Relationship between Complete Pathologic Response to Neoadjuvant Chemotherapy and Survival in Triple-Negative Breast Cancer

Christos Hatzis1, W. Fraser Symmans2, Ya Zhang2, Rebekah E. Gould2, Stacy L. Moulder3, Kelly K. Hunt4, Maysa Abu-Khalaf1, Erin W. Hofstatter1, Donald Lannin5, Anees B. Chagpar5, and Lajos Pusztai1

Abstract

Purpose: Pathologic complete response (pCR) to neoadjuvant chemotherapy reflects the cytotoxic efficacy of a drug, but patient survival is influenced by many other factors. The purpose of this study was to assess the relationship between increased pCR rate and trial-level survival benefit in triple-negative breast cancer (TNBC).

Experimental Design: We used bootstrap resampling from a neoadjuvant trial to simulate trials with different pCR rates. We used estimates from Adjuvant!Online to simulate trial populations with different baseline prognosis and estimated survival improvements associated with changes in pCR rate.

Results: Assuming that survival is similar for patients with pCR regardless of treatment arm, a linear relationship exists between increasing pCR rate and increasing recurrence-free survival (RFS). The slope is equal to the difference in survival between those with pCR and residual disease, which in turn is influenced by (i) the baseline prognosis of the trial population, (ii) interactions between prognostic variables and pCR, and (iii) the efficacy of the postneoadjuvant therapies. For example, if the pCR rates are 30% and 60% (OR = 3.5) and the 10-year RFS of the control arm is 0.74, the trial would require 3,550 patients per arm, whereas if the RFS is 0.54, the trial would require only 425 patients per arm to detect significant survival benefit.

Conclusions: We provide a framework for understanding the relationship between pCR and overall survival benefit that can help inform the design of neoadjuvant trials aiming to demonstrate improved survival from a regimen that results in higher pCR rate. Clin Cancer Res; 22(1): 26–33. ©2015 American Association for Cancer Research. See related commentary by Berry, p. 3

Introduction

Breast cancer patients who achieve pathologic complete response (pCR), defined as complete eradication of invasive cancer from the breast and lymph nodes, have excellent survival regardless of hormone receptor status or molecular subtype, although those with residual invasive disease (RD) have variable prognosis, depending on disease subtype and clinical stage (1–4). Triple-negative breast cancer (TNBC) patients with RD have a poor prognosis that is proportionate to the extent of residual disease, but some patients with estrogen receptor (ER)-positive cancer have excellent survival even with RD after neoadjuvant chemotherapy (3, 5, 6). One could expect that more effective cytotoxic therapies that produce higher pCR rates will lead to improved survival compared with less effective cytotoxic therapies, as they push more patients into the favorable survival group. Indeed, inclusion of taxanes in neoadjuvant chemotherapy increased pCR rates and also increased survival in much larger adjuvant trials (7, 8). The inclusion of trastuzumab with a broad range of chemotherapies doubled the pCR rates compared with chemotherapy alone, and also increased survival (9, 10). The NOAH study for locally advanced breast cancer, directly demonstrated improvement in pCR rate (38% vs. 19%) followed by an improvement in disease-free survival (DFS; 71% vs. 56%) in the same trial (11). Based on these and other supportive data, the FDA is currently considering pCR as an endpoint to support accelerated drug approval for early-stage breast cancer (12, 13).

However, the relationship between improvement in pCR rate and trial-arm level survival benefit is less than straightforward. Even studies that directly or indirectly support a positive association show substantially greater absolute improvements in pCR rate than in survival. Furthermore, the largest randomized neoadjuvant trial, the NSABP B-27 (n = 2,411), failed to demonstrate a significant improvement in 5-year DFS (71% vs. 68%) or overall survival (OS), despite a significant improvement in pCR rate (26% vs. 13%) with the inclusion of docetaxel with anthracycline/cyclophosphamide neoadjuvant therapy (2). Several other randomized neoadjuvant trials with sample sizes between 200 and 2,000, and including both ER-positive and -negative cases, also failed to demonstrate improvements in survival despite...
Translational Relevance

Pathologic complete response (pCR) to neoadjuvant chemotherapy is indicative of the cytotoxic efficacy of a drug, but patient survival is influenced by several other factors. The answer to the often-quoted question “Is pCR a surrogate for long-term survival?” is yes at the individual patient level. For the subtly different question “Is increased pCR rate a surrogate for improved survival in a trial arm?” the answer depends on the absolute improvement in pCR rate, baseline prognosis of the trial population, interaction of pCR with prognostic variables, and efficacy of postneoadjuvant treatment modalities. Our framework helps estimate the impact of increased pCR rate on survival and could help in designing and powering the next generation of neoadjuvant trials, particularly if the goal is to demonstrate improved survival from a regimen that results in higher pCR rate.

