Table 3. Test accuracy of photofluorography for gastric cancer screening

Author	Reported year	Follow-up strategy	Follow-up period (year)	Sensitivity (%)	Specificity (%)	Positive predictive value (%)
Murakami et al.	1989	Cancer registry	1	88.5	92.0	1.40
Sugahara N et al.	1991	Cancer registry	1	70.4	90.1	1.60
Fukao A et al.	1992	Cancer registry	1	69.3	88.8	2.00
Ishida T et al.	1994	Cancer registry	1	84.1	81.3	0.78
Ishida T et al.	1994	Cancer registry	2	70.1	81.3	0.90
Hattori M et al.	1998	Cancer registry	1	68.6–72.5	?	?
Abe S et al	2000	Cancer registry	1	56.8	90.7	2.00

85 gastric cancers were detected, of which 29 cases were early cancers. However, compared with the overall mortality for Linqu County, the standard morality ratio was 1.01 (95% CI: 0.32–1.37).

Only two studies related to the use of endoscopy as a diagnostic test reported the accuracy of endoscopy (34,35). In the first study, the sensitivity of endoscopy was found to be 77.8% based on 3-year follow-up using the cancer registry system in Fukui prefecture (34). However, the target population of this study was patients who had symptoms. In addition, the specificity was not reported. In another study based on a follow-up survey of individual participants, Otsuji et al. (35) reported that the sensitivity of endoscopy was 84.0%. No studies that have compared the survival of patients with gastric cancer between screened and non-screened group.

Adverse effects related to endoscopy are reported every 5 years by the Japanese Association of Gastroenterological Endoscopy (36). However, these results are not separately reported and classified by the purpose of endoscopy (screening, diagnostic test and treatment). The details of the adverse effects associated with endoscopic screening are unclear.

Serum pepsinogen test (level of evidence: 2 −)

A one-arm cohort study was conducted in a small distinct of Tokyo (37). On the basis of death certificates, three gastric cancer deaths were identified. Compared with gastric cancer deaths nationwide, the relative risk for the screening group was 0.34 (95% CI: 0.07–0.98). However, there are several limitations to this study related to target selection, lack of an appropriate comparator within the same population, lack of previous screening history and the short follow-up period. These issues were not discussed in the published paper.

Several studies reported the sensitivity of serum pepsinogen testing compared with the results of endoscopy done at the same time (38–45). The sensitivity of the serum pepsinogen test ranged from 40% to 80%, but the specificity was below 80%. Several studies reported that, compared with the results of endoscopy, the sensitivity of the serum pepsinogen test was higher than that of radiography. When the accuracy of a new method is compared with that of photofluorography, it is important to recognize that gastric cancer screening using radiography is popular in Japan, and that most participants have been previously screened by radiography. Although the detection rate of radiographic screening is low because it is being used as an incidence screen, it was the first time that most participants had been screened using serum pepsinogen testing; thus, the serum pepsinogen screen was a prevalence screen. The sojourn time of the serum pepsinogen testing, which is diagnostic test for atrophic gastritis, is probably longer than that of radiography. If one wishes to compare the two methods, they must be analysed under the same conditions.

Helicobacter pylori antibody (level of evidence: 2 −)

A high-risk group of patients could be identified using a combination of the *H. pylori* antibody and serum pepsinogen tests. Watabe et al. (46) followed 9293 participants who were screened using both *H. pylori* antibody and serum pepsinogen tests for 4.7 years. Compared with gastric cancer cases detected in the group that had a negative serum pepsinogen test and a negative *H. pylori* antibody (n =3324), the hazard ratios were higher in the following groups: the positive serum pepsinogen test and negative *H. pylori* antibody group (n =2134), 1.1 (95% CI 0.4–3.4); the normal serum pepsinogen test and positive *H. pylori* antibody group (n =1082), 6.0 (95% CI 2.4–14.5) and the positive serum pepsinogen test and positive *H. pylori* antibody (n =443), 8.2 (95% CI 3.2–21.5). Yamanoi et al. (43) reported that the sensitivity of *H. pylori* antibody testing was 87.1% and the specificity was 40.8%.

An RCT dealing with the prevention of gastric cancer using *H. pylori* eradication therapy was reported from China (47). The incidence of gastric cancer was similar between participants receiving *H. pylori* eradication therapy and those receiving placebo over a 7.5-year period (hazard ratio 1.10, 95% CI 1.05–1.15).

Recommendations

On the basis of the balance of benefits and harms, recommendations were formulated for population-based and opportunistic screening. Benefits were defined as evidence that mortality from a specific cancer was reduced by a cancer screening

program. As well, the harms of various methods were compared (Table 4) (15–28,34–36,38–45,48–57).

Gastric cancer screening using photofluorography was a grade B recommendation for population-based and opportunistic screening (Table 5). The other methods were not recommended as population-based screening due to insufficient evidence. In opportunistic screening, if individuals request screening they should be given appropriate information, and decision-making should be made at the individual level.

DISCUSSION

A guideline for gastric cancer screening was developed using a standardized method. The details of the guideline development method have been provided elsewhere, and the differences in the guideline development method have also been described elsewhere. Although the efficacy of gastric cancer screening has been evaluated in previous reports without providing recommendations (4), in our guideline, recommendations have been clearly defined based on the evidences. In the previous guideline, photofluorography was recommended for population-based gastric cancer screening. However, *H. pylori* antibody screening was not recommended, and the evidence for serum pepsinogen testing was insufficient to be able to make a recommendation either for or against its use. In addition, endoscopic screening was not targeted for evaluation.

Over the years, there have been four reports that have made suggestions about gastric cancer screening. In 1990, the UICC report concluded that screening programs should continue in regions with a high incidence of gastric cancer

Table 4. Harms of gastric cancer screening

Harms	Radiography	Gastroendoscopy	Serum pepsinogen	*Helicobacter pylori* antibody
False negative rate	20–30%	16%	16–50%	17.9%
	>10%	No report	20–30%	59.2%
	No	No	No (may be affected by nutrition)	No
False positive rate	No	Anticoagulant	Affected by use of proton pump inhibitor	No
Nutrition restriction	Photofluorography: No	Pharynx anesthetic sedation	No	No
Adverse effects of premedication	Direct radiography: possible (antispasmodic agent) shock, blood pressure failure, respiratory failure	Antispasmodic agent shock, hypotension, respiratory failure	—	—
Adverse effects of premedication (Death)	Possible	0.0001% (14/12,844,551)*	—	—
Adverse effects of screening test	False swallowing of barium meal constipation ileus	Bleeding, perforation, etc.	—	—
Frequency of adverse effects	False swallowing of barium meal 0.08 0.17%	0.012% (997/826,313)	No	No
	Defecation delay: 4–11%			
Adverse effects of screening test	False swallowing of barium meal constipation ileus	Bleeding, perforation, etc.	—	—
Death due to adverse effect	Cases reported	0.00076% (63/826,313)	—	—
Infection	No	Possible	No	No
Radiation exposure (effective dose)	Direct radiography	No	No	No
	Males: 4.9 mSv			
	Females: 3.7 mSv			
	Photofluorography			
	Males: 0.6 mSv			
	Females: 0.6 mSv			
Others	—	—	May be affected by stomach resection, renal failure and HP eradication	Antibiotic resistance, diarrhoea, soft stool

*Included in colonoscopy and laparoscopy.

Table 5. Recommendation for gastric cancer screening

Screening methods	Recommendation grade	Recommendations (language)	
		Population-based screening	Opportunistic screening
Gastroglurography	B	Recommend	Recommend
Gastroendscopy	I	Not recommend[a]	Decision making at individual[b]
Serum pepsinogen	I	Not recommend[a]	Decision making at individual[b]
Helicobacter pylori antibody	I	Not recommend[a]	Decision making at individual[b]

[a]There is insufficient evidence to recommend.
[b]If required, the health professional should explain that the evidence regarding mortality reduction by cancer screening is unclear. In addition, information about the harms is required. In such situations, the decision regarding cancer screening should be made on the individual level.

where they were already under way, but that gastric cancer screening could not be recommended as public health policy in other countries (58). The NCI-PDQ concluded that there was insufficient evidence to suggest that gastric cancer screening, including endoscopy, reduced mortality from gastric cancer (6). The European Code against Cancer (3rd edition) stated that there were no evidences to support gastric cancer screening using radiography, endoscopy or *H. pylori* antibody testing (59). The Medical Screening Association stated that the efficacy of radiographic screening was uncertain (60). The evaluations of endoscopy and *H. pylori* antibody testing reached the same conclusion that we reached. It is possible that serum pepsinogen testing may be effective due to its high test accuracy, as reported by Kitahara et al. (42). In all of the previous reports, no methods were recommended for gastric cancer screening.

The requirements of cancer screening programs differ among countries due to differences in cancer incidence and mortality. In Japan, although the incidence and mortality of gastric cancer have decreased in the last decade, gastric cancer screening is a major issue because the incidence and mortality remain high (1). The screening rate for gastric cancer has flattened, and the effectiveness of gastric cancer screening has been limited. However, endoscopic screening is expected to be an alternative strategy to radiography. No studies have evaluated whether endoscopic screening reduces gastric cancer mortality. Although most people consider that endoscopy has a high detection rate, its sensitivity compared with that of radiography is unclear. To prevent premature death from gastric cancer, evidence-based screening should be promoted. To achieve this aim, it is necessary to determine the mortality reduction that is associated with endoscopic screening. An RCT would be the most preferred strategy, but it would be difficult for gastric cancer screening to be conducted in Japan due to widespread of the screening programs nationwide. A case–control or cohort studies may be expected as alternative methods for evaluating mortality reduction by endoscopic screening.

