Gut, 1991, **32**, 220–224

PROGRESS REPORT

Screening for gastrointestinal cancer: an epidemiological review

J Weil, M J S Langman

The principle underlying screening is simple. Disease detected early will be treated better and the patient derive benefit. Since most people to whom screening tests are applied will never actually develop the disease for which the test is proposed, and those who would develop the disease are at the time free of symptoms, the test must meet certain basic criteria.[1 2] These are essentially of simplicity, safety, acceptability, reproducibility, and validity. Complex procedures are like to be impracticable because they are excessively time consuming, or costly, and those which have an element of risk will be unacceptable because, even if the disease in question were dangerous, it has to be borne in mind that most of the subjects would not have developed it. However simple and safe a procedure may be, it must also be acceptable. Thus in the United Kingdom sigmoidoscopic screening is less likely to be acceptable to the general public than in North America. Screening procedures for cervical cancer are less acceptable to the poor and less knowledgeable than to the well off, a point of considerable importance given the greater propensity of poorer women to develop cervical cancer. In addition, the test must be reproducible and valid: an examination which gives differing results at different times or when performed by different people is obviously defective. To obtain a useful conclusion it may need repeating several times and even then the weight attached to the conclusion is diminished. Occult blood tests are examples of procedures which give variable results, and which also have poor patient acceptability.[3]

Measures of test value

SENSITIVITY AND SPECIFICITY

These are the classical measures. The former denotes the proportions of true cases of the disease which are actually detected, while the latter measures the ability of a test to identify correctly those who are disease free. Ideally any test combines high specificity and sensitivity, and so defines accurately those with and without the characteristic under consideration. In practice, however, sensitivity and specificity are inversely related, and therefore any test which indicates all those who have the disease characteristic under consideration is likely to include a large number of people who are actually free of it. By contrast, a test which confidently indicates those who are disease free will also suggest that others are disease free when in fact they are not.

Predictive value of occult blood testing in populations of high and low frequency of true disease

	True disease	
	Positive	*Negative*
Occult blood:		
Positive	60	40
Negative	10	390
Total	70	430
Test sensitivity 60/70=85·7%		
Test specificity 390/430=90·7%		
Positive predictive values 60/100=60%		
Occult blood:		
Positive	60	400
Negative	10	3900
Total	70	4300
Test sensitivity 60/70=85·7%		
Test specificity 3900/4300=90·7%		
Positive predictive values 60/460=13%		

PREDICTIVE VALUE

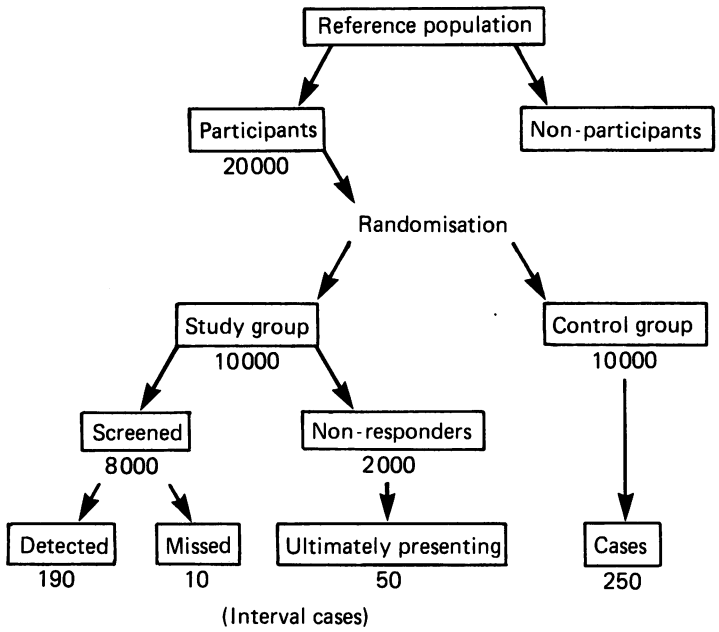
After a test is first introduced it will, if it shows adequate sensitivity and specificity, be more widely applied. This change of circumstances will, however, affect performance. A test is likely to be first used in a situation where the frequency of the disease under consideration is high. Let us suppose the test incorrectly labels as diseased 5% of healthy people. In the initial study the impact of this apparently small false positive rate might be low. If, however, the test is applied more widely in situations where disease frequency is much lower, and the number of healthy people screened is much higher, then the test will identify an undesirably large number of people who are actually healthy as diseased. The Table gives a theoretical example of such an occurrence where a faecal occult blood test is applied initially to a hospital group and then generalised to the population at large. Although sensitivity and specificity remain unchanged, and apparently high, the enlargement of the true disease free group by tenfold results in a tenfold increase in the numbers of healthy people labelled as diseased. This will have major implications for the cost effectiveness of the screening test.

