Pergamon

European Journal of Cancer, Vol. 35, No. 2, pp. 262–271, 1999

Printed in Great Britain
0959-8049/99/$ - see front matter

PII: S0959-8049(98)00379-7

Review

Rationale for Randomised Trials of Prostate Cancer Screening

The International Prostate Screening Trial Evaluation Group*

Screening for prostate cancer has been advocated by a number of organisations largely because there is good evidence that administration of the test for prostate specific antigen (PSA) results in the detection of cancers at an early stage. However, the mere fact that a cancer can be detected earlier in its natural history by screening is no guarantee that benefit will follow. Further, screening for prostate cancer can substantially impair the quality of life of those with detected and treated cancer, that would not otherwise have reduced life expectancy. The only established mechanism to evaluate the efficacy of screening is the randomised controlled trial. In this paper we review the trials contributing to our collaboration, the advantages that will flow from them, and the reasons why decisions on the introduction of population-based screening for prostate cancer cannot be made before these trials have come to fruition. © **1999 Elsevier Science Ltd. All rights reserved.**

Key words: prostate cancer, prostate specific antigen (PSA), screening, randomised controlled trials
Eur J Cancer, Vol. 35, No. 2, pp. 262–271, 1999

INTRODUCTION

IN AN EARLIER publication [1], we described our collaboration and what we hoped to derive from it. As we continue to recruit into the funded trials, and seek funding for those that have not yet received it, it becomes increasingly obvious that there are several issues that are still leading to confusion as to the reasons for undertaking randomised trials to determine whether or not screening for a disease is more beneficial than harmful in the population. In this paper, we present these issues for prostate cancer screening, and attempt to show how our collaboration will help to resolve them, and why we feel it is so important to take our trials to fruition. Failure to complete these studies may make it difficult to form firm conclusions concerning the value of population screening. The widespread introduction of early detection or screening in an unevaluated manner will 'contaminate' the study population and make it impossible to obtain the scientific answers required.

It is, therefore, the purpose of this paper to review the evidence that supports research to establish the efficacy of prostate cancer screening, and to describe in more detail the collaborative endeavor in which we are engaged.

THE BASIC PROBLEMS

To be effective, screening for cancer requires a valid and acceptable screening test, capable of detecting a sufficiently high proportion of cancers in the detectable preclinical phase (DPCP), and treatment for the detected abnormalities which is relatively free of side-effects and more effective at this time in the cancers' natural history than later.

This is a complicated series of requirements, and each must be valid for screening for prostate cancer to achieve its objective of reducing mortality from the disease. Unfortunately, we do not have the data available from research studies which confirm that each step of this process is effective. In spite of this, several individuals and at least one professional group (the American Urological Association) and one cancer

*F.E. Alexander†, Edinburgh; G.L. Andriole, St. Louis; G. Aus, Goteburg; A. Auvinen, Helsinki; P. Beensterboer, Rotterdam; B.G. Blijenberg, Rotterdam; R. Bonardi, Florence; S. Buys, Salt Lake City; B. Candas, Quebec; S. Ciatto, Florence; E.D. Crawford, Denver; R. Demers, Detroit; H.J. de Koning, Rotterdam; L. Denis, Antwerp; M.A. Dijk, Rotterdam; F. Figueiredo, Lisbon; E.P. Gellman, Washington; J. Gohagan, Bethesda; M. Hakama, Tampere; J. Hugosson, Goteborg; W.I. Kirkels, Rotterdam; R. Kranse, Rotterdam; Th.H. van der Kwast, Rotterdam; H. Lilja, Lund; L. Määttänen, Helsinki; J. Mandel, Minneapolis; A.B. Miller†, Toronto; H. Monteiro, Lisbon; V. Nelen, Antwerp; A. Oberman, Alabama; A. Paez, Madrid; P.C. Prorok, Bethesda; D. Reding, Marshfield; J.W.B.W. Rietbergen, Rotterdam; D.B. Rose, Antwerp; A.B. Sanchez, Madrid; F.H. Schröder, Rotterdam; E.C. da Silva, Lisbon; N. Simpson, Bethesda; P.H. Smith, Leeds; B. Stadaert, Brussels; U.H. Stenman, Helsinki; L. Teppo, Helsinki; A. Turkes, Cardiff; J.L. Weissfeld, Pittsburgh; L. Yokochi, Hawaii.
†Writing Committee.
Correspondence to A.B. Miller, International Agency for Research on Cancer, 150 cours Albert Thomas, 69372 Lyon cedex 08, France, e-mail: miller@iarc.fr
Received 6 May 1998; revised 18 Sep. 1998; accepted 26 Oct. 1998.

society (American) have concluded that the data are sufficient to promote screening. However, several other groups, including the American College of Physicians, the U.S. Preventive Services Task Force, the Canadian Task Force on the Periodic Health Examination, the National Cancer Institute of Canada, many European health authorities and the U.K. Department of Health, have concluded that the requisite data are not available. Indeed, it has been suggested that the adverse effects of prostate screening are large enough to make research on the issue unethical [2]. The main reason for these discrepancies is that for some, the fact that a screening test finds early cancer in an individual is sufficient to conclude that the test will be beneficial. However, in practice, this does not tell us if that individual has been helped by finding the cancer 'early', as explained below, nor whether screening is beneficial for the population as a whole.

For prostate cancer, there are three potential screening tests, digital rectal examination (DRE) by a trained health care professional or urologist, a blood test for prostate specific antigen (PSA) and transrectal ultrasound (TRUS). Treatment believed to be potentially curative includes radical prostatectomy and radical radiotherapy. Although conservative therapy (watchful waiting) has been recommended [3], and is appropriate for those whose disease has a very low probability of progression, such a policy can only be curative for those with progressive disease if the fact of progression is recognised at a stage when radical prostatectomy or radical radiotherapy is still applicable. Currently, a trial is ongoing in the U.S.A., the Prostatectomy Intervention versus Observation Trial (PIVOT), and another in Scandinavia, designed to evaluate whether radical prostatectomy leads to improved survival when compared with a policy of observation in patients with a clinically localised prostate cancer.

There is no question that the screening tests for prostate cancer are capable of detecting asymptomatic disease, although there are considerable reservations whether DRE achieves much in the way of early detection. TRUS has been advocated for screening, but it is expensive, and now, almost universally, it is reserved for use as part of the process of further investigation in those with an abnormal PSA. Previously it was used as a screening test in two centres in the early years of the European Randomized Study of Screening for Prostate Cancer (ERSPC) in order to evaluate its value. In general, therefore, we will refer to the PSA screening test.

