

THE CASE FOR EARLY DETECTION

Ruth Etzioni, Nicole Urban, Scott Ramsey, Martin McIntosh, Stephen Schwartz, Brian Reid, Jerald Radich, Garnet Anderson and Leland Hartwell

Early detection represents one of the most promising approaches to reducing the growing cancer burden. It already has a key role in the management of cervical and breast cancer, and is likely to become more important in the control of colorectal, prostate and lung cancer. Early-detection research has recently been revitalized by the advent of novel molecular technologies that can identify cellular changes at the level of the genome or proteome, but how can we harness these new technologies to develop effective and practical screening tests?

PROTEOMICS

The characterization and quantification of proteins and protein systems. Proteomics methods allow for the comparison of patterns of proteins isolated from bodily fluids or cells, in normal versus diseased subjects.

Cancer exacts a tremendous toll on society. In addition to the devastating effects on patients and their families, the economic costs of cancer are enormous, both in terms of direct medical-care resources for its treatment and in the loss of human capital due to early mortality.

The concept of early detection — finding tumours early, before they spread and become incurable — has tantalized cancer-control researchers for many years. Until now, however, relatively few early-detection approaches have proven sufficiently effective and practical for mass use. Recent advances in genomics and PROTEOMICS have altered the landscape of early detection, promising to vastly expand the pool of potentially useful screening tests.

In this article, we provide a rationale and a plan for early-detection research in the post-genome era. In doing so, we address three primary questions: why do we need early detection and what does it offer that other cancer-control approaches do not; what are the requirements for early detection to be effective and practical, and why do most tests that are in use at present fail to satisfy these requirements; and what kinds of studies must be conducted to fully and accurately evaluate the many emerging candidates for cancer screening tests?

Why do we need early detection?

Therapies for advanced cancers are elusive. Advances in cancer treatment and improvements in cancer outcomes over the past few decades have been modest, despite significant investment in cancer research. A great deal of research is invested in improving treatment for advanced disease, because most people who develop

cancer have advanced disease at the time of diagnosis. For example, among those with **lung, colorectal and breast cancers** in the United States, 72%, 57% and 34%, respectively, have regional or distant spread of their disease at the time of diagnosis¹. With a few notable exceptions (mostly childhood cancers), survival rates for people diagnosed with advanced cancer have changed little over the past 20 years. **FIGURE 1** shows the survival rates for persons diagnosed from 1973 to 1997 with distant, regional or distant, and localized lung, breast, **prostate** and colorectal cancer. Only modest gains in survival have occurred over the past few decades, particularly for those with distant metastases at the time of diagnosis. The data also illustrate the poor survival that is typical of advanced cancer. By contrast, as shown, survival is relatively good when these cancers are diagnosed at an early stage¹.

A tremendous amount of public and private research is directed, at present, at finding potentially curative therapies for advanced cancer, and translating these findings into benefits for patients. These efforts include research to identify molecular targets and to develop new therapeutics, and clinical trials to evaluate new therapeutics and combinations of standard therapeutics. This research is vitally important, but there is little evidence that the historical pattern of modest, incremental gains in survival among patients with disseminated disease will be improved. Indeed, a key insight provided by use of new molecular technologies for cancer — such as expression array analysis and proteomic profiling — is that we have greatly underestimated the heterogeneity of the disease. This knowledge can be expected to lead to a

Fred Hutchinson Cancer Research Center,
1100 Fairview Avenue
North, Seattle,
Washington 98109, USA.
Correspondence to R. E.
e-mail: retzioni@fhcrc.org
doi:10.1038/nrc1041

Summary

- The promise of early detection is that it will identify cancer while still localized and curable, preventing not only mortality, but also reducing morbidity and costs.
- Cervical cancer is a historical illustration of the promise of early detection; countries with broad screening programmes have markedly reduced disease-related deaths. However, the efficacy and practicality of screening tests for most other cancers remain controversial.
- The advent of new technologies — including transcript (gene-expression) analysis, genomic DNA-based methods and proteomics — offer many new opportunities for developing biomarker-based tests that are less expensive and more accurate than existing screening tests.
- To develop and fully evaluate a new screening test requires attention to all phases of biomarker development, including identification of promising biomarkers, production of assays that can detect both clinical and pre-clinical disease, development of tests that combine sensitive biomarkers to achieve greater diagnostic accuracy, and evaluation of the impact of the tests on disease mortality and costs.
- With many potential biomarkers in the early-detection pipeline, it will be important to develop strategies for evaluating the benefits and costs of biomarker-based tests within a reasonable time frame.
- The dissemination of screening tests that have been inadequately evaluated can have grave consequences, including invasive follow-up of healthy individuals, morbidity from unnecessary treatment and vastly increased costs to the medical system. Although randomized screening trials remain the ultimate test of screening efficacy in preventing disease-specific mortality, it will be important to develop these and other analytical approaches so that inferences about screening costs and benefits can be made in an efficient and timely fashion.

partitioning of cancer treatment research into more narrowly defined indications. As a result, incremental gains might ultimately be realized for increasingly narrowly defined cancer subpopulations, such as trastuzumab (Herceptin) for women with **ERBB2** (also known as HER2/neu)-positive breast cancer. Moreover, the genetic instability of invasive cancer means that combinations of agents will probably be required to prevent a cell from developing resistance to the targeted therapeutics, and this further complicates their development. The public's hope that scientists will find organ-wide 'cures' seems even less likely today than it did 20 years ago.

