Higashi *BioPsychoSocial Medicine* 2010, **4**:14
http://www.bpsmedicine.com/content/4/1/14

BIOPSYCHOSOCIAL MEDICINE

REVIEW Open Access

Lessons learned in the development of process quality indicators for cancer care in Japan

Takahiro Higashi

Abstract

In Japan, attention has increasingly focused on ensuring the quality of care, particularly in the area of cancer care. The 2006 Basic Cancer Control Act reinforced efforts to ensure the quality of cancer care in a number of sectors, including the role of government in ensuring quality. We initiated a government-funded research project to develop quality indicators to measure the quality of care for five major cancers (breast, lung, stomach, colorectal, and liver cancer) in Japan, and palliative care for cancers in general. While we successfully developed a total of 206 quality indicators, a number of issues have been raised regarding the concepts and methodologies used to measure quality. Examples include the choice between measuring the process of care versus the outcome of care; the degree to which the process-outcome link should be confirmed in real-world measurement; handling of exceptional cases; interpretation of measurement results between quality of care versus quality of documentation; creation of summary scores; and the optimal number of quality indicators for measurement considering the trade-off between the measurement validity versus resource limitations. These and other issues must be carefully considered when attempting to measure quality of care, and although many appear to have no correct answer, continuation of the project requires that a decision nevertheless be made. Future activities in this project, which is still ongoing, should focus on the further exploration of these problems.

Introduction

In Japan, interest in ensuring the quality of patient health care has recently increased. Although Japanese citizens have enjoyed universal health insurance coverage for more than 40 years, concern has been expressed at the lack of an efficient system for monitoring the quality of care [1,2]. To date, quality monitoring has been sporadic at best [3,4], and organized efforts to improve quality have yet to be established. In the area of cancer care, this concern led to the enactment of the Cancer Control Act in 2006, which mandated that the government adopt a leadership role and take responsibility in ensuring the quality of cancer care nationwide [5]. However, ensuring quality care first requires an efficient means of measuring it.

To this end, in 2006 the Japanese government funded a research project aimed at developing a system for measuring the quality of cancer care, focusing primarily on the five major types of cancer in Japan, namely breast, lung, stomach, colorectal, and liver cancer, as well as palliative care. Given that the project was aimed at measuring quality in terms of how current best practice was applied, rather than the general suitability of services (such as waiting times or the comfort of hospital beds), we sought extensive involvement from nationally recognized clinical experts in respective clinical areas. Methodology was under the direction of an epidemiologist and a health services researcher, while the contents of quality measurement were primarily defined by clinicians.

An overview of the study was published in Japanese [6]. Briefly, we developed a total of 206 process-of-care quality indicators for cancer care using methodology developed by the researchers at University of California, Los Angeles and RAND Corporation [7-12]. This involved creating a set of candidate quality indicators, compiling evidence to support these indicators, and having a panel of multidisciplinary experts examine them for validity in two rounds of rating (scale of 1-9), once before and once after a face-to-face discussion. After discarding those indicators judged as having low validity, 206 indicators remained, all of which described the standards of care which define the target patients and the care processes that need to be provided to them. Example indicators are presented in Table 1.

Correspondence: higashi@m.u-tokyo.ac.jp
Department of Health Policy/Public Health, Graduate School of Medicine, the University of Tokyo, Tokyo, Japan

Table 1 Example of quality indicators developed in our project

Denominator (target patients)	Numerator (care processes indicated)
Patients with stage 3 colorectal cancer who have undergone surgical resection	Patients who have received standard chemotherapy within 8 weeks of surgery or who have reasons for not receiving chemotherapy
Patients with colorectal cancer who have undergone surgical resection	Patients who have received total colonoscopy or who have reasons for not receiving total colonoscopy
Patients with gastric cancer who have undergone endoscopic resection and have one of the following: • positive vertical margin • lymphovascular invasion • invasion to the SM2 layer (≥500 μm)	Patients who have received surgical treatment with lymph-node dissection
Patients diagnosed with breast cancer	Patients who have had their Her-2/neu status examined
Patients receiving treatment for liver cancer	Patients who have had the levels of α-fetoprotein and protein in vitamin K absence-II checked before the start of therapy
Patient with lung cancer who have undergone surgical resection or radiation therapy on their lungs	Patients whose lung function has been assessed via spirometry
Patient receiving narcotic analgesics	Patients who have received education or medication to prevent constipation

Since most Japanese clinicians have no experience with quality measurement, such assessment must be approached with care, while conscientiously addressing a number of major issues. Although the study is still underway, we describe here several issues of note and our methods of addressing them in the hope that they may be of value to researchers and policy makers developing similar quality measurement systems. Before continuing, however, we caution that there are no right solutions or answers to these issues and questions and instead remind readers that the decisions made here were those which we believe best served the present purpose.

Choosing to examine process vs. outcome

Quality of care is typically measured with regard to structure, process, or outcome [13,14]. Structural quality refers to organizational and facility quality, such as the number of specialists on staff, staff-patient ratio, and the availability of high-technology equipment. Process quality refers to the appropriateness of care provided during patient encounters, such as the appropriateness of medications and the selection of therapies and follow-up. Outcome measurement estimates the quality of care that patients receive by examining what happens to patients as a result of care, such as mortality after surgery and readmission after discharge.

Establishing a quality monitoring system first requires a decision on which level(s) should be addressed, taking into consideration the relative strengths and weaknesses of each. Given that the government has already enacted structural requirements for hospitals seeking designation as "designated cancer care hospitals," [15] the choice for the present quality measurement lies between process and outcome. Outcome in cancer care is typically measured by five-year survival [16,17], and given that improvement of outcome is the general objective of medicine, its importance is clear. Member hospitals of the Japanese Association of Clinical Cancer Centers *(Zen-gan-kyo)* acknowledged this in 2007 through their initiation of public reporting of five-year survival rates by facility [18].

However, outcome reflects not only the medical care provided but other factors as well, such as patient baseline health and compliance with medical advice. Any comparison of quality across populations and facilities or over time needs to statistically adjust for these factors, particularly patient-case mix [19]. Moreover, the calculation of five-year survival is necessarily a five-year process, a time lag which weakens the measurement's value in improving care [13].

In contrast, evaluation of process, involving the specific focus on what is done to patients, is the most direct of the three measurement schemes. Process quality indicators describe what should be done to what type of patient under which clinical conditions, and the results accordingly highlight direct targets for improvement. Unlike outcome, process does not require a lengthy five-year period to obtain results once the measurement system is established. Indeed, when integrated with electronic order entry and medical records, process values can even be used prospectively, as clinical reminders [20].

One challenge to using process measures is criteria definition. Although quality indicators should be based on clinical evidence, direct clinical evidence among important populations (e.g. older persons) is occasionally limited [10], and criteria must therefore be extrapolated from studies in other populations (e.g. younger persons) based on expert consensus. In this situation, we considered that the optimal methodology to examine the consensus was that developed by the RAND Corporation, as described above [7-12]. While this methodology is not perfect [21], we expect that it will provide the least biased set of quality indicators when appropriately performed,

given previous findings that quality criteria developed using this methodology is reproducible [22,23] and agree with the opinions of practicing physicians [24] and have predictive validity [25].

Proving the process-outcome link

To qualify as a valid process quality indicator, a particular standard of care should help to improve patient outcome; if it does not, the provision of such care cannot be considered high-quality. Donabedian called this the "contributional validity" of the process measure [13]. The lack of direct clinical evidence from randomized controlled trials encountered in many clinical situations sometimes renders contributional validity ambiguous, and clinicians can present different interpretations of evidence from those of the expert panel members who helped to develop an indicator, particularly with indicators developed based on indirect evidence. Typical examples of such lack of direct evidence occur with quality indicators that target diagnostic standards: although computer tomography (CT) of the liver before colon cancer surgery is a standard of care, for example, on the basis that CT scan results affect treatment decisions and can thus be a quality indicator, no randomized controlled trial has examined the utility of CT of the liver before surgery. Indeed, few diagnostic procedures have been subject to randomized controlled trials.

Several points must be carefully considered before examining the relationship between the process quality measured and outcomes. First, the unit of analysis must be identified. When comparing outcomes of patients who do or do not receive a certain therapy, for example, the patient is the unit of analysis [25,26]. In comparing adherence rates, in contrast, the facility is the unit [27,28]. Although the relationship at the aggregate level does not necessarily imply the relationship at the individual level, owing to the *ecological fallacy* [29], analysis at the facility level will have acceptable validity if it is accepted that the adherence rate represents the quality of care provided by the facility, and incorporates other aspects of care than those specifically measured in the quality indicators. The appropriate unit of analysis will depend on the unit's comparability and the level at which the quality data are available.

Second, as quality indicators may target different types of outcomes, the outcomes themselves must be carefully chosen [14]. The most common patient outcome in cancer care is five-year survival, but not all high-level care prolongs life, or aims to prolong life. To list three examples, suitable explanation of the risks and benefits of treatment options is surely an aspect of quality, specifically the provision of respect to patient autonomy, but it is unlikely to improve survival; administration of anti-emetic medications before chemotherapy is aimed at alleviating adverse symptoms during therapy, and not to make any contribution to survival; and patient education about drug regimens may increase compliance or reduce medication errors. Although these may improve patient health or prevent adverse events, their benefits may be too small to detect from observation, yet they are strongly supported as quality indicators.

Another point in considering the process-outcome link may be the timing of outcome observation. This is seen in any epidemiological study which examines the cause-effect relationship [29]: outcomes of acute conditions may be observed relatively early, while preventive care processes will require some time to show results. Evaluation of 100-year survival will show no impact of quality no matter how high it may be. Though absurd, this example highlights the importance of timing, and the difficulty often met in determining the right timing for a particular outcome.

Finally, interpretation of process-outcome links observed in the real world requires caution. As randomization of patients to high or low quality care is ethically untenable, studies examining quality-outcome relationships are necessarily observational; and since persons receiving high- and low-quality care may differ, the level of evidence from such studies is lower than that from randomized controlled trials, where optimum comparability can be expected. If the content of a quality indicator (i.e., standard of care) is supported by randomized controlled trials, its validity as an indicator may not be refuted by a lack of relationship to outcomes observed in clinical application, unless the target population is extremely different.

