Organization-based performance measures of cancer care quality: core measure development for breast cancer in Taiwan

K.-P. CHUNG, PhD, Graduate Institute of Health Care Organization Administration, National Taiwan University, and Center for Health Insurance Research, College of Public Health, National Taiwan University, Taiwan, M.-S. LAI, MD, PhD, Graduate Institute of Preventive Medicine, National Taiwan University, and Center for Health Insurance Research, College of Public Health, National Taiwan University, Taiwan, S.H. CHENG, MD, Department of Radiation Oncology, Koo Foundation Sun Yat-Sen Cancer Center, S.-T. TANG, PhD, Graduate School of Nursing, Chang-Gung University, C.C. HUANG, PhD, Department of Health Care Management, National Taipei College of Nursing, A.-L. CHENG, MD, PhD, Department of Internal Medicine, National Taiwan University Hospital, & P.-C. HSIEH, PhD, Department of Health Care Management, National Taipei College of Nursing, Taiwan


Organization-based performance measures of cancer care quality: core measure development for breast cancer in Taiwan

The purpose of the study was to develop organization-based core performance measures (CPMs) for breast cancer patients treated in hospitals that participated in cancer quality improvement programmes in Taiwan. CPMs were developed in three stages that included a preparation, a consensus building stage, and two stages of stakeholder feedback. Three criteria and seven subcriteria were applied in the development process. Indicators listed in a Delphi questionnaire were based on a literature search, indicators developed by relevant institutions and discussion by authors. Each indicator needed to meet inclusion criteria as a final indicator. Evidence-based guidelines, expert opinions from panel group, 27 hospitals and empirical data were all applied to develop and revise the core measures. Fifteen out of 28 indicators were selected and modified after the three stages. There were two pre-treatment indicators for screening and diagnosis, nine treatment-related indicators, and four monitoring-related indicators. Six indicators were supported by evidence level I, and four indicators by level II evidence. The CPMs for breast cancer can be developed systematically and be applied for internal quality improvement and external surveillance. Our experience can be extended to other cancer sites and adapted to link with pay for performance or certification program in cancer care.

Keywords: organization-based measures, core performance measures, breast cancer, the modified Delphi technique, cancer quality improvement programme.

INTRODUCTION

The quality of clinical practice for cancer care has been shown to vary even for patients with similar conditions (Guadagnoli et al. 1998; Schrag et al. 2001; Malin et al. 2002; Oliveria et al. 2004). The Institute of Medicine (IOM) report entitled ‘Ensuring Quality Cancer Care’ concluded that ‘for many Americans with cancer, there is a wide gulf between what could be construed as the ideal
and the reality of their experience with cancer care’ (IOM 1999, p. 211 lines 8–10). To improve the quality of cancer care, the IOM committee recommended a number of steps including the development of a core set of cancer care quality measures (IOM 1999).

Many previous researchers have discussed the core measures concept but results of these initiatives to date have been limited. The Joint Commission on Accreditation of Healthcare Organizations (JCAHO) in January 1997 initiated the ‘ORYX’ programme, designed to incorporate quality measurement and performance improvement systems into the accreditation process (Epstein 1998). Core measures are standardized sets of valid, reliable and evidence-based measures that have been developed in acute care. Examples include indicators developed for acute myocardial infarction, and heart failure, etc. (JCAHO 2003). Several studies have explored generic performance measures such as a standardized core measurement set (Eddy 1998), common measures (McGlynn 2003) and evidence-based guideline transferred indicators (Grol 2001), as well as disease-specific measures in cancer care such as core cancer measures (McGlynn & Malin 2002), core process and outcome measures (Lipscomb & Snyder 2002) and core performance measures (CPMs) (Schneider et al. 2004). There are some previous efforts to develop cancer care indicators that cover part of the care continuum (Malin et al. 2000; Spencer et al. 2003) or include a large number of measures with burdensome data collection (Mandelblatt et al. 1999). However, disease-specific, organization-based CPMs that cover the whole spectrum of cancer care are still lacking. In addition, indicators cannot simply be transferred between countries without an intermediate process to allow for variations in professional cultures and/or clinical practice (Marshall et al. 2003).

The purpose of this study is to develop organization-based performance measures for improving cancer care quality in Taiwan. The National Health Insurance (NHI) scheme, a single payer system, has covered over 96% of the population since 1995. The insurance programme provides comprehensive coverage from ambulatory care to inpatient care. It also covered catastrophic diseases such as cancers. Any hospital can provide care to cancer patients without undergoing certification for cancer care. Since 2001, the Bureau of Health Promotion (BHP) under the Department of Health initiated a quality improvement programme of cancer care. The BHP provided financial support for 17 hospitals to join the programme and set the rules of participation. Each hospital is required to provide cancer prevention, cancer screening, diagnosis and treatment of cancer, education and clinical audits, cancer research and a quality assessment system. The number of hospitals approved by the BHP increased from 17 to 27 in year 2004.

Breast cancer was chosen as the first target area for the following reasons. First, breast cancer has a highly ranked burden of disease in terms of its incidence rate, new cases, mortality rate, person-years of life lost and estimated national direct costs (Brown et al. 2001). Second, the extensive clinical research on breast cancer care provides a firm grounding for process–outcomes links. A solid foundation of scientific research provides an extremely favourable vantage point from which to begin to evaluate the quality of care for breast cancer (Schuster et al. 1998; Malin et al. 2002).

METHODS

Based on previous studies of performance indicator development in general (Rubin et al. 2001; Mainz 2003), in specific diseases (Campbell et al. 2002; JCAHO 2003), and in cancer care (Malin et al. 2000; Spencer et al. 2003), this study combined the Research and Development (RAND) appropriateness method (Brook et al. 1986) and core measure development approach from JCAHO-ORYX (JCAHO 2003) with modifications in order to develop cancer-specific core measures (Campbell et al. 2002). The cooperation whose name is RAND, which is a non-profit institution renown for helping improve policy and decision-making through research and analysis, and ORYX initiative represents one of the JCAHO’s first steps in focusing the accreditation process on key patient care, treatment and service issues. We developed organization-based CPMs for breast cancer in three stages: preparation stage, consensus building stage and two stages stakeholder feedback.