These observations are not meant to argue for an end to molecular therapeutics, but rather to indicate that the plethora of new specifically targeted therapeutics promised by full genome analysis will be decades in coming. It is noteworthy that most patients are, at present, still treated with nonspecific cytotoxic agents and that most of the new targeted agents that are being developed by pharmaceutical companies are directed against a small number of targets.

Early detection — an alternative approach. Cancer is a heterogeneous disease. Even within a single cancer site, tumours can show markedly different behaviours and can be associated with widely varying prognoses. However, one observation transcends disease type and site for practically all cancers: cancers detected at advanced stages are far more likely to cause death than those detected while the cancer is still confined to the organ of origin. As shown in FIG. 1a, b and d — lung cancer is the exception (FIG. 1c) — survival is excellent

for the main cancers when early-stage disease is treated with existing therapies. It is possible that a significant fraction of the overall improvement in cancer outcomes over the past decade is attributable to earlier diagnosis and prevention.

The potential impact that combining current therapies with effective early detection would have on survival is significant, as outcomes are much improved when treatment is applied to disease that is confined to the organ of origin. There are two ways to consider the significance of early detection. For the fraction of patients who would currently not have been identified until their disease was advanced, the benefit of early detection is profound (FIG. 1). For the overall population, shifting all cases to early detection would have a significant impact on overall cancer mortality and economic burden (TABLE 1). Tests that can detect precursor lesions or *in situ* disease hold even more promise, namely the possibility of eliminating the invasive condition entirely and with it the burden of the disease. An example of this being done successfully is that of **cervical cancer** (BOX 1).

Criteria for effective early detection

The World Health Organization has enumerated conditions for early detection to be an appropriate disease-control approach. First, the disease must be common and associated with serious morbidity and mortality. Second, screening tests must be able to accurately detect early-stage disease. Third, treatment after detection through screening must have been shown to improve prognosis relative to usual diagnosis. Finally, evidence must exist that the potential benefits outweigh the potential harms and costs of screening². The expectation that these conditions could be satisfied for many cancers has made early detection a topic of intensive research for the past several decades.

For early detection to be an effective and practical approach, screening tests must satisfy four basic requirements. First, screening tests should be able to distinguish healthy individuals from cancer cases with a high degree of accuracy, showing both low FALSE-NEGATIVE and low FALSE-POSITIVE RATES. Second, detection should be possible before the disease progresses to an advanced stage, when treatment is less effective. Third, screening or formal diagnostic tests should ideally allow differentiation between lesions that are aggressive, requiring treatment, and those that ultimately will do no harm, avoiding the problem of OVERDIAGNOSIS. Fourth, tests should be inexpensive and well accepted by the population that is targeted for screening.

Although screening tests are in use for a range of cancers, almost none of the available tests satisfy all of these requirements. For instance, in the case of colorectal cancer screening, many guidelines recommend SIGMOIDOSCOPY OR COLONOSCOPY, which are neither inexpensive nor well accepted because of the time and discomfort involved as well as the risk of adverse outcome. Biomarkers that are in use, at present, for **ovarian cancer** screening — primarily CA125 — have false-positive rates that lead to an unacceptably high

FALSE-NEGATIVE RATE
The proportion of diseased subjects that test negative.

FALSE-POSITIVE RATE
The proportion of non-diseased (healthy) subjects that test positive.

OVERDIAGNOSIS
The detection by screening of disease that, in the absence of screening, would not have been diagnosed within the lifetime of the patient.

SIGMOIDOSCOPY
A test that is used to detect colorectal cancer. A thin, flexible, hollow tube (sigmoidoscope) is inserted into the rectum for imaging of the lower part of the colon and rectum.

COLONOSCOPY
Similar to sigmoidoscopy, but examines the entire length of the colon.

