Fecal Occult Blood Screening in the Minnesota Study: Sensitivity of the Screening Test

Timothy R. Church, Fred Ederer, Jack S. Mandel*  

Background: In the Minnesota Colon Cancer Control Study, which used guaiac slides to annually screen stool samples for blood, mortality from colorectal cancer was reduced by 33.4%. The reported sensitivity of this test for colorectal cancer was about 90%. However, results from another study estimated the sensitivity to be 25%–33%; other investigators have reported intermediate values. Given these contradictions, we examined screening sensitivity for colorectal cancer in the Minnesota study by several direct and indirect methods. Methods: In this reanalysis of data from the Minnesota study, we distinguished between sensitivity for colorectal cancer of the screening test (composed of six slides) and of the screening program (a series of such tests). We estimated screen sensitivity by adjusting the crude estimate from the final tests in each screening phase for colorectal cancer incidence in 5 years of follow-up, by modeling guaiac slide results at each screen as a function of the presence of occult blood, and by incorporating sensitive detection into a modification of a mathematical model developed by Lang and Ransohoff. Program sensitivity was estimated from the fraction of screen-detected cancers among all cancers diagnosed in screened individuals. Results: The crude estimate of program sensitivity was 89.4%, whereas the modified Lang–Ransohoff model estimates screen sensitivities at 94.1%–96.2%, consistent with the estimates from the other methods. Indirect measures, such as the association between the number of positive slides among the six slides in each set and the positive predictivity for colorectal cancer, are consistent with these estimates. Conclusions: The Minnesota study reduced mortality from colorectal cancer through use of a screening test with average screen and program sensitivities of about 90%. [J Natl Cancer Inst 1997;89:1440–8]

The Minnesota Colon Cancer Control Study, a randomized trial of mass screening for colorectal cancer in an asymptomatic population of 46,551 adults older than 50 years, demonstrated that offering screening annually for fecal occult blood with a colonoscopic examination in those who test positive results in a 33.4% decrease in mortality from colorectal cancer over a 13-year period when compared with usual care (1). From the fraction of screen-detected cancers among all cancers discovered within one screening interval of the last screen, sensitivity of the fecal occult blood test was estimated to be between 81% and 92% (2).

However, Ahlquist et al. (3) concluded that “... fecal occult blood is at best a flawed marker for colorectal neoplasia” after finding sensitivities for colorectal cancer that ranged from 25% to 33% in studies of two groups, one of patients whose colorectal cancer was recently resected and the other of relatives of those colorectal cancer patients. In mass population screening, sensitivity is the probability that a screening test given to persons with undiscovered disease will yield a positive result. Sensitivity is the proportion of prevalent cases that will be discovered by screening and thus will potentially benefit from early detection and treatment. Note that to be useful a test must discover the disease when prognosis can still be improved. A test sensitive for preclinical disease only after the disease has progressed too far to improve prognosis would be useless. The fecal occult blood test can detect a bleeding cancer in any stage. However, only for Dukes’ stage A, B, or C is there significant chance of improving the patient’s outcome. Hence, sensitivity alone is inadequate to determine the efficacy of the test.

Adenomatous polyps, widely believed to be precursor lesions, are sometimes discovered in individuals screened for colorectal cancer; the removal of these lesions potentially reduces a person’s chance of subsequently developing colorectal cancer. A screening test can be sensitive for polyps; however, since polyps are much less likely to bleed, they may be less detectable by fecal occult blood tests.

A complete definition of sensitivity must address what is being detected (e.g., colorectal cancer versus polyps), the nature of the test (e.g., a single guaiac slide, a set of six slides, or a 10-year program of annual screens using six slides), and whether sensitivity varies over time within an individual (e.g., stage-specific sensitivity). For fecal occult blood tests, each guaiac slide has a sensitivity, each set of slides has a sensitivity, and a program of periodic screenings has a sensitivity; the three sensitivities are distinct, although related. Furthermore, each may vary among populations, from time to time, and within the natural history of a tumor. If the definitions and assumptions are not stated explicitly, potential for confusion and misinterpretation results.

The results obtained by Ahlquist et al. (3) contradicted the estimates of sensitivity that were determined in the Minnesota study (2). In addition, their results have

*Affiliations of authors: T. R. Church, J. S. Mandel, Division of Environmental and Occupational Health, School of Public Health, University of Minnesota, Minneapolis; F. Ederer, Division of Biostatistics, School of Public Health, University of Minnesota, and The EMMES Corporation, Potomac, MD.

Correspondence to: Timothy R. Church, Ph.D., Division of Environmental and Occupational Health, School of Public Health, University of Minnesota, Box 807 Mayo, 420 Delaware St., Minneapolis, MN 55455. E-mail: trc@cccs.umn.edu

See “Notes” following “References.”

© Oxford University Press
led many authors to conclude (a) that the fecal occult blood test is insensitive to colorectal cancer (4–11) and (b) that most of the effect seen in the Minnesota study was from serendipitous discovery of colorectal cancer in the presence of a positive test due to other causes (12–15). We address the contradiction by clearly defining sensitivity, by refining estimates of sensitivity and of the contribution to mortality reduction of the fecal occult blood test in the Minnesota study, and by evaluating models to reconcile the results of Ahlquist et al. and the Minnesota study. In the process, we examine the validity of conclusions (a) and (b).