Measuring the value of tests

The randomised controlled trial is the classical means of showing test efficacy. To give a full perspective the study requires that a random

Department of Medicine, Queen Elizabeth Hospital, Birmingham B15 2TH
J Weil
M J S Langman
Correspondence to:
Professor M J S Langman.

Accepted for publication
26 March 1990



Outcome good	95	2	10	50
Outcome poor	95	8	40	200

Theoretical outcome of a controlled intervention study for early detection of colorectal carcinoma using Haemoccult slides of 95% sensitivity and specificity.

sample be taken from a reference population and divided into a study group of invitees and a control group who are not subject to any form of intervention and who undergo normal medical care. The Figure shows the full layout of a theoretical study of a disease which has a prevalence of 2·5% and where early treatment might be expected to reduce mortality from 80% to 50%, figures which might apply for cancer of the large intestine. Outcome has to be compared in all 10000 invitees, including the 2000 non-responders, and in the 10000 control subjects. Assuming a test of 95% specificity and sensitivity then the 8000 participants will include 200 with the disease in question, and in 190 this will be correctly detected. The outcome will be good in half and bad in half. We, however, must also add in the outcomes in those falsely labelled as negative and add in the 2000 non-responders, who incidentally may have a poorer prognosis than responders, although in the figure we have assumed them to be similar. Taken overall in this idealised study we will have approximately doubled the number in the study group with good outcome compared with the outcome in the control subjects (107 with good outcome compared with 50 among the controls). The design of the study has allowed some important comparisons. Firstly, the equivalent disease frequency in test and control groups makes it likely that the data are generalisable to the reference population. This overall comparison between test and control is only possible because outcome in responding invitees and in non-responders has been studied and not solely that in responding invitees and control subjects. But even if a full data set is available, caution must be observed before assuming that improved outcome is indeed obtained after screening. Outcome must be measured when it is decisively known. Thus at least five years must elapse for the outcome of treated intestinal cancer to be clear and so lead time bias to be eliminated.

Incident and prevalent disease

The frequency of disease is often expressed clinically as (say) 5% of patients who have had disease x will suffer from y afterwards. Although this may be accurate, it gives no information about time elapsed during which disease complications occur. Greater precision is needed, and it is necessary to express figures as incidence or prevalence rates. These terms have precise meanings yet are often taken to be equivalent measures of overall disease frequency. Incidence rates measure the number of new cases of x per unit of population in a defined time period – for instance, as the number of new cases of colonic cancer per 100000 population per year in those aged 55 to 59 years. Prevalence, by contrast, measures the number of cases extant in the community at a given point in time, however long they have been there.

It therefore follows that a single screening procedure, assuming complete efficiency, will detect all prevalent cases of a set disease however long present. A second screening will detect incident (new) cases since the last investigation plus any missed initially. The effect of the second screening procedure will therefore be related to the rate at which new cases develop. That cannot be judged from the pickup at initial screening because there is no means of knowing how long the disease detected initially has been present. Furthermore, although one disease may develop twice as frequently (or have double the incidence rate of another), the amount of disease in the community will depend on their time courses. If the disease with the high frequency runs its course twice as fast then the actual amount of both diseases in the community will be the same at any one time. Therefore an informed decision about the timing of repeat screening requires knowledge of the rate at which disease develops and of the time for which it can be left undetected without imperilling the patient.