Materials and Methods

Patient data

We used survival data from 127 clinical stage I-III TNBC cases from our previously published cohort from MD Anderson Cancer Center (6) with recently updated survival outcomes (19). TNBC status was defined as ER and PR <1% positivity by routine IHC and lack of HER2 amplification by FISH (HER2 CEP17 copy number ratio <2.0) or HER2 IHC score of 0. All patients received preoperative chemotherapy with weekly paclitaxel and fluorouracil, doxorubicin (or epirubicin), and cyclophosphamide, as previously reported. pCR was defined as no residual invasive cancer in the breast or lymph nodes; the pCR rate for TNBC patients was 35%. The median follow-up for event-free survivors was 9.3 years (range, 0.5–13.4 years). The 5-year distant recurrence-free survival (DRFS) for the pCR and RD groups was 0.93 [95% confidence interval (CI), 0.86–1.00] and 0.58 (95% CI, 0.49–0.69), respectively. Patient characteristics are presented in Supplementary Table S1.

We also generated simulated survival data sets for TNBC using AO predictions (20) to vary the baseline prognosis of the trial population. In these data sets, all patients were designated as 60 years of age with average comorbidities, grade 3, ER-negative cancers and treated with third-generation adjuvant chemotherapy (i.e., control arm). Clinical tumor size and clinical nodal status were considered in brackets of increasing disease burden: T1cN0, T2N1, T3N1, T3N2. AO only provides DFS estimates and not DRFS, therefore in these data sets we report RFS as outcome. The overall 10-year RFS associated with the 4 distinct prognostic groups were 0.741, 0.545, 0.456, and 0.350, respectively (Table 1).
Statistical analysis

We used biased bootstrap resampling to simulate 1:1, two-arm randomized clinical trial scenarios with \( N = 127 \) patients per arm. For each scenario, 1,000 bootstrap replicates were drawn from the reference cohort with replacement, stratified by response group to keep the pCR rate at 35% in the control arm. To simulate the experimental arm, 1,000 different data sets were generated from the same TNBC reference cohort using biased resampling of the pCR and RD groups from the TNBC reference cohort to achieve various prespecified pCR rates. This simulation strategy requires two assumptions: (i) the survival of cases with pCR is similar regardless of which treatment arm they are in and (ii) the survival of patients with residual disease is also similar in both arms. These assumptions are supported by observations from several randomized neoadjuvant clinical trials (2, 17). We used the Kaplan–Meier estimator and univariate Cox regression to estimate the 5-year DRFS for the control and experimental arms and the treatment-related DRFS HR, respectively. Means and 95% CI were estimated from the 1,000 bootstrapped data sets for each scenario. The power to detect a significant DRFS HR was estimated by the proportion of data sets for which the Wald statistic for the HR was significant (\( P < 0.05 \)). The bootstrap procedure is equivalent to generating event and censoring times from the Kaplan–Meier estimator of the survival and censoring distributions (21, 22).

To evaluate the effect of baseline prognosis on survival, we generated cohorts for the control arm that matched the 10-year RFS predicted by AO for the specified patient characteristics. We further assumed independent exponential relapse times for the pCR and RD patient groups in either arm (23). We assumed that the pCR rate for the control arm (TNBC treated with third-generation chemotherapy) is 30% and that the 10-year RFS of patients who achieved pCR is 82% for both arms corresponding to an exponential hazard rate \( \lambda_{pCR} \) of 0.0198 (events/year; ref. 19). Under these assumptions, the survival function of the entire cohort for each arm will be the mixture of the survival functions of the pCR and RD groups. Based on this, we then calculated the 10-year RFS of patients with RD as

\[
S_{0}(10|RD) = \left( 1 - \frac{s_{0}(10)}{s_{0}(10)} \right) + \left( \frac{\pi_{0}}{1 - \pi_{0}} \right) S_{0}(10|pCR)
\]

