



Published in final edited form as:

Contemp Clin Trials. 2013 January ; 34(1): 1–9. doi:10.1016/j.cct.2012.09.003.

Integrating Comparative Effectiveness Design Elements and Endpoints Into a Phase III, Randomized Clinical Trial (SWOG S1007) Evaluating OncotypeDX-Guided Management for Women With Breast Cancer Involving Lymph Nodes

Scott D. Ramsey, MD, PhD^a, William E. Barlow, PhD^{a,b}, Ana M. Gonzalez-Angulo, MD^c, Sean Tunis, MD^d, Laurence Baker, DO^e, John Crowley, PhD^{a,f}, Patricia Deverka, MD, MS^d, David Veenstra, PharmD, PhD^g, and Gabriel N. Hortobagyi, MD^c

^aFred Hutchinson Cancer Research Center, Seattle, WA

^bCancer Research and Biostatistics, Seattle, WA

^cThe University of Texas, MD Anderson Cancer Center, Houston, TX

^dCenter for Medical Technology Policy, Baltimore, MD

^eUniversity of Michigan, Ann Arbor, MI

^fSWOG Statistical Center, Seattle, WA

^gUniversity of Washington, Seattle, WA

Abstract

Women with breast cancer involving the lymph nodes are typically treated with cytotoxic chemotherapy. Retrospective evaluations of prior studies suggest that the 21-gene test (OncotypeDX®), may allow identification of those who can safely avoid chemotherapy. To better understand the performance of the 21-gene test, the RxPONDER (Rx for Positive Node, Endocrine Responsive breast cancer) study was designed, a multicenter Phase III trial randomizing women with hormone receptor-positive and HER2-negative breast cancer involving 1–3 lymph nodes and a 21-gene assay recurrence score (RS) of 25 or less to endocrine therapy alone versus chemotherapy followed by endocrine therapy. As one of the first large-scale comparative-effectiveness studies in oncology, RxPONDER utilized an external stakeholder group to help inform the design of the trial. Stakeholders met with representatives of SWOG over several months through a structured discussion process. The stakeholder engagement process resulted in several changes being made to the trial design. In addition, stakeholder representatives from the health insurance industry provided guidance regarding a mechanism whereby the costs of OncotypeDX® would be paid by the majority of health insurers as part of the trial. The process may serve as a template for future studies evaluating the comparative effectiveness of genomic tests in oncology, particularly those that are conducted within cooperative clinical trials groups.

© 2012 Elsevier Inc. All rights reserved.

Address for Correspondence: Scott D. Ramsey, MD, PhD, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. North, M3-B232, Seattle, WA 98109-1024, Phone (206) 667-7846, Fax (206) 667-5977, sramsey@fhcrc.org.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

breast cancer; comparative effectiveness research; OncotypeDx; everolimus; stakeholder

Introduction

Women with early stage breast cancer involving the lymph nodes are routinely offered chemotherapy in addition to surgery and radiation therapy. Although these recommendations are based on evidence from multiple randomized trials and a worldwide meta-analysis showing improvements in overall survival with chemotherapy,[1] there is also evidence that some, and perhaps many, patients do not benefit from chemotherapy; specifically, those with tumors that are well-differentiated, low grade, and with estrogen receptor positivity. In a retrospective evaluation of subsets of patients with hormone-receptor-positive tumors who received adjuvant endocrine therapy in two randomized clinical trials evaluating adjuvant endocrine therapy regimens, the 21-gene assay (OncotypeDX®) recurrence score (RS) combined with pathologic assessment was shown to have predictive value in identifying women who will benefit from chemotherapy with either node-positive[2] or node-negative breast cancer. [3]

OncotypeDX® is now being marketed as an option for patients with hormone-receptor-positive breast cancer involving 0–3 positive nodes, although genomic predictors such as OncotypeDX® are not yet widely employed in node-positive disease. Testing could potentially spare thousands of women from chemotherapy-related morbidity while also reducing chemotherapy expenditures by millions of dollars. However, it is not clear whether the RS can be calibrated to provide a balance between the benefits and harms of chemotherapy that would be acceptable to women and providers.[4]

The comparative effectiveness of management using OncotypeDX® versus current practice is an important question for this population. The Institute of Medicine defines comparative effectiveness research (CER) as “the generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat and monitor a clinical condition, or to improve the delivery of care.”[5] The inclusion of stakeholders to provide real-world insights into unmet information needs and inform the prioritization and design of clinical studies is considered to be a crucial feature of CER.

In this manuscript, we describe the design of SWOG S1007 (clinical trial registry: NCT01272037) given the acronym RxPONDER (Rx for Positive Node, Endocrine Responsive breast cancer). This multicenter Phase III trial randomizes to endocrine therapy alone versus chemotherapy followed by endocrine therapy for women with hormone receptor-positive and HER2-negative breast cancer involving 1–3 lymph nodes and a 21-gene assay RS of 25 or less. We also describe a unique aspect of this SWOG-led cancer cooperative group study: the participation of an external stakeholder group prior to finalizing the trial design and endpoints for the purpose of ensuring that the study results would be informative to patients, clinicians and payers while balancing considerations such as internal validity and trial feasibility. In addition, because reimbursement for genomic tests in the context of a clinical trial is unclear, we describe how a joint funding arrangement involving health insurers and the National Cancer Institute (NCI) was created for the RxPONDER study. The process we describe may serve as a template for future studies evaluating the comparative effectiveness of genomic tests in oncology, particularly those that are conducted within cooperative clinical trials groups.