Subjects and Methods

Fecal Occult Blood Test and Sensitivity

The details of the design of the Minnesota study were reported previously (16). For the fecal occult blood test used in the study (Hemoccult; SmithKline Diagnostics, Palo Alto, CA), subjects were asked to place two fecal samples from each of three consecutive stools on six individual guaiac-impregnated slides. If at least one positive slide resulted, the screen was considered positive and the subject was offered a diagnostic regimen that included colonoscopy. Because no difference in incidence was evident in the Minnesota study after 13 years of screening, the difference in mortality must have been primarily due to the early detection of cancer. Thus, we have stressed the sensitivity of the fecal occult blood test to detect colorectal cancer and have placed less emphasis on the sensitivity of the test to detect polyps. Unless stated otherwise, sensitivity will refer to detecting colorectal cancer.

Definitions

Within a single screen, the probability that a particular guaiac-impregnated paper slide will be positive for a case of detectable, preclinical colorectal cancer is defined as the slide sensitivity. Screen sensitivity is defined as the probability that at least one of the six slides for a case of detectable, preclinical colorectal cancer will be positive at the time of a single screen. Screen sensitivity is a function not only of the slide sensitivity but also of the degree of dependence between the individual slide results. Program sensitivity is defined as the probability that a case in the detectable, preclinical stage at any time during an ongoing screening program (and ending with the last screen) will test positive in at least one of the screens; this definition is the expected value of the program sensitivity given by Morrison (17). Program sensitivity is a function of the screen sensitivity, the frequency and duration of screening, and the degree of dependence between the screens. For a particular disease, it also depends on the distribution among the population of preclinical duration, which is defined as the interval of time from when the disease becomes detectable until it is clinically diagnosed in the absence of screening.

We are concerned primarily with screen sensitivity and program sensitivity; the slide sensitivity is important only in that it determines the other two. Note that each case of colorectal cancer may have its own sensitivity and what is usually thought of as sensitivity is an average of these individual sensitivities over the population and over time. We will focus on estimating this average sensitivity rather than on attempting to determine what the sensitivity for an individual colorectal cancer case might be.

To describe how the sensitivity changes as a colorectal cancer progresses, we define a screen sensitivity function relating the probability of detection by screening to disease progression (18–20). In the ”Appendix” section, we show under some general assumptions that the probability that an ongoing program of screening will detect a cancer with a preclinical duration of k screens is

\[ 1 - \prod_{i=1}^{k} [1 - \xi(t_i)], \]  

where \( \xi(t_i) \) is the sensitivity function and the \( i \)th screen occurs at time \( t_i, i = 1, \ldots, k \), of the disease’s progression. The average program sensitivity is a weighted sum of individual versions of expression 1 over all cases in the program, but data from a screening trial are generally inadequate for specifying a distribution of preclinical durations over the population. However, expression 1 can be used to calculate program sensitivity for a typical or plausible sensitivity function.

Sensitivity Estimates in the Minnesota Study

Currently, no single optimal method exists for estimating the screen sensitivities of individual screens in a program, but program sensitivity can be estimated simply by the ratio of the number of screen-detected cancers to the total number of cancers diagnosed during periodic screening among all subjects complying with the program. The sensitivity estimates of 80.8% and 92.2% given by Mandel et al. (2) are such “crude” estimates of program sensitivity from the unrehydrated and rehydrated slide test results, respectively, during the first phase of the study; overall sensitivity was given as 89.3%. In the current report, crude program sensitivity estimates, defined as the proportion of all cancers diagnosed during periodic screening that were detected by a positive test, will be derived instead from the combined data collected during both phases of the study. This estimate will apply to the actual screens used, which were a mix of screens involving both rehydrated and unrehydrated test slides. For the Lang–Ransohoff model, a colorectal cancer diagnosed within one screening interval (1 year) of the patient’s previous screen, regardless of the test result, is considered to be diagnosed during periodic screening. A crude estimate of screen sensitivity can be defined as the number of cancers detected at a particular screen divided by the total number of cancers diagnosed in screened individuals within 1 year of their screen.

Follow-up-Adjusted Screen Sensitivity Estimate

The crude estimate of screen sensitivity assumes that the number of cancers clinically diagnosed in the screening interval is an estimate of the number of cancers missed by the screen. However, some cases detectable but missed at the time of a screen may not be clinically diagnosed within the following year due to the test. Thus, a crude estimator may overestimate sensitivity. On the other hand, cases not detectable at the time of the screen, but becoming detectable and causing symptoms within a year following the screen, will cause underestimation of sensitivity. We assume that cases subsequently discovered in those with false-positive screens (“false-positives”) represent incident cases developing after screening, and we adjust incidence among those with negative screens (“negatives”) for this new incidence. The follow-up-adjusted estimator \( \hat{\xi}_{fa} \) of screen sensitivity that adjusts in this way for the follow-up bias is derived in the “Appendix” section.

\[ \hat{\xi}_{fa} = \frac{d}{d + m(l_m - l_m)}, \]  

where \( d \) is the number of colorectal cancer cases detected and \( m \) is the number of subjects testing negative at the screen, and \( l_m \) and \( l_m \) are estimated 5-year cumulative incidence rates for negatives and false-positives, respectively, from the last screens of each phase in both the annual and the biennial arms of the Minnesota study.