Handling exceptions

No rule is free of exceptions. Although quality criteria define which patients should receive which care processes, some patients will not receive a particular item of care for any of a number of reasons: for example, patients with acute myocardial infarction should, by standard, receive aspirin on admission, but this is of course foregone if the patient is allergic to aspirin. In this case, "high quality" can be defined as care that appropriately distinguishes exceptions from regular cases and tailors treatment to the individual patient. Unfairly penalizing such cases as "failures" to provide standard care should be avoided.

Three methods of handling exceptions to quality indicators can be considered. First, these cases can be excluded from the denominator in calculating the adherence rate, thereby providing a "pure" sample of standard patients. In our study also, denominators for some indicators were made narrower than the population to which the numerator care is usually applied in regular practice. This is because we wanted a sample to whom the care is clearly applicable. However, this method not only fails to give

proper credit to the provider who makes these important decisions of tailored medicine, but also lowers the observed quality score (i.e. adherence rate to the quality indicator), because subtracting the same number from both the numerator and denominator actually decreases the fraction when the fraction is less than 1 (for example, (8-1)/(10-1) = 7/9 = 0.78 is less than 8/10 = 0.8). Further, relatively small target populations can render quality scores statistically unstable, with large standard errors.

A second method of handling exceptions is to treat them as if the care had in fact been provided, provided that the reason for the exception is properly documented. In this way, atypical patients can be kept in the sample and the care they receive can be evaluated in the same way as for regular patients. Quality scores using this method tend to be more stable than those using the first method above, but patient eligibility for the quality indicator becomes heterogeneous, which may then dilute the link between the quality and the outcome by introducing noise into the biological process-outcome link. While acknowledging this limitation, we basically adopted this second method of handling exceptions in the present study. This can be seen in the fact that many of our quality indicators explicitly demand documentation of the reason why the indicated care was not provided (Table 1).

The third way of handling exceptions is to reduce target adherence by the expected number of exception patients [30]. If 5% of patients ostensibly applicable to a quality indicator are estimated to become exceptions, the quality target can be set at 95% instead of 100%. Given that identifying individual exceptions requires that the researcher abstracting the medical record has knowledge of treatment choices, which is occasionally difficult due to a lack of detailed documentation [31], a simple reduction in target enables the judgment of individual cases to be avoided, and also thus the risk of errors in such judgment. Further, gaming with the system can be discouraged by labeling patients who did not receive the indicated care as exceptions, perhaps retrospectively. In comparison of quality scores across populations, however, national variations in the proportion of exception cases may introduce large random errors into the comparison.

The above clearly demonstrates the difficulty in handling exceptions. Even when considering only individual cases, we may well refrain from stating that the target quality score is 100%, or instead state outright that the target is slightly lower than 100%. Indeed, the UK pay-for-performance initiative (Quality and Outcomes Framework) gives full points even for quality scores less than 100% [32,33], although they also allowed providers to limit the denominators reporting exceptions. At the least, quality information should be interpreted together with data regarding the nature and number of exceptions reported by providers [30,32,33].

Quality of care vs. quality of documentation

Quality scores depend to some extent on the quality of available data. While process-of-care quality should ideally be evaluated via direct observation, or perhaps in standardized patients [34], information is typically collected from medical records [9,35-37]. In addition, documentation of essential clinical information, such as cancer stage and follow-up review of drug regimens, consists of quality indicators based on the notion that such documentation represents an aspect of quality of care [31,38]. Both a reliance on documentation for the implementation of quality indicators and a specific focus on documentation among several quality indicators leads to the impression that quality measurement overly emphasizes documentation, and thereby begs the question of whether quality evaluation measures the quality of medical care or the quality of documentation.

Physicians who believe that documentation is a largely separate issue from medical care may be reluctant to accept that pertinent documentation is part of quality care. Indeed, during our informal discussion about quality indicators, some clinicians practicing in urban areas gave the example of providers in rural areas where a small number of health professionals handle most medical issues. They noted that rural physicians may feel particularly strongly about this aspect, as they are usually too busy to keep detailed records and may not need to communicate with the few other health professionals in their area through documentation, opting instead for face-to-face conversations. However, having sufficient documentation ensures smooth sharing of information among health care professionals, supplementing face-to-face conversation. Given that a miscommunication is a major cause of medical errors [39], quality of documentation can be safely deemed as falling under the umbrella of quality of care, which will thereby increase the probability of safe practices and improved outcomes.

Scope of quality measures and number of indicators

Medical care has various aspects, including not only medical intervention but also patient education and coordination of providers, and the long continuum spanning prevention, diagnosis, treatment, and follow-up. Narrow evaluation captures only limited aspects and cannot represent the quality of care [40]. In addition, evaluation based on a small number of quality criteria can easily be gamed by providers allocating resources only to satisfy the defined quality indicators and achieve

high scores, leaving out or even sacrificing other, perhaps more important aspects of care [41].

In contrast, broad measurement of quality covering various aspects of care is resource- and labor-expensive. Insurance claims data can be a suitable alternative, provided the necessary information is available; the range of information is limited to the original utilization, however, and the data tend to lack details regarding patient condition, such as laboratory and imaging results [42]. An empirical study examining broad aspects of quality based on medical records found that quality indicators that can be measured from claims data tended to result in better scores than those requiring information from medical records [43]. This finding suggests that selecting quality indicators based on the availability of information in insurance claims may fail to detect and solve problems in quality of care. Instead, theoretically, selection should be primarily based on the importance of the care process. Significant attention must be paid to the balance between the validity of the quality measurement and the resources spent on measurement.

Quality indicators in rare but important situations

Several issues must be considered when selecting quality indicators by the priority of measurement. One such issue is the expected effectiveness of the care indicated in the quality measure, i.e. the potential for the care to improve the patient's outcome. Another issue is the expected room for improvement in the practice being evaluated; care processes which are known to be always performed, for example, do not need to be re-examined.

A particularly controversial perspective is the number of patients in whom the quality indicator is applicable. One example is the need to perform additional surgical resection after endoscopic resection of a cancer which turned out to be more deeply extended than initially estimated, with a high risk of lymph node metastasis. Given that such cases are relatively rare, however, including these quality indicators in the measurement set may be inefficient, because information needs to be collected from the entire target patient population to determine whether the quality indicator is applicable to them. In addition, when the number of applicable patients is small, the denominator of the quality score as the percent adherence is also small, making scores calculated using adherence unstable. This latter problem can be solved by obtaining a larger denominator through application of the indicator to a larger population or for a longer duration; for example, even if one quality indicator applies to only three patients in a hospital annually, it may be applicable to 30 patients in a town with 5 hospitals over a 2-year period. The importance of a quality indicator should therefore be judged on a global scale; if the care process described in the indicator strongly affects even a few patients, a decision should be made based on contextual factors, such as whether or not the indicator can be applied on a broader level to achieve a sufficient number of applicable patients.

Creating a summary score

Quality indicators are scored based on the percentage of patients who receive the care described in them. Our cancer project produced 206 quality scores. When reviewing the results, the research team felt that interpreting such many scores was difficult. To summarize them, we therefore produced an overall number of adherence to the indicators by dividing the total number of patients to whom the quality indicator care was provided (sum of all numerators) by the total number of times the quality indicator was applicable to the sample patients (sum of all denominators). Although this score has the conceptually reasonable meaning of "overall performance of standards by the provider," we then speculated whether all indicators should have equal weight in calculating the summary score, given that quality indicators appeared to have different degrees of importance. In particular, for those who do not think documentation is important, documentation indicators must have smaller weights than other indicators.

Creating summary score weighting for quality indicators presents a challenge [44]. While the overall performance is one-way, this orientation ignores the natural importance of the care processes. An indicator's weight can relate to the comparative importance of the care process; for example, providing oxygen to a hypoxic patient is more important–at that moment–than documenting stage of cancer within one month of diagnosis.

Assessment of comparative importance can be aided by reviewing the expected improvement in outcome if the care is provided. However, when quality indicators target different stages of management of a disease, the comparison requires clinical judgment. For example, when determining whether or not prescribing antiemetic medications before starting chemotherapy is more (or less) important than examining the entire colon before surgery, we must consider the difference in the basic nature of the outcomes expected to be improved by these care processes. Clinical judgment can be elicited either directly in the context of quality measurement by expert ratings of overall importance of quality indicators, as done by Ashton et al [45]., or by focusing solely on outcome by asking clinicians about the overall degree of outcome improvement on a single scale, integrating the expected different outcomes such as survival and quality of life, as is done by the indicator selection at the National Comprehensive Cancer Network [46], albeit that this is not to assign weights but to rank quality indicators by their priority. Future studies should assess the validity and reliability of such methodologies.

One way to circumvent the difficulty in assigning weights is to create a rule which integrates multiple quality indicators, such as the all-or-none rule [47]. Under this rule, several quality indicators are grouped together, and patients who receive all the indicated care according to the quality indicator set are counted. For example, if a patient is eligible for quality indicators A, B, and C, s/he is counted as having received quality care only if s/he has received all the care described in indicators A, B, and C. If s/he receives care in quality indicator A and B, but not C, s/he is not counted as having received the care. The quality score for a facility or patient group is then defined as the proportion of patients who received "perfect" care in the quality indicator set. Use of this strategy is supported by the notion that care is interrelated, and that only receiving all necessary care in the target area is acceptable. However, the score then depends on the grouping of the quality indicators, which must be theoretically justifiable, and most importantly, determined before the start of measurement. Further, if multiple such groups are created, weighting of scores may again become an issue, presenting new problems. In addition, handling of exceptions to quality measurement becomes increasingly important, as mistakenly entering one quality indicator event that should have been excluded reduces the whole count for a patient.

In our present study, we have not yet decided on the best method of calculating the summary score. Several approaches will likely need to be tested, and discussions among members of the research team and related clinicians will be conducted to address the possible options.