Initial Designs of RxPONDER Trial

The initial RxPONDER trial design was developed by SWOG and submitted to the NCI's Breast Cancer Steering Committee for review, comment, and approval. Genomic Health, the manufacturer of the 21-gene assay (OncotypeDX®) had no role in the conception or design of the study. Reviewers included representatives of the Cancer Clinical Trials Cooperative Groups, the advocacy community, and other scientific personnel involved in breast cancer research. This study proposal involved a 2×2 factorial design with one factor centered on demonstrating non-inferiority of no chemotherapy versus chemotherapy for patients thought to be at low risk of cancer recurrence, as determined by the 21-gene assay RS and nodal status. The second factor was an efficacy comparison of everolimus in the adjuvant setting. This design required 6,000 patients to achieve its aims primarily due to the non-inferiority design.

The Breast Cancer Steering Committee expressed concern about the original inclusion criteria, which allowed women with up to 9 positive lymph nodes to be included, the lack of safety data for everolimus in breast cancer, and that the design allowed little latitude for addressing patient adherence to everolimus or a possible interaction of chemotherapy and everolimus. In response to these concerns SWOG removed the comparison of everolimus from the trial. SWOG also recognized that the non-inferiority design failed to take advantage of the continuous nature of the 21-gene RS. By considering the interaction of treatment and RS as the primary hypothesis, a new study design was employed that had the advantage of placing the issue of testing more centrally into the study.[6] The new design also allowed for identification of a RS cutpoint above which chemotherapy would be clinically beneficial. Finally, the redesign reduced the sample size to 4,000 patients randomized to chemotherapy or no chemotherapy, and included a margin for noncompliance. To address concerns about nodal status, the final design allowed inclusion of women with up to 3 positive lymph nodes. The final design will require screening of 9,400 women, taking into account the exclusion criteria and consideration that some women will refuse randomization after learning their RS.

The primary objective of the study is to test whether the difference in disease-free survival for patients treated with chemotherapy compared to no chemotherapy depends directly on the magnitude of the RS. If benefit depends on the RS score, the trial will determine the optimal cutpoint for recommending chemotherapy or not. The primary endpoint selected for the study was invasive disease-free survival (i.e. time to recurrence, new primary, or death due to any cause), with overall survival as a secondary endpoint. This endpoint was chosen in accordance with the STEEP system, a standardized approach to outcome definition that has been accepted for all major breast adjuvant trials.[7]

Incorporation of Comparative Effectiveness Research Elements into the Trial

From a comparative-effectiveness perspective, it could be argued that the preferable design is one that randomizes patients to one of two management options: “treat all” with chemotherapy (current standard of care) versus 21-gene assay RS-guided treatment. While this approach provides a direct test of the 21-gene assay versus the current standard of care, such designs are inefficient,[8, 9] in this case increasing the sample size requirements to randomize 9,000 women and to screen over 20,000. A second factor favoring the test-all then randomize strategy is that it greatly increases the number of women for whom tissue is collected. From a practical standpoint, it is likely that a larger proportion of women in the usual care arm would not have tissue collected in a way that permits centralized storage for future studies. Thus, the test-all strategy provides far greater numbers to test future assays

for predictive ability compared to the 21-gene RS assay or other risk-stratification approaches.

A key component of CER is the integration of perspectives from external stakeholders who represent constituencies that will be affected by research findings into the clinical trial design process.[10] The Center for Comparative Effectiveness Research in Cancer Genomics (CANCERGEN)—a multidisciplinary consortium involving the Fred Hutchinson Cancer Research Center (FHCRC), SWOG, the University of Washington (UW), and the Center for Medical Technology and Policy (CMTP), a non-profit organization involved in CER and stakeholder engagement activities—was created to incorporate external stakeholder input into RxPONDER and other SWOG-based clinical studies.

CANCERGEN's External Stakeholder Advisory Group (ESAG) included representatives from patient advocacy groups, clinicians, health insurers, test manufacturers, regulators and policy-makers. After reviewing the trial protocol approved by the Breast Cancer Steering Committee (see Figure 1), the ESAG met to discuss the RxPONDER trial design via conference call in January, 2010 and also completed a questionnaire that covered topics related to the trial design and implementation.(Table 1)

ESAG Input and Trial Modifications

The ESAG acknowledged the importance of the study and many felt, after discussing the rationale behind the design choices made for the trial (as approved by the Breast Cancer Steering Committee), that the objectives, study design, exclusion and inclusion criteria, and endpoints were appropriate. Many initial comments raised by the ESAG were considered and discussed by the Steering Committee during the formative stage of the design [e.g., randomizing to usual care (chemotherapy) vs. management based on the test results]. However, the ESAG had several suggestions for modifications to the trial that they felt would improve its quality as a CER study (Table 2). Particular aspects related to each issue are described below.

Participant Selection and Accrual

Several ESAG members noted that there may be factors correlated with a patient's decision to participate in the study that could be correlated with the outcomes observed. Specific concerns included family history of breast cancer, education, and disease knowledge. A suggestion was made to provide surveys including these and other factors to patients offered into the trial to see if external factors were influencing this and perhaps causing the study population to differ from the community. A related concern that was raised (also noted in the study description itself) is the number of women who decided not to be randomized after receiving their RS. There was a strong sense that all potential enrollees should be offered the chance to complete a survey to better understand their thinking and preferences surrounding the test (see **Measurement of Patient Preferences** below). There was also concern about the anticipated length of time needed for recruitment, the likelihood that new treatments will become available during this time, and possibility that this may influence accrual in later years of the study.