As implied above, the mere fact that a PSA screening test will detect prostate cancer early does not mean that those who have had their cancers detected as a result of the test have received benefit. There are two reasons for this, representing opposite types of disease natural history. Firstly, the cancer may not be curable nor have its natural history modified by available treatment. Secondly, the cancer might never have become life-threatening during the patient's lifetime. Even when a screening test detects an 'early' cancer, this is in fact very late in the life of that cancer in that individual. We do not know at what point a prostate cancer produces sufficient PSA for the elevation to be detectable, nor whether this is associated with the biological potential to progress and to metastasise. However, PSA levels appear to be associated with tumour volume and grade [4]. For cancers that have an inbuilt tendency to metastasise early, the tumour volume may still be quite low at the time metastatic spread has occurred. Thus, it is probable that many such cancers will not have an elevated PSA at the time metastases occur.

In practice, people who have their cancers detected by screening fall into one of four different groups, although at the time of their detection it is not possible to determine which group. Only the passage of time provides this information with certainty. These four groups are:

Group A

People whose cancer would have been cured by treatment given after clinical diagnosis even if the cancer had not been detected by screening. The screening test has found the cancer earlier, but has not extended their life, they would not have died of their cancer anyway.

Group B

People whose cancer is incurable because at the time it was detected by the screening test, they already had incurable locally extensive or metastatic disease. For some of these, the screening test may have detected the tumour at a stage where prolongation of life is possible with modern therapy, and if so, this may provide an opportunity for death to occur from a competing cause. For the remainder of this group, no prolongation of life is possible with therapy and, therefore, the screening test has not extended their life, they would have died of prostate cancer anyway.

Group C

People whose cancer would not have been found in the absence of the screening test, but who would have died of other causes before they had lived long enough for the cancer to show itself with symptoms. The screening test has not extended their life, they will still die of another cause, but the quality of their remaining life has been reduced because they have been labelled with, and possibly treated for, cancer.

Group D

People whose cancer would have been incurable if it had been diagnosed after symptoms had developed, but is curable if found early as a result of its detection by a screening test. These are the main group of people who benefit from the screening test, i.e. death from the cancer detected by screening has been averted.

In the population at the time the screening test is administered, there are likely to be people who have already developed prostate cancer, but whose cancer is not detected by the screening test, either because they have not yet entered the DPCP, or because they had a false-negative screening test. If the test is re-administered later, many of these will probably be detected and will then fall into one of the four groups described.

We do not know the proportions that fall into these four groups of detectable prostate cancer, although because of the long natural history of the disease, in countries where there is widespread interest in the early diagnosis of cancer many may fall into Group A, while because of the older ages of many screened, many will fall into Group C. So the opportunity for a large proportion to fall into Group D is bound to be small.

But what constitutes Group D? Is it those with a low, intermediate, or high PSA? It seems unlikely to comprise people with a high PSA as they are usually found to have advanced disease and are not amenable to treatment by radical prostatectomy or radiotherapy. Thus, the assumption will be that benefit resides with those with intermediate or low PSA levels that can be treated radically.

This brings us to the issue of treatment. Because of the absence of completed randomised controlled trials, there is currently no conclusive evidence that radical treatment for prostate cancer prolongs life, although most clinicians believe this and some observational data suggest that the balance of evidence may have shifted to suggest that early radical prostatectomy improves outcome [5, 6]. However, evidence derived from clinical series is not sufficient, we need evidence from patients with screen-detected cancers, as screen-detected cancers may be more likely to respond to existing treatments than clinically diagnosed cancers. We cannot obtain such evidence from randomised trials of screen-detected cancers, as it would be regarded as unethical to screen someone for cancer and then fail to treat it, especially if he was eligible for radical therapy, although of course different forms of radical therapy could be compared among screen-detected cases. It is the randomised trials of screening that provide the answer we need on the efficacy of treatment of screen-detected cancers. If prostate cancer mortality is lower in the group randomly allocated screening than in the control group, the treatment given to those in Group D must have been effective. It is because the answers sought to the effectiveness of treating screen-detected cancers can only be obtained from screening trials that screening trials do not need to be delayed until the answer to the treatment question is obtained from clinical series.

EVALUATION OF SCREENING TESTS

The validity (i.e. the sensitivity and specificity) of PSA for prostate cancer screening is not yet established. Indeed, we expect that the current screening trials will provide the necessary data. To reach an agreed determination on sensitivity we have to agree on the type of disease that the screening test is intended to detect. In general in the screening literature, this is any disease in the DPCP, as the cancers in Group D will be in this phase of the disease. A screening test that only detected disease in its clinical (symptomatic) stage would be virtually useless, and this may be true for DRE. There are some indications that the sensitivity of the PSA test for preclinical disease is less than 80%, possibly as low as 60%. We are used to assuming that the diagnostic tests applied to those with an abnormal result from a screening test are capable of detecting all the true-positives that are present. For prostate cancer screening this may not be true, as some of those who are reported negative to all diagnostic tests (DRE, TRUS and sextant biopsies) and, therefore, assumed to be false-positives, could be true-positives with as yet undiagnosed disease. This should be clarified by extended follow-up of those with abnormal screening tests in our collaboration. However, the majority of those who are regarded as not having disease after the diagnostic process is completed will be false-positives, and these determine the specificity of the test. Current data suggest that the specificity is no more than 90%. The values of sensitivity and specificity cited are dependent on the cut-off point used to declare the test positive. Reducing the cut-off point from 4 ng/ml (as is often recommended) to 3 ng/ml or lower will increase the sensitivity of the test, but reduce the specificity, to the order of 85% or less. Such low specificity increases substantially the cost of screening. A specificity level of 85% means that large numbers of men will receive additional investigations, and many random prostate needle biopsies, approaches that are not complication free. Determining an appropriate cut-off point for PSA will only be possible once the current generation of randomised trials are completed, and indeed this is one of the strengths of our collaboration (see below).