Conclusion

A number of issues must be addressed when developing process quality indicators for cancer care, several of which are reviewed here. While our project specifically targets cancer care, many of these issues may also apply in other clinical areas. Future researchers should not expect to find the "right answers" to these issues and questions, but rather should make decisions based on best judgment, and thereby ensure progress. While different systems can call for different decisions, comparison of quality across patient populations (e.g., several facilities or over time) requires recognition of the fact that these decisions must be consistent across the compared groups.

Acknowledgements
The author thanks collaborators to the Research Group for the Development of Quality Measurement Systems for Cancer Care for their support in the complex process of quality measurement.

Authors' contributions
The author wrote the manuscript and hold responsibility for the content of the manuscript.

Competing interests
The author declares that they have no competing interests.

Received: 14 October 2010 Accepted: 5 November 2010
Published: 5 November 2010

References

1. Yamasaki Y, Ogata H: *Healthcare Reform and the role of Insurers (Iryoseido-kaikaku-to-hokenja-kino in Japanese)* Tokyo: Nippon-keizai shinbun; 2002.
2. Ikegami N: *Basic Health Care Problems (Basic iryo-mondai in Japanese).* 3 edition. Tokyo: Nippon-keizai shinbun; 2006.
3. **Evaluation Project for Outcomes of Care (Shinryo-Autokamu-Hyoka-Jigyo, in Japanese).** [http://www.ajha.or.jp/hms/outcome/jigyou_1.html].
4. Fukui T: *Quality Indicaotr: An novel approach by St. Luke's International Hospital* Tokyo: Intermedica; 2009.
5. **Cancer Control Act.** [http://law.e-gov.go.jp/announce/H18HO098.html].
6. **Quality Indicators for Cancer Care (in Japanese).** [http://qi.ncc.go.jp].
7. Fitch K, Bernstein SJ, Aguilar MD, Burnand B, LaCalle JR, Lazaro P, van het Loo M, McDOnnel J, Vader JP, Kahn KL: *The RAND/UCLA Appropriateness Method User's Manual* RAND; 2001.
8. Park RE, Fink A, Brook RH, Chassin MR, Kahn KL, Merrick NJ, Kosecoff J, Solomon DH: **Physician ratings of appropriate indications for six medical and surgical procedures.** *Am J Public Health* 1986, **76**:766-772.
9. McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A, Kerr EA: **The quality of health care delivered to adults in the United States.** *N Engl J Med* 2003, **348**:2635-2645.
10. Shekelle PG, MacLean CH, Morton SC, Wenger NS: **Acove quality indicators.** *Ann Intern Med* 2001, **135**:653-667.
11. Wenger NS, Shekelle PG: **Assessing care of vulnerable elders: ACOVE project overview.** *Ann Intern Med* 2001, **135**:642-646.
12. Kerr EA, Asch SM, Hamilton EG, McGlynn EA: *Quality of Care for General Medical Conditions* Santa Monica: RAND Health; 2000.
13. Donabedian A: *The Definition of Quality and Approaches to its Assessment* Ann Arbor: Health Administration Press; 1980.
14. Donabedian A: **Evaluating the quality of medical care.** *Milbank Mem Fund Q* 1966, **44(Suppl)**:166-206.
15. **About the designation of cancer care hospitals (in Japanese).** [http://www.mhlw.go.jp/topics/2006/02/tp0201-2.html].
16. Berrino F, De Angelis R, Sant M, Rosso S, Bielska-Lasota M, Coebergh JW, Santaquilani M: **Survival for eight major cancers and all cancers combined for European adults diagnosed in 1995-99: results of the EUROCARE-4 study.** *Lancet Oncol* 2007, **8**:773-783.
17. Coleman MP, Quaresma M, Berrino F, Lutz JM, De Angelis R, Capocaccia R, Baili P, Rachet B, Gatta G, Hakulinen T, *et al*: **Cancer survival in five continents: a worldwide population-based study (CONCORD).** *Lancet Oncol* 2008, **9**:730-756.
18. **Collaborative Study of Cancer Survivals (in Japanese).** [http://www.gunma-cc.jp/sarukihan/seizonritu/].
19. Iezzoni LI: **Risk Adjustment.** Chicago, IL: Health Administration Press; 2003.
20. Shojania KG, Jennings A, Mayhew A, Ramsay C, Eccles M, Grimshaw J: **Effect of point-of-care computer reminders on physician behaviour: a systematic review.** *CMAJ* 2010, **182**:E216-225.
21. Coulter I, Adams A, Shekelle P: **Impact of varying panel membership on ratings of appropriateness in consensus panels: a comparison of a multi- and single disciplinary panel.** *Health Serv Res* 1995, **30**:577-591.
22. Hemingway H, Chen R, Junghans C, Timmis A, Eldridge S, Black N, Shekelle P, Feder G: **Appropriateness criteria for coronary angiography in angina: reliability and validity.** *Ann Intern Med* 2008, **149**:221-231.
23. Shekelle PG, Kahan JP, Bernstein SJ, Leape LL, Kamberg CJ, Park RE: **The reproducibility of a method to identify the overuse and underuse of medical procedures.** *N Engl J Med* 1998, **338**:1888-1895.
24. Ayanian JZ, Landrum MB, Normand SL, Guadagnoli E, McNeil BJ: **Rating the appropriateness of coronary angiography–do practicing physicians agree with an expert panel and with each other?** *N Engl J Med* 1998, **338**:1896-1904.
25. Higashi T, Shekelle PG, Adams JL, Kamberg CJ, Roth CP, Solomon DH, Reuben DB, Chiang L, MacLean CH, Chang JT, Young RT, Saliba DM, Wenger NS: **Quality of care is associated with survival in vulnerable older patients.** *Ann Intern Med* 2005, **143**:274-281.
26. Fonarow GC, Abraham WT, Albert NM, Stough WG, Gheorghiade M, Greenberg BH, O'Connor CM, Pieper K, Sun JL, Yancy C, Young JB:

Association between performance measures and clinical outcomes for patients hospitalized with heart failure. *JAMA* 2007, **297**:61-70.
27. Bradley EH, Herrin J, Elbel B, McNamara RL, Magid DJ, Nallamothu BK, Wang Y, Normand SL, Spertus JA, Krumholz HM: **Hospital quality for acute myocardial infarction: correlation among process measures and relationship with short-term mortality.** *JAMA* 2006, **296**:72-78.
28. Werner RM, Bradlow ET: **Relationship between Medicare's hospital compare performance measures and mortality rates.** *JAMA* 2006, **296**:2694-2702.
29. Rothman K, Greenland S, Lash T: *Modern Epidemiology.* 3 edition. Philadelphia: Lippincott Williams & Wlkins; 2008.
30. Doran T, Fullwood C, Reeves D, Gravelle H, Roland M: **Exclusion of patients from pay-for-performance targets by English physicians.** *N Engl J Med* 2008, **359**:274-284.
31. Malin JL, Schneider EC, Epstein AM, Adams J, Emanuel EJ, Kahn KL: **Results of the National Initiative for Cancer Care Quality: how can we improve the quality of cancer care in the United States?** *J Clin Oncol* 2006, **24**:626-634.
32. Doran T, Fullwood C, Gravelle H, Reeves D, Kontopantelis E, Hiroeh U, Roland M: **Pay-for-performance programs in family practices in the United Kingdom.** *N Engl J Med* 2006, **355**:375-384.
33. Campbell SM, Reeves D, Kontopantelis E, Sibbald B, Roland M: **Effects of pay for performance on the quality of primary care in England.** *N Engl J Med* 2009, **361**:368-378.
34. Peabody JW, Luck J, Glassman P, Dresselhaus TR, Lee M: **Comparison of vignettes, standardized patients, and chart abstraction: a prospective validation study of 3 methods for measuring quality.** *JAMA* 2000, **283**:1715-1722.
35. Jencks SF, Cuerdon T, Burwen DR, Fleming B, Houck PM, Kussmaul AE, Nilasena DS, Ordin DL, Arday DR: **Quality of medical care delivered to Medicare beneficiaries: A profile at state and national levels.** *JAMA* 2000, **284**:1670-1676.
36. Jencks SF, Huff ED, Cuerdon T: **Change in the quality of care delivered to Medicare beneficiaries, 1998-1999 to 2000-2001.** *JAMA* 2003, **289**:305-312.
37. Wenger NS, Solomon DH, Roth CP, MacLean CH, Saliba D, Kamberg CJ, Rubenstein LZ, Young RT, Sloss EM, Louie R, Adams J, Chang JT, Venus PJ, Schnelle JF, Shekelle PG: **The quality of medical care provided to vulnerable community-dwelling older patients.** *Ann Intern Med* 2003, **139**:740-747.
38. Jacobson JO, Neuss MN, McNiff KK, Kadlubek P, Thacker LR, Song F, Eisenberg PD, Simone JV: **Improvement in oncology practice performance through voluntary participation in the Quality Oncology Practice Initiative.** *J Clin Oncol* 2008, **26**:1893-1898.
39. Woolf SH, Kuzel AJ, Dovey SM, Phillips RL Jr: **A string of mistakes: the importance of cascade analysis in describing, counting, and preventing medical errors.** *Ann Fam Med* 2004, **2**:317-326.
40. Brook RH, McGlynn EA, Cleary PD: **Quality of health care. Part 2: measuring quality of care.** *N Engl J Med* 1996, **335**:966-970.
41. Casalino LP: **The unintended consequences of measuring quality on the quality of medical care.** *N Engl J Med* 1999, **341**:1147-1150.
42. Iezzoni LI: **Assessing quality using administrative data.** *Ann Intern Med* 1997, **127**:666-674.
43. MacLean CH, Louie R, Shekelle PG, Roth CP, Saliba D, Higashi T, Adams J, Chang JT, Kamberg CJ, Solomon DH, Young RT, Wenger NS: **Comparison of administrative data and medical records to measure the quality of medical care provided to vulnerable older patients.** *Med Care* 2006, **44**:141-148.
44. Reeves D, Campbell SM, Adams J, Shekelle PG, Kontopantelis E, Roland MO: **Combining multiple indicators of clinical quality: an evaluation of different analytic approaches.** *Med Care* 2007, **45**:489-496.
45. Ashton CM, Kuykendall DH, Johnson ML, Wun CC, Wray NP, Carr MJ, Slater CH, Wu L, Bush GR: **A method of developing and weighting explicit process of care criteria for quality assessment.** *Med Care* 1994, **32**:755-770.
46. Hassett MJ, Hughes ME, Niland JC, Ottesen R, Edge SB, Bookman MA, Carlson RW, Theriault RL, Weeks JC: **Selecting high priority quality measures for breast cancer quality improvement.** *Med Care* 2008, **46**:762-770.
47. Nolan T, Berwick DM: **All-or-none measurement raises the bar on performance.** *JAMA* 2006, **295**:1168-1170.

doi:10.1186/1751-0759-4-14
Cite this article as: Higashi: Lessons learned in the development of process quality indicators for cancer care in Japan. *BioPsychoSocial Medicine* 2010 **4**:14.