where \( S_{0}(10) \) is the 10-year RFS for the overall cohort as predicted by AO, \( S_{0}(10|pCR) \) and \( S_{0}(10|RD) \) are the 10-year RFS for the pCR and RD subgroups and \( \pi_{0} \) is the pCR rate. The exponential hazard rate for the RD group is then estimated as \( \lambda_{RD} = - \frac{\ln S_{0}(10|RD)}{10} \). The values used in the simulated scenarios are summarized in Table 1. We modeled censoring times for both pCR and RD cohorts as being distributed according to an exponential distribution with rate \( \lambda_{C} = 0.5 \lambda_{RD} \), which corresponded to the censoring pattern observed in our reference cohort. The R (v.3.2.0) statistical software was used for all computations (24).

Results

The relationship between trial-level pCR rate and survival benefit

We consider a two-arm randomized neoadjuvant trial where the experimental treatment has greater cytotoxic efficacy compared with the control treatment, as reflected by an increase in pCR rate of \( \Delta t \) (Fig. 1). If the survival functions for the control and experimental arms are \( S_{0}(t) \) and \( S_{1}(t) \), then

\[
S_{0}(t) = \pi_{0} S_{0}(t|pCR) + (1 - \pi_{0}) S_{0}(t|RD) \quad \text{and} \quad S_{1}(t) = (\pi_{0} + \Delta t) S_{0}(t|pCR) + (1 - \pi_{0} - \Delta t) S_{0}(t|RD).
\]

Assuming that the survival outcome of patients with pCR is the same in both treatment arms, the survival benefit at the study level is

\[
\Delta EFS(t) = S_{1}(t) - S_{0}(t) = \Delta t (S_{0}(t|pCR) - S_{0}(t|RD)) + (1 - \pi_{0}) (S_{0}(t|RD) - S_{0}(t|RD)),
\]

indicating a linear association between trial-level survival benefit (\( \Delta EFS \)) and improvement in pCR rate (\( \Delta t \)), conditional on all other prognostic factors between the two arms being held equal. The magnitude of the trial-level survival benefit derived from an improvement in pCR is driven by the patient-level survival benefit between the pCR and RD groups in the experimental arm. If the baseline prognosis of the overall patient cohort is good, the difference in outcome between pCR and RD groups can only be small and therefore even a large \( \Delta t \) translates into a minimal trial-level survival benefit. A more effective cytotoxic therapy, in addition to pushing more patients into the pCR response category, may also reduce the extent of residual disease in those patients who did not achieve pCR. The second term in the expression is the secondary contribution of this residual disease downstaging on the trial-level survival benefit.

Association between odds for pCR and DRFS benefit in TNBC

Figure 2A and B illustrate a scenario where doubling of the pCR rate (from 35% to 70%) is accompanied by a sizable patient level survival benefit (Fig. 2A) yet it translates into only a small improvement in trial-arm level survival (Fig. 2B).

The impact of pCR OR on the DRFS HR and DRFS difference at 5 years was determined through simulations of a large number of trial scenarios. Higher OR results in smaller DRFS HRs (Fig. 2C) or greater DRFS difference at 5 years (Fig. 2D). Power analysis suggests that an experimental treatment that increases the pCR rate from 35% to 65% (pCR OR of 3.45) would have a 50%, 76%, and 90% power to detect a significant survival benefit at the 95% confidence level in 1:1 randomized trials with 127, 250, or 350 patients per arm, respectively (Fig. 2E). These results suggest that the power to detect significant survival HR effect is quite low (<50%) for studies with less than 350 patients per arm unless the pCR OR is high (OR > 2).

Power to detect significant survival benefit depends not only on pCR OR but also on the baseline prognosis of patients

To evaluate how baseline prognostic risk affects the trial-level survival benefit, we simulated a series of scenarios where the baseline prognosis of patients differed by adjusting anatomical risk variables (\( T \) size and \( N \) stage). We assumed a pCR rate of 30% in the control arm for all baseline risk groups, as the effect of tumor size and nodal status on pCR rate is modest (16). The study-level survival benefit, measured by the difference in 10-year RFS, increased with increasing pCR rate effect but was considerably larger in patients with the worst prognosis (T3N2; Fig. 3A). A 100
A patient per arm randomized trial that includes only T1cN0 patients, or any patient population with similar good baseline prognosis, would fail to show significant survival benefit even if the absolute pCR rate increase is >60% (corresponding to an OR > 20; Fig. 3A). Even studies with 1,000 patients per arm would require an absolute pCR increase >40% (OR > 5.4) to demonstrate a significant survival benefit in a good risk patient population (Fig. 3A). Figure 3B provides a more systematic exploration of the effect of baseline prognostic characteristics on the power to detect significant survival benefit. It shows that if a trial targets high-risk patients only, clinically T3N2 or similar high-risk population, then 300 patients per arm could yield 80% power to detect significant survival benefit for an absolute pCR rate increase of >20% corresponding to an OR of 2.3.

