

# Reducing mortality from colorectal cancer by screening for fecal occult blood

Snover DC, Bradley GM, Schuman LM, Ederer F for the Minnesota Colon Cancer Control Study

*New England Journal of Medicine* 328(19):1365-1371, 1993

## Introduction

Colorectal cancer is a major contributor to both cancer incidence and mortality. In Australia in 1991, there were 3343 new cases of colorectal cancer diagnosed, and 14439 deaths<sup>49</sup>. Epidemiological research has revealed general risk factors affecting the whole population (food types), and some specific risks of small populations (e.g., familial polyposis coli, inflammatory bowel disease). Hope has been placed in screening tests as a means of detecting the disease earlier. Earlier stage disease (Dukes' A) has a much more favourable prognosis. Attempts to substantially improve the prognosis of Dukes' B and C disease have been disappointing.

Screening tests for colorectal cancer include faecal occult blood tests, routine rectal examination, barium enema (single or double contrast), sigmoidoscopy and colonoscopy. The highly specific tests used to assess the whole bowel are also labour-intensive and expensive. The inexpensive, automatable tests are non-specific. Faecal occult blood (FOB) may become positive after a meat meal, or brushing teeth. Despite the suggested use of FOB in colorectal cancer screening there have been no trials of screening on which to justify the use of the test.

## Trial Summary

The University of Minnesota instituted in a Phase III randomised control trial of the effectiveness of faecal occult-blood screening in the reduction of colorectal cancer mortality within the general population. The trial commenced in 1975, and by 1977 had accrued 46,551 volunteer participants to annual, biennial and no screening arms. The trial continued for 17 years. The study purported to "present conclusive evidence ... of the effectiveness of fecal occult-blood screening in reducing mortality from colorectal cancer".

## Trial Design

### *Hypothesis*

The proposed null hypothesis was not stated. The implied null hypothesis was that, in the general population, the annual or biennial use of faecal occult-blood tests in the setting of a screening program does not reduce the subsequent mortality from colorectal cancer.

### *Sample size and Statistical Power*

It was not possible to determine from this paper whether prospective sample size determinations were undertaken. The group's initial paper describing the protocol did carry out these estimations (assumed power of 90% at 0.05 level for a 50% reduction in mortality rates for annually screened population compared with the assumed control group mortality rate of 0.001)[50]. Adjustments in sample size was also made based on assumed attrition within the populations(50).

## Trial Conduct

### *Eligible Population*

Patients included were male or female volunteers aged 50-80 years from Minnesota. The volunteers were drawn from contacts with the American Cancer Society, and fraternal, veteran and employee groups.

Patients were not eligible if they had a previous history of colorectal cancer, familial polyposis coli, or ulcerative colitis. Patients also were excluded if "bedridden or otherwise disabled".

The eligible population is poorly defined for several reasons :

- the socioeconomic profile of this group does not match the general population. They are likely to be of higher socioeconomic standing as a result of a higher educational standard (fraternal) and a higher average wages (employee). Members of these groups are likely to be better educated about health risks such as colorectal cancer, and also by virtue of socioeconomic status, be more likely and more able to access the health system. Inclusion in this study may prompt control members to undertake their own screening.

No measures of socioeconomic status are presented.

- the treatment of previous colonic polyps is not specifically addressed
- the definition of "otherwise disabled" requires further definition. The mortality from cancer is frequently related to performance status, thus excluding the poor performance candidates may select longer surviving patients. Patients with physical disabilities, e.g., veterans who are amputees, may be expected to have a normal lifespan and should be included in a Phase III trial of screening.

The initial report also reported a change in eligibility criteria that occurred in 1980 when the age groups 40 - 50 and >80 were excluded, but no reasons were given for this change<sup>50</sup>. Presumably the colorectal cancer rates were too small in one group, and length of followup too short in the other.

### ***Patient Selection***

Participants were "selected" after agreeing to participate after invitation. The size of the invited population is not given.

Volunteers are not representative of the population from which they are drawn. They are more likely to comply with the investigators wishes (i.e., return for follow-up, fill out questionnaires, respond to letters) and to undertake extracurricular activities that impinge on the validity of the study (i.e., conduct their own screening).

### ***Patient Stratification and Randomization***

The paper states that stratification occurred for age, sex and place of residence.

The original report<sup>50</sup> describes the stratification method in detail. Participants were stratified according to residence within Minnesota (there were six regions), and then according to sex. Within these 12 groups, participants were ranked for age, and starting from oldest, were assigned sequentially into groups of three. These trios were then randomised to one of six permutations of treatment assignment within the group (Permutation 1 - control, annual, biennial; Permutation 2 - control, biennial, annual; etc. until all combinations were assigned). Once the treatment had been assigned, the participant was notified of required actions by mail.