Preparation stage

Developing a list of quality indicators

We started to search the literature using the MEDLINE database and review references. The search terms used included the Medical Subject Heading terms ‘outcome and process assessment health care’, ‘quality indicator health care’, ‘quality of health care’, ‘practice guidelines’, ‘evidence-based medicine’, ‘breast neoplasms’ and the keyword ‘performance measures’ from January 1985 to January 2002. Those terms were based on reference (Malin et al. 2002) and our research purpose. We downloaded the abstracts and retrieved full text articles related to breast cancer care. In addition, we collected materials on indica-
tors and guidelines from well-known cancer-related organizations such as: the European Society of Surgical Oncology, National Health Service (NHS), RAND Corporation, IOM, Foundation for Accountability (FACCT), National Committee of Quality Assurance, National Cancer Institute, Agency for Healthcare Research and Quality, and the Scottish Intercollegiate Guidelines Network (SIGN) (McGlynn & Malin 2002). We collected a total of 150 indicators and classified them along the care continuum into pre-treatment (including screening and diagnosis), treatment-related and monitoring-related (including surveillance and outcome). In this study, we focus on effectiveness of care supported with evidence for organizational performance. We held six meetings from March to June 2003 and decided to retain eight pre-treatment, 11 treatment-related and four monitoring-related indicators. The reasons for deleting indicators included similarity among indicators, structural indicators, the lack of attribution to quality of cancer care in an organization-based setting, patient-centred indicators limited by data availability, and system differences between health insurance plans (e.g. waiting time is usually not a problem in Taiwan). In addition, we wished to keep the indicator list of reasonable length in order to obtain full cooperation in the rating process.

The format of the Delphi Technique questionnaire was adapted from previous studies (Nadzam et al. 1993; Rubin et al. 2001; Campbell et al. 2002; JCAHO 2003) and included for each indicator the indicator type, indicator name and definition, rationale for selection, data source, level of evidence, references and related guidelines, and recommendations (Table 1).

The level of evidence was determined for each reference according to the SIGN (1998) rating system as follows:

- Ia – Evidence obtained from meta-analysis of randomized controlled trials.
- Ib – Evidence obtained from at least one randomized controlled trial.
- IIa – Evidence obtained from at least one well-designed controlled study without randomization.
- IIb – Evidence obtained from at least one other type of well-designed quasi-experimental study.
- III – Evidence obtained from well-designed non-experimental descriptive studies, such as comparative studies, correlation studies and case studies.
- IV – Evidence obtained from expert committee reports or opinions and/or clinical experiences of respected authorities.

Table 1. Example of Delphi technique questionnaire

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>I Importance and necessity</td>
<td>1 The indicator suggests the importance of measuring breast cancer care</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 The indicator suggests the necessity of measuring breast cancer care</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>II Scientifically acceptable</td>
<td>1 The indicator has sufficient rationale (supported by clinical evidence)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 The indicator has clear explanation or definition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 The indicator can distinguish good and bad quality of breast cancer care</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>III Usability</td>
<td>1 The indicator is compelling within the decision-making framework of physician</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 The indicator’s result could help in facilitating a change in physician’s behaviour</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Recommendation:

MDT, multidisciplinary team; NHS, National Health Service.
Selection of evaluation criteria used to rate quality indicators

After searching the literature and discussion by authors, we adopted three criteria and seven subcriteria for evaluating potential quality indicators (McGlynn & Malin 2002; McGlynn 2003). The first criterion, importance and necessity, included two subcriteria. The second criterion, scientific acceptance, included three subcriteria. The third criterion, usability, included two subcriteria (Table 1). These criteria can be seen as a hierarchy for assessing measures. If a measure is not important, its other characteristics are less meaningful. If a measure is not scientifically acceptable, its results may be at risk for improper interpretations. If a measure is not usable, we probably do not care if it is feasible (McGlynn 2003).

Inclusion criteria

The expert panel members were asked to rate each of the indicators according to the seven subcriteria using a 5-point scale ranging from a score of 1 (totally disagree) to 5 (totally agree). We applied median equal or greater than 4 in 1–5 scales, and range equal or less than 2 as cut-off scores (Brook et al. 1986). In addition, another criterion for assessing the content validity of an item when using seven experts is to accept an item when the percentage with agree or totally agree is higher than 86% (Lynn 1986). Each indicator needed to attain the cut-off score for all seven evaluation subcriteria to be included as a final indicator (McGlynn 2003).

Assembly of an expert panel

In order to facilitate the full involvement of panel members, we limited the number of experts to seven and invited them to rate the indicators twice by mail and to attend a face-to-face meeting. The Taiwan Cooperative Oncology Group (TCOG) of the National Health Research Institute was in charge of developing and disseminating consensus guidelines regarding cancer diagnosis and treatment as well as cancer clinical guidelines (TCOG 2002). The committee of the TCOG was composed of experts from different professional associations. We bypassed the nomination process by professional societies and selected 15 experts directly from the TCOG expert committee and from various hospitals on the basis of their field of expertise, geographical location and time availability. After an introduction to the details of the research project, seven experts joined the Delphi panel including two general surgeons, one medical oncologist, one pathologist, one radiologist, one oncology nurse scholar and one public health expert with preventive medicine and breast cancer research experience. The professional backgrounds of the seven expert panel members and their affiliated organizations are listed in Appendix 1.

