Screening

P. Boyle

Screening. P. Boyle. ©ERS Journals Ltd 2003.

ABSTRACT: Screening for cancer is currently entering a very promising period as new technologies become available for assessment. However, this great opportunity also presents significant problems, starting with how do you select the likeliest method for evaluation? How do you do trials where the technology will be out-moded by the time the result is known? Indeed, how can all promising technologies be evaluated?

There is a need for a renewal of methodological thinking on this topic at present. Randomised trials, which currently rely on death as the primary outcome, are very long and very complicated to undertake and run. There is an urgent need for biological markers, which are true surrogate end-points, to help reduce sample size and make trials more feasible, so that more trials can be done.

For those screening modalities of demonstrated effectiveness, the key issue at present is how to maximise the outcome in population screening when the technology is transferred from the research setting to the public health setting. It is essential to remember that screening, per se, has never prevented a cancer nor prevented a cancer death. It is impossible to separate the issues of screening with those of adequate treatment: the best results in a screening programme will come when treatment is excellent.

Screening is the first, and sometimes key, stage in a global management programme for the cancer in question. As the aim at present is to strive to maximise the effectiveness of population-based programmes, this aspect should be foremost in all thought processes.
Eur Respir J 2003; 21: Suppl. 39, 3s–15s.

Correspondence: P. Boyle
Division of Epidemiology and
Biostatistics
European Institute of Oncology
via Ripamonti 435
20141 Milan
Italy

Keywords: Early detection
lung cancer screening
preventative

Received: September 30 2002

Accepted after revision: November 12 2002

Many of the ideas presented in this manuscript have been published previously and elsewhere by the author. Thus, the author does not claim this as an original work but merely as a compilation of ideas brought together for educational purposes.

The large majority of human cancers diagnosed each year may be avoidable, but avoidable causes of many common cancers have not yet been clearly identified. A prerequisite of cancer prevention lies in identifying the determinants of cancer risk. Cancer Control embraces a number of important elements with the aim of reducing the incidence of cancer and, failing primary prevention, reducing mortality either by finding disease at an earlier and more "curable" stage or by improving the survival stage-for-stage through improvements in therapy. There are a number of disciplines involved within this embrace, including epidemiology, clinical science, behavioural science and health education. It is a complex and at times uncoordinated package.

Screening for cancer involves the examination of, but more frequently the performance of, a test on asymptomatic individuals in order to classify them as likely or unlikely to have the disease in question. Those who appear likely to have the disease are investigated further to arrive at a final diagnosis and those who are found to have the disease are treated. The organised application of early diagnosis and treatment activities in large groups is often described as population screening. The goal of screening generally is to reduce mortality from the disease among the people screened *via* early treatment of the cases discovered. Screening calls attention to the likelihood of disease before symptoms appear.

"Screening" in connection with early diagnosis and treatment should be clearly distinguished from other uses of the term in epidemiology and clinical practice. In particular, screening is commonly used to describe a series of tests done on a symptomatic patient for whom a diagnosis is not yet established. This type of screening is part of the practice of clinical medicine rather than public health or preventive medicine. Screening procedures may also be used to estimate the prevalence of various conditions without immediate disease-control objectives. Screening could be used both to refer to the identification of people at high risk of a disease but who do not yet have it, and to the application of tests to this group for the early detection of disease. It is clearly important to know exactly what is meant and what is not meant by screening.

It is also important to distinguish between diagnostic tests and screening tests. MORRISON [1] illustrates this using the example of diabetes. The use of a glucose tolerance test is considered to be a diagnostic test, while use of the (random) blood sugar method may be considered a screening test. Thus, a liver biopsy would be considered a diagnostic test for liver cancer and a biopsy of the prostate a diagnostic test for prostate cancer. It is also essential to distinguish between what is meant by screening and "case-finding". Screening is aimed at the general population and not merely those who have sought some medical attention.

Principles of screening trials

There are a number of important considerations that should be taken into account before even undertaking a screening trial (for cancer). The increased demand for services created by the screening programme should not be underestimated.

Subjects with positive results from screening tests will need rapid further assessment, since their anxiety will be increased by any recall. The decision of whether they have cancer, or not, should be arrived at as soon as possible. Most of the workload from a screening programme is likely to come from subjects who are positive at the screening test but who do not have cancer (the false positives). The proportion of those screened without cancer who are positive at the screening test will be determined by the specificity of the screening test used.

Thus, an important prerequisite to any trial would be the evaluation of the sensitivity of the proposed testing protocol relative to the "gold standard". This may be from pilot studies or literature review. It is essential that the management of cancer does not differ between the screened and the unscreened group, since it would be difficult to separate the benefits of treatment and screening between the two groups. A protocol for treating patients entered into the trial would be a major advantage, this should be best standard clinical practice.

The only conclusive end-point for the evaluation of a screening test for cancer is a significant reduction in cancer mortality in the screened group compared to the control group. In this comparison, the "screened" group must include all subjects randomised to be offered screening not merely all those who are screened. Other possible end-points are susceptible to the effects of lead time bias, length time bias or selection bias; these terms are explained by MORRISON [1].

Design of screening trials

A screening trial is a scientific study whose objective is to determine whether a screening programme can achieve its aim, which for prostate and most other cancers, though not for cervical cancer for example, is to reduce mortality from the disease. These studies are expensive and few are ever mounted to investigate each particular intervention so that their results have enormous impact on public health practice. In addition, since public health decisions may involve choices between alternative costly interventions, estimations of the magnitude of the benefits should be available from the completed trial data. Therefore, careful attention to the design at the outset is absolutely essential and the principles of control and randomisation should be adhered to.

Inappropriate designs

It may serve to clarify some of the important design issues by giving brief consideration to designs that grow naturally out of exploratory and pilot studies.

Hopefully, this will serve to illustrate major problems that can arise and that must be avoided.

Comparison of attendees and nonattendees

It is quite common for a screening programme to be introduced into the whole of a geographical community or another defined population, either on a pilot basis or because local public health professionals are themselves convinced of the efficacy of the planned intervention. The population for whom screening is available (e.g. men aged 55–74 yrs) can be divided into two groups: those who have accepted the offer of screening and those who have not. Mortality from the specific cancer type in the two groups can then be compared. Lower mortality rates in the men who have been screened suggests that screening is efficacious. However, these comparisons are biased since the two groups of men being compared are self-selected and there is no way of knowing the complex spectrum of factors which make some men accept and others reject screening, nor how these might affect their underlying risk of death from prostate cancer. There has been considerable methodological work applied to this basic design [2, 3] and it is useful in situations where screening trials are unethical (because screening is generally believed to be beneficial and cannot be withheld from a control group). This is one of the reasons why cervical cancer screening has never been evaluated in a randomised trial. The basic problem of selection bias has never been overcome and, for breast cancer at least, it appears that these comparisons consistently overestimate the benefit of screening [4, 5].

Before and after comparisons

An alternative to the situation described above is to compare mortality rates from prostate cancer in the total population (of men in the target age range) after the introduction of screening with those who have previously been observed. This type of analysis is also an important part of the statisticians' repertoire for use when screening trials are inappropriate. There are, however, several major draw-backs: 1) the basic assumption that the underlying mortality experience would be the same in the absence of screening is suspect, since therapy may have improved and "prostate-awareness" increased, thereby leading to earlier stage at presentation of the disease; 2) the benefits of screening emerge slowly and the mortality of all men diagnosed before screening was introduced and even some men diagnosed in the early years of the screening programme will be uninfluenced by screening, the simplistic use of time of death will inaccurately reflect the potential influence of screening and yield large conservative biases; and 3) no attention is paid to immigration and emigration from the region.

