

X-ray screening seems to reduce gastric cancer mortality by half in a community-controlled trial in Costa Rica

L Rosero-Bixby^{*,1,2} and R Sierra²

¹Centro Centroamericano de Población, Universidad de Costa Rica, San José 2060, Costa Rica; ²Instituto de Investigaciones en Salud (INISA), Universidad de Costa Rica, San José 2060, Costa Rica

X-ray screening of gastric cancer is broadly used in Japan, although no controlled trial has proved its effectiveness. This study evaluates the impact of an X-ray screening demonstrative intervention to reduce gastric cancer mortality in a Costa Rican region. The evaluation follows a quasi-experimental, community-controlled design, with measures before and after. About 7000 individuals participated by invitation in the two-wave screening programme. X-ray screening was followed by videoendoscopy and gastric biopsies. Treatment included resection with or without lymph node dissection. Comparisons with two control groups estimate that gastric cancer mortality was halved in the period from 2 to 7 years after the first screening visit. Validity of X-rays as used in this intervention had 88% sensitivity, 80% specificity, and 3% predictive value for individuals with two screening visits. Incidence in the screened group increased up to four times. Case survival was 85% in the intervention group after 5 years, compared to 12% among the controls before the intervention and 35% among the controls in the same region after the intervention. Although X-ray mass screening seems able to reduce stomach cancer mortality, its high cost may be an obstacle for scaling up this intervention in a non-rich country like Costa Rica.

British Journal of Cancer (2007) **97**, 837–843. doi:10.1038/sj.bjc.6603729 www.bjcancer.com

Published online 2 October 2007

© 2007 Cancer Research UK

Keywords: gastric cancer; gastric cancer screening; impact evaluation; Costa Rica

Gastric cancer is the second highest cause of cancer death worldwide (Parkin *et al*, 2001). The therapy of this cancer is early extirpation, whose outcome is closely related to the stage of disease progression. Japan, the country with the highest incidence in the world, has been broadly using X-ray screening to pick early gastric tumours since the 1960s (Inaba *et al*, 1999; Tsubono and Hisamichi, 2000). Some pilot programmes elsewhere sporadically used this Japanese technique as well (Llorens, 1991; Pisani *et al*, 1994). The effectiveness of this intervention is, however, controversial as it was in the past for lung cancer screening (which has proved to be ineffective) and currently as it is for prostate cancer (Collins and Barry, 1996; Parkin and Pisani, 1999; Genta, 2004). No controlled trial has been conducted, in our knowledge, to determine the effectiveness of the Japanese model of gastric cancer screening. Japan has made considerable investments in X-ray mass-screening programme, achieving important reductions in mortality (Oshima *et al*, 1986). Some critics, however, argue that these reductions may be due to other reasons and that the only effect of the screening programme has been to inflate the incidence rates with cases that never would have evolved into advanced cancer and death (Everett and Axon, 1998).

The impact of cancer screening interventions is difficult to evaluate (Adami *et al*, 2001). Biases occur when screened

individuals are volunteers and probably self selected in terms of lifestyles, including diet and preexistence of gastric problems (Pisani *et al*, 1994). Over-diagnosis or pseudo-disease bias is a problem too, that is, a disease discovered in screening campaigns, but which would not have affected the patient had it not been detected (Everett and Axon, 1998). When death is not the evaluation end point, multiple biases occur as well and the evidence is 'inadmissible' (Sackett *et al*, 1991; Alibhai, 2006).

Costa Rica has the second highest mortality rate and the fifth highest incidence rate of stomach cancer in the world (Ferlay *et al*, 2001). Within Costa Rica, there are large differences, with the highest rates of gastric cancer and of precancerous lesions occurring in the country's highlands (Salas, 1975; Sierra *et al*, 1989, 1995). Case fatality of stomach cancer in Costa Rica is very high with a mortality/incidence ratio of 85%, which is more than twice higher than the corresponding ratio in Japan (Ferlay *et al*, 2001). Case fatality is high because most diagnoses occur at advanced stages of the disease: only 3–7% of cases are diagnosed in early stages and survival after 5 year is less than 15% (Sasagawa *et al*, 1999).

A demonstrative gastric cancer, mass-screening programme in a high-risk population, with the Japanese method as model, was carried out in Costa Rica from 1996 to 2000 (Sasagawa *et al*, 2002). This study evaluates the impact of this intervention, which was designed as a community controlled, non-randomised, screening programme. It presents results regarding the validity of the screening programme. Then, following a cohort, quasi-experimental design with measurements before and after, it assesses the effect upon incidence, case survivorship, and mortality.

