ELSEVIER

Critical Reviews in Oncology/Hematology 32 (1999) 155–165

Critical Reviews in
Oncology
Hematology

www.elsevier.com/locate/critrevonc

Cost-effectiveness of faecal occult blood screening for colorectal cancer: results of the Nottingham trial

David K. Whynes*

School of Economics, University of Nottingham, Nottingham NG7 2RD, UK

Accepted 21 October 1998

Contents

1. Introduction

Colorectal cancer is a major cause of mortality in most industrialised countries. In the USA, 6% of the population will develop the disease during their lifetimes and 40% of these will die as a result of it. Around 140 000 new cases are registered each year [1]. In the European Union during the 1980s, annual deaths from colorectal cancer amounted to around 85 000 [2]. Indeed, the disease is the second leading cause of cancer death in both Western Europe and North America [3]. As with many cancers, risk increases with age.

In the case of two other cancers, breast and cervix, programmes of mass population screening have been adopted as instruments of public health policy in many industrialised countries. In each of these cases, it would appear that the decision-makers concerned felt that the extra costs of screening were more than outweighed by the benefits of early detection and treatment of the disease. Now, equivalent arguments are emerging with respect to colorectal cancer.

1.1. The case for screening

The clinical case for the introduction of mass population screening for colorectal cancer rests fundamentally on the 'stage shifting' hypothesis:

1. Although debate persists as to the possibility of de novo carcinomas, it is believed that the preponderance of colorectal cancers will evolve through a

* Tel.: +44-115-951-5463; fax: +44-115-951-4159.
E-mail address: david.whynes@nottingham.ac.uk (D.K. Whynes)

1040-8428/99/$ - see front matter
PII: S1040-8428(99)00021-9

sequence of morphologic alterations in the mucosal structures of the large bowel, from adenoma through to a series of carcinoma stages [4–6]. A variety of staging typologies for carcinomas have been devised, the Dukes' four-stage classification of A (carcinoma confined within the bowel wall) through D (unresectable tumours and/or metastases in distant sites) remaining the most widely used in prognostication.
2. Cancer risk has been shown to be significantly reduced in patients whose adenomas have been excised by polypectomy [7,8]. Patients whose cancers are diagnosed and treated at earlier stages exhibit significantly superior survival [9].
3. The symptoms of colorectal cancer—pain, irregular bowel habits and the visible passing of blood—tend to become apparent only when the disease has reached the more advanced stages of tumour growth, and when the probability of spread to distant sites is increased. The early stages of cancer are commonly found in only 10–15% of patients with symptoms [10].
4. It accordingly follows that, in comparison with an unscreened population, the screening of asymptomatic individuals known to be at risk should result in the detection of abnormalities with a superior stage distribution. Put simply, a case which *would* have presented at a later stage will, as a result of screening, be detected and treated at an earlier stage, with corresponding improvements in prognosis.

Beyond the potential for achieving net survival gains for patients with the disease, screening may also impart wider psychological benefits to the community. Colorectal cancer remains a relatively rare disease in the population at large, with the result that the great majority of the subjects of a mass screening programme are likely to achieve negative test results. These results will offer reassurance, especially valued by those experiencing concern over their likelihood of contracting the disease [11].

Although it is available to individuals via private health care schemes in both North America and Europe, routine screening for colorectal cancer remains outside the public sector health care programmes in both these regions. This omission has essentially been due to the lack of scientific evidence pertaining to the efficacy of screening. Over the past 10 years, however, the results of major clinical trials have been published. These trials include those conducted in Minnesota, USA, [12], Gothenborg, Sweden [13], Funen, Denmark [10] and Nottingham, England [14]. The evidence from all of these trials confirms the view that mass population screening can contribute to a significant reduction in mortality from colorectal cancer. The Nottingham and Funen trials, which employ very similar protocols, have thus far detected mortality reductions of 15 and 18%, respectively [3]. Using a more sensitive test on a volunteer population, the Minnesota reduction was even higher [1].

1.2. Economic evaluation

Considerations of clinical efficacy alone are insufficient to justify the adoption of screening or, for that matter, any new medical technique. In a world where health care budgets are strictly finite, 'value for money' is equally important. If introduced as a public health initiative, a mass screening programme would require the mobilisation of financial and human resources which might otherwise have been expended on alternative health care interventions. To support the introduction of screening, it is therefore necessary to demonstrate that the expected costs of the screening programme can be justified in the light of the benefits or outcomes which the programme ultimately produces. Screening for colorectal cancer must be shown not simply to be clinically effective but also to be cost-effective.

Given that we inhabit a world where money matters, it is improbable that any major new cancer prevention strategy would be implemented anywhere in the absence of an assessment of economic viability. The research teams associated with all the European screening trials naturally have a concern for cost-effectiveness evaluations, although that being undertaken within the Nottingham trial has been particularly detailed. As will become clear, evaluation is not simply confined to the global level, i.e. assessing the cost-effectiveness of screening as opposed to not screening, but also at the specific level, involving a comparison of different modalities of screening. This is necessary, because there is no unique way of offering screening.

The principal result to be obtained from the economic evaluation of any screening modality is an estimate of its incremental cost-effectiveness ratio (ICER). In comparison with the status quo (no screening), we anticipate additional gains or benefits from a screening programme, although at the expense of additional costs. The ICER of the modality is defined as the extra costs incurred by employing the modality, expressed per unit of the extra benefits anticipated. By way of illustration, if we have to expend an extra £1 million to achieve an expected gain of 100 life-years in a cohort of subjects, the modality's ICER amounts to £10 000 per life-year gained. It is the ICER which is reported to the health care decision-maker, who then judges whether the public health care system is willing to pay *that* cost for the unit gain.

