Barriers to Physician Adherence to Practice Guidelines

- □ Lack of Awareness
- □ Lack of Familiarity
- □ Lack of Agreement
- □ Lack of Outcome Expectancy
- □ Lack of Self-Efficacy
- □ Inertia of Previous Practice
- □ External Barriers

Cabana MD *et al. JAMA* 1999; 282: 1458–1465

41

標準的初回化学療法のオプション

ドセタキセル＋カルボプラチン （DC 療法）	•ドセタキセル：60～70mg／m2 •カルボプラチン：AUC5～6 静注，day1，3～4週間隔で投与
パクリタキセル＋カルボプラチン 毎週投与法（weekly-TC療法）	•パクリタキセル：60～80mg／m2，day1，8，15 •カルボプラチン：AUC 6，day1，3週間隔で投与 またはAUC=2，day1，8，15，4週間隔で投与
イリノテカン＋シスプラチン （CPT-P療法）	•イリノテカン：60mg／m2，day1，8，15 •シスプラチン：60mg／m2，day1 静注，4週間隔で投与
シクロホスファミド＋ ドキソルビシン＋シスプラチン （CAP療法）	•シクロホスファミド：500mg／m2 •ドキソルビシン：30～50mg／m2（ピラルビシン：30mg／m2，エピルビシン：50mg／m2） •シスプラチン：50～75mg／m2 静注，day1，3～4週間隔で投与
シクロホスファミド＋シスプラチン （またはカルボプラチン） （CP療法またはCC療法）	•シクロホスファミド：800～900mg／m2 •シスプラチン：60～75mg／m2（カルボプラチン：AUC5～6） 静注，day1，3～4週間隔で投与
シスプラチン単剤 または カルボプラチン単剤	•シスプラチン：75～100mg／m2 または •カルボプラチン：AUC5～6，静注，day1，3～4週間隔で投与

42

腫瘍専門医と化学療法

- duBois *et al.*
  - 臨床試験参加病院での治療は標準化学療法を受ける機会が増す

    Int J Gynecol Cancer 2005; 15: 183–191.

- Cress RD *et al.*
  - 婦人科腫瘍医による治療は化学療法を受ける機会が増す

    J Clin Oncol 2003; 21(8) 1530-1535

結語

- DPCデータの臨床情報には制約があるため、病状との関連についての解析は困難であるが、治療の個別化が進んでおらず、化学療法の領域で効果の高い標準治療が確立している疾患では、Administrative dataはより多くの施設を対象に標準化の実態調査が比較的簡便に可能である
- 本研究はDPCデータの分析がEBMの普及状態、ガイドラインの周知状況等の検討に有用であり、今後の医療向上のための示唆を与え得ることを示している

International Journal for Quality in Health Care 2009; Volume 21, Number 5: pp. 372–378
Advance Access Publication: 22 August 2009

10.1093/intqhc/mzp037

Change in clinical practice after publication of guidelines on breast cancer treatment

HARUHISA FUKUDA[1,2], YUICHI IMANAKA[2], TATSURO ISHIZAKI[2], KAZUHIDE OKUMA[2] AND TAKAKO SHIRAI[2]

[1]Institute for Health Economics and Policy, 1-5-11-2F, Nishi-Shinbashi, Minato-Ku, Tokyo 105-0003, Japan, and [2]Department of Healthcare Economics and Quality Management, School of Public Health, Kyoto University Graduate School of Medicine, Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan

Abstract

Objective. Several studies raise questions about whether clinical practice guidelines actually guide practice. We evaluated patterns of use of breast-conserving surgery (BCS) over time to examine the effect of guideline publication.

Design. Retrospective analysis of time-series data on breast cancer treatment. Multiple logistic regression analysis was performed, adjusting for covariates including the patient's age, comorbidity status and admission year, to assess whether the use of BCS was higher after publication of treatment guidelines.

Setting. Five teaching hospitals participating in the Quality Improvement/Indicator Project (QIP) in Japan.

Participants. Female breast cancer patients who received surgical treatment at five teaching hospitals from January 1996 through December 2007 ($n = 2199$).

Main Outcome Measure. Rates of use of BCS.

Results. The proportion of BCS use increased from 26.4% before guideline publication to 59.9% after guideline publication in Japan. After controlling for other characteristics, the use of BCS has increased significantly over time, especially since 2001. Women aged 70 years and older ($P=0.004$) and those with any comorbidity ($P < 0.001$) were significantly less likely to receive BCS.

Conclusions. This study demonstrated that the adjusted proportion of BCS has increased dramatically since 2001, 2 years after guideline publication in Japan and this is consistent with a relationship between guideline publication and a change in this clinical practice.

Keywords: breast cancer, breast-conserving surgery, guideline, practice variations

Introduction

Clinical practice guidelines are defined as systematically developed statements to assist practitioners' and patients' decisions about appropriate health care for specific clinical conditions [1]. Successful implementation of clinical practice guidelines should improve quality of care by decreasing inappropriate practice variation and promoting reasonable decisions for effective practice [2]. A consistent finding in health services research, however, is the gap between scientific evidence and actual clinical practice [3]. Past studies suggest that 30–40% of patients do not receive current scientific evidence-based care [4]. These findings raise questions about whether clinical practice guidelines actually guide practice.

Randomized clinical trials performed in the 1980s demonstrated that women with early-stage breast cancer treated with either modified radical mastectomy or breast-conserving surgery (BCS) followed by radiation therapy had equivalent rates of survival and recurrence [5, 6]. In 1990, a National Institute of Health (NIH) Consensus Development Conference recommended BCS followed by radiation therapy for the majority of women with Stage I or II breast cancer [7]. After publication of the guideline, the use of BCS was statistically significantly higher, but not dramatically so [8–11].

These prior studies [8–11] cannot definitively separate effects of the 1990 NIH Consensus Development Conference from earlier events that might also increase the use of BCS; results of earlier clinical trials [5, 6] and popular

Address reprint requests to: Yuichi Imanaka, Department of Healthcare Economics and Quality Management, School of Public Health, Kyoto University Graduate School of Medicine, Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan. Tel: +81-75-753-4454; Fax: +81-75-753-4455; E-mail: imanaka-y@umin.net

publications may also have stimulated the increased use of BCS. Therefore, simple comparison of BCS use in the Western countries before and after the Consensus may not be appropriate. To separate the effects of Consensus recommendations on the use of BCS from the effect of scientific publications, it is necessary to account for scientific publications.

In July 1999, a steering committee convened by The Japanese Breast Cancer Society published evidence-based clinical practice guidelines for treatment of early-stage breast cancer in Japanese women [12]. To our knowledge, these Japanese guidelines were based on clinical evidence and existing guidelines from Western countries and not on the results of randomized clinical trials or re-analyses of previous studies that have evaluated the efficacy of BCS in Japan. In addition, because of language barriers, several large clinical trials published in Western countries seemed to have less impact on knowledge of the effectiveness of BCS in Japan compared with the impact in English-speaking countries [5, 6]. Before the publication of the Japanese guideline, therefore, it was possible that Japanese women might be unaware of this treatment choice. Because of this unique situation in Japan, any change in the use of BCS before and after guideline publication in Japan might reflect the effect of consensus recommendations rather than other factors. The aim of this study was to evaluate whether publication of clinical guidelines was associated with a change of treatment practices for breast cancer patients through the use of secondary administrative data from Japanese hospitals.

Methods

Database

We used a database from the Quality Indicator/Improvement Project (QIP), which includes more than 10 privately owned teaching hospitals in Japan. These hospitals are located in urban cities in Hokkaido (in north Japan), throughout Honshu (the main island of Japan), and in Kyushu (in south Japan). Records of all patients discharged from these hospitals have been kept since 1995. In this particular study, we selected 5 out of 10 tertiary care community hospitals that had more than 50 breast cancer patients between 1996 and 1999 [13], and have submitted medical claims data since 1996–2007. In 2000, the average number of general beds in the five hospitals was 671 (range 450–925). All had full-time surgeons and four out of five had radiotherapy machines. The average number of surgeons who worked in the hospitals was 10.6 (range 7–13) in 2000. All five hospitals were private teaching hospitals for board certification of general surgeons accredited by the Japan Surgical Society. This study was approved by the Institutional Review Board of Kyoto University Graduate School of Medicine Faculty of Medicine in Japan.