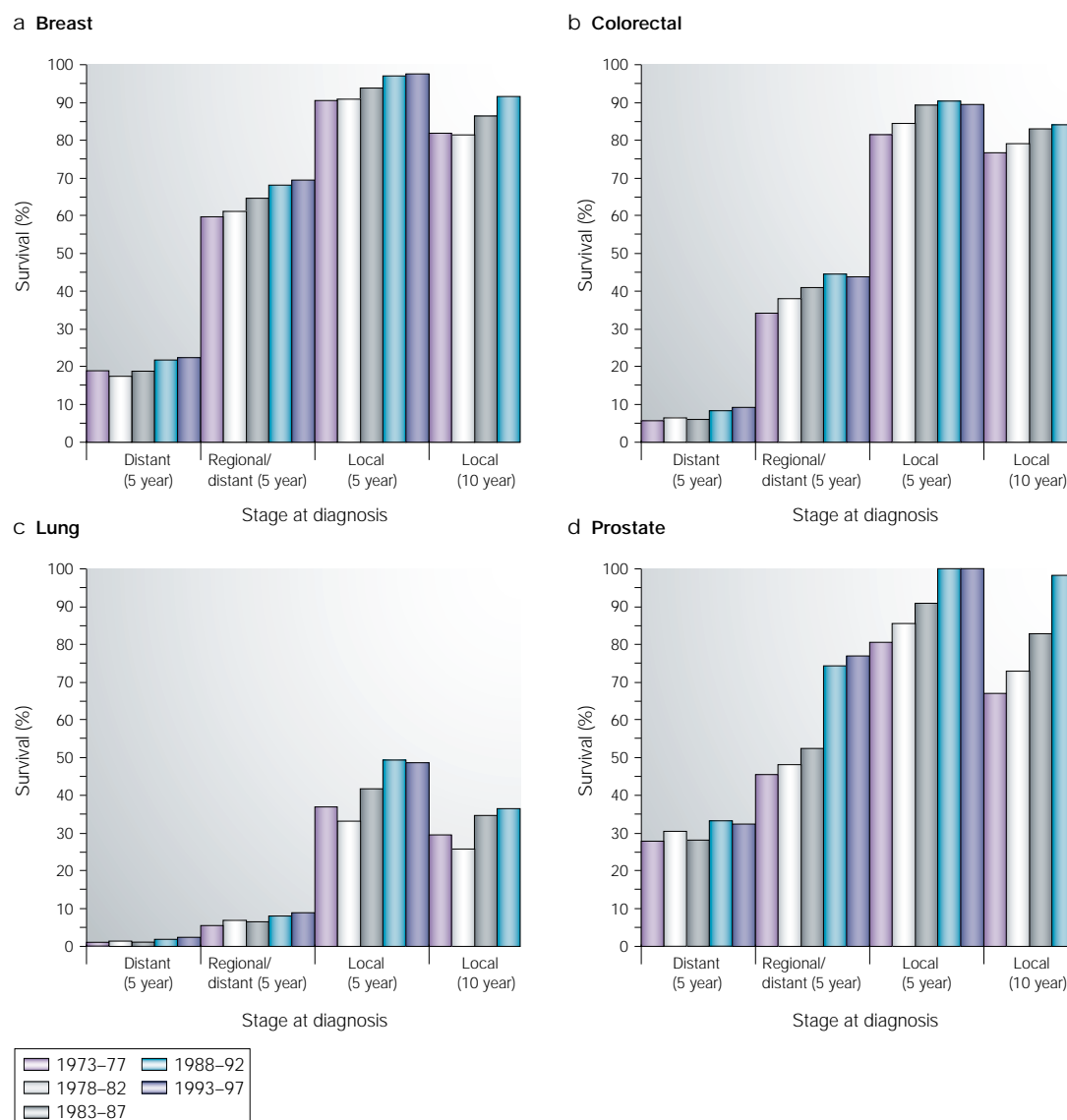


Figure 1 | Relative survival (5 year or 10 year) among cancer cases diagnosed with distant, regional or distant, and localized disease by year of diagnosis. a | Breast cancer; b | colorectal cancer; c | lung cancer; d | prostate cancer. Source: Surveillance, Epidemiology and End Results (SEER)¹. Stage is SEER historic stage.

SENSITIVITY

The proportion of diseased subjects that test positive.

SPECIFICITY

The proportion of non-diseased (healthy) subjects that test negative.

PROSTATE-SPECIFIC ANTIGEN (PSA)

A glycoprotein that is produced primarily by the epithelial cells of the prostate gland. PSA levels in serum are generally low but increase in most patients with prostate cancer.

GENE-EXPRESSION ANALYSIS

The measurement of the expression of thousands of genes simultaneously.

ratio of surgeries conducted (for confirmation of disease) to cancers detected and fail to identify many early-stage cancers³. Similarly, tests that have been evaluated for lung cancer screening show poor SENSITIVITY and SPECIFICITY and do not identify tumours sufficiently early to affect prognosis. Recent findings have indicated that helical computed tomography screening (spiral CT) might significantly improve on older modalities for the early detection of lung cancer^{4,5}, but there are grave concerns about its costs^{6,7}. In prostate cancer screening, the PROSTATE-SPECIFIC ANTIGEN (PSA) test carries a non-trivial risk of overdiagnosis due to the test's inability to clearly differentiate indolent cases from more aggressive cancers⁸.

New technologies, such as GENE-EXPRESSION ANALYSIS and serum proteomics, have already produced hundreds of potential biomarkers for detecting and

classifying cancers. Some of these biomarkers lead directly to novel diagnostics that promise to overcome the deficiencies of existing screening tests. For example, a recent analysis of proteomic patterns in the serum of ovarian cancer patients yielded a profile that distinguished cancer cases from controls with near-perfect sensitivity and specificity⁹. Similarly, several biomarkers that have been identified through expression array analysis have been shown to be predictive of the risk of biochemical recurrence (rise in PSA levels) after initial treatment for prostate cancer¹⁰. Not only do these approaches promise new models for disease discovery, but they could also improve the future availability of multiple markers for early detection, allowing more complete coverage than ever before of the spectrum of diseases that we call cancer (BOX 2).

Table 1 | Projected changes in survival with early detection

Cancer site	Tumours localized when detected (%)	5-year survival rate (%)	5-year survival rate if all tumours were localized when detected (%)
Colorectal	41	64	90
Lung	19	16	49
Breast	65	87	97
Prostate	65	90	100

Based on data from SEER¹ for cases diagnosed between 1990 and 1999 inclusive. Cases with *in situ* or unstaged disease have been excluded. The favourable overall 5-year survival among breast and prostate cancer patients is partly due to the prevalence of screening for these cancers during the calendar years considered.

Early-detection studies

The challenges faced by early-detection researchers could be classified in terms of the steps that are required to produce a useful population screening test: discovery, development and evaluation.