Lessons for screening for gastrointestinal cancer

OESOPHAGUS

In hospital based studies 10% of patients with Barrett's oesophagus were found to have oesophageal adenocarcinoma.[4,5] Such data, however, give no information about the rate at which cancer develops not can it be generalised to the total population with Barrett's oesophagus, which is likely to be much larger than the group presenting at hospital. Clinical behaviour may differ and, furthermore, the data are subject to inflation in accordance with the postulate of Berkson's fallacy.[6] The chances of detecting coincidental disease are multiplied owing to the possibility of detecting the two together because of examination of cases with Barrett's disease found to have cancer, and, separately, of cancer cases found to have Barrett's disease. In two

large studies where the incidence of oesophageal cancer was examined by follow up of patients with Barrett's oesophagus by yearly endoscopy[7] or postal questionnaire[8] an incidence of approximately one case every 170 patient years was found. Although increased over the general population frequency by 40-fold the chances remained low in absolute terms because oesophageal cancer per se is relatively uncommon. In the first study neither patient was fit for surgery, and in the second there was no difference in survival compared with that of an age and sex matched control population. A further factor inhibiting surveillance is that the natural history of progression from Barrett's epithelium to adenocarcinoma is not understood so that it is impossible to concentrate on patients at particular risk. Thus high grade dysplasia is common in patients who have developed cancer but is otherwise rare.[9-11] The case for screening is therefore faulty because the natural history of the disease is not understood, the early stage is inadequately recognised, and early treatment is not of clear benefit.

STOMACH

Endoscopic screening of asymptomatic patients is undertaken on a wide scale in Japan where benefits are believed to outweigh the costs and discomfort. In European populations where gastric cancer is less common endoscopic screening has been undertaken in patients with pernicious anaemia or after partial gastrectomy for benign ulcer disease where the risks of developing cancer are perceived to be raised.[12-14] Though the extent of risk is contested,[13-17] it seems likely that any material change only becomes evident some 20 years after gastric resection.[15 16] Even then the rise in risk, of the order of a doubling, is small, and furthermore is likely to occur in a fairly elderly population where the benefits of curative surgery in terms of prolonging life are diminished and the risks of operative complications and postoperative death must be raised. In absolute terms the risk is still low – even after 30 years the risk is of the order of one case for every 300 patient years of follow up.

If prevalent screening studies indicate a high rate of appreciable abnormality then this may suggest that the latent asymptomatic phase may be prolonged, indicating that in an elderly population with a limited life expectancy the gain to the patient from surgery after detection of an asymptomatic lesion may be limited.

Most gastric cancers occur in patients without predisposing disease and therefore a better strategy might be to consider case finding in patients aged 40 and over with recent onset of symptoms to determine if reducing patient and doctor delay before diagnosis will confer benefit.[18]

LARGE BOWEL

Ulcerative colitis

The raised risk of cancer in patients with extensive colitis of 10 or more years' duration justifies attempts at screening.[19] The technique used, repeated colonoscopy with multiple biopsies,[20] is unpleasant for the patient, assumes that colonoscopy can reliably detect cancer if present, or that biopsy proved dysplasia is a marker for it, that early case finding improves prognosis, and that the number of colonoscopies required in the detection programme is small. Patient discomfort inhibits compliance. Thus 15% of patients failed to attend in one study.[21] The reliability of colonoscopy will also be likely to vary. Dysplasia as a marker for the presence of cancer is of low sensitivity and specificity. Just under half of all patients undergoing resection who have severe or high grade dysplasia are found ultimately to have cancer in the resected colon.[22] Systematic blinded pathological study suggested that high or low grade distant dysplasia was detectable in just under three quarters of resected specimens with colon cancer, and high grade distant dysplasia in half of these.[23] Though the overall detection rate of dysplasia in those with cancer may seem high, it is likely that there are many more people with low grade dysplasia who do not have cancer. Dysplasia may also regress or pathologists' opinions on what is or is not dysplasia may vary. Regression is well described for cervical cancer,[24] and there is no inherent reason why the same should not be true in the colon. A greater problem may be sampling variability within the colon. Apart from interobserver variation in interpretation of biopsy specimens we have to take account of sampling variation. It seems plausible to suggest that the higher the proportion of dysplastic specimens the greater the chances of neoplastic change, but evidence is lacking.