Consensus building stage: modified Delphi technique

The method for consensus building of our research is to adopt the modified Delphi technique because of being applied to some empirical studies (Normand et al. 1998; Campbell et al. 2002; Hermann et al. 2004). First of all, we mailed out the first round of questionnaires and related materials in mid-June, 2003. The expert panel rated the indicators and provided comments when they wished. In addition, they added two new indicators that included ‘side effects of treatment, including arm symptoms, breast symptoms, body image, sexual functioning, alopecia, and perspectives on the future’ and ‘pre-menopausal women receiving adjuvant chemotherapy with adequate dosages’.

In the face-to-face meeting (24 July 2003), we presented statistical results and related opinions for each indicator for discussion. Two authors presented the original and newly added indicators one by one to highlight the comments made and questions raised by the different experts. Eight indicators were modified with respect to their titles and definitions after discussion in the meeting and five new indicators were added to the core measure list. Finally, we modified the first questionnaire and mailed the second version to the panel for a second round of ratings.

Stakeholder feedback stage

This step was modified from the approaches of JCAHO to solicited input from a wide variety of stakeholders for core measures development for hospital and from several references to perform preliminary test in developing and implementing a clinical performance measure (Rubin et al. 2001; Campbell et al. 2002; JCAHO 2003). Hospitals and clinical professionals are major stakeholders in our research. Following the modified Delphi technique, a draft of the core measures set was created. This draft was presented at a meeting on 22 November 2003. In addition to introducing our development method for creating core measures, we presented the results of our efforts as well as half-year indicator data to experts from 17 hospitals and invited their comments. The second stage stakeholder feedback was held at 5 April year 2005 to modify CPMs specifically from empirical results. It is designed to
provide national level of panel data (year 2002) to experts from 27 hospitals and discuss the empirical indicator results. It is not just to obtain consensus from expert and hospitals, but more specific opinions could also add to refine the definition and data collection for numerator, denominator, as well as rationale of indicators with results.

RESULTS

The response rate was 100% for both rounds of expert ratings but the attendance rate at the face-to-face meeting was less than 40 percent. After the modified Delphi technique, 17 out of 28 indicators passed the evaluation criteria. Those selected included two pre-treatment indicators, nine treatment-related indicators and six monitoring-related indicators (Table 2).

One pre-treatment indicator, ‘Proportion of women aged over 50 years who received bilateral mammography three months before surgery’, was highly rated with all 100% either agreeing or strongly agreeing with the seven subcriteria. Two out of six treatment-related indicators were even more highly rated, with all raters strongly agreeing with all seven subcriteria. These treatment indicators included the ‘Proportion of invasive breast cancer patients with oestrogen receptor analysis results in the medical record’ and the ‘Proportion of pre-menopausal breast cancer women with positive lymph nodes receiving adjuvant chemotherapy’. One of the four highly rated monitoring-related indicators – the ‘annual mammography rate for breast cancer patients after treatment’ – also had unanimous strong agreement with all seven subcriteria.

Eleven indicators were excluded when rated according to these evaluation subcriteria as well as inclusion criteria (Table 3).

Three indicators were excluded for not meeting one of the seven subcriteria: they either lacked sufficient rationale and supporting clinical evidence according to some of the raters, or lacked a clear explanation or definition. Nine of the 11 indicators failed due to lack of agreement regarding their sufficient rationale and supporting clinical evidence. In general, the subcriteria regarding scientific acceptability were the major reasons for indicator exclusion. Specifically, most of the indicators were excluded due to their failure to reach the threshold of 86% of the raters in agreement or strong agreement with the subcriteria.

In a subsequent meeting (22 November 2003) with 50 representatives from 17 hospitals, we presented our indicators, including the development process and the preliminary results. Some concerns about the definitions were discussed; others mentioned the importance of including indicators that had been deleted. Two additional monitoring-related indicators were excluded because of the difficulties anticipated in data collection (Table 3). This decision was based on discussions with experts from participating hospitals and site visits to four hospitals. Ninety-one medical professionals and related persons from 27 hospitals joined the second stage stakeholder feedback (5 April 2005). We provided information on 11 indicators (two pre-treatment and nine treatment-related indicators) for year 2002 including the operational definition, its rationale and the indicator results for discussion. After the discussion, seven indicators were modified either by the title, rationale of indicator or operational definitions. As a result of all of these efforts, 15 indicators were selected and revised as organization-based core measures for breast cancer in Taiwan (Table 4).

DISCUSSION

Methodological issues

Panel size and assembly of a panel

In the RAND appropriateness method, they applied a 9-point scale and therefore needed to have at least nine expert raters (Hicks 1994). Since the Delphi method is heavily dependent upon the sample having the time to commit to the process, it is also important that those who agree to participate maintain involvement until the process is complete (Buck et al. 1993). In summary, the range of panel size should be between 7 and 15 (Brook 2001). In this pilot study, we used a 5-point scale instead of nine categories out of concern for the number of expert panelists that would be needed. Cooperation also needs to be secured from experts with the required range of professional backgrounds. We decided to invite seven experts and expected their full involvement. The 100% response rate for two rounds of the Delphi process provided validation for our approach.