Discovery is the process by which candidate genes, proteins, antigens or imaging tools are identified. Novel technologies for measuring gene and protein expression have produced two powerful new approaches for identifying biomarkers. The first approach identifies genes or proteins that are either overexpressed or underexpressed in most tumours of a given type, as candidates for early-detection markers (transcript or gene-expression analysis). The second approach bypasses the analysis of tumour specimens and searches the serum directly for protein signatures that distinguish cases from controls (proteomics). At present, transcript analysis is faster and more comprehensive than proteomics, and also avoids searching through thousands of normal serum proteins that are unrelated to disease. However, important protein differences between cases and controls might result from post-transcriptional regulation and disease signatures

that arise from non-tumour cells (for example, angiogenesis or immune responses).

Although the studies published so far have generated enormous excitement, a great deal of further investigation is required to move from a finding of differential gene or protein expression to a clinically viable screening test (development) and to conclusively show that the test is effective and practical for mass use (evaluation).

Pepe *et al.* have organized the types of studies that are required into five phases, which we now describe¹¹ (FIG. 2). Although the five phases are not necessarily sequential, they are ordered according to strength of evidence from weakest to strongest, and results from earlier phases will typically be required to justify conducting later-phase studies. This formalization of these 'phases of biomarker development' is fairly recent, and standards for determining whether a test has successfully qualified to proceed to the next phase in the sequence have not yet emerged. It is likely that, in the future, fairly diverse groups — including academic researchers, regulatory agencies and administrators of biorepositories whose material is being requested for study — will all be important in determining whether new tests might be deemed sufficiently successful to progress.

Phase 1: Preclinical exploratory studies. Most of the molecular studies that have been conducted so far have been Phase 1 studies that evaluated the expression of thousands of genes or proteins in tumour and comparable healthy organ tissue to identify candidates for early detection. However, new proteomics technologies will allow discovery to be performed directly in fluids of interest, such as serum or urine, which will greatly facilitate the process of early-detection biomarker research.

Box 1 | Cervical cancer — a success story

Cervical cancer provides an excellent example of the power of early detection, and subsequent treatment, in reducing the burden of cancer. At the beginning of the twentieth century, mortality due to invasive cervical cancer was among the highest for women. By the middle of the twentieth century, pathologists had shown that the natural history of cervical cancer progressed through stages of increasingly severe cervical intraepithelial neoplasia, and that these stages could be identified histologically using exfoliated cells. The elucidation of the natural history of cervical neoplasia led to the development of the 'PAP SMEAR', and the subsequent introduction of programmes and policies in developed countries to implement widespread early detection of pre-neoplastic cervical lesions in populations⁵⁸. Since 1950, there has been an approximately 70% decline in the incidence of, and mortality due to, invasive cervical cancer in the United States^{59,60}. In developing countries where Pap smear screening is not widespread, cervical cancer remains a major public-health problem⁶¹.

Cervical cancer also illustrates the potential power of using molecular tests to enhance both the accuracy and dissemination of early detection. The reading of Pap smears requires expensive cytotechnologists and/or computers that are beyond the economic means of the developing countries that now bear the highest burden of years of life lost to cervical cancer, but Pap smears are performed on millions of women each year, even in developed countries, so a large number of both false-negatives and false-positives occur⁶². The development of molecular methods to augment, or possibly replace, Pap smears has been spurred by the recognition that cervical neoplasia is caused by persistent infection by oncogenic HUMAN PAPILLOMAVIRUSES (HPVs)^{62,63}. Since the late 1990s, studies have shown that relatively inexpensive, easy-to-use, molecular tests for the presence of HPV can be performed on cervical swabs collected either by a practitioner or by a woman herself, and will detect pre-invasive cervical cancer with higher sensitivity and no (or slight) loss of specificity in comparison to Pap smears^{62,64–66}.

PAPANICOLAOU (PAP) SMEAR
An exfoliative cytological staining procedure that can detect premalignant and malignant changes in the cervical epithelium and that is named after its founder.

HUMAN PAPILLOMAVIRUS (HPV). A virus that causes genital warts. It has also been shown to cause cervical cancer.

Box 2 | The use of multiple markers

In the future, cancer screening tests will probably combine the results of multiple diagnostic tests or biomarker assays with prior testing histories and patient-specific risk factors. Recent advances in quantitative methods use all of this information to identify individuals with positive test results who are most in need of surgical intervention, referral to imaging or early recall.

The ability to optimally combine information on multiple markers is important because single markers typically lack the sensitivity and specificity that is necessary for useful mass screening. The power of combining multiple markers has been recognized for some time^{67–69}. For example, in prostate cancer, the need to reduce false-positive rates that are associated with prostate-specific antigen (PSA) screening prompted an extensive study of the ways to combine total PSA with information about its complexed and free isoforms^{70–73}. Until recently, quantitative approaches to identify optimal combination tests were informal and suboptimal, but several statistical studies have now addressed this problem^{74–77}.

Taking biomarker behaviour over time into account can also greatly improve test performance. In prostate cancer, PSA velocity is commonly used as an adjunct to absolute PSA level to determine whether prostate biopsy is warranted. Similarly, in ovarian cancer, longitudinal algorithms that base test results on prior and present marker levels have been shown to improve sensitivity and specificity relative to tests based on a single observation^{18,21}.

In translating biomarker discoveries from the laboratory to the clinic, diagnostic performance can be further enhanced by interpreting test results in context — that is, by taking into account information that is specific to the patient, including clinical indicators, patient risk factors and screening histories. For example, screening for prostate cancer is improved by formally combining results of digital rectal examination (DRE) and PSA in combination with race and age. All four of the variables contribute significantly to the prediction of cancer at biopsy in an individual patient⁷⁸. Basing the screening decision on formal combinations of all relevant information can improve both the sensitivity and specificity of a screening intervention, and so can control costs as well as save lives^{79–81}.