Survival analyses

A study that is particularly attractive to clinicians involves comparison of survival among people with

cancer according to the method of detection. For prostate cancer appropriate methods could be: 1) symptomatic presentation, 2) screen detected, prostate specific antigen (PSA) positive, 3) screen detected, TRUS positive, and 4) screen detected, digital rectal examination (DRE) positive. In addition 1) could be further divided according to screening history and 2), 3) and 4) could be modified to include combined test results. The basic method of analysis is to compare survival from date of diagnosis in each of the groups. Improved survival for screen-detected cases and for those detected by specific modalities is often regarded as persuasive evidence of the efficacy of screening and of the superior performance of one (or more) modality. These analyses can be useful at the early stage and as aids to interpretation of a completed trial, but they should be approached with extreme caution since they are influenced by several biases, such as the following.

Selection bias. Selection bias occurs as above, since the screen-detected cases all arise in men who have selected themselves for the application of a screening test.

Lead-time bias. Lead-time bias occurs as a result of measuring survival from date of diagnosis. As a result of the advancement of this date in the screen-detected cases they will have an apparent survival advantage even if their actual date of death has not been altered by the screening process. It is important to note that this same bias affects any comparison of the screening modalities, a test which has greater capacity to advance the diagnosis will have an apparent advantage even if the benefits it confers are identical to other screening tests. This has not yet been assessed for prostate cancer and remains an important argument against recommending any of the available tests, the DRE or, more importantly, the PSA, as a population screening procedure for prostate cancer at this time.

Length bias. Length bias occurs too since the screen-detected cancers arising amongst the men actually screened will be a sample whose average time in the detectable preclinical phase is longer than that of the totality of disease. It is likely (though difficult to prove without withholding treatment from human subjects) that slower progression through the preclinical stage of disease correlates with better prognosis later. If this is the situation, then the screen-detected cases would have improved survival, without the screening process having conferred any benefit on the men who had been screened. To avoid this particular bias it is necessary to combine the symptomatic cases arising in screened men (e.g. the "interval" cases) with the screen-detected cases.

One would normally wish to stratify an analysis of survival by values of age and certain other confounding variables. As a consequence of lead-time bias, stratification by variables known only at diagnosis is not valid. In particular these analyses cannot be stratified by age at diagnosis. It should be noted, in passing, that for this analysis of screening, data can be stratified by stage at diagnosis.

Screening trials: the basic design

Randomised controlled trials of drugs and other therapies are standard in modern medicine. The Health Insurance Plan (HIP) trial of breast cancer screening was one of the first to apply the same principles to evaluate prophylactic procedures and the first application to screening for cancer [6–8]. Although designed and commenced more than 40 yrs ago, this study remains the paradigm for evaluation of screening.

It is appropriate to highlight some basic features of the design. The study population is identified and randomised at the outset; this is a very large group of healthy (*i.e.* free from diagnosed cancer) individuals, the healthy status is typically confirmed retrospectively, but is important to ensure that individuals who so obviously cannot benefit from screening are excluded. The date of randomisation precedes diagnosis of cancer in general by many years and the majority of the study population will never suffer from (symptomatic) disease. To avoid lead-time bias all durations of time are measured from this date of randomisation, which is uninfluenced by status within the study.

The study population is randomised to an intervention group and a control group usually of equal sizes. All members of the intervention group are offered the opportunity to participate in a screening programme. Some will accept and some will not, and amongst those who do accept some cancers will be screen-detected whilst others will surface clinically. Neither the method of detection nor the decision to participate is relevant to the eventual analysis, which compares the number of deaths from cancer in the entire intervention group with the number in the control group. Note, also, that opportunistic screening by members of the control group is likely, but is ignored in the analysis. Thus, in the familiar terminology of the randomised clinical trial, the analysis is based on the "intention to treat" principle. "Protocol deviants", men offered screening who refuse it and men not offered screening who procure it, will dilute the power of the study [9], and efforts should be made to minimise them.

The statistical power of the study depends not on the size of the study population but on the expected number of events (*i.e.* deaths from cancer) during the follow-up period. It is important to note here that screening trials typically encounter two problems regarding numbers: very large numbers of individuals for follow-up and small numbers of events for analysis.

Finally, it must be emphasised that a screening trial with all its resources reports directly only on the particular screening programme being applied. This includes: the screening protocol, the choice of screening tests, the decision rules governing selection for further evaluation and a variety of other factors, including the frequency of screening, the expertise and training of staff, the technical quality of the equipment and the management protocol.

The only methods of glean information concerning other possible programmes will be by computer or statistical modelling [10] or further analyses of trial

data, which are themselves subject to bias (*e.g.* survival comparisons and inspection of interval cancer rates [11]). It is thus imperative that careful attention be paid to the choice of screening tests and the specification of the screening programme at the outset.

Problems associated with the basic design

Protocol deviants. It has already been noted that there will be a dilution that will result from the screening of control men and the failure to screen men in the intervention group. The main implications are for the statistical power of the study [9]. For trials of breast cancer screening neither of these have presented major problems; attendance rates among women invited to participate in the screening trial have been $\geq 60\%$ and only at Malmö [12] has significant numbers of the control population been screened. There are reasons to believe that the situation for prostate cancer screening may be quite different. First, the target population is older and attendance at even breast screening decreases sharply with age. Secondly, pilot studies have tended to give disappointing attendance figures. Finally, there is some evidence that professionals in primary healthcare are encouraging their patients to undergo tests using some that would be part of a screening programme (especially the PSA). The extent of this problem may be such that alternative designs are required.

It must also be realised that reporting trial results based on small percentages of attendees at screening yields little information on the ability of the screening process to prevent deaths since it has been delivered successfully: a negative report is consistent with the following "if only men could be persuaded to be screened for prostate cancer then a major impact on mortality from the disease could be made".

Zelen's single consent design. The basic design is analogous to Zelen's single-consent design (although the analysis strategy differs); informed consent is in effect sought only from members of the intervention group and is given only by those who agree to the intervention (*i.e.* who attend for screening). Detailed follow-up is nevertheless undertaken for all members of the control group and for those who refuse the offer of screening. It is likely that developments in medical ethics and data confidentiality in a number of countries will now render this approach impossible. It will often be necessary to obtain informed consent from all members of the study population. A further problem then arises since it may then be considered unethical to offer no benefit to those entered into the study population but randomised to the control arm of the trial.

Dilution effect of existing cancers. The criteria for entry to the trial excludes subjects with diagnosed cancer, but under the basic design subjects with asymptomatic disease will be recruited into both arms of the trial. For some of these the disease may already be so far developed that little or no benefit can be conferred by advancing the diagnosis. The effect of eventual deaths of these cases in both arms of the trial will be: 1)

to increase the number of events for analysis, but 2) to reduce the percentage differences in mortality between the two arms, and 3) to reduce the statistical power of the study to detect the same "real" benefit of screening. This problem has not been too important for breast cancer screening trials, since, at least for short periods of follow-up, significant reductions in mortality were observed based on the mortality experience of cases who had had, at most, one screen. In addition, locally advanced or metastatic but asymptomatic breast cancer is extremely rare. However, pilot studies have shown that a substantial proportion of prostate cancers detected at initial screening are already advanced. Some of the alternative designs discussed below have the capacity to ameliorate this problem.

Determination of end-points. Good clinical trial practice requires that determination of the end-points should be as objective as possible and, ideally, all cause mortality is taken as outcome. Where subjectivity is essential then bias is avoided by "blindness" of the assessor. Since trial subjects in screening trials are initially healthy, disease specific mortality is low both in absolute terms and as a proportion of all cause mortality. It follows that all cause mortality is an unsuitable end-point [13, 14]. Thus, determination of "death from prostate cancer" is required; this can use either routine (death certificate) data or special ascertainment for the study. Of these the second is vastly more expensive and since case-note review cannot be "blind" it is potentially biased. Use of death certificates provides somewhat inaccurate (though unbiased) data which results, at worst, in a conservative bias for the results.