Truncated Product-Binomial Model

A truncated product-binomial model (21) of the guaiac test, which takes advantage of the replications of the test, is applied to the Minnesota study results to estimate the screen sensitivity. The model assumes that for a cancer to be missed, for each of the three stools, either blood is present and both slides fail to react or blood is absent. The probability of detecting the cancer (the screen sensitivity) is 1 minus the probability of missing it. The truncated product-binomial estimate \( \hat{\xi}_{tpb} \) of screen sensitivity is computed from equation 3 by replacing parameters with estimates that maximize the likelihood based on data from the Minnesota study. Equation 3 and a more detailed derivation are given in the “Appendix” section.

Dual-Effect Model (Modified Lang–Ransohoff Model)

In estimating the theoretical effect of chance detection, the Lang–Ransohoff model (22) assigns those receiving colonoscopy a reduced colorectal cancer mortality rate for the following 5 (or 10) years, while the remaining subjects experience the same rate as the U.S. population. No sensitive screening effect from the fecal occult blood test is included. To determine how large a sensitivity would be needed to explain the full Minnesota study result in the presence of chance detection, we modified the Lang–Ransohoff model to include, in addition to chance colonoscopy, screen sensitivity of the fecal occult blood test, and we adjusted the sensitivity until the relative reduction in mortality equaled the 33.4% observed in the Minnesota study. We assumed the 5-year colorectal cancer mortality of a person whose colorectal cancer was sensitively detected was reduced by 51%, equal to the unadjusted ratio of the 5-year colorectal cancer mortality for the screen-detected cases to that for the non-screen-detected cases in the Minnesota study (1). As in the Lang–Ransohoff model, each time a screen occurs, reductions in the standardized colorectal cancer mortality rates are applied to the appropriate cases.
fraction of the surviving participants for the corresponding duration of the hypothesized effect. The details of the model are given in the “Appendix” section.

Positive Predictivity by Number of Positives

If the fecal occult blood test is insensitive to colorectal neoplasms, then the probability of cancer being discovered should be the same regardless of the test outcome; hence, the number of the six guaiac slides that are positive for the presence of blood in the sample will be independent of the presence or absence of a tumor and the positive predictivity of the test will be the same regardless of the number of positive slides. Although, for a sensitive test, the positive predictivity depends strongly on the specificity, for an insensitive test, the predictivity will be equal to the prevalence regardless of the specificity. If, on the other hand, the test is sensitive, then the probability of a positive slide from a subject with preclinical cancer should be greater than the probability of a positive slide from a person without cancer. The proportion of cases (the positive predictivity) for a given number of positive slides should increase as the number of positive slides increases. Although the specificity will determine the absolute level, predictivity should monotonically increase with the number of positive slides. For this not to occur, either a negative relationship between slide results needs to be introduced (e.g., one slide is more sensitive to the presence of blood than two slides) or dependence between individuals in a population needs to occur (e.g., person A being positive decreases the chances that person B is).

Since we expect cancers to bleed much more often than nonmalignant polyps, we examined the relationship between positive predictivity and number of slides separately for cancer and for polyps. The positive test results were grouped by the number of positive slides (e.g., results with only one of six slides showing positive results for the presence of blood), and the positive predictivity for colorectal cancer and for nonmalignant polyps was calculated as the fraction of screens with a given number of positive slides leading to detection of a lesion. For example, for all screens with exactly one slide positive, the number of cancers detected is divided by the total number of such screens. The association between the number of positive slides and the positive predictivity can be tested statistically by means of a two-sided rank-test for trend (23). A greater relative association (as measured by Spearman’s $\rho$) indicates greater relative sensitivity.

Comparative Incidence by Test Result

As indicated above, an insensitive test makes no distinction between those with and those without cancer, so the prevalence of cancers among those who test positive will be no greater than that among those who test negative. Therefore, the cumulative incidence rate among test-positives, after a sufficient period of time, should be no higher than that among test-negatives. In fact, if removal of polyps during a diagnostic workup is preventive, the rate among test-negatives will eventually exceed that among test-positives. Contrapositively, if the cumulative incidence rate among test-positives persists at a higher level than among test-negatives, the test must have been doing something to separate persons with cancer from those without cancer and so must be sensitive. We computed and compared the cumulative incidence in the positive and negative test groups as an indication of the fecal occult blood test sensitivity.

Relating Screen Sensitivity Functions to Program Sensitivity

Published research (24–28) suggests that it may take a colorectal cancer in the range of 5 years to progress from inception to clinical manifestation. In an annual screening program then, a tumor could undergo as many as five screens before surfacing clinically, and during that time a positive test could occur at any screen. Fig. 1, A, plots both a constant hypothetical screen sensitivity function, $\xi(t) = 0.26$, where $t$ is the time of the $i$th screen from inception of the cancer in years, and the program sensitivity computed at $t_i = 1, \ldots, 7$ from expression 1 versus the preclinical duration of the tumor. Fig. 1, B, does the same for an increasing (logistic) hypothetical screen sensitivity function,

$$\xi(t) = (1 + \exp(-3t + 4))^{-1}.$$  

Although hypothetical, the curves plausibly relate screen and program sensitivity to the natural history of colorectal cancer and qualitatively address constant and increasing sensitivity functions.

Results

Estimating Sensitivity in the Minnesota Study

The crude estimate of program sensitivity for colorectal cancer is 89.4%, based on 178 cases of colorectal cancer...
detected by screening divided by 199 cases discovered during screening through 1 year after each screening phase. Table 1 gives the crude estimates of individual screen sensitivities and standard errors.