This method of randomisation will achieve its purpose.

### ***Patient Exclusion***

No patients were excluded once entered into the trial. No mention is made of protocol violations at entry.

## **Treatment Description**

### ***Screening Protocol***

After a special diet designed to reduce false positives, participants presented two smears from three consecutive stools (6 slides) which were processed in standardised controlled procedure by the University of Minnesota Pathology Department by personnel dedicated to this task.

Prior to 1982 only delayed slides were rehydrated. Later all slides were rehydrated. Rehydration altered positivity (2.4% v 9.8%), especially among the elderly patients and men. Sensitivity was increased (80.8% v 92.2%), specificity decreased (97.7% v 90.4%), and the positive predictive value (PPV) fell (5.6% v 2.2%). The dual methods of slide examination introduced a confounding variable. The outcomes of these two screened populations are not compared, nor are the reasons for the change described. The rates of rehydration in each group are not available.

The end result was however to increase the yield, detecting a higher proportion of cases. In the 2x2 table below, PPV reflects the number of true-positive cases a as a proportion of positive tests a+b, while sensitivity reflects the number of true-positive cases a as a proportion of the total cases a+c.

	Disease +	Disease -
Test +	a	b
Test -	c	d

When the total number of cases, a+c (e.g., 100 cases in the population) is static, a higher proportion of cancers are detected (92 v 81) in a smaller total number of tests (42,704 v 59,447), but the total number of positive tests (i.e, recall for diagnostic work-up) increases (Table 2).Table 2: Effect of rehydration on FOB slide performance

	Dry slides	Rehydrated slides
Sensitivity=a/(a+c)	80.8%	92.2%

Specificity=d/(b+d)	97.7%	90.4%
PPV=a/(a+b)	5.6%	2.2%
if a+c (total cases)	100	100
a (true positives)	81	92
b (false positives)	1365	<b>4090</b>
c (false negatives)	19	8
d (true negatives)	57982	38514

The authors do not supply absolute numbers in the text. Total FOB tests done was not reported, however the potential maximum for annual and biennial screening with six slides over the 13 years was 1,401,810 slides  $(((15570*10) + (15587*10/2))*6)$ .

### **Case finding**

If a single slide was positive, the participant was notified and asked to return for review to either the University of Minnesota Hospital or private physician. The protocol on return was detailed in the Methods, and variations with respect to bowel imaging (barium enema, double contrast barium enema, upper GIT series, rigid proctosigmoidoscopy) are noted.

All visible lesions were biopsied. The variation in diagnostic procedures does not introduce a confounding variable into this study of a screening modality. The use or disuse of a barium enema does not affect the positivity rates of the FOB screen. The therapeutic strategies applied on finding a lesion are not detailed. Definitive treatment of a detected malignancy appears to be left to the clinician's discretion. Variations in therapy would not confound the analysis, unless reports of advances cause participants to self-screen.

### **Follow up**

All participants were contacted annually by mail and specifically asked about colorectal cancer and polyp detection in the previous 12 months. It is not clear if different questionnaires were sent to each group. A protocol for telephone follow-up of non-responders was defined at trial inception(50).Where lesions were reported, formal reports and tissue review was arranged.

### **Pathology and Case Verification**

1. Central pathology review at the University of Minnesota Hospital was undertaken of pathological material and reports. One pathologist, blinded to randomization, undertook this role. The central pathology review provided for consistent histological diagnosis. Potential difficulties occur when staging relied on the macroscopic appearance of the tumour, which may not be reviewable, but microscopic appearances are taken from other areas, e.g., invasion through bowel wall macroscopically but no sections showing muscle or serosa.
2. A Deaths Review Committee of blinded clinicians (including the review pathologist) assessed all relevant data in determining cause of death included hospital notes and death certificates. This mechanism also enhances the study's veracity.

## **Trial Analysis**

### **Statistical Analysis**

The endpoints studied included death (all causes), death (colorectal cancer) and incidence of colorectal cancer and were analysed by life-table methods. Cox regression analysis was used in adjusting for stratified variables only, and did not affect conclusions.

The original statistical design did not require interim analysis looking specifically for a difference between the three screening arms. The sequential log-rank statistic used, which required a one-sided  $p = 0.025$  with 80% power for a 25% reduction in mortality from colorectal cancer, appeared to have been instituted between 1980-1985. Its periodicity was not defined.

Stopping guidelines are a technique used to permit the early cessation of trials with a conclusive result thereby reducing expected sample size and allowing early dissemination of information, at a cost of short followup and wider confidence intervals. Sequential analyses are an unreliable statistical method without special controls, since uncorrected sequential analyses have an increased likelihood of discovering a falsely positive " $p < 0.05$ " result, i.e., a type I ( ) error (7,28,51).