Even when derived from a well-designed observational study, an ICER is a point estimate, based on experimental data, possibly augmented by assumptions and modelling. It is accordingly necessary to provide some indication of the confidence which can be attached to the result. It has become conventional in economic evalua-

tions [15] to subject the estimated ICER to sensitivity analysis. In such analysis, relevant components of the calculation are changed by plausible amounts, or perhaps to 'worst cases', and corresponding new ICER values are estimated. Comparison between the original and new estimates demonstrate which particular parameters are critical to the results.

1.3. Screening modalities

Advocating the introduction of mass population screening pre-supposes the availability of a viable screening test. In this respect, a number of candidates have been proposed:

1. double-contrast barium enema (DCBE) X-ray, allowing radiological visualisation of the colon;
2. endoscopy, permitting direct visualisation of abnormalities within the range of the chosen instrument. This would be approximately 19 cm from the anal margin for the rigid proctosigmoidoscope, 30 or 60 cm for the flexible sigmoidoscope and, in 95% of cases, the entire colon for the colonoscope;
3. faecal occult blood (FOB) tests. Bowel tumours are known to bleed sporadically and to deposit blood in the stools. Although FOB tests vary in their chemical functioning, they all aim to detect the presence of such occult blood and are thus predictors of the presence of abnormalities.

Each of these individual testing methods has strengths and weakness. For example, colonoscopy offers high sensitivity and specificity, although at the prospect of considerable cost if used as the initial screen in a mass programme. Amongst the range of available FOB tests, increased sensitivity usually appears to entail decreased specificity, requiring a trade-off to be made between yield and the cost of managing false-positives. The more invasive procedures are likely to achieve poorer subject compliance, especially if they are employed on a recurrent basis. In consequence, cases have, in the past, been made for each of a wide variety of possible screening protocols, involving one or other of the possible tests, used alone or in combination with others, and offered over a range of time intervals [16]. Examples include annual FOB plus sigmoidoscopy [17], radiological investigation every 3–5 years [18] and 'once-only' colonoscopy [19] or sigmoidoscopy [20].

1.4. The Nottingham FOB trial

The pilot phase of the Nottingham trial ran over the period 1981–1983, whilst recruitment to the main study ran between 1985 and 1991. Working through general practices in the surrounding area, the investigators randomised approximately 153 000 people to equally-sized screening and control groups. The age range for eligibility for inclusion was 45–74 years in the pilot phase, and 50–74 years in the main study. Every 2 years since entry, members of the screening group have been sent the Haemoccult guaiac FOB test kit, accompanied by instructions and a letter of invitation. Compliers take two samples from each of three consecutive stools and return their completed tests for development. In an effort to minimise the false-positive rate, subjects with a positive test result are invited to complete a second test under a regime of dietary restriction. Those testing positive a second time are invited for investigation, those testing negative being invited to complete a third FOB test, with positives again proceeding to investigation. In the pilot phase, investigation was by means of DCBE and flexible sigmoidoscopy. Since 1985, colonoscopy has been the principal investigation method, with DCBE being used in the minority of cases where complete colonoscopic visualisation proves impossible.

All adenomas and carcinomas in screening subjects have been treated by surgical procedures, including polypectomy if feasible. Thereafter, patients have been transferred to follow-up programmes and exited from the trial.

At a median follow-up of 7.8 years, by which time subjects had been received between three and six invitations depending upon time of recruitment, overall subject compliance had reached around 57%. Over this period, 893 cancers had been diagnosed in the group offered screening. Of these 236 were screen-detected, whilst 400 presented in subjects who failed to respond to screening invitations. Interval cancers, those presenting between screening rounds, accounted for 249 cases, with 236 occurring after a negative FOB result. By the same endpoint, 856 cancers had presented in the control group. The proportion of stage A cancers was significantly higher (20 versus 11%), and the proportion of stage C and D cancers significantly lower (46 versus 52%), in the screening group compared with the controls. Nearly three times as many adenomas were detected in the screening group (1001 versus 370) and histological assessment revealed severe dysplasia in 97 of these cases. The number of verified deaths attributable to colorectal cancer was significantly lower in the group offered screening (360 versus 420), implying—as noted earlier—a 15% relative reduction in disease-induced mortality [14].

2. The economic structure of screening

Screening is a complex process, in which a number of technical parameters interact. To understand the nature of the interactions, we now consider a simplified model of mass population screening, based loosely on the first round of the Nottingham protocol. In this model, each subject will receive an FOB screening test kit (of unit

cost, S) through the post. Only a proportion, n, however, will be willing to complete the test. All such subjects recording negative test results exit the protocol, and all positives are investigated (at unit cost, I). For simplicity at this stage, we assume that the investigation is definitive.

We proceed to define three key parameters, disease prevalence (P), test sensitivity (X) and test specificity (Y). In a cohort of N subjects with n complying, the number of expected true positive results is $NnPX$, whilst the number of expected false positives is $Nn(1-P)(1-Y)$. It accordingly follows that the expected programme cost of offering screening to N subjects and consequently detecting cancers in those complying is the cost of the tests issued, plus the costs of investigating the positives, i.e.

$$NS + NnPXI + Nn(1-P)(1-Y)I \tag{1}$$

On detection, each true positive then proceeds to treatment and gains the expected net unit benefit of earlier detection (B). However, true positives are incurring treatment costs earlier that those presenting symptomatically, which we represent as a potential, additional unit cost, T. Confirmed false positives on investigation exit the protocol, incurring no further costs. Total programme plus treatment costs therefore become:

$$NS + NnPX(I+T) + Nn(1-P)(1-Y)I \tag{2}$$

The expected health gains for the cohort derive only from the true positives:

$$NnPXB \tag{3}$$

The ICER is the quotient of these two expressions, Eq. (2)/Eq. (3), which, after dividing throughout by N, is:

$$\frac{S + nPX(I+T) + n(1-P)(1-Y)}{nPXB} \tag{4}$$

For clarity, this can be written:

$$\frac{S}{nPXB} + \frac{(I+T)}{B} + \frac{I(1-P)(1-Y)}{PXB} \tag{5}$$

from which we deduce that the ICER will *fall*, i.e. cost-effectiveness will improve, if:

1. test costs (S), investigation costs (I) or net treatment costs (T) *fall*,
2. test sensitivity (X), test specificity (Y), prevalence (P), compliance (n) or the expected unit benefit from screening (B) *increase*.

