



Published in final edited form as:

Med Care. 2008 August ; 46(8): 762–770. doi:10.1097/MLR.0b013e318178ead3.

Selecting High Priority Quality Measures for Breast Cancer Quality Improvement

Michael J. Hassett, MD, MPH,

Department of Medical Oncology, Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA 02115; P: 617-632-6631; F: 617-632-3161

Melissa E. Hughes, MSc,

Department of Medical Oncology, Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA 02115; P: 617-632-2268; F: 617-632-3161

Joyce C. Niland, PhD,

Division of Information Sciences, City of Hope National Medical Center, 1500 E. Duarte Road, Duarte, CA 91010-3000; P: 626-359-8111; F: 626-301 8802; jniland@coh.org

Rebecca Ottesen, MS,

Division of Information Sciences, City of Hope National Medical Center, 1500 E. Duarte Road, Duarte, CA 91010-3000; P: 805-594-0441; F: 805-594 0442

Stephen B. Edge, MD,

Department of Breast and Soft Tissue Surgery, Roswell Park Cancer Institute, Elm & Carlton Streets, Buffalo, NY 14263; P: 716-845-5789; F: 716 845-3434

Michael A. Bookman, MD,

Fox Chase Cancer Center, 333 Cottman Avenue, W11/OPD Dept, Philadelphia, PA 19111; P: 215-728-2987

Robert W. Carlson, MD,

Stanford Hospital & Clinics, Stanford Cancer Center, 875 Blake Wilbur Drive; Room #2236, Stanford, CA 94305-5826; P: 650-725-6457; F: 650-498-4696

Richard L. Theriault, DO, MBA, and

University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Unit 424, Houston, TX 77030; P: 713-792-2817; F: 713-794-4385

Jane C. Weeks, MD, MSc

Department of Medical Oncology, Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA 02115; P: 617-632-2509; F: 617-632-2270

Michael J. Hassett: mhassett@partners.org; Melissa E. Hughes: melissa_hughes@dfci.harvard.edu; Rebecca Ottesen: rottesen@coh.org; Stephen B. Edge: stephen.edge@roswellpark.org; Michael A. Bookman: michael.bookman@fcc.edu; Robert W. Carlson: rcarlson@smi.stanford.edu; Richard L. Theriault: rtheriau@mdanderson.org; Jane C. Weeks: jane_weeks@dfci.harvard.edu

Abstract

Background—While many quality measures have been created, there is no consensus regarding which are most important. We sought to develop a simple, explicit strategy for prioritizing breast cancer quality measures based on their potential to highlight areas where quality improvement efforts could most impact a population.

Methods—Using performance data for 9,019 breast cancer patients treated at 10 National Comprehensive Cancer Network institutions, we assessed concordance relative to 30 reliable, valid breast cancer process-based treatment measures. We identified four attributes that indicated there was room for improvement and characterized the extent of burden imposed by failing to follow each measure: number of non-concordant patients, concordance across all institutions, highest concordance at any one institution, and magnitude of benefit associated with concordant care. For each measure, we used data from the concordance analyses to derive the first 3 attributes and surveyed expert breast cancer physicians to estimate the fourth. A simple algorithm incorporated these attributes and produced a final score for each measure; these scores were used to rank the measures.

Results—We successfully prioritized quality measures using explicit, objective methods and actual performance data. The number of non-concordant patients had the greatest influence on the rankings. The highest-ranking measures recommended chemotherapy and hormone therapy for hormone-receptor positive tumors, and radiation therapy after breast-conserving surgery.

Conclusions—This simple, explicit approach is a significant departure from methods used previously, and effectively identifies breast cancer quality measures that have broad clinical relevance. Systematically prioritizing quality measures could increase the efficiency and efficacy of quality improvement efforts and substantially improve outcomes.

Keywords

Quality measurement; quality improvement; clinical practice guidelines; breast cancer and medical intervention

Introduction

There is widespread consensus that the quality of health care in the United States is suboptimal.¹⁻³ However, there is little consensus regarding how best to address this problem. Many different quality improvement strategies have been tested, with varying degrees of success.⁴⁻⁸ One common strategy has been to develop quality measures – estimates of the degree of adherence to practice standards – to serve as evidence-based tools for evaluating and improving quality of care. While quality measures have been proposed for nearly all aspects of medical care, few have been used widely, and even when measures have been implemented it is hard to assess the impact they have had on health care.

The process of creating quality measures usually involves a panel of experts who consider a number of factors – such as clinical evidence, feasibility of measurement, and potential impact on patients – and then come to consensus.⁹⁻¹³ However, in many cases consensus panels do not analyze clinical evidence or potential impact on patients in a consistent and explicit manner, and do not consider actual practice performance data when creating quality measures. As a result, the measures created by these panels are not always based on high quality evidence^{14,15}, may not identify situations for which practice performance is clearly sub-optimal^{16,17}, and frequently cannot highlight which processes of care should be targeted to effect improvements in quality¹⁴.

Measuring practice performance is an expensive and time-consuming process, and resources are limited. So there are strong incentives to identify which of the available and scientifically acceptable measures are most likely to impact quality of care and should therefore be given the highest priority. If practice performance is already high or a guideline applies to very few patients, then measuring quality relative to that guideline may be an inefficient use of resources because it will never translate into a significant improvement in outcomes at the population level.

Just as there are a growing number of quality measures, there are also a growing number of ways to use quality measures to influence practice performance. For example, they have been used to inform providers and institutions about their performance¹⁸, direct payment incentives¹⁹⁻²¹, and publicly report performance grades²²⁻²⁴. However, these efforts have yielded only limited success and there is no agreement regarding which approach is best. Moreover, it is possible that quality measures are not one-size-fits all tools – that different quality measures are required for different purposes. Some have suggested that quality measures could instead be used to identify where to target quality improvement efforts.²⁵ In the past, quality improvement programs have selected areas for intervention based on perceived importance, anecdotal evidence, financial impact, or intuition. Developing a systematic, explicit method of prioritizing a set of measures based on how well they identify high-priority targets for quality improvement could offer significant advantages, but would require detailed patterns-of-care data and accurate knowledge regarding the impact of recommended treatments on patients' outcomes.