Current situation in screening for cancer

Cancer can be detected earlier than usual either when an individual recognises symptoms and then quickly consults and is diagnosed by a physician, or through the application of a screening test, aimed at diagnosing pre-cancerous changes or cancer itself, in generally asymptomatic individuals. Public education can be used to increase awareness of symptoms and their importance, but the effects of such strategies have not been consistently evaluated. Despite not having detailed scientific evaluation, there are important reasons for improving public knowledge and awareness of abnormal signs and symptoms. However, at the present time, issues in population screening are once more of critical concern to physicians and public health specialists.

Cervical cancer

In 1996, the National Institutes of Health (NIH) Consensus Statement [15] concluded that carcinoma of the cervix is causally related to infection with the human papillomavirus (HPV). Reducing the rate of HPV infection by changes in sexual behaviour in young people and/or through the development of an effective HPV vaccine would reduce the incidence of this disease.

Cytological ("Pap smear") screening remains the best available method of reducing the incidence and mortality of invasive cervical cancer. The test is named after George Nicolas Papanicolaou, who was born in 1883 in Kymi, Greece. In 1920, Papanicolaou began his study of the vaginal cytology of the human. After becoming familiar with normal cytology changes, he found some cases of malignancy and remarked "the first observation of cancer cells in a smear of the uterine cervix was one of the most thrilling experiences of my scientific career". In 1928 he published a paper about the results of his work entitled "new cancer diagnosis".

At the New York Hospital (New York, NY, USA) in 1939, the re-evaluation of the vaginal smear for cancer detection began with all women patients required to take a routine vaginal smear. Herbert Traut, a gynaecological pathologist, collaborated with Papanicolaou to validate the diagnostic potential of the vaginal smear. In 1943, they published their findings and conclusions in the famous monograph, *Diagnosis of Uterine Cancer by the Vaginal Smear* and the diagnostic procedure was named the Pap test.

Screening for cervical cancer by examination of a cervical smear leads to a reduction in the incidence, and subsequently mortality from cervical cancer [16]. It has also been demonstrated to be cost-effective in older women, particularly among those who have not been screened regularly [17]. The impact is greatest where organised screening programmes exist with personal letters of invitation; this leads to an improved attendance particularly among those women who are at high risk of cervical cancer [18].

It has been shown, particularly in the Nordic countries, that a population-based and well-organised screening programme, with a valid target age range, the right frequency of screening and built-in quality assurance programmes at each stage of the screening process, is more successful than opportunistic screening, and that it can be effective in reducing both the incidence of and mortality from invasive cervical cancer. The most successful programme in terms of reduction in risk of cervical cancer is in Finland, with an official recommendation that a screening programme be started at age 30 and that the smear be repeated every 5 yrs [18].

While cytological screening programmes are effective in preventing invasive cervical cancer and cervical cancer mortality, various reports have highlighted that the method may fail to detect a certain number of cervical cancers, mainly of the glandular type [19]. For instance, screening histories were examined in a case-controlled study where the cases were all women with invasive cervical cancer in 24 health districts of Health Boards of Scotland, England and Wales in 1992. It was estimated that the number of cases of invasive cervical cancer would have been 57% greater if there had been no previous screening, in women <70 yrs it would have been ~75% [20]. The authors estimated that full adherence to current screening guidelines could have prevented 1,250 cases of invasive cervical cancer in the UK in the same year. However, further steps would have to be sought to prevent some of the remaining 2,300 cases in women <70 yrs.

The most frequently evoked reasons to explain the lack of sensitivity of cytological screening surround inadequate cell sampling with the spatula and errors in the reading of smear slides. However, even in the best hands, a certain number of false-negative cytology tests cannot be explained by sampling or reading problems. Besides searching to improve screening coverage, there is a clear need for additional ways to improve screening methods for cervical cancer. The first area of improvement should be in the spatula used for cell sampling (with current preference for instruments like the extended tip spatula) and in the automatization of cytological reading. However, it remains to be assessed whether these improvements in the cytological methods will solve all the problems with false-negative results.

Methods based on other approaches to cervical cancer screening are under investigation and some may eventually be employed as adjuncts to the pap-smear test. One of them is cervicography, which is popular in the USA and in some screening clinics in Europe. Although numerous groups have published results on screening with cervicography, well-conducted studies to evaluate cervicography (as compared to cytology) are rare [21]. The principal findings of these studies are that cytology and cervicography detect different pre-malignant lesions of the cervix. Hence, cervicography represents an interesting adjunct test to cytology, able to reduce the false-negative results of the screening process. Cervicography is, however, not a simple technique as it requires a large reading experience and experience is essential to keep the unnecessary referrals to colposcopy-biopsy as low as possible.

Several other new technologies have shown some early promise, although their potential value in the early detection of cervical cancer is still in the early stage of assessment. For example, the Polarprobe is an instrument that is based on a mathematical recognition algorithm for detecting anomalies in the cervix. An initial pilot study with a prototype of the Polarprobe instrument has shown false-positive and false-negative rates in the order of 10% for the detection of premalignant lesions of the cervix [22]. This provides strong encouragement to develop studies to further examine the potential role of the Polarprobe in the detection of high-grade cervical lesions. Speculoscopy was developed in the USA in ~1992, as an attempt to increase the sensitivity of the pap smear by using chemical-luminescence.

While the latter is being investigated as a tool in assisting the collection of a cervical smear, the technology may also have a potentially important application to cervical cancer detection in developing countries, where it could be very useful in improving the quality of a visual inspection of the cervix. Any way to advance the detection of cervical cancer or pre-invasive cervical lesions in the developing world would be particularly important, as such regions typically have the highest rates of invasive cervical cancer [23].

Given the implication of HPV infection in cervical cancer, detecting HPV could represent an appealing screening method. A study of 2,009 women, having routine screening in England and Wales, revealed that

44% of cervical intraepithelial neoplasia (CIN) lesions of grade 2/3 detected had a negative cytology and were found only by HPV testing (for types 16, 18, 31 and 33); a further 22% were positive for HPV but demonstrated only borderline or mild cytological changes [24]. However, 25% of CIN 2/3 lesions were not detected by the four HPV tests. Although appealing, routine HPV testing for cervical cancer screening is still controversial, as HPV infection is very common in women <30 yrs, and what is really important are those women >30 yrs with a HPV infection that persists over a long period of time. As it is currently impossible to identify those women with a HPV infection who will develop cervix cancer, HPV testing is proposed to be used in various ways. For instance, as an adjunct to cytology for sorting out the cytological results classified as ASCUS (atypia of squamous cells of undetermined significance), referring ASCUS lesions positive for a HPV infection to colposcopy-biopsy. Another proposal consists of testing all women >30 yrs for HPV and referring only those positive for HPV to cytology [25]. Hence, HPV testing is still to be thoroughly evaluated in order to find the exact role it could play in cervical cancer screening. Certainly, the development of an effective HPV vaccine will accelerate the development of HPV screening programmes.

United Kingdom National Cervix Cancer Screening Programme. The United Kingdom National Cervix Cancer Screening programme was established in 1986. The target population is women aged 25–64 yrs in England, aged 20–64 yrs in Wales and aged 20–59 yrs in Scotland, who have been screened in the previous 5 yrs (5.5 yrs in Scotland). The Health of the Nation [26] target for cervix cancer is to reduce the incidence of invasive cervix cancer by at least 20% by the year 2000 (baseline 1986).