Follow-up Adjustment

The 3483 false-positive and the 33 321 negative screens yield a screen sensitivity estimate adjusted for 5-year follow-up of 74.8%, with approximate 95% confidence limits of 65.2%–84.4%, which are consistent with the crude estimate of sensitivity from those combined screens of 84.1%.

Truncated Product-Binomial Model

The estimates of screen sensitivity from the truncated product-binomial model are also shown in Table 1. Although variable, most are above 90% (median, 95.7%), and even the lowest sensitivity of 67.8% is consistent with the crude estimate for the same screen.

Dual-Effect Model (Lang–Ransohoff Model) With the Use of Minnesota Study Data

The estimated sensitivity from the dual-effect model under various assumptions about the size and duration of the colonoscopy effect ranges from 94.1% to 96.2%, regardless of the assumptions (Table 1). Thus, the sensitivity in the presence of chance detection effects must be about 95% to account for the 33.4% overall reduction in colorectal cancer mortality.

Positive Predictivity

Fig. 2 shows positive predictivity for colorectal cancers and for polyps according to the number of slides per test that were positive for cancer and polyps. In Fig. 2, A, the predictivity for cancer reliably increases as the number of positive slides in the test increases (Spearman’s ρ = 0.94; P = .0167); in Fig. 2, B, although the power is low, the observed association between the predictivity for polyps and the number of positive slides is consistent with a low sensitivity (Spearman’s ρ = 0.77; P = .1028). These results are consistent with the fecal occult blood test being sensitive for colorectal cancer but relatively insensitive for non-malignant neoplasms.

Comparative Incidence by Test Results

Fig. 3 is a plot of the cumulative incidence rate over time for three subgroups of subjects defined by their participation in and results on the first screening test. Not only is the highest rate of cancer diagnosed among those whose tests were positive in the first year following the screen (“screen-positives”), but also this higher rate persists by about the same absolute amount over the 17 years of follow-up available. If the fecal occult blood test were relatively insensitive, the cumulative incidence among those whose tests were negative (“screen-negatives”) would eventually approximate that among screen-positives. The incidence among those who refused to be screened is intermediate between that among the screen-negative and screen-positive groups, but closer to that of the screen-negatives.

Simulation Reconciling Results and Conclusions of the Minnesota Study and of the Studies by Ahlquist et al.

The first study by Ahlquist et al. (3), in which patients whose colorectal cancer was recently resected were annually screened and examined structurally for colorectal cancer, yielded a sensitivity of 26%. To generate Fig. 1, A and B, the sensitivity estimated by the first study by Ahlquist et al. was used as a constant and as the initial value of an increasing sensitivity function, respectively, to compute program sensitivity. In panels A and B of Fig. 1, the program sensitivity at any screen other than the first is higher than the screen sensitivity and quickly approaches the 90% range. In panel B of

Table 1. Estimates of screen sensitivity in the Minnesota Colon Cancer Control Study by using crude estimates, truncated-binomial estimates, and dual-effect model estimates

<table>
<thead>
<tr>
<th>Screen No.</th>
<th>Crude estimates*</th>
<th>Truncated product-binomial estimates†</th>
<th>Dual-effect model estimates‡</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Phase I‡ sensitivity ± standard error</td>
<td>Phase II§ sensitivity ± standard error</td>
<td>Phase I‡ sensitivity</td>
</tr>
<tr>
<td>1</td>
<td>85.7% ± 6.6%</td>
<td>96.0% ± 3.9%</td>
<td>97.4%</td>
</tr>
<tr>
<td>2</td>
<td>86.7% ± 8.8%</td>
<td>94.4% ± 5.4%</td>
<td>95.0%</td>
</tr>
<tr>
<td>3</td>
<td>93.8% ± 6.1%</td>
<td>83.3% ± 7.6%</td>
<td>99.5%</td>
</tr>
<tr>
<td>4</td>
<td>100.0% ± 0.0%</td>
<td>92.9% ± 6.9%</td>
<td>97.1%</td>
</tr>
<tr>
<td>5</td>
<td>83.3% ± 8.8%</td>
<td>85.7% ± 13.2%</td>
<td>99.1%</td>
</tr>
<tr>
<td>6</td>
<td>80.0% ± 10.3%</td>
<td>N/A</td>
<td>93.9%</td>
</tr>
<tr>
<td>Phase average</td>
<td>89.6% ± 3.1%</td>
<td>89.3% ± 3.0%</td>
<td>N/A</td>
</tr>
<tr>
<td>Annual average</td>
<td>89.4% ± 2.2%</td>
<td>N/A</td>
<td></td>
</tr>
</tbody>
</table>

*Crude estimates of screen sensitivity based on the ratio of true-positives to the sum of true-positives and true-negatives and approximate standard errors based on the normal approximation to the binomial.

†Estimates of screen sensitivity for the fecal occult blood test, computed from the truncated product-binomial model. N/A = not available from the model.

‡Phase I, from December 1975 through February 1982, comprised five screens in the annually screened group.

§Phase II, from February 1986 through February 1992, comprised six screens in the annually screened group.

¶Estimates of screen sensitivity from the dual-effect model (modified Lang–Ransohoff model) with the use of various values for the duration and size of the colonoscopy effect. Values for positivity, duration of screening and follow-up, and compliance with screening and diagnosis are given in Ederer et al. (29).
Fig. 1, the program sensitivity is only slightly higher than the screen sensitivity.