Quantitative methods for Phase 1 studies are an area of active development. Data produced by high-throughput technologies present special challenges because of their high dimensionality, typically yielding expression levels for thousands of genes or proteins. Retaining a low false discovery rate while identifying promising genes or markers becomes progressively more challenging as the number of candidate markers increases. In its generic form, this is simply a classification problem; the goal is to differentiate accurately between two groups by selecting from a collection of candidate measurements, or patterns. In the case of Phase 1 studies, however, this collection is typically extremely large.

Several methods have been adapted from the statistical and machine learning literature on classification for the purposes of gene/biomarker discovery. For

gene-expression studies, the simplest approaches attempt to identify individual genes that are differentially expressed in disease cases and healthy controls. Candidate genes are effectively ranked according to some summary statistic such as a T-TEST statistic or the area under a RECEIVER-OPERATING CHARACTERISTIC (ROC) CURVE that has been computed for each gene^{12,13}. Other approaches, such as CLUSTER ANALYSIS OF SUPPORT VECTOR MACHINES — a technique adapted from computer science — identify groups of genes that seem to discriminate between cases and controls when used together^{14–16}.

As with gene-expression studies, protein expression or proteomics studies are also characterized by high-dimensional continuous data — the sample ‘mass spectra’. Here, a blend of computational and statistical methods has been found to be highly successful in identifying relevant features of the molecular signature. A recent study combined statistical clustering algorithms with computational optimization algorithms and identified peaks from protein mass spectra that provided excellent discrimination between ovarian cancer cases and healthy controls⁹.

Phase 2: Assay development and validation. Phase 2 studies focus on the development of clinical assays to measure markers in specimens that can be obtained non-invasively, such as serum, urine or sputum. Once a set of markers has been identified that correlates with disease, assays can be developed that are more efficient than those used for discovery. Many types of test can be used, including antibody assays such as ‘sandwich’ ELISA’S to detect proteins in fluids, polymerase chain reaction tests to identify circulating tumour cells, methylation tests and proteomic profiles, such as those produced by SELDI-TOF. The goals of Phase 2 studies are to develop clinical assays that are reproducible within and between laboratories, to

T-TEST

A statistical procedure for comparing measurements in two groups or samples. The result of a t-test provides an assessment of the difference between the average value in each sample relative to the variability in the two samples.

RECEIVER-OPERATING CHARACTERISTIC (ROC) CURVE

A graph of the false-negative rate versus the false-positive rate corresponding to a biomarker-based test, as the threshold biomarker level (or cutoff) for declaring the test positive varies.

CLUSTER ANALYSIS

A technique for grouping a collection of objects into subsets or clusters such that those within each cluster are more closely related to one another than are objects assigned to different clusters.

SUPPORT VECTOR MACHINES

A technique for separating data points into classes. Support vector machines derive nonlinear boundaries to optimally separate clouds of points.

ELISA

(Enzyme-linked immunosorbent assay). A widely used technique for determining the presence or amount of protein in a biological sample, using an enzyme that bonds to an antibody or antigen and causes a colour change.

SELDI-TOF

(Surface-enhanced laser desorption — time of flight). A method for profiling a population of proteins in a sample according to the size and net electrical charge of the individual proteins. The position of an individual protein in the spectrum produced corresponds to its ‘time of flight’ because the small proteins fly faster and the large proteins fly more slowly.

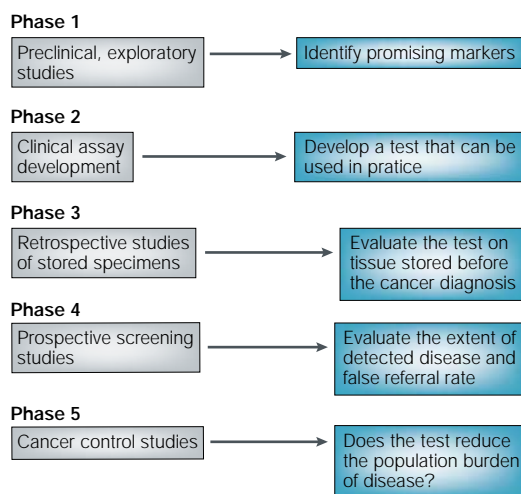


Figure 2 | Phases of biomarker development. See REF. 11.

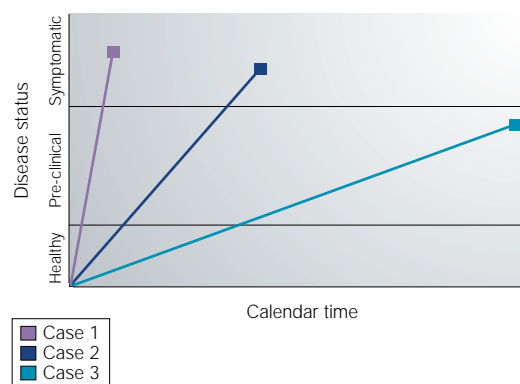


Figure 3 | Length bias and overdiagnosis in cancer screening studies. The figure shows three hypothetical cases progressing from healthy to preclinical (asymptomatic but detectable by screening) disease, to disease with clinical symptoms. The solid square indicates each case's time of death. Case 1 progresses relatively quickly through the preclinical period and therefore has less of an opportunity to be detected by screening than either case 2 or case 3. Screening preferentially detects those patients with slower-growing disease such as cases 2 or 3. Note that if case 3 is screen-detected, this would be considered as an overdiagnosis; in the absence of screening, his/her disease would never surface clinically because death due to other causes would have occurred prior to the end of the preclinical period.

confirm the correlation between these assays and the corresponding Phase 1 studies, and to evaluate their ability to discriminate between patients with clinically established disease and population controls.