In spite of representing a simplification of the screening process, Eq. (5) does serve to alert us to important implications for protocol design. For example:

(i) As the ICER clearly rises when prevalence (P) falls, it is likely to be economically unrealistic to screen populations with particularly low prevalences. In real-world programmes, the value of prevalence can, in effect, be adjusted by the appropriate selection of the target population. As both age and family history are known risk factors, for example, screening programmes which exclude both the young or those with no such history would typically record lower ICERs than programmes where such low-risk groups are not excluded.

(ii) Different FOB tests have different sensitivities. Whilst a protocol involving a test with a lower sensitivity would necessarily generate a smaller cancer yield in comparison with one with a higher test sensitivity, such a protocol is not necessarily the less cost-effective. From Eq. (5), lowering the sensitivity (X) lowers the denominator which, other things remaining equal, causes the ICER to rise overall. However, if, at the same time, the lower-sensitivity test were to be sufficiently cheap (i.e. lower S), it would be possible for the resulting lower numerator to more than compensate for the lower denominator. Put the other way around, it is possible for higher-yield, clinically-effective tests to represent relatively poorer value for money than cheaper, less effective tests, if the former are disproportionately expensive.

(iii) In addition to sensitivity differences, FOB tests also display specificity differences and, in general, sensitivity and specificity are inversely related [16]. Setting the test threshold to detect as many cancers as possible (high sensitivity) typically entails accepting a high rate of false positives (low specificity), whilst any attempt to reduce the numbers of false positives (high specificity) typically means that many more true positives are incorrectly classified as negatives (low sensitivity). In Eq. (5), sensitivity (X) and specificity (Y) operate in the same direction, i.e. more of either improves cost-effectiveness. As with the cost effect noted above, therefore, moving to a more sensitive FOB test will certainly lead to improvements in yield but, if the concomitant fall in test specificity is sufficiently large, the ICER overall could be inferior.

(iv) With respect to the ICER Eq. (5), the effect on variations in compliance (n) impact only on the first term involving unit test cost (S). It accordingly follows that, other things remaining equal, lower levels of compliance have a proportionally greater impact on raising the ICER when the unit test cost is high in relation to the other cost components in the equation. Put the other way round, the smaller the unit cost of the test, the less significant does the compliance rate in influencing the ICER become.

(v) Moving the model into a dynamic framework, both a high test sensitivity and high compliance in the initial screening round will contribute to a reduction in disease prevalence for any later rounds. In real-world programmes, intervals between screening rounds can be adjusted. Offering a repeat screen 'too soon' in such circumstances runs the risk of cost-ineffectiveness, as with (i) above, whilst 'too late' implies lower health gains than might otherwise have been obtainable.

The effects of parameter interactions, of the form outlined above, were a major area of investigation in the Nottingham evaluation, to which we now turn.

3. Programme costs

One of the early goals of the economics research programme was to model the likely resource costs of employing the Nottingham screening protocol in a real-world context [21]. Using contemporary trial evidence of compliance (58%), positive rate (1.3%) and cancer detection rate (0.2%), it was estimated that issuing one round of FOB tests to a target population of 75 000 subjects would yield 559 positive test results requiring further investigation, with 85 cancers subsequently being detected. This protocol would require the expenditure of around £45 000 on administrative and medical staff, £120 000 on FOB tests (including postage and development) and, based on earlier unit cost estimates [22], around £66 000 for investigations. The total cost of £231 000 (1989/1990 prices) is equivalent to around £2700 per cancer detected, or £5.30 per person screened.

For subsequent screening rounds, the compliance rate amongst those offered tests was observed to increase, although the impact on costs was largely offset by a falling positive rate for the test. Sensitivity analysis revealed that, for example, a 10% increase in each of the unit cost of the FOB test and investigations would raise all costs by 4 and 3%, respectively. Doubling the detection rate approximately halved the costs per cancer detected, whilst achieving 100% compliance reduced such costs by around 30%. Overall, the sensitivity analyses revealed that programme costs were far more likely to be sensitive to changes in clinical variables, especially detection and compliance rates, than they were to variations in the costs of resource inputs.

3.1. Alternative protocols

At one stage during the trial, a sub-study of screening using 6-day Haemoccult was undertaken, allowing the cost-effectiveness of this particular modality to be modelled. The experimental evidence showed the 6-day modality to be more successful in detecting cancers per se, but it did so at the expense of lower compliance, higher overall unit test costs and more follow-up investigations of positive test results. In comparison with 3-day testing, the cost of the extra cancers detected using the 6-day modality was estimated at £6484. The expected cost per person screened in the first round was £8.60, 62% higher than that under the 3-day modality [21].

Owing to dehydration during storage, faecal occult blood tests have been found to lose sensitivity. Accordingly, test rehydration prior to development has been advocated and was employed in the Gothenborg [23], but not in the Nottingham, trial. The above model was accordingly confronted with published results from both of these trials, to assess the cost-effectiveness of rehydration [24]. It was found that the higher sensitivity resulting from rehydration (Gothenborg's 85%, compared with Nottingham's 65%) is accompanied by losses in specificity (Nottingham's 99% against Gothenborg's 95%). By implication, approximately 30% more cancers can be detected via rehydration, although the cost of investigations (largely those of false positives) increases more than fourfold. More specifically, and in comparison with the Nottingham results above, rehydration would appear to double the cost per person screening, and increase the cost per cancer detected by around 50%.