After the publication of our guidelines, two cohort studies dealing with radiographic screening were reported (61,62). The results of these studies with respect to the mortality reduction of photofluorography screening are similar. Therefore, we did not need to change our recommendations. In addition, Yoshihara et al. (63) reported the results of a case–control study of serum pepsinogen testing. In this study, although mortality reduction by serum pepsinogen testing was suggested, there were several serious issues that could affect the interpretation of the result. More than half of the cases were over 70 years of age (mean age 71.9 years, ranged up to 92 years). Since the reference date was not clearly defined, the history of exposure to serum pepsinogen testing was not reliable. Therefore, it is difficult to judge the efficacy of serum pepsinogen screening based on this low-quality study. However, both serum pepsinogen testing and *H. pylori* antibody testing are expected to be used to identify individuals at high risk of gastric cancer. A recent study by You et al. (64) reported that *H. pylori* eradication reduced the prevalence of precancerous gastric lesions Gastric cancer prevention may become possible if the efficacy of *H. pylori* eradication were to be proven. Furthermore, a new screening method involving endoscopy may be expected in the near future. Therefore, we are planning to revise the guideline within 5 years, given that new evidence may become available.

Acknowledgments

We thank Kanoko Matsushima and Junko Asai for secretarial support.

Funding

This study was supported by Grant-in-Aid for Cancer Control from Ministry of Health, Labor and Welfare of Japan (Grant number 15-3).

Conflict of interest statement

None declared.

References

1. Nomura K, editor. Cancer Statistics in Japan 2005. Tokyo: Foundation for Promotion Cancer Research 2006.
2. Statistics and Information Department, Ministry of Health, Labour, and Welfare. Vital Statistics of Japan 2004. Tokyo: Health and Welfare Statistics Association 2006.
3. Statistics and Information Department, Ministry of Health, Labour, and Welfare. National Reports on Cancer Screening Programs 2004. Tokyo: Health and Welfare Statistics Association 2006.
4. Hisamichi S, editor. Guidelines for Cancer Screening Programs. Tokyo: Japan Public Health Association 2001 (in Japanese).

5. Hamashima C, Saito H, Nakayama T, et al. Standardized development method of Japanese Guidelines for Cancer Screening. *Jpn J Clin Oncol* (submitted for publication).
6. NCI: PDQ (Physician Data Query); gastric cancer screening. http://www.cancer.gov/cancertopics/pdq/screening/gastric/HealthProfessional/page1 (2005.9.6 access).
7. Oshima A, Hirata N, Ubukata T, Umeda K, Fujimoto I. Evaluation of a mass screening program for stomach cancer with a case-control study design. *Int J Cancer* 1986;38:829–33.
8. Fukao A, Tsubono Y, Tsuji I, Hisamichi S, Sugahara N, Takano A. The evaluation of screening for gastric cancer in Miyagi Prefecture, Japan: a population-based case-control study. *Int J Cancer* 1995;60:45–8.
9. Abe Y, Mitsushima T, Nagatani K, Ikuma H, Minamihara Y. Epidemiological evaluation of the protective effect for dying of stomach cancer by screening programme for stomach cancer with applying a method of case-control study.-A study of a efficient screening programme for stomach cancer. *J Gastroenterol* 1995;92:836–45 (in Japanese).
10. Tsubono Y, Hisamichi S. Case-control studies of screening for gastric cancer in Japan. *J Gastroenterol Mass Surv* 1999;37:182–5 (in Japanese).
11. Pisani P, Oliver WE, Parkin DM, Alvarez N, Vivas J. Case-control study of gastric cancer screening in Venezuela. *Br J Cancer* 1994;69:1102–5.
12. Inaba S, Hirayama H, Nagata C, Kurisu Y, Takatsuka N, Kawakami N, et al. Evaluation of a screening program on reduction of gastric cancer mortality in Japan: preliminary results from a cohort study. *Prev Med* 1999;29:102–6.
13. Mizoue T, Yoshimura T, Tokui N, Hoshiyama Y, Yatsuya H, Sakata K, et al. Prospective study of screening for stomach cancer in Japan. *Int J Cancer* 2003;106:103–7.
14. Yoshida Y, Tamura K, Arisue T, Tebayashi A, Yamaguchi Y, Hiyama S, et al. Precision of routine examination in the gastric carcinoma mass survey. *Stomach and Intestine* 1985;20:943–7 (in Japanese).
15. Yoshida Y, Tamura K. Estimation of false negative case by the ADC method. *J Gastroenterol Mass Surv* 1986;50:6–11 (in Japanese).
16. Sugahara N, Sibuki S, Hirasawa Y, Morimoto T. Characteristics of false negative cases. *Stomach and Intestine* 1991;26:1357–62 (in Japanese).
17. Fukao A, Hisamichi S, Takano A, Sugawara N. Accuracies of mass screening for gastric cancer-Test sensitivity and program sensitivity. *J Gastroenterol Mass Surv* 1992;97:59–63 (in Japanese).
18. Ishida T, Suematsu T, Oomayashi K, Takada Y, Kimura S, Suematsu C. Measurement of accuracy of stomach mass screening by population-based cancer registration. *J Gastroenterol Mass Surv* 1994;32:9–16 (in Japanese).
19. Hattori M, Fujita M, Hosokawa O, Yamazaki S. A clinico-pathological evaluation of false negative cases in gastric cancer mass survey. *J Gastroenterol Mass Surv* 1998;36:468–75 (in Japanese).
20. Abe S, Shibuya D, Noguchi T, Shimada T. An estimate of the false-negative rate of mass-screening for gastric carcinoma. *J Gastroenterol Mass Surv* 2000;38:475–82 (in Japanese).
21. Murakami R, Tsukuma H, Ubukata T, Nakanishi K, Fujimoto I, Kawashima T, et al. Estimation of validity of mass screening program for gastric cancer in Osaka, Japan. *Cancer* 1990;65:1255–60.
22. Tamura K, Hayami H, Suzuki S, Saito H. Sodium picosulfate's effect of constipation induced barium intake for radiographic screening for gastric cancer. *J Gastroenterol Mass Surv* 1985;69:92–101 (in Japanese).
23. Sugawara N, Hirasawa Y, Morimoto T, Sibuki S, Kogane T, Sato H, et al. An investigative report about health state of old age groups at first screening for gastric mass survey. *J Gastroenterol Mass Surv* 1992;95:184–6 (in Japanese).
24. Watanabe Y, Yokoshima T, Sato M, Akihama T, Orii S, Noro A, et al. A case of anaphylactoid reaction occurring after upper gastrointestinal series by barium sulfate compound. *J Iwate Prefectural Hosp Assoc* 1999;39:37–41 (in Japanese).
25. Sano M, Wada N, Katai H, Maeda K, Sakai S, Kou J, et al. Two cases of barium peritonitis after upper gastrointestinal barium radiography-a review of Japanese literature 44 reported caces. *J Abdominal Emerg Med* 1995;15:423–7 (in Japanese).
26. Kato K, Antoku S, Sawada S, Wada T, Russell WJ. Organ doses to atomic bomb survivors during photofluorography, fluoroscopy and computed tomography. *Br J Radiol* 1991;64:728–33.
27. Maruyama T, Iwai K, Nishizawa K, Noda Y, Kumamoto Y. Organ or tissue dose, effective dose and collective effective dose from X-ray diagnosis, in Japan. *Radioisotopes* 1996;45:761–73 (in Japanese).
28. Maruyama T, Iwai K. Exposure dose and effective dose equivalent for gastric cancer screening. Committee of studies on problems from several view points among mass screening programs for cancer. Annual Report 1993 (Primal researcher Hisamichi S) March 1994,74–80 (in Japanese).
29. Ueda H, Mai M, Asai T, Ohta T, Kann T, Ueno M, et al. A clinico-pathological evaluation of detected cancer between screening and no-screening groups. *J Gastroenterol Mass Surv* 1986;71:52–6 (in Japanese).
30. Fujiya T, Komatsu S, Yamanami H, Mikuni J, Kakugawa Y, Kamiyama Y, et al. Treatment of gastric cancer in Miyagi prefecture; Characteristics of gastric cancer detected by mass-screening and current surgical outcome in Miyagi Cancer Center. *Jpn Soc Gastroenterol Surg* 1998;31:2118–22 (in Japanese).
31. Moki F, Imai T, Abe K, Saotome C, Kawamura O, Takagi H, et al. The situation of stomach cancer in Gunma prefecture that we examined on the basis of population-based cancer registry. *J Jpn Assoc Cancer Detect Diag* 2003;10:145–50 (in Japanese).
32. Kampschöer GH, Fujii A, Masuda Y. Gastric cancer detected by mass survey. Comparison between mass survey and outpatient detection. *Scand J Gastroenterol* 1989;24:813–7.
33. Riecken B, Pfeiffer R, Ma JL, Jin ML, Li JY, Liu WD, et al. No impact of repeated endoscopic screens on gastric cancer mortality in a prospectively followed Chinese population at high risk. *Prev Med* 2002;34:22–8.
34. Hosokawa O, Hattori M, Takeda T, Watanabe K, Fujita M. Accuracy of endoscopy in detecting gastric cancer. *J Gastroenterol Mass Surv* 2004;42:33–9 (in Japanese).
35. Otsuji M, Kouno Y, Otsuji A, Tokushige J, Shimotatara K, Arimura K, et al. Assessment of small diameter panendoscopy for diagnosis of gastric cancer: comparative study with follow-up survey date. *Stomach and Intestine* 1989;24:1291–7 (in Japanese).
36. Kaneko E, Harada H, Kasugai T, Kogosi K, Niwa H. Adverse effect survey committee of Association of Gastroenterological endoscopy. The fourth report of adverse effects for gastrointestinal endoscopy from 1998 to 2002. *Gastroenterol Endosc* 2004;46:54–61 (in Japanese).
37. Watase H, Inagaki T, Yoshikawa I, Furihata S, Watanabe Y, Miki K. Five years follow up study of gastric cancer screening using the pepsinogen test method in Adachi city. *J Jpn Assoc Cancer Detect Diagn* 2004;11:77–81 (in Japanese).
38. Dinis-Ribeiro M, Yamaki G, Miki K, Costa-Pereira A, Matsukawa M, Kurihara M. Meta-analysis on the validity of pepsinogen test for gastric carcinoma, dysplasia or chronic atrophic gastritis screening. *J Med Screen* 2004;11:141–7.
39. Shiga T, Nomoto K, Nisizawa M, Yamaki G, Nagahama R. A study of accuracy of serum pepsinogen method in mass screening and a mass screening using serum pepsinogen. *J Gastroenterol Mass Surv* 2000;38:490–5 (in Japanese).
40. Ikuma H, Mitsushima T. Evaluation of the screening of stomach cancer using serum pepsinogen, compared with eodoscopic findings as the gold standard. *J Gastroenterol Mass Surv* 1998;36:136–44 (in Japanese).
41. Inoue K, Miyoshi E, Aoki S, Itakura S, Mihara O, Yoshihara M, et al. The evaluation of cut-off levels of serum pepsinogens-comprising with gastric endoscopy on the same time in human dry dock. *J Gastroenterol Mass Surv* 1997;35:495–500 (in Japanese).
42. Kitahara F, Kobayashi K, Sato T, Kojima Y, Araki T, Fujino MA. Accuracy of screening for gastric cancer using serum pepsinogen concentrations. *Gut* 1999;44:693–7.
43. Yamanoi A, Hayashi T, Ishihara A, Torisu R, Fujimoto S, Nakamoto Z, et al. Study of the screening of test on gastric cancer. *J Gastroenterol Mass Surv* 1997;35:485–94 (in Japanese).
44. Hattori Y, Tashiro H, Kawamoto T, Kodama Y. Sensitivity and specificity of mass screening for gastric cancer using the measurement of serum pepsinogens. *Jpn J Cancer Res* 1995;86:1210–5.
45. Goto N, Nishiya T, Sakurai Y. A study of gastric mass survey in the employee using serum pepsinogen test (Second Report). *J Gastroenterol Mass Surv* 2004;42:412–7 (in Japanese).