Measurement of Patient Preferences

Several respondents suggested capturing information related to patient decision making prior to knowing their RS and/or prior to deciding whether or not to enter the trial. This information would be particularly useful in helping practitioners understand influences on patient decision making related to breast cancer treatment choices. Some noted that it is known that some (perhaps many) patients ignore the results of their 21-gene RS assay, thus reducing the potential clinical utility of the test. Currently, there are no studies gathering

information around the time of the decision of whether or not to take the test, and following the results, the factors that most influence patient decisions about treatment. The ESAG noted that this would involve “upstream” surveys of people who eventually might not participate in the randomized trial, but felt the opportunity was too important to miss.

Estimating the Cost-Effectiveness of the 21-Gene RS Assay for Node-Positive Breast Cancer

Members expressed interest in understanding the cost-effectiveness of the 21-gene RS assay, particularly in light of the issues noted above. There was concern with ascertaining the cost-effectiveness of management with the 21-gene assay versus usual care given the way the trial was designed (i.e., not a direct comparison of management with vs. without the test, overall survival not a primary endpoint). Collecting patient preferences was mentioned as a way to facilitate a cost-effectiveness study. To ensure adequate power to identify potential differences in preferences, the sample size of the quality of life component of the study was increased from 360 to 1,000 individuals with the sample including women screened, but not randomized, either because of refusal or ineligibility (RS>25).

Patient Inclusion Criteria

A concern was raised by the ESAG clinicians and patient advocates that in clinical practice, some women decline full axillary node dissection due to fear of potential complications (e.g., pain, edema). The trial protocol originally required the full procedure, and thus might have made some ineligible to participate in the trial even though they had established nodal involvement. After exploring this issue with site investigators, the option was added to allow sentinel node biopsy as an alternative to determine nodal status if that represented the local standard of care and was the preferred procedure by the patient and her physician. In addition, ESAG members and others involved with RxPONDER were concerned with the original inclusion criteria, which covered women with up to 9 positive lymph nodes. Many felt that few women or clinicians would feel comfortable forgoing adjuvant chemotherapy when more than 3 lymph nodes were involved, regardless of the 21-gene assay score.

Tissue Storage

There was a suggestion to consider storing additional tissue, or collecting Peripheral Blood Mononuclear Cells prospectively, in a subset of (or all) patients to allow additional comparative-effectiveness studies in the future from the same resources. There was also a question about where samples will be housed, and the process of patient consent for future studies with their specimens.

Notification of Results to the Participants

Concerns were raised that the study did not mention whether or not participants will receive the results of the study or whether/how the results will be made available to trial participants, particularly if the results suggest that the 21-gene assay has a low negative predictive value.

The Value of Comparative Effectiveness Research

On the basis of the previous retrospective analysis of RSs in the node-positive population, others have generated comparative effectiveness analyses of use of OncotypeDx®. [11] These analyses note the uncertainty of their estimates and the need for replication. Many of these elements will be addressed in RxPONDER. A general issue encountered by the stakeholder group, trial designers, and research team was the relative benefits and costs of modifying the original study design to address comparative effectiveness issues, as outlined above. While cost estimates were provided for changes to the study design such as sample

size and addition of patient surveys, the benefits were significantly more difficult to quantify. An emerging method in CER is value of information (VOI) analysis. This approach, based in health economics and decision analysis, can be used to estimate the societal value of decreasing uncertainty about the benefits and harms of an intervention by conducting future research. The CANCERGEN team developed a VOI analysis of the RxPONDER study to assess its overall value in relation to its cost, and is described in a separate study [12]. However, given the time and resources required to conduct these analyses, they were not available to stakeholders at the time of decision making. The integration of VOI into the research prioritization and trial design process requires initiation at an early stage, as has been done within the CANCERGEN project for several other genomic tests.

Funding Sources

Trials of the size of RxPONDER are very costly. A potential advantage of CER is in leveraging trial funds through the participation and support of multiple stakeholder groups, including health insurers and the manufacturers of the technologies of interest. An important issue was finding funds to pay for the test itself (at a retail price of approximately \$4,000, the 21-gene assay is the largest single component of the study after study personnel costs). OncotypeDX® has been developed and marketed without FDA review, (CLIA approved lab developed tests are not required to undergo FDA review) and there is no requirement that it be provided free of charge as a part of clinical trials. The manufacturer of OncotypeDX®, Genomic Health Incorporated, did not agree to provide the test free of charge to trial participants. It was noted, however, that the National Comprehensive Cancer Network (NCCN) guidelines include the 21-gene RS assay as a potential management strategy for women with node-negative and pN1 mi or 2mm axillary node-positive disease. Although the NCCN criteria are more restrictive than the enrollment criteria for RxPONDER, because of limited health insurer ability to distinguish among types of nodal involvement, it was felt that the 21-gene assay was likely to be covered by most private and public insurers. This provided a potential avenue of billing patients' insurers for the test, since it would be part of a recognized management approach for the women in the study. One issue of interest was whether the Centers for Medicare and Medicaid Services (CMS) would pay for the assay under its Coverage with Evidence Development (CED) policy (approximately 30% of patients enrolled in RxPONDER were estimated to have Medicare or Medicaid insurance). Discussions with CMS, however, made it clear that this was not possible due to the fact that several local CMS contractors already paid for the 21-gene test for both node-negative and positive disease (CMS would have to issue a national no-coverage decision in order to make the test eligible for coverage under the CED rule). In addition, discussions with commercial insurers suggested that coverage of the 21-gene RS assay varied, but was generally accepted by the majority of carriers. It was not feasible to contact every carrier to address coverage prior to the start of the trial.