Study population

Because the QIP database involved all inpatient admissions, patients with either primary or recurrent breast cancer were included. A total of 2199 women who received either BCS or mastectomy between January 1996 and December 2007 were selected. Breast cancer surgeries are almost never performed in an outpatient setting in Japan. The database contains all clinical procedures with medical claims as well as demographic data on inpatients. Patients who underwent both surgeries and patients with distant metastasis were not included. The International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) and International Classification of Diseases, Tenth Revision (ICD-10) were used to identify women with breast cancer (ICD-9-CM codes 174.0–174.9 and ICD-10 codes C50) treated by either BCS (ICD-9-CM codes 85.20–85.23) or mastectomy (ICD-9-CM codes 85.41–85.48).

Statistical analysis

Data were analysed using either the Student's *t*-test for continuous variables or the χ^2 test for categorical variables. We examined trends in the use of BCS between admission years by using a non-parametric test for trend. A multiple logistic regression analysis with robust standard errors (robust cluster) was performed to identify trends in the use of BCS over time, after adjusting for potential covariates. The dependent variable was the type of surgical procedure. Independent variables were the patient's age, comorbidity status and time period of admission. The hospital in which each patient was admitted was defined as a cluster. We used an adaptation of the Charlson comorbidity index to assess comorbidities [14, 15]. A patient was identified as having comorbidity if any of comorbidities were coded in the diagnosis (present vs. absent). All analytical procedures were performed using the STATA 9.2 statistical package (StataCorp. College Station, TX, USA). All reported *P*-values were two-tailed, and level of significance was $P < 0.05$.

Results

Table 1 shows the distribution of characteristics in patients treated by either BCS or mastectomy before or after guideline publication. In both time periods, women who underwent BCS were significantly younger than those who underwent mastectomy ($P = 0.001$ and <0.001, respectively). For women with BCS, comorbidity status was not significantly different between types of surgery before guideline publication ($P = 0.49$), whereas the difference in comorbidity status was statistically significant after guideline publication ($P < 0.001$).

Table 2 represents trends in BCS use among breast cancer patients with surgical treatment diagnosed before and after guideline publication. The percentage of patients receiving BCS almost doubled between the two time periods ($P < 0.001$). In all subgroups of women classified by age, the percentage receiving BCS increased substantially in the later time period. In addition, women with and without comorbidity had more BCS in the later time period ($P < 0.05$ and <0.001, respectively).

Table 1 Characteristics of breast cancer patients who underwent either BCS or mastectomy at five teaching hospitals ($n = 2199$)

	Guideline publication (1999)					
	Before publication (1996–99)			After publication (2000–07)		
	BCS	Mastectomy	*P*-value	BCS	Mastectomy	*P*-value
Age [mean (SD)]	53.4 (12.4)	57.3 (13.2)	0.001[a]	55.9 (13.3)	59.1 (13.8)	<0.001[a]
Age group (%)			0.006[b]			<0.001[b]
<50	46.9	34.3		38.2	28.8	
50–59	25.3	23.0		26.0	25.7	
60–69	13.0	21.7		19.2	21.3	
70+	14.8	21.0		16.6	24.3	
Comorbidity (%)			0.49[b]			<0.001[b]
Absent	95.7	94.3		91.6	83.3	
Present	4.3	5.8		8.4	16.7	
Total [*n* (%)]	162 (100%)	452 (100%)		950 (100%)	635 (100%)	

[a]Student's *t*-test. [b]χ^2 test.

Table 2 Percentage of the use of BCS before and after guideline publication by age group, comorbidity and hospital among breast cancer patients ($n = 1112$)

	Guideline publication (1999)				*P*-value[a]
	Before publication (1996–99)		After publication (2000–07)		
	n	% BCS	*n*	% BCS	
Total	162	26.4	950	59.9	<0.001
Age group					
<50	76	32.9	363	66.5	<0.001
50–59	41	28.3	247	60.2	<0.001
60–69	21	17.7	182	57.4	<0.001
70+	24	20.2	158	50.6	<0.001
Comorbidity					
Absent	155	26.7	870	62.2	<0.001
Present	7	21.2	80	43.0	<0.05
Hospital					
Hospital A	48	50.5	44	41.1	0.18
Hospital B	20	28.2	109	60.6	<0.001
Hospital C	50	41.3	115	73.7	<0.001
Hospital D	22	9.1	618	64.9	<0.001
Hospital E	22	25.9	64	33.7	0.20

[a]χ^2 test.

Hospitals demonstrated wide variation in BCS use. Before guideline publication, Hospital A had the highest BCS use (50.5%), whereas Hospital D had the lowest (9.1%). The pattern of increased BCS after guideline publication differed between hospitals. Change in BCS at Hospitals A and E was not statistically significant, whereas the other three hospitals showed significant increases in BCS ($P < 0.001$). After guideline publication, an increase in BCS at Hospital D, which had exhibited the lowest use of BCS before guideline publication, was the second highest among all hospitals analysed.

Table 3 shows the results of a multiple logistic regression analysis to identify factors associated with utilization of BCS among all breast cancer patients who received surgical treatment in five hospitals analysed. For sensitive detection of changes after the 1999 guideline publication, we divided the 12-year interval annually, rather than into pre- and post-publication periods. A multiple logistic regression analysis demonstrated that the use of BCS has increased dramatically since 2001.

The use of BCS demonstrated linear correlation with patient age; younger women received more BCS than did

Table 3 Factors associated with the use of BCS among breast cancer patients ($n = 2199$)

Parameter	Unadjusted			Adjusted[a]		
	Odds ratio	95% CI	*P*-value	Odds ratio	95% CI	*P*-value
Age group						
<50	1.00			1.00		
50–59	0.83	0.67–1.03	0.096	0.78	0.54–1.13	0.19
60–69	0.67	0.53–0.85	0.001	0.65	0.36–1.16	0.14
70+	0.56	0.44–0.71	<0.001	0.56	0.38–0.83	0.004
Comorbidity						
Absent	1.00			1.00		
Present	0.61	0.46–0.82	0.001	0.51	0.43–0.60	<0.001
Admission year[b]						
1996	1.00			1.00		
1997	1.42	0.79–2.57	0.25	1.46	0.99–2.16	0.058
1998	2.68	1.54–4.68	<0.001	2.92	1.67–5.08	<0.001
1999[c]	2.53	1.43–4.49	0.001	2.60	1.44–4.72	0.002
2000	3.31	1.82–5.99	<0.001	3.67	2.31–5.81	<0.001
2001	6.27	3.22–12.20	<0.001	6.68	4.57–9.77	<0.001
2002	7.98	4.53–14.04	<0.001	8.55	3.88–18.84	<0.001
2003	7.11	3.57–14.16	<0.001	7.63	2.23–26.17	0.001
2004	7.15	4.05–12.65	<0.001	8.06	2.77–23.45	<0.001
2005	9.66	5.77–16.15	<0.001	10.87	4.66–25.33	<0.001
2006	8.89	5.41–14.63	<0.001	9.89	3.22–30.42	<0.001
2007	8.12	4.91–13.44	<0.001	9.61	3.84–24.05	<0.001

CI, confidence interval. [a]A multiple logistic regression analysis with robust standard errors (robust cluster). Hospitals in which each patient admitted were defined as clusters. [b]A non-parametric test for trend across admission years: $P < 0.001$. [c]Year of guideline publication.

older women. The odds ratio (OR) for patients with comorbidities was significantly less than for those lacking comorbidities (OR = 0.51 vs. no comorbidities; 95% confidence interval =0.43–0.60).

Discussion

We observed a significant increase in BCS during 2001–07 (after guideline publication) compared with 1996–99 (before guideline publication) consistent with a positive relationship between guideline publication and change in clinical practice. A change in practice patterns with a time lag (i.e. 2 years in our study) following guideline publication is consistent with the theory of diffusion of innovation [16], a process by which new technology or practices spread through social networks of physicians and patients over time. The trends in the use of BCS over time in this study were consistent with a study based on questionnaire data collected from 352 hospitals in Japan which showed an increase in BCS during 1995–2000, a plateau during 2000–01, followed by a second increase in the period after 2002, by using the crude proportion of BCS [17].

Although the impact of publication of guidelines or recommendations from expert consensus on practice changes is difficult to study in the absence of a control group, our study had two main advantages. First, because of the significant language barrier, it is unlikely that publication of results of several large clinical trials conducted in Western countries in the 1980s would have influenced practice in Japan, making the inference of causation from an epidemiological study more credible [5, 6]. Although the proportion of BCS use in Japan at the time of the Fisher's report [5] was 0.4% [17], BCS use in the USA had increased 16–17% [8, 9] during the same period. The NIH Consensus Development Conference recommended BCS for the majority of women with early-stage breast cancer [7]. Several studies have assessed effects of this recommendation on the use of BCS [8–11]. However, past studies could not definitively separate the effect of the Consensus recommendation from that of prior published clinical trials on BCS.

