Comparison of Observed and Expected Numbers of Detected Cancers in the Research Center for Cancer Prevention and Screening Program

Chisato Hamashima, Tomotaka Sobue, Yukio Muramatsu, Hiroshi Saito, Noriyuki Moriyama and Tadao Kakizoe

Research Center for Cancer Prevention and Screening, National Cancer Center, Tokyo, Japan

Received September 7, 2005; accepted February 12, 2006; published online May 30, 2006

Background: The Research Center for Cancer Prevention and Screening program is a one-arm prospective study designed to evaluate the effect of multiple modalities for cancer screening. Basic programs consist of screening tests for cancer of the lung, esophagus, stomach, colon, rectum, liver, gall bladder, pancreas and kidneys, in addition to prostate cancer screening for males and breast, cervical, endometrial and ovarian cancer screenings for females.

Objective: To investigate the possibility of overdiagnosis, we compared the observed numbers with expected numbers based on the model.

Methods: We calculated the expected number of cancers on the basis of negative or positive history of screening tests within the previous year, based on assumed sensitivity and sojourn time. Observed numbers of screen-detected cases for stomach, colorectal, lung, prostate and breast cancer were compared with expected numbers.

Results: From February 2004 to January 2005, 3786 participants were enrolled in our study. The overall cancer detection rate was 5.8% (119/2061) for males and 4.1% (71/1725) for females. No statistically significant difference was found between observed and expected cases for colorectal cancer screening, gastric cancer screening for females and lung cancer screening for males. Observed numbers of breast, prostate and lung cancer for females exceeded those expected ($P < 0.05$).

Conclusions: Although cancer screening programs in the present study increased the detection of potentially curable cancers, these modalities, particularly lung, breast and prostate screening, might detect cancers which would not necessarily be clinically significant. We should therefore weigh up benefit and harm for such cancer screening programs.

Key words: cancer screening – detection rate – sensitivity – sojourn time – overdiagnosis

INTRODUCTION

In an attempt to prevent premature death, the Health Service Law for the Aged introduced cancer screening programs in Japan for all residents over the age of 40 in 1983. Screening for gastric and cervical cancer was introduced initially, and colorectal, lung and breast cancer screening programs followed. At present, five cancer screening programs are conducted nationwide, and over 25 million people are screened annually (1). Although the research group for cancer screening in Japan recommended six cancer screening programs (2) in 2001, new modalities for cancer screening have been introduced in several local municipalities without evaluation by reliable studies. To reduce mortality from a specific cancer, effective, evidence-based screening should be conducted and appropriate management of quality assurance is required.

In 2004, the Japanese Government initiated the Third-Term Comprehensive 10-Year Strategy for Cancer Control, aimed at reducing the incidence and mortality of cancer in Japan. The Research Center for Cancer Prevention and Screening (RCCPS) was established at the campus of the National Cancer Center, Tokyo, in the same year. Although development of the new modalities is worthwhile, a systematic approach for the evaluation of cancer screening programs is required. In order to investigate the efficacy of cancer screening, programs using new modalities have been conducted. Variable cancers were detected in the past year, but might consist of overdiagnosis.

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

© 2006 Foundation for Promotion of Cancer Research
cases. To investigate its possibility, we compared the observed numbers with expected numbers based on the model.

SUBJECTS AND METHODS
CANCER SCREENING PROGRAMS

The RCCPS Cancer Screening Program is a one-arm prospective study designed to evaluate the effect of multiple cancer screening modalities. This is a hospital-based program and participants are enrolled on a voluntary basis. Age for the target group was 50 years and over for males and 40 years and over for females. Exclusion criteria were previous diagnosis of cancer and followed-up for pre-cancerous disease based on self-reporting. The research and screening methods were explained to all participants using written materials and face-to-face presentations by health-care professionals. In addition, participants signed informed consent documents approved by the National Cancer Center. All participants responded to a questionnaire concerning life style, smoking, alcohol intake, nutrition, past history of disease including cancer, family history and previous investigations within a year. These participants will be followed using a questionnaire survey after the baseline screening year. Follow-up studies include a hospital survey to investigate medical records of cancer patients detected by cancer screening and interval cancer rates based on the participant’s response. In addition, these participants are asked to attend repeat screening 5 years after the baseline.