*Correspondence: Dr L Rosero-Bixby;

E-mail: lrosero@ccp.ucr.ac.cr

Received 31 October 2006; revised 12 March 2007; accepted 12 March 2007; published online 2 October 2007

PARTICIPANTS AND METHODS

The screening programme

The Costa Rican government, with assistance from the Japan International Cooperation Agency (JICA), carried out a pilot stomach cancer mass-screening intervention from 1996 to 2000. It covered the region of *Cartago* and *Los Santos*, which is well known for its high incidence of this disease (Sierra *et al*, 1995). The *Max Peralta* Hospital, which is the regional, public hospital that covers this region, hosted it. The 'Cartago project' followed procedures developed in Japan for screening and treatment of gastric cancer.

The project conducted X-ray mass screening using upper gastrointestinal tract high-resolution, double-contrast studies. Trained X-ray technicians did the studies, using effervescent granules and barium meal of high density. Two radiologists examined the films. Participants with a suspicious abnormality (cancer or, in most cases, other stomach lesions such as ulcers) were sent for additional examination using high-resolution upper videoendoscopy and, if necessary, gastric biopsies. Those with gastric cancer were treated and followed up in the *Max Peralta* hospital. Treatment included resection with or without lymph node dissection. All studies and surgeries were performed with the participation of Japanese staff using methods that are identical to those used in Japan for detection, diagnosis, treatment, and stage classification of gastric cancer (Sasagawa *et al*, 1999). The screened individuals were invited to repeat the protocol after 2 years. Regarding the radiation dose, the project's equipment was authorised by the corresponding authority as meeting the norms. There was, however, no requirement of measuring or keeping records of the radiation dose used on each individual.

Participants

Participants were not self-selected volunteers but those invited from the general population. The project sent invitation letters to the residents in the intervention region born between 1921 and 1945. In addition, eligible participants must have had no previous diagnosis of gastric cancer. The intervention region was covered taking one 'health sector' at a time. Health sectors are small areas of less than 1000 households served with primary health care by a team of health professionals named EBAIS. Invited individuals were drawn from a census conducted earlier by the project or by the EBAIS. To increase statistical power, only half of the eligible women were randomly invited. From 15 to 20 participants were invited daily. The project provided transportation to and from the hospital. Participants gave informed consent in writing. The Ethics Committee of the University of Costa Rica approved the study.

The first wave of mass screening took place from March 1996 to March 1999, and the second wave, from March 1998 to June 2000. Screening activities were suspended for 9 months in 1998–1999 because of construction works at the hospital.

The original study design was aimed at 12 000 participants to have statistical power to detect 50% reduction in mortality after 4 years. The actual study, however, had only 7000 participants because of the 9-month suspension, and productivity was lower than the 30 daily screenings originally planned according to Japanese parameters. We compensated this reduction by observing mortality for a longer period: 7 years instead of 4. Also because of this reduction in sample size, the intervention was not able to cover the entire *Cartago - Los Santos* region.

Evaluation design and the control groups

The evaluation follows a quasi-experimental design (Mohr, 1988) with an intervention group (the cases) and two control groups, as well as *before* and *after* measurements. It is thus a nonrandomised community-controlled trial. The first control group includes the cohorts born in 1921–1945 and residing in the region of *Pérez Zeledón* (PZ), a region that shares with *Cartago* the highest incidence of stomach cancer in Costa Rica. PZ is located more than 100 kilometres South of *Cartago*; mountains separate the two regions. The second control arm came from the same intervention area of *Cartago* and *Los Santos*. This arm includes all intervention cases until the date of their first screening visit (which on average took place in May 1997) as well as residents in the originally selected region who were not invited to the mass screening; that is about 3000 women randomly left out and about 8000 eligible residents in sub-areas that the project was not able to cover. We did not include in this control arm approximately 2000 non-respondents to the invitation.

For the impact evaluation exercise, we imposed two additional eligibility conditions: (1) being a resident in the study regions in 1990, and (2) being a Costa Rican citizen. The study individuals should have been included in the 1990 voting registry of the study regions and all must have a unique identification number (the *cedula*), which would allow us to follow them in computer files of the death registry and the national tumour registry.

There are thus four control arms, as described in Table 1: (1) *Cartago-before*, with about 20 000 individuals; (2) *Cartago-after*, 11 000; (3) *PZ-before*, 12 000; and (4) *PZ-after*, 11 000. Each individual contributes 6 or 7 years of observation to each arm. The number of screening cases in the evaluation (6200) is smaller than the total number of participants due to the additional eligibility criteria of being a Costa Rican citizen and a resident in the region by 1990. The *after* control groups, as well as the intervention group, are about 6 years older than the *before* groups, reflecting the time elapsed. The intervention group has more men (64%) and the *Cartago-after* has fewer men (42%) by design, as 50% of women were randomly excluded from the intervention. These age and sex differences must be controlled for in the analyses. Socioeconomic status, as measured by the percentages

Table 1 The case and control groups for the evaluation

Indicator	Cases	Control 1	Control 2	Control 3	Control 4
Region	<i>Cartago</i>	<i>Cartago</i>	<i>Cartago</i>	PZ	PZ
Timing	After	Before	After	Before	After
Entry date	X-ray visit 1	January 1991	May 1997	January 1991	May 1997
Exit date	December 2004	Visit 1 or May 1997 ^a	December 2004	May 1997	December 2004
N participants	6206	20 030	11 190	11 915	11 318
N person-years	45 231	123 121	80 630	73 666	82 408
Males	64%	51%	42%	51%	51%
Mean age	64.3	57.9	64.3	58	64.6
High school education	7.8%	—	8.6%	—	8.1%
Have refrigerator	86%	—	90%	—	83%

PZ = *Pérez Zeledón*. ^aObservation in this group ends (it is censored) at the date of the first screening visit if the individual became a 'case' or in May 1997 if he did not.

with high school education and possessing a refrigerator, differs little among groups (Table 1).