Second, considerable room for change existed in 1999. Several studies have addressed effects of the Consensus on treatment of breast cancer [8–11, 18], prostate cancer [19] and other conditions [20–22]. Only the breast cancer guidelines appeared to affect physicians' behaviour, perhaps because the practices recommended in guidelines for conditions other than breast cancer had already been adopted prior to publication of the guidelines. Our study demonstrated that only 26.4% of female breast cancer patients underwent

BCS before guideline publication in Japan (Table 2); therefore, there was significant room for practice improvement following guideline publication.

Patient-, hospital- and surgeon-related factors have been associated with the use of BCS. Many previous researchers have considered the patient's age. Our observation that younger patients tended to undergo BCS rather than mastectomy is in accordance with previous reports [8, 23–25]. Several reasons may account for this association. First, younger patients seek more information than do older patients [26–29]. Second, older patients may be more sensitive to physicians' opinions and recommendations than are younger patients [28]. Third, older patients may wish to avoid the inconvenience of long courses of daily radiation therapy and, therefore, may prefer mastectomy [23, 27, 29].

As for other patient-related factors, we also examined whether comorbidity status was associated with type of surgery. In Table 1, the percentage of women who had any comorbidity after 2000 was higher than that before 1999 in both surgery groups, suggesting possible biases in the data. One possible source of bias could be an economic incentive. However, we feel that this may not be relevant. Although all participant hospitals in this study introduced the new per-diem payment system after 2004, which departed from the previously used fee-for-services system, our results show that the proportion of patients who had any comorbidity was lower before 1999 than after 2000. A more probable explanation for this inconsistency may be a difference in the format of our database. Before 2001, there was no column in the database specifically for patients' comorbidities; instead, there was the option of recording a maximum of five separate existing diseases for each patient at the time of admission. We identified the reported pre-existing diseases, excluding breast cancer, as each patient's comorbidities. However, since 2002, providers could input up to four comorbidities in our database.

Despite the variability in our measurement of comorbidity status, our multivariable logistic regression found that women who had any comorbidities at the time of diagnosis were less likely to receive BCS than those with no comorbidities, even after taking patient's age into account. Few previous studies have examined the relationship between comorbidity status or age and BCS use. Ballard-Barbash *et al.* [30] reported that older women and women who had any comorbidities at the time of diagnosis were more likely to undergo BCS without radiotherapy, and also noted that women were more likely to receive radiotherapy followed by BCS, after adjustment for other factors, if they were younger or had fewer comorbidities. Our results were concordant with previous findings [31] that younger women and women who had no comorbidity at the time of diagnosis were more likely to undergo BCS, without consideration of adjuvant radiotherapy status after BCS. There are many possible reasons why BCS use would be used less frequently in women with comorbidities. Women or their physicians may have wanted to avoid the challenge of radiotherapy after BCS due to severe symptoms of comorbidities, or to avoid the challenge of a second surgery if there were a recurrence of disease. Additional research using more detailed clinical data might clarify the relationship.

Also, we observed a variation in the use of BCS among participant hospitals. Percentage of BCS and pattern of increase in BCS differed significantly among five hospitals, perhaps due to surgeon- and hospital-related factors. According to medical literature, high-volume surgeons [24, 26, 29, 32] and female surgeons [24] favoured BCS. Causes of practice variation that were hospital-related included ownership of teaching hospitals [25, 32] and hospital location [24, 26, 32], as well as the type of teaching hospital (teaching hospital vs. non-teaching hospital). Regarding the association of hospital characteristics and BCS proportion in this study, we could not compare BCS use among different types of hospital ownership or reimbursement system because all five hospitals were private teaching hospitals, and had adopted the same reimbursement system at the same year. In terms of volume of surgery, however, our results suggest an association between high-volume hospitals and high proportion of BCS use (Table 2). Further research is necessary to identify reasons for practice variation among hospitals and to estimate the differences in guideline diffusion, such as rapidity and depth among different type of hospitals using a larger sample database.

Some limitations must be considered when interpreting the results of this study. First, since we estimated the impact of the surgery year on the use of BCS and separately assessed the year of guideline publication, this study cannot definitively demonstrate causality between the publication of guideline and practice changes. Second, we do not have information on tumour size, staging or other pathological features of breast cancer for each patient that might explain the choice of BCS, although it seems unlikely that these clinical characteristics would have changed as much the change in BCS rate suggests. According to *The report of clinical statistical studies on registered mammary cancer patients in Japan* edited by the Joint Committee for Mammary Cancer in Japan, the proportion of patients with early breast cancer (Stages 0–II) to which the BCS guidelines apply have not changed over the study period. Third, this study could not identify patients with recurrent breast cancer. Although the guideline covers local treatment of patients with Stages I and II breast cancer, our study population also involved patients with Stage III to whom the guidelines do not apply (patients with Stage IV have been excluded in our study by use of the information regarding distant metastasis).

Although these limitations make it difficult to know whether our estimates of BCS use by hospital reflect differences in clinical presentation or unwarranted variation in treatment, we believe that guideline publication could be associated with the change in clinical practice we observed. The proportion of patients with Stage III cancer, who are not eligible for BCS, has been $<10\%$ of all breast cancer patients during the study period (findings from the cancer registry reports). Therefore, even without clear information on stage, we believe that our results would not be essentially different if all data regarding patients' severity of cancer were available. Fourth, because the female patients received surgical treatment at five teaching hospitals in Japan, rather than being randomly selected from all hospitals in Japan, and the

number of hospitals in our study was relatively small, this may reduce the external validity of our findings. However, our results regarding the use of BCS before and after guideline publication are consistent with a crude proportion of BCS use estimated by questionnaire survey at over 300 hospitals in Japan [17]. Therefore, the findings of our study may have face validity for hospitals throughout Japan. Last, since our database used administrative medical claims data, our analysis was not adjusted by surgeon factors and patients' preferences. Tarbox *et al.* [33] showed that not all surgeons believed that BCS and mastectomy had an equivalent survival rate. Especially in Japan, because there have been no clinical trials to examine whether early breast cancer patients in Japan are equally well treated with BCS as with mastectomy, we believe that the decision to use BCS for early breast cancer patients might depend on the preferences of surgeons and/or patients. Also, since both surgical procedures are equivalent in terms of survival, women may opt out of BCS by other factors such as a woman's individual preferences and personal reasons. For example, concerns about the burden of hospital visit for radiotherapy and the perceived risk of recurrence are supposable. On the contrary, of course, some patients with information regarding treatment strategy from media coverage would ask for BCS treatment and therefore the use of BCS might increase. The information, however, might be mostly derived from guideline publication in Japan.

Clarifying the impact of guideline dissemination alone cannot actually be achieved by the use of data from observational studies. Data from the clinical trials underlying these guidelines would have been published several years ahead of the guideline dissemination. Consequently, information from other sources such as medical literature, academic conference or seminars based on the results of clinical trials may spread to clinicians without passing through a specific guideline. Similarly, even after guideline dissemination, major events such as public education campaigns or patient advocacy campaigns could influence the rate of surgery. The results of several previous studies [8–11, 18–22], which aimed to explore the impact of guideline publication did not take into account biases involving information derived or available from other information sources (such as the results of clinical trials and dissemination of clinical guidelines). Our study is unique in potentially minimizing the direct and indirect impacts of published clinical trials on clinical practice (through a language barrier).

Conclusion

Several studies have assessed the effects of consensus recommendations on clinical practice, but have failed to show the impact of the consensus alone perhaps because information from scientific findings spread prior to the recommendations. Owing to the language barrier, however, we believe that by placing focus on the Japanese population, our study may reduce the impact of these publications. Our study suggested that the use of BCS has been increased dramatically in Japan since 2001, 2 years after guideline publication in Japan. This increase is consistent with a relationship between guideline publication and a change in clinical practice.

Acknowledgements

The authors are grateful to the anonymous QIP participant hospitals.