Journal of Evaluation in Clinical Practice
International Journal of Public Health Policy and Health Services Research

Journal of Evaluation in Clinical Practice ISSN 1356-1294

Concordance of hospital-based cancer registry data with a clinicians' database for breast cancer

Mingji Zhang MD PhD,[1] Takahiro Higashi MD PhD,[2] Hiroshi Nishimoto MD,[3] Takayuki Kinoshita MD PhD[4] and Tomotaka Sobue MD PhD[5]

[1]Research Resident, [3]Section Head, [5]Head, Cancer Information Services and Surveillance Division, Center for Cancer Control and Information Services, National Cancer Center, Tokyo, Japan

[2]Associate Professor, Cancer Information Services and Surveillance Division, Center for Cancer Control and Information Services, National Cancer Center and Department of Public Health and Health Policy, University of Tokyo, Tokyo, Japan

[4]Chief, Breast Surgery Division, National Cancer Center Hospital, Tokyo, Japan

Keywords
breast cancer, clinicians' database, concordance of data, exchange information, hospital-based cancer registry

Correspondence
Dr Mingji Zhang
National Cancer Center
5-1-1 Tsukiji
Chuo-ku
Tokyo 104-0045
Japan
E-mail: mzhang@ncc.go.jp

Accepted for publication: 7 October 2010

doi:10.1111/j.1365-2753.2010.01600.x

Abstract

Objective Reliable information is essential to both clinical and policy decision making. We aimed to shed lights on the similarity and differences between a hospital-based cancer registry with a clinicians' database for breast cancer by comparing the registered data on the same year.

Methods We performed a head-to-head comparison of breast cancer cases extracted from the hospital-based cancer registry and the clinicians' database maintained by the Division of Breast Surgery at the National Cancer Center Hospital in 2004.

Results The hospital-based cancer registry reported 827 cases of newly diagnosed breast cancer patients in 2004, while the clinicians' database contained 366 surgically treated cases from 2004. Of these, 276 cases overlapped. Presence or absence of treatment modality was discordant in 15% for radiation therapy, 19% for chemotherapy, and 24% for hormone therapy between the two data sets. Furthermore, the recorded disease pathology was discordant in 13% for pathology and 28% for staging, with 22% for T-stage, 7% for N-stage, 7% for M-stage.

Conclusions Although information contained in hospital-based cancer registry and clinicians' database are generally accurate, some important differences were revealed as a result of varying interpretations of clinical information. Analyses of these data sets must be made with attention to details such as eligible patients, registered treatment, and timing of registration.

Introduction

Effective cancer control policies require accurate information on epidemiology and practice patterns. Cancer registries can theoretically serve these purposes, but Japan has been delayed in establishing such systems with national coverage. Traditionally, two types of cancer registries have been developed: population-based and site-specific cancer registries. Population-based cancer registries, run by prefectural government health departments, aim to assess cancer incidence, while site-specific registries, managed by professional societies, focus more on collecting detailed clinical information [1]. Although both have more than 30 years of history, individual efforts are dependent on prefectures and cancer sites with no existent system that can provide a national picture of cancer incidence or practice patterns.

The Cancer Control Act [2] enacted in 2007 and the National Basic Cancer Plan mandate government promotion of cancer registries. To systematically enhance such activities, the government created a third system, hospital-based cancer registries, by mandating designated cancer centres to register all cancer cases diagnosed or treated at their facility. The items are standardized centrally and cover basic information including disease localization, clinical and pathological stages, and initial treatment provided. Data are collected by registrars trained by the Center for Cancer Control and Information Services at the National Cancer Center. Because standardized items are common for all types of cancer, the collected information lacks site-specific data such as types of surgery and names of chemotherapeutic drugs. However, when a facility participates in both registry schemes, basic information is collected twice: once by clinicians for site-specific

cancer registry and again by tumour registrars for hospital-based cancer registry. In the future, these two systems may be integrated within facilities for the sake of efficiency. In the meantime, however, the current situation provides a unique opportunity to confirm the accuracy of case identification and information in both databases (DBs). Knowledge of the difference and exchangeability of information is important because public agencies and the government tend to use hospital-based cancer registries, while clinicians use site-specific cancer registries to answer the same questions of outcome, survival, and patterns of care.

To understand the similarity and differences of the hospital-based cancer registry and site-specific cancer registry, we compared the hospital-based cancer registry with the clinicians' DB, which supplies data to a site-specific cancer registry at the National Cancer Center Hospital (NCCH), taking breast cancer cases as an example. The comparison provided basic information on exchangeability and differences between the two systems for future discussion of possible integration.

Methods

We extracted breast cancer cases from the hospital-based cancer registry and the clinicians' DB maintained by the Division of Breast Surgery in NCCH in 2004. Although the hospital-based cancer registry and clinicians' DB both collect breast cancer cases, they are independent data collection schemes with several differences.

Hospital-based cancer registry

For the hospital-based registry, information is collected by trained tumour registrars, who systematically extract cases based on pathology reports and other sources to register all cancer cases in the hospital according to the national standard. The unit of registration is the number of tumours. Thus, if one patient has two independent tumours (e.g. breast cancer and colon cancer, or two histologically different breast cancers), he/she is registered twice to represent both cancers. The index date is the date of initial diagnosis if the patient underwent definitive diagnostic test in the hospital or date of first visit to the facility if the patient was already definitively diagnosed with cancer before the first visit. The information collected is common to all cancer types, including cancer site, pathology, route of referral, presentation, clinical and pathological staging [based on the International Union Against Cancer (UICC) system], initial treatment provided in the hospital, and treatment outcomes. The registry does not collect information specific to individual cancers (e.g. hormone receptor status for breast cancer).

Cases were entered in the hospital-based cancer registry at approximately 6 months after diagnosis. Among the treatments provided, only initial therapy planned at diagnosis was documented.

Clinicians' database

In the clinicians' DB for breast cancer, practising physicians collect data in accordance with a template provided by the Japanese Society for Breast Cancer. Almost all patients surgically treated in the Division of Breast Surgery are included. The unit of registration is the patient, and the index date is the date of surgery in the hospital. Patients are registered each time they undergo an operation. Data recorded include clinical findings (e.g. cancer site, tumour size, stage), imaging findings (e.g. mammography, ultrasonography), pathology, complications, hormone receptor status, and specific therapeutic methods (e.g. types of surgery, regimen, dose of radiation, and chemotherapeutic dosage). Staging is assessed using Japanese General Rules for Clinical and Pathological Recording of Breast Cancer. The system differences between hospital-based cancer registry and site-specific cancer registry are summarized in Table 1.

Unlike in the hospital-based cancer registry, all provided therapies are registered in the clinicians' DB. No fixed timing is specified for data entry in the clinicians' DB, but it is presumed the information is updated continuously.

Analytic methods

We compared the concordance of case data registered in the hospital-based cancer registry versus the clinicians' DB. We also compared basic clinical information across overlapping cases, including clinical staging, the tumour-node-metastasis (TNM) classification, date of surgery, and presence or absence of radiation therapy, chemotherapy, and hormone therapy. For patients with discrepancies in documented information, we reviewed the medical record to determine accuracy of data and underlying reasons for the differences.

Results

Differences in registered subjects

The hospital-based cancer registry contained 827 cases diagnosed as breast cancer at the NCCH in 2004, while the

Table 1 Characteristics of the two types of cancer registries

	Hospital-based cancer registry	Site-specific cancer registry
Primary purpose	To assess current status of cancer care	To collect in-depth information for the advancement of clinical cancer management
Managing entity	Hospital	Academic society
Subjects	All diagnosed cancer patients at first visits to the hospital	Patients in major hospitals with cancers of specific sites
Data items	Diagnosis, initial treatment; follow-up 74 items	Variable by cancer site; 100–300 items
Data entry	Mainly tumour registrars	Clinician (physicians)
Problem	Lack of clinical details, especially site-specific information such as hormone status; shortage of tumour registrars	Incomplete follow-up; burden to clinicians

Table 2 Comparison of adjuvant therapy between hospital-based cancer registry and the clinicians' database

Hospital-based cancer registry		Clinicians' database							
			Radiation therapy		Chemotherapy		Hormone therapy		Concordance rate (%)
		All	Yes	No	Yes	No	Yes	No	
	All	276							
Radiation therapy	Yes		97	14					85
	No		13	138					
Chemotherapy	Yes				85	30			81
	No				5	139			
Hormone therapy	Yes						95	12	76
	No						32	123	

Yes: information present; no: information absent.

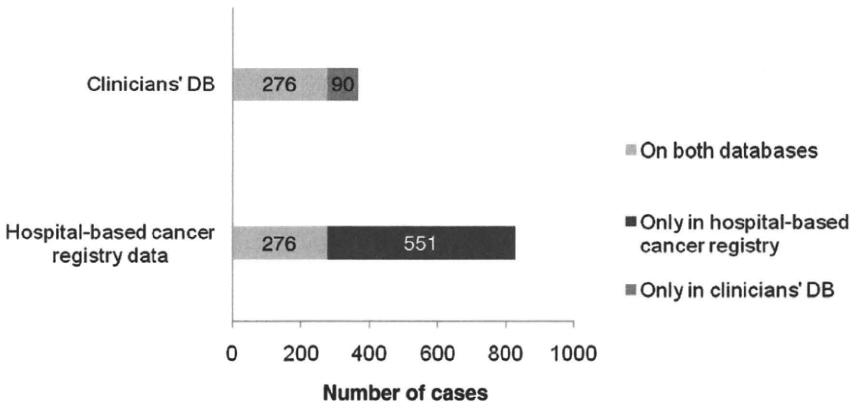

Figure 1 Comparison of subjects from the hospital-based cancer registry and clinicians' database (DB) in 2004.

clinicians' DB documented 366 cases who underwent breast cancer surgery in 2004. Among these, 276 cases were found in both DBs, 551 cases were registered only in the hospital-based cancer registry, and 90 cases were found only in the clinicians' DB (Fig. 1).