Basic programs consisted of screenings for esophageal, gastric, colon, rectal, lung, hepatic, gall bladder, pancreatic and renal cancer. Cancer screening modalities were as follows: gastrofiberscopy (GFS) for the esophagus and stomach; total colonofiberscopy (TCF) or barium enema (BE) for the colon and rectum; computed tomography (CT) and sputum cytology for the lung; and abdominal ultrasonography (US) for the liver, gall bladder, pancreas and kidneys. The participants could choose TCF or BE based on their preferences. For males, prostate cancer screening was performed using an assay of prostate specific antigen (PSA) serum levels with a cut-off value of 2.7 ng/ml. For females, a combination of modalities was performed: two-view mammography (MMG), US and physical examination (PE) for the breasts, Pap smear for the cervix, and magnetic resonance imaging (MRI) for the endometrium and ovaries. Moreover, whole body scanning using positron emission tomography (PET) with injection of 2.78 MBq/kg fluorine-18-FDG was provided as an optional investigation. This study was approved by the Institutional Review Board of the National Cancer Center.

COMPARISON OF OBSERVED AND EXPECTED DETECTION NUMBERS

Numbers of subjects recruited into the program from February 2004 to January 2005 and observed numbers of detected cancers were classified by 5-year age group and by gender. In the questionnaire survey, we collected information on the following investigations performed within the previous year as follows: photofluorography, GFS, fecal occult blood test (FOBT), TCF, BE, chest radiography and MMG. We could not obtain information regarding previous investigation of CT for lung and PSA because these indicators were lack of the questionnaire.

Since screening detects cancer in a large prevalence pool, detection rate is influenced by previous investigations. Sojourn time (ST) is the duration of the detectable, preclinical phase of cancer (Fig. 1). The ST depends both on the natural history of the cancer and performance of screening modalities. Maximum lead time would therefore be achieved if screening was performed at the beginning of the ST. Although ST and sensitivity (SE) vary with age on individual cases, we used estimated mean values obtained from literatures. For simplicity of the present study, we assumed the following conditions: (i) ST and SE were constant in all age groups and (ii) SE was constant throughout ST.

We calculated the expected numbers of gastric, colorectal, lung, prostate and breast cancers in patients. The subjects are divided into three groups based on the previous history as follows: (i) subjects with no history of screening, (ii) subjects with history by the same test and (iii) subjects with history by the different test. In the first group, given that $I$ represents underlying incidence and $P$ target population numbers, expected numbers ($E$) at prevalence screening, which corresponds screening without previous investigation, can be derived from the following formula: $E = I \times (P/100 \,000) \times ST \cdot SE$. PSA screening is applicable to this case because previous history cannot be obtained from the questionnaire. In the second group, the expected numbers ($Ex$) is the sum of incidence and false-negative cases of previous investigation (Fig. 2). The sensitivity of modality1 assumed SE1 and ST1 for its sojourn time. Ex is calculated as follows: $Ex = I \times (P/100 \,000) \times (ST1-(ST1-1) \times SE1) \times SE1$. The modality2 was previous investigation, which is different from the modality of RCCPS screening program. Similarly, the sensitivity of modality2 assumed SE2 and ST2 for its sojourn time. These cases are the participants who have a screening history using other modalities in colorectal, gastric and lung cancer screening. When participants had history of previous investigation
using modality 2, the expected number (Ey) including false-negative cases of previous screening is as follows: 
\[ Ey = I \times \left( P/100,000 \right) \times (ST1 - (ST2 \times SE2)) \times SE1 \] (Fig. 3).

The incidences of gastric, colorectal, lung, prostate and breast cancer were obtained from estimations calculated by cancer registries (4), while the ST and SE of breast cancer screening were assumed based on published reports (3,5–9). The ST or lead time of prostate cancer screening was determined from published articles and it ranged from 5 to 15 years (10–17). In other modalities, SE has been reported without adjustment for ST (18–20). In the baseline analysis, SE was assumed as follows: 70% for GFS; 70% for BE; 70% for TCF; 80% for CT; 80% for the combination of MMG, US and PE; 70% for MMG; 70% for PSA; 50% for chest radiography; 50% for FOBT; and 60% for photofluorography. ST was assumed as follows: 5 years for GFS; 5 years for BE; 10 years for TCF; 5 years for CT; 5 years for a combination of MMG, US and PE; 4 years for MMG only; and 10 years for PSA screening. In colorectal cancer screening, ST of immunological FOBT was assumed to be 2 years [published reports which reported the range from 2 to 4.70 years using various estimation models (21–23)]. The ST of chest radiography is 1 year based on previous reports (24,25). No references to ST of photofluorography could be found; this was assumed to be 3 years in the present study. We estimated E of detected cancers and compared these with observed numbers (O) to calculate the ratio O/E. The observed and expected numbers of detected cancer were compared using the chi-squared test. A sensitivity analysis was used to assess the effect of varying individual model parameters during the construction and testing of the models; this was performed to assess the effects of changes in our assumptions regarding ST and SE. We conducted a sensitivity analysis in the cases in which difference of the ratio O/E was significant.