Participation in the United Kingdom Cervix Cancer Screening Programme is high in England (84.0%), Wales (81.9%) and Scotland (86.6%). The most recent incidence data available from SOCRATES (the Scottish Cancer Registration System) is for 1998, and between the baseline year of 1986 and 1998 there was a decrease in incidence of 30.9%.

In conclusion, near maximal effectiveness in reducing incidence and mortality from cervix cancer can be achieved by an organised programme of cervical smear testing with high coverage, in which screening is initiated at the age of 25 yrs and is repeated at 3- or 5-yrly intervals until the age of 60 yrs. Extension of this approach should be considered only if maximal coverage has been attained, the resources are available and the marginal cost-effectiveness of the recommended changes has been evaluated. HPV testing and other new technologies have still to be thoroughly evaluated, although some of these could potentially usefully augment conventional cytology.

Breast cancer screening

There is considerable evidence that breast cancer screening with mammography is effective at reducing mortality from breast cancer, especially in

circumstances where the quality of the mammography is high with good quality control. The best estimates from randomised trials suggest that the size of the reduction may be ~30%, if take-up of screening in the population is good and quality control standards high. An overview of the Swedish trials reported relative risks of death of 0.71 in the group randomised to the offer of screening, with a 95% confidence interval (CI) of 0.57–0.89 for women aged 50–59 yrs at entry. Results for women aged 60–69 yrs were almost identical. When applied to a population, it could be expected that a well-organised programme with a good compliance could lead to a reduction in breast cancer mortality of the order of 20% in women aged >50 yrs [27]. There is, as yet, no clear evidence that screening benefits older women and it is certain that they are less willing to attend for screening.

More importantly, results for younger women (<50 yrs) are ambiguous, with no trials having large enough statistical power to analyse these women separately. The issue of screening women aged 40–49 yrs is an important social problem; >40% of the years of life lost due to breast cancer diagnosed before the age of 80 yrs is attributable to cases presenting symptomatically at ages 35–49 yrs, frequently an age of maximal social responsibility for women. In 1993 it was considered that there were no statistically significant results for this age group reported, but point estimates, including both reductions and increases in breast cancer mortality in women offered mammographic screening while aged <50 yrs [27], were reported. Since then it has become clear that the natural history of breast cancer among women younger and older than 50 yrs may be different. Re-analysis of the Swedish Two-County Trial has shed considerable light on two important issues. First, the mean sojourn time is estimated to be between 3 and 4 yrs for women aged ≥ 50 but only ~20 months for women aged <50 yrs, after adjustment for tumour size and nodal status [28]. Furthermore, dedifferentiation often occurs at an early stage for women <50 but later for women >50 yrs. This implies that the poorer performance of mammographic screening in younger women might be due to rapid progression and failure to arrest dedifferentiation in this age group because the screening interval is too long [29]. It has also been demonstrated mathematically that the same benefit in terms of breast cancer mortality reduction could be expected among younger women if they were screened approximately every 18 months [29]. At Falun, Sweden, in 1996, it was concluded that mammographic screening of women aged 40–49 could reduce subsequent mortality from breast cancer and that it was probably necessary to screen every 12–18 months in this age range, with two-view mammography and double reading of films, to obtain substantial benefits [30].

This position was not accepted by the United States National Cancer Institute Consensus meeting held in January 1997, which ended in controversy and some acrimony [31], but was upheld as a basis for recommendations to women by the American Cancer Society in March 1997. Clearly, there are some remaining uncertainties surrounding this issue, but

women, and their families, deserve better ways to receive reliable information and clear recommendations than are available at present about such an important issue.

With regard to other methods of breast cancer screening, a recent large trial of taught breast self-examination demonstrated a reduction in mortality among women who had been trained in breast self-examination, although the authors concluded that the study would continue and that stronger results may emerge during follow-up [32]. The effect of taught breast self-examination is also currently being evaluated in a randomised trial in Russia [33].

The evidence from these studies was sufficiently strong that the European Code Against Cancer [34] included the following recommendation: check your breasts regularly and participate in organised mammographic screening programmes if you are >50 yrs. With the premise that breast cancer screening with mammography is effective at reducing mortality from breast cancer, especially in circumstances where the quality of the mammography is high with good quality control, a National Mammographic Screening Programme was introduced into the UK in 1988 [35]. It has been estimated that, by 2004, the second round of screening in East Anglia should reduce breast cancer mortality by ~7% in women <55 yrs at diagnosis and by ~19% in those aged 55–64 yrs [36].

United Kingdom National Mammographic Screening Programme. A National Breast Cancer Screening programme was launched in the UK in 1988, following publication of the Forrest Report [35]. The target population of the programme is women aged 50–64 yrs who have been screened in the previous 3 yrs. Currently, participation is good: England (67.6%), Wales (69.9%), Scotland (71.1%) and Northern Ireland (70.5%).

The Health of the Nation Target [26] for breast cancer was to reduce the death rates for breast cancer in the population invited for screening by at least 25% by the year 2000 (baseline 1990). The most recent mortality data available from the General Registry Office (Scotland) is for 2000, and between the baseline year of 1990 and 1999 there was a decrease in mortality of 24.4% by Poisson regression (mortality rates have been age standardised to the European Standard Population and covers the age group 55–69 yrs, which is based on a recommendation of the evaluation group of the National Health Service Breast Screening Programme and demonstrated a 25.7% decline without using regression estimation).

Global summit on mammographic screening. Clinical and pathological considerations clearly demonstrate that survival following the diagnosis and treatment of breast cancer at an early stage is very much better than when the disease is locally advanced or metastatic. Mammography can detect tumours at a clinically undetectable stage, such tumours have a very good prognosis and many can be cured by appropriate treatment. The results from the early randomised trials of mammographic screening were sufficiently promising to lead to the introduction of organised

national programmes of screening in several countries in 1986–1988. Reports from seven trials involving >500,000 women subsequently indicated a reduction in mortality from breast cancer of 20–30% in women invited to be screened. The reduction of mortality in those actually attending for screening is clearly greater.

However, doubts regarding the validity of five of these trials have recently been raised by GOTZSCHE and OLSEN [37], firstly in an article to the *Lancet* and then as a Cochrane review with a research letter published in the *Lancet* [38]. Their conclusions have been vigorously debated. Such uncertainty regarding the efficacy of mammography is clearly an important public health issue which must be resolved. If the conclusions of GOTZSCHE and OLSEN [37] are correct, women participating in screening programmes may have been harmed; if they are incorrect encouragement to avoid screening may cost lives.

Following the controversial publication of GOTZSCHE and OLSEN [37], Swedish workers have conducted an overview of four of their trials. Their conclusions, published in the *Lancet* [39], indicate that the benefit of breast screening, in terms of a reduction in breast cancer mortality of 21%, persisted for a median time of 15.8 yrs. They also argued convincingly that many of the criticisms made against the Swedish trials by GOTZSCHE and OLSEN [37] are misleading and scientifically unfounded.

In addition to this overview, two working groups have been convened, the proceedings of which are as yet unpublished. A working group of the International Agency for Research on Cancer (IARC), which met in Lyon on 5–12 March 2002, consisted of 24 experts from 11 countries. The quality of the seven trials was carefully assessed, as a result of which it was concluded that many of the criticisms raised by GOTZSCHE and OLSEN [37] were unsubstantiated. Further, those criticisms of substance did not invalidate the evidence that screening by mammography reduced mortality from breast cancer in women of 50–69 yrs of age. In women who participated in screening programmes the reduction was estimated as 35%. For women of 40–49 yrs, evidence for a reduction in mortality was limited. It was recognised that the effectiveness of national programmes of screening would vary according to differences in coverage and compliance, the quality of the mammograms, methods of assessment and treatment and many other factors. However, such organised programmes were more likely to be effective in reducing the rate of death than the sporadic screening of selected groups of women.