To minimize the burden of the 21-gene RS assay to the trial and also to keep it from being a barrier to entry for patients, it was agreed that the primary insurers for patients who participated in the study would be billed for the OncotypeDX® test. In the case where insurance rejected the claim, a fund would be established to reimburse patients for test-related expenses. The National Cancer Institute, the Breast Cancer Research Foundation[13], and the Komen Foundation[14] agreed to contribute to this fund.

The NCI provides primary funding for other expenses related to the RxPONDER study.

Figure 1 shows the trial design at the launch of the study (February, 2011). Table 3 shows the final study design, primary, and secondary endpoints and hypotheses.

Sample Size and Power Analyses

The primary hypothesis tests an interaction of randomized treatment (chemotherapy versus no chemotherapy) and RS. The actual statistical model for the test uses a Cox regression model and a test of this interaction. The underlying model and sample size computations are too lengthy to be included here, but are described in the following handbook.[15] The sample size yields 81% power to test the interaction and to determine the optimal cutpoint for recommending chemotherapy.

Discussion

Prospective experimental studies are a crucial source of CER information, and for questions that can be addressed effectively with these methods, study designs and infrastructure must be available to generate credible and relevant information as quickly, efficiently, and inexpensively as possible.[16] The Cancer Cooperative Clinical Trials Program offers a compelling infrastructure for meeting these goals; however, several challenges must be addressed before the full potential of this network for CER can be realized. Our experience with adapting the RxPONDER study to address CER goals illustrates many of the opportunities and issues that will be faced by investigators involved in future CER studies in this setting, and offer some insight into how it is possible to address some of these challenges.

Issue 1: Conceptualization and Timing—The initial study was not conceptualized as a CER study, and the process by which stakeholders were brought in was influenced by the timing of funding of the CANCERGEN study, which started about 8 months after planning for the clinical trial began. ESAG members thus required time to learn about the clinical context while at the same time understanding the study design itself. During this process, ESAG members raised many issues that the clinical trial team felt had already been resolved. Ultimately, the ESAG input had only a modest impact on the study design, but an important influence on the study endpoints and the size of the quality of life study.

Phase III studies in oncology are in some ways an imperfect match for CER, because they are designed to test hypotheses about efficacy rather than comparative effectiveness. In addition, Phase III studies are usually conceived by clinical trialists who do not necessarily have the same perspectives as other stakeholders with interest in the intervention. On the other hand, Phase III trials are typically the most influential studies in changing clinical practice. In some cases successful Phase III studies can fail to change practice due to the restricted nature of the population studied or the high cost of the new treatment. It is reasonable to accept the differences between Phase III trials and CER and find ways to amend these studies so that the needs of a broader group of stakeholders are met. The two issues that offer the most potential for flexibility, in our view, are ensuring that the eligibility criteria are as broad and representative as possible and including study endpoints that provide evidence that is of interest to the greatest number of stakeholders. The RxPONDER clinical team responded very favorably to suggestions on these issues.

Issue 2: Planning and Initiating Studies—In the case of RxPONDER, the stakeholder engagement process probably had little impact on the timing of the study initiation, in part because some of the issues raised by the stakeholder group were also issues raised by the NCI's Breast Cancer Steering Committee (approval by this committee is required before a cooperative group study can begin), and partly because the need to secure funding for the 21-gene RS test (after approval) was a vital issue, given the size of the trial. It could be argued that ESAG involvement actually facilitated the start of the trial, because some of the stakeholders were able to help SWOG work through the issues of insurance reimbursement

for the 21-gene assay. The study design was first submitted by SWOG to NCI on April 3, 2009, but was not approved initially. ESAG meetings concerning S1007 took place on January 8, 2010 and June 8, 2010. The final form of the study was submitted to NCI on May 26, 2010, approved on June 9, 2010, and the first patient was accrued on February 28, 2011.

Clinical researchers sometimes express the concern that the process of soliciting stakeholder feedback might potentially delay the start date of a study. We believe that this issue can be minimized by ensuring the following: (1) proper planning to establish a trained, standing stakeholder advisory group (as opposed to *ad hoc* groups) allows stakeholders to familiarize themselves with cooperative group processes and clinical researchers to adjust protocol development timelines for effective interactions and (2) developing processes for triaging potential Phase III trials for CER studies. Many more Phase III study concepts are introduced within the cooperative group system than are ultimately approved and implemented, and not all Phase III studies warrant CER. Thus, while it is probably not possible to incorporate CER into all Phase III trial discussions, developing a checklist similar to that used by the NCI's Biomarker, Imaging, & Quality of Life Studies Funding Program (BIQSFP) may provide opportunities for researchers to identify promising studies and for better intergroup coordination of these activities.[17] Such checklists may be informed by concepts from VOI analyses. Table 4 shows an adaptation of the BIQSFP cost-effectiveness criteria for CER. Once a trial has been selected for potential modification as a CER study with the inclusion of an external stakeholder group, investigators would need to answer a standard series of screening questions to determine suitability for further investment.