**RESULTS**

Table 1 presents the distribution of all participants by 5-year age group and by gender. From establishment of the study in February 2003 to January 2004, 3786 participants were enrolled: 2061 males and 1725 females. In both genders, most participants (over 25%) were in the 60- to 64-year age
groups. Of participants over 70 years of age, 5.5% (114/2061) were males and 3.9% (67/1725) were females. Almost 90% of participants came from the Tokyo metropolitan area and the seven surrounding prefectures. Regarding colorectal cancer screening, TCF was performed in 83.6% (1723/2061) of male participants and 77.8% (1342/1725) of female participants, and the remaining 15.4% (317/2061) of male and 19.9% (343/1725) of female participants had BE. PET scans were performed for 79.0% (1629/2061) of males and 74.3% (1282/1725) of females. In the first year of the RCCPS programs, 190 cancers were detected (Table 2). The detection rate for all cancers was 5.8% (119/2061) for males and 4.1% (317/1712) for females. In males, expected numbers of cancers were as follows: gastric cancer, 15.3 cases; colorectal cancer, 2.3 cases for BE and 21.9 cases for TCF; lung cancer, 10.9 cases; and prostate cancer, 7.0 cases. In females, expected numbers were as follows: gastric cancer, 3.7 cases; colorectal cancer, 1.1 cases for BE and 7.6 cases for TCF; lung cancer, 2.4 cases; and breast cancer, 6.2 cases. For TCF screening, observed numbers were almost equal. The observed numbers for gastric cancer were almost two times than expected numbers. But, in females, it was not significantly different. On the other hand, lung cancer was observed seven times more often in females but nearly equal in males. Prostate cancer and breast cancer were both detected over two times more frequently than expected. On the sensitivity analysis of prostate, breast and lung cancer screening for females, expected numbers of prostate and lung cancer increased in accordance with ST and SE. For prostate cancer screening, O/E ratio ranged between 5.36 and 16.07 according to SE values from 30 to 90% when ST was set at 5 years;
observed numbers of prostate cancer always exceeded expected numbers at any cases if ST was changed from 5 to 15 years. For lung cancer screening for females, O/E ratio ranged between 6.72 and 12.10 according to SE values from 50 to 90% when ST was set at 5 years; observed numbers of breast cancer were three times more than expected at any cases if ST was changed from 5 to 10 years.

**DISCUSSION**

The efficacy of reducing mortality rates from cancer has been established for several cancer screening programs. Based on these studies, the research group for cancer screening in Japan recommended the following six cancer screening programs (2): photofluorography for gastric cancer, fecal occult blood
test for colorectal cancer, chest radiography and sputum cytology 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. Recently, the guideline for colorectal cancer screening has been revised, and chemical and immunological fecal occult blood tests have been recommended as population-based screening (20). Both TCF and BE could be introduced in opportunistic screening as long as well-controlled risk management is performed. Although these guidelines follow evidence-based cancer screening programs, new modalities which show no evidence of mortality reduction have rapidly been disseminated. These new modalities, such PET, CT and GFS, possess high sensitivity and are therefore anticipated to detect early cancer; however, while they are useful for cancer detection, their effectiveness in cancer screening is unclear.

The detection rates in our study were higher than those of population-based screening (20). There are two possibilities for this difference. First, for over 70% of participants, it was the first experience that they were examined by GFS, TCF, BE, CT and MMG. When screening is initiated, an apparent excess of diagnosed cancers is inevitable, because in the first round of screening a large number of cancers that would have occurred in future are diagnosed earlier. Second, the sensitivity of the modalities in our study was superior to those of population-based screening (18–20). Population-based screening programs have been conducted using chest radiography and sputum cytology for individuals at high risk of lung cancer, while similar programs using photofluorography for gastric cancer and immunological fecal occult blood testing for colorectal cancer have also been performed. Considered these conditions, we calculated the expected numbers of detected cancers in our cohort based on assumptions of sensitivity and sojourn time in several modalities. The difference of observed and expected numbers could be changed according to use of the data. We conducted a sensitivity analysis to investigate the robustness because it was possible that our conclusion would be changed according to the data used for the analysis. For example, we used incidence rates obtained from population-based cancer registries. The incidence rate from cancer registries is the weighted average of incidence among the population with and without previous history of screening. These assumptions might introduce under- or overestimation.