Evaluation criteria and establishing threshold scores

There are no published guidelines that we are aware of on the selection of quality indicators. Core measures are standardized performance measures for specific diseases that cover the processes and outcomes of care and that should meet established evaluation criteria (McGlynn 2003). Since our research focus was to develop core cancer measures, we needed to have evaluation criteria more than validity and feasibility criteria (Spencer et al. 2003) that could help to keep the core measure set parsimonious.
<table>
<thead>
<tr>
<th>Indicators</th>
<th>Evidence level</th>
<th>Selection criteria and items</th>
<th>I-1</th>
<th>I-2</th>
<th>II-1</th>
<th>II-2</th>
<th>II-3</th>
<th>III-1</th>
<th>III-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion of women aged over 50 who received bilateral mammography 3 months before surgery</td>
<td>III [USPSTF 2002]</td>
<td></td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
</tr>
<tr>
<td>Proportion of breast cancer patients who have diagnosis in cytology and histology before surgery</td>
<td>III [Blichert-Toft et al. 1997; SIGN 1998]</td>
<td></td>
<td>5 ± 1.0</td>
<td>5 ± 1.0</td>
<td>4 ± 1.0</td>
<td>4 ± 1.0</td>
<td>5 ± 1.0</td>
<td>5 ± 1.0</td>
<td>5 ± 1.0</td>
</tr>
<tr>
<td>Treatment-related</td>
<td></td>
<td></td>
<td>2, 86</td>
<td>2, 86</td>
<td>2, 86</td>
<td>1, 100</td>
<td>2, 86</td>
<td>2, 86</td>
<td>2, 86</td>
</tr>
<tr>
<td>Proportion of breast cancer patients who had been discussed by multi-disciplinary team</td>
<td>III [Blichert-Toft et al. 1997; SIGN 1998]</td>
<td></td>
<td>5 ± 1.0</td>
<td>5 ± 1.0</td>
<td>4 ± 1.0</td>
<td>4 ± 1.0</td>
<td>5 ± 1.0</td>
<td>5 ± 1.0</td>
<td>5 ± 1.0</td>
</tr>
<tr>
<td>Proportion of zero-stage breast cancer patients with 10 or more lymph nodes on pathology report</td>
<td>III [SIGN 1998]</td>
<td></td>
<td>5 ± 1.0</td>
<td>5 ± 1.0</td>
<td>5 ± 1.0</td>
<td>5 ± 1.0</td>
<td>5 ± 0.0</td>
<td>5 ± 1.0</td>
<td>5 ± 1.0</td>
</tr>
<tr>
<td>Proportion of stage I and II patients who undergo breast-conserving surgery (BCS)</td>
<td>IIIb [Nettinger et al. 2000; Fisher et al. 2002]</td>
<td></td>
<td>5 ± 1.0</td>
<td>5 ± 1.0</td>
<td>5 ± 1.0</td>
<td>5 ± 1.0</td>
<td>5 ± 0.0</td>
<td>5 ± 1.0</td>
<td>5 ± 1.0</td>
</tr>
<tr>
<td>Proportion of breast cancer patient with pathology report of tumour-size in the medical record after surgery</td>
<td>IIIb [AJCC 2002]</td>
<td></td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
<td>5 ± 1.0</td>
<td>5 ± 0.0</td>
<td>5 ± 1.0</td>
<td>5 ± 1.0</td>
<td>5 ± 0.0</td>
</tr>
<tr>
<td>Proportion of invasive breast cancer after surgery with 10 or more lymph nodes removed on pathology report</td>
<td>IIIb [Blichert-Toft et al. 1997; SIGN 1998]</td>
<td></td>
<td>5 ± 1.0</td>
<td>5 ± 1.0</td>
<td>5 ± 1.0</td>
<td>5 ± 1.0</td>
<td>5 ± 1.0</td>
<td>5 ± 1.0</td>
<td>5 ± 1.0</td>
</tr>
<tr>
<td>Proportion of invasive breast cancer patients with oestrogen receptor analysis results in the medical record</td>
<td>IIIb [EBCTCG 1995, Gudagnoli et al. 1998]</td>
<td></td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
</tr>
<tr>
<td>Proportion of patients with invasive cancer who receive radiation treatment after BCS</td>
<td>IIIa [EBCTCG 2000; NIHDC 2001]</td>
<td></td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
<td>5 ± 1.0</td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
</tr>
<tr>
<td>Proportion of pre-menopausal breast cancer women with positive lymph node receiving adjuvant chemotherapy</td>
<td>IIIa [EBCTCG 1998a; NIHDC 2001]</td>
<td></td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
</tr>
<tr>
<td>Proportion of post-menopausal breast cancer women with positive lymph node receiving adjuvant hormone therapy or chemotherapy</td>
<td>IIIa [EBCTCG 1998b, 2000]</td>
<td></td>
<td>5 ± 1.0</td>
<td>5 ± 1.0</td>
<td>5 ± 1.0</td>
<td>5 ± 1.0</td>
<td>5 ± 1.0</td>
<td>5 ± 1.0</td>
<td>5 ± 1.0</td>
</tr>
<tr>
<td>Monitoring-related</td>
<td></td>
<td></td>
<td>2, 86</td>
<td>2, 86</td>
<td>2, 86</td>
<td>1, 100</td>
<td>2, 86</td>
<td>2, 86</td>
<td>2, 86</td>
</tr>
<tr>
<td>Annual mammography rate for breast cancer patients after treatment</td>
<td>IIIb [Smith et al. 1999]</td>
<td></td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
</tr>
<tr>
<td>Five-year local recurrence rate after surgery</td>
<td>III [Blichert-Toft et al. 1997]</td>
<td></td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
</tr>
<tr>
<td>Five-year disease-free survival</td>
<td>III [NHSE 2002]</td>
<td></td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
<td>5 ± 1.0</td>
<td>5 ± 0.0</td>
<td>5 ± 1.0</td>
<td>5 ± 1.0</td>
<td>5 ± 1.0</td>
</tr>
<tr>
<td>Five-year overall survival rate</td>
<td>III [FACCT 1997; NHSE 2002]</td>
<td></td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
<td>5 ± 0.0</td>
<td>5 ± 1.0</td>
</tr>
<tr>
<td>Newly added indicators</td>
<td></td>
<td></td>
<td>1, 100</td>
<td>1, 100</td>
<td>1, 100</td>
<td>1, 100</td>
<td>1, 100</td>
<td>1, 100</td>
<td>1, 100</td>
</tr>
<tr>
<td>Arm oedema rate for breast conservation surgery patient after radiation therapy</td>
<td>IV</td>
<td></td>
<td>4 ± 1.0</td>
<td>4 ± 1.0</td>
<td>4 ± 1.0</td>
<td>4 ± 1.0</td>
<td>4 ± 0.0</td>
<td>4 ± 1.0</td>
<td>4 ± 1.0</td>
</tr>
<tr>
<td>Local recurrence rate for breast conservation surgery patient after radiation therapy</td>
<td>IV</td>
<td></td>
<td>5 ± 1.0</td>
<td>5 ± 1.0</td>
<td>5 ± 1.0</td>
<td>5 ± 1.0</td>
<td>5 ± 1.0</td>
<td>5 ± 1.0</td>
<td>5 ± 1.0</td>
</tr>
</tbody>
</table>