Cancer care is no exception to the quality problem. Investigators have identified many circumstances in which cancer patients do not receive treatments proven effective.^{15,17,26,27} Breast cancer has been a prime focus of quality assessment efforts because it is prevalent, effective treatments exist, and public interest is strong. Despite extensive efforts to measure the quality of breast cancer care^{13,17,28,29} and disseminate treatment recommendations via clinical practice guidelines³⁰⁻³², significant disparities between recommended treatments and actual patterns of care persist³³⁻³⁶.

The National Comprehensive Cancer Network (NCCN), an alliance of twenty U.S. cancer centers, has developed a comprehensive set of evidence-based cancer guidelines^{12,37} and a prospective database on patterns-of-care at member institutions. Using these resources, it has generated an extensive set of evidence-based breast cancer quality measures.³⁸ We wanted to build upon this work and develop an integrated approach to quality improvement. Our objectives were (1) to develop an explicit, transparent, and simple strategy for prioritizing quality measures based on their potential to highlight areas where quality improvement strategies could improve the outcomes of a population, and (2) to apply this methodology to a comprehensive set of breast-cancer quality measures to identify those that offer the greatest potential to improve disease-free survival (DFS) and quality of life (QOL).

Methods

Developing Quality Measures

The NCCN has produced and regularly updates a comprehensive set of evidence-based cancer treatment guidelines.^{12,37} The NCCN also supports a prospective database for women with breast cancer treated at participating member institutions that collects all the information needed to assess concordance relative to the guidelines, including detailed patient, treatment, and outcomes variables.^{39,40} The eligibility criteria and data collection procedures for the database have been described previously.^{41,42} Some data elements, including socio-demographic features, are collected from a survey administered to patients when they first present. Other variables, including stage and treatments, are collected through a series of routine chart reviews by trained, dedicated abstractors. Rigorous quality assurance processes ensure the data are reliable.

To assess practice performance, we applied the methods originally developed by Weeks and colleagues.³⁸ First, a quality measure is created for each definitive guideline recommendation. Then, information from the outcomes database are used to calculate concordance with each measure – the number of patients who receive a recommended treatment divided by the number eligible to receive that treatment. As the guidelines are

updated to reflect new and emerging data, the measures are modified so they always conform with the most current version of the guidelines. Concordance is always assessed relative to the recommendations in place when a patient is treated. The reliability and validity of each measure is evaluated by the NCCN as part of its yearly internal concordance report.

We assessed concordance with 30 quality measures derived from the 2003 NCCN guidelines (Table 1) – the most current guidelines for which at least one year of follow-up data are available. Twelve recommendations apply to chemotherapy, eight to hormonal therapy, six to radiation therapy, and four to surgery. Twenty-one recommend for and nine recommend against a treatment. We assessed concordance among women with newly diagnosed stage 0-III breast cancer. Women 70 or older were excluded, because the NCCN guidelines report there are insufficient data to define chemotherapy recommendations for this cohort. Data from ten centers who volunteered to participate in the outcomes database are included in this analysis: City of Hope, Dana-Farber, Fox Chase, M.D. Anderson, Roswell Park, University of Michigan, Ohio State University, H. Lee Moffitt, University of Nebraska, and Stanford University. Institutional Review Boards (IRBs) from each center approved the data collection, transmission and storage protocols.

Prioritizing Quality Measures

Organizations that develop quality measures, such as the Institute of Medicine, recommend considering factors such as the extent of burden imposed by a condition, the extent of the gap between current practice and evidence-based practice, the likelihood the gap can be closed, and the relevance of an area to a broad range of individuals,⁴³ when creating measures. Our goal was to develop a systematic and explicit method of incorporating these factors into the historically subjective, consensus-based quality measure development process. We endeavored to prioritize a set of feasible and scientifically acceptable quality measures using factors similar to those enumerated by the Institute of Medicine.

We identified four key attributes that could be used to help prioritize quality measures. The overall concordance, defined as the number of patients who receive a recommended therapy divided by the number eligible to receive that therapy, serves as a measure of performance across all institutions and helps characterize the extent of the gap between current and evidence-based practice. We considered using mean institutional concordance instead, to reduce the influence of large centers, but wanted the overall assessment to reflect performance at the population rather than the institution level. The number of patients who did not receive concordant care, calculated as the number of eligible patients minus the number who receive the recommended treatment, shows how relevant a measure is for a broad range of individuals. The highest concordance achieved by any one institution represents a realistic benchmark that all institutions should strive to achieve, and helps identify an achievable goal. Concordance values from institutions enrolling fewer than 10 patients on a recommendation were considered unreliable, and were excluded when determining the highest concordance for that recommendation.

Physician Survey

The impact of a quality improvement program on a population depends not only on the number of patients who could benefit, but also on the magnitude of benefit experienced by each patient. To fully consider the extent of burden imposed by a condition when prioritizing quality measures, a fourth feature was identified – how much better would a population's outcomes be if patients who had received non-concordant care instead received concordant care. To estimate the impact specific treatments have on outcomes, we surveyed a panel of physicians who have expertise in breast cancer and familiarity with the NCCN

guidelines. The panel included 22 medical, radiation and surgical oncologists from 19 U.S. cancer centers who participated in the NCCN breast cancer guidelines panel. The survey was developed using an iterative approach, with serial rounds of question development followed by feedback from test subjects. The Dana-Farber IRB reviewed the survey and its associated methodologies.

The goal of the survey was to have respondents consider both DFS and QOL, and generate a single estimate of the benefit experienced by patients who received the recommended treatment instead of a common non-concordant treatment. First, participants were asked to estimate the improvement in DFS as the percent absolute benefit at five years, and the improvement in QOL as “greatly favors the recommended treatment,” “slightly favors the recommended treatment,” “no difference,” “slightly favors the non-concordant treatment,” or “greatly favors the non-concordant treatment.” Then, participants were asked to consider both factors and report one estimate of the magnitude of benefit using a seven point scale: none (0), minimal (1), small (2), moderate (3), large (4), very large (5), or substantial (6). The mean magnitude-of-benefit for each recommendation was divided by six (the maximum value) to produce a fractional magnitude-of-benefit estimate.