46. Watabe H, Mitsushima T, Yamaji Y, Okamoto M, Wada R, Kokubo T, et al. Predicting the development of gastric cancer from combining *Helicobacter pylori* antibodies and serum pepsinogen status: a prospective endoscopic cohort study. *Gut* 2005;54:764–8.
47. Wong BCY, Lam SK, Wong WM, Chen JS, Zheng TT, Feng RE, et al. *Helicobacter pylori* eradication to prevent gastric cancer in a high-risk region of China: a randomized controlled trial. *JAMA* 2004;291:187–94.
48. Koyama I, Hoshino K, Kubota H, Nakamura J, Machida T, Arai T, et al. Research report on exposure dose in fluoroscopy system. *Jpn J Radiol Technol* 1997;53:609–20 (in Japanese).
49. Kato H, Isobe T, Takagi S, Ochi S, Anzai T, Okumura K, et al. Assessment of institutional differences in radioactivity levels during abdominal x-rays. *Jpn J Radiol Technol* 1999;55:655–64 (in Japanese).
50. Fraiser AG, Lam WM, Luk YW, Sercombe J, Sawyerr AM, Hudson M, et al. Effect of ranitidine bismuth citrate on postprandial plasma gastrin and pepsinogens. *Gut* 1993;34:338–42.
51. Jansen JB, Klinkenberg Knol EC, Meuwissen SG, De Bruijne JW, Festen HP, Snel P, et al. Effect of long-term treatment with omeprazole on serum gastrin and serum group A and C pepsinogens in patients with reflux esophagitis. *Gastroenterology* 1990;99:621–8.
52. Tokieda M, Kodama K, Ito A, Fujiyama K, Kodama R, Kawasaki H, et al. Serum pepsinogen response to therapy for *Helicobacter Pylori* associated gastro-duodenal disease. *J Gastroenterol* 1995;92:1825–31 (in Japanese).
53. Biemond I, Rieu PN, Jansen JB, Joosten HJ, Lamers CB. Prospective study of the effect of gastrectomy with and without bile reflux on serum pepsinogens. *Digestion* 1989;44:124–30.
54. Murakawa M. Influence of impaired renal function and *Helicobacter pylori* infection on serum pepsinogen concentrations. *Jpn J Nephrol* 1999;41:399–405 (in Japanese).
55. Miwa H, Ohkura R, Murai T, Nagahara A, Yamada T, Ogihara T, et al. Effectiveness of omeprazole-amoxicillin-clarithromycin (OAC) therapy for *Helicobacter pylori* infection in a Japanese population. *Helicobacter* 1998;3:132–8.
56. Asaka M, Sugiyama T, Kato M, Satoh K, Kuwayama H, Fukuda Y, et al. A multicenter, double-blind study on triple therapy with lansoprazole, amoxicillin, clarithromycin for eradication of *Helicobacter pylori* in Japanese peptic ulcer patients. *Helicobacter* 2001;6:254–61.
57. Kato M, Yamaoka Y, Kim JJ, Reddy R, Asaka M, Kashima K, et al. Regional differences in metronidazole resistance and increasing clarithromycin resistance among *Helicobacter pylori* isolates from Japan. *Antimicrob Agents Chemother* 2000;44:2214–6.
58. Miller AB, Chamberlain J, Day NE, Hakama M, Prorok PC. Report on a workshop of the uicc project on evaluation of screening for cancer. *Int J Cancer* 1990;46:761–9.
59. European Code Against Cancer, Third edition. http://www.cancercode.org/add_items.htm (2005.9.6 access).
60. Association of Medical Screening. Screening brief: radiographic screening for stomach cancer. *J Med Screen* 2000;7:120.
61. Lee KJ, Inoue M, Otani T, Iwasaki M, Sasazuki S, Tsugane S, et al. Gastric cancer screening and subsequent risk of gastric cancer: a large-scale population-based cohort study, with a 13-year follow-up in Japan. *Int J Cancer* 2006;118:2315–21.
62. Miyamoto A, Kuriyama S, Nishino Y, Tsubono Y, Nakaya N, Ohmori K, et al. Lower risk of death from gastric cancer among participants of gastric cancer screening in Japan: a population-based cohort study. *Prev Med* 2007;44:12–9.
63. Yoshihara M, Hiyama T, Yoshida S, Ito M, Tanaka S, et al. Reduction in gastric cancer mortality by screening based on serum pepsinogen concentration: a case-control study. *J Scan Gostroenterol* 2007;42:760–4.
64. You WC, Brown LM, Zhang L, Chang YS, Ma JL, Pan KF, et al. Randomized double-blind factorial trial of three treatments to reduce the prevalence of precancerous gastric lesions. *J Natl Cancer Inst* 2006;98:974–83.

Appendix:

Review Committees for Gastric Cancer Screening

Hamashima C, Saito H, Sano H (National Cancer Center), Tanaka J (Hiroshima University), Shibuya D (Miyagi Cancer Association), Honjyo S (Tochigi Cancer Center), Matsuda K (Fukui Health Management Association), Nishida H (Matsushita Health Management Center), Yamazaki H (Osaka Cancer Screening Center), Inoue K (Matsue Red Cross Hospital) and Sasaki S (St Ruka Hospital).

Peer Review Committee for the Japanese Gastric Cancer Screening Guideline

Koizumi M (Saga University), Ishikawa T (Tochigi Cancer Center), Imamura K (Yokohama City Hospital), Suko H (Saiseikai Kumamoto Hospital), Yoshino J (Fujita Medical University), Muto K (Shinsyu University), Kakihara H (Ritsumeikan University), Sato T and Ibukuro C (Yamagata Tuberculosis Prevention Association).

Japanese Research Group for Development of Cancer Screening Guidelines

Sobue T (Chief, National Cancer Center), Hamashima C, Saito H (National Cancer Center), Nakayama T (Kyoto University), Fukao A (Yamagata University), Tsuji I, Tsubono Y (Tohoku University), Onuki Y (Iwate Prefecture Central Hospital), Sagawa M (Kanazawa Medical University), Aoki D (Keio University), Nakayama T (Osaka Medical Center for Cancer and Cardiovascular Disease), Watanabe Y (Kyoto Prefecture Medical University), Tanaka J (Hiroshima University), Daisuke Shibuya, Shimada G (Miyagi Cancer Association), Honjyo S (Ohmuta National Hospital), Matsuda K (Fukui Health Management Association), Nishida H (Matsushita Health Management Center), Hideo Yamazaki (Osaka Cancer Screening Center), Kazuhiko Inoue (Matsue Red Cross Hospital) and Sasaki S (St Ruka Hospital).

Jpn J Clin Oncol 2008;38(4)288–295
doi:10.1093/jjco/hyn016

The Standardized Development Method of the Japanese Guidelines for Cancer Screening

Chisato Hamashima[1], Hiroshi Saito[1], Tomio Nakayama[2], Takeo Nakayama[3] and Tomotaka Sobue[4]

[1]Cancer Screening Technology Division, Research Center for Cancer Prevention and Screening National Cancer Center, Tokyo, [2]Department of Cancer Control and Statistics, Osaka Medical Center for Cancer and Cardiovascular Diseases, Osaka, [3]Department of Health Informatics, Kyoto University School of Public Health, Kyoto and [4]Cancer Information Services and Cancer Surveillance Division, Center for Cancer Control and Information Services, National Cancer Center, Tokyo, Japan

Received November 13, 2007; accepted February 10, 2008; published online March 12, 2008

Background: To reduce cancer mortality, effective screening should be implemented properly. In Japan, the Research Group for Cancer Screening developed screening guidelines; however, the development process was not well established.
Methods: Based on the development processes of other guidelines, an original method, unique to Japan, was established to develop the Japanese cancer screening guidelines.
Results: The guideline development process involved the following steps: topic selection, development of the analytic framework, systematic literature review, translation to recommendations, consultation and publication. Mortality reduction related to cancer screening was evaluated using both direct and indirect evidence. To select appropriate articles, an analytic framework for cancer screening program with key questions was developed. Direct evidence was defined as a single body of evidence that established the linkage between screening and health outcomes such as mortality and incidence. The use of indirect evidence to determine the level of evidence was limited to situations where test accuracy could be compared with that of a method whose evidence was supported by randomized, controlled trials. Eight levels of evidence were defined based on the study design and quality. The benefits of each screening modality were determined based on the level of evidence according to the results of the systematic review. Balancing the benefits and harms, five grades of recommendation were formulated for population-based and opportunistic screening. After organized consultations, three types of guidelines were published.
Conclusion: We developed a unique, standardized method for developing cancer screening guidelines in Japan. Based on this process, previously developed cancer screening guidelines have been revised.

Key words: cancer screening – guideline – mortality reduction – population-based screening – opportunistic screening

INTRODUCTION

Cancer control is a major issue, the goal of which is to reduce the burden of the disease. In 2006, the Cancer Control Act, which was aimed at reducing the incidence and mortality associated with cancer in Japan, was approved. Cancer screening programs are one of the effective strategies for reducing mortality; however, to achieve this goal, evidence-based screening programs need to be implemented properly.