However, these issues are further complicated by the increasing propensity for men to develop latent prostate cancer as they age, with autopsy series suggesting frequencies for men in their 60s of 20–30%, and higher for older men. Even if PSA screening is not sensitive to cancers of such low biological potential (as several claim), under the circumstance of a specificity of 85% or less, several men with latent prostate cancer will be diagnosed as having disease if they are subjected to random prostate needle biopsies purely by the chance of the needle hitting such latent cancer foci. Thus, it is probable that some of the prostate cancers found following PSA screening are potentially latent (i.e. in Group C above) , and, thus, effectively overdiagnosed. Evidence is accumulating from the retrospective evaluation of stored bloods tested for PSA [7], that those with a cut-off point of 10 ng/ml or more progress to clinically detectable cancer at a mean duration substantially less than those with a PSA level of < 10 ng/ml. Those with a PSA level of < 3 ng/ml have a longer mean time to progression than those with a PSA of 3–10 ng/ml. However, these means do not fully describe the distribution of progression times. There are some who never progress at any PSA level, and some who progress within a few years. We have no idea which of the patients will benefit by early detection, but clearly it cannot be those with cancers that would never have progressed if left undiagnosed. The general assumption is that it will be those with a low PSA level, rather than those with a high, but as discussed above such assumptions can be challenged.

THE CONTRIBUTION OF RANDOMISED CONTROLLED TRIALS

In assessing the impact of treatment of disease, it has become customary to compare the survival of patients treated with the new treatment with those treated with the old or standard treatment (which may be no treatment). Randomised controlled trials are accepted as the best methodology for this, as they avoid the biases inherent in comparison of outcomes between different institutions, or historical comparisons within the same institution. They should also ensure that those treated with the compared methods are identical for known and unknown confounders.

Using survival as the endpoint and the start of cancer treatment as the time of entry is satisfactory when it is possible to assemble a group of patients with cancer at comparable stages in their natural history, as is the situation for clinically diagnosed cancers carefully categorised (and stratified) by stage and if possible grade, and then allow the play of chance (randomisation) to decide which treatment group they are in. Randomisation, providing the numbers of units are large, will ensure balance of unrecognised as well as recognised prognostic factors that could act as confounders between the two arms of the trial. However, this balance is not possible when the outcome for screen-detected cancers is compared with clinically diagnosed cancers, as the process of screening, if at all effective in early detection, will find cancers at earlier stages in their natural history than the clinically diagnosed cancers. The biases that impact on survival and make it invalid for assessing the impact of screening are lead time, length bias, overdiagnosis bias and selection bias.

Lead time is the amount of time that diagnosis is advanced by the application of screening. We do not know at this time

the average lead time gained by PSA screening, but comparison of the detection rate on screening with the incidence rate expected in the population suggests it is at least 6 years, and in many cases much longer. For some cases the lead time extends beyond the lifetime of the individual, in that case this is equivalent to overdiagnosis bias. Of course, there is a distribution of lead times. Some cases, detected late in their natural history or with a short DPCP, will have short lead times. They are likely to be cases with relatively advanced disease, although it is possible, as discussed above, that they could benefit from screening. In any case it is not possible to correct survival data for lead time in individual cases as it is unknown.

Length bias results from the fact that screening occurs intermittently. People with rapidly progressive disease, that progress quickly through their DPCP, may completely miss being detected by screening, whereas people with more slowly progressive disease will be detected. Thus, compared with a clinical series, a screen-detected series will have a longer mean DPCP and, thus, potentially a longer survival, as many of the short survival cases will not be present because of length bias.

People who agree to be screened are a self-selected group of the population, often those who are aware of the disease and are health conscious, and who might not allow intrusive urinary symptoms to persist so as to permit a prostate cancer to grow to such a size that it would become incurable. Hence, the potential for screening to benefit them may be reduced. This is an example of selection bias. It is very likely that for prostate cancer screening selection bias will result in an increase in survival of screen-detected cancers. For the same reason, it is not valid to compare the outcome of men who accept the offer of screening with a control group, unless it has been demonstrated that all members of the control group would also have accepted the offer of screening.

Overdiagnosis bias is the result of a screen-detected lesion being labelled as cancer, that would not have progressed to be clinically diagnosed in the absence of screen detection. There are several reasons to believe that it is likely to be very important in prostate cancer screening, especially the very high detection rates of prostate cancer reported in many studies compared with the numbers expected from the incidence rates of the disease in the population.

Case survival, therefore, cannot be used to assess screening. The correct outcome is population mortality from the disease over a follow-up period beginning with randomisation, that is, the numbers of deaths from prostate cancer related not to the cases that occur, but to the total population offered screening.

The only way to determine the degree of benefit without bias is by comparing people offered screening with a group of truly comparable people who are not offered screening. We cannot make a valid comparison by comparing people screened with those who were unscreened in the past. This is because we cannot determine who, among those unscreened in the past, would have agreed to be screened, and who would have refused the test. Often an offer of screening is, at least initially, only accepted by a relatively small proportion of the population. These are highly selected people. Making a 'historical' comparison, as just described, would be fraught with uncertainty and could not be relied upon for valid policy decisions on whether to screen or not. Making comparisons in one area with everyone in the past and everyone now including those screen-detected and not is another historical approach. This may be more valid than the first method, but still fraught with uncertainty, as many factors change with time, including approaches to the treatment of cancer, and these could impact on cancer mortality. Making a comparison between people screened in one area, including not only those who accept screening but those who refuse it, with people in another area would be a slightly better approach. Even this approach remains unsatisfactory. We cannot be certain that the people in the two populations compared are truly similar in all the aspects that impact upon cancer occurrence and death from the disease.

Methodologies used in observational epidemiology, particularly case–control and cohort studies, are sometimes used to evaluate screening. Both approaches require that screening has been in place in a community for a sufficient length of time for a benefit to be detectable if it does occur. Thus, these methods are not applicable at the present time for evaluation of PSA screening, as almost certainly it has not been available sufficiently long for an effect on mortality from prostate cancer to be demonstrable. However, even if we were to plan to wait and then use such approaches in the future, they could not be conclusive. This is because they both suffer from the effects of selection bias, and this is not readily corrected. Case–control studies assess the screening histories in those who die from the disease with controls who have not died from the disease drawn from the population from which the cases arose. However, the controls may have had more screening than those who die of the disease because they were health conscious and self-selected themselves for screening. In cohort studies, the outcome in a self-selected group of people who were screened is compared with the outcome in the general population. Again, the effect of self-selection cannot be controlled. This is the major reason that case–control or cohort studies of screening cannot replace randomised controlled trials of screening, which compare by intention to screen, the intention being made unbiased by the random allocation.

The only methodology that avoids all biases is the randomised controlled trial. Randomised trials of screening are carried out by one of two approaches, randomisation of people to be invited to be screened or not, or randomisation of a group all of whom volunteer to be screened (here it is assumed that the majority will attend screening, if allocated to the intervention arm). The latter approach is more akin to an efficacy trial, the former to an effectiveness trial.