If screening is effective it should improve prognosis. Controlled comparisons with unscreened patients are not available and the limited clinical data do not help. Published reports have typically described sets of six or seven cases[25-30] and these have varied in Dukes's classification from almost all grade A to none of group A and mainly group C. Such an inconsistent pattern does not suggest that lesions have generally been detected early, and even if they were it would require prolonged follow up to take account of lead time bias before accepting improved outcome.

The costs of screening can be expected to be high. Collins and his colleagues estimated that given a perfect outcome, 50 cancers might be detected by a yearly colonoscopic surveillance programme requiring a total of nearly 10 000 colonoscopies in 1000 patients.[22] Put another way, in 20 centres each with 50 patients at risk, a colonoscopist might expect at best to detect one cancer every fourth year while conducting 50 examinations a year.

It seems logical to recommend a screening rate which would be inversely proportional to the hazard. Unfortunately, it is difficult to derive a practical scheme. Thus in Chicago 99 patients with pancolitis were screened yearly with biopsy specimens being taken every 10 cm.[31] Twenty six developed low or high grade dysplasia or cancer after 30 years from onset, half having low grade and a third high grade dysplasia.

Benefit has to be assumed; the authors do not present evidence on dysplasia progression rates nor do they give the number of true cancers or

discuss their outcome. They comment: 'it is assumed but not known for this analysis that the presence of a dysplastic lesion is associated with the development of colonic cancer, and earlier detection may be associated with favourable prognostic factors.'

Adenomatous polyps and cancer

It is here that screening strategies for cancer come closest to meeting requirements. The problem is plainly important – 17 300 people died of colonic cancer in England and Wales in 1985,[32] the five year survival rate having remained unchanged at 50% for several decades. Occult blood testing is safe and inexpensive, the polyp-cancer sequence is generally accepted,[33] and the method is capable of detecting asymptomatic disease.[34-36] Mathematically it can be argued that faecal occult blood screening could cut colorectal mortality by a third in a high risk model.[37] The test method, however, is imperfect, the proportions taking up the test have varied between 15% and 85% according to circumstances, with common figures being 55% to 65%. Apart from detecting cancers, large numbers of adenomas are found. The clinical importance of this is difficult to weigh because progression from polyp to cancer is not inevitable. Indications of benefit include the detection of high proportions of early stage tumours, although final assessment must be delayed to take account of detection bias (or detection of clinically irrelevant changes) and lead and length time biases (early detection which does not influence disease behaviour). The costs of programmes, if effective in preventing fatalities, have been estimated at approximately $250 000 at age 55 or $70 000 at age 65[38] per death prevented. Improvements in test specificity and sensitivity could plainly reduce these greatly.

It has, for example, been suggested[39] that Haemoccult slides be rehydrated before testing to overcome loss of sensitivity due to drying. Mandel and colleagues report a gain in sensitivity from 80·8% to 92·2% after the introduction of rehydration of Haemoccult slides in the University of Minnesota's colon cancer control study, with a fall in specificity from 97·7% to 90·4%.[40] If we used the data of Hardcastle *et al* as an example we can examine the results of the apparently large gain in sensitivity while only modestly reducing specificity.[41] In that study, in 27 000 screened subjects 63 cancers were detected, or roughly 20 cancers detected for every 10 000 screened. A gain of sensitivity of 12% will result in detecting two or three more tumours, but a loss of specificity of 7% will mean that 700 extra will need examination by colonoscopy. Such figures illustrate the care necessary in considering the beguiling case for surveillance and the relative technical merits of different procedures.

The virtues of screening programmes are not self evident; proof of benefit is generally lacking and indications of possible benefit are limited to colorectal cancer screening where improved test methods are needed.