Finally, the Eq. (1) model was employed to make a comparison of probable detection costs across a range of tests then available, using published data on test parameters [25]. In addition to Haemoccult, these included other FOB tests such as Hemeselect and Hemoquant. This analysis permitted the exclusion of a number of these alternatives on cost-effectiveness grounds: several of the then-existing tests were evidently inferior to their rivals, in that they appeared to detect the same number of cancers although at higher cost. Of the tests as sensitive, or less sensitive, than Haemoccult, none appeared to offer as low a an expected cost per cancer detected, when used in the appropriate protocol. Of the more sensitive tests, Hemeselect appeared to offer the best prospects of significantly superior yield at the lowest obtainable costs, although mean detection costs were evidently higher than for Haemoccult. This theoretical result was subsequently confirmed by trial evidence [26].

3.2. Compliance effects

Not surprisingly, compliance in the colorectal screening trials has been less than perfect, as is routinely the case in other, existing forms of cancer screening. In clinical terms, a screening compliance rate of 100% may be deemed optimal in that the number of abnormalities detected is thereby maximised. This proposition holds in cost-effectiveness terms also: the equations of the simple screening model presented earlier demonstrate that whilst compliance directly influences both the expected costs and the expected benefits, higher compliance necessarily lowers a given programmes ICER, other things remaining equal. It is therefore pertinent to ask: would it be economically justified to attempt to increase compliance, and thereby yield, by expending more resources on invitation? Using data from the European trials, the Nottingham researchers assessed the compliance and cost effects of utilising differing

methods of screening invitation, and proceeded to explore the incremental costs and yields associated with the compliance enhancement techniques currently being employed [27].

The cost and compliance observed from use of the Nottingham protocol (FOB test mailed from a central hospital) was compared with models of four earlier UK experiments in FOB test distribution, involving, for example, distribution by patients' general practitioners or by health visitors. It was found that, although it represented one of the more expensive methods of administration, direct FOB mailing would be likely to secure the highest compliance rate, resulting in a comparatively low figure for cost per test completed.

Whilst the Nottingham, Gothenborg and Funen trials all employed hospital-based postal invitation, each experimented with particular techniques aimed at enhancing compliance, that is, changing initial non-compliers into compliers. In Nottingham, reminder letters ere sent to non-compliers and these subjects' general practitioners were asked to prompt patients with a further reminder. In Gothenborg, both reminder letters and duplicate tests were posted. Two experiments were conducted in Funen, each involving combinations of additional reminder letters to, and direct telephone contacts with, non-compliers. The compliance increases resulting from each these methods were available from trail data which revealed, for example, that sending reminder letters in Nottingham caused compliance to rise by a further 10 percentage points, whilst prompting by general practitioners yielded a further 2.5% rise. Based on Nottingham cost structures, it was possible to model the cost implications of each of the methods. The study revealed that, as long as the monetary value placed on the gain of one lifeyear was greater than approximately £1000, each of the compliance enhancement methods employed thus far had been cost-effective. In particular, the Nottingham protocol with compliance enhancement was shown to be technically-efficient [28].

This having been said, by pooling the results across trial centres, it became clear that diminishing returns to compliance enhancement would be likely to prevail, as residual non-compliers would become increasingly difficult to persuade. A Nottingham study examined the two ends of the compliance continuum, by surveying the attitudes and characteristics of both persistent compliers and persistent non-compliers [29], in order to identify those characteristics most closely associated with persistent compliance behaviour. Persistent compliers were found, *inter alia*, to be of higher socio-economic classes than persistent non-compliers, to have more personal and family experiences of illness and to visit their dentists more regularly. The results suggest that generalised attempts at compliance enhancement might be less effectual against the prevailing background characteristics of the non-compliant population than the more overt targeting of efforts on particular subject groups.

4. Treatment costs

In assessing the overall cost consequences of implementing screening, the costs of the screening programme are only one element. As noted earlier, consideration must also be given to any additional costs of treatment. Throughout the Nottingham trial, treatment following diagnosis has been predominantly surgical. Accordingly, the Nottingham researchers initiated a study to compare the hospital costs of treating patients with colorectal cancers detected as a result of screening with those of patients whose cancers presented symptomatically [30]. Patient-specific cost estimates were made, using case records and hospital accounts, for 152 control group and 208 study group patients (those offered screening) over 3 years following the initial treatment episode. For the sample as a whole, the costs of treating cancers at stages A or D were significantly lower than those of treating cancers at stages B or C. In the case of the stage A cancers, the lower mean treatment costs were largely accounted for by the relatively higher usage of polypectomy, as opposed to resection, and fewer subsequent hospital re-admissions for complications and recurrence. In the stage D case, low hospital costs were largely accounted for by early patient death.

Although screen-detection appeared to offer treatment cost advantages, owing to superior staging, this effect in the study group overall was compensated for by the higher treatment costs incurred by the interval cancers and those presenting in non-compliers. The mean variable treatment costs for the control and study groups overall were £2966 and £3179, respectively (1990/1991 prices), with the difference between the means being insignificant. On the basis of trial evidence, it was therefore concluded that the introduction of mass screening was unlikely to give rise to substantial savings in the costs of surgical treatment.

4.1. Adenoma excision

In theory, the ostensible benefit of screening is not confined to the detection and treatment of cancers at an earlier stage than would be the case with symptomatic presentation. Screening should also be effective in detecting asymptomatic, pre-cancerous lesions. As noted in Section 1.1, there exists a widely-accepted hypothesis of a development sequence from benign polyp to carcinoma, implying that, were adenomas to be excised during colonoscopic investigation as a result of screening, potential cancers would be prevented from occurring.