Randomised trials of invitation to be screened are usually randomised population-based trials, in which through a register that identifies by age those in the population, the relevant members of the population are randomised to be invited to be screened or not, and site-specific cancer mortality compared in the groups randomised, whether or not those in the intervention group accepted the invitation, or those in the control group were screened (an 'intention to treat' analysis). Trials of this type are not practical in the absence of a population register. Unless the register serves only a segment of the population, they are more readily generalisable than volunteer trials. They incorporate within their design less than 100% compliance with screening in the intervention group, but this should not be a major problem unless compliance is substantially less than 70%, as those who accept the invitation to be screened will often be at higher risk of the disease than those who decline. Consent is typically not

obtained prior to randomisation. Pilot studies are usually required to determine the likely compliance, as this has an important impact on the planned sample size. Contamination is usually low in such trials, because the controls have not been alerted to the possible availability of screening. However, this is not invariable, as there is no mechanism to ensure that information on an individual basis is provided on the reason why screening is not currently recommended, as is possible in a trial of volunteers who give informed consent prior to randomisation. Further, there is usually no mechanism to obtain risk factor information on the control series and, thus, confirm the validity of the randomisation. Instead, it has to be assumed that the randomisation of large numbers has ensured equal distribution of risk factors. Another difficulty, with no individual contact with the controls, is inability to confirm at baseline that those randomised had not been previously diagnosed with prostate cancer. This will be possible for those in the intervention group who attend for screening, but such data cannot be used to exclude those with previously diagnosed cancers, as otherwise a bias in favour of demonstrating an artefactual mortality reduction in the intervention group will be introduced. One mechanism, which we favour, is to base such exclusions on linkage of the records of all trial participants to a population-based cancer registry or registries covering all the areas from which subjects are drawn. This will identify those subjects included in the trial who had a diagnosis of prostate cancer prior to their date of randomisation (known from the study records) in the trial. Another, is to determine for all deaths due to prostate cancer the exact date of diagnosis. Those diagnosed prior to the date of randomisation will be excluded from the analysis.

Trials based on volunteers have to be performed when it is deemed unethical for an individual to be randomised and/or placed in a control group in a human experiment without their consent (as in Canada, the Netherlands and the U.S.A.). They are also essential when it is likely that only a small proportion of those that might be invited to be screened would comply with the invitation (as would be usual in North America). Further, they are necessary when there is no population register that identifies people by age to be used as a sampling frame for randomisation, or when it would be impractical to develop such a register through the records of general practitioners. Randomised trials of volunteers are efficacy trials, they have internal validity, but may not be generalisable to the whole population, although they are often representative of those who readily participate in screening, essentially a voluntary activity. Randomised trials of volunteers not only permit full identification of participants and collection of data that will facilitate their being traced, but will usually enable data to be collected on risk factors for prostate cancer, and sometimes biological specimens from the control as well as the intervention group. They have the additional advantage that there is usually very close to 100% compliance with the first screen, but the disadvantage that contamination may be high because of disappointment in the controls that they were not randomised to be screened, although fully informed consent should serve to diminish this. In such a trial, healthy people are approached for their willingness to participate in a research study of screening. They understand that at the time the trial is started, although there may be evidence that screening tests can find cancer, it is not known if they are effective in finding cancer sufficiently early that death from cancer is prevented. The people approached to consider being in the trial are, therefore, told about current evidence, and then, if they have not had prostate cancer already, are asked to agree to being randomised. The person approached cannot choose to be screened, just accept the offer to participate in the trial, with the result that they would have, for example, a 50:50 chance of receiving the screening tests. Those who are randomised to be screened can then have the tests. Those randomised not to be screened do not.

In both types of trial the randomised groups are carefully followed, including those who do not accept the invitation to be screened, usually for 10 years or more, and whether they develop prostate cancer, and what eventually happens as a result of treatment, is carefully ascertained. After a defined interval, it will be possible to determine whether screening results in the reduction of prostate cancer mortality, and what proportion of people with cancer in an unscreened group would benefit were screening to be offered.

INTERPRETING TRENDS IN PROSTATE CANCER INCIDENCE AND MORTALITY

There have been some attempts to interpret recent trends in mortality from prostate cancer in the U.S.A. with regard to possible benefit from PSA screening [8]. If prostate cancer screening is effective, and a high proportion of the 'at risk' population is screened, mortality from the disease will eventually fall. However, it is highly unlikely that recent trends can be interpreted in this way. In the U.S.A., age standardised prostate cancer incidence shows a substantial rise, slow from 1973 and then more rapid from about 1987 in caucasians and 1989 in blacks. A decline in caucasians commenced in 1991 and in blacks in 1992. Age-standardised mortality from prostate cancer showed a very slow rise in caucasians from 1980, and stability in blacks, and then a more substantial rise in both races from 1988. A decline in caucasians commenced in 1991 and in blacks in 1992, and the rates are currently back to the 1988 level.

The rise in prostate cancer incidence was probably initially due to increasing use of transurethral resection of the prostate, the subsequent increase to use of PSA, but the changes in prostate cancer mortality have probably been too rapid to reflect effects of screening in the population, although they may in part reflect improvements in therapy. The fact that the trends in incidence as well as mortality occurred at the same period argues against an effect of screening, as we would expect mortality to fall after incidence, probably after at least a 5 year interval. These simultaneous effects suggest that changes in therapy cannot be the sole explanation for the recent mortality falls either. Rather, an artefactual explanation seems likely to be contributing, associated with changes in medical practice, especially over management of presumed benign prostate abnormalities, the introduction of PSA screening and recognition of greater curability of prostate cancer than formerly believed, accompanied by a labelling effect with individuals diagnosed as prostate cancer being more likely to be certified as dying from it. Recent falls in incidence are probably due to the reduction in new recruits to PSA screening and the lower detection rate associated with rescreening.

These considerations make it imperative that the trials comprising our collaboration continue to fruition. If the trends in incidence and mortality were to be misinterpreted by those responsible for decisions on funding, we might never learn whether or not the PSA test does reduce prostate cancer mortality, and if it does at what level and cost. If so, there is a

real possibility that a procedure leading to much harm without commensurate benefits could become part of routine health care.