Algorithm Development

The four attributes described above were incorporated into a single algorithm (Figure 1). The highest concordance minus the overall concordance was divided by 100% minus the overall concordance, and this result was multiplied by the number of non-concordant encounters. In essence, the algorithm computed the number of patients whose care would have to be converted from non-concordant to concordant to raise the overall concordance to the benchmark value. This number was then multiplied by the fractional magnitude-of-benefit estimate to derive a final score.

Statistical Analysis

To explore whether the algorithm gives too much or too little weight to one contributing factor, sensitivity analyses were performed to assess the effect of adjusting the weights of the four variables on the rankings. Adjustments included squaring each variable and performing a natural logarithmic transformation of the skewed number-of-discordant-patients variable. Spearman correlation coefficients were used to assess the relationship between the magnitude of benefit estimates and the survival and quality-of-life estimates, and to explore the association between the final scores and the four contributing variables. Trend lines were derived using simple linear regression. Statistical analyses were performed using SAS software version 9.1 (Cary, NC). A two-sided P value < 0.05 was considered significant.

Results

Evaluating Quality Measures

Concordance analyses were performed on 9,019 women with newly diagnosed breast cancer. A majority (56%) were 50-69 years old; the rest were under 50. At diagnosis, 14.3% had DCIS, 38.4% had stage I, 40.3% had stage II, and 7.1% had stage III breast cancer. The Eastern Co-operative Oncology Group performance status was 1 in 12.9%; the Charlson co-morbidity score^{44,45} was 1 in 18.4%. Most (83.1%) were Caucasian non-Hispanic; 7.2% were African-American and 9.8% were of another race-ethnic background.

Magnitude-of-benefit estimates came from the survey of expert breast cancer clinicians. Thirteen physicians from twelve institutions completed the survey (response rate 59%). To assess the validity of the magnitude-of-benefit estimates as a measure of DFS and QOL, we

compared magnitude-of-benefit estimates with DFS and QOL estimates separately for recommendations stating that treatments should and should not be administered (Figure 2). As expected, there were significant positive relationships between magnitude-of-benefit and DFS estimates among recommendations for treatment (Spearman correlation = 0.80; $p < 0.001$), and between magnitude-of-benefit and QOL estimates among recommendations against treatment (Spearman correlation = 0.92; $p < 0.001$). Respondents identified little DFS benefit, but reported a range of magnitude of benefit estimates for recommendations stating treatments should not be administered.

Prioritizing Quality Measures

Quality measures were ranked based on their final scores. The five highest-ranking measures, listed in Table 2, assess the care offered to 65% of the cohort. Together, they encompass 45% of all the episodes where a recommended treatment was not provided. Sensitivity analyses in which we adjusted the weights of the four variables, as described above, yielded rankings that correlated highly (Spearman correlation = 0.8; $P < 0.01$) with those generated by the simplest algorithm. Therefore, the simplest algorithm (Figure 1) was selected for further analyses.

The final scores and the four values used to generate these scores for each measure are presented in parallel in Figure 3. The number of non-concordant patients showed the greatest correlation with, and had the greatest influence on, the final scores (Spearman correlation = 0.90; $P < 0.001$). The magnitude-of-benefit estimates, highest concordance values, and overall concordance values demonstrated smaller, borderline significant correlations with the final scores (Spearman correlations = 0.75 [$P=0.06$], 0.77 [$P=0.06$], and 0.72 [$P=0.07$], respectively).

Discussion

We describe a systematic, explicit method of ranking quality measures using regularly updated clinical practice guidelines and prospectively collected performance data. Measures are ranked based on their potential to improve a prospectively defined outcome in a specified patient population, rather than on their ability to increase institutional concordance values. When applied to a comprehensive set of breast-cancer process-of-care measures, the highest-ranking measures recommend (1) chemotherapy for node-negative, hormone-receptor positive, tumors measuring 1.1-3 cm, (2) hormone therapy for node-positive, hormone-receptor positive tumors, (3) chemotherapy for node-positive, hormone-receptor positive tumors, (4) radiation therapy following breast-conserving surgery, and (5) hormone therapy for node-negative, hormone-receptor positive, tumors measuring 1.1-3 cm.

Higher-ranking measures tend to have more eligible patients and demonstrate a larger difference between the highest and overall concordance values. Sometimes measures with many eligible patients (#6) or many non-concordant patients (#14 and 16) do not rank highly, because they offer relatively limited potential for improvement. Since the rankings depend largely on the relative number of eligible patients per measure and this factor should be reasonably consistent across systems of care, the quality measures that rank highly in this analysis could have broad relevance beyond the institutions that provided performance data. However, additional studies should assess the reproducibility of the data used to assess concordance and the validity of the measures considered high priority by our analysis before these measures are implemented widely by other health care systems.

Treatments with few eligible patients rarely rank highly, in part because their corresponding measures cannot have a large number of non-concordant patients. This reinforces the need to appropriately scale quality measures to the population and organization being assessed.

Treatments that confer only modest improvements in outcomes for individual patients sometimes rank highly (#18). This occurs when many patients do not receive recommended treatments and benchmark concordance values are much greater than overall concordance values. The fact that such measures rank highly underscores the importance this approach places on improving the outcomes of a population rather than the outcomes of individuals.

Using the highest concordance achieved by an institution as the goal for all institutions is advantageous, because it defines a level of performance that is feasible and highlights circumstances where interventions may be more likely to work. However, this approach has its limitations. First, it fails to prioritize situations where care is universally non-concordant (i.e., all institutions perform below 100% and no institution demonstrates a significantly higher concordance). While this is a potential weakness of our approach, such situations do not necessarily represent areas where attention, and quality improvement resources, should be focused. There may be other explanations for consistently non-concordant care. Moreover, systematically identifying a realistic benchmark when all institutions exhibit the same level of care is difficult. Second, it inherently prioritizes situations for which there is substantial variability in performance from center to center. One could argue this often occurs when data are conflicting and experts disagree. However, deriving measures from consensus-based guidelines, as was done for this analysis, helps to minimize this risk.