In addition, the United States Preventive Services Task Force (USPSTF) has also assessed the current evidence on mammographic screening. Key elements in this assessment are an evaluation of the quality of the available evidence and the performance of a meta-analysis wherever possible. They concluded that mammographic screening could be recommended as a category B intervention on the grounds that the quality of evidence was fair and the net gain moderate. The reduction in breast cancer mortality among women invited to screening appeared to be 23%.

In response to the uncertainty over the efficacy of

breast screening, a global summit on mammographic screening was organised in Milan between 3–5 June 2002. The Summit was planned in association with the World Health Organization, the European Commission, the American Cancer Society, the Centers for Disease Control and Prevention, the American Italian Cancer Foundation, the European Society for Medical Oncology, the American Society for Clinical Oncology and the International Union Against Cancer.

The design and recent results from the seven randomised trials were presented and discussed in detail in the light of each criticism put forward by GOTZSCHE and OLSEN [37, 38]. Some were discarded as being wrong, others had been addressed by new analyses and shown to be of minor significance. It was appreciated that conducting such large trials over many years is difficult, particularly as technology, treatment and indeed public health policy can change during their course. But there was unanimity that the remaining minor considerations did not detract from the conclusion that screening mammography reduced the mortality from breast cancer in women receiving an invitation to be screened in well-organised clinical trials: the reduction in breast cancer mortality appeared to be between 21% and 23% according to recent estimates. Those participating fully could expect greater benefit.

Those attending the Milan summit believed that the criticisms which had been raised had been fully addressed, that the "book" on screening trials should now be closed and that future activities should concentrate on the evaluation of organised programmes of mammographic screening, on exploring methods to ensure full participation, particularly amongst deprived women, and on the development of new technologies for early diagnosis. During the meeting there were presentations of 14 such organised programmes of population screening; those of longer duration demonstrating tentative trends towards mortality reduction. Viable data from more recently established programmes tended to have similar values for many of the intermediate end-points (e.g. stage of disease) as seen in the longer-established programmes, which was encouraging.

Mammographic screening is only one step in the total management of women with breast cancer. As has been shown from long-term established programmes in the UK, Sweden, Finland and the Netherlands, recognition of the importance of the multidisciplinary team in the assessment of mammographic abnormalities has "spun over" into the symptomatic sector, leading to the development of integrated multidisciplinary breast care centres. Staffed by dedicated surgeons, radiologists and pathologists working alongside breast care nurses, counselling and other support personnel, these centres offer optimum care for women with breast cancer.

Mammographic screening: summary of the current situation. Forty years of clinical trials, the contribution of hundreds of scientists and health workers and the dedication of hundreds of thousands of women to participate in studies lasting for decades has resulted in adequate evidence to support the efficacy of

mammographic screening for breast cancer, which now allows its transfer to the arena of public health-care. Doctors and women can now be assured that participation in organised screening programmes, with high quality control standards, is of benefit, provided appropriate investigation and treatment is available. Special effort should be made to encourage screening among the more deprived. It is important not to overemphasise the benefit of screening and to appreciate that this is but one step in the total care of women with the disease. Women should, however, be informed clearly of the level of benefit and of potential risks and costs.

The Milan Global Summit, having examined recent results from all seven randomised trials of screening, concluded that evidence of benefit was convincing and that it was now time to move on. Attention should now focus on the further development of organised programmes of mammographic screening on a population basis, and insistence on quality assurance and meticulous evaluation.

Colorectal cancer

Colorectal cancer is the fourth commonest form of cancer that occurs worldwide, with an estimated 782,900 new cases diagnosed in 1990 [40]. The disease is not uniformly fatal, although there are large differences in survival according to stage of disease. In advanced colorectal cancer, in which curative resection is possible, 5-yr survival in Dukes' B is 45%, which drops to 30% in Dukes' C [41]. Five-year survival in resected Dukes' A is ~80% and survival following simple resection of an adenomatous pedunculated polyp containing carcinoma *in situ* (or severe dysplasia) or intramucosal carcinoma is generally close to 100%. Although it has been argued that death from colorectal cancer may be avoidable [42], it is estimated that there are still 394,000 deaths from colorectal cancer world-wide annually [43].

The identification of a well-determined pre-malignant lesion, the adenomatous polyp, together with the good survival associated with early disease, make colorectal cancer an ideal target for screening. In the past quarter century, great progress has been made in the ability to screen patients for colorectal cancer or its precursor state, using advances in imaging and diagnostic technology. WINAWER [44] and GREGOR [45] first employed the faecal occult blood guaiac test cards, and the flexible sigmoidoscopy was introduced in the mid-1970s to replace the rigid sigmoidoscope, which had been first introduced in 1870, and colonoscopy has been available since 1970 [46].

Four randomised trials have examined annual or biennial screening with faecal occult blood testing (FOBT), while there are only early data available regarding sigmoidoscopy and colonoscopy and little as yet from randomised trials. There is evidence from these randomised trials to support the use of FOBT [47–49], with a reduction in colorectal cancer mortality of ~16% (95% CI 7–23) from a meta-analysis (23% (95% CI 11–43) reduction among those screened) [50] and a reduction in incidence reported, but only after

18 yrs of follow-up [51]. Concerns remain about the high rate of false-positive results, the feasibility and the small clinical benefit of such screening. These concerns were outlined recently [52] and it was calculated that 1,173 individuals needed to be tested for 10 yrs to avoid one death from colorectal cancer.

Various organisations have considered recommendations for colorectal cancer screening. A recent report to the Europe Against Cancer Advisory Committee on Cancer Prevention [53] recommended that "FOBT should be seriously considered as a preventive measure". In an accompanying editorial, COEBERG [54] was not completely persuaded by this advice, pointing out that there was a contribution to the modest effects seen potentially due to the more intensive follow-up of controls in some of the trials [55].

Subsequently, there have been some important findings reported. LIEBERMAN *et al.* [56] examined the sensitivity of FOBT and sigmoidoscopy for detecting neoplasia. A total of 2,885 asymptomatic subjects provided stool specimens on cards, which underwent rehydration. They then underwent colonoscopy with sigmoidoscopy, defined as examination of the rectum and sigmoid colon during colonoscopy. Of the subjects, 23.9% with advanced neoplasia had a positive test for occult blood. As compared with subjects who had a negative test for faecal occult blood, the relative risk of advanced neoplasia in subjects who had a positive test was 3.5 (95% CI (2.8–4.4)). Sigmoidoscopy identified 70% of all subjects with advanced neoplasia. Combined FOBT and sigmoidoscopy identified 75.8% of subjects with advanced neoplasia. The authors concluded that one-time screening with both faecal occult blood (with rehydration) and sigmoidoscopy fails to detect advanced colonic cancer in 24% of subjects with the condition [56]. This is an important finding that requires further confirmation and detailed investigation.

DETSKY [57] considered that there were five important reasons why colonoscopy was not routinely recommended as a screening tool: the standard of evidence, adherence, risk, economics and availability. The issue of standard of evidence is one which requires much attention in epidemiology at the present time. FOBT has been evaluated in randomised trials whereas colonoscopy has not. DETSKY [57] concluded that the higher sensitivity of colonoscopy plus the evidence that early detection improves survival is sufficient to conclude that colonoscopy is more effective than FOBT. In addition, there is support from observational studies.