Issue 3: Funding Opportunities CER in Phase III Cancer Trials—Phase III clinical trials are expensive, and paying for large trials such as RxPONDER will necessarily involve multiple funding sources. Because CER studies typically involve treatments that are already in practice, there are fewer barriers to health insurance reimbursement. In the case of Phase III trials, the treatment of interest may not be considered standard and therefore not covered by most insurers. In the case of RxPONDER, insurance payment policies for OncotypeDX® for women with lymph node positive breast cancer were unclear, although the NCCN practice guidelines suggested that most insurers are “enabled” to pay for the test in this clinical setting. In cases where interventions are not typically reimbursed and the intervention is not addressed in practice guidelines, CED policies may be the most useful approach for study organizers to engage with health insurers. It is important to note that while Medicare has established procedures for considering whether its CED applies in a particular setting, very few commercial insurers have established programs for CED. We believe that this is an important area for developmental work, and an area for engagement between cancer cooperative clinical trials groups and the health insurance community. VOI analyses may be helpful in the discussions between health insurers and cancer clinical trials groups, as the method includes projections of use of the experimental therapy with and without information from the trial, and the economic impact of that use on health plan budgets. In many cases, the value of CER will increase as oncology treatments become more expensive.

Other Benefits of Stakeholder Engagement for CER in Phase III Trials

Meaningful involvement of stakeholders in the design of Phase III clinical trials can improve their relevance and importance to a wide variety of audiences. This may enhance the impact of Phase III studies, and speed the uptake of relevant findings into clinical practice. For example, if health insurers are involved in clinical trials of a new test (or new application of a test, such as is the case for OncotypeDX®), their participation may better enable them to modify coverage policies to facilitate use of the test if the study demonstrates meaningful

clinical and economic benefit for the intervention vs. usual care. Similarly, evaluation of patient preferences around testing may enable clinicians to identify and address potential barriers to acceptance of test results, thereby improving the clinical utility of testing.

Research Agenda

Incorporating CER design elements and endpoints into Phase III cancer trials is a work in progress. The best approach to effectively engage external stakeholders in this process has yet to be defined. One important element for future research will be to further test a framework whereby stakeholders can most effectively translate their individual or organizational perspectives into the trial design process in a way that promotes productive discussions.[18] For example, manufacturers will want to expand access to their products, health insurance plans will be mindful of allocating resources under constrained health budgets, and patient advocates desire reduced barriers to products they feel offer potential benefits to individuals with the condition of interest, but not all of these goals can be met in the context of a clinical trial. An additional issue is selecting CER research priorities from the larger portfolio of potential and planned Phase III studies. This will also require both a framework for stakeholder input and a methodological approach. VOI methods, which quantify the amount decision makers would be willing to pay for information prior to making a decision, may be a useful tool to allocate scarce research resources for CER as part of Phase III studies.[19] VOI has been proposed as an alternative to traditional methods for finding study sample sizes that are sufficient to address the multiple objectives and hypotheses in combined Phase III + CER studies, but there are few applications of these methods to date.[20–22] Finally, approaches for reporting the results of CER, either alongside the parent trial or separately requires understanding of the needs of the many stakeholders involved. The Patient Centered Outcomes Research Institute (www.pcori.org) has identified communication and dissemination as a priority research area. This may be one avenue for financing studies aimed at advancing our understanding of communicating the results of these trials to diverse audiences.

Acknowledgments

Funding Source: This work was funded by the Center for Comparative Effectiveness Research in Cancer Genomics (CANCERGEN) through the American Recovery and Reinvestment Act of 2009 by the National Cancer Institute, National Institutes of Health under Agency Award #5UC2CA148570-02 and by PHS Cooperative Agreement grants CA32102 and CA38926 awarded to SWOG by the National Cancer Institute. The content of this manuscript is solely the responsibility of the authors and does not necessarily reflect the views or policies of the National Cancer Institute, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

References

1. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet*. 2005; 365:1687–717. [PubMed: 15894097]
2. Albain KS, Barlow WE, Shak S, Hortobagyi GN, Livingston RB, Yeh IT, et al. Prognostic and predictive value of the 21-gene recurrence score assay in postmenopausal women with node-positive, oestrogen-receptor-positive breast cancer on chemotherapy: a retrospective analysis of a randomised trial. *Lancet Oncol*. 2010; 11:55–65. [PubMed: 20005174]
3. Tang G, Cuzick J, Costantino JP, Dowsett M, Forbes JF, Crager M, 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; 29:4365–72. [PubMed: 22010013]
4. Puzstai L. Anatomy and biology: two complementary sides of breast cancer prognostication. *J Clin Oncol*. 2011; 29:4347–8. [PubMed: 22010020]