Funding

The work described in this article was funded in part by the Health Sciences Research Grants for the Research on Policy Planning and Evaluation from the Ministry of Health, Labor and Welfare of Japan and the Grant-in-aid for Scientific Research A from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

1. Field MJ, Lohr MJ (eds). *Clinical Practice Guidelines: Directions for a New Program.* Washington, DC: National Academy Press, 1990.
2. Audet AM, Greenfield S, Field M. Medical practice guidelines: current activities and future directions. *Ann Intern Med* 1990;**113**:709–14.
3. Grol R, Grimshaw J. From best evidence to best practice: effective implementation of change in patients' care. *Lancet* 2003;**362**:1225–30.
4. Schuster MA, McGlynn EA, Brook RH. How good is the quality of health care in the United States? *Milbank Q* 1998;**76**:517–63.
5. Fisher B, Bauer M, Margolese R *et al.* Five-year results of a randomized clinical trial comparing total mastectomy and segmental mastectomy with or without radiation in the treatment of breast cancer. *N Engl J Med* 1985;**312**:665–73.
6. Veronesi U, Saccozzi R, Del Vecchio M *et al.* Comparing radical mastectomy with quadrantectomy, axillary dissection, and radiotherapy in patients with small cancers of the breast. *N Engl J Med* 1981;**305**:6–11.
7. NIH Consensus Conference. Treatment of early-stage breast cancer. *JAMA* 1991;**265**:391–5.
8. Lazovich D, Solomon CC, Thomas DB *et al.* Breast conservation therapy in the United States following the 1990 National Institutes of Health Consensus Development Conference on the treatment of patients with early stage invasive breast carcinoma. *Cancer* 1999;**86**:628–37.
9. Du X, Freeman DH Jr, Syblik DA. What drove changes in the use of breast conserving surgery since the early 1980s? *Breast Cancer Res Treat* 2000;**62**:71–9.
10. Gilligan MA, Kneusel RT, Hoffmann RG *et al.* Persistent differences in sociodemographic determinants of breast conserving

treatment despite overall increased adoption. *Med Care* 2002; **40**:181–9.

11. Gaudette LA, Gao RN, Spence A *et al.* Declining use of mastectomy for invasive breast cancer in Canada, 1981–2000. *Can J Public Health* 2004;**95**:336–40.
12. Japanese Breast Cancer Society. Practice guideline: breast-conserving surgery. *Jpn J Breast Cancer* 2000;**15**:147–56 [in Japanese].
13. Ishizaki T, Imanaka Y, Hirose M *et al.* A first look at variations in use of breast conserving surgery at five teaching hospitals in Japan. *Int J Qual Health Care* 2002;**14**:411–8.
14. Charlson ME, Pompei P, Ales KL *et al.* A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987;**40**:373–83.
15. Sundararajan V, Henderson T, Perry C *et al.* New ICD-10 version of the Charlson comorbidity index predicted in-hospital mortality. *J Clin Epidemiol* 2004;**57**:1288–94.
16. Rogers EM. *Diffusion of Innovations*, 4th edn. New York, NY: Free Press, 1995.
17. Sonoo H, Noguchi S. Results of questionnaire survey on breast cancer surgery in Japan 2004–2006. *Breast Cancer* 2008;**15**:3–4.
18. Kosecoff J, Kanouse DE, Rogers WH *et al.* Effects of the National Institutes of Health Consensus Development Program on physician practice. *JAMA* 1987;**258**:2708–13.
19. Sherman CR, Potosky AL, Weis KA *et al.* The Consensus Development Program: detecting changes in medical practice following a consensus conference on the treatment of prostate cancer. *Int J Technol Assess Health Care* 1992;**8**:683–93.
20. Gleicher N. Cesarean section rates in the United States: the short-term failure of the National Consensus Development Conference in 1980. *JAMA* 1984;**252**:3273–6.
21. Lomas J, Anderson GM, Domnick-Pierre K *et al.* Do practice guidelines guide practice? *N Engl J Med* 1989;**321**:1306–11.
22. Thamer M, Ray NF, Henderson SC *et al.* Influence of the NIH Consensus Conference on *Helicobacter pylori* on physician prescribing among a Medicaid population. *Med Care* 1998;**36**:646–60.
23. Chagpar AB, Studts JL, Scoggins CR *et al.* Factors associated with surgical options for breast carcinoma. *Cancer* 2006;**106**:1462–6.
24. Mandelblatt JS, Berg CD, Meropol NJ *et al.* Measuring and predicting surgeons' practice styles for breast cancer treatment in older women. *Med Care* 2001;**39**:228–42.
25. Morrow M, White J, Moughan J *et al.* Factors predicting the use of breast-conserving therapy in stage I and II breast carcinoma. *J Clin Oncol* 2001;**19**:2254–62.
26. Katz SJ, Lantz PM, Janz NK *et al.* Surgeon perspectives about local therapy for breast carcinoma. *Cancer* 2005;**104**:1854–61.
27. Liang W, Burnett CB, Rowland JH *et al.* Communication between physicians and older women with localized breast cancer: implications for treatment and patient satisfaction. *J Clin Oncol* 2002;**20**:1008–16.
28. Cyran EM, Crane LA, Palmer L. Physician sex and other factors associated with type of breast cancer surgery in older women. *Arch Surg* 2001;**136**:185–91.
29. Mandelblatt JS, Hadley J, Kerner JF *et al.* Patterns of breast carcinoma treatment in older women: patient preference and clinical and physical influences. *Cancer* 2000;**89**:561–73.
30. Ballard-Barbash R, Potosky AL, Harlan LC *et al.* Factors associated with surgical and radiation therapy for early stage breast cancer in older women. *J Natl Cancer Inst* 1996;**88**: 716–26.
31. Hall SE, Holman CD, Hendrie DV *et al.* Unequal access to breast-conserving surgery in Western Australia 1982–2000. *ANZ J Surg* 2004;**74**:413–9.
32. Nattinger AB, Gottlieb MS, Veum J *et al.* Geographic variation in the use of breast-conserving treatment for breast cancer. *N Engl J Med* 1992;**326**:1102–7.
33. Tarbox BB, Rockwood JK, Abernathy CM. Are modified radical mastectomies done for T1 breast cancers because of surgeon's advice or patient's choice? *Am J Surg* 1992;**164**:417–20.

Accepted for publication 29 July 2009

Influence of Verification Bias on the Assessment of MRI in the Diagnosis of Meniscal Tear

Haruo Nishikawa[1]
Yuichi Imanaka
Miho Sekimoto
Kenshi Hayashida
Hiroshi Ikai

OBJECTIVE. Previous studies of the sensitivity and specificity of MRI in the diagnosis of meniscal tear have not included correction for verification bias. The purpose of this study was to investigate the extent to which verification bias affected assessment of the utility of MRI in the diagnosis of meniscal tear.

MATERIALS AND METHODS. The patients included in the study were outpatients who from April 2006 through July 2008 consecutively visited a single institution for MRI of the meniscus for evaluation of knee pain. For patients who underwent arthroscopy in addition to MRI, the sensitivity and specificity of MRI were calculated. Global sensitivity analysis of data on patients who did not undergo arthroscopy was performed to estimate the influence of verification bias. Global sensitivity analysis is a method for graphically determining whether a particular pair of sensitivity and specificity estimates is compatible with observed data.

RESULTS. Eighty-two patients (23%) underwent arthroscopic verification. The sensitivity and specificity of MRI were 85% and 31%. When the possibility of meniscal tears in patients who did not undergo arthroscopy was subjected to global sensitivity analysis, the sensitivity of MRI ranged from 29% to 95% and the specificity ranged from 3% to 92%. All combinations of sensitivity and specificity produced a butterfly-shaped curve, but the base case was not inside the curve.

CONCLUSION. Verification bias greatly affected assessment of the utility of MRI in the diagnosis of meniscal tear. Sensitivity and specificity from previous studies may be incompatible with our data owing to verification bias.

Meniscal tear, one of the most common causes of knee pain [1, 2], can occur in any age group. Young people are affected in sports injuries [3, 4], the elderly are affected by degenerative processes [5, 6], and persons of all ages can be affected by trauma [7, 8]. Meniscal tear causes substantial disability [9]. Because early diagnosis of meniscal tear can be of great benefit in improvement of functional outcome and quality of life, most authors advocate early treatment [10, 11].

MRI has been widely used for screening; it is a noninvasive and accurate diagnostic tool [1, 12]. Both the sensitivity and the specificity of MRI in the diagnosis of meniscal tears have been reported to be greater than 80% [1, 12]. Some researchers [13–15] have argued that MRI is an excellent, cost-effective diagnostic tool. There is serious concern, however, about the methods used in studies of the diagnostic performance of MRI. One problem, verification bias, also known as workup bias, posttest referral bias, and selection bias [16, 17], can occur if patients are not equally likely to have the diagnosis verified with a reference standard technique. When a diagnostic test is evaluated against a definitive reference standard test, which can be invasive and expensive, one often finds that not all patients undergo the reference standard test. If only patients with verified disease participate in assessments of test performance, sensitivity and specificity estimates are likely to be biased because patients with verified disease often are not a random sample of the population in which the diagnostic test is used. This bias is referred to as verification bias [18].