*IQR, interquartile range.
†Range, difference of maximum and minimum score.
‡Agreement, the proportion of raters responding agree [4] and strongly agree [5].
Table 3. Excluded indicators (shaded values indicate failure to meet threshold)

<table>
<thead>
<tr>
<th>Indicator category and title</th>
<th>Source</th>
<th>Selecting Criteria and Items [median ± IQR*, range†, agreement [%]‡]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interval between receipt of abnormal screening result and the first visit to a physician</td>
<td>IOM [1999]</td>
<td>I-1 4 ± 1.0, 2, 71 I-2 4 ± 1.0, 2, 57 II-1 4 ± 1.0, 1, 57 II-2 4 ± 0.0, 1, 86 III-1 3 ± 2.0, 2, 43 III-2 3 ± 1.0, 1, 71 3 ± 1.0, 2, 29</td>
</tr>
<tr>
<td>Rate of confirmation follow-up</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recall rate</td>
<td>FACCT [1997]</td>
<td>I-1 3 ± 0.0, 2, 14 I-2 3 ± 1.0, 1, 0 II-1 3 ± 0.0, 1, 0 II-2 4 ± 2.0, 3, 57 III-1 3 ± 1.0, 2, 14 III-2 3 ± 2.0, 2, 29 3 ± 1.0, 1, 0</td>
</tr>
<tr>
<td>Proportion or number of patients whose breast cancer was detected at stage 0 or I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion of diagnosis informed by the attending physicians</td>
<td>NHSE [2002]</td>
<td>I-1 5 ± 1.0, 1, 100 I-2 4 ± 1.0, 1, 100 II-1 5 ± 2.0, 2, 71 II-2 5 ± 1.0, 1, 100 III-1 5 ± 1.0, 1, 100 III-2 5 ± 1.0, 2, 86</td>
</tr>
<tr>
<td>Satisfaction rate of diagnosis informing process</td>
<td>NHSE [2002]</td>
<td>I-1 4 ± 2.0, 2, 57 I-2 4 ± 2.0, 2, 71 II-1 4 ± 2.0, 3, 71 II-2 4 ± 2.0, 2, 71 III-1 4 ± 2.0, 2, 71 III-2 4 ± 1.0, 2, 86</td>
</tr>
<tr>
<td>Proportion of receiving surgery within 21 days from the time of first surgical outpatient department</td>
<td>USPSTF [2002]</td>
<td>I-1 4 ± 0.0, 1, 86 I-2 4 ± 0.0, 1, 86 II-1 4 ± 1.0, 1, 57 II-2 4 ± 0.0, 2, 86 III-1 4 ± 1.0, 1, 71 III-2 4 ± 0.0, 1, 86 4 ± 1.0, 1, 71</td>
</tr>
<tr>
<td>Proportion of post-surgery chemotherapy given by medical oncologist</td>
<td>Blichert-Toft et al. [1997]</td>
<td>I-1 4 ± 0.0, 2, 86 I-2 4 ± 1.0, 2, 86 II-1 4 ± 1.25, 2, 71 II-2 5 ± 1.0, 2, 86 III-1 4 ± 1.0, 2, 86 III-2 4 ± 1.0, 2, 86</td>
</tr>
<tr>
<td>Pre-menopausal women receiving adjuvant chemotherapy with adequate dosages</td>
<td>NIHCD [2001]</td>
<td>I-1 5 ± 1.0, 1, 100 I-2 5 ± 1.0, 1, 100 II-1 4 ± 1.0, 1, 100 II-2 4 ± 2.0, 2, 71 III-1 5 ± 1.0, 1, 100 III-2 5 ± 1.0, 1, 100 5 ± 1.0, 1, 100</td>
</tr>
<tr>
<td>Proportion of new breast cancer patients treated by MDT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion of severe complications after radiotherapy of patient with breast-conserving surgery</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

I-1, the indicator suggests the importance of measuring breast cancer care; I-2, the indicator suggests the necessity of measuring breast cancer care; II-1, the indicator has sufficient rationale (supporting by clinical evidence); II-2, the indicator has clear explanation or definition; II-3, the indicator can distinguish good and bad quality of breast cancer care; III-1, the indicator is compelling within the decision-making framework of physician; III-2, the indicator’s result could help in facilitating a change in physician’s behaviour.

MDT, multidisciplinary team.

*IQR, interquartile range.
†Range, difference between maximum and minimum score.
‡Agreement, the proportion of responding raters who agreed [4] and strongly agreed [5].