Screening with sigmoidoscopy has been demonstrated in case-controlled and nonrandomised studies to reduce the incidence and mortality from colorectal cancer by >50% [58–60]. FOB testing does not appear to reduce the incidence of colorectal cancer although there are important questions to be resolved [61], as follows. Should FOB testing now be recommended as a population screening method? Should consideration be given to other screening modalities for colorectal cancer? Since a large proportion of individuals tested for FOB have positive tests and are referred for colonoscopy, could it prove effective to bypass FOB

testing and go directly to screening colonoscopy? Or flexible sigmoidoscopy? This latter strategy is currently being assessed in a large, randomised trial and it is a clear reflection of the tremendous potential for colorectal cancer early detection by screening, which is clearly outlined in detail elsewhere [62]. This should continue to be a priority research activity at present.

To bring all the arguments about colorectal cancer screening with FOBT to the readership, the *Annals of Oncology* invited twelve international groups to outline their position; nine accepted immediately and three declined to do so (on the grounds of pressing priorities). The groups were chosen on their expertise in the area of colorectal cancer and groups who have been involved in trials of colorectal cancer screening were excluded from consideration.

One common thread in all these articles relates to the economics of FOBT. LA VECCHIA [63] outlined the basic epidemiological data, both from observational studies and randomised trials, and indicated where there are key gaps in knowledge. MCARDLE [64] noted key issues, such as compliance among deprived members of the community and the utility of two pilot studies on-going in the UK at present to address such issues.

LOWENFELS [65] invited reflection on "why don't we screen for this potentially preventable cancer?" emphasising that this is a major practical issue. LOWENFELS [65] lays out the arguments in an epidemiological manner as to why screening is necessary, why FOBT may not be the best test and weighs the potential benefits of other tests. Leading gastroenterologists, BLEIBERG [66], CRESPI and LISI [67] and STRUL and ARBER [68] adopted a more clinical approach and presented good overviews and interpretations of the available data. STRUL and ARBER [68] also gave an interesting glimpse of new stool-based tests that may soon be available for population assessment and use.

AUTIER [69] and BARRY [70] took a broader public health perspective. Arguing that FOBT produces modest changes in mortality rates and AUTIER [69] concludes that FOBT is less efficient than screening tests for other cancers, such as pap-smear for cervix cancer and mammography for breast cancer. BARRY [70] points out that "most Americans have not been screened for colorectal cancer by any means" and the situation is identical throughout the rest of the world. Unfortunately, as BARRY [70] emphasises, there is a gap between what the (public health) doctor prescribes and what the patient is willing to do. While this situation persists and scientific squabbles continue about the best way to screen populations for colorectal cancer, the chance is being missed to prevent a significant number of the 400,000 colorectal cancer deaths which occur each year throughout the world.

Screening for other forms of cancer

While there are screening tests proposed for a number of different forms of cancer, there are no randomised trial data to support screening at other sites as a public health measure. There has been

apparent success in Japan, with screening for stomach cancer, but this has not been carefully evaluated [16]. There are screening tests available and being evaluated for oral cancer, nasopharynx cancer and neuroblastoma [16]. The issues in screening have been well outlined in two recent monographs [71, 72].

Screening for prostate cancer

At the present time there is great pressure to screen for prostate cancer, although widespread implementation of screening programmes for prostate cancer cannot be recommended based on the available evidence; little has changed since the most recent review of the topic by an expert committee of the IUAC (International Union Against Cancer), which came to the same conclusion [16]. Unfortunately, there are national practices which are at complete variance; in the USA screening with prostate specific antigen (PSA) is widespread [73], while in the UK there is a strong bias against screening with PSA [74], essentially on evidence-based grounds.

The main reason for this situation is that no useful results are available from randomised trials assessing screening for prostate cancer. These are the only methods of evaluation that avoid bias and, as a consequence, it is not known whether screening by whatever of the available modalities or their combinations is effective in leading to a reduction in the mortality rate of prostate cancer. Obviously, this is a necessary pre-requisite for embarking on population screening or even screening high-risk groups (even if such a group can be defined for prostate cancer).

Screening for prostate cancer at a population level would be expensive and consume a large proportion of available resources for health; it is essential to have some indication of effectiveness and efficacy before embarking on such programmes. Although there is no evidence available at present from randomised trials to indicate that any lives will be saved by such screening, it is logical to suppose that early detection and effective treatment could be effective. The current situation in the USA, in favour of PSA screening, has been compared with the enthusiasm and the similarity of the arguments put forward in favour of lung cancer screening several years ago [75], and this bright promise did not subsequently materialise.

Routine PSA testing was made freely available to the male population, aged 45–74 yrs, of Tyrol (Austria) from 1993 onwards [76]. By comparing prostate cancer mortality in Tyrol, where PSA testing was introduced, with the rest of Austria, where it was not, the impact of screening could be monitored in a natural experiment. Initially only total PSA was measured, but free PSA measurement was added in 1995. The IMx assay was used. DRE was not part of the screening examination.

There has been a reduction in mortality rates in the rest of Austria from 1993 onwards, with a greater reduction in Tyrol. Trends in prostate cancer mortality rates since 1993 differ significantly between Tyrol and the rest of Austria ($p=0.006$) [76]. The mortality reduction has remained significantly reduced in 2000.

Quite similar trends in prostate cancer mortality are found in geographically adjacent states to Tyrol (Vorarlberg, Carinthia, Salzburg), where there was some participation in the screening programme, while there is a similar and lower reduction in mortality in the remaining six states, where there was no participation.

These findings are consistent with the hypothesis that the policy of making PSA testing freely available and the wide acceptance by men in the population is associated with a sustainable reduction in prostate cancer mortality in an area where high-quality diagnostics, urology and radiotherapy are available freely to all patients [76]. This latter situation makes Tyrol somewhat unique and is a *sine qua non* for considering implementation of such programmes.

Screening for lung cancer

It has long been established that the best way to control lung cancer is to reduce cigarette smoking in the population, foremost through prevention and secondarily through smoking cessation. However, even after stopping smoking, long-term smokers remain at high risk for lung cancer. Although prevention and cessation strategies are obvious investments for intervention, presently there is no agreed upon control policy for subjects already at high risk due either to prolonged exposure to tobacco smoke or occupational exposures. Lung cancer, when clinically diagnosed, has a poor outcome with 10–16% survival at 5 yrs. If the tumour is small enough to be removed surgically, the outcome is much better, >70% for stage I tumours. This has led to speculation in the past as to whether long-term smokers or others at high risk might benefit from earlier detection.

In the 1970's lung cancer screening focussed on the chest radiograph and several studies were established. These trials had discouraging results, suggesting that screening by chest radiography did not lead to a significant reduction in lung cancer mortality. Increased numbers of tumours in the screened arms of the Czech Trial and Mayo Lung Project, both of which compared chest radiographs to usual care, suggested a degree of overdiagnosis of histologically confirmed lung cancers due to the screening [77, 78]. The Mayo project, in particular, had an excess of early stage tumours in the screened arm, but no deficit of late stage. Although no clear evidence of benefit from early detection emerged from these studies, they had a number of methodological shortcomings, so significant that an international lung cancer screening conference in Varese, Italy (1998) concluded that they were an "imperfect basis for public policy" [77]. The studies suggested shortcomings in chest radiography as a screening tool; including doubts about sensitivity. This and other aspects of chest radiograph screening should be clarified when the lung cancer results of the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO) are published [78].

With the development of low-dose spiral computerised tomography (CT) scanning, there is new hope for a sensitive screening tool for lung cancer [79–81].