For evaluation purposes, we further distinguish in the intervention group the 2 years following the first screening visit from the period afterwards. Cancer gastric mortality in this initial period is probably biased by the eligibility requirement of being free of the disease. (Because the intervention group cannot include cases of individuals with gastric cancer diagnosis, very few deaths, if any, due to this disease will occur in the months following the screening.) In turn, early detection of cancer lesions will help avoid deaths only after several months or years have passed since the screening date. In an extreme situation, if the screening detects very early lesions, which would have progressed to metastatic disease in say, 5 years, we should wait at least 5 years to evaluate the impact. In this study, the waiting period is 2 years.

The data

We used four major sources of information.

1. The *Cartago* project databases, which provided the dates of the screening procedures, waves 1 and 2, and the diagnoses of these procedures.
2. The 1990 Voting Registry (*padrón electoral*) to identify the study individuals, cases, or controls, with their identification number. The *Tribunal Supremo de Elecciones (TSE)* provided the database.
3. The 1990–2004 Death Registry to determine the dates and causes of death of deceased individuals. The TSE also provided this database.
4. The 1990–2004 National Tumor Registry to find out all the stomach cancer cases and their diagnoses dates. The Ministry of Health provided the computer data file.

Validity assessment

All participants in the intervention were included in a validity assessment of sensitivity, specificity, and positive predictive value. With data from the tumour registry, we consider incident cases of all those diagnosed within 18 months from the X-ray screening visit. We consider ‘screening positives’ to individuals who went to endoscopy, regardless of whether or not they were ‘cancer suspect.’ In a second, hypothetical estimate, we consider positives only to those rated ‘cancer suspect’ in the X-ray exam. Note that many cancers were detected in participants sent to endoscopy because of no cancer gastric pathologies.

Survival

We compare Kaplan–Meier survival curves of stomach cancer cases to assess the effect of the intervention upon survivorship. These curves show the proportion of individuals alive over time starting at the time the cancer was diagnosed and with correction for ‘censoring’ (Hosmer and Lemeshow, 1999).

Incidence and mortality

We compared incidence or mortality rates in the intervention group with the standardised rates in the four control groups. The age (5-year brackets) and sex structure of the intervention group was the standard. We used ‘robust’ estimates of the 95% confidence intervals for these rates, which takes into account that several observation years come from the same individual (Statacorp, 2005).

The final estimate of impact comes from a proportional hazard model, also known as Cox regression, on the gastric cancer death hazard function (Hosmer and Lemeshow, 1999). We used the Stata software to estimate the model, with birth date as the origin and

January 1991 as the entry date of observation. Data were censored at the following three events: (1) death by causes other than stomach cancer, (2) first screening visit or May 1997 for the ‘before’ groups, and (3) December 2004 for the ‘after’ groups. In addition, for the first two years of the intervention, the censoring date was the second anniversary of the first screening visit. The death rate ratios estimated with the model measure the impact of the intervention. Since no second-order interaction terms were found significant in preliminary estimates of the regression model, we excluded them.

RESULTS

About 7000 participants went through the first X-ray screening procedure (Table 2). Five thousand returned for a second wave of screening 2 years later. Non-response to the first invitation and attrition in the second were reasonably low: 22 and 25%, respectively.

The X-ray series detected gastric pathologies meriting an endoscopy in 34% of participants in the first wave and in 20% in the second wave (Table 2). Most lesions were gastritis, peptic ulcers, and hernias. Our broad operational definition of positive test is the sending of an individual to endoscopy. Most of these ‘positive X-rays’, however, had no diagnosis of cancer. After endoscopy and histopathology diagnoses, the project detected 59 gastric cancers in the first wave and 28 in the second wave. Ten cancers went undetected by the project: six in the first wave and four in the second wave. These undetected cases were diagnosed outside the project within 18 months of the X-ray examination. It is worth noting, however, that all but one of these ‘false negatives’ went through the endoscopy procedure, and four cases had a biopsy. One case had a histopathology diagnosis of ‘suspicious cancer’, but did not return to surgery for a final diagnosis.

After surgery, and following Japanese classification criteria, 62% of cancers were at early stage, 55% in the first wave and 80% in the second wave (Table 2).