Among the 90 cases registered only in the clinicians' DB, the medical record was not available for one case. The other 89 cases were not in the hospital-based registry because they were diagnosed with breast cancer before 2004 and presented for repeat surgery in 2004.

In comparison, 551 cases were found only in the hospital-based cancer registry, among whom 85 later received surgery in 2005. Reasons for why these cases were not found in the clinicians' DB are detailed in Fig. 2.

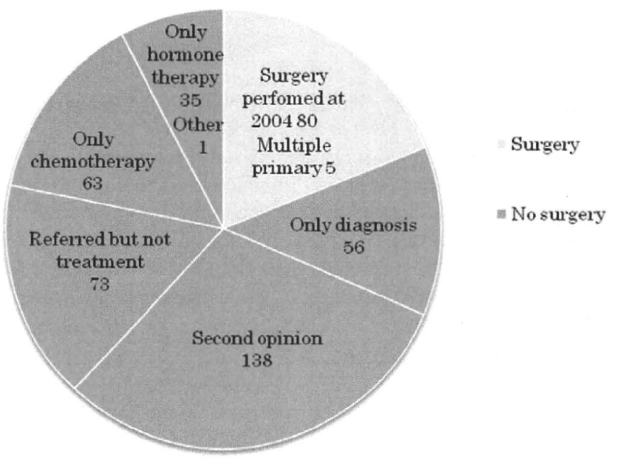

Figure 2 Description of 551 cases only in the hospital-based cancer registry.

Data concordance across overlapping cases

Treatment modality

Among 276 surgically treated patients, the date of surgery matched perfectly across the two DBs. Concordance rates of the presence of adjuvant therapies between the two DBs were also analysed (Table 2). Documentation on administration of radiation therapy was concordant in 235 cases (85%), chemotherapy in 224 (81%), and hormone therapy in 218 (76%). Among discrepant cases, the number of cases documented with receiving radiation therapy only in the hospital-based cancer registry (14) was similar to the number found only in the clinicians' DB (13). However, chemotherapy was logged more frequently in the hospital-based cancer registry (30 vs. 5) and hormone therapy was documented more frequently in the clinicians' DB (32 vs. 12).

Confirmation using the medical record revealed that reasons for the discordance included timing of registration, registration

Table 3 Reasons for discordance in adjuvant therapy between hospital-based cancer registry and clinicians' database

	Reason		
Adjuvant therapy	Timing of registration	Registration criteria	Human error
Radiation therapy			
14 (HD – yes, CD – no)	6 (43%)	0	8 (57%)
13 (HD – no, CD – yes)	2 (15%)	4 (31%)	7 (54%)
Chemotherapy			
30 (HD – yes, CD – no)	2 (7%)	3 (10%)	25 (83%)
5 (HD – no, CD – yes)	1 (20%)	3 (60%)	1 (20%)
Hormone therapy			
12 (HD – yes, CD – no)	3 (25%)	1 (8%)	8 (67%)
32 (HD – no, CD – yes)	17 (53%)	7 (22%)	8 (25%)

HD, hospital-based cancer registry database; CD, clinicians' database.

criteria, and human error (Table 3). For radiation therapy, the most frequent reasons for discrepancy, for both cases only in the hospital-based registry and only in the clinicians' DB were human errors (15 of 27 discrepant cases, 56%). For chemotherapy, the major reason for case only in the hospital-based cancer registry was human errors (25 cases, 83%), while the reason for case only in the clinicians' DB was most frequently different registration criteria (i.e. clinicians' DB register all treatments provided, three cases, 60%). For hormone therapy, human error again accounted for a majority of cases that were only in the hospital-based cancer registry (eight cases, 67%), while timing of registration accounted for cases being only in the clinicians' DB (17 cases, 53%).

Staging

Concordance in disease staging documentation among the 276 cases was 200 (72%), 216 (75%), 256 (93%), and 257 (93%) for clinical stage, T-stage, N-stage, and M-stage, respectively. Major causes for discrepancy included difference in sources of information for staging (i.e. hospital-based cancer registry has the rule to base on imaging, while clinicians' DB sometimes chooses other source based on the clinical judgment), timing of staging (hospital-based cancer registry uses stages before neo-adjuvant therapy, while clinicians' DB tends to document stage between neo-adjuvant therapy and surgery), staging rules (i.e. supraclavicular lymph-node metastasis is coded M0 in hospital-based cancer registry according to the UICC staging while the same situation is classified as M1a-stage in Japanese rule) and human error (Table 4). The most frequent reasons for disagreement are source of information for T-stage (17 cases, 42%), human error for N-stage (10 cases, 67%), and staging rule for M-stage (four cases, 67%).

Pathology

Pathohistology was concordant between the two DBs in 239 (87%) cases (Table 5). The major reason for discrepancy lies in the different timings of pathological reports used for staging (15 cases, 43%) between the two data sets. When neo-adjuvant therapy was

Table 4 Comparison of staging between hospital-based cancer registry and clinicians' database and reasons for differences

		The reason for disagreement			
	Concordance rate	Source of information	Timing of staging	Staging criteria	Human error
Stage	200 (72%)				
T-stage	216 (78%)	17 (42%)	7 (28%)		16 (40%)
N-stage	256 (93%)		5 (33%)		10 (67%)
M-stage	257 (93%)			4 (67%)	2 (33%)

Table 5 Comparison of pathology between hospital-based cancer registry and clinicians' database

	Clinicians' database				
Hospital-based registry	Invasive	Non-invasive	Paget's disease	Missing	Total
Invasive	235	32	0	2	269
Non-invasive	3	3	0	0	6
Paget's disease	0	0	1	0	1
Total	238	35	1	2	276

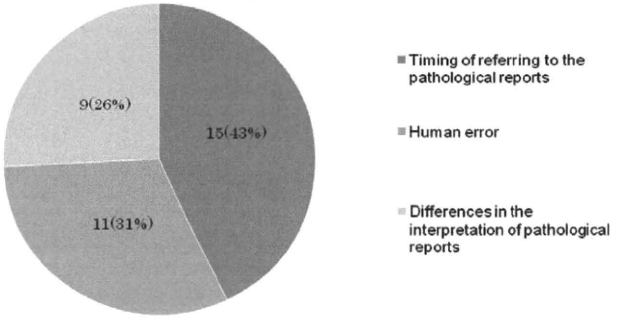

Figure 3 Major reasons for discrepancy in pathologies.

administered, the hospital-based cancer registry refers to pathological reports before neo-adjuvant therapy, while the clinicians' DB refers to pathological reports after neo-adjuvant therapy. In nine cases (26%), differences in the interpretation of pathological reports existed. For example, intraductal apocrine carcinoma is recorded as apocrine carcinoma in the hospital-based cancer registry but as non-invasive ductal carcinoma in the clinicians' DB. Eleven cases (31%) differed in the records as a result of human error (Fig. 3).

Discussion

This study revealed a moderate rate of concordance between hospital-based cancer registry data and the clinicians' DB for breast cancer. Differences in data for the same cases can be attributed to different registration timing, varying information sources, and human errors.

When differences in registered subjects are considered, it should be noted that the target subject differs between the hospital-based cancer registry and the clinicians' DB. While the former documents all diagnosed cancer patients seen at the NCCH, the latter focuses on patients who underwent surgery in the Division of Surgery. Capturing all cases is an important role of the hospital-based cancer registry as it supplies data to the population-based cancer registry. Therefore, more emphasis is given to finding all cases than to collecting detailed information on specific cancers, compared to the clinicians' DB. On the other hand, clinicians' DB has more emphasis on recording clinically detailed information. While the sample is limited to the patients in the Division, the DB has more than twice the items than the hospital-based cancer registry.

Data discrepancies were examined in the 276 cases contained in both data sets. The date of surgery matched perfectly, but substantial discordance was found in the administration of radiation therapy, chemotherapy, and hormone therapy. As the hospital-based cancer registry only documents initial therapy, adjuvant therapy not initially planned but added later based on surgical or pathological findings (i.e. those not initially planned) are not registered. This resulted in narrower coverage of information in the hospital-based cancer registry compared to that in the clinicians' DB, which contains all treatment provided in the facility. The timing of data entry is also different; while the hospital-based cancer registry waits 4 to 6 months after diagnosis to allow the initial therapy be completed [3], the clinicians' DB usually begins registering within 3 months of discharge from the NCCH, and updated when additional later on.

Care received in other hospitals may be underreported in both registry DBs [4,5]. Current cancer registry systems do not follow up patients who transfer care to other facilities. Similarly, the US National Cancer Institute study on Patterns of Care reported that the Surveillance, Epidemiology and End Results data on adjuvant therapy was also somewhat underreported [6–9]. To gain a comprehensive picture of patient care, a preferred approach may be integrating multiple data sources, including insurance claims.

Comparison of TNM stages also revealed the substantial discordance between the hospital-based cancer registry and the clinicians' DB, which may be attributed to different definition criteria. When neo-adjuvant therapy is provided, the hospital-based cancer registry documents the clinical stage prior to initiation of chemotherapy, while the clinicians' DB documents the cancer stage between chemotherapy and surgery. Furthermore, the preferred bases for staging (physical exams, ultrasonography, mammography) is precisely defined in the training of tumour registrars for the hospital-based cancer registry (i.e. use ultrasonography findings over other imaging studies and physical exams). In contrast, the clinicians' DB uses clinical judgment for staging rather than predefined rule. Similar mechanisms produced discordances in pathology documentation.