Amongst the body of evidence favouring the sequence hypothesis is the correlation between adenoma preva-

lence and deaths from colorectal cancer in different populations [31] and those studies reporting reductions in cancer risks in cohorts where adenomas have been excised [32–34]. In a small number of studies [35–37], adenomas have remained both untreated and under observation. These studies typically report a cumulative risk of a cancer diagnosis at the index polyp, rising from time of diagnosis. This having been said, necropsy studies suggest that the prevalence of benign polyps, especially in elderly populations, is very many times higher than the prevalence of bowel cancer. It is accordingly seems likely that the vast majority of polyps do not, in fact, become cancerous and that, even if they do, most do not present a lifetime risk [38].

Throughout the Nottingham trial, adenomas have been detected and excised as a result of FOB screening and subsequent colonoscopic investigation. A smaller number have presented symptomatically in the control group, and these have also received treatment. Using the published estimates of cumulative risk of cancer in untreated adenoma sites, it was accordingly possible to estimate the expected number of cancers which would have developed over the longer term in both cohorts, had excision not taken place [39]. In detecting more adenomas, screening and subsequent excision obviously prevents more cancers than does symptomatic presentation. In consequence, screening economises on the costs of treating the cancers which would have developed, had the adenomas not been so excised.

The potential for cost savings was found to represent a sizeable discount on the overall costs of mass population screening for colorectal cancer. More specifically, based on an estimate of the costs of screening the trial population, expected cost savings as a result of not having to treat cancers which would have arisen in adenomas, had they not been detected by screening and subsequently excised, amounted to approximately 20% of total screening costs to date.

Although not formally measured in the above study, cancer prevention would be expected to yield survival gains in addition to costs savings. The adenoma detection capability therefore substantially improves the economic case for screening or, put the other way round, the expected ICER for FOB screening would be significantly higher were the benefits of adenoma detection to be ignored. This point is especially relevant when considering the age at which screening should commence. For a cohort of individuals aged 40–49, for example, the very low prevalence of cancer implies a very high ICER for screening, if cancer detection alone is the programme's objective. However, members of this age group face, in comparison with those aged, for example, 75–85 years, a far higher cumulative risk of carcinoma development in detected polyps, owing to considerably longer life expectancy. Whilst screening only the elderly means more cancers detected per capita, screening the less-elderly means proportionately more cancers prevented per capita [40].

5. Outcomes

It was always the intention of the Nottingham researchers to estimate the ICER of colorectal cancer screening as the incremental cost per expected quality-adjusted-life-Year (QALY) gained. The QALY is a composite measure of outcome, in that it weights expected survival gains (lifeyears) by a quality of life adjustment coefficient [41]. Although not free from controversy, the QALY is being increasing employed both by health economists conducting evaluations and by health authorities in planning services. The case for using the QALY as the outcome measure in this particular evaluation was especially strong, in view of the fact that it had been employed in the evaluation of breast cancer screening in the UK [42], the so-called *Forrest Report*. That evaluation calculated that the resource cost of screening per QALY gained lay in the range £3–5000, approximately (1984 prices). Following the publication of that report, a breast cancer screening programme was implemented in the UK. Accordingly, this ICER estimate would appear to be a not un-natural comparator for that of the Nottingham colorectal trial.

The results for the lifeyear component of the QALY have emerged naturally from the clinical trial, relative survival being one of its primary endpoints. Research on quality of life proceeded independently, initially with an exploration of appropriate measurement instruments [43]. Thereafter, a total of 418 survivors of the trial's test and control groups, and 33 randomly-selected cancer patients, independently completed two quality of life questionnaires, the Nottingham Health Profile [44] and the Health Measurement Questionnaire [45]. From the responses to the latter instrument, subjects may be allocated to one of 29 health states, defined by degrees of disability and distress [46]. Each of these states may, in turn, be represented as a numerical value on a zero-one scale, with unity representing no observed or perceived disability/distress.

The mode of entry to diagnosis and treatment (screening versus no screening) appeared to exert no major impact on post-intervention quality of life, as measured by both of the instruments [47]. Moreover, the stage of cancer progression was not significantly related to outcome life quality. A quality of life coefficient for surviving patients based on the Health Measurement Questionnaire was estimated to lie within the range 0.95–0.98, and this figure accords well with the estimates of other studies of interventions in populations of similar age. Overall, therefore, there are no grounds for believing that faecal occult blood screening for colorectal cancer per se significantly influences patients' post-intervention quality of life.

An assessment of 53 patients pre- and post-intervention showed that the high levels of psychological distress experienced prior to treatment were largely dissipated by three months. A logistic regression model showed that certain pre-operative symptoms, such as rectal bleeding and heartburn were more common in patients with late-, as opposed to early-, stage cancer [48].

In restricting the assessed benefit of the FOB screening programme to net QALY gains, the economic evaluation has necessarily excluded any consideration of the social and psychological impacts of the screening programme on the vast majority of subjects found not to have abnormalities. Examples include the potential value of reassurance, anxiety, process disutility (distaste at undergoing the test), plus any consequences for subsequent health-related behaviour. Whilst the potential importance of these impacts of screening programmes can be sensed intuitively, it has so far proved very difficult both to measure such health consequences formally and to include them in economic evaluations. For example, the heightened anxiety of those registering false positive results in other screening programmes has already been documented [49], yet the possibility of a counter-balancing effect of anxiety reduction in true negatives appears to have been neglected. Indeed, there is speculation that the latter subjects may experience a 'certificate of health effect'. As a result of being shown to be cancer-free, subjects may engage in activities more risky than if they were ignorant of their status and thus render themselves more susceptible to the disease which the test has shown they currently do not have!