Table 2 Main results and validity of the screening

Item	Wave 1	Wave 2
<i>N</i> invited	8705	6769
<i>N</i> screened	6828	5046
% non-response/attrition	22%	25%
<i>Screening results</i>		
% sent to endoscopy – total	34%	20%
Because cancer suspect	3%	2%
With other pathologies	31%	18%
<i>Stomach cancer cases</i>		
<i>N</i> detected	59	28
<i>N</i> undetected	6	4
Cancer detected at early stage	55%	80%
<i>X-ray validity indicators^a</i>		
Sensitivity	91%	88%
Specificity	67%	80%
Positive predictive value	3%	3%
<i>Hypothetical restricted^b screening</i>		
Sensitivity	58%	47%
Specificity	98%	98%
Positive predictive value	19%	16%

^aAll sent to endoscopy are taken as positives. ^bIf only cancer suspicious participants were sent to endoscopy.

Validity

Taking the screening programme in its entirety, sensitivity was 91% in the first round and 88% in the second wave. As noted, most false negatives occurred in the endoscopy procedure or in the histopathology diagnosis. Specificity of the X-ray procedure was 67% in the first wave, and improved to 80% in the second wave. The positive predictive value of the intervention was just 3% (Table 2).

This moderate specificity comes from our operational definition of X-ray 'positives' as all who went to endoscopy for any reason. In actuality, only about 10% of them went under suspicion of cancer. If we define positives as those who had the diagnosis 'suspicious of cancer' in the X-ray examination (202 participants in the first round and 96 in the second), the specificity climbs to 98% (last panel of Table 2). The positive predictive value also improves to 19%. The problem, however, is that many of the cancers diagnosed by the programme were among participants sent to endoscopy because of other pathologies. In the hypothetical situation that the programme would have sent to endoscopy only 'suspicious of cancer' subjects, sensitivity would have dropped to a dismaying 58%.

Incidence

If a screening programme is having an impact, incidence must increase in the intervention group, at least in the short term. About 100 gastric cancer cases occurred in the screened group between the entry-time in the project and December 2002. The corresponding incidence rate of 56 per 10 000 person-years is clearly higher than the rates in the four control groups (Figure 1). Incidence in the screened group is about four times higher than the rate before the intervention in the two study areas. The higher incidence narrows when the comparison is with the *after* control groups. The confidence intervals in Figure 1 show that the higher incidence in the screened group is statistically significant, except when compared to the *Cartago after* group. In both *Cartago* and *PZ* control areas, incidence increased over time.

Case survival

If a screening programme is having an impact, it must show that case survival has improved. Figure 2 shows the Kaplan–Meier survival curve (adjusted for age and sex) for cases in the intervention and the four control groups (group size ranges between 70 and 129 cases). The survival curve of the screened group is remarkably higher than those of the control groups. Five-year survival in the intervention arm was 80% or 86% after adjustment for age and sex. Comparable figures for the control groups range from 12% in *Cartago before* to 35% in *Cartago after*. Further distinguishing early and late cancers in the intervention group (Japanese classification of stage of disease, Sasagawa *et al*, 1999), 5 year survival is comparatively high in both groups: 95% for early cases and 77% for late cancer stage at diagnosis (figures adjusted for age and sex). Both early- and late-stage survival curves of the intervention group (not shown in figures) are significantly higher than the four control groups. Survival in early cancers at detection is also significantly higher than that in cancers detected at late stages. Unfortunately, there is no comparable information on stage of disease for the control groups, although from earlier studies we know that the great majority of gastric cancers normally diagnosed in Costa Rica are at late or advanced stage.

As with incidence, case survival increased over time in the two control areas, especially in *Cartago*. The improvement in survival in *Cartago* (excluding the screened group) is statistically significant; the improvement in *PZ* is not so. There are no significant differences between *PZ* and *Cartago* in the curves of the *before* period.

Mortality

Twenty-six screened individuals died of gastric cancer during the evaluation period of 2–7 years post-screening, at a rate of 8 per 10 000 person-years (Figure 3). This rate is substantially lower than the rates in the four control groups. In relative terms, it ranges between 0.41 and 0.52 of the control rates. The confidence intervals show that these effects are statistically significant. Depending on which control arm is taken as comparison, the intervention would have avoided between 24 and 37 deaths over the 5-year evaluation period. Figure 3 also suggests a decline over time in gastric cancer

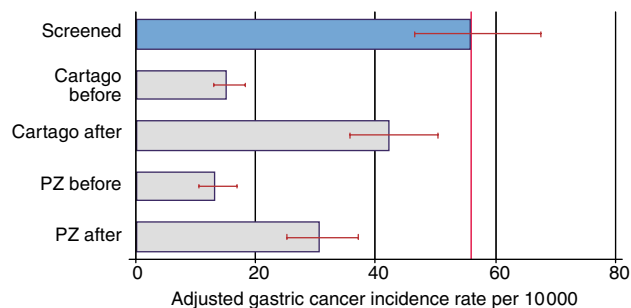


Figure 1 Gastric cancer incidence rates adjusted by age and sex to the screened population (95% confidence intervals shown).