Phase 3: Retrospective, longitudinal studies. Phase 1 and 2 studies focus on discriminating between established cases and healthy controls. However, to yield diagnosis at an earlier stage, a screening test must be able to identify the disease before it would become clinically apparent. Phase 3 studies therefore focus on biomarker measurements in cases before diagnosis. To do this, Phase 3 studies rely on the existence of repositories of clinical specimens, typically serum, that have been routinely collected and stored. Samples obtained from individuals before they were diagnosed with the cancer of interest are compared with samples from healthy age-matched controls. Because the cases' samples have been obtained before their diagnosis, they allow for the evaluation of biomarker levels during the preclinical phase of the disease.

Phase 3 studies are vitally important because they provide a window into the natural history of the disease and how it relates to the measurement of the biomarker under study. In the case of PSA screening, for example, Phase 3 studies provided uniquely valuable information about the amount of time by which measuring PSA could advance prostate cancer diagnosis (the lead time)¹⁷, and the sensitivity of the test¹⁷, which is typically impossible to infer from prospective screening studies.

Phase 3 studies are also important because they provide information on how marker levels change over time in disease cases and in healthy individuals. Knowledge of these patterns of change can be used to

develop tests that are both more sensitive and more specific than tests based on marker measurements at a single point in time¹⁸. Algorithms for using repeated marker levels over time have been developed for prostate^{19,20} and ovarian cancer screening^{18,21} and yield markedly improved detection rates over single threshold rules. However, developing these algorithms requires the availability of serum drawn at different intervals prior to clinical diagnosis of the disease.

Fortunately, some serum repositories already exist. In the United States, these include the Women's Health Initiative (WHI), the Baltimore Longitudinal Study of Aging (BLSA) and the Physicians' Health Study (PHS). The BLSA and the PHS have already been used for Phase 3 studies of PSA and related biomarkers in prostate cancer^{17,22}. However, developing and maintaining additional repositories for Phase 3 studies is a vitally important endeavour. Existing repositories cannot accommodate all requests for specimen samples, and the need for these resources will surely only increase as more markers of disease are discovered.

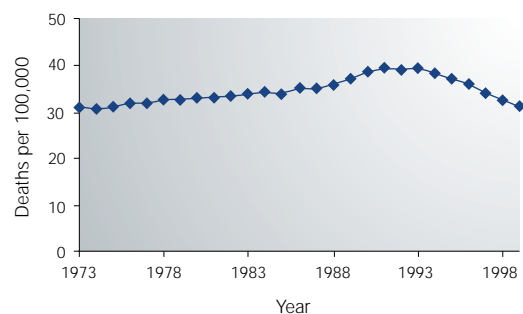
Phase 4: Prospective screening studies. Phase 3 studies can determine how long before normal clinical diagnosis a tumour marker might be able to detect disease, but they cannot infer disease characteristics at this time. Therefore, even if Phase 3 studies have shown that a candidate marker detects a cancer early with a considerable lead time, prospective studies are necessary to determine whether the marker is, in fact, able to detect the disease while it is still localized. Prospective Phase 4 studies serve this purpose.

By prospectively screening an asymptomatic population and rigorously following up individuals who test positive, Phase 4 studies provide important information about the prevalence of detectable disease in the population and the test's specificity. As a significant proportion of screening costs arise from definitive follow-up for positive results, a low false-positive rate is a prerequisite for a screening test to be eligible for implementation in a population setting. Phase 4 studies allow for the assessment of whether a candidate marker passes this test and provide information that can be used to develop preliminary estimates of expected screening costs.

As most Phase 4 studies provide definitive follow-up only in the case of a positive screening result, they cannot produce unbiased estimates of test sensitivity. Estimating test sensitivity requires knowledge of the proportions of the study population with and without the disease at the time of the test. This is only possible if all subjects are definitively checked for the presence of disease — for example, through biopsy. So, both Phase 3 and Phase 4 studies are required to fully determine the diagnostic profile of a candidate screening test.

Phase 5: Cancer control studies. Phases 1–4 focus exclusively on developing tests that are feasible for widespread use and evaluating their diagnostic performance. Even if a test performs well through to Phase 4, this does not necessarily imply that the test will reduce the population burden of disease in a meaningful way. It must be shown

a USA prostate cancer mortality



b

Study	Area of high PSA use	Area of low PSA use	Results: prostate cancer mortality
Oliver <i>et al.</i> ⁴⁸	USA; Australia	UK	Inconsistent: mortality declines in UK and USA, but does not decline in Australia
Bartsch <i>et al.</i> ⁴⁶	Tyrol, Austria	Rest of Austria	More rapid mortality decline in Tyrol
Shibata <i>et al.</i> ⁸⁵	USA	UK	More rapid mortality decline in USA
Crocetti <i>et al.</i> ⁸⁶	Florence, Italy	Prato, Italy	Similar mortality declines
Lu-Yao <i>et al.</i> ⁴⁷	Seattle, USA	Connecticut, USA	Similar mortality declines

PSA, prostate-specific antigen.