5. Sox HC, Greenfield S. Comparative effectiveness research: a report from the Institute of Medicine. *Ann Intern Med.* 2009; 151:203–5. [PubMed: 19567618]
6. Crowley, J.; Hoering, A.; Ankerst, D. *Handbook of Statistics in Clinical Oncology*. 2. Boca Raton, FL: CRC Press; 2006.
7. Hudis CA, Barlow WE, Costantino JP, Gray RJ, Pritchard KI, Chapman JA, et al. Proposal for standardized definitions for efficacy end points in adjuvant breast cancer trials: the STEEP system. *J Clin Oncol.* 2007; 25:2127–32. [PubMed: 17513820]
8. Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics.* 1975; 31:103–15. [PubMed: 1100130]
9. Hoering A, Leblanc M, Crowley JJ. Randomized phase III clinical trial designs for targeted agents. *Clin Cancer Res.* 2008; 14:4358–67. [PubMed: 18628448]
10. Hoffman A, Montgomery R, Aubry W, Tunis SR. How best to engage patients, doctors, and other stakeholders in designing comparative effectiveness studies. *Health Aff (Millwood).* 2010; 29:1834–41. [PubMed: 20921483]
11. Hall PS, McCabe C, Stein RC, Cameron D. Economic evaluation of genomic test-directed chemotherapy for early-stage lymph node-positive breast cancer. *Journal of the National Cancer Institute.* 2012; 104:56–66. [PubMed: 22138097]
12. Wong W, Ramsey SD, Barlow WE, Garrison LP, Veenstra DL. The value of comparative effectiveness research: Projected return on investment of the RxPONDER trial (SWOG S1007). *Contemporary Clinical Trials.* 2012; 33:1117–23. [PubMed: 22981891]
13. The Breast Cancer Research Foundation. 2012.
14. Susan, G. *Komen for the Cure*. 2012.
15. Barlow, WE. Design of a clinical trial for testing the ability of a continuous marker to predict therapy benefit. In: Crowley, J.; Hoering, A., editors. *Handbook of Statistics in Clinical Oncology*. 3. Boca Raton: CRC Press; 2012. p. 293-304.
16. Tunis SR, Benner J, McClellan M. Comparative effectiveness research: Policy context, methods development and research infrastructure. *Statistics in medicine.* 2010; 29:1963–76. [PubMed: 20564311]
17. Services DoHaH. Biomarker, Imaging, & Quality of Life Studies Funding Program (BIQSFP). Vol. 11. BQSFP Funding Announcement; 2011.
18. Deverka PA, Lavalley DC, Desai PJ, Esmail LC, Ramsey SD, Veenstra DL, et al. Stakeholder participation in comparative effectiveness research: defining a framework for effective engagement. *Comparative Effectiveness Research.* 2012; 1:1–14.
19. Claxton K, Posnett J. An economic approach to clinical trial design and research priority-setting. *Health Econ.* 1996; 5:513–24. [PubMed: 9003938]
20. Willan A, Eckermann S. Optimal clinical trial design using value of information methods with imperfect implementation. *Health Econ.* 2010 May; 19:549–61. [PubMed: 19399753]
21. Eckermann S, Karnon J, Willan A. The value of value of information: best informing research design and prioritization using current methods. *Pharmacoeconomics.* 2010; 28:699–709. [PubMed: 20629473]
22. Koerkamp BG, Spronk S, Stijnen T, Hunink MG. Value of information analyses of economic randomized controlled trials: the treatment of intermittent claudication. *Value Health.* 2010; 13:242–50. [PubMed: 19818058]