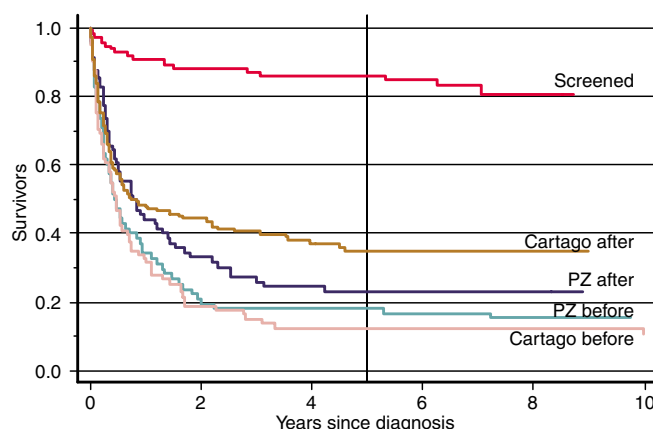


Figure 2 Kaplan–Meier survival curves of stomach cancer cases in the screened and control groups. Curves adjusted to age 60 and males.

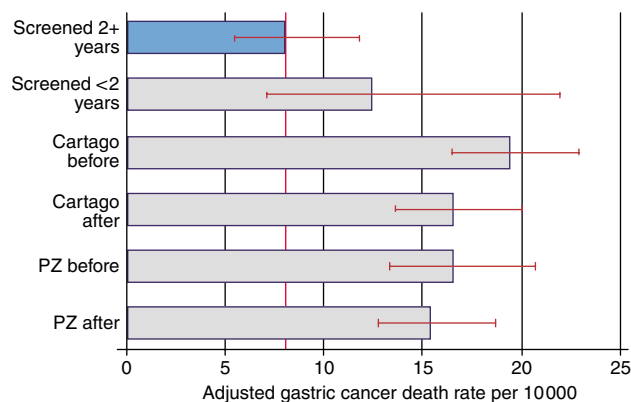


Figure 3 Gastric cancer death rates adjusted by age and sex to the screened population (95% confidence intervals shown).

mortality in both control areas, especially in *Cartago*, although this trend is not statistically significant.

The Cox regression results provide more precise estimates. The death rate ratios estimated with the regression model show that the risk of death by gastric cancer among controls is 2.1 and 1.7 times higher than in the cases *before* and *after* the intervention, respectively (Table 3). The model also shows that males die at rates more than twice those of women. It also suggests that mortality is a bit higher in *Cartago* than in *PZ* and that mortality is declining in more recent cohorts, although these effects are not statistically significant. The regression model controls the potentially confounding effects of cohort, sex, and region of residence. Cox-regression models also implicitly control the effect of ageing.

In the Cox model, as well as in Figure 3, mortality in the first 2 years' post-screening is lower than in the control groups and slightly higher than in the evaluation period of 2–7 years. These differences are not, however, statistically significant, which in part is due to the lack of statistical power (large confidence intervals).

DISCUSSION

An X-ray screening pilot intervention apparently halved gastric cancer mortality in a community controlled trial in Costa Rica. The two-wave, double-contrast X-ray examination also increased the incidence of gastric cancer diagnosis and substantially improved case survival. We have estimated these effects using a quasi-experimental, community evaluation design with *before* and *after* measures.

Two improvements in the design of this study set it apart from previous evaluations of X-ray screening of gastric cancer. First, the intervention group comes from the general population. They did not volunteer but were invited to participate as a sample of the general population, minimising self-selection biases. Second, the end-point evaluation is mortality, as mandatory for assessing the impact of screening interventions.

Incidence (and stage of disease)

The *Cartago* project improved detection of gastric cancer. There are two ways of reaching this conclusion. One is by looking at the stage of the detected cancers, which has been done in an early report of this intervention (Sasagawa *et al*, 1999). If most of them were in early stages, there is an indication that the programme is indeed highlighting cases that would have gone undetected without the programme. Sasagawa *et al* (1999) show that 64% of stomach cancer cases were diagnosed in early stages in the intervention group, a proportion similar to the results obtained in Japan, whereas in the rest of the population in the same region, served by the same medical team, and in the same hospital, only 30% of gastric cancer cases were detected in early stages. This proportion was only 3–7% before the intervention in *Cartago*. Unfortunately

Table 3 Cox regression estimates of mortality rate ratios in the screening programme

Variable	Rate ratio (95% CI)
Birth cohort (1 birth year)	0.97 (0.93 – 1.02)
Males (reference females)	2.13 (1.77 – 2.59)
<i>Cartago</i> (reference <i>PZ</i>)	1.12 (0.93 – 1.35)
Screening groups	
Unscreened before	2.06 (1.23 – 3.46)
Unscreened after	1.73 (1.18 – 2.53)
Screened <2 years	1.24 (0.63 – 2.45)
Screened 2–7 years	1.00 Reference

CI = confidence interval; *PZ* = Pérez Zeledón.

we do not have information on the stage of disease in the control groups for analysing this topic in detail.