**Discussion**

**Minnesota Study**

The crude overall positive predictivity for colorectal cancer was 2.8%. That most positive tests were from subjects without colorectal cancer is consistent with low specificity. Because the slides are meant to be sensitive to the presence of blood in the stool or any substance causing the same chemical reaction, many sources—e.g., consumed red meat, fresh fruit, or vegetables; diverticuli; hemorrhoids; menstruation; polyps or cancer—can cause a positive test (30). However, among those colorectal cancer cases discovered in the Minnesota study among subjects who had been offered and who complied with regular annual screening, 89.4% were found by means of a positive fecal occult blood test. Moreover, among those subjects with positive tests, 28.1% had at least one polyp found during consecutive examination.

**High Estimated Sensitivity of Fecal Occult Blood Tests**

Screening tests are designed to detect disease before it causes symptoms and motivates the patient to seek medical care but after it becomes diagnosable. An effective screening test will be sensitive to a tumor before its course becomes unalterable. Thus, sensitivity is a good, but an imperfect, indicator of potential effectiveness.

We have shown that the direct, crude estimates of individual screen sensitivities from the Minnesota study ranged from 80.0% to 100% and that the estimated program sensitivity was 89.4%. The estimate of sensitivity adjusted for over-ascertainment and under-ascertainment of cases was consistent with the crude estimate, and the truncated product-binomial estimates based on the pattern of positive tests were equally high for individual screen sensitivities. By modifying the Lang–Ransohoff model to incorporate sensitive detection explicitly, we obtained estimates of the screen sensitivity from 94% to 96%, whether we assumed the same 70% effectiveness of colonoscopy as Lang and Ransohoff (22) or a 60% effectiveness that is consistent with that

![Fig. 2. A) Positive predictivity for colorectal cancer of all positive sets, by the number of positive slide results per set. Sets of fewer than six slides are excluded, as are those with any slides of indeterminate outcome. B) Positive predictivity for polyps, including those that were adenomatous, hyperplastic, and undetermined, of all positive sets, by the number of positive slide results per set. The same sets are excluded as were excluded in A.](image)

![Fig. 3. Cumulative incidence of colorectal cancer (CRC) for subjects who had been randomly assigned to receive annual screening and offered an initial screen, plotted against the time since their random assignment by the results of the test. Values are Kaplan–Meier estimates at the end of each whole year. SE = standard error.](image)
given by Selby et al. (31). Since our model estimates sensitivity in the presence of chance finding to explain the Minnesota study results, it accounts for the high positivity rate of rehydrated fecal occult blood test slides. The consistency of the results with the use of four methods of estimation points to a sensitivity in the range of 85%–95%.

Most slides (83.3%) were rehydrated with deionized water during the Minnesota study to yield higher laboratory sensitivity to blood. This method is particularly prone to false-positivity (32) and contributed to the large number of colono-scopies.

Indirect Evidence

The markedly higher incidence among subjects who tested positive than among those who tested negative and the increasing positive predictivity associated with higher numbers of positive slides provided indirect evidence that the fecal occult blood test is sensitive to colorectal cancer. In particular, these results are inconsistent with an insensitive test.

Reconciling the Study by Ahlquist et al. With High Program Sensitivity

In one study, Ahlquist et al. (3) began screening for fecal occult blood and structural examination in 1217 colorectal cancer patients within the year of their colon resection; therefore, cancers available for screen detection during the initial screen were, on average, only 6 months into the detectable, preclinical phase, assuming roughly uniform incidence. Since the screening and examination were annual, the same is true of cancers available at subsequent screens. However, in general screening, some people will have tumors that have been detectable for more than one screening interval. It is probable that sensitivity increases as the tumor progresses and bleeds more (33). The test used by Ahlquist et al., which was similar (but not identical) to that used in the Minnesota study, had an estimated sensitivity of only 26% (95% confidence interval = 13%–39%), based on 29 recurrent and 18 new primary tumors in 46 patients.

In the other fecal occult blood test study conducted by Ahlquist et al. (3), in which 12312 relatives of the colorectal cancer patients were screened and a small, systematic sample of test-positives with-out cancer (300 of 1339 subjects) and test-negatives (600 of 10931 subjects) were interviewed annually by telephone for up to 3 years, only 74% of the screen-positives received a thorough bowel examination. Furthermore, it is unclear how cancers among the screen-positives without examinations were used in the calculations. We therefore focused on the first study.

A fairly simple model utilizing an increasing sensitivity function is consistent with low sensitivity in the first year of detectability, high average screen sensitivity, and high program sensitivity for the duration of the detectable, preclinical phase (Fig. 1, B). Thus, low sensitivity early in the detectable, preclinical phase was found in the first study conducted by Ahlquist et al. does not necessarily conflict with high overall screen or program sensitivity.

Consistency of Sensitivity Estimates

By estimating sensitivity for colorectal cancer in several ways, we obtained consistently high estimates for the test in the Minnesota study. The sensitivity of the fecal occult blood test varies, of course, with the type (e.g., rehydrated versus non-rehydrated slide tests), duration (e.g., 5 years versus 10 years of testing), and frequency (e.g., annual versus biennial) of screening. Other randomized trials of fecal occult blood screening have published estimates of sensitivity ranging from 48% to 83% (34–36), with the highest sensitivity associated with rehydration and consistent with the results presented here. Furthermore, in a recently published comparison of two guaiac tests and an immunochemical test (37), the most sensitive guaiac test had a positivity rate of 13.6%, comparable to the 9.8% for screens that used rehydrated slides in the Minnesota study, and a screen sensitivity of 79.4% (95% confidence interval = 64.8%–94.0%). Although based on only 34 cases, this result is consistent with the high sensitivity in our study.