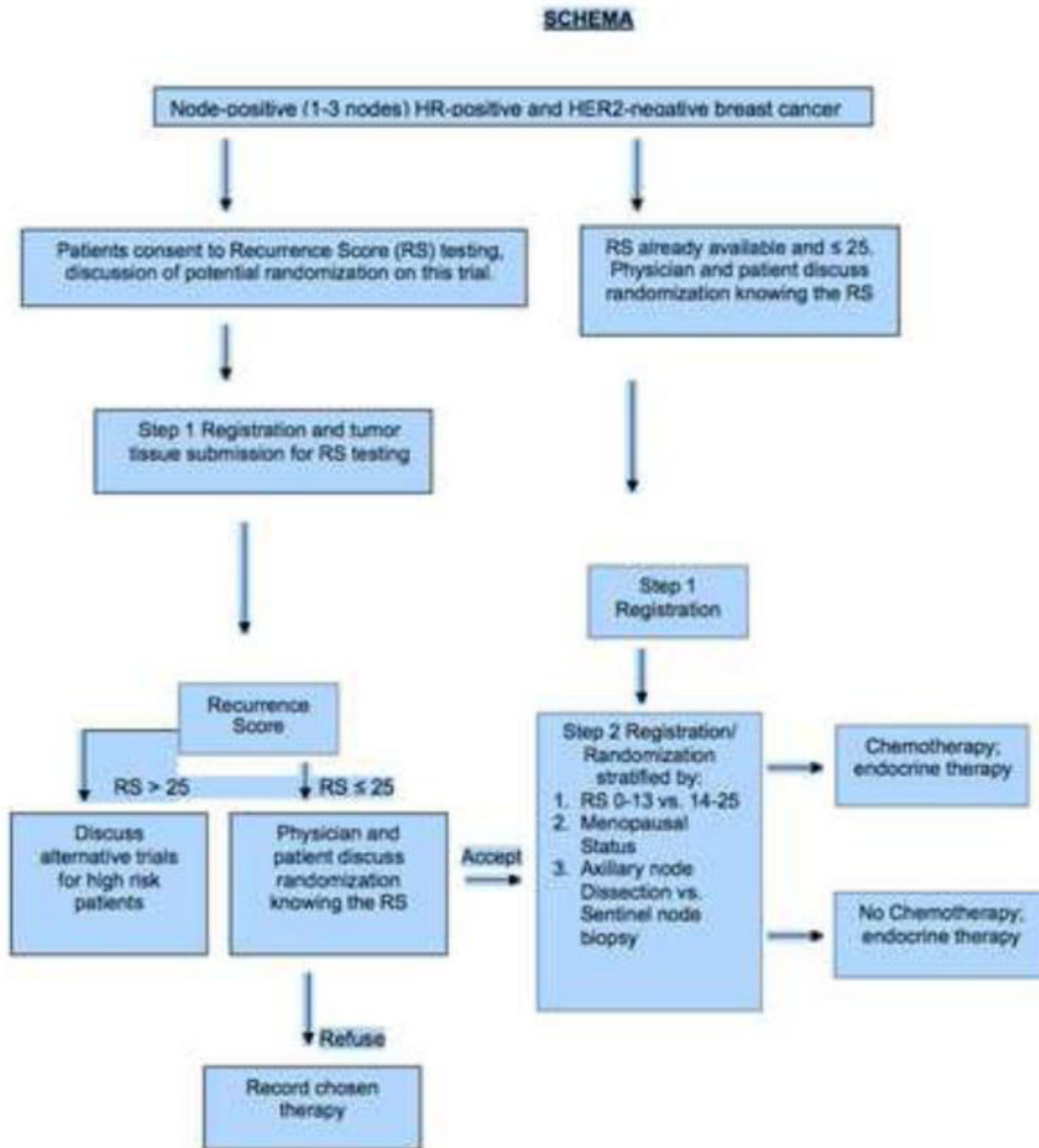


Figure 1. Study design for eligibility, recruitment, and randomization for the RxPONDER study.

Table 1

CANCERGEN External Stakeholder Advisory Group questions regarding design and endpoints for the RxPONDER (Rx for Positive Node Endocrine Responsive Breast Cancer) study.

<p>1 Are the objectives, hypotheses, study design, and endpoints constructed in a way that will provide clear guidance to a number of decision makers (e.g. patients, physicians, insurers) regarding the clinical utility of the Oncotype DX test relative to current best practice for women with lymph node-positive, HR-positive, HER2-negative breast cancer? If not, what changes would you recommend regarding the following?</p> <p>Objectives:</p> <ul style="list-style-type: none"> i. Primary: Should the study consider other specific stratification/prediction approaches beyond the PAM 50, SET index, Adjuvant Online that are currently included in the protocol? ii. Secondary: Are the secondary objectives feasible given that the patients are not randomized to Oncotype DX testing or not? <p>Hypotheses:</p> <p>Study design:</p> <ul style="list-style-type: none"> i. <i>Exclusion/Inclusion Criteria:</i> Do the inclusion/exclusion criteria allow for a study population that reflects the population that is likely to receive Oncotype Dx in practice? ii. <i>Exclusion/Inclusion Criteria:</i> Should patients enrolled in the RxPONDER study be allowed to enroll in other trials? iii. <i>Exclusion/Inclusion Criteria:</i> Should high RS scores (>25) be excluded? <p>Endpoints:</p> <ul style="list-style-type: none"> i. Are particular outcomes not currently mentioned that should be included in the study? ii. Should a question about patient preference be included in the study to determine their preference for chemotherapy before and after receiving RS scores? <p>2 What additional comments do you have on the study or its design?</p> <p>3 Are the objectives, hypotheses, study design, and endpoints constructed in a way that will provide clear guidance to a number of decision makers (e.g. patients, physicians, insurers) regarding the clinical utility of the Oncotype DX test relative to current best practice for women with lymph node-positive, HR-positive, HER2-negative breast cancer? If not, what changes would you recommend regarding the following?</p>

Table 2

Suggestions from the External Stakeholder Advisory Group (ESAG) for consideration in the Design of the RxPONDER trial.

ESAG Suggestion	Modification of Trial Design in Response ^a
<ul style="list-style-type: none"> <li data-bbox="266 396 711 464">i. Monitoring those who offer to participate and are enrolled to the trial to ensure they match the general patient population for this disease. <li data-bbox="266 472 711 604">ii. Capturing patient preferences prior to receiving the Recurrence Score as well as background data (education, knowledge of disease, etc.) that might influence patient decision making regarding testing and willingness to be randomized. <li data-bbox="266 613 711 680">iii. Clarification of data elements necessary to will allow estimation of the cost effectiveness of the 21-gene assay. <li data-bbox="266 688 711 783">iv. Requiring full axillary nodal dissection may dissuade some eligible women from participation due to patient concerns regarding potential for procedure-related morbidity. <li data-bbox="266 791 711 861">v. Including more information for patients about tissue storage and how this tissue would be used in the future. <li data-bbox="266 869 711 919">vi. Stating how participants will be informed of the results of the study upon completion. 	<ul style="list-style-type: none"> <li data-bbox="773 396 1367 464">i. No change in the trial design. The goal is to maximize participation. Investigators will compare the enrolled population to the general population at the conclusion of the trial. <li data-bbox="773 472 1367 919">ii. Secondary objectives added: To determine the impact of management with Oncotype DX® on patient-reported anxiety (co-primary Health-Related Quality of Life [HRQL] outcome) prior to screening, afterdisclosure of test results, and during the randomized trial To determine the impact of management with Oncotype DX® on patient-reported decision conflict, perceptions regarding Oncotype DX® testing, and survivor concerns prior to screening, after disclosure of test results, and during the randomized trial. HRQOL questionnaires added: <ul style="list-style-type: none"> <li data-bbox="867 747 1367 774">• Perceptions regarding Oncotype DX testing <li data-bbox="867 783 1367 810">• Decisional Conflict Scale¹¹ <li data-bbox="867 819 1367 846">• EQ-5D¹² <li data-bbox="867 854 1367 919">• Emotional distress (PROMIS emotional distress-anxiety Short Form)¹³ <li data-bbox="773 928 1367 1144">iii. Secondary objectives added: Using modeling and DFS information from the trial, to estimate the cost-effectiveness of management with Oncotype DX® vs. usual care. Data collection added: <ul style="list-style-type: none"> <li data-bbox="867 1077 1367 1104">• Collection of health insurance records to estimate costs <li data-bbox="867 1113 1367 1140">• Health state utilities (EQ-5D) <li data-bbox="773 1152 1367 1203">iv. After consulting with sites, added option for sentinel node biopsy to determine nodal status as per local standard of care. <li data-bbox="773 1211 1367 1239">v. None. <li data-bbox="773 1247 1367 1274">vi. Will be implemented when the trial has progressed sufficiently.