Another way of approaching this issue is by comparing incidence in the screened group with incidence in the control groups. The higher incidence (up to four-fold) in the intervention group indicates that the project is catching additional cases. This estimate comes from comparisons with the control groups *before* the intervention. Comparison with controls *contemporary* to the intervention renders a smaller increment, although this result may be biased by 'contamination' among the controls, especially in the *Cartago* area, as suggested by the increased incidence from the before to after periods within each control arm. The intervention was accompanied by ample diffusion of information about the project, as well as about gastric cancer in general, between both health professionals and lay public. In addition, the *Cartago* project trained several technicians and physicians from other hospitals in gastric cancer screening procedures. This information diffusion could motivate individuals and physicians not in the project to increase the use of screening practices resulting in the aforementioned increased incidence with time.

Case survival

A screening programme that just detects more cases of early cancer would be useless if it is not complemented with proper treatment that extends life expectancy. The increased 5-year case survival to 85% in the intervention group may be the result of both early detection (see discussion below about the 'lead time bias') and improved treatment. Prognosis of early gastric cancer is very good worldwide. The Japanese treatment model with radical surgery seems to be successful in extending life expectancy (Everett and Axon, 1997). In this regard, the substantial increase of 5-year survival among the *Cartago controls* may be a positive side effect (a sort of contamination) from the intervention. The surgery technique brought by the Japanese surgeons to the project was used in the *Cartago* hospital in all gastric cancer patients and not confined to the project's cases.

An intriguing result is the high survival (77% after 5 years) of late stage cancer cases diagnosed in the intervention, which was much higher than in any control arm. A plausible explanation for this is that even these late stage cancers were less advanced compared to those normally diagnosed in Costa Rica and in the control arms.

Mortality

The high survival curve of screened cases, although encouraging, is not a proof of an impact. 'Lead time bias' is a well-known problem with survivorship of cases uncovered with screening: they live longer because they have been diagnosed earlier than unscreened cases (as noted before, 64% of cancers in the intervention group were in early stages). There could even be an 'overdiagnosis bias' if the screening procedure detects early pseudo-cancers that, left alone, never will evolve into advanced stages or they will evolve so slowly that the person will die of some other disease. This is the main reason why it is so important to evaluate mortality impact. About 50% lower mortality in the screened group compared to the control arms is a strong evidence that the screening intervention had an impact. This effect is a bit lower when we compare mortality with controls *after* the intervention. If lead time and overdiagnosed biases are operating, most of the reported results (higher incidence, higher survival, and earlier disease stage) would be present. The drop in mortality, however, cannot be explained for these common biases in screening interventions.

How robust is this impact estimate? We used the second best evaluation design: quasi-experimental (Bertrand *et al*, 1996). Although individuals were not randomly assigned to the intervention or control group (although women were randomly invited to

participate), participants were not allowed to volunteer. We are not able to identify any reason, other than the screening itself, for the intervention group to have 50% lower gastric cancer mortality. Four pieces of information suggest that we have not selected an intervention group that is biased with low gastric cancer mortality: (1) Data for the period *before* the intervention show that the intervention area of *Cartago* had a slightly higher mortality than the control area of PZ, in contrast with the lower mortality of the screened group compared to PZ. (2) Mortality did not increase in the *Cartago* control arm, as would have happened if we had selected a low-mortality subgroup for the intervention (Figure 3). (3) The groups are comparable in terms of age, socioeconomic status, and, most importantly, gastric cancer incidence and mortality. Moreover, the Cox regression model controlled for differences in gastric cancer mortality by age, sex, or region. (4) The high response rate of 78% makes self-selection or volunteer bias unlikely, although it is still an issue. In this regard, the data for non-respondents shows that their gastric cancer mortality is higher than among the screened individuals and slightly higher (non significantly) than the control group. Including about 2000 refusals in the 'after' control arm, the rate ratio of this group in the Cox regression is 1.85 compared with the 1.73 original estimate, that is results are not sensitive to the decision of excluding non-respondents from the control-after arm.

In spite of all this strong evidence, our conclusions cannot be definitive as the intervention was not a randomised trial. In addition, the length of observation for mortality (from 2 to 7 years) may not be long enough if the cases were at a very early stage. We will continue following the study arms several years in the future to assess the long-term effects.

Readers from developed countries may question that we used the national registries of deaths and tumours to track incidence and mortality. It is very difficult to prove that both registries have complete coverage in Costa Rica even though no study has proved the contrary. Not only is coverage complete but also the quality of information on cause of death and tumour location is valid given that 98% of deaths have a medical certificate and the tumour registry is histology based. The good quality of the Costa Rican statistical system goes along with an outstanding and universal healthcare system that has achieved the second highest national life expectancy in the continent (Canada has the highest), higher even than the United States (World Bank, 2006). Moreover, even if these registries had deficiencies, it is not clear how these could affect our results given that we are using the same sources of information for both the intervention and control groups. If errors in recording gastric cancer deaths were biased within our groups, this would be in the direction of omitting relatively less deaths in the intervention group given the fact that these individuals were under close scrutiny by the project.