As well as clinical practice guidelines, cancer screening guidelines can assist physicians and other health professionals and policy-makers to make effective decisions. In 2001, the research group for cancer screening funded by the Ministry of Health and Welfare in Japan recommended the following six cancer screening programs (the Hisamichi reports) (1): gastrofluorography for gastric cancer, fecal occult blood testing (FOBT) for colorectal cancer, a combination of chest radiography and sputum cytology (focused

For reprints and all correspondence: Chisato Hamashima, Cancer Screening Technology Division, Research Center for Cancer Prevention and Screening, National Cancer Center, 5-1-1 Tsukiji Chuo-ku, Tokyo 104-0045, Japan. E-mail: chamashi@ncc.go.jp

on high-risk group) for lung cancer, Pap smear for cervical cancer, a combination of physical examination and mammography for breast cancer and hepatitis virus markers for hepatocellular carcinoma. Although the Hisamichi reports were based on the evidence for cancer screening, they were, in some respects, immature as guidelines. The details of the development process, including the systematic review process, were not standardized. Although the recommendations were not graded, the level of evidence was shown based on the method used in the US Preventive Task Force (USPTSF) 2nd edition. However, the method used to evaluate the efficacy of cancer screening was limited to the study design (2). Since there are insufficient rules for guideline development, different conclusions could be reached with respect to each of the cancer screening programs.

A good guideline should be the product of an established process. Details about the process of guideline development have been published by several guideline development institutes in the US and Europe (3–7). To promote evidence-based screening in Japan, there is a need to standardize the development method of cancer screening guidelines and to disseminate appropriate information thorough the guidelines.

METHODS

To clarify the evidence related to each cancer screening program, the new guideline development process was standardized. The USPTSF (3), the Scottish Intercollegiate Guideline Network (SIGN) (4) and the Appraisal of Guidelines, Research and Evaluation in Europe instrument (AGREE) (8) were used as the primary sources of information. On the basis of these guideline development processes, an original method, unique to Japan, was created for the evaluation of cancer screening programs.

RESULTS: A STANDARDIZED METHOD FOR GUIDELINE DEVELOPMENT

Outlines

Our guidelines target the public, health professionals working in cancer screening programs and policy-makers. An outline of our guideline development process is shown in Fig. 1; it includes topic selection, panel composition, development of the analytic framework and setting of key questions, systematic literature review (literature search, abstract review and full text review), determination of the level of evidence, translation to recommendations, formulation of the draft guideline, consultation (peer review and national open meeting) and publication of the guideline. Overall, these processes require 12–18 months. All of the guidelines are scheduled for revision within 5 years, so that new evidence can be incorporated.

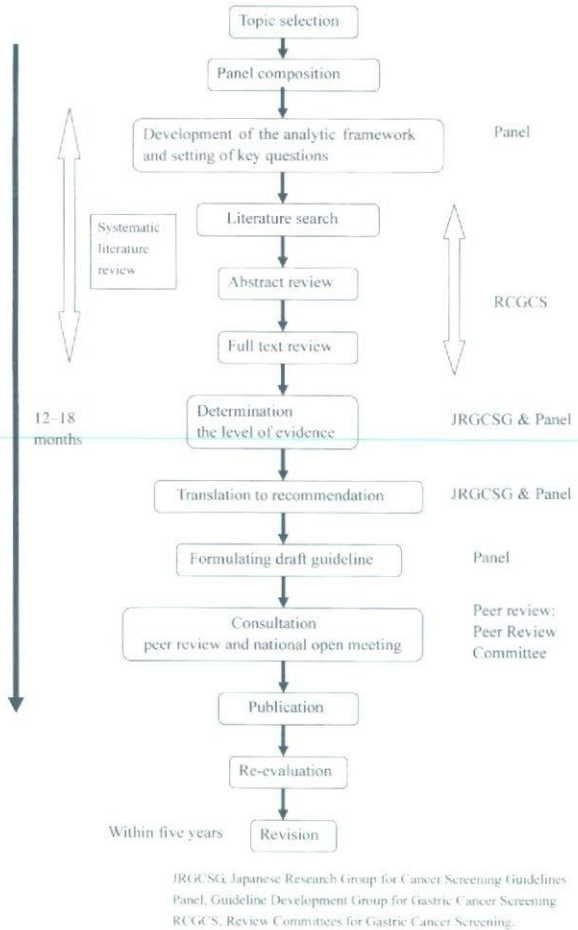

Figure 1. Cancer screening guideline development process.

Topic Selection

The topics of the cancer screening guidelines deal with the current population-based screening programs: colorectal, gastric, lung, breast and cervical cancer screening. In every screening program, several methods that have been used for cancer screening are evaluated. Then, a schedule is developed to assess the common screening programs used by certain local municipalities and in various clinical settings (e.g. prostate cancer screening using prostate specific antigen).

Panel Composition

The members of the guideline development group (Panel) include physicians working in cancer screening programs, researchers including epidemiologists and health economists and board members of the Japanese Research Group for the Development of Cancer Screening Guidelines (JRGCSG). A systematic literature review is conducted by the members of the review committees, including the members of the Panel,

for each specific cancer screening program. The recommendations are assessed in conjunction with the board members of the JRGCSG.

Analytic Framework

An analytic framework demonstrates the chain of logic that must be supported by evidence that links screening to improved health outcomes. Both the USPTSF and the Community Guide use an analytic framework to select evidences (3,7). We use basically the same flowchart to evaluate the cancer screening programs (Fig. 2). If needed, the flowchart is modified so as to be appropriate to the Japanese context. This could be a useful tool for mapping out the plan used to evaluate each screening program, which guides the search for evidence. For each stage of the analytic framework, key questions are prepared; these questions must be clear and focus on the main issues related to each stage of the analytic framework. The key questions are based on the population, intervention, comparison and outcome format.

Determination Of The Level Of Evidence

Direct and indirect evidences are used to evaluate mortality reduction related to cancer screening. Direct evidence is defined as a single body of evidence that establishes the connection between screening and health outcomes, such as mortality and incidence (mainly late-stage cancer) (Fig. 2, arrow 1). Direct evidence could be provided by randomized, controlled trials (RCTs) and observational studies such as cohort and case–control studies. Other studies that provide indirect evidences are selected based on the key questions related to other stages of the analytic framework (Fig. 2, arrows 2–8). Intermediate outcomes are often used as indicators of efficacy, but they are not direct measures of mortality reduction; studies that use intermediate outcomes may be useful for offering indirect evidence of efficacy. The use of indirect evidence to determine the level of evidence was limited to situations where the test accuracy could be compared with methods for which the evidence was supported by RCTs. However, the presence of harmful effects of screening and treatment is a significant factor for deciding the recommendation grade. Data from studies that dealt with the key questions of the analytic framework are collected (Fig. 2, arrows 4 and 6).

Literature Search

A systematic literature review is performed to identify articles relevant to assessing specific cancer screening programs. Literature searching is done from 1985 to the present using MEDLINE, EMBASE, CINHAL, the Cochrane Collaboration Library, the Database of Abstracts of Reviews of Effectiveness and the Japanese Medical Research Database (Igaku-Chuo-Zasshi). At a minimum, both MEDLINE and Igaku-Chuo-Zasshi are searched. A manual search of key journals related to cancer screening, mainly published in Japan, is performed. In addition, reference lists of the Hisamichi reports, as well as other systematic reviews and guidelines dealing with the same topic, are identified and included as needed.

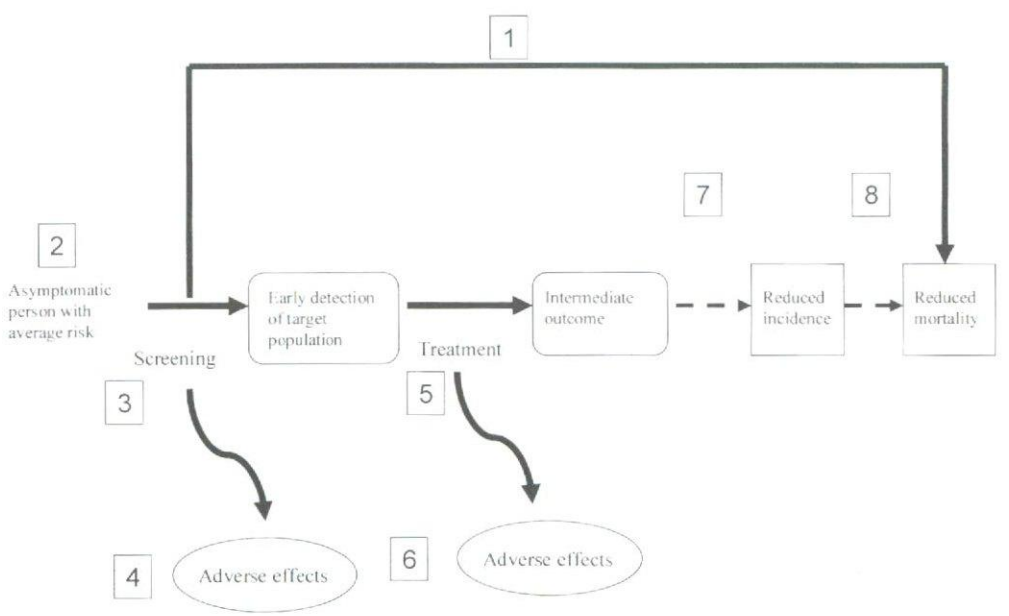

Figure 2. Analytic framework and key questions.

Literature Review (Abstract and Full Text Review)

The abstracts of all articles identified by the literature search are reviewed. Each abstract is independently evaluated by two members of the review committee to assess the evidences. The articles for which the abstracts meet the inclusion criteria of both reviewers are retrieved. If the reviewers do not agree on the assessment of a particular article, then adoption of the article is discussed at a meeting of the guideline development group. The main inclusion criteria are: (i) original article, not including abstracts or meeting/conference proceedings; (ii) the target group studied in the article must be limited to asymptomatic persons with an average risk, not symptomatic patients and (iii) the article adequately answers the key questions. The reports dealing solely with detection rates in a local area or in a clinical setting are excluded.