STATISTICAL ISSUES

In designing randomised trials, it is important to decide what level of benefit it is reasonable to expect, and what level of benefit would be important in practice. This enables computation of how many individuals will have to be recruited to demonstrate whether or not such a level of benefit exists. Given the caveats relating to the different types of cases detected by screening discussed above, it is likely that the degree of benefit will not be large, say a 20% reduction in prostate cancer mortality in the screened group compared with an unscreened group at 10 years. Even if that is the actual benefit from screening, in a population-based trial, the observed benefit in the group offered screening will be less, to the extent that those invited fail to attend for screening. Detecting such relatively small benefits is very difficult, as it requires a carefully designed trial with large numbers of subjects recruited. Even so, such a reduction would have an important public health impact. However, if a statistically significant reduction in prostate cancer mortality of this order was demonstrated, this estimate will have a range of confidence about it. Such a range is usually set at the 95% confidence level. The narrower the confidence interval, the more our confidence that the point estimate is close to the truth. To achieve such precision, larger sample sizes will be required, such that we can say, for example, that we demonstrated a 20% reduction in prostate cancer mortality, and the 95% confidence interval around that estimate is say, from as much as a 30% reduction, to as little as a 10% reduction.

The calculation of numbers of men that will have to be recruited into the trial is based on the required number of events, i.e. the number of prostate cancer deaths in the control arm and, thus, depends upon a number of factors. First, one must bear in mind that those recruited to a screening trial do not have a history of previously diagnosed prostate cancer. Yet, it is in these individuals that the majority of deaths from the disease will accumulate in the base population in the next few years. In the trial, cases of prostate cancer will have to occur, and they will have to run through their natural history, whether or not that is influenced by screening and treatment, before deaths occur. So determining the expected number of prostate cancer deaths in a group of men recruited into a screening trial requires computing the expected incidence of the disease, the expected survival of the men from prostate cancer and the influence of competing causes of death, before the expected numbers of deaths in the control group can be determined.

Another factor is the age range of the men to be recruited, and the expected proportion of subgroups by age. The younger the average age of participants in a trial, the larger the sample size required in terms of numbers of men that will have to be recruited to yield the required number of events. However, younger men have a longer life expectancy, so that the potential years of life saved from a successful screening strategy would be greater than for the older participants. Further, there is a greater tendency not to propose radical treatment for older men, especially over the age of 70 years, because of the lesser average life expectancy. Nevertheless, prostate cancer is not an optimal route to death, and even at older ages in men largely free of co-morbidity there will be benefits from preventing death from the disease. Further, with increasing proportions of relatively older men in all our populations, it becomes increasingly difficult, and in some countries unacceptably discriminatory, not to evaluate means to prevent or delay a very important cause of death. Clearly, however, subjects should only be recruited who appear to have sufficient life expectancy to derive benefit from screening, should such a benefit exist. That implies a minimum life expectancy on entry of 10 years.

There has to be a compromise, therefore, between recruiting samples of older men because more deaths would be expected and recruiting younger men, who might derive more benefit in terms of life years gained from screening if screening is beneficial. In general, the investigators in our collaboration have tended to aim to recruit men with the lower age boundary ranging from 50 to 60 years and the older from 65 to 74 years.

A further consideration relating to the required numbers of subjects in a trial is the extent to which subgroup analyses will be conducted, e.g. by age, ethnicity or race, or other risk factors for the disease. Few trials can be large enough to conduct large numbers of subgroup analyses, which in any case can usually only be exploratory (hypothesis generating), unless they are an integral part of the design of the trial. However, it is one of the strengths of our collaboration, discussed in more detail below, that more subgroup analyses will be possible than in the individual trials.

Further factors that influence the power of a screening trial include compliance with screening, the extent of screening in the control group (contamination) and the efficiency of follow-up of randomised subjects. The first two of these factors may have to be estimated in prior feasibility studies. If during the course of the trial the estimates are found to be in error, midcourse corrections to the numbers to be recruited may be needed.

Once decisions on the age group to be entered have been taken, knowing the expected incidence and survival from prostate cancer in the relevant area, and having decided on the degree of mortality reduction that we should like to demonstrate (or more accurately exclude, as the statistical tests that will be applied are designed to determine whether we can reject the null hypothesis), what subgroup analyses will be performed and what compliance and contamination we can expect, we can determine the required sample size in the trial.

THE TRIALS COMPRISING OUR COLLABORATION

The US PLCO trial is based on volunteers who sign a consent form to be randomised, and has a planned sample size of 74 000 men aged 55–74 years, half randomised to receive annual PSA and DRE screening to a total of four screens and half to usual care in the community. The PSA cut-off is 4 ng/ml. The PLCO trial was planned on the assumption that it would have 90% power to show a statistically significant 20% reduction in prostate cancer mortality at 10 years. This benefit is sought in the presence of whatever non-compliance and contamination exists in the population. Currently the expectation is for 90% compliance with screening, and approximately 12% contamination in the usual care group. The latter may be optimistic, given the extent to which prostate cancer screening tends to be advocated in many of the areas where the screening centres operate. To date over half the planned sample has been recruited.

The Canada Quebec trial is population-based. A total of 46 193 men aged 45–80 years were identified through electoral lists as residing in the area of Quebec city, 30 956 were randomised to be invited to be screened, of which 7155 accepted the invitation and were screened. In the control group, 982 men were screened. The cut-off point is a PSA level of 3 ng/ml. Rescreening is annual. A recent analysis [9] compared the death rate from prostate cancer among those screened with those unscreened (combining those unscreened in the control group with those who did not accept the invitation to be screened). There was a 69% lower prostate cancer mortality in the screened than in the unscreened men. However, there was no reduction in prostate cancer mortality if those randomised to be invited to be screened was compared with the control group not invited (the intention to treat analysis, a crude ratio of 1:1) [10]. It is only the intention to treat analysis which avoids selection bias; the plan is that all IPSTEG collaborative analyses will be restricted to intention to treat comparisons.

The European Study of Screening for Prostate Cancer so far has centres in The Netherlands, Belgium, Finland, Italy, Portugal, Spain and Sweden. Additional centres in France, Norway and Switzerland are under consideration. The trial as a whole has been planned with a total sample size of approximately 190 000 men. Given the different expected incidence and mortality rates, and the different designs in the subsegments, as described below, the trial is expected to attain 90% power to detect a statistically significant reduction in prostate cancer mortality of the order of 20% at 10 years.

In Rotterdam, The Netherlands, it is planned to recruit 40 000 volunteers aged 55–74 years. Initially, DRE, PSA and TRUS were offered. However, with the experience from the first 23 000 randomised it has now been decided to concentrate on PSA screening, with a cut-off for investigation of 3 ng/ml. Rescreening after 4 years is planned to a total of two screens.