A variety of stopping rule methods exist, including predetermined or sequential analysis, and variability in the type I

error determination. The p value chosen for significance will however be less than 0.05. Pocock suggests that  $p = 0.01$  should be satisfied for major trials using sequential analysis that will affect clinical practice(7). The methods of assigning type I levels include the choice of a very small p value ( $<0.001$ ) for early stopping, a changing p value proportional to the trial's duration, or a prespecified discount rate at which the p value is expected to change ("spending rate"), and from this deriving significance levels (a stopping boundary) for each analysis time<sup>7,28</sup>. Some authors have suggested that analyses should occur at regular time or event intervals(57).

The trial was recommenced in the third year of followup because the mortality rate in the control group was lower than expected. Given that no patients were accrued after the decision to recommence screening, prognostic factor and patient management imbalances are not likely to affect results.

The authors selected the "spending rate" method to control for increased number of analyses, however they have employed early stopping rules for a late stopping circumstance. Although the trial was initially to run for 10 years, the recommencement of screening effectively permits the trial to run until a positive result ensues or all patients die. The results were prepared for publication when the first statistically significant sequential analysis occurred, some 13 years after commencement. The publication of results after 10 years was not undertaken as planned. At analysis in 1985, the control and annual screening arms had identical mortality and incidence, presumably despite a stage shift in initial presentation. A reasonable explanation at this time would have been the lack of efficacy of FOB screening, or that the differences were unlikely to be as large as a 50% reduction in mortality. An analysis of the likelihood of achieving a significant result (e.g., stochastic curtailment method) would possibly have revealed the same conclusion. The data committee did not believe this explanation or undertake such an analysis. The statistical analysis is suboptimal in the following aspects :

- a. the trial design included a predetermined sample size aimed at detecting a 50% reduction in colorectal cancer mortality. The use of a 25% reduction in the interim analysis after accrual was closed, calls into question the sample size.
- b. the trial design called for 5 years accrual and minimum 5 year followup. Accrual occurred over the years 1977-1982, and a final report should have appeared in 1987. Three years into the followup, the trial was reinstated since the control and annual screening arms were not significantly different, and the mortality rate expected in the control group had not eventuated. This indicates the presence of a confounding variable, however no attempt to define or discuss this variable occurred. Publication at this time would have resulted in a negative trial, and evidence that FOB screening has no apparent value in this group.
- c. the number of colorectal cancer cases within the report are not in agreement.

Table 3: Comparison of reported patient numbers within the trial report.

	Table 4 (colorectal cases)	Figure 3 (colorectal cases)	Figure 4 (colorectal cases)	Table 5 (colorectal cases)	Text Figure 1 (all cases)	Figure (all cases)
Annual screen	323	354	177+17 (354)		15570	15534
Biennial screen	323	368	145+223 (368)		15587	15551
No screen	356	394			15394	15364
Total group			1116	1002		

- d. interim analyses are not designed for circumstances of late stopping.
- e. performance analysis of FOB screening assumes that a cancer found within a year of a positive test is a "true positive". This assumption will prejudice results within the annual screening group since all FOB positive patients who have a cancer diagnosed are true positives. The FOB performance rates should be determined on the results of diagnostic work-up on recall, not on events which may occur 11 months later. This assumption may explain the excellent sensitivity and specificity of FOB described.
- f. there are other uncertainties within the FOB performance data. Sensitivity, the test positive rate for the diseased population (true positive rate), is quoted at 80.8%-92.2%. The authors also quote that "only 1.8 percent of the cases in the control group was the diagnosis made after a positive fecal occult-blood test, as compared with 49.5 and 38.3 percent of those in the annually and biennially screened groups, respectively". In the annually screened, only 49.5% of detected cancers occurred within the setting of a very high sensitivity, and a very high screening rate. Given the lack of raw data, further delineation is not possible.
- g. the mortality rates analysed are cumulative rates averaging over the 13 year period. The changing rates of mortality resulting from the operation of unknown factors are therefore hidden rather than identified.
- h. there are a small number of events within the groups.

### Confounding Variables

a. the comparison of colorectal cancer mortality in annual, biennial, and no screening arms was the aim of the trial. A Phase III trial of screening should leave followup and treatment variables open to change, so long as screening detection was not associated with different diagnostic or therapeutic procedures. There was an adequate randomization and stratification method.

b. the comparison of rehydrated and dehydrated FOB slides. The change in this test during the trial was not optimal. With the increased sensitivity, it was more likely that the desired result would be achieved.

c. the comparison of short and long term screening. Screening occurred for 10 years although the trial only called for 5 years. Short term screening was shown not to produce positive results.

d. There was a three year delay without screening, the effect of this variable is unknown.

e. communication between screened and non-screened participants. It was highly likely that the performance of annual screening was a motivating factor in the actions of non-screened participants. The individuals attended the same groups regularly, this allowed their inclusion in the trial. The rates of off-trial screening (FOB or colonoscopy) were not sought or reported.

f. an unidentified variable operating to produce an abnormally low colorectal cancer mortality rate in the control group. This may be a serendipitous event, or may reflect an alteration in the behaviour of the control group brought on by the screening trial.