Once the articles have been selected for full text review, independent reviewers appraise the quality of each article using a checklist for each study design. The checklist is a modified version of SIGN (4), which is used to evaluate the efficacy of cancer screening program. There are seven types of checklists: systematic review, RCT, case–control study, cohort study, test accuracy, ecological and time-series studies and others. Case series studies regard to harm are evaluated using the 'other' checklist. Each checklist item is scored using five categories: adequately addressed (score 4), addressed (score 3), poorly addressed (score 2), not reported (score 1) and not applicable. Articles with a total score of >60% are adopted. Although these criteria are the basic requirement for selecting adequate articles, low-quality articles could be chosen if there is a general lack of evidence on the topic. Although most studies concerning harms tend to be low quality (e.g. case series), these studies could be accepted in the absence of other evidences.

Level of Evidence

For each key question of the analytic framework, the body of evidence for each screening method is summarized in an evidence table. Evidence is obtained from studies that evaluate mortality reduction as a result of cancer screening. There are eight levels of evidence based on the study design and quality (Table 1). The level of evidence of the SIGN (4) was modified. The highest level of evidence is considered to be a high-quality RCT; expert opinion is considered to provide the lowest level of evidence. Although the most relevant study is considered to be an RCT, the level of evidence is divided into three levels based on the quality of the study. Ecological and time-series studies are limited due to attenuating effects or difficulties in controlling confounding factors; however, in both the UK National Institute for Clinical Excellence [presently, National Institute for Health and Clinical Excellence (NICE)] public health guidelines and the Community Guide (6,7), these studies have been defined as providing significant evidence. In our guidelines, ecological and time-series studies are considered to provide the same level of evidence as medium-quality case–control and cohort studies. The use of indirect evidence to determine the level of evidence is limited to situations where the test accuracy can be compared with that of methods for which evidence is supported by RCTs.

Definition of Cancer Screening Programs

There are two types of screening: population-based and opportunistic screening. The features are summarized in Table 2. In Japan, as in most countries, both types of cancer screening are prevalent (9). Although the aim of both screening programs is to reduce cancer mortality, they are different in many aspects, particularly with respect to their significance in the anti-cancer strategy. Population-based screening programs have been mainly conducted as a preventive policy in local municipalities with government support. Organized screening, which is considered to be an ideal system of population-based screening, involves centralized responsibility for the program's process, such as registration for eligibility, quality assurance follow-up and evaluation. In this regard, population-based screening in Japan has not matured as organized screening. There has been firm evidence that organized screening could achieve mortality reduction from the corresponding cancer. In contrast, opportunistic screening depends on individual members of the public requesting screening or their health advisors recommending it. This type of program is also common and has been conducted using various modalities in clinical settings, even when there is insufficient evidence for efficacy. In addition, quality assurance is lacking in opportunistic screening. Therefore, this type of screening is generally not preferred as a screening strategy to reduce mortality.

Translation into Recommendation

On the basis of balance of benefits and harms, recommendations are formulated. The benefit of each screening modality is determined based on the level of evidence. In contrast, the harms, including the false-negative rate, the false-positive rate and the burden for cancer screening participants, are compared among the various methods. If there is a serious issue relating to harms, any recommendation should be tempered by considering the impact of the harm.

Five grades of recommendation are defined and applied for the two types of screening (Table 3). Since there is sufficient evidence, both Grade A and B recommendations could be directly applied to the target population as both population-based and opportunistic screening programs. However, a Grade D recommendation implies that the method should not be used for either population-based or opportunistic screening programs. A grade C recommendation implies that the method should not be used for population-based screening; although there is sufficient evidence, important harms that cannot be ignored could occur. Thus, considering the balance

Table 1. Level of evidence

Level	Study design	Language
1++	RCT	High-quality RCTs with a very low risk of bias. Overall consistency of the results is needed.
	Systematic review	High-quality meta-analysis or systematic review of RCTs
1+	RCT	Medium-quality RCTs with a low risk of bias. Overall consistency of the results is needed.
	Systematic review	Medium-quality meta-analysis or systematic review of RCTs, case–control or cohort studies.
	Combination based on selected studies by each stage on analytic framework	Medium-quality RCTs dealing with an important stage of the analytic framework and high-quality case–control or cohort studies
1−	RCT	Low-quality RCTs with a high risk of bias
	Systematic review	Low-quality systematic review of RCTs, case–control or cohort studies with a high risk of bias
2++	Cohort study/case–control study	High-quality case–control or cohort studies with very low risks of bias, confounding or chance and a high probability that the relation is causal. Overall consistency of the results is needed
2+	Cohort study/case–control study	Medium-quality case–control or cohort studies with a low risk of bias, confounding or chance and a moderate probability that the relation is causal. Overall consistency of the results is needed
	Time-series/ecological study	Well-conducted time-series or ecological study studies. Overall consistency of the results is needed
	Combination based on selected studies by each stage on analytic framework	RCTs dealing with an important stage of the analytic framework and several studies concerning test accuracy and survival rate, etc at various stages of the analytic framework
2−	Cohort study/case–control study	Low-quality case–control or cohort studies with a high risk of confounding, bias, or chance and a high probability that the relation is not causal
	Time-series/Ecological study	Low-quality time-series or ecological studies
	Combination based on selected studies by each stage on analytic framework	Several studies concerning test accuracy and survival rate, etc at various stages of analytic framework
3	Non-analytic study	Case reports, case series, etc
4	Expert opinion	

RCT, randomized controlled trial

Table 2. Comparison of population-based and opportunistic screenings*

Screening program	Population-based screening**	Opportunistic screening
Aim	Reduce cancer mortality at the population level	Reduce cancer immortality at the individual level
Provider	Decision-maker for health plan of local municipalities and workplace	Variable
Outline	Public service for prevention	Spontaneous medical services in clinical settings
Target population	Specified: all members limited to specific age range.	Variable: limited to asymptomatic person
Screening test	Method to reduce mortality from specific cancer: chosen based on the cancer screening guidelines by local municipality and workplace	Variable: chosen based on individual preference***
Sensitivity of test	The most sensitive test may not be chosen	The most sensitive test is usually chosen
Specificity of test	High specificity is important for reducing unnecessary work up of false-positive results and associated adverse effects	High specificity is less important at individual level
Quality assurance	Continues monitoring	Not be monitored
Available financial resources	Limited at the population level in relation to the politics of health spending, taking into account all aspects of health care	Limited at the individual level
Benefits	Maximized for the population within available resources	Maximized for the individual
Harms	Minimized for the population to avoid	Not necessarily minimized

*The table was customized for the Japanese situation based on the article by Miles et al. (*Cancer* 2004;101:1201–13).
**Organized screening is conducted systematic invitation based on specific target and monitoring for quality. In Japan, population-based screening is still in its early stages as a form of organized screening.
***In opportunistic screening, the screening method should be chosen based on the cancer screening guidelines. If this is impossible, the benefits and harms of specific cancer screening programmes must be explained.

Table 3. Recommendation grades

Recommendation	Population-based screening	Opportunistic screening	Level of evidence
A	The JRGCSG strongly recommends the use of the method based on sufficient evidence which evaluated mortality reduction from the specific cancer by cancer screening	The JRGCSG strongly recommends the use of the method based on sufficient evidence that evaluated mortality reduction from the specific cancer by cancer screening	1++/1+
B	The JRGCSG recommends the use of the method based on fair evidence that evaluated mortality reduction from the specific cancer by cancer screening	The JRGCSG recommends the use of the method based on fair evidence that evaluated mortality reduction from the specific cancer by cancer screening	2++/2+
C	The JRGCSG recommends against the use of the method.	The JRGCSG judges that this method could be used; both adequate risk management and informed consent about harms are needed.	1++/1+
			2++/2+
D	The JRGCSG recommends against routine use based on sufficient evidence.	The JRGCSG recommends against routine use based on sufficient evidence.	1++/1+
			2++/2+
I	The JRGCSG recommends against routine use due to insufficient evidence.	The JRGCSG recommends against routine use due to insufficient evidence. If the patient requests, decision-making should be done at the individual level based on appropriate information.	1−/2−/3/4

JRGCSG, Japanese research group for cancer screening guidelines

of the benefits and the harms, these could not be recommended for population-based screening. However, a Grade C recommendation implies that the method could be used in clinical settings if both adequate risk management and informed consent about the harms were assured. Programs for which there is insufficient evidence are graded as I; they are not recommended for routine population-based screening or as screening methods in clinical settings, although the decision to undergo screening could be made at the individual level based on appropriate information provided by health professionals in clinical settings.

The guideline development group and all members of the JRGCSG are involved in allocating the recommendation grades. All members must agree with the recommendation; any major issues discussed in the meeting are noted in the guidelines.

Formulating Guideline

A draft guideline is prepared based on the recommendation of the Research Group for Cancer Screening Guidelines. The members of the Panel for each guideline program write the draft. The draft is evaluated by peer reviewers and at a national open meeting. All guidelines are reviewed in draft form by eight independent referees from two expert groups: an expert group dealing with screening, diagnosis and treatment for the specific cancer and another specialty group, such as general practitioners or experts in public health, disease management, epidemiology and health technology assessment. For peer reviews, the AGREE instrument (8) and the checklist proposed by the Conference on Guideline Standardization (10) were modified to evaluate the cancer screening guidelines. Questions about the appropriateness of the level of evidence and the recommendation grade were added. A 27-item checklist is used for peer review. The draft guideline is available on our web site (Promoting Evidence-based Cancer Screening: http://canscreen.ncc.go.jp/) for a limited period prior to official publication. The Research Group for Cancer Screening Guidelines holds a national open meeting to discuss each draft guideline. Taking into account the comments received, the appropriateness of the recommendation is discussed and the guideline is refined.

Publication

The guidelines are published in several forms. The full text version contains the recommendations, details of how they were developed and information about the evidence on which they were based. The concise version includes the recommendations and short comments about the evidence. The non-specialist version is mainly targeted at public health nurses who work in local municipalities; it can be easily understood and can be used to provide information to cancer screening participants. All of the guidelines are posted on the following website: Promoting Evidence-based Cancer Screening and Research Center for Cancer Prevention and Screening, National Cancer Center (http://ganjoho.ncc.go.jp/pro/index.html).