Whilst excluded from the evaluation, social and behavioural considerations may, of course, appear as qualitative arguments in any future implementation decisions. It is also possible that emerging valuation techniques, such as estimating subjects' willingness-to-pay for screening, might, in the future, allow such considerations to be considered more formally [50]. Indeed, a willingness-to-pay study is presently being conducted in Nottingham.

6. The cost-effectiveness of FOB screening

The trial's clinical data, and the economic data obtained from the above studies, were combined to construct a mathematical model of the screening process [51]. This model comprised three sectors and used a semi-Markov framework. In the conventional Markov model, the disease process is interpreted as a sequence of possible health states, with subjects moving from one state to the next according to defined transition probabilities. With the semi-Markov formulation, the probabilities depend, in part, on how long the patient has been in the current state.

The first sector began by calculating the expected number of cancers occurring in a reference (unscreened) population with a pre-specified age–sex distribution and thereafter generated the expected number of years lost from normal life expectancy as a result of the disease. Following the imposition of the screening modality, estimates were then made of the reduction in annual incidence occasioned by screening. Such a reduction comes about as a result of both cancer detection at earlier stages and pre-emptive adenoma excision. Expected lifeyears lost under the screening modality were calculated and, from the combined results, year-by-year survival gains from screening were estimated. These data were augmented with the estimated quality of life coefficients, and allowed us to represent expected screening yield in terms of QALYs gained.

Disease incidence by stage thereafter determines expected treatment costs over time. The first sector also determined the number of cancers and adenomas detected at each screening round and the number of subjects passing through to subsequent rounds. These first sector outputs of the model drove the remaining two sectors. The second calculated the year-by-year costs of screening, investigations, polypectomies and adenoma surveillance, whilst the third estimated the net cost of treating latent cancers, those detected by screening but which would not have presented symptomatically during the expected lifetime of the patient.

Being consistent with observed clinical and economic data, the model, at the simplest level, permitted us to derive an ICER for the Nottingham protocol to date. However, the flexibility allowed by the mathematical approach allowed to simulate likely outcomes beyond the fixed parameters of the trial and to examine the sensitivity of results to variations in such parameters.

At a median follow-up of 8 years, the Nottingham FOB trial, employing the protocol described in Section 1.4 above, recorded an ICER of approximately £5700 per QALY gained for males, and approximately £5000 for females (1995–1996 prices) [52]. However, these estimates are, arguably, unrealistic as a representation of an implemented programme as they are truncated. At 8 years, many of the subjects recruited would remain within the eligible screening age range, implying that further opportunities for detection in this group would exist. Put the other way around, analysis at 8 years means that costs and outcomes occurring beyond this time are being disregarded.

The model enabled us to simulate trial costs and outcomes for many years further, up until the time when all subjects initially invited could be assumed to have died from colorectal cancer or from other causes. Assessed in the longer term, the estimated ICER reduces to around £2000 for males and £1400 for females. These estimates are evidently very much lower than

those for the truncated data, although such a result is to be expected. Repeated testing over a longer time period increases the probability of a given abnormality being detected and treated. Moreover, costs accrue in the short term whilst survival gains accumulate over a far longer period. The analysis of truncated results naturally understates expected benefit relative to expected costs.

Simulations under varying assumptions reveals that these ICERs are relatively insensitive to quite sizeable changes in certain of the key screening parameters, for example, the age at which screening should start. The ICER estimates vary only within a range of around 20% for starting ages from 40 to 60 years. This insensi tivity can be attributed to compensating effects. At lower starting ages, the net costs of the programme are higher as more persons have to be tested and investigated. Although lowering the starting age does not increase cancer yield to any great extent, the preventive potential of adenoma detection becomes considerable. As the starting age is increased, programme costs fall and the average cancer yield rises, but the potential for effecting survival gains by early detection declines. The impact of variations in compliance on the ICERs also appears relatively modest, in part, because more compliers mean greater yields but also greater investigation and treatment costs. The sunk costs of FOB tests being issued but not used is relatively insubstantial in the total programme costs, owing to the low unit cost of the test.

At the opposite extreme in terms of impact, a lower test specificity would greatly increase the number of false positives requiring unnecessary investigations and would thereby raise programme costs substantially. Indeed, simulation reveals that a fall in test specificity of only 10% would be sufficient to approximately double the estimated ICERs. Given that the evaluation is comparing short-term costs against longer-term benefits, it was not surprising to discover that the ICERs vary considerably with the choice of the discount rate. For example, halving the assumed discount rate on benefits, from 6 to 3%, improves both of the ICERs by around 50%.

7. Conclusions

Possibly the most informative way of summarising the results of the economic evaluation of colorectal cancer screening using the Nottingham protocol is to note that, under plausible assumptions, the programme would offer the prospect of ICERs superior to those identified for breast cancer screening in the UK [42,53,54]. Interestingly enough, the economic evaluation of the Danish FOB trial has produced the equivalent conclusion for that country also, whilst employing a slightly different economic evaluation methodology [55]. Nottingham's evaluation model has been based on an epidemiological description of the disease process, with trial results and observations supplying the necessary parameter values, such as test characteristics. In the Danish evaluation, however, trial prevalence and incidence data have been used to estimate the test parameters statistically.

Economic evaluation results inform, rather than make, decisions. Even so, we might expect that, because breast cancer has been recently implemented in the UK on the basis of *its* evaluation results, FOB screening for colorectal cancer would quickly follow, in view of the superior prospective cost-effectiveness findings. Policy makers might well prefer less haste, however, which might be justified or explained in a variety of ways. Perhaps the most obvious would be an implicit belief that earlier implementation decision had been misguided [56].