In the cases of prostate, breast and gastric cancer for males and lung cancer for females, the observed numbers exceeded expectation and were similar to those expected in the other cases. High detection rate is a consequence of the screening itself, i.e. overdiagnosis, especially in prostate and lung cancer for females. Overdiagnosis has been pointed out and was a major harm in both screening programs (26). Although the test was conducted using the same modality for lung cancer screening, the results were different between males and females in our study. The difference of two groups might be explained by the difference of the history of chest radiography. Strauss et al. (27) state that the overdiagnosis hypothesis is counter to virtually all known data on the natural history and biological behavior of lung cancer. In recent screening studies, both detection rate and stage I cancer by CT exceeded that of chest radiography (28,29). For the very reason, overdiagnosis could be a more serious problem for CT screening. On the other hand, the cut-off point for prostate cancer screening is controversial. PSA value of 4.0 ng/ml is a popular cut-off point for prostate cancer screening; 2.7 ng/ml was used in the present study. However, only two cases (8.3%) of the detected prostate cancers exhibited PSA levels below 4.0 ng/ml. In the European Randomized Study of Screening for Prostate Cancer, the cut-off PSA level was changed from 4.0 to 3.0 ng/ml (30). Krumholtz and colleagues (31) found a prostate cancer incidence rate of 22% in patients with 2.6–4.0 ng/ml PSA based on biopsies of 94 patients with clinical stage T1c. Recently, the prevalence of prostate cancer was reported to be 14.9% for those with PSA values below 4.0 ng/ml (32). Of these tumors, 15% contained Gleason pattern 4, indicating that high-grade cancer occasionally occurs in the presence of low PSA. Disagreement exists as to the best cut-off value for PSA. Greater detection of prostate cancer increases the risk of overdiagnosis and overtreatment, which can cause erectile dysfunction and urinary incontinence. The risk of overdiagnosis has been reported as more than 48% within a screening population with a 4-year screening interval (13). Etzioni and colleagues calculated the overdiagnosis rates of prostate cancer screening as 29% for whites and 44% for blacks, based on SEER-Medicare database (14). Men with low-grade prostate cancer (Gleason score of 2–4) have minimal risk of dying from prostate cancer during 20 years of follow-up compared with men with high-grade prostate cancer (Gleason score of 8–10) (33). On the other hand, Bill-Axelson et al. (34) reported that radical prostatectomy reduces disease-specific mortality and overall mortality compared with watchful waiting. Including selection of therapy, the efficacy of prostate

### Table 3. Proportion of having previous investigations within a year by screening modalities

<table>
<thead>
<tr>
<th>Examination</th>
<th>Modality</th>
<th>Male (%)</th>
<th>Female (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stomach</td>
<td>XP</td>
<td>43.5</td>
<td>30.0</td>
</tr>
<tr>
<td></td>
<td>GFS</td>
<td>28.7</td>
<td>23.3</td>
</tr>
<tr>
<td>Colon and rectum</td>
<td>FOBT</td>
<td>52.7</td>
<td>40.7</td>
</tr>
<tr>
<td></td>
<td>BE</td>
<td>4.9</td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td>TCF</td>
<td>15.4</td>
<td>8.3</td>
</tr>
<tr>
<td>Lung</td>
<td>Chest X-ray</td>
<td>73.9</td>
<td>62.0</td>
</tr>
<tr>
<td></td>
<td>MMG</td>
<td></td>
<td>18.5</td>
</tr>
</tbody>
</table>

The percentage of previous examination compared males and females using the chi-squared test.

XP: gastrophotofluorography; FOBT: fecal occult blood test; GFS: gastrofibroscopy; BE: barium enema; TCF: total colonoscopy.

PSA: prostate specific antigen; MMG: mammography; US: ultrasonography; PE: physical examination.
cancer screening programs is still unclear. Although the cancer screening programs in the present study increased the detection of potentially curable cancers, these modalities might detect tumors that would not be clinically significant. We should accordingly weigh up the benefits and harms of cancer screening using these modalities, and such information should be given to the participants of our study.

The present study is the first report from the RCCPS and has several limitations. First, our cohort of around 4000 volunteers is insufficient to observe reduction of mortality rates from specific cancer and no comparable group was included. Second, participants were volunteers who were receptive to screening by the new modalities. Hence, a self-selection bias could not be excluded. In the present study, we estimated expected numbers using a simple model based on approximate assumptions. However, to estimate correct sojourn time accurately and to modify our model accordingly, lengthy follow-up is needed. We have started follow-up studies, which include an annual questionnaire survey of participants and a hospital survey to acquire information on cancer patients. Information concerning interval cancer can be obtained through this survey, and sensitivity and sojourn time of several cancers can be reinvestigated based on the new model. In addition, we aim to investigate all participants using the same modalities after 5 years and are planning further programs to evaluate the accuracy of the screening modalities.

Acknowledgments
This study was supported by a Grant-in-Aid for the Third Term Comprehensive 10-Year Strategy for Cancer Control (H16-cancer control-016) from Ministry of Health, Labour and Welfare of Japan. The authors thank Ms Kanoko Matsushima and Ms Junko Asai for the manuscript preparations.

References