In Antwerp, Belgium, men aged 55–74 years selected from population registries are randomised, with a planned target of 30 000. The initial approach was similar to Rotterdam, with nearly 9000 men screened to date. A cut-off for PSA screening of 3 ng/ml will now be adopted. Rescreening every 4 years is planned to a total of two screens.

Finland (Helsinki and Tampere), population registries are the bases for randomised invitations (in a ratio of 1 screen:2 control), aged 55–67 years. Approximately 43 000 have been randomised, and 10 000 screened, with a planned total sample of 80 000. Rescreening every 4 years is planned to a total of three screens, except for those enrolled at 67 years of age who will have two screens.

In Italy (Florence), the population registry is the basis for randomisation, with a total of 15 000 men age 50–69 years planned to be randomised. Screening started in October 1996. So far 2851 controls have been included, 2931 allocated screening invited, of which 1788 have been screened. A cut-off of 4 ng/ml will prompt random biopsy. Biopsy will also be performed for suspicious abnormalities found at DRE/TRUS, routinely performed in subjects with PSA between 2.5 and 4 ng/ml. Rescreening every 4 years is planned to a total of three to four screens.

In Portugal, recruitment is through general practitioners, with a target group of 15 000 men, aged 50–74 years. So far 2059 men have been enrolled, 1027 to the control group, and 1006 screened. Rescreening annually is planned to a total of four screens.

In Sweden, population registries in Malmo and Gothenburg are the bases for randomisation, at a ratio of 1 invited:2 control; the target group is 32 400 men aged 50–66 years. So far 32 290 have been randomised; 22 298 to the control group, 10 000 to the screen group with 9800 invited and 5846 screened. Rescreening every 2 years is planned to a total of three screens.

THE OVERVIEW ANALYSIS: DESIGN ISSUES

The strength of advance planning and co-ordinated quality control

Overview analyses using individual data have several advantages over meta-analyses of published data in groups, or even analyses of data specially provided by the investigators of studies to the analysts, as some of us have provided for recent analyses of breast cancer screening. These analyses, however, were almost invariably conducted after the results of the individual trials were known, and this appears to have influenced the structure of some of these 'meta-analyses', even to the extent of selectivity in the trials to be included or excluded. Overview analyses require that individual subject records are available, provided to a standard format, so that they can be shown to be consistent, or if not entirely so, control can be attempted for some of the differences between studies. A useful overview analysis in the screening literature is that conducted of the Swedish trials of breast cancer screening [11]. A particular feature of this analysis was that all the possible breast cancer deaths were reviewed by a special panel preparatory to the analysis, so that consistency and blindness in determining the underlying cause of death was ensured. However, many other relevant features differed between the trials and historical reconstruction was often not possible.

We have progressed several steps further in our plan for an eventual overview analysis of prostate cancer screening trials. Early on during the initiation of the trials, meetings were held to plan an eventual overview analysis of findings. This has enabled us to plan to collect similar data, and pay particular attention to quality control issues [1]. We are addressing issues related to comparability of standards for PSA testing, facilitated by the common use of a test from one manufacturer, the Hybritech (Europe SA, Liege, Belgium) tandem-E assay, as well as exchanging standards between the national laboratories performing the tests. We are attempting to obtain uniformity in staging classifications, similar classifications of pathology with exchange of material planned and uniform application of criteria for cause of death review. We are well on the way to defining a minimal data set for the purposes of the final overview analysis, and we have addressed issues such as publication policy. We plan to meet on an annual basis, to ensure that the advantages gained through this international collaboration are maintained.

The main differences between the trials

In spite of these advantages derived from advance planning, there are several differences between the trials that will have to be borne in mind if and when any heterogeneity in the findings is eventually identified. In part because of different standards in requirements for human experimentation between the countries participating, some trials are based on volunteers, enrolled after informed consent, and some on population randomisation followed by invitations to screening. As described above, several centres in the European Randomized Study of Screening for Prostate Cancer are based on randomised invitations, including the Finland, Italy,

Portugal and Sweden centres, the Canada Quebec trial and a proposed U.K. trial, whilst the volunteer trials in this collaboration are the US. PLCO trial and the participation from Rotterdam and Antwerp in the European Trial of Prostate Cancer Screening.

The second major difference between the trials is in the baseline mortality from prostate cancer. The countries with the highest mortality are Sweden, Finland, The Netherlands, Belgium and the U.K., the lowest Portugal, Spain and Italy, with the U.S.A. and Canada intermediate. These differences will impact on the power of the studies, but providing in the overview analysis that comparisons are always stratified by country and trial, will not affect the validity of the analysis. It is likely that the differences in underlying mortality may be reflected in differences in apparent sensitivity. This is because the underlying differences in incidence that lead to the mortality differentials may be reflected in different ratios of progressive to latent prostate cancer. Although if PSA in practice detects few truly latent cancers, the fact that all centres use the same PSA test will minimise these differences. However, there are likely to be differences in interval cancer rates, and this may be reflected in different sensitivities, since estimates of sensitivity depend on comparison of interval cancers with expected incidence in the absence of screening. Fortunately, the expected incidence will be derivable from the control groups, as basing expected incidence on population based registers in the country would be invalid for the volunteer-based trials, and would be dependent on efficiency of registration and representativeness of the area in which the trial was conducted for the population-based trials.

A third difference in outcome could relate to different policies over application of radical therapy or watchful waiting of confirmed disease and immediate or delayed hormone therapy for more advanced disease. However, all centres encourage active treatment of screen-detected cancers and any residual effects due to differences in treatment policies can be regarded as a strength of our collaboration, as further discussed below. One of the requirements for our uniform database is therapy information which will document the extent to which screen-detected cancers are treated actively, and the extent to which cancers in the control group receive similar therapy according to stage and grade at detection and age.

There are also differences in the trials in the frequency of rescreening, ranging from annual PSA tests in the US, Canada and Portugal to 4 yearly in The Netherlands, Belgium, Finland and Italy. We do not know how much of any benefit from screening relates to the first or prevalence screen, compared with subsequent screens. Although these differences may result in different estimated effectiveness, exploring this in the overview analysis will help to determine the optimal frequency of rescreening.

The other difference between the trials is the fact that two of them, Finland and Sweden, randomise larger numbers of control subjects than in the intervention arms, a process which increases power with minimal increases in costs. Again, this will not affect the validity of the overview analysis.

