DC, Britton PD, et al. Screen-detected vs symptomatic breast cancer: is improved survival due to stage migration alone? *Br J Cancer*. 2008;98(11):1741–1744.
30. Berry D. The screening mammography paradox: better when found, perhaps better not to find. *Br J Cancer*. 2008;98(11):1729.
31. Geyer CE, Tang G, Mamounas EP, et al. 21-Gene assay as predictor of chemotherapy benefit in HER2-negative breast cancer. *NPJ Breast Cancer*. 2018;4(1):37.
32. Kurian AW, Munoz DF, Rust P, et al. Online tool to guide decisions for BRCA1/2 mutation carriers. *J Clin Oncol*. 2012;30(5):497–506.