Leaving the possibility of such second thoughts aside, it is evident that, even if an intervention is accepted as being cost-effective, it might not necessarily be seen as being affordable. The UK population aged between 45 and 74 years—those who would be the likely targets for FOB screening—amounts to around 18 million persons. Based on the 1989–1990 cost estimates presented in Section 3 above, and allowing for subsequent inflation of 30%, the offer of FOB screening to one half of these each year would entail annual programme costs approaching £40 million in current prices. This figure ignores the extra costs of treatment and the very substantial, and immediate, investment necessary to establish the national screening infrastructure. This investment would include the provision of additional clinic capacity for investigations, manpower training and the development of call-recall systems for the invitation of subjects. The UK's public health service is cash-limited and, given existing health care commitments and government priorities, the resources required by the programme simply might not be available in the short term.

Second, it might be held that, despite the volume of research to date, the information remains incomplete. As noted earlier, the psychological costs of FOB screening have yet to be incorporated into the formal evaluation and some have argued that implementation decisions must necessarily await such evidence [49]. Again, it might be anticipated that technical change and further research could eventually reveal a screening modality even more cost-effective that the one already identified. Indeed, a new trial has recently been initiated in the UK, examining the cost-effectiveness of offering screening by flexible sigmoidoscopy on a once-only basis [20]. The ICER from this trial will be directly comparable with that already obtained from the Nottingham FOB study, allowing us to rank the modalities in terms of cost-effectiveness.

8. Reviewers

This paper was reviewed by Dr Jacqueline Brown, Health Economics Research Group, Brunel University, Uxbridge, Middlesex UB8 3PH, UK and Dr Mark Schulpher, Centre for Health Economics, University of York, Heslington, York Y01 5DD, UK.

References

[1] Lieberman GP. Mass screening: North American perspective. In: Young GP, Rozen P, Levin B, editors. Prevention and early detection of colorectal cancer. London: WB Saunders Company Ltd, 1996:289–300.

[2] Jensen OM, Esteve J, Moller H, et al. Cancer in the European Community and its member states. Eur J Cancer 1990;26:1167–256.

[3] Lieberman D, Sleisenger MH. Is it time to recommend screening for colorectal cancer? Lancet 1996;348:1463–4.

[4] Morson BC. The polyp-cancer sequence in the large bowel. Proc R Soc Med 1974;67:451–7.

[5] Morson BC. The evolution of colorectal carcinoma. Clin Radiol 1984;35:425–31.

[6] Lotfi AM, Spencer RJ, Ilstrup DM, Melton LJ. Colorectal polyps and the risk of subsequent carcinomas. Mayo Clinic Proc 1986;61:337–43.

[7] Atkin WS, Morson BC, Cusick J. Long-term risk of colorectal cancer after excision of rectosigmoid ademomas. N Engl J Med 1992;326:658–62.

[8] Winawer SJ, Zauber AG, Ho MN, et al. Prevention of colorectal cancer by colonoscopic polypectomy. N Engl J Med 1993;329:1977–81.

[9] Jatzko G, Lisborg P, Wette V. Improving survival rates for patients with colorectal cancer. Br J Surg 1992;79:588–91.

[10] Kronborg O, Fenger C, Olsen J, Jorgensen OD, Sondergaard O. Randomised study of screening for colorectal cancer with faecal-occult-blood test. Lancet 1996;348:1467–71.

[11] Jansen JH. Participation in the first and second round of mass-screening for colorectal cancer. Soc Sci Med 1984;18:633–6.

[12] Mandel JS, Bond JH, Church TR, et al. Reducing mortality from colorectal cancer by screening for fecal occult blood. N Engl J Med 1993;328:1365–71.

[13] Kewenter J, Brevinge H, Engaras B, Haglind E, Ahren C. Results from screening, rescreening and follow-up in the prospective randomised study for detection of colorectal cancer by fecal occult blood testing: results for 86,308 subjects. Scand J Gastroenterol 1994;29:468–73.

[14] Hardcastle JD, Chamberlain JO, Robinson MHE, et al. Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. Lancet 1996;348:1472–7.

[15] Drummond MF, Brandt A, Luce B, Rovira J. Standardising methodologies for economic evaluation in health care. Int J Technol Assess Health Care 1993;9:26–36.

[16] Young GP, Macrae FA, St John DJB. Clinical methods for early detection: basis, use and evaluation. In: Young GP, Levin B, editors. Prevention and early detection of colorectal cancer. London: WB Saunders Company Ltd, 1996:242–70.

[17] Winawer SJ, Flehinger BJ, Schottenfeld D, Miller DG. Screening for colorectal cancer with fecal occult blodd testing and sigmoidoscopy. J Natl Cancer Inst 1993;85:1311–8.

[18] Eddy D. Screening for colorectal cancer. Ann Intern Med 1990;113:373–84.

[19] Ransohoff DF, Lang CA. Screening for colorectal cancer. N Engl J Med 1991;325:37–41.

[20] Atkin WS, Cuzick J, Northover JMA, Whynes DK. Prevention of colorectal cancer by once-only sigmoidoscopy. Lancet 1993;341:736–40.

[21] Walker AR, Whynes DK, Chamberlain JO, Hardcastle JD. The cost of screening for colorectal cancer. J Epidemiol Community Health 1991;45:220–4.

[22] Walker AR, Whynes DK, Hardcastle JD, Chamberlain JO. The hospital costs of diagnostic procedures for colorectal cancer. J Clin Epidemiol 1991;44:907–14.

[23] Kewenter J, Bjork S, Haglind E, Smith L, Svanvik J, Ahren C. Screening and rescreening for colorectal cancer: a controlled trial of feacal occult blood testing in 27,700 subjects. Cancer 1988;62:645–51.