In a Japanese study, 5-yr survival of lung cancer cases diagnosed by CT screening was around ~85% [81]. In the Early Lung Cancer Action Project (ELCAP), the stage of the nonsmall cell cancers diagnosed suggest that a similar outcome will be observed in these cases after 5-yr follow-up [79, 80]. Due to potentially manageable costs, acceptable levels of radiation exposure and improved detection sensitivity, there are grounds for hope that this new technology might allow for detection at a sufficiently early stage to allow successful treatment of lung malignancies that would be certainly fatal otherwise. The ELCAP study demonstrated that spiral CT was able to identify very small lung cancers in high-risk volunteers, with a resectability rate of 96% and a proportion of stage I >80% [79]. However, in order to achieve those excellent results, high-resolution CT had to be applied to a high proportion of subjects with a complex algorithm of three-dimensional reconstruction for minimal growth assessment, with a diagnostic period extending up to 2 yrs and major expertise in fine-needle biopsy of small lesions. A recent meta-analysis has highlighted the diagnostic value of positron emission tomography in undetermined pulmonary nodules [82].

The great improvements in diagnostic imaging offer real possibilities to develop better screening technologies. This is particularly true in lung cancer where something needs to be done, especially for those who have quit smoking. Such advances in technology raise other issues, notably how to evaluate a technology that will be outdated by the time the study is completed. The need to seek alternative, reliable methods of evaluation, should be a major research focus at present.

Acknowledgements. This study was conducted within the framework of support from the Italian Association for Cancer Research (AIRC). It is a pleasure to note the contributions to the authors thinking on screening of many people, particularly F. Alexander (UK), J. Cuzick (UK), P. Sasieni (UK), S. Daffy (UK), O. Brawley (USA) and G. Bartsch (Austria). Of course, they have no responsibility for any aspect of the text printed here.

References

- Morrison AS. Screening in chronic disease. New York, Oxford University Press, 1985.
- Collette HJ, Day NE, Rombach JJ, de Waard F. Evaluation of screening for breast cancer in a non-randomised study (the DOM project) by means of a case-control study. *Lancet* 1984; 1: 1224-1226.
- Sasco AJ, Day NE, Walter SD. Case-control study for the evaluation of screening. *J Chronic Dis* 1986; 39: 399-405.
- Wald N, Frost C, Cuckle H. Breast cancer screening: the current position. *BMJ* 1991; 302: 845-846.
- Moss SM. Case-control studies in screening. In: Miller AB, Chamberlain J, Day NE, *et al*, eds. Cancer Screening. England, Cambridge University Press, 1991; pp. 419-428.
- Shapiro S, Strax MD, Venet L. Periodic breast cancer screening in reducing mortality from breast cancer. *JAMA* 1971; 215: 1777-1755.
- Shapiro S. Evidence on screening for breast cancer from a randomised trial. *Cancer* 1977; 39: 2772-2782.
- Hurley SF, Kaldor JM. The benefits and risks of mammographic screening for breast cancer. *Epidemiol Revs* 1992; 14: 101-130.
- Zelen M. Are primary cancer prevention trials feasible? *J Natl Cancer Inst* 1988; 80: 1442-1444.
- Baker SG, Connor RJ, Prorok PC. Recent developments in cancer screening modelling. In: Miller AB, Chamberlain J, Day NE, *et al*, eds. Cancer Screening. England, Cambridge University Press, 1991; pp. 404-418.
- Tabar L, Fagerberg G, Duffy SW, Day NE. The Swedish two county trial of mammographic screening for breast cancer: recent results and calculation of benefit. *J Epidemiol Community Health* 1989; 43: 107-114.
- Andersson I, Aspegren K, Janzon L, *et al*. Mammographic screening and mortality from breast cancer - the Malmo Mammographic screening trial. *BMJ* 1988; 297: 943-948.
- Chamberlain J, Clifford RE, Nathan BE, Price JL, Burn I. Repeated screening for breast cancer. *J Epidemiol Comm Health* 1984; 38: 54-57.
- Wilson JMG, Jungner G. Principles and practice of screening for disease. Geneva, WHO, Public Health Paper No. 34, 1969.
- National Institutes of Health. Cervical Cancer. NIH Consensus Statement. Bethesda, MD, April 1996. 1-3; 43: 1-26.
- Miller AB. Issues in screening for prostate cancer. In: Miller AB, Chamberlain J, Day NE, Hakama M, Prorok PC, eds. Cancer Screening. Cambridge, UICC, Cambridge University Press, 1991.
- Fahs MC, Mandelblatt J, Schechter C, Muller C. Cost effectiveness of cervical cancer screening in the elderly. *Ann Intern Med* 1992; 117: 520-527.
- Hakama M, Magnus K, Pettersson F, Storm H, Tuhnius H. Effect of organised screening on the risk of cervix cancer in the nordic countries. In: Miller AB, Chamberlain J, Day NE, Hakama M, Prorok PC, eds. Cancer Screening. Geneva, International Union Against Cancer, 1991.
- Koss LG. The Papanicolaou test for cervical cancer detection - A triumph and a tragedy. *J Am Med Acad* 1989; 261: 737-743.
- Sasieni PD, Cuzick J, Lynch-Farmery E. Estimating the efficacy of screening by auditing smear histories of women with and without cervical cancer. The National Co-ordinating Network for Cervical Screening Working Group. *Br J Cancer* 1996; 73: 1001-1005.
- Coibion M, Autier P, Vandam P, *et al*. Is there a role for cervicography in the detection of the pre-malignant lesions of the cervix uteri? *Br J Cancer* 1994; 70: 125-128.
- Coppleson M, Reid BL, Skladnev VN, Dalrymple JC. An electronic approach to the detection of pre-cancer and cancer of the uterine cervix: a preliminary evaluation of Polarprobe. *Int J Gynecol Cancer* 1994; 4: 79-83.
- Parkin DM, Ferlay J, Pisani P. Estimates of the worldwide incidence of eighteen major cancers in 1985. *Int J Cancer* 1993; 54: 594-606.
- Cuzick J, Szarewski A, Terry G, *et al*. Human