Implications

Annual fecal occult blood testing (followed by a thorough diagnostic regimen for those responding positively and standard treatment for those diagnosed) has been shown to reduce mortality from colorectal cancer by a third through annual screening in asymptomatic individuals over the age of 50 years. How sensitive is the test? The methods used herein demonstrate that program sensitivity of the test that used rehydrated fecal occult blood test slides is in the range of 90% for a program that is 5 years or more in duration. The answer is important because the diagnostic workups associated with the high positivity rate of fecal occult blood test increase costs considerably (38–40).

As Winawer (41) pointed out, the specificity of the test should be improved. Recently, smaller (yet statistically significant) reductions in colorectal cancer mortality (15%–18%) have been obtained by screening biennially with more specific tests using unhydrated slides (42,43). Whether tests that use unhydrated slides can reduce colorectal cancer mortality as much as one third when annual screening is offered remains undetermined; as yet, no trials are studying such a program.

Appendix

Program Sensitivity and Sensitivity Functions

A tumor in the detectable, preclinical phase through several annual screening tests will be discovered by those tests with a probability that can exceed the maximum sensitivity of any of the individual tests. To illustrate, assume that the probability of a positive test depends only on the sensitivity function \(\xi(t_i)\), giving the sensitivity of the \(i\)th screen at time \(t_i\). Then, a subject whose detectable, preclinical phase spans \(k\) screens has program sensitivity given by expression 1. If \(\xi_m\) is the maximum value of \(\xi(t_i)\) over all \(t_i\), then, by virtue of the fact that \(1 - \xi \leq 1\),

\[
\xi_m = 1 - (1 - \xi_m) \prod_{i=1}^{k} [1 - \xi(t_i)] = expression 1.
\]

Of course, those case subjects exposed to only one screen at \(t_i\) of the preclinical phase will have program sensitivity \(\xi(t_i)\), and those exposed to no screens will have sensitivity 0. To calculate the overall program sensitivity requires integration of in-
individual versions of expression 1 over all cases in the program, weighted by their frequency in the population. Prorok (44) derived the estimated program sensitivity (‘‘proportion detected by the entire screening program’’) from a preclinical incidence that is constant over each interval, a constant screen sensitivity, and a distribution of preclinical durations. To extend those results to sensitivity functions was beyond the scope of this report and was not attempted. However, the above heuristic presentation suggests that, if the fraction of preclinical cases with very short preclinical durations is very small and if the majority of cases have preclinical durations long enough to span more than a single screening interval, expression 1 will dominate the sensitivity.

Follow-up-Adjusted Screen Sensitivity Estimate

The sensitivity of a particular screen could be estimated by the ratio of the number of cases discovered by the screen divided by the total cases available for detection, i.e., the sum of d and the number of cases missed by the screen u. Formally, \( \hat{\xi}^* = d/(d + u) \). Unfortunately, u is not observable. However, following a screen, the cumulative incidence rate of colorectal cancer, \( I_n \), among those testing negative derives from two kinds of cases: 1) those that were in the detectable, preclinical phase at the time of the screen (false-negatives), with cumulative incidence rate \( I_{na} \); and 2) those entering the detectable, preclinical phase after the screen, with \( I_{nr} \). Hence, \( I_n = I_{na} + I_{nr} \). In those subjects testing positive but with a negative diagnostic work-up (false-positives for cancer), cumulative incidence \( I_{fp} \) following the screen consists only of cases becoming detectable after the screen. This approach contrasts with the approach of Allison et al. (26), wherein cases discovered in the false-positive group are considered missed cancers and added to both the numerator and the denominator.

If we assume that the preventive effect of removing polyps does not lead to substantial changes in incidence for a specified period after screening (e.g., 5 years) and that those persons with false-positive tests (for cancer) do not have substantially greater propensity for developing cancers in that specified period, then the cumulative incidence of new cases over that period in those testing negative should be equal to the incidence of new cases in false-positives, i.e., \( I_{na} = I_{fp} \). The difference in incidence rates, \( I_n - I_{fp} = I_{na} + I_{nr} - I_{fp} = I_{nr} \) is the cumulative incidence rate of cancer detectable at the time of screening. The Kaplan–Meier estimates \( \hat{I}_n \) and \( \hat{I}_{fp} \) of the cumulative incidence rates among those testing negative and false-positive, respectively, can be used to estimate the number of cases missed by screening. If m is the number of persons who screened negative, then \( m(\hat{I}_n - \hat{I}_{fp}) \) estimates \( u \), the number of missed cases. We can substitute this estimate for \( u \) in the denominator of the definition of \( \hat{\xi}^* \) given above, yielding the estimate of sensitivity in equation 2. If the incidences \( \hat{I}_n \) and \( \hat{I}_{fp} \) in equation 2 are cumulated through enough follow-up time after screening so that all the false-negative cancers have been clinically diagnosed, the estimate will be unbiased (barring significant prevention and selection effects in that period).