In previous studies [8, 19–21], MRI has been evaluated in reference to the findings at arthroscopy, which is considered the reference standard in the diagnosis of meniscal tear. However, arthroscopy is not performed on all study subjects because the procedure is invasive and is sometimes accompanied by

Keywords: meniscal tear, MRI, sensitivity, specificity, verification bias

DOI:10.2214/AJR.08.2223

Received December 9, 2008; accepted after revision May 12, 2009.

[1]All authors: Department of Healthcare Economics and Quality Management, School of Public Health, Kyoto University Graduate School of Medicine, Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan. Address correspondence to Y. Imanaka (imanaka@pbh.med.kyoto-u.ac.jp).

AJR 2009; 193:1596–1602

0361–803X/09/1936–1596

serious complications [22]. Arthroscopy usually is performed on patients who have abnormal findings at MRI. A difference between the distribution of meniscal tear among patients who undergo arthroscopy and the distribution among patients who do not undergo arthroscopy can lead to serious verification bias, resulting in overestimation of the sensitivity and specificity of MRI. To our knowledge, previous studies of the sensitivity and specificity of MRI in the diagnosis of meniscal tear have not included correction for verification bias.

Kosinski and Barnhart [18] described global sensitivity analysis, a novel method of graphically determining whether a particular pair of sensitivity and specificity estimates is compatible with the observed data [23]. Global sensitivity analysis is a practical approach to estimation of the influence of verification bias. The purpose of our study was to use global sensitivity analysis to determine the effects of verification bias on the diagnostic performance of MRI in the evaluation of meniscal tear.

Materials and Methods

Patients

The study setting was an outpatient clinic at a single institution. Patients who consecutively visited the hospital for MRI evaluation of the meniscus from April 2006 through July 2008 were included in the study. Patients were excluded from the study if they had undergone knee surgery, if more than 240 days had elapsed between MRI and arthroscopy, and if MR images had poor resolution. All information regarding patient characteristics and clinical findings was collected retrospectively in a chart review. The study was approved by the institutional review board.

Diagnosis

All MRI examinations were performed with the same protocol on the same 1.5-T unit (Excelart with Pianissimo, Toshiba Medical Systems) with an extremity quadrature coil and the spin-echo method. The parameters for sagittal T2-weighted images were TR/TE, 3,628/94; flip angle, 90° and 160°; field of view, 20 × 20 cm; slice thickness, 3.5 mm; interslice gap, 1.0 mm; and matrix size, 224 × 288. The parameters for sagittal T1-weighted images were 495/15; flip angle, 90° and 180°; field of view, 20 × 20 cm; slice thickness, 3.5 mm; interslice gap, 1.0 mm; and matrix size, 176 × 272. Sagittal STIR images were obtained with the following parameters: 5,635/80; flip angle, 90° and 160°; field of view, 20 × 20 cm; slice thickness, 3.5 mm; interslice gap, 1.0 mm; and matrix size, 224 × 304. The parameters for the sagittal T2*-weighted images were 535/15; flip angle, 25°; field of view, 20 × 20 cm; slice thickness, 3.5 mm; interslice gap, 1.0 mm; and matrix size, 160 × 304. The parameters for coronal STIR images were 5,635/80; flip angle, 90° and 160°; field of view, 20 × 20 cm; slice thickness, 3.0 mm; interslice gap, 1.0 mm; and matrix size, 224 × 272. The parameters for coronal T2*-weighted images were 535/15; flip angle, 25°; field of view, 20 × 20 cm; slice thickness, 3.0 mm; interslice gap, 1.0 mm; and matrix size, 160 × 304. The parameters for axial T2-weighted images were 3,628/94; flip angle, 90° and 160°; field of view, 18 × 20 cm; slice thickness, 4.0 mm; interslice gap, 2.0 mm; and matrix size, 224 × 400. The imaging time for each of the seven sequences was 2–3 minutes.

MR images were interpreted by one of two radiologists with more than 10 years of experience. A meniscus was considered possibly torn if a sign of tear was found on one image and torn if more than one image showed abnormal findings suggestive of tear according to previously published criteria. An abnormal finding was defined as intrameniscal signal intensity in contact with an articular surface [24–26]. For this study, all menisci with the diagnosis of possibly torn or torn were considered torn, because an abnormality on one or more images is the standard MRI criterion for the diagnosis of meniscal tear. We used arthroscopy as the reference standard because of its previously reported [8, 19–21] accuracy of greater than 95%. For our study, arthroscopy was performed by one orthopedic surgeon with more than 15 years of experience.

TABLE 1: Characteristics of Patients With Suspected Meniscal Tear (*n* = 356)

Characteristic	Value
Age (y)	
Mean	51
SD	20
Range	7–93
Sex (*n*)	
Male	182
Female	174
Location of suspected tear (*n*)	
Medial	183
Lateral	63

Analysis

The outcome measure was meniscal tear, classified as medial or lateral. The results of MRI were compared with the results of arthroscopy. We calculated the sensitivity and specificity of MRI in the sample who underwent arthroscopy. We performed global sensitivity analysis to assess the influence of verification bias [18]. We simulated the complete range of possible prevalence (0–100%) of meniscal tear for the MRI-positive and MRI-negative subgroups of patients who did not undergo arthroscopy. We calculated and graphically plotted the sensitivity and specificity in order to depict all possible combinations of sensitivity and specificity (Appendix 1).

We compared sensitivity and specificity between subgroups of patients. Subgrouping was based on the following factors known to influence the diagnostic performance of MRI: age (younger than 45 years, 45 years and older), sex (male, female), interval between MRI and arthroscopy (less than lower quartile for this study population, lower quartile or greater for this study population) [24], and location of tear (medial, lateral) [24, 27–29]. Chi-square tests were performed to compare the rates of positive findings between subgroups of patients. Statistical analysis was performed with a software package (Stata version 10, StataCorp).

Results

Five of the initial 361 patients were excluded from the final analysis because of previous knee surgery (n = 3), poor resolution of MR images (n = 1), and more than 240 days between MRI and arthroscopy (n = 1). The general characteristics of the patients are shown in Table 1. Of the 356 patients included (182 men, 174 women; mean age, 51 years), 82 (23%) underwent MRI and arthroscopy, and 274 underwent MRI only. Among the 356 patients, 214 had abnormal (positive) test results, and 142 had normal (negative) test results. The distribution of test results is shown in Table 2. Unlike in previous studies, the point estimate of specificity in our study was quite low. Among the cases verified with

TABLE 2: Frequency of Test Results

MRI Result	Verified Meniscal Tear		Not Verified	Total
	Present	Absent		
Positive	56	11	147	214
Negative	10	5	127	142
Total	66	16	274	356

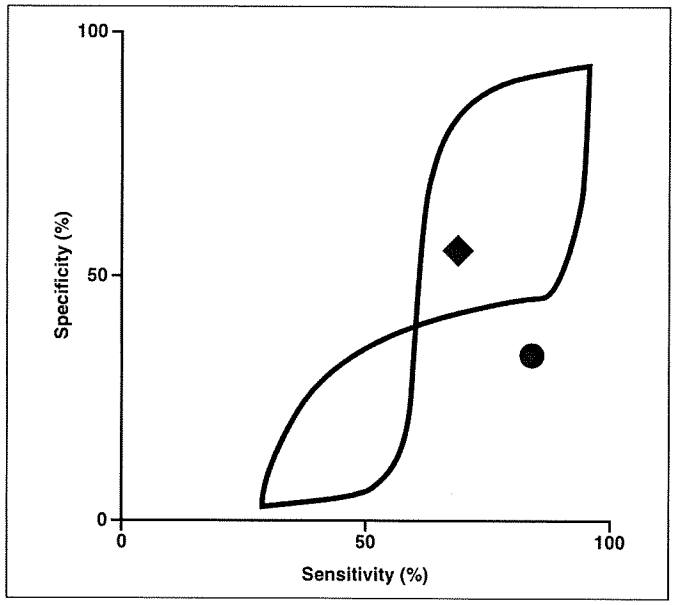

Fig. 1—Results of global sensitivity analysis. Areas outlined by butterfly curve represent all possible values of sensitivity and specificity in consideration of the effects of verification bias. Circle indicates point of sensitivity and specificity based on verified data only (base case); diamond, estimated point of sensitivity and specificity under missing-at-random assumption.

arthroscopy, the specificity was 31%, and the sensitivity was 85%. The area of all possible pairs of sensitivity and specificity based on the results of global sensitivity analysis is illustrated in Figure 1. The point estimates of sensitivity and specificity in the base case were not included in this region. The variation of sensitivity (29–95%) was smaller than the variation of specificity (3–92%).