- papillomavirus testing in primary cervical screening. *Lancet* 1995; 345: 1533–1536.
25. Meijer CJ, van den Brulle AJ, Snijders PJ, *et al.* The epidemiology of cervical cancer and human papillomavirus. Lyon, IARC Sci Public 119, 1992; pp. 271–281.
 26. Department of Health. Health of the Nation. Her Majesty's Stationary Office, London, January 1993.
 27. Wald NJ, Chamberlain J, Hackshaw A, *et al.* Report of the European Society of Mastology (EUSQMA) Breast Cancer Screening Evaluation Committee. *The breast* 1993; 2: 209–216.
 28. Chen HH, Duffy SW, Tabar L, Day NE. Markov chain models for progression of breast cancer. Part I: tumour attributes and the pre-clinical screen-detectable phase. *J Epi Bio* 1997; 2: 25–36.
 29. Chen HH, Duffy SW, Tabar L, Day NE. Markov chain models for progression of breast cancer. Part II: tumour attributes and the pre-clinical screen-detectable phase. *J Epi Bio* 1997; 2: 37–44.
 30. Swedish Cancer Society and the Swedish National Board of Health and Welfare. Breast-cancer screening with mammography in women aged 40–49 years. *Int J Cancer* 1996; 68: 693–699.
 31. News and Comments. The breast-screening brawl. *Science* 1997; 275: 1056–1059.
 32. Thomas DB, Gao DL, Self SG, *et al.* Randomized trial of breast examination in Shanghai: methodology and preliminary results. *J Natl Cancer Inst* 1997; 99: 355–365.
 33. Semiglazov VF, Sagaidak VN, Moiseyenko VM, Mikhailov EA. Study of the role of breast self-examination in the reduction of mortality from breast cancer. The Russian Federation/World Health Organization Study. *Eur J Cancer* 1993; 29: 2039–2046.
 34. Boyle P, Veronesi U, Tubianu M, *et al.* European School of Oncology Advisory Report to the European Commission for the "Europe Against Cancer Programme": European Code Against Cancer. *Eur J Cancer* 1995; 9: 1395–1405.
 35. Forrest P. Breast cancer screening. Her Majesty's Stationary Office, London, 1986.
 36. McCarrn J, Duffy S, Day NF. Predicted long-term mortality reduction associated with the second round of breast screening in East Anglia. *Br J Cancer* 2001; 84: 423–428.
 37. Gøtzsche PC, Olsen O. Is screening for breast cancer mammography justifiable? *Lancet* 2000; 355: 129–134.
 38. Olsen O, Gøtzsche PC. Cochrane review on screening for breast cancer with mammography justifiable. *Lancet* 2001; 358: 1340–1342.
 39. Nystrom L, Andersson I, Bjurstam N, Frisell J, Nordenskjöld B, Rutqvist LE. Long-term effects of mammography screening: updated overview of the Swedish randomised trials. *Lancet* 2002; 359: 909–919.
 40. Parkin DM, Whelan S, Ferlay, *et al.* Cancer incidence in five continents. Volume VII. Lyon, IARC Sci Publ, 1997.
 41. Morson BC. Gastrointestinal Pathology. Oxford, Blackwell Scientific Publications, 1979.
 42. Boyle P. Progress in preventing death from colorectal cancer (editorial). *Br J Cancer* 1995; 72: 528–530.
 43. Pisani P, Parkin DM, Bray F, Ferlay J. Estimates of the worldwide mortality from 25 major cancers in 1990. *Int J Cancer* 1999; 83: 18–29.
 44. Winawer SJ. A quarter century of colorectal cancer screening: Progress and prospects. *J Clin Oncol* 2001; 19: Suppl. 18, 6s–12s.
 45. Gregor DH. Diagnosis of large-bowel cancer in the asymptomatic patient. *JAMA* 1967; 201: 123–125.
 46. Winawer SJ, Fletcher RH, Miller L, *et al.* Colorectal cancer screening: clinical guidelines and rationale. *Gastroenterology* 1997; 112: 594–642.
 47. Mandel JS, Bond JH, Church TR, *et al.* Reducing mortality from colorectal cancer by screening for fecal occult blood. *New Engl J Med* 1993; 328: 1365–1371.
 48. Kronberg O, Fenger C, Olsen J, Jorgensen OD, Sondergaard O. Randomised study of screening for colorectal cancer with faecal-occult-blood test. *Lancet* 1996; 348: 1467–1471.
 49. Hardcastle JD, Chamberlain JO, Robinson MHE, *et al.* Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. *Lancet* 1996; 348: 1472–1477.
 50. Towler B, Irwig L, Glasziou P, Kewenter J, Weller D, Sllagy C. A systematic review of the effects of screening for colorectal cancer using the faecal occult blood test, Hemocult. *BMJ* 1998; 317: 559–565.
 51. Mandel JS, Church TR, Bond TH, *et al.* The effect of fecal occult-blood screening on the incidence of colorectal cancer. *N Engl J Med* 2000; 343: 1603–1607.
 52. Canadian Task Force on Preventive Health Care. Colorectal Cancer Screening. *Can Med Assoc Journal* 2001; 165: 206–207.
 53. Recommendations on Cancer Screening in the European Union. Advisory Committee on Cancer Prevention. *Eur J Cancer* 2000; 36: 1473–1478.
 54. Coeberg JWW. Challenges and pitfalls of mass-screening in the European Union. *Eur J Cancer* 2000; 36: 1469–1472.
 55. Autier P, Bleiberg H. Should there be mass screening using faecal occult blood tests for colorectal cancer. *Eur J Cancer* 1998; 34: 773–780.
 56. Lieberman DA, Weiss DG. One-time screening for colorectal cancer with combined fecal occult-blood testing and examination of the distal colon. *New Engl J Med* 2001; 345: 555–560.
 57. Detsky A. Screening for colon cancer - Can we afford colonoscopy? *New Engl J Med* 2001; 345: 607–608.
 58. Selby JV, Friedman GD, Quesenberry CP, *et al.* A case-control study of screening sigmoidoscopy and mortality from colorectal cancer. *New Engl J Med* 1992; 26: 653–657.
 59. Greenberg R, Baron J. Prospects for preventing colorectal cancer death. *J Natl Cancer Inst* 1993; 85: 1182–1184.
 60. Boyle P. Progress in preventing death from colorectal cancer (editorial). *Br J Cancer* 1995; 72: 528–530.
 61. Boyle P. Faecal occult blood testing (FOBT) as screening for colorectal cancer: the current controversy. *Ann Oncol* 2002; 13: 16–18.
 62. Mandel J. Colon and rectal cancer. In: Reintgen DS, Clark RA, eds. Cancer Screening. St Louis, Mosby, 1996; pp. 55–96.
 63. La Vecchia C. Faecal occult blood screening for colorectal cancer: open issues. *Ann Oncol* 2002; 13: 32–34.
 64. McArdle CS. Faecal occult blood testing for colorectal cancer. *Ann Oncol* 2002; 13: 35–39.
 65. Lowenfels AB. Fecal occult blood testing as a screening procedure for colorectal cancer. *Ann Oncol* 2002; 13: 40–43.

66. Bleiberg H. Hemocult should no more be used for the screening of colorectal cancer. *Ann Oncol* 2002; 13: 44–46.
67. Crespi M, Lisi D. Is colorectal cancer screening by faecal occult blood feasible? *Ann Oncol* 2002; 13: 47–50.
68. Strul H, Arber N. Fecal occult blood test for colorectal cancer screening. *Ann Oncol* 2002; 13: 51–56.
69. Autier P. Should there be organised faecal-occult-blood test screening established? *Ann Oncol* 2002; 13: 57–60.
70. Barry MJ. Fecal occult blood testing for colorectal cancer: a perspective. *Ann Oncol* 2002; 13: 61–34.
71. Chamberlain J, Moss S. eds. Evaluation of cancer screening. Berlin, Springer, 1996.
72. Reintgen DS, Clark RA. eds. Cancer screening. Mosby, St Louis, 1996.
73. Catalona WJ. Prostate cancer. In: Reintgen DS, Clark RA, eds. Cancer screening. Mosby, St Louis, 1996; pp. 97–117.
74. Chamberlain J, Melia J. In: Chamberlain J, Moss S, eds. Evaluation of cancer Screening. Berlin, Springer, 1996; pp. 117–136.
75. Collins MM, Barry MJ. Controversies in prostate cancer screening. Analogies to the early lung cancer screening debate. *JAMA* 1996; 276: 1976–1979.
76. Bartsch G, Horninger W, Klocker H, *et al.* Prostate cancer mortality after introduction of prostate-specific antigen mass screening in the federal State of Tyrol, Austria. *Urology* 2001; 58: 417–424.
77. Dominioni L, Strauss G. eds. International conference on prevention and early diagnosis of lung cancer. Italy, Varese, 1998.
78. Marcus PM, Bergstralh EJ, Fagerstrom TAM, *et al.* Lung cancer mortality in the Mayo Lung Project: impact of extended follow-up. *J Natl Cancer Inst* 2000; 92: 1308–1316.
79. Henschke CT, McCauley DI, Yankelevitz DF, *et al.* Early lung cancer action project: overall design and findings from baseline screening. *Lancet* 1999; 354: 99–105.
80. Henschke CT, Naidich DP, Yankelevitz DF, *et al.* Early lung cancer action project: initial findings on repeat screening. *Cancer* 2001; 92: 153–9.
81. Kaneko M, Kusumoto M, Kobayashi T, *et al.* Computed tomography screening for lung carcinoma in Japan. *Cancer* 2000; 89: 2485–2488.
82. Gould MK, Maclean CC, Kuschner WG, Rydzac CE, Owens DK. Accuracy of positron emission tomography for diagnosis of pulmonary nodules and mass lesions. A meta-analysis. *JAMA* 2001; 285: 914–924.