DISCUSSION

Cancer screening guidelines are required to promote effective cancer screening to primarily achieve mortality reduction from a specific cancer, or to reduce the incidence of late-stage cancer in some cases, such as cervical cancer

screening. In Japan, the Research Group for Cancer Screening Guidelines refined the development process for cancer screening guidelines. Compared with previous reports, the Japanese Guidelines for Cancer Screening development process has certain unique characteristics. Our guidelines include a system of recommendation grading that is divided into population-based and opportunistic screening programs. It provides a consistent and transparent process for evaluating cancer screening based on a standardized method. Using the recommendations, effective cancer screening programs can be developed.

Screening program requirements differ among countries due to differences in the incidence and mortality of specific cancers. Although global evidence is a preferable starting point for evaluating the efficacy of a particular screening test, local evidence is needed to determine what should be done in a specific setting (11). In Japan, gastric and lung cancer screening are major issues due to the high incidence and mortality of these cancer (12). Although an RCT is considered to be the most reliable study design to evaluate cancer screening, the results of these studies are limited to screening for specific cancers. Thus, we also use evidence obtained from case–control and cohort studies conducted in Japan. Additionally, indirect evidence is used to formulate recommendations; however, the use of indirect evidence is strictly limited to the situation where the accuracy of a new screening method can be compared with that of an established method that is supported by RCTs. For example, this process was followed for determining the recommendation for immunological FOBT. The efficacy of chemical FOBT screening was evaluated by three RCTs. However, the evidence for the efficacy of immunological FOBT was considered weak, because the efficacy was evaluated by observational studies. On the basis of this process, the accuracy of both tests was compared to assess the value of immunological FOBT screening. Thus, immunological FOBT screening could be adopted as a Grade A recommendation.

The recommendations grading systems are based on the best available evidence. If the evidence is strong, the process should be straightforward, and the evidence should translate directly into recommendations. However, when there is disagreement among the members of the guideline development group, a consensus must be reached. For several guidelines, if there is poor evidence, formal consensus methods are adopted (13,14). Using a formal approach that is explicit and transparent, it is possible to trace how the group reached a decision. In contrast, the GRADE group defined a new system that takes into account the study design, study quality, consistency and directness (15). We experienced no difficulties in our process of grading the recommendations; however, a clear process is needed that considers all of these methods to ensure that difficulties can be resolved.

Despite the use of the latest methods, our guideline development process does have certain limitations. First, economic evaluations are not used to determine recommendations. NICE clarified the need to introduce health policy from the perspective of health economics (6). It is important to consider costs when developing guidelines, because resources are always limited (16). An economic analysis can provide useful information for improving health outcomes in an environment with limited resources. Secondly, service user involvement needs to be considered. Although several comments from the general public can be received at the national open meetings, the opportunity for service user involvement is limited in our process. On the other hand, many clinical guidelines have included patient involvement (3,17). Thirdly, to promote informed decision-making for cancer screening, a version of the guidelines that is accessible to the public needs to be developed (18,19).

Finally, our guideline development process does not deal with issues surrounding implementation. Evidence-based medicine requires evidence-based implementation (20). In Japan, population-based screening has not matured due to a lack of key elements required for organized screening. It is impossible to reduce mortality from a specific cancer without an appropriate implementation system that includes quality assurance. Implementation of a guideline should be monitored and evaluated through clinical audit (4). A monitoring system could be conducted that would be limited to population-based screening programs. The European Community Guidelines defined performance indicators for mammography screening (21); they have been used to assess the implementation of cancer screening programs (22,23). Thus, appropriate performance indicators need to be developed for the quality assurance of cancer screening programs that are based on our guidelines.

To reduce cancer mortality, effective screening is required that is based on the cancer screening guidelines. Recommendations based on a rigorous process of guideline development are needed. We standardized the original method that was used for the development of the Japanese Guidelines for Cancer Screening. Based on this method, colorectal, gastric and lung cancer screening guidelines were revised. Nevertheless, the development process needs to be refined further based on international collaboration of guideline development and its applicability to the Japanese Guidelines for Cancer Screening.

Acknowledgment

We thank Kanoko Matsushima and Junko Asai for secretarial support.

Funding

This study was supported by Grant-in-Aid for Cancer Control from Ministry of Health, Labor and Welfare of Japan (Grant number 15-3).

Conflict of interest statement
None declared.

References

1. Hisamichi S. editors. Guidelines for Cancer Screening Programs. Tokyo: Japan Public Health Association 2001 (in Japanese).
2. U.S. Preventive Services Task Force. Guide to Clinical Preventive Services. 2nd edn. Baltimore: Williams & Wilkins 1996.
3. Harris RP, Helfand M, Woolf SH, Lohr KN, Mulruw CD, Teuysh SM, et al. Current methods of the U.S. Preventive Services Task Force: a review of the process. *Am J Prev Med* 2001;20(Suppl 3):21–35.
4. Scottish Intercollegiate Guidelines Network. SIGN50: a Guideline Developer's Handbook. Edinburgh: SIGN 2001.
5. National Institute for Clinical Excellence. The Guideline Development Manual. London: NICE 2006.
6. National Institute for Health and Clinical Excellence. Methods for Development of NICE Public Health Guidance. London: NICE 2006.
7. Briss PA, Zarra A, Psappaioanou M, Fielding J, Hopkins DP, Woolf SH, et al. Developing an evidence-based guide to community preventive services-methods. *Am J Prev Med* 2000;18(Suppl 1):35–43.
8. AGREE Collaboration. Appraisal of Guidelines for Research and Evaluation (AGREE) Instrument. London: St. George's Hospital Medical School 2001.
9. Miles A, Cockburn J, Smith RA, Wardle J. A perspective from countries using organized screening programs. *Cancer* 2004;101: 1201–13.
10. Shiffman RN, Scekelle P, Overhage JM, Slutsky J, Grimshaw J, Deshpande AM. Standardized reporting of clinical practice guideline: a proposal from the conference on guideline standardization. *Am Intern Med* 2003;139:493–8.
11. Oxman AD, Schunemann HJ, Fretheim A. Improving the use of research evidence in guideline development: 7. Deciding what evidence to include. *Health Res Policy Syst* 2006;4:19.
12. Nomura K. editor. Cancer Statistics in Japan 2005. Tokyo: Foundation for Promotion Cancer Research 2006.
13. Murphy MK, Black NA, Lamping DL, Mckee CM, Sanderson CF, Askham J, et al. Consensus development methods, and their use in clinical guideline development. Health Technol Assess 1998;2(3):i–iv, 1–88.
14. National Institute for Clinical Excellence. Preoperative Tests: The Use of Routine Preoperative Tests for Elective Surgery. London: NICE 2003.
15. Grade of Recommendation, Assessment, Development, and Evaluation (GRADE) Working Group. Grading quality of evidence and strength of recommendation. *BMJ* 2004;328:1490–4.
16. Guyatt G, Baumann M, Pauker S, Halperin J, Maurer J, Owens DK, et al. Addressing resource allocation issues in recommendations from clinical practice guideline panels. *Chest* 2006;129:182–7.
17. National Institute for Health and Clinical Excellence. A guide for patients and carers: contributing to a NICE clinical guideline. 2006.
18. Sheridan SL, Harris RP, Woolf SH. Shared decision making about screening and chemoprevention. *Am J Prev Med* 2004;26(1): 56–66.
19. Briss P, Rimer B, Reilley B, Coates RC, Lee NC, Mullen P, et al. Promoting informed decisions about cancer screening in communities and healthcare systems. *Am J Prev Med* 2004;26(1):67–80.
20. Grol R, Grimshow J. Evidence-based implementation of evidence-based medicine. *Jt Comm J Qual Improv* 1999;25:503–13.
21. The European Community. European Guidelines for Quality Assurance in Mammography Screening. Luxembourg: Europe Against Cancer Programme, Office for official Publications of the European Communities 2006.
22. Sarkeala T, Anttila A, Forsman H, Luostarinen T, Saarenmaa I, Hakama M. Process indicators from ten centers in the Finnish breast cancer screening programme from 1991 to 2000. *Eur J Cancer* 2004;40:2116–25.
23. Blanks RG, Moss SM, Wallis MG. Monitoring and evaluating the UK National Health Services Breast Screening Programme: evaluating the variation in radiological performance between individual programmes using PPV-referral diagrams. *J Med Screen* 2001;8:24–8.

Japanese Research Group for Development of Cancer Screening Guidelines

Sobue T (Chief, National Cancer Center), Hamashima C, Saito H (National Cancer Center), Nakayama T (Kyoto University), Fukao A (Yamagata University), Tsuji I, Tsubono Y (Tohoku University), Ohnuki Y (Iwate Prefecture Central Hospital), Sagawa M (Kanazawa Medical University), Aoki D (Keio University), Nakayama T (Osaka Medical Center for Cancer and Cardiovascular Disease), Watanabe Y (Kyoto Prefecture Medical University), Tanaka J (Hiroshima University), Shimada G (Miyagi Cancer Association), Honjyo S (Ohmuta National Hospital), Matsuda K (Fukui Health Management Association), Nishida H (Matsushita Health Management Center).