Given the differences between hospital-based cancer registry and site-specific cancer registry revealed in the study, a DB should be chosen according to the purpose of the analysis. The hospital-based cancer registry DB provides a large amount of detail on pathological types with ICD-O-3 coding, and covers both medical and surgical cases but lacks information specific to breast cancer, such as surgical method (breast conserving vs. mastectomy) and hormone receptor status. In contrast, the clinicians' DB provides detailed clinical information, but only contains surgical cases. Care must be taken when interpreting clinical stages with neo-adjuvant therapy because the clinicians' DB may document the pre-surgery staging as the clinical stage. It may be more appropriate to use the hospital-based cancer registry to analyse medical aspects of care or detailed pathology across cancers. However, surgery-related research questions appear to be better served by the clinicians' DB.

Our study has several limitations. First, all overlapping cases were proved to be surgical cases. If the clinicians' DB starts collecting data on medical cases, new sources of discrepancy may arise and warrant examination. Second, we only studied patients who were diagnosed with breast cancer in a highly specialized cancer hospital with many tumour registrars and breast surgeons, which may not be generalizable to other facilities. Our level of concordance may be overestimated in comparison to average Japanese cancer hospitals. Third, this study focused on patients diagnosed in 2004. Registry items and manuals are updated frequently. Also, the ability of tumour registrars may have improved. Progressive changes in the accuracy of DBs warrant evaluation.

In conclusion, our study found that data were generally accurate in both registries. However, important differences in the scope of registry, documented therapies, and timing of registration were highlighted, which may affect the results of research analyses. These data sets must be used with attention to the definition of collected items. As both data sets have characteristics unique to their purposes, integration of these two systems will require cautious standardization of items with collaboration from both sides.

Acknowledgements

Dr Mingji Zhang has received a Research Resident Fellowship from the Foundation for Promotion of Cancer Research (Japan) for the Third Term Comprehensive 10-year Strategy for Cancer Control. We thank the staff of the Division of Breast Surgery and Hospital Information Services at the National Cancer Center Hospital for their cooperation in this study.

Conflict of interest

None to declare.

References

1. Japanese Association of Cancer Registries (2007) Handbook on Population-Based Cancer Registration in Japan, fifth revision, 2–3. Tokyo: Japanese Association of Cancer Registries.
2. Ministry of Health, Labour and Welfare (2006) *Cancer Control Act*. Available at: http://law.e-gov.go.jp/announce/H18HO098.html (last accessed 8 June 2010).
3. Center for Cancer Control and Information Services, National Cancer Center (2008) *Manual of Hospital-Based Cancer Registry*, 7–10.
4. Bickell, N. A. & Chassin, M. R. (2000) Determining the quality of breast cancer care: do tumor registries measure up? *Annals of Internal Medicine*, 132 (9), 705–710.
5. Warren, J. L. & Harlan, L. C. (2003) Can cancer registry data be used to study cancer treatment? *Medical Care*, 41 (9), 1003–1005.
6. Cress, R. D., Zaslavsky, A. M., West, D. W., Wolf, R. E., Felter, M. C. & Ayanian, J. Z. (2003) Completeness of information on adjuvant therapies for colorectal cancer in population-based cancer registries. *Medical Care*, 41 (9), 1006–1012.

7. Mariotto, A., Feuer, E. J., Harlan, L. C., Wun, L. M., Johnson, K. A. & Abrams, J. (2002) Trends in use of adjuvant multi-agent chemotherapy and tamoxifen for breast cancer in the United States: 1975–1999. *Journal of the National Cancer Institute*, 94 (21), 1626–1634.
8. Harlan, L. C., Clegg, L. X. & Warren, J. L. (2003) Chemotherapy in women with breast cancer. *Annals of Internal Medicine*, 139, 868.
9. Du, X. L., Key, C. R., Dickie, L., Darling, R., Delclos, G. L., Waller, K. & Zhang, D. (2006) Information on chemotherapy and hormone therapy from tumor registry had moderate agreement with chart reviews. *Journal of Clinical Epidemiology*, 59 (1), 53–60.

JJCO Japanese Journal of Clinical Oncology

Jpn J Clin Oncol 2010;40(11)1021–1030
doi:10.1093/jjco/hyq091
Advance Access Publication 9 June 2010

Evaluation and Revision of Checklists for Screening Facilities and Municipal Governmental Programs for Gastric Cancer and Colorectal Cancer Screening in Japan

Takahiro Higashi[1], Ryoko Machii[2], Ayako Aoki[2], Chisato Hamashima[2] and Hiroshi Saito[2,*]

[1]Department of Public Health/Health Policy, University of Tokyo and [2]Cancer Screening Research Division, Research Center for Cancer Screening and Prevention, National Cancer Center, Tokyo, Japan

*For reprints and all correspondence: Hiroshi Saito, Cancer Screening Research Division, Research Center for Cancer Prevention and Screening, National Cancer Center, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan

Received December 12, 2009; accepted May 7, 2010; published online June 9, 2010

Objective: To evaluate the appropriateness of current checklists created by a governmental committee to assess screening programs run by municipal governments and service provider facilities for gastric and colorectal cancer, and to accumulate expert opinions to provide insights aimed at the next revision.
Methods: We convened an expert panel that consisted of physicians nominated by regional offices of the Japanese Society for Gastrointestinal Cancer Screening and radiology technicians nominated by the technician chapter of the society. The panel rated the appropriateness of each checklist item on a scale of 1–9 (1, extremely inappropriate; 9, extremely appropriate) twice, between which they had a face-to-face discussion meeting. During the process they were allowed to propose modifications and additions to the items.
Results: In the first round of rating, the panelists rated all 57 and 56 checklists items for gastric and colorectal cancer, respectively, as appropriate based on an acceptance rule determined *a priori*. During the process of the face-to-face discussion, however, the panel proposed modifications to 23 (40%) and 22 (39%) items, respectively, and the addition of 27 new items each. After integrating overlapping items and rating again for appropriateness, 66 and 64 items, respectively, were accepted as the revised checklist set.
Conclusions: The expert panel considered current checklists for colorectal and gastric cancer-screening programs and facilities to be suitable. Their proposals for a new set of checklist items will help further improve the checklists.

Key words: gastric cancer – colorectal cancer – quality assurance – mass screening

INTRODUCTION

Cancer screening aims to detect cancer in the early stage and to reduce mortality. To achieve this goal, screening programs must be organized properly, from the recruitment of eligible populations to the provision of effective screening tests, and referral to complete diagnostic evaluation when necessary.

In Japan, reliable evidence has confirmed that screening methods for gastric and colorectal cancer, among the most prevalent cancer types in Japan, reduce mortality (1–3). The Health and Medical Services Law for the Aged introduced screening programs for these cancers for all residents aged 40 years or older in 1983 and 1992, respectively. The need for major quality control improvements in these screenings has since been identified, however: currently, only 29 and 25% of eligible population receive screening for gastric cancer and colorectal cancer, respectively (4), whereas only 55–79% of examinees with positive screening results subsequently undergo complete diagnostic evaluation (5,6). Further, screening (4,7) and complete diagnostic evaluation rates (5,6) vary widely across geography and screening settings.

One reason for suboptimal quality control is the fragmented delivery system and variation in cancer-screening programs. In Japan, cancer-screening programs are managed by municipal governments and employers. Although employers provide screening services as an employment benefit to

their employees, municipal governments organize population-based screening programs within their municipalities. Each municipal government contracts with one or more local screening facilities for the provision of services to local residents. Depending on the contract, service providers may specialize in mass screening, or be a local medical association which provides individualized screening services to residents simultaneously with conventional care of sick patients. Funding mechanisms for screening are independent of the health insurance provided for usual medical services.

Although the fragmented nature of this system hampers quality control efforts, improvement is nevertheless dependent on effective quality control. In 2007, the government issued a set of 'Checklists for Evaluation of Cancer Screening Programs' for breast, cervical, colorectal, gastric cancer screening (8). Three sets of checklists were published, one each targeting prefectural governments, municipal governments and screening facilities. These governments and screening facilities are expected to be soon held accountable for the quality of the screening services they provide.

Improvements in the use of these checklists require continuous revision and updating based on feedback from frontline providers working at the local government or individual screening facility level. In this study, we aimed to evaluate the current checklists for gastric cancer and colorectal cancer screening, and to develop a new set of checklists which incorporates the views of frontline physicians and technicians.

PATIENTS AND METHODS

Original Checklist

The checklists ask whether the cancer-screening facilities, municipal government programs and prefectural governments meet the standard procedures and structural criteria (8,9). We did not examine the checklists for prefectural governments because the checklist items for municipal and prefectural governments greatly overlapped, and because the role of the prefectural governments is generally limited to supervising the municipal governments in organizing the programs. Municipal governments play a role in organizing the screening programs through contracting with screening service providers. They also determine the extent of services covered financially in accordance with the requirements of the Ministry of Health, Labour and Welfare. The current checklists are presented in the left columns of Tables 1 and 2. The gastric and colorectal cancer checklists for municipal governments are identical except for one item, which asks whether the government records the names of fecal occult blood tests. In contrast, since the officially recommended screening methods for gastric and colorectal cancer differ—photofluorography tests for gastric cancer and fecal occult blood tests for colorectal cancer—the checklists for screening facilities included both common and different domains. Common domains were 'Explanation to examinees' and 'Quality control systems'. In contrast, items for gastric cancer-screening providers included domains of 'Quality control of history taking and imaging' and 'Quality control of X-ray reading', whereas those for colorectal cancer-screening providers had domains of 'Quality control of screening tests' and 'Handling of samples'.

Composition of the Expert Panel

To examine the appropriateness of the current checklist and develop the revision, we invited 9 and 10 expert panelists for gastric cancer and colorectal cancer, respectively, as nominated by the regional offices and the technician chapter of Japanese Society of Gastrointestinal Cancer Screening. The gastric cancer panel consisted of six physicians and two technicians, and the colorectal cancer panel consisted of seven physicians and two technicians. Since many physicians provide screening for both, four physicians served in both panels.