[24] Walker AR, Whynes DK, Hardcastle JD. Rehydration of guaiac-based faecal occult blood tests in mass screening for colorectal cancer: an economic perspective. Scand J Gastroenterol 1991;26:215–8.

[25] Walker AR, Whynes DK, Hardcastle JD. Filtering strategies in mass population screening for colorectal cancer: an economic evaluation. Med Decis Making 1992;12:2–7.

[26] Robinson MHE, Marks CG, Farrands PA, Whynes DK, Bostock K, Hardcastle JD. Is an immunological faecal occult blood test better than Haemoccult?—a cost benefit study. Eur J Surg Oncol 1995;21:261–4.

[27] Walker AR, Whynes DK. Participation and screening programmes for colorectal cancer: more would be better? J Health Econ 1991;10:207–25.

[28] Whynes DK, Walker AR, Hardcastle JD. Cost-effective screening strategies for colorectal cancer. J Public Health Med 1992;14:43–9.

[29] Neilson AR, Whynes DK. Determinants of persistent compliance with screening for colorectal cancer. Soc Sci Med 1995;41:365–74.

[30] Whynes DK, Walker AR, Chamberlain JO, Hardcastle JD. Screening and the costs of treating colorectal cancer. Br J Cancer 1993;68:965–8.

[31] Jass JR. Do all colorectal carcinomas arise in preexisting adenomas? World J Surg 1989;13:45–51.

[32] Gilbertson VA. Proctosigmoidoscopy and polypectomy in reducing the incidence of rectal cancer. Cancer 1974;34:936–9.

[33] Kronborg O, Fenger C. Prognostic evaluation of planned follow-up in patients with colorectal adenomas. Int J Colorectal Dis 1987;2:203–7.

[34] Murakami R, Tsukuma H, Kanamori K, et al. Natural history of colorectal polyps and the effect of polypectomy on occurrence of subsequent cancer. Int J Cancer 1990;46:159–64.

[35] Eide TJ. Risk of colorectal cancer in adenoma bearing individuals within a defined populations. Int J Cancer 1986;38:173–6.

[36] Hoff G, Foerster A, Vatn MH, Sauar J, Larsen S. Epidemiology of polyps in the rectum and colon: recovery and evaluation of unresected polyps 2 years after detection. Scand J Gastroenterol 1986;21:853–62.

[37] Stryker SJ, Wolff BG, Culp CE, Libbe SD, Ilstrup DM, MacCarty R. Natural history of untreated colonic polyps. Gastroenterology 1987;93:1009–13.

[38] Pollock AM, Quirke P. Adenoma screening and colorectal cancer. Br Med J 1991;303:3–4.

[39] Whynes DK, Walker AR, Hardcastle JD. Cost savings in mass population screening for colorectal cancer resulting from the early detection and excision of adenomas. Health Econ 1992;1:53–60.

[40] Whynes DK, Walker AR, Hardcastle JD. Effect of subject age on costs of screening for colorecal cancer. J Epidemiol Community Health 1992;46:577–81.

[41] Williams A. Economics of coronary artery bypass grafting. Br Med J 1985;291:326–9.

[42] DHSS Working Group (Chairman-Sir Patrick Forrest). Breast cancer screening: report to the health ministers of England, Wales, Scotland and Northern Ireland. London: HMSO, 1986.
[43] Whynes DK, Neilson AR. Convergent validity of two measures of the quality of life. Health Econ 1993;2:229–35.
[44] Hunt SM, McKenna SP. The Nottingham Health Profile User's Manual. Manchester: Galen Research and Consultancy, 1991.
[45] Kind P, Gudex C. The HMQ-measuring health status in the community. Discussion Paper 93. York: Centre for Health Economics, University of York, 1991.
[46] Rosser RM. A health index and output measure. In: Walker SR, Rosser RM, editors. Quality of life assessment: key issues in the 1990s. Dordrecht: Kluwer Academic Publishers, 1993:151–78.
[47] Whynes DK, Neilson AR, Robinson M, Hardcastle JD. Colorectal cancer screening and quality of life. Qual Life Res 1994;3:191–8.
[48] Whynes DK, Neilson AR. Symptoms before and after surgery for colorectal cancer. Qual Life Res 1997;6:61–6.
[49] Stewart-Brown S, Farmer A. Screening could seriously damage your health. Br Med J 1997;314:533.
[50] Johansson P-O. Evaluating health risks: an economic approach. Cambridge: Cambridge University Press, 1995.
[51] Neilson AR, Whynes DK. Cost-effectiveness of screening for colorectal cancer: a simulation model. IMA J Math Appl Med Biol 1995;12:355–67.
[52] Whynes DK, Neilson AR, Walker AR, Hardcastle JD. Faecal occult blood screening for colorectal cancer: is it cost-effective? Health Econ 1998;7:21–9.
[53] Department of Health. Breast cancer screening: evidence and experience since the Forrest Report. Sheffield: NHS Breast Cancer Screening Programme, 1991.
[54] Wolstenholme JL, Smith SJ, Whynes DK. The costs of treating breast cancer in the United Kingdom: implications for screening. Int J Technol Assess Health Care 1998;14:277–89.
[55] Gyrd-Hansen D, Sogaard J, Kronborg O. Colorectal cancer screening: eficiency and effectiveness. Health Econ 1998;7:9–20.
[56] Rodgers A. The UK breast cancer screening programme: an expensive mistake. J Public Health Med 1990;12:197–204.

Biography

David K. Whynes holds degrees from the University of York and St Andrews, UK, and is currently Professor of Health Economics at the University of Nottingham, UK. In addition to directing the economic evaluation of the Nottingham FOB trial, he works on other aspects of both cancer and screening, for example, a prospective evaluation of cervical screening and cancer treatment costing studies. In addition he has, for some years, been engaged in a long-term analysis of the economics of UK primary health care.