がん検診の現状と課題

国立がんセンター がん予防・検診研究センター
検診研究部　濱島　ちさと

はじめに

がん検診によってがん死亡を減少させるためには、有効性の確立した検診を正しく実施する必要がある。このため、有効性評価に基づくがん検診ガイドラインの作成が進められている。また、がん検診の精度には都道府県格差があり、精度管理のためのシステムが整備されつつある。

がん検診の現状

1982年(昭和57)の老人保健法施行以来、市区町村では老人保健事業によるがん検診が行われてきた。82年度から、胃がん及び子宮頸がん検診が開始し、続いて、肺がん、乳がん、大腸がん検診が行われている。1998年度(平成10)から、がん検診は一般財源化され、検診の実施、検査方法の選択

図1　がん検診受診率の推移

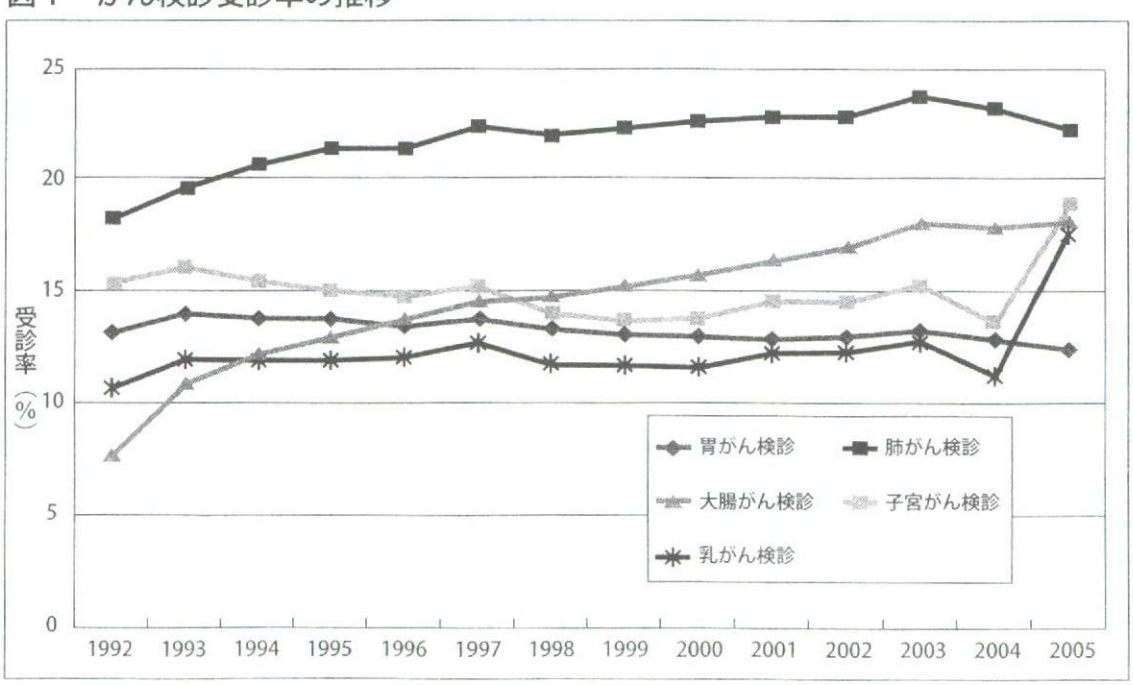

図2　諸外国との受診率の比較

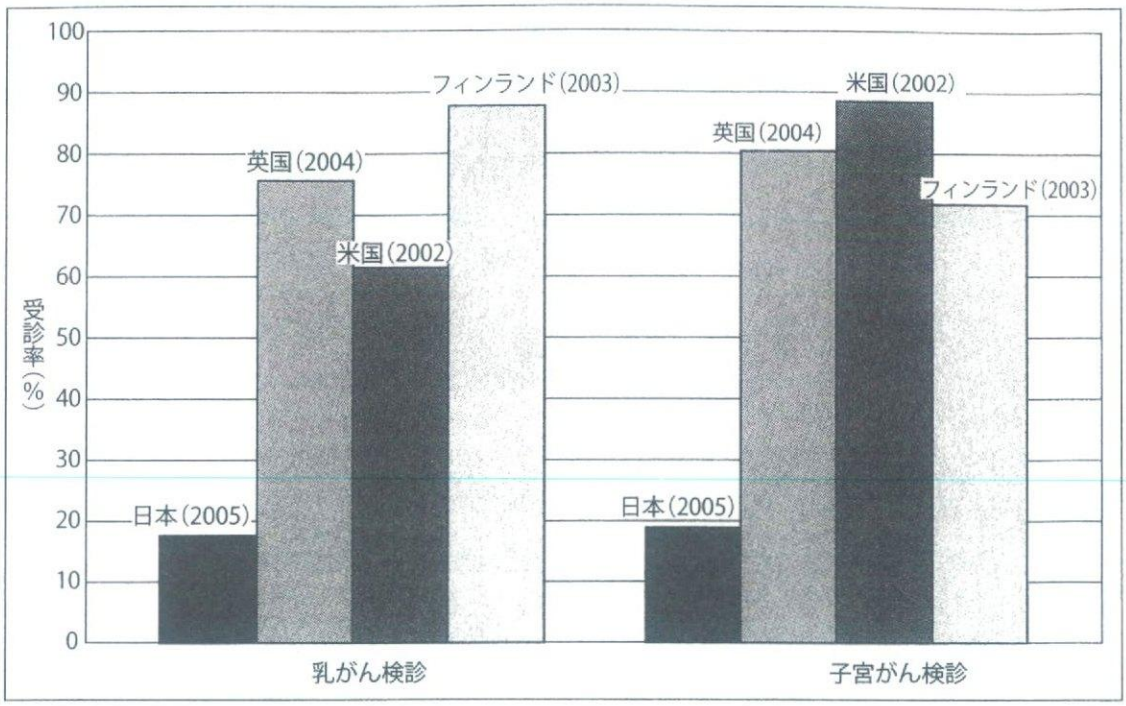

表1　がん検診の実績

がん検診	胃がん	大腸がん	肺がん	乳がん	子宮頸がん
検査方法	胃X線	便潜血	胸部X線	視触診及びマンモグラフィ	細胞診
受診者数(人)	4,344,918	6,630,503	6,963,844	1,604,557	3,439,094
がん発見率(%)	0.15	0.17	0.04	0.27	0.06
要精検率(%)	10.8	7.2	2.8	8.9	1.2
精検受診率(%)	74.6	54.5	72.3	79.9	62.6

(平成17年　地域保健・老人保健事業報告)

などは市区町村の判断に委ねられている。

いずれの検診においても、受診率は、最近10年間は横ばいである[1](**図1**参照)。英国、米国における受診率は、乳がん検診では60%以上、子宮がん検診では80%以上と高い[2~4](**図2**参照)。検診の対象年齢や算出方法の相違はあるが、わが国におけるがん検診の受診率は極めて低い。

2005年(平成17)の地域保健・老人保健事業報告によるがん検診の実績を、**表1**に示した。

対策型検診と任意型検診

わが国におけるがん検診の実施体制は、住民検診型の対策型検診と人間ドック型の任意型検診に大別される[5](**表2**参照)。

対策型検診とは、集団全体の死亡率減少を目的として実施するものを指し、公共的な予防対策として行われる。偶発症や受診者の心理的・身体的負担などの不利益を最小限とすることが基本条件となる。市区町村が行う老人保健事業による集団検診・個

表２　対策型検診と任意型検診の比較

検診方法	対策型検診　注1） （住民検診型）	任意型検診　注2） （人間ドック型）
	Population-based screening	Opportunistic screening
目的	対象集団全体の死亡率を下げる	個人の死亡リスクを下げる
概要	予防対策として行われる 公共的な医療サービス	医療機関・検診機関等が 任意で提供する医療サービス
検診対象者	構成員の全員 （一定の年齢範囲の住民など）	定義されない
検診費用	公的資金を使用	全額自己負担
利益と 不利益	限られた資源の中で、 利益と不利益のバランスを考慮し、 集団にとっての利益を最大化	個人のレベルで、 利益と不利益のバランスを判断

注１）対策型検診では、対象者名簿に基づく系統的勧奨、精度管理や追跡調査が整備された組織型検診（Organized Screening）を行うことが理想的である。
ただし、現段階では、市区町村や職域における対策型検診の一部を除いて、組織型検診は行われていない。

注２）任意型検診の提供者は、死亡率減少効果の明らかになった検査方法を選択することが望ましい。
がん検診の提供者は、対策型検診では推奨されていない方法を用いる場合には、死亡率減少効果が証明されていないこと、及び、当該検診による不利益について十分説明する責任を有する

別検診や、職域の法定健診に付加して行われるがん検診が該当する。

対策型検診では、対象者名簿に基づく系統的勧奨、精度管理や追跡調査が整備された組織型検診（Organized Screening）を行うことが理想的である。北欧や英国では、乳がん検診や、子宮頸がん検診の組織型検診（Organized Screening）が行われている[6]。

任意型検診は、個人の死亡リスクの減少を目的とし、医療機関や検診機関が任意で提供するがん検診を意味する。検診機関や医療機関などで行われている総合健診や人間ドックなどに含まれているがん検診が該当する。

有効性評価

わが国におけるがん検診の有効性評価は、1998年(平成10)3月の厚生省老人保健推進費補助金・老人保健福祉に関する調査研究等事業「がん検診の有効性評価に関する研究班」報告書（主任研究者　久道茂）[7]をはじめとし、過去３回にわたる評価が行われた。

2003年度(平成15)から、厚生労働省がん研究助成金「がん検診の適切な方法とその評価法の確立に関する研究」班（主任研究者　祖父江友孝）では、これらの成果を踏まえ、わが国独自のがん検診ガイドラインの作成手順を定式化した[8]。

有効性評価に基づくガイドライン

がん検診の有効性評価に基づくガイドラインは以下の経緯を経て、作成されている。科学的根拠となる文献を抽出し、系統的総括を行い、死亡率減少効果についての証拠のレベル（**表３**参照）を判定した。

不利益は、受診者の負担や偶発症について、検査方法間の対比を行い、さらに両者の評価から、推奨のレベル（**表４**参照）を決定する。系統的総括の結果に基づき、各検診方法の死亡率減少効果と不利益に関する科学的根拠を明確にし、わが国における対策型検診と任意型検診の実施について、推奨として総括する。