For the second analysis, the subgroup was the 82 patients who underwent both MRI and arthroscopy. Characteristics of the 82 patients (mean age, 52 years) are shown in Table 3. The subgroup consisted of 48 men and boys and 34 women and girls. In all cases, the interval between MRI and arthroscopy was less than 6 months. The number of meniscal tears was 72. The chi-square results of stratified comparisons are shown in Tables 4 and 5. Statistically significant differences were observed in sensitivity for medial and lateral location (84% vs 59%; $p = 0.02$) and in specificity also for medial and lateral location (56% vs 85%, $p < 0.01$).

Discussion

Using global sensitivity analysis, we investigated the influence of verification bias on the sensitivity and specificity of MRI in the diagnosis of meniscal tear. Although previous studies of the diagnostic performance of MRI have shown promising results, those studies predominantly included patients with arthroscopic referrals. Several methods have been proposed to handle patient selection bias. The easiest way is to examine the results only for patients undergoing both MRI and arthroscopy. The assumption is that disease status is independent of patients' decisions to undergo arthroscopy [16, 30]. Alternatively, negative test results can be verified with a different, often less thorough standard, such as follow-up findings and physical signs [31]. However, such methods cannot exclude bias completely, and use of those methods may lead to inaccurate conclusions about the accuracy of a test. Physical signs and a medical history are clinically essential but are not sensitive enough to exclude meniscal tear [2, 32].

Global sensitivity analysis is the most robust approach [18]. It simulates the behavior of sensitivity and specificity, but disease prevalence among patients not undergoing the reference standard test takes all possible values. Therefore, global sensitivity analysis can be used to determine graphically whether a particular pair of sensitivity and specificity estimates is compatible with observed data. In our study, the presumed pair represented a butterfly shape, as depicted in Figure 1. If the verification bias did not exist or had a small influence, the point estimate would be supposed to be included in the presumed region, which was not the case in our study. Given that the point estimate was not included in the presumed region, we conclude that this estimate is not compatible with observed data and that there was verification bias with great influence. Other studies [1, 12] have shown sensitivities and specificities greater than 80% in the MRI diagnosis of meniscal tear. Owing to verification bias, indexes from those studies may be incompatible with our data.

TABLE 3: Characteristics of Patients Undergoing Both MRI and Arthroscopy ($n = 82$)

Characteristic	Value
Age (y)	
Mean	52
SD	18
Range	13–79
Sex (*n*)	
Male	48
Female	34
Interval to reference test (d)	
Mean	41
SD	38
Range	1–167
Location of suspected tear (*n*)	
Medial	50
Lateral	22

TABLE 4: Results of Stratified Comparisons of Sensitivity Values in Subgroups of Patients Undergoing Both MRI and Arthroscopy ($n = 82$)

Characteristic	Sensitivity (%)	*p*[a]
Age (y)		0.40
< 45	79	
≥ 45	87	
Sex (*n*)		0.07
Male	92	
Female	76	
Interval to reference test (d)		0.37
< 16	77	
≥ 16	87	
Site of meniscal tear (*n*)		0.02
Medial	84	
Lateral	59	

[a]Chi-square test.

One easy method is to calculate verification bias–corrected point estimates on the basis of the assumption that the prevalence of disease in unverified cases of positive and negative results is the same as in the corresponding groups of verified cases [18, 33] (Appendix 2). The sensitivity and specificity of the missing-at-random estimate are 65% and 57%, respectively. This point estimate is included in the presumed region and is consistent with observed data. Compared with

TABLE 5: Results of Stratified Comparisons of Specificity Values in Subgroups of Patients Undergoing Both MRI and Arthroscopy (n = 82)

Characteristic	Specificity (%)	p^a
Age (y)		0.09
< 45	60	
≥ 45	18	
Sex (n)		0.51
Male	36	
Female	20	
Interval to reference test (d)		0.20
< 16	14	
≥ 16	44	
Site of meniscal tear (n)		< 0.01
Medial	56	
Lateral	85	

[a]Chi-square test.

this specific missing-at-random point estimate, the uncorrected sensitivity is overestimated and the uncorrected specificity is underestimated. Bias correction is a complex field, and the missing-at-random calculation is certainly not perfect. Given the compatibility with observed data, the missing-at-random method yields a better estimate of actual sensitivity and specificity than do estimates obtained with traditional methods.

Studies by other researchers [13–15] have suggested that the cost-effectiveness of MRI is favorable. Those earlier analyses, however, were based on the results of studies in which verification bias was not considered [34]. If the sensitivity of MRI is not adequate, the cost-effectiveness of MRI may be different from that concluded in the past. Therefore, research on MRI with high internal validity is mandatory before cost-effectiveness analyses are conducted in the future.

A major finding in our study was that the presumed region of sensitivity and specificity was remarkably wide. One possible explanation is that the number of cases not verified was three times the number of cases verified. Another major finding is that the specificity in our study was lower than that reported in previous studies by other investigators. In many past studies by other researchers, medial meniscal tears were less common than lateral meniscal tears. Our patient population, however, included two times as many medial meniscal tears as it did lateral tears. MRI is less sensitive for lateral meniscal tears and less specific for medial tears, as reported previously [24, 28]. We believe that the difference in patient population explains the low specificity observed in our study.

We conclude that verification bias has substantial influence on the assessment of MRI in the diagnosis of meniscal tear. Using global sensitivity analysis, we found that the bias can result in overestimation of sensitivity and underestimation of specificity. Therefore, clinicians should not place too much trust in MRI in the diagnosis of meniscal tear. Future studies should be designed to include follow-up of patients who do not undergo arthroscopy and to determine the effects of verification bias by use of global sensitivity analysis.

References

1. Jackson JL, O'Malley PG, Kroenke K. Evaluation of acute knee pain in primary care. *Ann Intern Med* 2003; 139:575–588
2. Solomon DH, Simel DL, Bates DW, Katz JN, Schaffer JL. The rational clinical examination: does this patient have a torn meniscus or ligament of the knee? Value of the physical examination. *JAMA* 2001; 286:1610–1620
3. Terzidis IP, Christodoulou A, Ploumis A, Givissis P, Natsis K, Koimtzis M. Meniscal tear characteristics in young athletes with a stable knee: arthroscopic evaluation. *Am J Sports Med* 2006; 34:1170–1175
4. Shellock FG, Deutsch AL, Mink JH, Kerr R. Do asymptomatic marathon runners have an increased prevalence of meniscal abnormalities? An MR study of the knee in 23 volunteers. *AJR* 1991; 157:1239–1241
5. Christoforakis J, Pradhan R, Sanchez-Ballester J, Hunt N, Strachan RK. Is there an association between articular cartilage changes and degenerative meniscus tears? *Arthroscopy* 2005; 21:1366–1369
6. Low AK, Chia MR, Carmody DJ, Lucas P, Hale D. Clinical significance of intrasubstance meniscal lesions on MRI. *J Med Imaging Radiat Oncol* 2008; 52:227–230
7. Wagemakers HP, Heintjes EM, Boks SS, et al. Diagnostic value of history-taking and physical examination for assessing meniscal tears of the knee in general practice. *Clin J Sport Med* 2008; 18: 24–30
8. Vaz CE, Camargo OP, Santana PJ, Valezi AC. Accuracy of magnetic resonance in identifying traumatic intraarticular knee lesions. *Clinics* 2005; 60:445–450
9. McAlindon TE, Cooper C, Kirwan JR, Dieppe PA. Knee pain and disability in the community. *Br J Rheumatol* 1992; 31:189–192
10. Lysholm J, Gillquist J, Liljedahl SO. Arthroscopy in the early diagnosis of injuries to the knee joint. *Acta Orthop Scand* 1981; 52:111–118
11. Hou XK. Early arthroscopy in diagnosis and treatment of acute injury of the knee [in Chinese]. *Zhonghua Wai Ke Za Zhi* 1992; 30:7–9, 61
12. Oei EH, Nikken JJ, Verstijnen AC, Ginai AZ, Myriam Hunink MG. MR imaging of the menisci and cruciate ligaments: a systematic review. *Radiology* 2003; 226:837–848
13. Weinstabl R, Muellner T, Vecsei V, Kainberger F, Kramer M. Economic considerations for the diagnosis and therapy of meniscal lesions: can magnetic resonance imaging help reduce the expense? *World J Surg* 1997; 21:363–368
14. Williams P. MRI can prevent unnecessary arthroscopy. (commentary) *J Bone Joint Surg Br* 1998; 80:371
15. Bui-Mansfield LT, Youngberg RA, Warme W, Pitcher JD, Nguyen PL. Potential cost savings of MR imaging obtained before arthroscopy of the knee: evaluation of 50 consecutive patients. *AJR* 1997; 168:913–918
16. Begg CB. Biases in the assessment of diagnostic tests. *Stat Med* 1987; 6:411–423
17. Revesz G, Kundel HL, Bonitatibus M. The effect of verification on the assessment of imaging techniques 1983. *Invest Radiol* 1990; 25:461–464
18. Kosinski AS, Barnhart HX. A global sensitivity analysis of performance of a medical diagnostic test when verification bias is present. *Stat Med* 2003; 22:2711–2721
19. DeHaven KE, Collins HR. Diagnosis of internal derangements of the knee: the role of arthroscopy. *J Bone Joint Surg Am* 1975; 57:802–810
20. Fischer SP, Fox JM, Del Pizzo W, Friedman MJ, Snyder SJ, Ferkel RD. Accuracy of diagnoses from magnetic resonance imaging of the knee: a multi-center analysis of one thousand and fourteen patients. *J Bone Joint Surg Am* 1991; 73:2–10