Our approach to prioritizing quality measures relies on qualitative estimates of the benefits associated with treatments as determined by a survey of a relatively small group of expert breast cancer clinicians. We considered using the results of clinical trials to estimate these benefits, or to calculate the incremental quality-adjusted life years generated by treatments. However, the published data on breast cancer outcomes were too inconsistent to estimate these benefits reliably and consistently for each recommendation. Clinical trials rarely select the same outcomes (DFS, recurrence-free-survival, etc.), end-points (5 years, 10 years, etc.), or patient populations. Furthermore, the estimates provided by clinical trials often compare the outcomes associated with recommended treatments to the outcomes associated with experimental treatments, not the outcomes associated with common non-concordant treatments. Our priority was to use the same estimation method for each recommendation. The approach we chose is simple, practical, and reproducible. It is reassuring that we identified an association between magnitude-of-benefit estimates and DFS, and important to note that the final rankings are only modestly sensitive to the magnitude-of-benefit estimates.

Our goal was to prioritize quality measures based on their potential to improve DFS and QOL. Certainly, these are not the only outcomes that need to be considered. We realize our rankings would have been different if the goal had been different. For example, if we had prioritized measures based on their potential to improve overall survival, then some measures would have ranked lower (#6) and others would have ranked higher (# 22). Moreover, treatment effectiveness is not the only important component of health care quality that needs to be addressed. The Institute of Medicine considers patient safety, patient centeredness and timelines-of-care to be equally important aspects of health care quality.⁴⁶ While some of the measures included in our analysis, such as the 'over-use' measures, address these other components of quality, these 'over-use' measures were often not prioritized highly by our methodology. If one believes all components of quality should receive balanced attention, then it may be necessary to develop unique measures for each component of quality and prioritize them separately. Doing so, however, would be challenging because there are relatively few reliable measures and it is hard to define clear, quantifiable goals for these other aspects of health care quality.

While we used quality measures to help identify where potentially ameliorable gaps in quality of care exist, there are other applications for quality measures (e.g., public reporting, grading providers and paying-for-performance). The measures identified as high priority in our analysis may not be ideally suited for these other applications. Unfortunately, quality measures are frequently not tailored to the different purposes for which they are used or the groups to which they are applied. To make quality measurement more efficient and effective, one may have to develop unique measures for these different applications.

It is important to recognize that our prioritization methodology requires a comprehensive set of quality measures and an ability to estimate the impact recommended treatments have on outcomes. Unfortunately, it is not always possible to define an extensive set of measures or estimate the impact of treatments. Our approach also requires a detailed patterns-of-care database – a resource that may not be available in many centers. If non-NCCN centers exhibit different patterns of care than NCCN centers, then all institutions will have to repeat the analysis to identify their own, unique high priority quality measures. However, the resources required to do this could be prohibitive. Finally, this methodology does not preclude the need to reevaluate practice performance as clinical evidence, practice patterns, and quality measures change. The recommendation for chemotherapy in hormone-receptor-positive, node-negative, breast cancer was in line with the highest-level evidence when it was created, but emerging data now suggest chemotherapy may only benefit a subset of these patients. While the measure based on this recommendation (#18) ranked highly in this analysis, it might rank differently in the future, as evidence and practice patterns change.

A few organizations have described criteria for identifying where quality improvement efforts should focus their resources. In addition to those enumerated by the Institute of Medicine (discussed above)⁴³, authors have recommended considering impact on health, meaningfulness to consumers, potential for quality improvement, and susceptibility to influence by the health care system.⁴⁶ Some researchers have proposed selecting quality measures based on their clinical impact, reliability, feasibility, scientific acceptability, usefulness, and potential for improvement.^{10,47} Each set of criteria could be used to generate quality measures, and the last set has been used to identify several widely accepted measures. However, we are not aware of any previous efforts that use explicit criteria to prioritize a set of measures in a systematic way or that identify which measures are most likely to help achieve a particular outcome.

Several organizations have described quality measures for breast cancer.^{9,14,16,48-50} The National Quality Form recommended four: needle biopsy before excision, radiation therapy following breast conserving surgery for women under 70, combination chemotherapy within 60 days of surgery for hormone-receptor negative breast cancer > 1 cm, and axillary node dissection or sentinel node biopsy for stage I-IIb breast cancer.¹⁶ The RAND corporation endorsed three: offer modified radical mastectomy or breast-conserving surgery, radiation therapy within 6 weeks of surgery or chemotherapy for women who have breast conserving surgery, and adjuvant systemic therapy (combination chemotherapy and/or tamoxifen) for women over age 50 with positive nodes.¹⁴

These quality measures have limitations. Some are not supported by high-quality clinical evidence. Others do not clearly define a population of eligible patients or recommend a specific treatment. Several relate to aspects of care for which it is hard to identify a measurable process that a quality improvement program could target. Most importantly, all were selected as consensus measures by expert panels, without considering actual patterns-of-care data or impact on outcomes. While they overlap somewhat with the recommendations prioritized by our analysis, we identified several unique measures (e.g., #18 and 19). Moreover, some of the measures selected by other organizations and supported

by high-quality evidence did not rank near the top of our list (e.g., # 22 and 23), because few patients were eligible for these recommendations and there was not much room for improvement. All of the measures included in our analysis were derived from evidence and consensus-based clinical practice guidelines. Analyses performed by the NCCN pre hoc ensure the measures are feasible and reliable. Most importantly, the highest-ranking measures in our analysis identify clinical areas where practice performance is sub-optimal and a change in practice performance can substantially improve outcomes.

The systematic method of prioritizing quality measures that we describe represents a significant departure from previous efforts to identify priority areas for quality improvement. The methodology is simple and flexible, and could easily be applied to other practice settings, data sources, and diseases, or used it to rank measures across different diseases. The breast cancer quality measures that ranked highly in our analysis represent key leverage points that may have broad relevance beyond the institutions that contributed performance data. In conjunction with the NCCN, the American Society of Clinical Oncology used the results of our analysis to help select their breast cancer quality measures.¹¹ Widespread use of the methods described above could increase the efficiency and efficacy of quality improvement efforts and improve the outcomes of people who rely on our health care system.

Acknowledgments

This work was supported in part by grant P50 CA89393 from the National Cancer Institute to Dana-Farber Cancer Institute. Dr. Hassett received salary support from R25 CA092203. The sponsors had no direct influence on the design of the study, analysis of the data, interpretation of the results, or writing of the manuscript.