1 Wilson JMG, Junger G. *Principles and practice of screening for disease.* Geneva: World Health Organisation, 1968. (WHO public paper 34.)
2 Barker DJP, Rose G. *Epidemiology in medical practice.* 2nd ed. Edinburgh: Churchill-Livingstone, 1979.
3 Simon JB. Occult blood screening for colorectal carcinoma: a critical review. *Gastroenterology* 1985; **88:** 828–37.
4 Haggit RC, Tryzelaar J, Ellis FH, Colcher H. Adenocarcinoma complicating columna epithelial lined (Barretts) esophagus. *Am J Clin Pathol* 1978; **70:** 1–5.
5 Naef AP, Savary M, Ozello I. Columna lined lower esophagus an acquired lesion with malignant predisposition: report on 140 cases of Barretts esophagus with 12 adenocarcinomas. *J Thorac Cardiovasc Surg* 1975; **70:** 826–35.
6 Berkson J. Limitation of the application of fourfold table analysis to hospital data. *Biometrics* 1946; **2:** 47.
7 Spechler SJ, Robbins AH, Rubins HB, *et al.* Adenocarcinoma and Barrett's esophagus. An over-rated risk? *Gastroenterology* 1984; **87:** 927–33.
8 Van Der Veen AH, Dees J, Blankensteijn JD, *et al.* Adenocarcinoma in Barrett's oesophagus: an overated risk. *Gut* 1989; **30:** 14–8.
9 Hamilton SR, Smith RRL. The relationship between columna epithelial dysplasia and invasive adenocarcinoma arising in Barrett's oesophagus. *Am J Clin Pathol* 1987; **87:** 301–12.
10 Lee RG. Dysplasia in Barrett's esophagus: a clinical-pathologic study of six patients. *Am J Surg Pathol* 1985; **9:** 845–52.
11 Reid BJ, Lewin K, Van Deventer G, *et al.* Barrett's esophagus: high-grade dysplasia and intramucosal carcinoma detected by endoscopy biopsy surveillance [Abstract]. *Gastroenterology* 1986; **90:** 1601.
12 Stockbrugger RW, Menon GG, Beilly J, *et al.* Endoscopic screening in patients with pernicious anaemia. In: Cotton P, ed. *Early gastric cancer.* Welwyn Garden City, Hertfordshire: Smith Kline & French Laboratories, 1981: 64–6.
13 Huigbregste K. Endoscopic screening for malignancy in the gastric remnant. In: Cotton PB, ed. *Early gastric cancer.* Welwyn Garden City, Hertfordshire: Smith Kline & French Laboratories, 1981: 63–8.
14 Farrands PA, Blake JRS, Ansell ID, Cotton RE, Hardcastle JD. Endoscopic review of patients who have had gastric surgery. *Br Med J* 1983; **286:** 755–8.
15 Lundegardh G, Adams H, Hemlick C, Zack M, Meirik O. Stomach cancer after partial gastrectomy for benign ulcer disease. *N Engl J Med* 1988; **319:** 195–200.
16 Offerhaus GJA, Tersmette C, Hurbredtse K, *et al.* Mortality caused by stomach cancer after remote partial gastrectomy for benign conditions: 40 years of follow-up of an Amsterdam cohort of 2633 post-gastrectomy patients. *Gut* 1988; **29:** 1588–90.
17 Langman MJS, Logan RFA. Screening for gastric cancer after gastric surgery. *Lancet* 1983; ii: 17.
18 Aste H, Amadori D, *et al.* Early gastric cancer detection in four areas at different gastric cancer death rate. *Acta Endoscopa* 1981; **11:** 123–32.
19 Gyde SN, Prior P, Allan RN, *et al.* Colorectal cancer in ulcerative colitis: a cohort study of primary referrals from 3 centres. *Gut* 1988; **29:** 206–17.
20 Morson BC, Pang LSC. Rectal biopsy as an aid to cancer control in ulcerative colitis. *Gut* 1967; **8:** 423–34.
21 Lennard-Jones JE. Cancer risk in ulcerative colitis. Surveillance or surgery. *Br J Surg* 1985; **72** (suppl): 584–6.