Finally, there are likely to be differences between the centres in compliance and contamination rates. The extent of these differences is not known at present, but it is possible that there will have to be some attempt to control for such differences in the overview analysis.

Other potential differences we hope to overcome as our collaboration matures, by agreeing upon common (uniform) criteria, include differences in pathology classification and potential differences in cause of death classification.

THE OVERVIEW ANALYSIS: STRENGTHS

There are many questions over prostate cancer screening that one would like to evaluate in carefully designed research studies, if it is demonstrated that screening is effective in reducing prostate cancer mortality. These include the optimal age to initiate screening, the optimisation of the screening procedure, the frequency to rescreen after the initial screen and the appropriate age to terminate screening. A single multicentre trial with randomised block design incorporating evaluation of such parameters would be preferable to many trials looking at different segments of these questions, but such a trial would be impracticable and probably unacceptable to funding agencies because of its size and expense. The complexity of such a trial might also make it unacceptable to the population to be recruited into the trial.

That being so, the only practical way to obtain answers to such questions is to draw comparisons between different trials with different approaches to these issues. Therein lies some of the major strengths of the planned overview, although perhaps the major strength is the increased sample size that will be available to address the mortality question.

Sample size

The exact total sample size that will eventually be available from our collaboration is not precisely known, but will almost certainly exceed 300 000 men enrolled, with a minimum follow-up period of 10 years. Even though some of the individual trials are being planned within a projected sample sufficient to detect a 20% reduction in prostate cancer mortality, the 95% confidence intervals surrounding the point estimate of effect will be quite wide. The overview will permit far greater precision on the estimate of effect. Further, if it transpires that the true effect is less than 20%, the overview may permit determination of the likely range of effect and, thus, facilitate policy decisions on whether it is appropriate to introduce population screening for prostate cancer.

However, the large sample size available will have greater advantages than those simply derived from the amalgamation. It will be possible to decide whether consistency in estimated effect has been obtained across the different countries and trial designs. If there is similarity in effect, this will go a long way to confirm that the association between screening and reduction in prostate cancer mortality is truly causal, an impossible task if there was only one trial result available.

Outcomes for volunteer studies

The volunteer randomised studies in the collaboration will provide estimates of screening efficacy, i.e. the degree of benefit in those who agree to be screened. The PLCO trial will also permit analysis of whether benefit is concentrated on certain risk groups, e.g. as defined by family history, again facilitating policy decisions in the future.

Outcomes for population-based studies

The population-based studies in the collaboration will provide estimates of screening effectiveness, i.e. the degree of benefit in a population invited to be screened, taking into consideration the extent that those invited agree to attend for screening. Although the degree of compliance is specific to the culture involved, and the public's appreciation of the state

of scientific knowledge, this will help to determine, through comparison with the volunteer trials, whether those who agree to be screened comprise a low, average or high risk group in terms of risk of prostate cancer mortality. This will be a very important consideration in deciding upon policy recommendations for screening.

Multiplicative/additive effect on mortality

The collaboration will help to determine which statistical model best fits the effect, an absolute reduction in risk of death from prostate cancer or a relative reduction in risk of death, when studies from populations with different underlying rates of death from prostate cancer are amalgamated.

Timescale of effect

At this time it is not possible to know when an effect upon prostate cancer mortality can be anticipated. The rough estimates of lead time already possible suggest that it will take many years to detect an effect, possibly even longer than the 10 years incorporated into all current protocols. However, this is largely dependent upon the stage of natural history where effective treatment is possible. If the effect is largely upon poorly differentiated cancer, a benefit might be detectable within 5–7 years [5]. Even if a greater benefit occurs from treatment of cases with a longer projected time to death, the overview analysis might facilitate detecting an earlier effect in time than might be detectable in the individual trials, while also demonstrating an increasing effect with longer time from initiation of screening.

Age at screening

It is possible that the effect of screening for prostate cancer may vary by age, partly because of different validity of the PSA test by age, different rates of progression of disease by age and/or different approaches when first screened to the utilisation of radical therapy by age. The sample size provided at different ages may facilitate exploring effectiveness of screening by age subgroups (55–59, 60–64, 65–69, 70–74 years), thus providing guidance on the optimal age to initiate and terminate screening.

Rescreening frequency

The fact that the rescreening frequency ranges in different studies from 1 to 4 years and the cut-off level for action on PSA positivity from 3 to 4 ng/ml, will facilitate answering questions in relation to these parameters. Important in this regard will be the determination and comparison of interval cancer rates, facilitated by comparative mortality analyses, as interval cancer rates are only weakly predictive of prostate cancer mortality. The examination of stage and grade distributions of cancers detected at second and later screens will also be important. Precise answers to these questions may not be possible, because of a relationship in the trial designs between rescreening frequency and PSA cut-off values. Thus, one trial with a PSA cut-off of 4 ng/ml or more (PLCO) has an annual rescreening frequency.

Other effects of screening protocol

Other exploratory analyses will be facilitated by the heterogeneity between trials. Sensitivity and specificity at different PSA cut-offs will be readily explored. It is probable that the issue of the appropriate cut-off point to achieve sufficient sensitivity to obtain maximal mortality reduction while minimising unnecessary investigations in men free of prostate cancer will be resolved from the analyses that will be possible. Different policies of screening unrepresented in individual trial designs could be explored by statistical and microsimulation modelling.

Effect of 'decision to treat' protocol

We cannot determine from the overview analysis which were the effective treatments given in the different centres, in part because it is possible that treatments that are relatively ineffective when given to patients with clinically diagnosed cancer may be more effective when given to screen-detected cancers, as discussed earlier. However, we may be able to address such questions by comparing the outcome of trials in areas where different treatment policies operate. This could be important in helping to resolve the controversies over radical and watchful waiting treatment policies.

Quality of life and cost-effectiveness

A critical evaluation measure is the extent to which overall quality of life is impaired by screening compared with usual care. Decision making for health care policy is not possible if information is not available on quality of life of screened and unscreened participants as well as mortality reduction from screening. It requires an optimistic estimate of screening effectiveness to derive an overall benefit from screening [12]. Pilot quality of life and cost-effectiveness studies have already been initiated by the PLCO and European trials. Issues concerning health-related quality of life and cost-effectiveness will be more readily understood with reference to different cultural value systems and different health care systems. The issues are complex and it is unlikely that all could be evaluated in any one trial. Our collaboration will facilitate resolving these issues.