References

1. Basch P. Quality of health care delivered to adults in the United States. *New England Journal of Medicine*. 2003; 349:1866–8. comment. author reply 1866-8. [PubMed: 14606461]
2. Kohn, LT.; Corrigan, J.; Donaldson, MS. *To err is human: building a safer health system*. Washington, D.C.: National Academy Press; 2000.
3. McGlynn EA, Asch SM, Adams J, et al. The quality of health care delivered to adults in the United States. *New England Journal of Medicine*. 2003; 348:2635–45. see comment. [PubMed: 12826639]
4. Ray-Coquard I, Philip T, de Laroche G, et al. Persistence of medical change at implementation of clinical guidelines on medical practice: a controlled study in a cancer network. *Journal of Clinical Oncology*. 2005; 23:4414–23. [PubMed: 15994151]
5. Borenstein J, Badamgarav E, Henning JM, et al. The association between quality improvement activities performed by managed care organizations and quality of care. *American Journal of Medicine*. 2004; 117:297–304. see comment. [PubMed: 15336578]
6. Goldberg HI, Wagner EH, Fihn SD, et al. A randomized controlled trial of CQI teams and academic detailing: can they alter compliance with guidelines? *Joint Commission Journal on Quality Improvement*. 1998; 24:130–42. [PubMed: 9568553]
7. Grol R. Improving the quality of medical care: building bridges among professional pride, payer profit, and patient satisfaction. *JAMA*. 2001; 286:2578–85. see comment. [PubMed: 11722272]
8. Shojania KG, Grimshaw JM. Evidence-based quality improvement: the state of the science. *Health Affairs*. 2005; 24:138–50. [PubMed: 15647225]
9. Greenberg A, Angus H, Sullivan T, et al. Development of a set of strategy-based system-level cancer care performance indicators in Ontario, Canada. *International Journal for Quality in Health Care*. 2005; 17:107–14. [PubMed: 15665069]
10. McGlynn EA, Asch SM. Developing a clinical performance measure. *American Journal of Preventive Medicine*. 1998; 14:14–21. [PubMed: 9566932]
11. Network ASOCONCC. *ASCO/NCCN Quality Measures: Breast and Colorectal Cancers*. 2007

12. Winn RJ, National Comprehensive Cancer N. The NCCN guidelines development process and infrastructure. *Oncology (Huntington)*. 2000; 14:26–30.
13. Malin JL, Asch SM, Kerr EA, et al. Evaluating the quality of cancer care: development of cancer quality indicators for a global quality assessment tool. *Cancer*. 2000; 88:701–7. [PubMed: 10649266]
14. Asch, SM. Rand Corporation., United States. Agency for Healthcare Research and Quality. Quality of care for oncologic conditions and HIV: a review of the literature and quality indicators. Santa Monica, CA: Rand; p. 2000
15. Neuss MN, Desch CE, McNiff KK, et al. A process for measuring the quality of cancer care: the Quality Oncology Practice Initiative. *Journal of Clinical Oncology*. 2005; 23:6233–9. see comment. [PubMed: 16087948]
16. Forum, TNQ. Quality of Cancer Care Measures. Project Steering Committee Meeting; 2005;
17. Malin JL, Schneider EC, Epstein AM, et al. Results of the National Initiative for Cancer Care Quality: How Can We Improve the Quality of Cancer Care in the United States? *Journal of Clinical Oncology*. 2006; 24:626–634. [PubMed: 16401682]
18. Beck CA, Richard H, Tu JV, et al. Administrative Data Feedback for Effective Cardiac Treatment: AFFECT, a cluster randomized trial. *JAMA*. 2005; 294:309–17. see comment. [PubMed: 16030275]
19. Grossbart SR. What's the return? Assessing the effect of “pay-for-performance” initiatives on the quality of care delivery. *Medical Care Research & Review*. 2006; 63:29S–48S. see comment. [PubMed: 16688923]
20. Rosenthal MB, Fernandopulle R, Song HR, et al. Paying for quality: providers' incentives for quality improvement. *Health Affairs*. 2004; 23:127–41. see comment. [PubMed: 15046137]
21. Rosenthal MB, Frank RG, Li Z, et al. Early experience with pay-for-performance: from concept to practice. *JAMA*. 2005; 294:1788–93. see comment. [PubMed: 16219882]
22. Barr JK, Giannotti TE, Sofaer S, et al. Using public reports of patient satisfaction for hospital quality improvement. *Health Services Research*. 2006; 41:663–82. [PubMed: 16704506]
23. Hibbard JH, Stockard J, Tusler M. Hospital performance reports: impact on quality, market share, and reputation. *Health Affairs*. 2005; 24:1150–60. [PubMed: 16012155]
24. Werner RM, Asch DA. The unintended consequences of publicly reporting quality information. *JAMA*. 2005; 293:1239–44. see comment. [PubMed: 15755946]
25. Snyder C, Anderson G. Do quality improvement organizations improve the quality of hospital care for Medicare beneficiaries? *JAMA*. 2005; 293:2900–7. see comment. [PubMed: 15956635]
26. Harlan LC, Greene AL, Clegg LX, et al. Insurance status and the use of guideline therapy in the treatment of selected cancers. *Journal of Clinical Oncology*. 2005; 23:9079–88. see comment. [PubMed: 16301598]
27. Hewitt, ME.; Simone, JV. National Cancer Policy Board (US). Ensuring quality cancer care. Washington, D.C.: National Academy Press; 1999.
28. Moher D, Schachter HM, Mamaladze V, et al. Measuring the quality of breast cancer care in women. Evidence Report: Technology Assessment (Summary). :1–8. 2004.
29. Winn RJ, Botnick WZ, Brown NH. The NCCN guideline program--1998. *Oncology (Huntington)*. 1998; 12:30–4.
30. Winn RJ, Botnick W, Dozier N. The NCCN Guidelines Development Program. *Oncology (Huntington)*. 1996; 10:23–8.
31. Senn HJ, Thurlimann B, Goldhirsch A, et al. Comments on the St Gallen Consensus 2003 on the Primary Therapy of Early Breast Cancer. *Breast*. 2003; 12:569–82. [PubMed: 14659136]
32. Adjuvant Therapyfor Breast Cancer. NIH Consensus Statement. 2000 Nov 1-3; 17(4):1–35.
33. Malin JL, Schuster MA, Kahn KA, et al. Quality of breast cancer care: what do we know? *Journal of Clinical Oncology*. 2002; 20:4381–93. [PubMed: 12409338]
34. Du XL, Key CR, Osborne C, et al. Discrepancy between consensus recommendations and actual community use of adjuvant chemotherapy in women with breast cancer. *Annals of Internal Medicine*. 2003; 138:90–7. see comment erratum appears in *Ann Intern Med* 2003 Nov