Figure 4 | **Population studies of prostate cancer screening.** **a** | Prostate cancer death rates in the United States, where prostate-specific antigen (PSA) screening is common, have declined¹. **b** | But international studies of mortality trends across areas with differing PSA screening rates show mixed results. The Table summarizes recent studies comparing prostate cancer mortality rates in areas with low- versus high-PSA screening frequencies. The inconsistencies could be due to a number of factors as well as lack of test efficacy, including the use of surrogate measures of screening frequency, variability in population mortality statistics, and insufficient calendar time to observe test efficacy in the population setting.

conclusively that interventions that are used as a result of a positive test reduce mortality. In addition, the test should be cost-effective when used to screen the population. Phase 5 studies directly evaluate the impact of screening on population disease morbidity and mortality.

Phase 5 studies include randomized, controlled cancer screening trials, as well as a number of other study designs that provide important supporting information about the impact of a specific screening test on the population disease burden, including case-control studies, computer modelling studies and population studies. The goal of evaluation is to document or refute efficacy. Evaluating early-detection interventions is extremely challenging from an analytical point of view. By its very nature, early detection affects the normal diagnosis of disease. The impact of this is twofold.

First, early-detection advances disease diagnosis even if it is not life-saving or prolonging. Consequently, survival after a fixed interval — for example, five years — following early-detection will always seem to be more favourable than in the absence of early-detection. Survival for a fixed interval — the most frequently used cancer outcome — is therefore an unreliable indicator of benefit when evaluating early-detection modalities. Other measures that do not depend on survival from diagnosis must be used in practice. This phenomenon is known as lead-time bias.

Second, early detection is, in effect, a selection process. Early-detection interventions selectively identify cancer cases from the preclinical population. Those cases with longer preclinical durations are preferentially identified through early-detection techniques (FIG. 3). This phenomenon, known as length bias, is perhaps the greatest obstacle to the drawing of inferences about the value of early detection from observational studies. As an example, consider the recent, well-publicized findings about screening for infant neuroblastoma. The development of a simple, non-invasive and sensitive test for infant neuroblastoma fuelled hopes that mortality due to the disease could be markedly reduced. However, a Canadian study that tested all infants born in Quebec found no reduction in mortality over a 9-year period even though the incidence of the disease doubled²³. The researchers concluded that the test was not finding the dangerous cancers, but, instead, was finding tumours that later would stop growing or even regress. This phenomenon, known as overdiagnosis, might be thought of as an extreme case of length bias.

Because of the difficulties in assessing early-detection interventions, the standard of evidence for efficacy of a screening test is the randomized controlled trial (RCT). The importance of RCT methodology in the presence of uncertainty cannot be underestimated. Only in the context of a randomized trial can the mortality reduction due to screening be directly estimated and the costs, including overdiagnosis, directly assessed. However, the difficulties of randomized trials for evaluating screening interventions are many. An early-detection RCT is extreme in many aspects — extremely time consuming, extremely expensive and extremely vulnerable to irrelevance due to either technological advancement or adoption of the intervention by an impatient clinical community and public. Additionally, screening trials, like prevention trials, can only test a minimal number of intervention strategies. Finally, screening trials test interventions in a highly rigorous and controlled environment. Prescribed screening protocols might not always be faithfully implemented outside this setting. An example of this is found in **oesophageal carcinoma** — a highly lethal cancer with a mortality rate of more than 90% (REF. 24). Although Barrett's oesophagus is a known precursor to oesophageal adenocarcinoma, most patients with Barrett's oesophagus do not progress to cancer^{25,26}. Specialty centres with multidisciplinary expertise have established rigorous standards for endoscopic biopsy surveillance of Barrett's oesophagus patients²⁷⁻³⁰. Surveillance programmes that do not meet these standards seem unable to consistently detect cancers at an early, curable stage²⁹⁻³¹. Although in the case of oesophageal carcinoma, research to develop more accurate and efficient screening biomarkers is ongoing, the point remains that screening interventions as they are implemented in practice might not yield quite the same benefits as those observed in a clinical-trial setting. Conversely, as in the case of breast cancer, technology improvements in the years following screening trials might actually lead to greater benefits being realized in practice than were observed in the trial setting.

Box 3 | The economics of cancer screening

The economic burden of cancer is enormous. In the United States, the overall costs of cancer in the year 2002 have been estimated at US \$171.6 billion — US \$60.9 billion for direct medical costs, US \$15.5 billion for indirect morbidity costs (lost productivity due to illness) and US \$95.2 billion for indirect mortality costs (cost of lost productivity due to premature death)⁸². Cancer treatment is one of the fastest-growing cost segments of the health economy, even accounting for the rising prevalence of cancer in an ageing society. So, acknowledging the successes that have occurred in treatment, early detection holds great potential for reducing the economic burden of cancer, as it can reduce both the direct and indirect losses.

Cost-effectiveness analysis (CEA) is valuable to decision-makers who must choose between alternative health interventions, given a limited health expenditure budget⁸³. Indeed, cost-effectiveness has become an “important, even crucial, part of health policy decision-making”⁷. The goal of CEA is to select a single ‘best’ intervention among competing strategies; that is, the strategy that yields the greatest incremental health benefits per additional dollar spent. CEA is useful for the evaluation of screening, particularly when several alternative screening interventions exist, each varying in efficacy, safety, discomfort and expense.