Our decision to exclude from the evaluation the first 2 years after screening may raise objections. Reestimating the Cox regression model without this exclusion results in effects that are only slightly smaller than those originally reported: a mortality reduction of 48% with respect to the controls *before* and of 38% with respect to the controls *after*, compared to the 51 and 42% originally reported.

Validity of X-rays as screening test

It is obviously desirable to have a screening test that is both highly specific and highly sensitive. In practice, however, there is usually a trade-off between the sensitivity and specificity of a test. The Japanese screening model implemented in Costa Rica opted for high sensitivity (about 90%) at the expense of a somehow poor specificity (67% in first visit, 80% after two visits). This result was a conscious choice of the intervention physicians and came from the practice of sending to endoscopy not only those who appeared suspicious of cancer but many participants with gastric patholo-

gies (mostly ulcers, ulcer scars, hernias, and gastritis). This practice had adverse effects in terms of cost and the discomfort caused to many, but it also allowed more cancers to be detected and treated and helped improve the health of many individuals. These positive impacts of the screening intervention were not evaluated in this study.

Other adverse effects may come from the radiation of the X-ray examination, although the project used cutting edge technology equipment that warrants very low doses of radiation. The unnecessary stress in the many 'false-positive' individuals also can be accounted as an adverse effect, although participants were clearly informed that those not suspicious of cancer would also be sent to endoscopy. And there is the possibility of complications from the endoscopy and biopsy examinations, including bleeding, perforation, infection, and others. It is estimated that one in 1000 procedures has these type of complications. The project did not record systematically the occurrence of these adverse events. The only serious complication reported in the project was an oesophagus perforation in a participant with advanced cancer.

It must be noted, however, that our estimates of sensitivity and specificity are not based on a parallel gold standard for all screened participants, but on national tumour registry and death data, which are not perfect.

Policy implications

A health intervention with the potential for reducing gastric cancer mortality by 50% seems extraordinarily attractive in particular, considering that this is the second most lethal cancer worldwide, and the first one in some countries like Costa Rica. Some cautionary notes are however in place. First, the *Cartago* project was a *pilot* intervention, implemented under quite controlled circumstances and with an unusual wealth of resources, including manpower and equipment provided free by the government of Japan as well as unusually abundant academic and medical resources provided by Costa Rican institutions. Second, the 50% impact is restricted to just the screened population. A full scale, mass-screening programme will hardly be able to cover the totality of the population. Third, the cost of an X-ray mass-screening programme including treatment and other side activities is not cheap.

A cost estimate of the pilot programme resulted in about US\$1.000 per examination or, assuming 30 avoided deaths, \$300.000 per saved life (Schram, 2005). These figures are, of course, inflated by the high costs of Japanese personnel. It seems feasible to halve the costs by using only national manpower and improving productivity and screening technology. However, the law of decreasing yields will start to operate when the intervention is scaled up to cover populations with lower cancer incidence than the pilot group (i.e., lower ages, more women, and regions other than *Cartago*). In any scenario, the costs of the screening program seem far too high in a country like Costa Rica where yearly health expenditures are in the order of \$300 per capita (World Bank, 2006). A new promising, affordable screening test for gastric cancer is serum pepsinogens (Miki *et al*, 2003; Dinis-Ribeiro *et al*, 2004; Oishi *et al*, 2006). This new screening test must, however, be evaluated in studies like the one in this study.

ACKNOWLEDGEMENTS

The *Max Peralta* Hospital of the *Caja Costarricense de Seguro Social* (CCSS) carried out the mass-screening project. The Japan International Cooperation Agency (JICA) supported the project with equipment, manpower, and know-how. H Solano was the project director. T Sasagawa was the Japanese team leader. R Con-Wong and F Mirambell were key volunteers. F Mena, M Salazar, F Saenz, L Fonseca, and the project's staff were key for completing

the 5-year intervention. The Costa Rican Ministry of Health provided the databases of the National Cancer Registry. The *Tribunal Supremo de Elecciones* provided the databases of the

death and voting registries. Funding for this evaluation came from the *Universidad de Costa Rica*, as well as from a Wellcome Trust Grant No. 072406.