The following indicators were excluded because of perceived difficulties in data collection: arm oedema rate for breast conservation surgery patient after radiation therapy, local recurrence rate for breast conservation surgery patient after radiation therapy.
<table>
<thead>
<tr>
<th>Indicator name</th>
<th>Indicator identification and rationale</th>
</tr>
</thead>
</table>
| Proportion of women aged over 50 who received bilateral mammography 3 months before surgery | Numerator: the number of breast cancer patients aged over 50 who receive bilateral mammography 3 months before surgery  
Denominator: the number of breast cancer patients aged over 50 who have undergone surgery  
Rationale: avoid the risk of contralateral breast cancer |
| Proportion of breast cancer patients who have diagnosis in cytology and histology before surgery | Numerator: the number of breast cancer patients who at least 1 day before surgery have diagnosis by cytology and histology (exclude: metastasis breast cancer and persons undergoing surgery and diagnosis during the same day)  
Denominator: the number of patients who have undergone surgery  
Rationale: avoid same day surgery and have patient’s fully understanding and consent |
| Proportion of breast cancer patients who were discussed by multidisciplinary team | Numerator: the number of patients discussed by multidisciplinary team  
Denominator: the number of new patients who receive treatment  
Rationale: ensure appropriate care and better outcome |
| Proportion of zero-stage breast cancer patients with 10 or more lymph nodes on pathology report | Numerator: the number of zero-stage breast cancer patients with 10 or more axillary lymph nodes removed according to pathology reports  
Denominator: the number of zero-stage breast cancer patients with tumour size less than 2.5 cm (exclude: the patients who receive chemotherapy or radiation therapy before surgery)  
Rationale: avoid overuse or unnecessary care |
| Proportion of stage I and II patients who undergo breast-conserving surgery [BCS] | Numerator: the number of stage I and II breast cancer patients with BCS  
Denominator: the number of stage I and II breast cancer patients  
Rationale: indicate whether patients are involved in decision-making regarding the treatment plan |
| Proportion of breast cancer patients with pathology report of tumour-size in the medical record after surgery | Numerator: the number of breast cancer patients with pathology reports indicating tumour-size in the medical record after surgery  
Denominator: the number of breast cancer pathology reports  
Rationale: provide stage information, and evidence for treatment and subsequent follow-up |
| Proportion of invasive breast cancer after surgery with 10 or more lymph nodes removed on pathology report | Numerator: the number of breast cancer patients after modified radical mastectomy, radical mastectomy or BCS Combine axillary clearance with 10 or more axillary-lymph node removed on pathology reports  
Denominator: the number of breast cancer patients who have undergone modified radical mastectomy, radical mastectomy or breast-conserving surgery combine axillary clearance (exclude: the patients receiving chemotherapy or radiation therapy before surgery and Sentinel node biopsy)  
Rationale: avoid under-stage and subsequent under-treatment of patients |
Proportion of invasive breast cancer patients with oestrogen receptor analysis results in the medical record

Numerator: the number of female patients with invasive breast cancer who, after initial biopsy or resection, have oestrogen receptor analysis results in the medical record.
Denominator: the number of female patients with invasive breast cancer who have undergone biopsy or resection.
Rationale: indicate degree of tumour differentiation and provide evidence for future treatment plan.

Proportion of patients with invasive cancer who receive radiation treatment after BCS

Numerator: the number of invasive breast cancer patients after breast-conserving surgery with radiotherapy.
Denominator: the number of invasive breast cancer patients after breast-conserving surgery with radiotherapy.
Rationale: reduce local recurrence.

Proportion of breast cancer women aged less than and equal to 50 years (pre-menopausal) with positive lymph node receiving adjuvant chemotherapy

Numerator: the number of cancer women aged less than and equal to 50 years (pre-menopausal breast) with positive lymph nodes receiving adjuvant chemotherapy.
Denominator: the number of breast cancer women aged less than and equal to 50 years (pre-menopausal) with positive lymph nodes (exclude: male patients and metastasis patients).
Rationale: avoid distal metastasis.

Proportion of breast cancer women aged greater than 50 years (post-menopausal) with positive lymph node receiving adjuvant hormone therapy or chemotherapy

Numerator: the number of breast cancer women aged greater than 50 years (post-menopausal) with positive lymph nodes receiving adjuvant hormone therapy or chemotherapy.
Denominator: the number of breast cancer women aged greater than 50 years (post-menopausal) with positive lymph nodes.
Rationale: avoid distal metastasis.

Annual mammography rate for breast cancer patients after treatment

Numerator: the number of breast cancer patients who have mammography annually.
Denominator: the number of breast cancer patients (exclude: patients with excisions on two sides of the breast).
Rationale: detect ipsilateral or contralateral breast cancer.

Five-year local recurrence rate after surgery

Numerator: the number of breast cancer patients who have local recurrence after surgery within 5 years.
Denominator: the number of breast cancer patients who receive surgical operations within 5 years by stage.
Rationale: evaluate treatment effectiveness.

Five-year disease-free survival rate

Numerator: the number of breast cancer patients who survived and are disease-free in 5 years by stage.
Denominator: the number of breast cancer patients in 5 years by stage.
Rationale: evaluate treatment effectiveness.

Five-year overall survival rate

Numerator: the number of breast cancer patients who survived in 5 years by stage.
Denominator: the number of breast cancer patients in 5 years by stage.
Rationale: evaluate treatment effectiveness.
[McGlynn & Malin 2002]. This study included three criteria: importance and necessity, scientific acceptability, and usability [McGlynn 2003]. Areas for measurement are important when they link directly to a national goal of quality improvement, they represent a significant leverage point for achieving the goal, and/or there is considerable variation in the quality of care provided [JCAHO 2003; McGlynn 2003]. The study working group added the necessity criteria based on previous research [Kahan et al. 1994]. They considered care necessary if there is a reasonable chance of a non-trivial benefit to the patient and if it would be improper not to provide care.

There are three subcriteria under the scientific acceptability criterion. Several studies have applied an evidence-based criterion, mandating that indicators should have good evidence documenting the relationship between process of care or use of services and health outcomes [Mandelblatt et al. 1999; Rubin et al. 2001; Spencer et al. 2003]. Our second subcriterion is similar to ‘the measures should be clearly defined and operationalized’ [Mandelblatt et al. 1999]. The third subcriterion is similar to the validity criterion that is applied in other studies [Rubin et al. 2001; Spencer et al. 2003]. A measure is considered valid if the scores that are produced from the measure distinguish between good and bad quality and the construction of the measure adequately represents the concept of interest [McGlynn 2003].