定式化された作成手順に基づき、胃がん検診、大腸がん検診、肺がん検診、前立腺がん検診の推奨は**表５**として評価された[8~11]。

表３　証拠のレベル

証拠のレベル	主たる研究方法	内　　容
1++	無作為化比較対照試験	死亡率減少効果について一致性を認める、質の高い無作為化比較対照試験が複数行われている
	系統的総括	死亡率減少効果の有無を示す、質の高いメタ・アナリシス等の系統的総括が行われている
1+	無作為化比較対照試験	死亡率減少効果について一致性を認める、中等度の質の無作為化比較対照試験が複数行われている
	系統的総括	死亡率減少効果の有無を示す、中等度の質のメタ・アナリシス等の系統的総括が行われている
	ＡＦ組み合わせ	Analytic Framework の重要な段階において無作為化比較対照試験が行われており、2++ 以上の症例対照研究・コホート研究が行われ、死亡率減少効果が示唆される
1-	無作為化比較対照試験	死亡率減少効果に関する質の低い無作為化比較対照試験が行われている
	系統的総括	死亡率減少効果の有無を示す、質の低いメタ・アナリシス等の系統的総括が行われている
2++	症例対照研究 / コホート研究	死亡率減少効果について一致性を認める、質が高い症例対照研究・コホート研究が複数行われている
2+	症例対照研究 / コホート研究	死亡率減少効果について一致性を認める、中等度の質の症例対照研究・コホート研究が複数行われている
	地域相関研究 / 時系列研究	死亡率減少効果について一致性を認める、質の高い地域相関研究・時系列研究が複数行われている
	ＡＦ組み合わせ	死亡率減少効果の有無を示す直接的な証拠はないが、Analytic Framework の重要な段階において無作為化比較対照試験が行われており、一連の研究の組み合わせにより死亡率減少効果が示唆される
2-	症例対照研究 / コホート研究	死亡率減少効果の有無を示す、質の低い症例対照研究・コホート研究が行われている
	地域相関研究 / 時系列研究	死亡率減少効果について同様の結果を示す、中等度の質以下の地域相関研究・時系列研究が行われている
	ＡＦ組み合わせ	死亡率減少効果の有無を示す直接的な証拠はないが、Analytic Framework を構成する複数の研究がある
3	その他の研究	横断的な研究、発見率の報告、症例報告など、散発的な報告のみで Analytic Framework を構成する評価が不可能である
4	専門家の意見	専門家の意見

ＡＦ：Analytic Framework

注 1）研究の質については、以下のように定義する

質の高い研究：バイアスや交絡因子の制御が十分配慮されている研究。

中等度の質の研究：バイアスや交絡因子の制御が相応に配慮されている。

質の低い研究：バイアスや交絡因子の制御が不十分である研究。

表４　推奨のレベル

推奨	対策型検診（住民検診型）	任意型検診（人間ドック型）	証拠のレベル	証拠の内容
A	推奨する	推奨する	1++/1+	死亡率減少効果を示す十分な証拠（RCT）がある
B	推奨する	推奨する	2++/2+	死亡率減少効果を示す相応な証拠（観察研究）がある
C	推奨しない	条件付きで実施できる	1++/1+ 2++/2+	死亡率減少効果を示す証拠（RCT あるいは観察研究）があるが、無視できない不利益がある
D	推奨しない	推奨しない	1++/1+ 2++/2+	死亡率減少効果がないことを示す証拠（RCT あるいは観察研究）がある
I	推奨しない	個人の判断に基づく受診は妨げない	1-/2-	死亡率減少効果の有無を判断する証拠が不十分である（RCT・適切な観察研究がない、あるいは複数の研究結果が一致しない）

注）推奨 I と判定された検診の実施は、有効性評価を目的とした研究を行う場合に限定することが望ましい。

表５　有効性評価に基づくがん検診ガイドラインの推奨

がん検診	検診方法	推奨グレード	対策型検診	任意型検診
胃がん検診	胃Ｘ線	B	○	○
	胃内視鏡	I	×	△（条件付きで個人の判断）
	ペプシノゲン法	I	×	△（条件付きで個人の判断）
大腸がん検診	便潜血	A	○	○
	全大腸内視鏡	C	×	○（条件付実施可）
肺がん検診	非高危険群に対する胸部Ｘ線検査、及び高危険群に対する胸部Ｘ線検査と喀痰細胞診併用法	B	○	○
	胸部ＣＴ	I	×	△（条件付きで個人の判断）
前立腺がん検診	直腸診	I	×	△（条件付きで個人の判断）
	前立腺特異抗原（ＰＳＡ）	I	×	△（条件付きで個人の判断）

精度管理

がん検診の精度管理については、関連学会が技術的管理を中心に行い、ガイドラインなどを公表している。老人保健事業については、各都道府県では生活習慣病検診管理指導協議会がその任にあたっている[12]が、一部を除いて十分な機能を果たしていない。

一方、市区町村についても、対象者の把握と管理、記録の整備、発見がんの追跡調査などが求められている[10]。このため、「医療・介護関係事業者における個人情報の適切な取り扱いのためのガイドライン」では、公衆衛生上の目的として、医療機関の精密検査結果の情報提供は、本人の同意がなくても行える例外事項に含まれている[13]。

ＥＣでは、乳がん検診の精度管理のガイドラインを作成し、精度管理指標となる、がん発見率、要精検率などについて、一定の目標値を定めている[14]。わが国においても、厚生労働省がん検診検討会を経て、精度管理システムのチェックリストが公表された。

今後の課題

がん検診によってがん死亡を減少させるためには、有効性の確立した検診を正しく実施する必要がある。そのため、有効性評価を定期的に更新するための常設機関の必要性が、久道班第３版でも述べられている[7]。

精度管理を行うには、運用ガイドラインや目標値の設定など運営体制の整備が必要である。これまでに行われてきた技術ベースの医療者個人の評価や管理ではなく、プロセス管理・アウトカム管理に基づくシステムとしての管理が課題となる。

さらに、医療者のみならず、受診者に対する適切な情報発信が、がん検診への理解を深め、有効性の確立した適切な方法の選択に寄与すると考えられる。

（はましま・ちさと）

■参考文献■

1）厚生労働省大臣官房統計情報部編、平成4年～15年度地域保健・老人保健事業報告（老人保健編）、厚生統計協会、東京、1993年-2004年

2）Center for Disease Control and Prevention. Behavioral Risk Factor Survey. Atlanta, A : National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention ; 2002

3）NHS Health and Social Care Information Centre, Community Health Statistics. Breast Screening Program, England : 2004-2005, 2006

4）NHS Health and Social Care Information Centre, Community Health Statistics. Cervical Screening Program, England : 2004-2005, 2006

5）祖父江友孝、濱島ちさと、齋藤博他「有効性評価に基づくガイドライン作成手順（普及版）癌と化学療法」p32、p893- p900、2005年

6）Miles A, Cockburn J, Smith RA, Wardle J. A prospective from countries using organized screening programs. Cancer. 2004, 101(S5),1201-13

7）平成12年度厚生労働省老人保健事業推進費等補助金　がん検診の適正化に関する調査研究事業「新たながん検診手法の有効性評価報告書（主任研究者　久道茂）」公衆衛生協会、2001年

8）「平成17年度厚生労働省がん研究助成金　がん検診の適切な方法とその評価法の確立に関する研究班（主任研究者　祖父江友孝）有効性評価に基づく胃がん検診ガイドライン」、2006年

9）「平成16年度厚生労働省がん研究助成金　がん検診の適切な方法とその評価法の確立に関する研究班（主任研究者　祖父江友孝）有効性評価に基づく大腸がん検診ガイドライン」、2005年

10）「平成18年度厚生労働省がん研究助成金　がん検診の適切な方法とその評価法の確立に関する研究班（主任研究者　祖父江友孝）有効性評価に基づく肺がん検診ガイドライン」、2006年

11）「平成19年度厚生労働省がん研究助成金　がん検診の適切な方法とその評価法の確立に関する研究班（主任研究者　濱島ちさと）有効性評価に基づく前立腺がん検診ガイドライン」、2008年

12）厚生省老人保健福祉局老人保健課監修「老人保健法による健康診査マニュアル」日本医事新報社、東京、1998年

13）厚生労働省「医療・介護関係事業者における個人情報の適切な取り扱いのためのガイドライン」、2003年

14）The European Community Guideline for Quality Assurance in Mammography Screening. Luxemburg. Europe Against Cancer Programmed, Office for Official Publications of the European Communities, 2001

Cancer-Prevention and Early Detection

がん検診

Cancer screening

濱島ちさと
Chisato Hamashima
国立がんセンターがん予防・検診研究センター検診研究部検診評価研究室室長

はじめに

がん対策推進基本計画において，「すべての市町村において，精度管理・事業評価が実施されるとともに，科学的根拠に基づくがん検診が実施されることを目標とする」ことが謳われている。がん検診の目的は，早期発見・早期治療により当該がんによる死亡を減少させることであり，目的を確実に達成するには有効性の確立した検診を正しく実施する必要がある。このため，「有効性評価に基づくがん検診ガイドライン」の作成が進められ，精度管理のためのシステムが整備されつつある。

がん対策におけるがん検診

2006年に成立したがん対策基本法には，基本的施策として「がんの予防及び早期発見の推進」が織り込まれている。この法律に基づくがん対策推進基本計画でも，「がんによる死亡者の減少」が全体目標として掲げられていると同時に，「がんの早期発見」が個別目標の1つとなっている[1]。

がん対策推進基本計画では「がんによる死亡者の減少」を達成するため，今後10年で「がんの年齢調整死亡率(75歳未満)の20%減少」を目標としている。1995～2005年のがんによる年齢調整死亡率は毎年1%減少しているが，この傾向が今後10年間続くとさらに10%減少する[1]。また，たばこ対策，がん検診，がん診療の均てん化などのがん対策の推進により，さらに10%の上乗せが予測されている。なかでも，がん検診の寄与度がきわめて大きく，受診率が50%に増加した場合，がんによる死亡率が3.9%減少すると期待されている。

がん検診の現状

1982年の老人保健法施行以来，市区町村では老人保健事業によるがん検診が行われてきた。1982年度から胃がんおよび子宮頸がん検診が開始し，続いて肺がん，乳がん，大腸がん検診が行われている。1998年度からがん検診事業は一般財源化され，検診の実施，検査方法の選択などは市区町村の判断に委ねられている。平成17年度地域保健・老人保健事業報告によるがん検診の実績を表1に示した。

いずれの検診においても，受診率

表1　がん検診の実績

がん検診	胃がん	大腸がん	肺がん	乳がん	子宮頸がん
検査方法	胃X線	便潜血	胸部X線	視触診およびマンモグラフィ	細胞診
受診者数(人)	4344918	6630503	6963844	1604557	3439094
がん発見率(%)	0.15	0.17	0.04	0.27	0.06
要精検率(%)	10.8	7.2	2.8	8.9	1.2
精検受診率(%)	74.6	54.5	72.3	79.9	62.6

(平成17年度地域保健・老人保健事業報告)