- 18;139(10):873 summary for patients in *Ann Intern Med* 2003 Jan 21;138(2):I16; PMID: 12529113. [PubMed: 12529090]
35. Harlan LC, Abrams J, Warren JL, et al. Adjuvant therapy for breast cancer: practice patterns of community physicians. *Journal of Clinical Oncology*. 2002; 20:1809–17. [PubMed: 11919238]
 36. Palazzi M, De Tomasi D, D'Affronto C, et al. Are international guidelines for the prescription of adjuvant treatment for early breast cancer followed in clinical practice? Results of a population-based study on 1547 patients. *Tumori*. 2002; 88:503–6. [PubMed: 12597147]
 37. NCCN. NCCN Clinical Practice Guidelines in Oncology. v.1.2006. Jenkintown, PA: The National Comprehensive Cancer Network; 2006. pp www.nccn.org
 38. Weeks J, Niland J, Hughes ME, et al. Institutional Variation in Concordance with Guidelines for Breast Cancer Care in the National Comprehensive Cancer Network. Implications for Choice of Quality Indicators; AcademyHealth Annual Research Meeting; Seattle, WA. 2006;
 39. Weeks JC. Outcomes assessment in the NCCN. *Oncology (Huntington)*. 1997; 11:137–40.
 40. Niland JC. NCCN outcomes research database: data collection via the Internet. *Oncology (Huntington)*. 2000; 14:100–3.
 41. Weeks J. National Comprehensive Cancer Network. Outcomes assessment in the NCCN: 1998 update. *Oncology (Huntington)*. 1999; 13:69–71.
 42. Christian CK, Niland J, Edge SB, et al. A multi-institutional analysis of the socioeconomic determinants of breast reconstruction: a study of the National Comprehensive Cancer Network. *Annals of Surgery*. 2006; 243:241–9. [PubMed: 16432358]
 43. Adams K, Corrigan JM. Institute of Medicine (US). Priority areas for national action: transforming health care quality. Washington, D.C.: National Academies Press; 2003.
 44. Charlson ME, Pompei P, Ales KL, et al. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of Chronic Diseases*. 1987; 40:373–83. [PubMed: 3558716]
 45. Katz JN, Chang LC, Sangha O, et al. Can comorbidity be measured by questionnaire rather than medical record review? *Medical Care*. 1996; 34:73–84. [PubMed: 8551813]
 46. Hurtado MP, Swift EK, et al. Institute of Medicine (U.S.). Envisioning the national health care quality report. Washington, D.C.: National Academy Press; 2001. Committee on the National Quality Report on Health Care Delivery.
 47. Campbell SM, Braspenning J, Hutchinson A, et al. Research methods used in developing and applying quality indicators in primary care. *BMJ*. 2003; 326:816–9. [PubMed: 12689983]
 48. Schneider EC, Epstein AM, Malin JL, et al. Developing a system to assess the quality of cancer care: ASCO's national initiative on cancer care quality. *Journal of Clinical Oncology*. 2004; 22:2985–91. [PubMed: 15284249]
 49. Department of Health UK. Quality and Performance in the NHS Performance Indicators : July 2000. 2000
 50. Anonymous: Measuring the quality of breast cancer care in women. Evidence Report: Technology Assessment (Summary). 2004:1–8.

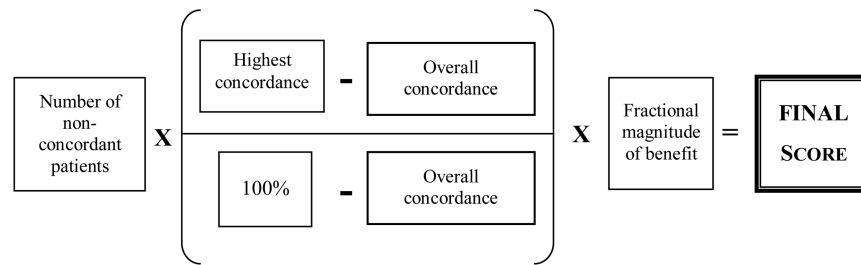


Figure 1. Algorithm Used to Generate Scores

Four attributes were incorporated into an algorithm that was used to generate scores for 30 quality measures. Quality measures were prioritized as targets for quality improvement based on their final scores. Three attributes – the number of non-concordant patients, highest concordance at any one institution, and overall concordance across all institutions – were derived from analyzing the treatments provided to 9019 women with breast cancer relative to the recommendations made by the National Comprehensive Cancer Network. One attribute – the fractional magnitude-of-benefit estimate – was derived by asking a panel of expert breast cancer clinicians to estimate the benefit experienced by patients who received a recommended treatment compared to patients who received a common non-concordant treatment.