One possible explanation revolves around the access that this group has to health care and their ability to conduct their personal screening program as they are reminded by the screening groups. It is possible that this self program was not taken up with such vigour after the three year break when screening was reinstated.

However against this explanation is the lack of change in mortality from colorectal cancer in the annually screened group. If the control group conducted their own screening, then trends in the three year period of no screening should be the same. If a single screening (however obtained) produced a fall in mortality rate for 5 years, the same trend should have been seen in the biennially screened group. In the end, no plausible biological explanation exists for the mortality rates.

## **Trial Outcome**

### ***Criteria for Evaluation Endpoints***

The patients were observed at well-defined intervals for clinically-relevant endpoints that included :

- occurrence of colorectal cancer
- death from colorectal cancer
- death from other causes

### ***Toxicity***

Only the complication rates for colonoscopy (15 serious side effects in 12,246 colonoscopies) were reported. There is no toxicity from FOB screening.

### ***Protocol Violations***

The participation rate in screening was exceptional. Participants completing all screening opportunities comprised 46% and 60% of the annual and biennial screening groups. The majority completed more than half (77% and 82%). This performance over 13 years demonstrates a trial population with special characteristics, non-representative of the entire population.

No protocol existed post-screening since patient attendance was optional. Diagnostic regimens at the University Hospital were standardised, but not in the private clinic where fewer colonoscopies were performed and 3% had no examination.

Verification of cause of death was also exceptional, with 99.9% of death certificates were obtained and 99% of medical records.

### ***Exclusions***

No patients were excluded from analysis.

## **Results**

The trial reports the following results :

1. colonoscopy is relatively safe when performed frequently
2. in 46,551 participants there were 1002 colorectal cancer cases in 13 years.
3. 10,097 deaths occurred, of which 320 were due to colorectal cancer.

4. each trial arm had an identical colorectal cancer incidence (23-26 cases/1000 patients/year)
5. the cumulative annual mortality rate was lower in the annually screened group (5.88 deaths/1000 patients/year) than the other groups (8.33-8.83 deaths/1000 patients/year)
6. patients detected by screening were of lower stage and had a better survival.
7. patients undetected by screening had a worse survival, irrespective of their trial arm
8. five year survival is proportional to pathological stage; Dukes' A (94%), Dukes' B (84%), Dukes' C (57%), "Dukes" D (2.4%).
9. rehydrated FOB slides are more sensitive, but less specific than 'dehydrated' slides.
10. FOB slides are rarely used, outside the screening context, in the diagnosis of colorectal cancer. k. more than half of the colorectal cancers were not screen detected.

## **Conclusion**

The authors conclude that FOB annual screening reduces mortality from colorectal cancer by 33%, by detecting cancers at an earlier stage. They point to the high rate of colonoscopy, its potential disease-modifying effects on polyps, although this did not affect incidence rates, and its large costs. A brief search for confounding variables revealed that FOB tests in the unscreened, different therapy in the University setting and undiagnosed cases in the control group did not affect the study. They acknowledge that the three year hiatus in screening has an unknown effect.

## **OVERVIEW**

**This trial is extremely and unnecessarily complicated. A number of substantial criticisms can be levelled that compromised both the external and internal validity of the trial.**

- a. **the sampled population are a subset with greater than average access to health information and resources, and are more likely to respond to the publicity of FOB by undertaking a personal screening program.**
- b. **while the trial protocol was adequate, the occurrence of an unusual pattern of mortality calls for further investigation, before the decision to recommence the trial during the followup can be justified.**
- c. **the statistical methods are inappropriate and the number of events are small.**
- d. **the lack of advantage in the biennially screened group merits little comment, especially since the reduction in the incidence of Dukes' D tumours (which is the implied mechanism of the reduced mortality) is similar.**
- e. **the false positive rates are very high (12,500 colonoscopies after FOB positive in the University patients to detect a total of 1000 cancers in all patients - 9.8% in Table 6).**
- f. **the screening protocol only successfully detected 50% of available cancers.**
- g. **the economic issues involved in a potential 100,000 FOB slides per year and 1000 colonoscopies per year per 30,000 patients, are only superficially addressed.**
- h. **the overall death rate for the 13 years was unchanged.**

**This trial does not present "conclusive evidence" as stated, there is certainly an indication that annual screening is better than biennial screening, but the control group prevent any meaningful comparison with the other arms. Whether any health economy could afford this program on a population-wide basis is very doubtful.**