22 Collins RH, Feldman M, Fortran J. Colon cancer, dysplasia and surveillance in patients with ulcerative colitis. *N Engl J Med* 1987; **316:** 1654–8.
23 Ransohoff DF, Riddell RH, Levin B. Ulcerative colitis and cancer; problems in assessing the diagnostic usefulness of mucosal dysplasia. *Dis Colon Rectum* 1985; **28:** 383–8.
24 Spriggs AI. Pre-cancerous states of the cervix uteri. In: Carter RL, eds. *Pre-cancerous states.* London: Oxford University Press, 1984: 317–55.
25 Lennard-Jones JE, Morson BC, Park FRC, Ritchie JK, Shove DC, Williams CB. Cancer in colitis: assessment of the individual risk by clinical and histological criteria. *Gastroenterology* 1977; **73:** 1280–9.
26 Lennard-Jones JE, Morson BC, Ritchie JK, Williams CB. Cancer surveillance in ulcerative colitis: experience over 15 years. *Lancet* 1983; ii: 149–52.
27 Nugent FW, Hagitt RC. Results of a long term prospective surveillance program for dysplasia in ulcerative colitis. *Gastroenterology* 1984; **86:** 1197.
28 Nugent FW. Surveillance of patients with ulcerative colitis. Lahey Clinic results. In: Winawer SJ, Schottenfeld D, Sherlock P. *Colorectal cancer prevention and screening.* New York: Raven Press, 1980: 375–80.
29 Rosenstock E, Farmer RG, Petras R, Sivak MVJr, Rankin GB, Sullivan BH. Surveillance for colonic carcinoma in ulcerative colitis. *Gastroenterology* 1985; **89:** 1342–6.
30 Blackstone MO, Riddell RH, Rogers BHG, Levin B. Dysplasia-associated lesion or mass (DALM) detected by colonoscopy in long-standing ulcerative colitis: an indication for colectomy. *Gastroenterology* 1981; **80:** 366–74.
31 Lashner B, Hanauer SB, Silverstein MD. Optimal timing of colonoscopy to screen for cancer in ulcerative colitis. *Ann Intern Med* 1988; **108:** 274–8.
32 Office of Population Censuses and Surveys. *Mortality statistics 1985. Series DH2.* London: HMSO, 1987.
33 Morson BC. Genesis of colorectal cancer. *Clin Gastroenterol* 1976; **3:** 505–26.
34 Hardcastle JD, Armitage NC, Chamberlain J, Amar SS, James PD, Balbour TW. Faecal occult blood screening for colorectal cancer in the general population. Results of a controlled trial. *Cancer* 1986; **58:** 397–403.
35 Allison JE, Feldman R. Cost benefits of hemoccult screening for colorectal carcinoma. *Dig Dis Sci* 1985; **30:** 860–5.

36 Gilbertson VA, McHugh RB, Shuman L, *et al*. The earlier detection of colorectal cancers. A preliminary report of the results of the occult blood study. *Cancer* 1980; **45**: 2899–907.
37 Eddy DM, Nugent FW, Eddy JF. Screening for colorectal cancer in high risk population. Results of a mathematical model. *Gastroenterology* 1987; **92**: 682–92.
38 Barry MJ, Mulley AG, Ritcher JM. Effect of workup strategy on the cost-effectiveness of fecal occult blood screening for colorectal cancer. *Gastroenterology* 1987; **93**: 301–10.
39 Wells HJ, Pagano JF. 'Haemoccult' test-reversal of false negative results due to storage. *Gastroenterology* 1977; **72**: 114.
40 Mandel JS, Bond JH, Bradley M, *et al*. Sensitivity, specificity and positive predictivity of the hemoccult test in screening for colorectal cancers. *Gastroenterology* 1989; **97**: 597–600.
41 Hardcastle JD, Thomas WM, Chamberlain J, *et al*. Randomised controlled trial of faecal occult blood screening for colorectal cancer. *Lancet* 1989; i: 1160–4.