Table 2 provides sample size estimates for 1:1 randomized trials to detect a significant improvement at 5% significance level in trial-level RFS with 80% or 90% power as a function of (i) the improvement in pCR in the experimental arm and (ii) the overall baseline risk of the trial population. Power estimates for the extreme hypothetical scenario of 100% pCR in the experimental group versus 0% pCR in the control group are provided in Supplementary Table S2 for comparison. Even in this unrealistic scenario, a study involving only low risk T1cN0 patients will require 333 patients to show significant survival benefit with 80% power. It is apparent that the size of a phase III trial powered to demonstrate significant survival benefit depends not only on the expected pCR rate improvement but also to a large degree on the baseline risk of the treated cohort.

Discussion
There is confusion and contradictory data in the literature about how increase in pCR rate may affect survival difference between arms in a randomized trial. We show that trial-level survival benefit is a complex function of many factors beyond the difference in pCR rate. We used survival data from real patients participating in neoadjuvant studies and simulated patient outcome derived from AO estimates to assess the relationship between pCR rate and survival for TNBC. Assuming that the survival outcome is similar for those who achieve pCR, regardless of which treatment arm they are in, a linear relationship exists between trial-level pCR rate effect and RFS benefit. The slope is

Figure 1.
Patient-level (first plot) versus trial-level (second plot) survival analysis for a 1:1 two-arm randomized clinical study. Patient-level analysis shows a survival benefit associated with patients who achieve pCR compared with patients who have residual disease (RD). The trial-level survival benefit in the experimental treatment arm depends on the increase in pCR rate and also on the patient-level survival benefit, which varies depending on baseline prognostic risk of the trial population. In both clinical trial scenarios shown, the pCR rate increase in the experimental arm is 20% (pCR OR = 2.33), but scenario A is assumed to involve a low-prognostic risk population (T1c, N0) and scenario B a high-risk population (T3, N2). The prognostic risk characteristics of these populations are shown in Table 1. The survival functions of the pCR and RD groups are assumed to be exponentials.
equal to the difference in survival between those with pCR and RD, or the patient-level survival benefit, in the experimental arm, which in turn is influenced by the baseline prognosis of the trial population and the effectiveness of postneoadjuvant therapies. Patients with favorable tumor–node–metastasis (TNM) stage derive little benefit from even highly efficacious neoadjuvant therapies. Similarly, patients with highly endocrine sensitive cancers will derive a large survival benefit from adjuvant endocrine therapy and therefore their initial response to neoadjuvant cytotoxic therapy is less important.

There are molecular variables that could simultaneously influence both the probability of achieving pCR and the prognostic risk of a patient (i.e., risk of recurrence in the absence of any systemic therapy). For example, high tumor proliferation rate is associated with higher probability of pCR but also with worse prognosis (25), although high tumor infiltrating lymphocyte count is associated with higher pCR rate and better prognosis (26, 27). Many similar, yet to be identified, interactions between prognostic markers and pCR may exist. Some could be breast cancer subtype specific and may also differ by treatment and hard to model (28). We focused our analysis on TNBC because the pCR and survival relationship is the clearest in this disease subset due to the lack of effective post-neoadjuvant systemic therapies. We also assumed that the experimental therapy behaves like a third-generation chemotherapy in terms of its hidden interaction with prognostic and predictive markers, but it is more effective. This implies that the power and sample size relationship that we report here could be different for neoadjuvant therapies that have a substantially different mechanism of action because different interactions could exist between the experimental therapy and prognostic and predictive markers. Our model relies on a commonly used constant