The RAND appropriateness method has been shown to be highly reproducible [Shekelle et al. 1998]. In the RAND approach, indicators were considered valid if the median panel rating was equal to or greater than 7 [in a 9-point scale] and there was no evidence of panel disagreement [defined to occur when at least one-third of the panelists rated at least one of the criteria in the 1–3 or 7–9 group of points on the scale] [Brook et al. 1986]. In this study, the requirement for a threshold score of 86% agreement removed more indicators from subcriteria consideration than use of the median and range [modification from disagreement] for establishing cut-off scores. With a panel of seven experts, it is necessary to have at least 86% agreement on endorsing each specific subcriterion in order to establish content validity beyond the 0.05 level of significance [Lynn 1986]. When using a 5-point scale and seven experts, this threshold may be better than using the median and the range of scores if a simpler decision rule is desirable.

**Face-to-face discussion**

There is no opportunity for participants to elaborate on their views in the traditional Delphi technique [Hicks 1994]. One key feature of the modified Delphi method is the face-to-face discussion among experts [Campbell et al. 2002]. Three experts joined the face-to-face meeting that was facilitated by two authors. The low attendance rate justified our concern that it would be difficult to have full attendance if there were even more panelists. Written opinions from experts unable to attend were available as supplemental materials for discussion. During the meeting, one author with a medical and academic background introduced the indicators one by one and led the discussion. Emphasis was placed on the statistical results from the first round and written comments, focusing specifically on any areas of disagreement.

Because this meeting was just part of our method, these discussion materials as well as the second round questionnaire were sent to the expert panel for the second round of ratings. Since most of these experts had clinical and academic backgrounds, this result should be acceptable as a basis for thorough discussion and consensus building within the medical and academic communities.

Overall, the approach to developing CPMs involves asking first what needs to be measured, and then ascertaining whether sufficient data are available to answer the questions posed [Loeb 2004]. One limitation is the lack of evidence on the reliability of the Delphi method: if the same information were given to two or more panels, would the same results be obtained [Hasson et al. 2000]? The use of participants who have knowledge and an interest in the topic may help to increase the content validity of the Delphi technique [Goodman 1987], and the use of successive rounds of the questionnaire helps to increase the concurrent validity. All these concerns were considered by experts with professional backgrounds, face-to-face discussion and two rounds of ratings in the modified Delphi technique.

**Two stages stakeholder feedback**

Two stages stakeholder feedback served the purpose to rectify potential weakness in the modified Delphi techniques with limited size of an expert panel in our study. The critical challenge is to select measures that are compelling to stakeholders. This is a necessary first step if quality is to routinely taken into account in making decisions by key actors in the healthcare system [McGlynn 2003]. The first stage stakeholder feedback included 50 experts, mostly clinical professionals, from 17 hospitals. Ten more hospitals were approved to join the programme at the end of year 2004. In the second stage, 91 representatives attend the meeting that included clinical professionals, case managers, cancer
registry staffs, etc. The small group design with longer duration could facilitate better discussion. Opinions in pros and cons would be more specific to the point based on information from 10 hospitals in year 2002 in the discussion session [Rubin et al. 2001]. These data were reported from 10 out of 17 hospitals voluntarily. The accuracy and process of data quality control will not be covered in the article in detail.

The findings

Our final core measures number for breast cancer are consistent with the experience of JCAHO-ORYX regarding acute care measures [McGlynn 2003]. Our final core set included six level I indicators, four level II indicators and five level III indicators (Table 2). Five level I indicators are process indicators in the treatment-related category. The proportion (40%) based on level I evidence [i.e. based on at least one randomized controlled trial] was higher than in a previous effort to develop cancer quality indicators [Malin et al. 2000]. Three of our indicators based on level I evidence were similar to the recommended quality indicators for breast cancer in the RAND study [Research & Development 2000]. The other seven indicators for screening, diagnosis and follow-up developed by RAND with only level IV evidence [expert opinions] were not included in our final indicators list.

Three of our indicators were the same as indicators developed by the FACCT (The Foundation for Accountability 1997). Their other indicators covered screening, diagnosis and patient satisfaction and were either not included even at the first stage of our effort, or are included in the deleted indicators listed in Table 3. Five of our indicators were similar to those included in a recent review [Malin et al. 2002] and at least three of our indicators are similar to the indicators developed by the American Society of Clinical Oncology, and the National Initiatives on Cancer Care Quality [NICCQ] [Schneider et al. 2004]. Our treatment-related indicators are supported by RAND, FACCT, NHS, SIGN, NICCQ and other references [Blichert-Toft et al. 1997; FACCT 1997; Schuster et al. 1998; Research & Development 2000; Malin et al. 2002; NHSE 2002; Schneider et al. 2004]. Our monitoring-related indicators are supported by RAND, FACCT, NHS and other references [FACCT 1997; Mandelblatt et al. 1999; RAND 2000; NHSE 2002]. Overall, these findings are consistent with previous research [Schuster et al. 1998; IOM 1999; Mandelblatt et al. 1999; Malin et al. 2002; Schneider et al. 2004].

Several indicators deserve more discussion. There is fairly strong evidence that a multidisciplinary service is likely to provide better care, and that multidisciplinary care is associated with better 5-year survival [SIGN 1998]. The core breast team should include a designated breast surgeon, oncologist, radiologist, pathologist, breast care nurse(s), coordinator and team secretary [NHSE 2002]. A multidisciplinary team (MDT) can work together to obtain the diagnosis, treatment options and cross-specialty/department referral for the patient to ensure an optimal treatment plan [Whelan et al. 2006]. This indicator could be controversial in terms of the definition of MDT. The definition of MDT functions can range from combined conferences and group discussions, to informal opinion exchanges regarding the patient’s diagnosis, treatment and/or referral. These discussions can be conducted concurrently as care is provided and/or prospectively or retrospectively. In Taiwan, the adoption of MDT for treating cancer patients is just beginning. We need to be careful in interpreting the results according to the suggestion of the second stage stakeholder feedback. It should be accompanied with meeting frequency, discussion contents and composition of the MDT to reflect its functional effectiveness.