REFERENCES

- Adami HO, Day NE, Trichopoulos D, Willet WC (2001) Primary and secondary prevention in the reduction of cancer morbidity and mortality. *Eur J Cancer* **37**: 118–127
- Alibhai S (2006) Cancer screening: the importance of outcome measure. *Crit Rev Onco Hematol* **57**: 215–224
- Bertrand JT, Magnani RJ, Rutenberg N (1996) *Evaluating Family Planning Programs*. The Chapel Hill, NC: Evaluation Project
- Collins MM, Barry MJ (1996) Controversies in prostate cancer screening. Analogies to the early lung cancer screening debate. *J Am Med Assoc* **276**: 1976–1979
- Dinis-Ribeiro M, Yamaki G, Miki K, Costa-Pereira A, Matsukama M, Kurihara M (2004) Meta-analysis on the validity of pepsinogen test for gastric carcinoma, dysplasia or chronic atrophic gastritis screening. *Med Screen* **11**: 141–147
- Everett SM, Axon ATR (1997) Early gastric cancer in Europe. *Gut* **41**: 142–150
- Everett SM, Axon ATR (1998) Early gastric cancer: disease or pseudo-disease? *Lancet* **351**: 1350–1352
- Ferlay J, Bray F, Pisani P, Parkin DM (2001) *GLOBOCAN 2000: Cancer Incidence, Mortality and Prevalence Worldwide, Version 1.0* Vol. IARC CancerBase No. 5. Lyon: IARC Press
- Genta RM (2004) Screening for gastric cancer: does it make sense? *Aliment Pharmacol Ther* **20**: 42–47
- Hosmer DW, Lemeshow S (1999) *Applied Survival Analysis*. New York: John Wiley & Sons
- Inaba S, Hiramaya H, Nagata C, Kurisu Y, Takatsuka N, Kawakami N, Shimizu H (1999) Evaluation of a screening program on reduction of gastric cancer mortality in Japan: preliminary results from a cohort study. *Prev Med* **29**: 102–106
- Llorens P (1991) Gastric cancer mass survey in Chile. *Semin Surg Oncol* **7**: 339–343
- Miki K, Morita M, Sasajima M, Hoshina R, Kanda E, Urita Y (2003) Usefulness of gastric cancer screening using the Serum Pepsinogen Test Method. *Am J Gastroenterol* **98**: 735–739
- Mohr RB (1988) *Impact Analysis for Program Evaluation*. Chicago: Dorsey Press
- Oishi Y, Kiyohara Y, Kubo M, Tanaka K, Tanizaki Y, Ninomiya Y, Doi Y, Shikata K, Yonemoto K, Shiota T, Matsumoto T, Iida M (2006) The serum pepsinogen test as a predictor of gastric cancer. *Am J Gastroenterol* **163**: 629–637
- Oshima A, Hirata N, Ubukata T, Umeda T, Fujimoto I (1986) Evaluation of a mass screening program for stomach cancer with a case-control study design. *Int J Cancer* **38**: 829–833
- Parkin DM, Bray FI, Devesa SS (2001) Cancer Burden in the year 2000. The global picture. *Eur J Cancer* **37**: 54–66
- Parkin M, Pisani P (1999) Gastric Cancer. In *Cancer Screening* Kramer B, Gohagan J, Prorok PC (eds) New York: Marcel Dekker Inc
- Pisani P, Oliver WE, Parkin DM, Alvarez N, Vivas N (1994) Case-control study of gastric cancer screening in Venezuela. *Br J Cancer* **69**: 102–105
- Sackett DL, Haynes RB, Guyatt GH, Tugwell P (1991) *Clinical Epidemiology: A Basic Science of Clinical Medicine*. Boston, MA: Little Brown & Co. Inc.
- Salas J (1975) Lesiones precancerosas del estómago en Costa Rica. *Patología* **15**: 63–79
- Sasagawa T, Solano H, Mena F (1999) Early detection of gastric cancer project in selected high-risk area in Costa Rica. *Gastrointest Endosc* **50**: 594–595
- Sasagawa Y, Sasagawa T, Takasaki K (2002) Mass screening for gastric cancer performed in Costa Rica (Abstract). *Nippon Shokakibyo Gakkai Zasshi* **99**: 577–583
- Schram A (2005) *Análisis de costo-efectividad y costo-beneficio Programa de Detección Temprana de Cáncer Gástrico En Costa Rica*. San José, Costa Rica: Instituto de Investigaciones en salud (INISA), Universidad de Costa Rica
- Sierra R, Parkin M, Muñoz-Leiva G (1989) Cancer in Costa Rica. *Cancer Res* **49**: 717–724
- Sierra R, Rosero-Bixby L, Antich D, Muñoz G (1995) *Cancer en Costa Rica: Epidemiología Descriptiva: Mortalidad 1970-1990. Incidencia 1984-1990*. San José, Costa Rica: Editorial de la Universidad de Costa Rica
- Statacorp (2005) *Stata Statistical Software: release 9.0*. College Station, TX: Stata Corporation
- Tsubono Y, Hisamichi S (2000) Screening for gastric cancer in Japan. *Gastric Cancer* **3**: 8–18
- World Bank (2006) HNPStats. Web page <http://devdata.worldbank.org/hnpstats/>, accessed 05/30/2006



This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 3.0 License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>.