Experience after the last screen in each phase of the study provides up to 5 years of follow-up uninterrupted by screening to which equation 2 can be applied. If the median preclinical duration for colorectal cancer is 5 years, it is reasonable to assume that most cancers detectable at screening will surface within 5 years and that the effect on incidence of polyp removal will not be observable in only 5 years. Since the interval incidence estimates may have large standard errors due to the small number of cases, the estimates of the annual incidence rates are smoothed before combining to calculate the estimate in equation 2.

Truncated Product-Binomial Model

A single screen using fecal occult blood testing consists of two samples each from three consecutive stools placed on six guaiac-impregnated paper slides, at least one of which must react positively for the screen to be positive. For a preclinical colorectal cancer case, there is some probability of the presence of blood in any particular stool, say \( \theta \). If the tumors randomly bleed into the lumen of the colon, then each stool is independent. In general, the preceding assumption applies not just to blood but to any guaiac-reactive substance, although in most cases this will be blood. Furthermore, there is some probability that a guaiac test will react positively to a sample from the stool with blood in it (call this probability \( p \)) and, if the samples are selected arbitrarily, the probabilities will be independent. In addition, assume that neither \( \theta \) nor \( p \) varies much from individual to individual.

The screen sensitivity is 1 minus the probability that a preclinical cancer is falsely negative; i.e., the cancer must be missed by the screen. Under the preceding assumptions, to miss a cancer, in each of the three stools either the tumor must bleed and both slides must fail to react [with probability \( \theta(1 - p)^2 \)], or it must fail to bleed (with probability \( 1 - \theta \)). Formally, the likelihood for a true-positive can be written

\[
\xi = 1 - [\theta(1 - p)^2 + (1 - \theta)]^3.
\]

By introducing a parameter for the probability of blood in the stool from a subject without cancer and a parameter for the prevalence of cancer, likelihoods can be generated for true-positives, false-positives, and all negatives, and estimates \( \beta \) and \( \delta \) can be computed that maximize the likelihood for the Minnesota study data and substituted for the parameters \( p \) and \( \theta \) in equation 3 to estimate the screen sensitivity (21). This procedure can be used to estimate the sensitivity at each screen by using data from those individuals participating in the screen.

Dual-Effect Model (Modified Lang–Ransohoff Model)

The Lang–Ransohoff model (22) assigns those who randomly undergo a colonoscopy a reduced colorectal cancer mortality rate over the next 5 or 10 years, while the remaining subjects experience the same rate as the U.S. population. It incorporates random noncompliance with screening and with colonoscopy for test-positives and total mortality taken from U.S. life tables (22). Let \( \rho_j \) be the underlying colorectal cancer mortality rate for year \( j = 1, \ldots, 13 \), and let \( e \) be the reduction in mortality due to random colonoscopy; for a particular screen, let \( \sigma \) be the fraction of eligible persons who comply with the screen, \( \gamma \) the probability of random selection for colonoscopy by the test (i.e., the positivity rate for the screen), and \( \delta \) the fraction of test-positives complying with the diagnostic workup. Then among the cohort surviving to the screen, in the fraction \( \sigma \gamma \delta \), the colorectal cancer
mortality rate for the next \( n = 5 \) (or 10) years following the workup is \((1 - e)p_n\), \(j = i, \ldots, i + n - 1\). After \( n \) years, the fraction is returned to the general cohort pool for further screening. The remaining \(1 - \sigma g \delta\) of the cohort experiences the underlying colorectal cancer mortality rate of \( \rho_i \). Each year the entire cohort is also depleted by the total mortality rate (minus the colorectal cancer mortality rate) for that year. To compute the total reduction in colorectal cancer mortality, for each year the colorectal cancer mortality is summed across strata defined by the year of the person’s last colonoscopy, including a stratum for those who never had a colonoscopy. These yearly mortality sums are summed across years and divided by the cumulative colorectal cancer mortality in an equivalent unscreened population.

In the models of Lang and Ransohoff (22) and Ederer et al. (29), this fraction is then divided by the reduction seen in the Minnesota study to get the (maximal) effect of chance selection for colonoscopy.

We modified the Lang–Ransohoff model by incorporating sensitive detection as well as chance to determine how large sensitivity must be in order to explain the full Minnesota study result. We included a parameter, \( \xi_{de} \), for the screen sensitivity of the fecal occult blood test and adjusted this parameter until the total reduction in mortality was equal to the 33.4% observed in the Minnesota study. A second parameter, \( \psi \) (the early detection effect), was included to represent a 51% reduction in mortality over a 5-year period, equal to the ratio, unadjusted for age and sex, of the 5-year risk of death from colorectal cancer for the screen-detected cases to that for the non-screen-detected cases in the Minnesota study (1). Thus, because of sensitive detection, the rate of dying from colorectal cancer for persons compliant with the screen was reduced for 5 years by the early detection effect \( \psi \) in the fraction of the population \( \xi_{de} \), so for the subset \( \sigma \xi_{de} \) the mortality rate was \((1 - \psi)p_i\). The reduction due to random colonoscopies applies only to the fraction of colorectal cancer cases not sensitively detected, \(1 - \xi_{de}\). Then, redefining \( \gamma \) to be the positivity rate, among the \( \sigma \cdot \gamma \cdot \delta \) test-positives who complied with their diagnostic workup, the mortality rate is \((1 - \xi_{de})(1 - e)p_{j\rho_i}\) for \( j = i, \ldots, i + n - 1\). As in the Lang–Ransohoff model, each time a screen occurs, reductions in the standardized colorectal cancer mortality rates are applied to the appropriate fraction of the surviving participants for the corresponding duration of the hypothesized effect. For a 10-year colonoscopy effect, the sensitive detection effect terminates after 5 years, and the colonoscopy effect remains for another 5 years.