The proportion of stage I and II patients undergoing breast-conserving surgery (BCS) could indicate whether patients are involved in decision-making regarding their treatment plans. The National Surgical Adjuvant Breast and Bowel Project [NSABP] study B-06 compared the outcomes of patients who received mastectomies with patients who received lumpectomy plus radiation. No significant differences were noted between these two groups of women regarding local control, disease-free survival (DFS) rates, distant DFS rates and overall survival rates [Fisher et al. 2002]. Breast-conserving surgery should be offered to patients with stages I and II breast cancers as one of the standard therapies [Lu et al. 2004]. Owing to the size difference of breast for oriental and western women, it would be difficult to operate BCS for the stage II women in Taiwan. This indicator would be presented in sub-indicators from different stages according to the second stage stakeholder feedback.

The proportion of patients with invasive cancers who receive radiation treatment after BCS could reduce the local recurrence of breast cancer [Fisher et al. 2002]. The risk of local recurrence after BCS alone could be as high as 40%; adjuvant radiotherapy can reduce the risk of recurrence to less than 10% [Fisher et al. 2002; Veronesi et al. 2002]. Early Breast Cancer Trials’ Collaborative Group obtained similar findings from a meta-analysis [Early Breast Cancer Trials’ Collaborative Group 2000]. Two recent studies have applied similar indicators [Malin et al. 2002; Schneider et al. 2004].
Use of indicators

Usefulness in revealing patterns of care for internal quality improvement

There are several sources of data in Taiwan that could be used for obtaining empirical results for these core measures. The claims data from the NHI cover inpatient and ambulatory care, and some palliative care. The Taiwan Cancer Data Base has cancer staging and treatment data for newly diagnosed patients using definitions of the American Joint Committee on Cancer (AJCC 2002; American College of Surgeons 2002). We can combine data from these sources across time to obtain a complete picture of CPMs for breast cancer. These indicators also could be used to identify regional differences [McGlynn 2003], to assure accountability [McGlynn & Malin 2002] and to incorporate measurement and feedback for facilitating change [Berwick et al. 2003]. Indicator could be used specifically to reveal patterns of care across hospitals, or as a whole to represent organization-based quality of care. System theory suggests that the trigger for change will be a clear, specific, measurable goal [Rhydderch et al. 2004]. Feedback from an assessment against the standard or goal can be a stimulus for guiding the movement forward [Van de Ven & Poole 1995] and triggering quality improvement [Grol 2001].

Serving as national benchmarks or external monitoring measures

In Taiwan, the Legislative Yuan passed the Cancer Control Act in 2003 (Bureau of Health Promotion 2003) and the BHP provided financial support to initiate cancer quality improvement programme with regulatory mechanisms of reporting and feedback. Each hospital must submit data twice a year for six cancers including breast, cervical, lung, oral cavity, colorectal cancer and hepatocellular carcinoma. Having a single set of measures for the entire country is the only way to set national benchmarks and compare plans that have national programmes (Eddy 1998; McGlynn & Malin 2002).

Limitations

Several limitations of this study need to be discussed. The number and content of the evaluation subcriteria, as well as the threshold values for assessing levels of agreement may have affected the results. The size and attendance in the face-to-face discussion of the expert panel may be another concern. The two stages stakeholder feedback could be supplement to the weakness regarding the consensus building process. This study focused more on the effectiveness aspect of quality, while other aims for improvement such as safety and patient-centred care were not considered simultaneously [IOM 2002]. Stakeholders other than medical professionals were not included in this study. End-of-life care was not covered in the pilot of CPMs.

Conclusions

This study attempted to modify methods from JCAHO and the RAND appropriateness method for oncology care. Two stages stakeholder feedback, comprehensive criteria and strict inclusion criteria were major contributions in developing core cancer measures from this study. Developing core measures for breast cancer is the first step to achieving standardized measures for external monitoring, as well as for providing feedback and serving as benchmarks for quality improvement. Future research can obtain complete results for these indicators, and can consider risk adjustment for comparing performance across different hospitals. It also may be desirable to apply fewer evaluation subcriteria and more flexible inclusion criteria for selection of core measures in other cancer sites when support from evidence-based results is not sufficiently strong. The application of core measures results linked with pay for performance or certification programmes in cancer care may be considered in the future.

ACKNOWLEDGEMENTS

The authors thank Dr Laura Morlock [Department of Health Policy and Management, Johns Hopkins University), Dr Shoou-Yih Lee [Department of Health Policy Administration, University of North Carolina at Chapel Hill] and Dr Jung-Der Wang [Institute of Occupational Medicine and Industrial Hygiene, National Taiwan University] for their thoughtful comments of this article. This study was supported by a grant from the BHP, Department of Health [DOH-92-HP-2501]. We would like to thank the cooperation from seven experts, 27 hospitals and four of them joining the on-site-visit voluntarily.

REFERENCES


National Institutes of Health Consensus Development Panel [NIHCD] (2001) National Institutes of Health Consensus...


Taiwan Cooperative Oncology Group [TCOG] (2002) *Diagnosis and Treatment Consensus for Breast Cancer*. Available at: http://www.nhri.org.tw/publish/


**APPENDIX 1**

Panel members and affiliations

- MD, PhD, Assistant Professor, Department of Surgery, National Taiwan University Hospital;
- MD, PhD, Professor of Institute of Preventive Medicine, College of Public Health, National Taiwan University;
- MD, Department of Radiation Oncology, Koo Foundation Sun Yat-Sen Cancer Center;
- MD, MS, Associate Professor, Director, Department of Pathology, Taipei Medical University Hospital; MD, Associate Professor, Director, Department of General Surgeon, Tri-General Service Hospital;
- MD, Associate Professor, Director, Department of General Surgeon, Tri-General Service Hospital;
- PhD, Assistant Professor, Institute of Nursing, Medical College, National Taiwan University;
- MD, Department of Medical Oncology, Taichung Veteran General Hospital.