21. Halbrecht JL, Jackson DW. Office arthroscopy: a diagnostic alternative. *Arthroscopy* 1992; 8:320–326
22. Reigstad O, Grimsgaard C. Complications in knee arthroscopy. *Knee Surg Sports Traumatol Arthrosc* 2006; 14:473–477
23. Danias PG, Parker JA. Novel Internet-based tool for correcting apparent sensitivity and specificity of diagnostic tests to adjust for referral (verification) bias. *RadioGraphics* 2002; 22:e4
24. De Smet AA, Mukherjee R. Clinical, MRI, and arthroscopic findings associated with failure to diagnose a lateral meniscal tear on knee MRI. *AJR* 2008; 190:22–26
25. De Smet AA, Norris MA, Yandow DR, Quintana FA, Graf BK, Keene JS. MR diagnosis of meniscal tears of the knee: importance of high signal in the meniscus that extends to the surface. *AJR* 1993; 161:101–107
26. De Smet AA, Tuite MJ. Use of the "two-slice-touch" rule for the MRI diagnosis of meniscal tears. *AJR* 2006; 187:911–914
27. Nikolić DK. Lateral meniscal tears and their evolution in acute injuries of the anterior cruciate ligament of the knee: arthroscopic analysis. *Knee Surg Sports Traumatol Arthrosc* 1998; 6:26–30
28. De Smet AA, Nathan DH, Graf BK, Haaland BA, Fine JP. Clinical and MRI findings associated with false-positive knee MR diagnoses of medial meniscal tears. *AJR* 2008; 191:93–99
29. Justice WW, Quinn SF. Error patterns in the MR imaging evaluation of menisci of the knee. *Radiology* 1995; 196:617–621
30. Begg CB, Greenes RA. Assessment of diagnostic tests when disease verification is subject to selection bias. *Biometrics* 1983; 39:207–215
31. Knottnerus JA. The effects of disease verification and referral on the relationship between symptoms and diseases. *Med Decis Making* 1987; 7:139–148
32. Zanetti M, Pfirrmann CW, Schmid MR, Romero J, Seifert B, Hodler J. Patients with suspected meniscal tears: prevalence of abnormalities seen on MRI of 100 symptomatic and 100 contralateral asymptomatic knees. *AJR* 2003; 181:635–641
33. Cecil MP, Kosinski AS, Jones MT, et al. The importance of work-up (verification) bias correction in assessing the accuracy of SPECT thallium-201 testing for the diagnosis of coronary artery disease. *J Clin Epidemiol* 1996; 49:735–742
34. Selesnick FH, Noble HB, Bachman DC, Steinberg FL. Internal derangement of the knee: diagnosis by arthrography, arthroscopy, and arthrotomy. *Clin Orthop Relat Res* 1985; 26–30

APPENDIX 1: Global Sensitivity Analysis

Global sensitivity analysis is a method for graphically showing whether a particular pair of sensitivity and specificity estimates is compatible with the observed data, including unverified cases. In this study, the complete range of possible prevalence (0–100%) of meniscal tears for the MRI-positive and MRI-negative subgroups of patients who did not undergo arthroscopy was simulated. Thereafter, sensitivity and specificity were calculated and graphically plotted to depict all possible combinations of sensitivity and specificity, including unverified cases. The frequency of test results is shown in Table 6, which includes data from unverified cases. Estimates of sensitivity and specificity are expressed in the following equations:

$$\text{Sensitivity}(E, F) = \frac{a + E \times e}{a + E \times e + c + F \times f} \times 100$$

$$\text{Specificity}(E, F) = \frac{d + (1 - F) \times f}{b + (1 - E) \times e + d + (1 - F) \times f} \times 100$$

E and F can be anywhere in the square area (0, 1) × (0, 1) independently. All values (0–1) of E and F are assigned. The sensitivity and specificity of any point value can be calculated with an Excel spreadsheet (Microsoft) (Table 7). Thereafter, pairs of calculated sensitivity and specificity are plotted on the graph.

TABLE 6: Calculation of Frequency of Test Results

MRI Result	Verified Meniscal Tear		Not Verified	Total
	Present	Absent		
Positive	$a + E \times e$	$b + (1 - E) \times e$	e	$n1$
Negative	$c + F \times f$	$d + (1 - F) \times F$	f	$n2$
Total				N

Note—E = proportion of the population with meniscal tear among patients with positive MRI result not undergoing arthroscopy, F = proportion of population without meniscal tear among patients with negative MRI result not undergoing arthroscopy, a = number of patients with positive MRI result and positive arthroscopy result, b = number of patients with positive MRI result and negative arthroscopy result, c = number of patients with negative MRI result and positive arthroscopy result, d = number of patients with negative MRI result and negative arthroscopy result, e = number of patients with positive MRI result not undergoing arthroscopy, f = number of patients with negative MRI result and not undergoing arthroscopy, n = subtotal, N = number of all included patients.

(Appendix 1 continues on next page)

TABLE 7: Spreadsheet of Calculation of Sensitivity and Specificity of Any Point Value

Row Number	A	B	C	D	E	F	G	H
1		Arthroscopy positive	Arthroscopy negative	Not verified		$a = 56$		
2	MRI positive	56	11	147		$b = 11$		
3	MRI negative	10	5	127		$c = 10$		
4				274		$d = 5$		
5						$e = 147$		
6		Arthroscopy positive	Arthroscopy negative	Not verified	Total	$f = 127$		
7	MRI positive	$a + E^* e$	$b + (1–E)^* e$	e	$n1$	$n1 = 214$		
8	MRI negative	$c + F^* f$	$d + (1–F)^* f$	f	$n2$	$n2 = 142$		
9					N	$N = 356$		
10								
11	E(%)	F(%)	$a + E^* e$	$b + (1–E)^* e$	$c + F^* f$	$d + (1 – F)^* f$	Sensitivity (%)	Specificity (%)
12	0	0	56.00	158.00	10	132	84.84848	45.51724
13	1	0	55.47	155.53	9	132	86.04002	45.90825
14	2	0	56.94	154.06	9	132	86.35123	46.14417
15	3	0	58.41	152.59	9	132	86.64887	46.38252
16	4	0	59.88	151.12	9	132	86.93380	46.62334
17	5	0	61.35	149.65	9	132	87.20682	46.86668
18	6	0	62.82	148.18	9	132	87.46867	47.11257
19	7	0	64.29	146.71	9	132	87.72002	47.36106
20	8	0	65.76	145.24	9	132	87.96148	47.61218
21	9	0	67.23	143.77	9	132	88.19362	47.86598
22	10	0	68.70	142.30	9	132	88.41699	48.12249
23	11	0	70.17	140.83	9	132	88.63206	48.38178
24	12	0	71.64	139.36	9	132	88.83929	48.64387
25	13	0	73.11	137.89	9	132	89.03909	48.90881
26	14	0	74.58	136.42	9	132	89.23187	49.17666
27	15	0	76.05	134.95	9	132	89.41799	49.44746
28	16	0	77.52	133.48	9	132	89.59778	49.72126
29	17	0	78.99	132.01	9	132	89.77156	49.99811
30	18	0	80.46	130.54	9	132	89.93964	50.27805
31	19	0	81.93	129.07	9	132	90.10228	50.56115
32	20	0	83.40	127.60	9	132	90.25974	50.84746
33	21	0	84.87	126.13	9	132	90.41227	51.13702

Note—See footnote to Table 6 for explanation of variables.