Identification and validation of intermediate outcome measures

There is a need for intermediate outcome measures that will permit an assessment of the likely benefit from prostate cancer screening in advance of the time that prostate cancer mortality can be used as the definitive endpoint. These will be essential if screening is effective and new trials are required to compare screening protocols. At present, it seems probable that only the cumulative prevalence of advanced (metastatic) prostate cancer could be suitable, and yet it is possible that this could only become available as an endpoint too close to the mortality endpoint to be valuable. Use of this also requires complex follow-up protocols which are both expensive and possibly unacceptable to control groups. It is conceivable that the prognostic molecular markers that are being developed for prostate cancer could fill this need, yet they would have to be carefully validated under controlled conditions in comparison with the mortality endpoint. Our collaboration will provide the mechanism to evaluate a number of possible candidate markers.

Microsimulation modelling

A number of different questions will be tackled by microsimulation modelling of the MISCAN type, utilising individual records that will be available from the collaboration. In such modelling, the natural history of the disease, the epidemiology, the design of the screening trial(s) and the performances of screening are incorporated. The natural history is modelled as a progression through a number of stages (local,

regional and distant and by grade). Key parameters of the performance of screening are mean duration and variability of duration of screen-detectable preclinical states, to be validated with the use of all screening data by age and screening round, and sensitivity, validated after linkage with cancer registries on interval cancers. First outcomes will be estimates of the mean duration of the screen-detectable period (by stage and age), i.e. the mean stage of advancement of the tumours, through comparisons of different screening policies applied in different periods or different centres. The induced changes in assessment, treatment and follow-up (years) can be estimated and linked to empirical quality of life estimates for each state induced or prevented by screening. Although the improvement in prognosis for screen-detected cases will be based on the end-results of the study, the opposite question will be answered earlier: what anticipated percentage of cancer mortality reduction outweighs the inevitably unfavourable effects of screening? Estimates on the proportion of Groups A–D, as mentioned previously, will be possible. The effects of contamination in the control group can be considered. With plausible estimates on key issues, known from clinical practice or validated by the trials, policy questions regarding age to start screening, to stop screening, screening interval, cost-effectiveness can be addressed.

Validation of complementary research

Complementary research is ongoing to identify (a) subsets of the populations at risk of progressive prostate cancer and, hence, in need of screening, and (b) tumour markers of early disease which is life threatening and, therefore, requiring therapy. Both of these will enable the observed benefits from the screening trials (if, indeed, they are observed) to be attainable at reduced cost, both financial and personal. With the wealth of data collected through our collaboration, many of such advances could be validated and their potential role in prostate cancer control determined.

CONCLUSIONS

Screening is not a particularly cost-effective way to control cancer mortality. This is in part because of its limited effectiveness and in part because even to obtain this limited effect, a very high proportion of the total at risk population has to be screened, with associated substantial costs. Improvements in therapy of cancer, widely anticipated over the next decade, could have important effects in improving the outcome of screening. However, such improvements could equally serve to make screening obsolete, if the same segment of disease is influenced by both improvements in treatment and early detection. Further, the adverse effects of current radical therapy for prostate cancer, including impotence in the majority and incontinence of varying degrees in many, as well as bowel dysfunction in some, make it imperative to be certain that the desired mortality reduction does indeed follow screening, and that this benefit is not outweighed by the disadvantages, complications and costs. This is why it is so important to wait for the outcome of the current generation of screening trials before advocating widespread population screening. As indicated in this special paper, the benefits may be far less than many imagine, while the burden and costs are substantial.

In the meantime, can nothing be done to facilitate answers to the questions sought? First, it is clear that the greatest contribution physicians and their 'at risk' patients can make in the areas where these trials are ongoing is to encourage participation in the trials. Second, physicians in these and other areas should take note of the scanty evidence supporting screening, and ensure that those patients who seek screening should understand the risks they are taking. There are medicolegal consequences to false-positive screening tests, and even more to overdiagnosis of cancer and adverse consequences of unnecessary treatment that many, especially in North America, have ignored potentially to their peril. The ongoing controversy over breast cancer screening among women age 40–49 years has alerted many, especially in the breast cancer activist community, to the potential adverse effects of unnecessary screening, and it will probably not take too long for this understanding to percolate into the prostate cancer activist community.

There is clearly no substitute for scientific evidence. The only means to derive conclusive scientific evidence are randomised screening trials. Our collaboration is designed to provide this evidence.

1. Auvinen A, Rietbergen JBW, Denis LJ, Schröder FH, Prorok PC. Prospective evaluation plan for randomized trials of prostate cancer screening. *J Med Screening* 1996, **3**, 97–104.
2. Adami HO, Baron JA, Rothman KJ. Ethics of a prostate cancer screening trial. *Lancet* 1994, **343**, 958–960.
3. Johansson JE, Holmberg L, Johansson S, *et al.* Fifteen-year survival in prostate cancer. A prospective population-based study in Sweden. *JAMA* 1997, **277**, 467–471.
4. Hoedemaeker RF, Rietbergen JBW, Kranse R, van der Kwast TH, Schröder FH. Comparison of pathologic characteristics of T1c and non-T1c cancers detected in a population-based screening study, the European randomized study of screening for prostate cancer. *World J Urol* 1997, **15**, 339–345.
5. Dearnaley DP, Melia J. Early prostate cancer—to treat or not to treat. *Lancet* 1997, **349**, 892–893.
6. Lu Yao GL, Yao SL. Population-based study of long-term survival in patients with clinically localised prostate cancer. *Lancet* 1997, **349**, 906–910.
7. Gann PH, Hennekins CH, Stampfer MJ. A prospective evaluation of plasma prostate-specific antigen for detection of prostatic cancer. *J Am Med Ass* 1995, **273**, 289–294.
8. Mettlin CJ, Murphy GP. Why is the prostate cancer death rate declining in the United States? *Cancer* 1998, **82**, 249–251.
9. Labrie F, Candas B. Decrease of prostate cancer death by screening: first data from the Quebec prospective and randomized study. American Society of Clinical Oncology, Annual Meeting, 18 May, 1998.
10. P. Boyle, Commentary on "Decrease of prostate cancer death by screening: first data from the Quebec prospective and randomized study". American Society of Clinical Oncology, Annual Meeting, 18 May, 1998.
11. Nyström L, Rutqvist LE, Wall S, *et al.* Breast cancer screening with mammography: overview of Swedish randomized trials. *Lancet* 1993, **341**, 973–978.
12. Krahn MD, Mahoney JE, Eckman MH, *et al.* Screening for prostate cancer: a decision analytic view. *JAMA* 1994, **272**, 781–786.