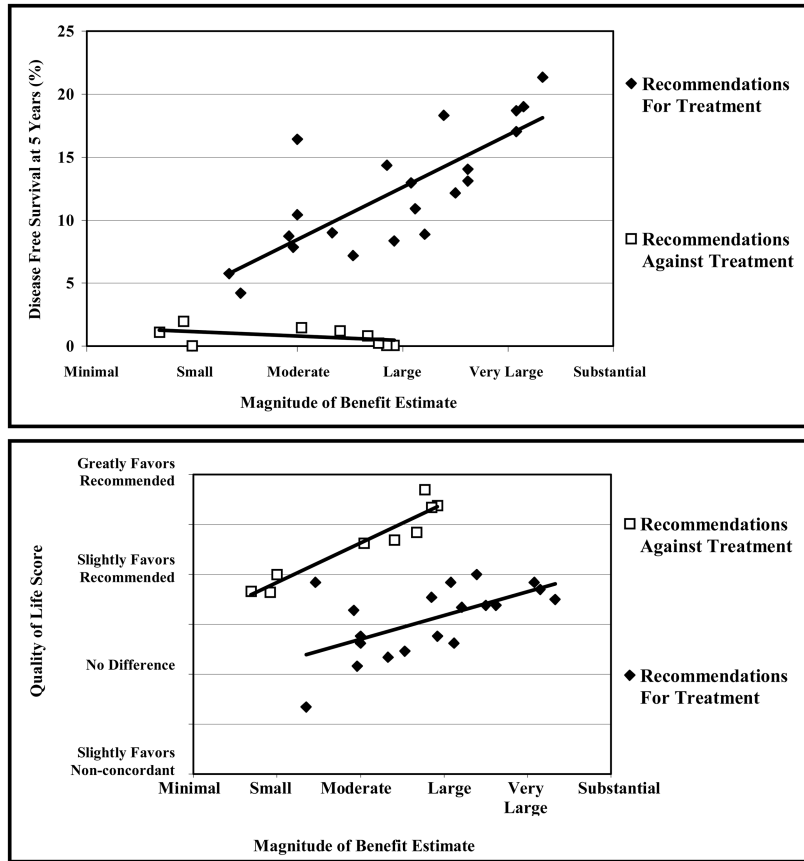


Figure 2. Five-year disease-free survival and quality-of-life scores versus magnitude-of-benefit estimates for each treatment recommendation

Recommendations for treatments (closed diamonds) are presented separately from recommendations against treatments (open squares). The estimated improvement in disease-free survival experienced by a population of patients receiving the recommended instead of a non-concordant treatment was reported as the percent absolute benefit at five years. The estimated difference in quality of life was reported as greatly favors recommended treatment, slightly favors recommended treatment, no difference, slightly favors non-concordant treatment, or greatly favors non-concordant treatment. Trend lines were generated using simple linear regression. Spearman correlation coefficients (r values) are presented with their corresponding P values.

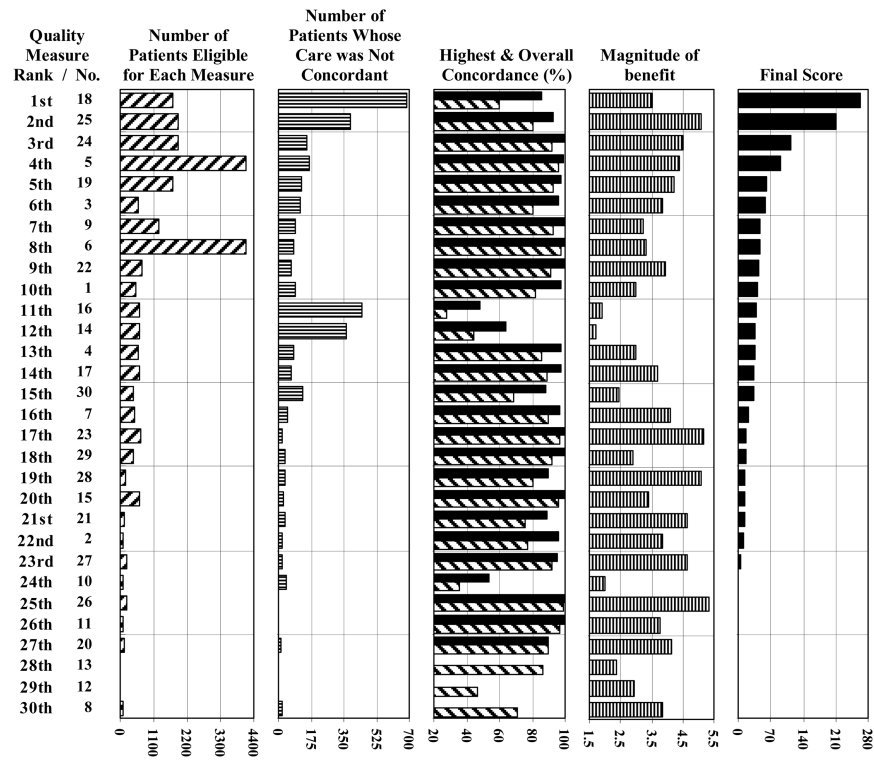


Figure 3. Comparison of values for each quality measure

Results are presented in parallel for each quality measure. Measures are listed in order from highest to lowest final score. The figure includes the measure number (as it appears in table 1), the total number of eligible patient encounters, the number of patients who received care that was not concordant with each treatment recommendation, the highest and overall percent concordance values (solid black and grey bars, respectively), the mean magnitude-of-benefit estimates, and the final scores. The values that appear in the second, third, and fourth graphs were used to calculate the final scores that appear in the fifth graph. Measures 13, 12 and 8 (ranked 28th, 29th and 30th, respectively) had fewer than 10 eligible patients at all centers, so reliable highest-concordance values could not be calculated (and therefore, final scores were not derived for these three measures).

Table 1

Quality Measures

Stage & Patient Group	Number	Recommended Treatment *	A Common Non-Concordant Treatment †	Number of Eligible Patients **	Overall Concordance
DCIS‡					
Widespread DCIS (in 2 or more quadrants)	1	Mastectomy	Breast conserving excision	509	82%
Unicentric DCIS treated with breast conserving surgery. Pathology shows low grade and negative margins.	2	No axillary lymph node surgery	Axillary lymph node surgery	84	77%
Unicentric DCIS treated with breast conserving surgery. Pathology shows negative margins, and intermediate-high grade or size 0.5 cm.	3	No axillary lymph node surgery	Axillary lymph node surgery	577	80%
	4	Whole breast radiation	No whole breast radiation	577	86%
Clinical Stage I or II Breast Cancer					
Treated with breast conserving surgery. Pathology shows negative margins.	5	Whole breast radiation	No radiation therapy	4,150	96%
	6	Axillary lymph node surgery	No axillary lymph node surgery	4,150	98%
Treated with mastectomy. Pathology shows 4 or more positive axillary nodes.	7	Chest wall radiation	No chest wall radiation	467	90%
Treated with mastectomy. Pathology shows tumor > 5cm or margins positive.	8	Chest wall radiation	No chest wall radiation	66	71%
Treated with mastectomy. Pathology shows tumor ≤ 5 cm, margins negative, & nodes negative.	9	No chest wall radiation	Chest wall radiation	1,275	93%
Treated with surgery. Pathology shows: tubular or colloid histology; nodes negative; and tumor ≤ 1 cm.	10	No adjuvant hormone therapy	Adjuvant hormone therapy §	67	36%
	11	No adjuvant chemotherapy	Adjuvant chemotherapy	67	97%
Treated with surgery. Pathology shows: tubular or colloid histology; nodes negative; and tumor > 3 cm.	12	Adjuvant hormone therapy	No adjuvant hormone therapy	15	47%
	13	Adjuvant chemotherapy	No adjuvant chemotherapy	15	87%
Treated with surgery.	14	No adjuvant hormone therapy	Adjuvant hormone therapy	647	44%