(Appendix 2 appears on next page)

APPENDIX 2: Missing-at-Random Point Estimates

If only patients with verified disease status are considered, the sensitivity and specificity estimates are $a/(a + c)$ and $d/(b + d)$, respectively. This condition is unlikely to be true in practice, and the estimates often are subject to verification bias. If selection of verified cases is independent of the unobserved variable, which we call the missing-at-random assumption, the verification bias–corrected sensitivity and specificity estimates can be calculated by assigning a/(a + b) to E and c/(c + d) to F. Verification bias–corrected sensitivity and specificity point estimates are expressed as follows:

$$\text{Missing-at-random sensitivity} = \frac{n1 \times a / (a + b)}{n1 \times a / (a + b) + n2 \times c / (c + d)} \times 100$$

$$\text{Missing-at-random specificity} = \frac{n2 \times d / (c + d)}{n1 \times b / (a + b) + n2 \times d / (c + d)} \times 100$$

研究成果の刊行に関する一覧表

研究成果の刊行に関する一覧表

原著論文（英文）

1. Umegaki T, Sekimoto M, Hayashida K, Imanaka Y. An outcome prediction model for adult intensive care. *Critical Care and Resuscitation* (in press)
2. Sekimoto M, Imanaka Y, Shirai T, Sasaki H, Komeno T, Lee J, Yoshihara K, Ashihara E, Maekawa T. Risk-adjusted assessment of incidence and quantity of blood use in acute-care hospitals in Japan: an analysis using administrative data. *Vox Sanguinis* (in press)
3. Regenbogen SE, Hirose M, Imanaka Y, Oh EH, Fukuda H, Gawande AA, Takemura T, Yoshihara H: A comparative analysis of incident reporting Lag times in Japan and the United States. *Quality & Safety in Health Care* 2009 (in press).
4. Fukuda H, Imanaka Y, Ishizaki T, Okuma K, Shirai T. Change in clinical practice after publication of guidelines on breast cancer treatment. *International Journal for Quality in Health Care.* 2009;21(5): 372-378.
5. Sasaki H, Imanaka Y, Sekimoto M , Lee J, Otsubo T. Antimicrobial prescription patterns for children hospitalized with pneumonia and compliance to guidelines in Japan: A multicenter study. *Journal of Evaluation in Clinical Practice* (in press)
6. Nishikawa H, Imanaka Y, Sekimoto M, Hayashida K, Ikai H. Impact of verification bias on the assessment of MRI for the diagnosis of meniscal tears. *American Journal of Roentgenology* (in press)
7. Nojo T, Imanaka Y, Ishizaki T, Sekimoto M, Yoshino M, Kurosawa T, Takao T, Ohtomo K. Lung cancer incidence in middle-aged men estimated by low-dose computed tomography screening. *Lung Cancer.* 2009;65:56-61.
8. Murakami G, Imanaka Y, Kobuse H, Lee J, Goto E. Patient Perceived Priorities between Technical Skills and Interpersonal Skills: Their Influence on Correlates of Patient Satisfaction. *Journal of Evaluation in Clinical Practice* (in press)
9. Shirai T, Imanaka Y, Sekimoto M, Ishizaki T, QIP Ovarian Cancer Expert Group. Primary chemotherapy patterns for ovarian cancer treatment in Japan. *The Journal of Obstetrics and Gynaecology Research* (in press)
10. Kawasaki K, Sekimoto M, Ishizaki T, Imanaka Y. Work stress and workload on full-time anesthesiologists of acute care hospitals in Japan. *Journal of Anesthesia.*2009;23:235-241.
11. Fukuda H, Imanaka Y, Hirose M, Hayashida K. Impact of system-level activities and reporting design on the number of incident reports for patient safety. *Quality & Safety in Health Care* (in press)

12. Sekimoto M, Imanaka Y, Kobayashi H, Okubo T, Kizu J, Kobuse H, Mihara H, Tsuji N, Yamaguchi A. Factors affecting performance of hospital infection control in Japan. *American Journal of Infection Control.* 2009;37(2):136-42.
13. Fukuda H, Imanaka Y, Hirose M, Hayashida K. Factors associated with system-level activities for patient safety and infection control. *Health Policy.* 2009;89(1):26-36.

国際学会発表

1. Sekimoto M, Imanaka Y, Shirai T, Sasaki H, et al. Risk-adjusted assessment of blood product use in acute-care hospitals in Japan: an analysis using administrative data. 25th PCSI Conference, Fukuoka, Japan. 11-14 November, 2009.
2. Umegaki T, Sekimoto M, Imanaka Y. Physician Staffing Patterns and Costs for Septic Patients in Intensive Care Units. 25th PCSI Conference, Fukuoka, Japan. 11-14 November, 2009.
3. Tanaka M, Sekimoto M, Imanaka Y. Development of a method for assessing operation room management based on Diagnosis Procedure Combination E and F-File data. 25th PCSI Conference, Fukuoka, Japan. 11-14 November, 2009.
4. Hirose M, Takemura T, Oh EH, Egami K, Shima H, Imanaka Y, Kuroda T, Yoshihara H. Comparison Among Three Teaching Hospitals In Japan Regarding Incident Reports Including Drug Names. ISPOR 12th Annual European Congress, Paris, 24-27 October, 2009.
5. Imanaka Y. 【招待】 Economics and Policy on Quality of Care in Japan. Workshop on Public Health: Health Policy, Legal Issues and Trade. Taipei: Taiwan, July 23 - 24, 2009.

学会発表

1. 福田治久，大隈和英，今中雄一．腹部・胸部大動脈瘤切除術における施設別手術件数と術後アウトカムの関連性．第 47 回日本医療・病院管理学会学術総会：東京，2009 年 10 月 17 日-18 日．
2. 濱田啓義，猪飼宏，関本美穂，今中雄一．妊娠・分娩にかかる個人および社会全体の費用の検討．第 47 回日本医療・病院管理学会学術総会：東京，2009 年 10 月 17 日-18 日．
3. 田中将之，関本美穂，今中雄一．手術室運営評価モデルを用いた多施設間比較．第 31 回日本手術医学会：東京，2009 年 10 月 16 日-17 日．
4. 田中将之，足立峻吾，濱田啓義，猪飼宏，林田賢史，今中雄一．日米の医療マネジメントの内容と方法の比較．第 41 回日本医学教育学会：大阪，2009 年 7 月 24 日．
5. 関本美穂，今中雄一．終末期における入院医療：疾患・診療パターンと医療費．医療経済学会総会・第 4 回研究大会：東京，2009 年 7 月 18 日．
6. 大隈和英，福田治久，関本美穂，猪飼宏，濱田啓義，今中雄一． DPC に基づく包括支払い制度導入後の乳癌治療への影響と変化．第 109 回日本外科学会学術集会総会：福岡，2009 年 04 月 04 日．(抄録：日本外科学会雑誌 Vol.110 Supplement: p327, 2009)
7. 関本美穂，今中雄一．急性胆管炎・胆嚢炎の診療ガイドライン：ガイドラインが診療に与える効果の検証について．第 45 回日本腹部救急医学会総会：東京，2009 年 3 月 12 日．
8. 関本美穂，今中雄一．大規模個票データを用いた診療パターンの分析．「大規模個票データを使った医療データの分析・統計分析手法」に関する国際シンポジュウム：東京，2009 年 3 月 4 日．
9. 関本美穂，梅垣岳志，今中雄一．Closed ICU と Open ICU —その定義と ICU の診療体制が診療プロセス・患者アウトカム・医療資源消費に与える影響について．第 36 回日本集中医療医学会学術集会：大阪，2009 年 2 月 26 日-28 日．(抄録：日本集中治療医学会雑誌 16 Supplement: p192, 2009)．
10. 梅垣岳志，関本美穂，猪飼宏，今中雄一．多施設 DPC データに基づく DIC 治療の現状．第 36 回日本集中治療医学会学術集会：大阪，2009 年 2 月 26 日-28 日．(抄録：日本集中治療医学会雑誌 16 Supplement: p219, 2009)．