Evaluation of Appropriateness and Revision of the Current Checklist

The panelists examined the current checklist and proposed a revised version through a formal process derived by adopting the RAND/UCLA appropriateness method (10). This method consists of convening a panel of multidisciplinary experts and following a formal process of reviewing proposed items using discussion as well as pre- and post-rating of the items. It is expected that the composition of the panel and encouragement of evaluation in a non-threatening environment through a concealed rating process will allow the integration of various perspectives. Further, the multiple-step rating and discussion of items will ensure that efforts spent in evaluating the items are pertinent. This method is extensively used to assess the appropriateness of indications for surgery and other invasive procedures (e.g. cardiac surgery and angiography) (11–13) and to develop quality indicators for a variety of medical and surgical conditions (14–16).

We first mailed the current checklists along with the rating scale of 1–9 to the individual panelists. The panelists individually rated the appropriateness of each checklist item on a scale of 1–9, with a low score indicating an inappropriate and a high score indicating an appropriate item. Panelists were encouraged to indicate in the designated space on the rating sheet whether they felt that an item required revision, or whether a new item was required. We then convened a face-to-face discussion meeting of all panelists, at which they were provided with tables presenting the distribution, median and summary of the level of agreement from the first-round ratings and the list of comments written with the ratings. Reviewing these materials, they discussed the checklist items one by one, including the newly proposed items.

Table 1. Expert panel ratings of checklists for gastric cancer screening provider facilities

	Current checklist item	Median, agreement, range	Revised checklist item	Median, agreement, range
1	Explanation for examinees			
(1)	Does the provider clearly inform examinees in advance that they will not be able to forego a complete diagnostic evaluation if such an evaluation is considered necessary?	8, I, (1–9)	Does the provider clearly inform all examinees in advance through personal interview, explanation meetings, or brochures or leaflets that they will not be able to forego a complete diagnostic evaluation if such an evaluation is considered necessary, and the details of diagnostics tests?	9, A, (7–9)
(2)	Does the provider explain to examinees the types and methods of the complete diagnostic evaluation	7, I, (1–9)	Does the provider explain to examinees who require complete diagnostic evaluation the types and methods of the complete diagnostic evaluation?	9, A, (8–9)
(3)	Does the provider adequately explain to examinees the policy of reporting their complete diagnostic evaluation results to the municipal government?	7, I, (1–9)	*No change*	9, A, (1–9)
(4)			Does the provider inform examinees in advance of possible complications caused by the screening test (e.g. constipation from barium)?	7, D, (1–9)*
(5)			Does the provider inform examinees in advance of the benefits and harm associated with screening?	1, I, (1–9)*
2	Quality control of history taking and imaging			
(1)	Does the service include history taking and photofluorography examination?	9, A, (5–9)	*No change*	9, A, (9–9)
(2)	Does the provider ask examinees to report past, present and family histories, and the prior receipt of screening?	8, A, (5–9)	*No change*	9, A, (7–9)
(3)	Does the provider maintain records of history for at least five years?	9, A, (5–9)	*No change*	7, I, (1–9)*
(3)′			Does the provider maintain records of history for at least three years?	8, I, (1–9)
(4)	Does the provider publicly report information on the type of imaging devices used (direct/indirect/DR, use of image intensifier, etc.)? As a rule, providers should use an image intensifier on files 10 cm × 10 cm or larger.	8, A, (5–9)	Is the provider able to clearly report the type of imaging devices used (direct/indirect/DR, use of image intensifier, etc.) in response to a request by a supervising body, such as a prefectural government?	9, A, (7–9)
(5)	Does the provider take at least seven pictures?	9, A, (7–9)	*Deleted*	
(6)	Does the provider make its position and imaging methods conform to the style set by the Japanese Society of Gastroenterological Cancer Screening (JSGCS)?	8, A, (7–9)	Does the provider make its position and imaging methods conform to the style set by the Japanese Society of Gastroenterological Cancer Screening (with modifications where necessary) ?	9, A, (7–9)
(7)	Does the provider ensure the appropriate concentration of contrast medium (180–220 W/V%), and pay adequate attention to complications?	9, A, (7–9)	*Deleted*	
(7)′			Does the provider maintain records of complications caused by the screening test?	9, A, (7–9)
(8)	Do the X-ray technicians complete the training conducted by the JSGCS?	8, A, (5–9)	Do the X-ray technicians complete the training conducted by the JSGCS or Japan NPO Organization for Quality Control Management in Gastroenterological Cancer Screening?	9, A, (7–9)
(9)	Does the provider report the number of X-ray technicians, both in total and those certified by the JSGCS?	8, I, (5–9)	Is the provider able to report the number of X-ray technicians, both in total and those certified by the JSGCS, in response to a request by a supervising body such as the prefectural government?	9, A, (7–9)

Continued

Table 1. *Continued*

	Current checklist item	Median, agreement, range	Revised checklist item	Median, agreement, range
3	Quality control of X-ray reading			
(1)	Does the provider report the number of X-ray reading physicians, both in total and those certified by the Japanese Society of Gastroenterological Cancer Screening?	8, I, (5–9)	Is the provider able to report the number of X-ray reading physicians, both in total and those certified by the JSGCS, in response to a request by a supervising body such as the prefectural government?	9, A, (7–9)
(2)	Is the reading performed by two or more physicians, at least one of whom is certified by the JSGCS? Depending on the finding of the primary reading, are comparisons made with previous images?**	8, A, (2–9)	Is the reading performed by two or more physicians, at least one of whom is certified by the JSGCS?**	9, A, (7–9)
(2)′			Depending on the finding of the primary reading, are comparisons made with previous images?**	7, I, (5–9)
(3)	Does the provider maintain photofluorography images for at least three years?	9, A, (2–9)	*No change*	9, D, (1–9)*
(3)′			Does the provider keep photofluorography images for at least five years?	8, I, (5–9)
(4)	Does the provider keep the results of the screening tests for at least five years?	9, A, (7–9)	*No change*	9, A, (5–9)
(4)′			Does the provider keep the results of the screening tests for at least three years?	2, I, (1–9)*
4	Quality control systems			
(1)	Does the provider receive the results of final complete diagnostic evaluations and treatments from the medical facilities that conducted these complete diagnostic evaluations and treatments?	9, A, (2–9)	Does the provider track the results of final complete diagnostic evaluations for their examinees?**	9, A, (8–9)
(1)′			Does the provider track the details of examinees diagnosed with cancer (treatment, stage)?**	9, A, (2–9)
(2)	Does the provider have a review committee or conference that includes third-party specialists in gastric cancer?	7, A, (5–9)	Does the provider have a review committee or conference that includes third-party specialists in gastric cancer aimed at improving diagnostic accuracy?	8, A, (7–9)
(3)	Does the provider submit sufficient data to allow the prefectural government to evaluate process indicators (screening rates, positive screening rate, cancer detection rate, and positive predictive values)?	9, A, (2–9)	Does the provider submit sufficient data to allow prefectural and municipal governments to calculate screening rates, positive screening rates, cancer detection rates, and positive predictive values?	9, A, (9–9)
(4)	Does the provider use the items required for the Community Health and Health Promotion Project Report for its statistical reports?	9, A, (2–9)	Does the provider use the items required for the Community Health and Health Promotion Project Report in its report to the municipal government?	9, A, (8–9)
(6)			Does the provider report the results to examinees (or to the municipal government if the results are processed by the government) within four weeks after the screening test was performed?	9, A, (7–9)
5	Tracking of complete diagnostic examination results			
(1)′			Does the provider keep the results of complete diagnostic evaluations for five years?	9, A, (4–9)
(2)′			Does the provider keep the results of complete diagnostic evaluations for three years?	9, D, (1–9)*

A indicates agreed; D, disagreed; and I, indeterminate.
*Not accepted because of low median rating or disagreement.
**Divided into two items.

Table 2. Expert panel ratings of checklists for colorectal cancer screening provider facilities

	Current checklist item	Median, agreement, range	Revised checklist item	Median, agreement, range
1	Explanation to examinees			
(1)	Does the provider clearly inform examinees in advance that they need to have a complete diagnostic examination with colonoscopy?	8, I, (1–9)	Does the provider clearly inform all examinees in advance through personal interview, explanation meetings, or brochures or leaflets that they cannot forego complete diagnostic evaluation with colonoscopy?	9, A, (7–9)
(2)	Does the provider explain to examinees the types and methods of the complete diagnostic evaluation (colonoscopy or barium enema)?	7, I, (1–9)	Does the provider explain to examinees who require complete diagnostic evaluation the types and methods of the complete diagnostic evaluation (basically colonoscopy)?	9, A, (7–9)
(3)	Does the provider adequately explain to examinees the policy of reporting their complete diagnostic evaluation results to the municipal government?	7, I, (1–9)	*No change*	9, A, (1–9)
2	Quality control of screening test			
(1)	Does the provider hold regular education sessions and/or workshops for technicians?	7, I, (2–9)	Does the provider have their technicians attend education sessions and/or workshops regularly?	7, A, (7–9)
(2)	Does the provider use two-day methods for FOBT?	9, A, (7–9)	*No change*	9, A, (9–9)
(3)	If the provider scales FOBT quantitatively, does it explicitly set the cut-off value?	9, A, (6–9)	Does the provider record FOBT methods (name of the FOBT kit) and explicitly set the cut-off value?	9, A, (9–9)
(4)	Does the provider follow the procedure described in the *Colorectal Cancer Screening Manual* (1992)?	9, A, (1–9)	*Deleted because the manual is out of print.*	
3	Handling of samples			
(1)	Does the provider prepare a pamphlet that clearly explains to examinees how to obtain a stool specimen	9, A, (8–9)	*No change*	9, A, (9–9)
(2)	Does the provider instruct examinees to submit the specimen immediately after they obtain the second sample?	8, A, (7–9)	*No change*	9, A, (8–9)
(3)	Does the provider instruct examinees to keep the stool sample in a cold dark place?	9, A, (2–9)	Does the provider instruct examinees to keep the stool sample in a cold dark place, such as in a styrene foam container with cooling agents?	9, A, (8–9)
(4)	Does the provider store the specimen in a refrigerator after receiving it from the examinee until it sends it to the laboratory?	9, A, (7–9)	*No change*	9, A, (9–9)
(5)	Does the provider store the specimen in a refrigerator until it sends it to the laboratory?	9, A, (1–9)	*Deleted because of overlap with (4)*	
(6)	Does the provider test the specimen within 24 hours after receipt?	9, A, (3–9)	*No change*	9, A, (8–9)
(7)	Does the provider report the result to the municipal government within 2 weeks of the receipt of the specimen?	7, I, (1–9)	Does the provider report the results to examinees (or to the municipal government if results are processed by the government) within four weeks after the screening test is performed?	9, A, (8–9)
(8)	Does the provider keep the test results for at least five years?	9, A, (7–9)	*No change*	9, A, (2–9)
(9)			Does the provider keep the test results for at least three years?	7.5, D*, (1–9)
(10)			Does the provider record on the report sheet the test results for the past three years?	2, I*, (1–9)