Stage & Patient Group	Number	Recommended Treatment *	A Common Non-Concordant Treatment †	Number of Eligible Patients **	Overall Concordance
Pathology shows: ductal, lobular, mixed, or metaplastic histology; nodes negative; and tumor 0.5 cm or microinvasive.	15	No adjuvant chemotherapy	Adjuvant chemotherapy	647	96%
Treated with surgery. Pathology shows: ductal, lobular, mixed, or metaplastic histology; nodes negative, tumor 0.6-1.0 cm, well differentiated, & no unfavorable features. ‡	16	No adjuvant hormone therapy	Adjuvant hormone therapy	624	28%
Treated with surgery. Pathology shows: ductal, lobular, mixed, or metaplastic histology; nodes negative; hormone receptors positive; and tumor 1.1 – 3 cm.	17	No adjuvant chemotherapy	Adjuvant chemotherapy	624	89%
Treated with surgery. Pathology shows: ductal, lobular, mixed, or metaplastic histology; nodes negative; hormone receptors positive; and tumor > 3 cm.	18	Adjuvant chemotherapy	No adjuvant chemotherapy	1,723	60%
Treated with surgery. Pathology shows: ductal, lobular, mixed, or metaplastic histology; nodes negative; hormone receptors positive; and tumor > 3 cm.	19	Adjuvant hormone therapy	No adjuvant hormone therapy	1,723	93%
Treated with surgery. Pathology shows: ductal, lobular, mixed, or metaplastic histology; nodes negative; hormone receptors positive; and tumor > 3 cm.	20	Adjuvant chemotherapy	No adjuvant chemotherapy	143	90%
Treated with surgery. Pathology shows: ductal, lobular, mixed, or metaplastic histology; nodes negative; hormone receptors positive; and tumor > 3 cm.	21	Adjuvant hormone therapy	No adjuvant hormone therapy	143	76%
Treated with surgery. Pathology shows: ductal, lobular, mixed, or metaplastic histology; nodes negative; hormone receptors positive; and tumor > 1 cm.	22	Adjuvant chemotherapy	No adjuvant chemotherapy	736	91%
Treated with surgery. Pathology shows: axillary nodes positive; hormone receptors negative.	23	Adjuvant chemotherapy	No adjuvant chemotherapy	662	97%
Treated with surgery. Pathology shows: axillary nodes positive; hormone receptors positive.	24	Adjuvant chemotherapy	No adjuvant chemotherapy	1,909	92%
Treated with surgery. Pathology shows: axillary nodes positive; hormone receptors positive.	25	Adjuvant hormone therapy	No adjuvant hormone therapy	1,909	80%
Clinical Stage III Breast Cancer					
Stage IIIA treated with mastectomy	26	Adjuvant Chemotherapy	No adjuvant chemotherapy	226	99%
Stage IIIA treated with mastectomy. Pathology shows: hormone receptors positive.	27	Radiation therapy [#]	No radiation therapy	226	92%
Stage IIIA treated with mastectomy. Pathology shows: hormone receptors positive.	28	Adjuvant hormone therapy	No adjuvant hormone therapy	175	80%
Stage IIIB	29	Anthracycline chemotherapy	Non-anthracycline chemotherapy	416	92%
	30	Neo-adjuvant chemotherapy	Adjuvant chemotherapy	416	69%

* The treatment recommended by the NCCN breast cancer clinical practice guidelines in 2003.

⁷ An alternative treatment that was provided to some patients but was not consistent with the recommendation made by the NCCN guidelines.

⁸ Ductal carcinoma in situ. All staging is based on the American Joint Committee on Cancer Staging Manual (6th edition).

⁹ Tamoxifen for women 50 or younger. Tamoxifen or an aromatase inhibitor for women older than 50.

// Cytotoxic chemotherapy excludes antibody therapy.

¹⁰ Unfavorable features: angio-lymphatic invasion, high nuclear grade, high histology grade, and HER-2 over-expression.

Radiation therapy to the chest wall and supra-clavicular areas.

^{**} The total number of patients whose care was stipulated by each measure, regardless of whether their treatment was concordant or non-concordant.

Table 2
Highest Ranking Quality Measures *

Treatment Recommendation & Patient Subgroup (Treatment recommendation #)	Quality Measure Rank & Final Score	Number of patients to whom each recommendation applies (% of 9,019 total patients)
Adjuvant chemotherapy For clinical stage I-II breast cancer treated with surgery. Pathology shows axillary nodes negative, hormone receptors positive, and tumor 1.1-3 cm (#18)	1 st – 264	1,723 (19%)
Adjuvant hormone therapy For clinical stage I-II breast cancer treated with surgery. Pathology shows axillary nodes positive, hormone receptors positive (#25)	2 nd – 210	1,909 (21%)
Adjuvant chemotherapy For clinical stage I-II breast cancer treated with surgery. Pathology shows axillary nodes positive, hormone receptors positive (#24)	3 rd – 114	1,909 (21%)
Radiation therapy For clinical stage I-II breast cancer treated with breast conserving surgery (#5)	4 th – 91	4,150 (46%)
Adjuvant hormone therapy For clinical stage I-II breast cancer treated with surgery. Pathology shows nodes negative, hormone receptors positive, and tumor 1.1-3 cm (#19)	5 th – 60	1,723 (19%)
Aggregate of the top 5 highest-ranking measures	NA	5,834 (65%)

* The total number of patients to whom the top 5 highest-ranking measures apply is less than the sum of the number of patients for each of the measures, because more than one measure can apply to each patient. For example, a patient could be eligible for a radiation therapy, chemotherapy, and hormonal therapy recommendation at the same time. NA = Not applicable.