^a Also see Table 3

Table 3**Primary and Secondary Objectives and Hypotheses for the RxPONDER Study.**

Objectives	
<i>Primary Objective:</i>	
1.1	To determine the effect of chemotherapy in patients with node-positive breast cancer who do not have high Recurrence Scores (RS) by Oncotype DX®. In patients with 1–3 positive nodes, and hormone receptor (HR)-positive, HER2-negative breast cancer with RS ≤ 25 treated with endocrine therapy we will test whether the difference in disease-free survival for patients treated with chemotherapy compared to no chemotherapy depends directly on the magnitude of RS. If benefit depends on the RS score, the trial will determine the optimal cutpoint for recommending chemotherapy or not.
<i>Secondary Objectives:</i>	
1.2	To compare overall survival (OS), distant disease-free survival (DDFS) and local disease-free interval (LDFI) by receipt of chemotherapy or not and its interaction with RS.
1.3	To compare the toxicity across the treatment arms.
1.4	To perform other assays or tests (in particular the PAM50 risk of relapse score), as they are developed and validated, that measure potential benefit of chemotherapy and compare them to Oncotype DX®.
1.5	To determine the impact of management with Oncotype DX® on patient-reported anxiety (co-primary Health-Related Quality of Life [HRQL] outcome) prior to screening, after disclosure of test results, and during the randomized trial. ^a
1.6	To determine the impact of Oncotype DX® on the initial management cost of node-positive, HR-positive, HER2-negative breast cancer. ^a
1.7	To compare patient-reported utilities (e.g. HRQL) for those randomized to chemotherapy versus no chemotherapy.
1.8	Using modeling and DFS information from the trial, to estimate the cost-effectiveness of management with Oncotype DX® vs. usual care.
1.9	To determine the role of other assays (e.g. PAM50) as predictors of DFS, DDFS and LDFI for patients randomized to chemotherapy versus no chemotherapy.
1.10	To determine the impact of treatment with chemotherapy versus no chemotherapy on patient-reported fatigue and cognitive concerns (secondary HRQL outcomes).
1.11	To determine the impact of management with Oncotype DX® on patient-reported decision conflict, perceptions regarding Oncotype DX® testing, and survivor concerns prior to screening, after disclosure of test results, and during the randomized trial (secondary HRQL outcomes). ^a
Hypotheses	
<i>Primary Clinical Questions:</i>	
H1	The 21-gene recurrence score (RS) will predict the benefit of chemotherapy in node positive (1–3 nodes), hormone receptor (HR)-positive breast cancer patients with RS ≤ 25 treated with state-of-the-art endocrine therapy. Chemotherapy benefit (if it exists) will increase as the RS increases.
H2	Chemotherapy is not beneficial for some patients in the range of RS 0–25 and that the point of equivalence between chemotherapy and no chemotherapy can be identified in this range. Above this point, patients begin to benefit from the addition of chemotherapy. We will identify a RS cutpoint for which there is clinically significant benefit of chemotherapy for all RS values above this cutpoint.
<i>Ancillary Comparative Effectiveness Research Hypotheses:</i>	
H1	Women who receive an RS that is > 25 will have greater anxiety than those whose RS is ≤ 25 when compared at the second assessment when RS results and treatment decisions are known.
H2	Women whose RS is ≤ 25 but do not consent to randomization will have lower anxiety scores than those who do agree to be randomized.
H3	EQ-5D Index scores will be worse for women whose RS is > 25 compared to women whose RS is ≤ 25 at the second assessment when RS and treatment decisions are known.
H4	Anxiety for women who are randomized to no chemotherapy will be greater than for those who are randomized to receive chemotherapy when compared at the second assessment and at all follow-up assessments.
H5	Patients receiving adjuvant chemotherapy will report worse EQ-5D utility or Index scores than patients who are receiving no chemotherapy.

^aItems modified or added following input from the CER external stakeholder advisory group.

\$watermark-text

\$watermark-text

\$watermark-text

Table 4**Concept Checklist for Incorporating Comparative Effectiveness Research into Randomized Phase 3 Clinical Trials with a Comparator (Control Arm).**

- 1 Describe and justify the rationale for including CER endpoints.
- 2 Explain the situations in which the CER outcomes of the clinical trial could substantially change clinical practice.
- 3 Describe the initial CER endpoints.
- 4 Describe the stakeholder groups most relevant to the trial, and process for engaging those groups in the study design, and the process for incorporating stakeholder suggestions into the trial design.
- 5 Describe the implication of the CER endpoints on the sample size requirements for the study.
- 6 Evaluate the return on investment of the CER study using value of information analysis, if feasible.
- 7 Describe the process of disseminating CER findings among stakeholder groups, particularly future patients who would be eligible for the investigational treatment.