Continued

Table 2. *Continued*

	Current checklist item	Median, agreement, range	Revised checklist item	Median, agreement, range
(11)			Does the provider record the quantity of hemoglobin so that it can report the values in response to a request from prefectural and/or municipal governments?	8, A, (5–9)
4	Quality control system			
(1)	Does the provider receive the results of complete diagnostic evaluations and treatments from medical facilities that conduct complete diagnostic evaluation and treatments?§	9, A, (7–9)	Does the provider track the results of complete diagnostic evaluations for their examinees?§	9, A, (9–9)
(1)′			Does the provider track the details of examinees diagnosed with cancer (treatment, stage)?§	9, A, (8–9)
(2)	Does the provider submit sufficient data to allow the prefectural government to evaluate process indicators (screening rates, screening positive rate, cancer detection rate, and positive predictive values)?	9, A, (5–9)	Does the provider submit sufficient data to allow prefectural and municipal governments to calculate screening rates, positive screening rate, cancer detection rate, and positive predictive values?	9, A, (9–9)
(3)	Does the provider use the items required for the Community Health and Health Promotion Project Report for its statistical reports?	9, A, (5–9)	Does the provider use the items required for the Community Health and Health Promotion Project Report in its report to the municipal government?	9, A, (9–9)
5	Tracking of complete diagnostic evaluation results			
(1)			Does the provider keep the results of complete diagnostic evaluations for five years?	9, A, (2–9)
(2)			Does the provider keep the results of complete diagnostic evaluations for three years?	8, A, (1–9)**
(3)			Can the provider report the methods and results of complete diagnostic evaluation in response to a request from the prefectural and/or municipal government?	7, I, (3–9)

A indicates agreed; D, disagreed; and I, indeterminate.
*Not accepted due to low median rating or disagreement.
**Not accepted because the competing item had a better rating distribution.
§Divided into two items.

They also proposed modifications when they considered that the wording of an item required improvement. After discussing each checklist item, the panel individually rated the appropriateness of the finalized checklist items again. The ratings of individual panelists were not disclosed to the other panelists unless they chose to do so themselves.

The appropriateness of each checklist item was judged in terms of both its validity and clarity. Each checklist describes the process required to be carried out by the targeted entity (i.e. screening organization or municipal government). If a checklist item is valid, failure to carry out the checklist process indicates that the quality of the screening service for which the evaluated entity is responsible is poor. Although the entity can assign another entity to actually carry out the described process, it is still responsible for ensuring that the process actually happens. For example, the municipal government can request the group of contracted screening facilities to produce a statistical report, but it is still held responsible if the process is not done. Not only the validity but also the clarity of the checklists was examined. Since the compliance of each entity (i.e. municipal government or screening facility) with the checklist items may eventually be publicly disclosed for accountability purposes, it needs to be clearly specified. We asked the expert panel to rate items that were not clearly specified as low, and then suggest a revision to clarify it. The level of panelist agreement was classified as 'agree' if no more than two panelists rated an item outside the 1–3, 4–6 and 7–9 range to which the median rating belonged, and 'disagree' if three or more members rated it 1–3, whereas another three or more rated it 7–9. Distributions that did not fall in either category were considered 'indeterminate'. In compiling the expert ratings, a checklist item was considered appropriate if it had a median rating of 7 or higher and less than two panelists rated it 3 or less.

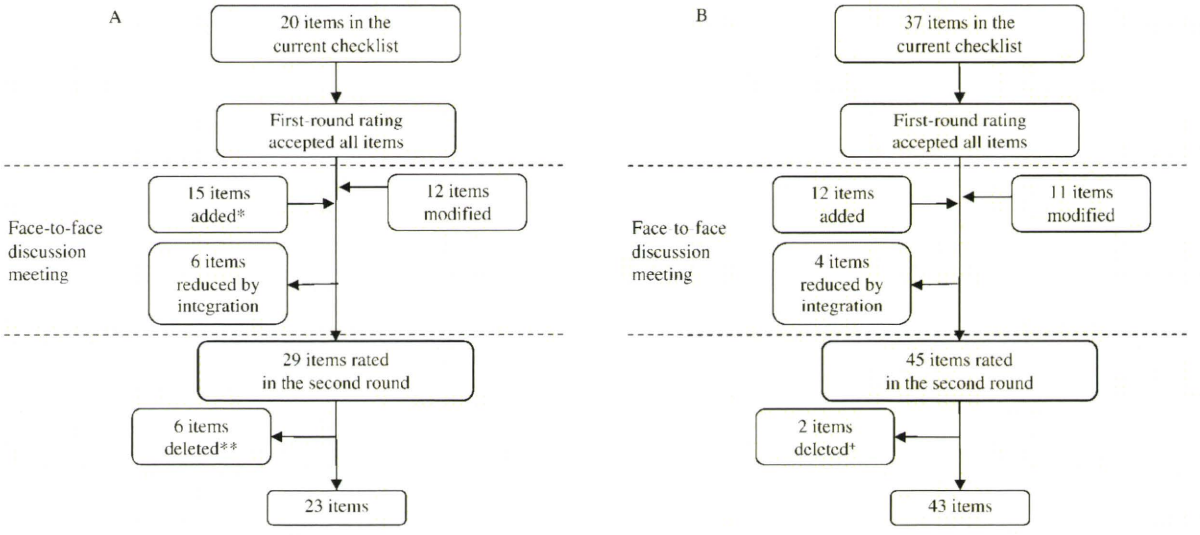

Figure 1. Flow of the rating process of checklist for gastric cancer screening. (A) Screening provider facilities. Single asterisk indicates inclusion of one item split from the original. Double asterisk indicates inclusion of one item that satisfied the acceptance criteria but was deleted because another item with a better rating was accepted. (B) Municipal government. Plus sign indicates inclusion of one item that satisfied the acceptance criteria but was deleted because another item with a better rating was accepted.

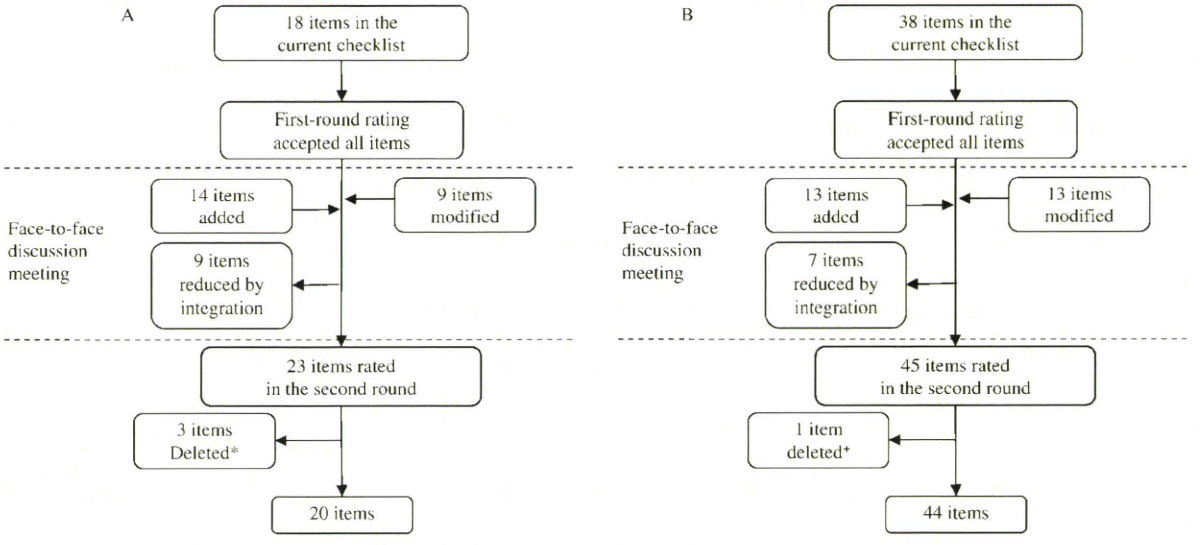

Figure 2. Flow of the rating process of checklist for colorectal cancer screening. (A) Screening provider facilities. Single asterisk indicates inclusion of one item that satisfied the acceptance criteria but was deleted because another item with a better rating was accepted. (B) Municipal government. Plus sign indicates items satisfying the acceptance criteria, but deleted because another item with a better rating was accepted.

RESULTS

The overall processes of checklist rating for both gastric and colorectal cancer are shown in Figs 1 and 2, respectively. Tables 1 and 2 present the first-round ratings for current checklist items before the discussion session (left columns) and the second-round ratings for the items after revision during the discussion for screening provider facilities (right columns). Supplementary Table S1 (available only as an online supplement) presents the ratings for the municipal government programs. The items for municipal governments for gastric and colorectal cancer screening are mostly common, and are thus presented in a single table. The numbers presented in the following text show the numbers totaled for both municipal government and screening provider facilities for each cancer.

Confirming the appropriateness of the current checklists, all 57 and 56 checklist items for gastric and colorectal cancer, respectively, were rated 7 or higher by more than half of the panelists, and no item was rated 3 or less by more than two panelists. However, the levels of agreement for the checklists for screening provider facilities were indeterminate for 5 of 20 (25%) for the gastric cancer facilities and 5 of 18 items (28%) for the colorectal screening