Given values from the Minnesota study for \( \sigma, \delta, \) and \( \gamma \), (for the \( i \)th screen) and \( \psi \) and given \( e \) from the conclusions of Selby et al. (31), cumulative colorectal cancer mortality rates with and without screening and their ratio are computed as in the Lang–Ransohoff model. The value of \( \xi_{de} \) that yields a relative reduction in mortality equal to 0.334 is used to estimate the screen sensitivity of the fecal occult blood test in the Minnesota study.

### References


32. Wells H, Pagano J. Hemoccult test: reversal of


Notes

Supported by Public Health Service (PHS) research contracts NO1-CB-95613, NO1-CB-961005, and NO1-CB-53862 from the National Cancer Institute (NCI), National Institutes of Health, Department of Health and Human Services; by PHS grant RO1CA65728 from the National Cancer Institute (NCI); and under a cooperative agreement from the Centers for Disease Control and Prevention through the Association of Schools of Public Health.

We thank Drs. Christopher A. Lang and David F. Ransohoff for making available to us the spreadsheet of the original Lang–Ransohoff mathematical model. We also thank Mindy S. Geisser for her computational support and preparation of the graphs and tables and Steven J. Mongin for his helpful review and comments. Furthermore, the suggestions of anonymous reviewers helped to clarify and focus the presentation.

Manuscript received March 14, 1997; revised July 29, 1997; accepted July 31, 1997.

Concordance of Genetic Alterations in Poorly Differentiated Colorectal Neuroendocrine Carcinomas and Associated Adenocarcinomas

Alexander O. Vortmeyer, Irina A. Lubensky, Maria J. Merino, ChaoYu Wang, Tina Pham, Emma E. Furth, Zhengping Zhuang

Background: The histopathologic spectrum of colorectal neuroendocrine tumors ranges from benign to highly malignant. In this spectrum, poorly differentiated neuroendocrine carcinoma (PDNC) is the most aggressive type, characterized by early dissemination and a rapidly fatal course. Since it is unclear whether PDNC originates from neoplastic transformation of pre-existing neuroectodermal cells, pluripotent epithelial stem cells, or adenocarcinoma precursor cells, we investigated the histogenesis of this type of cancer by performing genetic analyses on a series of colorectal tumors. Methods: Archived histologic sections of colorectal PDNC from nine patients were analyzed; gastrointestinal carcinoid tumor specimens from four patients were used as controls. The specimens were deparaffinized, microdissected, and analyzed genetically. After DNA extraction, polymerase chain reaction amplification was performed to investigate alteration (i.e., loss of heterozygosity [LOH]) of the APC (adenomatous polyposis coli), DCC (deleted in colorectal carcinoma), and p53 (also known as TP53) genes. Results: LOH of the APC, DCC, or p53 genes was observed in six of eight informative PDNC tumors; no LOH was detected in the carcinoid control specimens. Four of five informative PDNC tumors had associated adenocarcinoma; LOH of the APC and p53 genes in these tumors involved the same allele in both tissue components. Four of the five tumors with associated adenocarcinoma showed LOH of the DCC gene; in three of these four tumors, the PDNC and adenomatous components showed LOH of the same allele. Conclusions: PDNC and associated adenocarcinoma appear to be derived from the same cell of origin, which is most likely either a pluripotent epithelial stem cell or an adenocarcinoma precursor cell. [J Natl Cancer Inst 1997;89:1448–53]

Colorectal neuroendocrine tumors comprise a wide histopathologic spectrum that ranges from benign to highly malignant (1). Within this spectrum, poorly differentiated neuroendocrine carcinoma (PDNC) clearly represents the most aggressive neoplasm and is characterized by early dissemination and a rapidly fatal course (2).

In 1949, “anaplastic polygonal-celled carcinoma” or “carcinoma simplex” was recognized by Dukes (3) as a neoplasm with a particularly poor prognosis compared with that for all other types of rectal adenocarcinomas. The characteristic histopathologic features of PDNC have subsequently been interpreted as “neuroendocrine differentiation” based on ultrastructural and immunohistochemical studies. Neurosecretory-type dense core granules may be demonstrated upon ultrastructural examination (4–6), and the tumors may show positive immunoreactivity with anti-neuron-specific enolase (NSE) (2,6,7) and anti-chromogranin A (7).

Although the cells of colorectal PDNC tend to be small and have hyperchromatic nuclei containing coarse, clumped chromatin (hence, commonly termed “small-cell carcinoma”), the presence of scattered bizarre mononucleated and multinucleated tumor cells with large hyperchromatic nuclei (4,5) or large vesicular nuclei (2) is not unusual. The term “small-cell carcinoma” reflects pheno...

*Affiliations of authors: A. O. Vortmeyer, I. A. Lubensky, M. J. Merino, C. Wang, T. Pham, Z. Zhuang, Laboratory of Pathology, Division of Clinical Sciences, National Cancer Institute, Bethesda, MD; E. E. Furth, Department of Pathology and Laboratory Medicine, Hospital of the University of Pennsylvania, Philadelphia.

Correspondence to: Zhengping Zhuang, M.D., Ph.D., National Institutes of Health, Bldg. 10, Rm. 2A33, Bethesda, MD 20892.

See “Note” following “References.”

© Oxford University Press