Figure 2.
The Kaplan–Meier plots show (A) patient-level survival by responder group in each arm (pCR subgroups shown in solid lines and RD subgroups in dashed lines) and (B) the trial-level survival of the two arms for one typical bootstrap sample in the MDACC TNBC cohort. The control arms (n = 127) were generated through unbiased stratified bootstrap to have 35% pCR rate, and the experimental arms were generated by biased bootstrap resampling to yield a pCR rate of 70%. The association between pCR OR and trial-level survival benefit in terms of the DRFS HR or the difference in 5-year DRFS between the study arms is shown in (C) and (D), respectively. The points represent the means and the vertical lines the 95% CI of the survival outcome estimates obtained by 1,000 bootstrap replicates. E, the power to detect significance in DRFS HR at the 95% confidence level as a function of pCR OR for studies of different sample sizes (number of patients per arm: circles, N = 127; triangles, N = 250; squares, N = 350).
Hazard function to generate exponential survival estimates, although we recognize that the hazard of distant recurrence for TNBC peaks initially and then diminishes over time. We recognize that clinical trials accrue patients with different baseline prognoses, and we selected a set of specific T/N stage groups to illustrate the dependence on baseline prognosis of the hazard function to generate exponential survival estimates, although we recognize that the hazard of distant recurrence for TNBC peaks initially and then diminishes over time.

Table 2. Sample size per arm required for 1:1 randomized trials to detect improvement in trial-level RFS at a one-sided 5% significance level as a function of the overall baseline risk of the trial population and the improvement in pCR rate in the experimental arm (the pCR rate of the control arm is assumed to be 30% and the 10-year RFS of the pCR subgroup to be 82%).

<table>
<thead>
<tr>
<th>Baseline risk group</th>
<th>Power</th>
<th>Absolute improvement in pCR rate in experimental arm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.1</td>
<td>0.2</td>
</tr>
<tr>
<td>T1cN0</td>
<td>0.8</td>
<td>&gt;4,000</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>&gt;4,000</td>
</tr>
<tr>
<td>T2N1</td>
<td>0.8</td>
<td>3,078</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>&gt;4,000</td>
</tr>
<tr>
<td>T3N1</td>
<td>0.8</td>
<td>1,788</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>2,320</td>
</tr>
<tr>
<td>T3N2</td>
<td>0.8</td>
<td>1,122</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>1,488</td>
</tr>
</tbody>
</table>

NOTE: The power to detect significant survival benefit between treatment arms using unadjusted Cox proportional hazards regression analysis based on the Wald test (HR for RFS between arms <1) was estimated by 1,000 simulations of each baseline and pCR improvement scenarios with study sizes ranging from 30 to 3,200 patients per arm. Natural spline interpolation was then used to obtain the sample sizes for 80% and 90% power. No adjustments were made for patient accrual time.
current best therapies. An effective treatment could improve survival in the RD group by reducing the amount of residual disease without affecting the overall pCR rate resulting in a smaller trial-level survival benefit. Alternatively, a treatment could increase the pCR rate by shifting patients with minimal residual disease into the pCR category that would impact the net survival benefit only marginally as patients with minimal residual cancer burden have the same excellent prognosis as patients who achieve pCR [6]. The net effect of residual cancer downstaging on the association between pCR rate and trial-level survival will need to be evaluated in neoadjuvant studies that also monitor levels of residual disease.

In summary, our results indicate that modest improvements in pCR rate translate into very small improvements in survival in low-risk and moderate-risk trial populations, consistent with the existing literature. Almost all trials included in the CTNeoBC pooled analysis had pCR OR < 1.5 and included many ER-positive patients. These results do not undermine the value of pCR as an efficacy measure to identify more effective cytotoxic therapies in early-stage breast cancer but highlight the importance of selectively administering more effective novel treatments to those who remain high risk for recurrence despite current best therapies.

References

Relationship between Complete Pathologic Response to Neoadjuvant Chemotherapy and Survival in Triple-Negative Breast Cancer

Christos Hatzis, W. Fraser Symmans, Ya Zhang, et al.


Updated version
Access the most recent version of this article at:
doi:10.1158/1078-0432.CCR-14-3304

Supplementary Material
Access the most recent supplemental material at:
http://clincancerres.aacrjournals.org/content/suppl/2016/01/05/1078-0432.CCR-14-3304.DC1

Cited articles
This article cites 25 articles, 12 of which you can access for free at:
http://clincancerres.aacrjournals.org/content/22/1/26.full.html#ref-list-1

Citing articles
This article has been cited by 2 HighWire-hosted articles. Access the articles at:
/content/22/1/26.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.