Cost-effectiveness is commonly expressed as the ratio of costs per life year (LY) saved, either with or without adjustment for quality of life. In the case of cancer screening, the cost-effectiveness of a particular screening strategy — in this case, strategy A — can be expressed by the following formula:

$$\text{Cost-effectiveness}_{\text{strategy A}} = \frac{C_{\text{strategy A}} - C_{\text{strategy B}}}{LY_{\text{strategy A}} - LY_{\text{strategy B}}}$$

in which $C_{\text{strategy A}}$ and $C_{\text{strategy B}}$ are the expected costs per patient of the alternative screening programmes, including post-diagnosis treatment costs, and $LY_{\text{strategy A}}$ and $LY_{\text{strategy B}}$ are the expected LYs saved by the alternative screening programmes. Note that strategy B might be a strategy of not screening at all. In general, an intervention is considered to be cost-effective if the cost-effectiveness estimate is less than US \$50,000 per life year saved.

Projecting the cost-effectiveness of a cancer screening intervention requires detailed information on the cost of screening, including follow-up of positive results, the long-term costs of cancer care, the sensitivity and specificity of the screening test, and the impact of the test on survival.

Published cost-effectiveness projections for cancer screening interventions vary widely due, in part, to uncertainty about these quantities, but also due to differing target populations, screening strategies and modelling assumptions made by the investigators.

To explore reasons for the variability in published cost-effectiveness studies, Brown and Fintor⁵⁰ compared two apparently similar studies of the cost-effectiveness of screening mammography — one from the United States³⁴ and one from The Netherlands⁵⁴ — and documented reasons for their differing results (the United States cost-effectiveness estimate was US \$34,600 per LY saved; the Dutch estimate was US \$3,825). When both models were re-run under similar assumptions about screening and biopsy frequency, as well as survival benefits and costs, the results were almost identical. So, in interpreting the results of cost-effectiveness studies, it is important to be aware of the model specifics and, in particular, the values assumed by the investigators for intervention costs and assumed impact on survival.

Regardless of the difficulties, randomized screening trials are fundamental to establishing the value of early detection. From the societal perspective, implementation of an early-detection programme in the general population — no matter how inexpensive the test — would probably soon dwarf the costs of a randomized trial. The time-consuming nature of RCTs and their vulnerability to changes in the environment cannot be avoided; these factors simply argue for the timely initiation of these efforts and the need for clear support from those who are most anxious for the results. In certain

cases, RCTs might be avoidable; for example, when existing screening methods have already been proven to be effective. In these special cases, it might be possible to evaluate novel tests by comparing their diagnostic performance with existing tests.

Other research methods continue to offer important information both in support of RCTs and in expanding our understanding of the value of screening beyond what can be measured in RCTs. Non-randomized approaches, such as epidemiological case-control studies, have been adapted for assessing the efficacy of cancer screening tests^{32–34}. Population studies and computer models represent a growing body of work in this area^{35,36}. Both case-control and population studies allow for the evaluation of screening tests as they are implemented in practice, as opposed to in a controlled setting, thereby providing an assessment of public-health impact. Development of surrogate end points for disease-specific mortality is also an important endeavour with respect to the problem of timely evaluation of screening interventions^{37,38}.

Case-control studies are useful when screening is prevalent in the population and results of clinical trials are either unavailable or pending. Case-control studies compare the screening experience of ‘cases’ (individuals who died from the disease) with ‘controls’ (individuals from the same population who did not die from the disease). Case-control studies have been used to evaluate the efficacy of breast self-examination^{39,40}, mammography^{41,42} and sigmoidoscopy⁴³, and a large, case-control study of PSA screening, funded by the Centers for Disease Control (CDC), is now in progress. Case-control studies of screening present special challenges beyond those present in aetiological studies and must be carefully conducted to avoid biased results. In some settings, bias might be unavoidable^{32–34}.

Population studies attempt to make inferences about the likely effects of screening from trends in disease incidence and mortality in the population. This is particularly important in the case of diseases such as prostate cancer for which screening is common but controversy about test efficacy exists. PSA screening for prostate cancer has, by now, become widespread in the United States, as well as in several countries in Europe, but clinical trials of PSA efficacy are still ongoing, with results expected in 2008 (REFS 44,45). Several population studies have recently been published, comparing prostate cancer mortality across areas with different frequencies of routine PSA testing^{46–48} (FIG. 4). In spite of the lack of consistency between their results, a recent summary of the evidence reviewed population studies as the most extensive source of information that is available at present regarding the efficacy of the PSA test⁴⁹.

Computer models synthesize what is known about the natural history of the cancer, the biomarker of interest, survival, and costs of screening and treatment. By simulating a population in the absence of screening, and then superimposing a specific screening strategy, computer models project the impact of screening on disease stage at diagnosis, survival and costs. An important

application of computer models is to aid researchers in designing efficient screening studies. A limitation of computer models is their reliance on multiple input parameters, some of which might be unobservable or difficult to estimate from available data. This limitation can be partly addressed through conscious attention to the development of statistical methods and appropriate data resources for parameter estimation.

A principal use of computer models has been to compare the projected benefits and costs of competing screening strategies that cannot be evaluated in practice because of time and cost considerations. Computer models that have been developed for this purpose are typically referred to as 'decision analysis' models because their results are important in medical decision-making about screening practice (BOX 3). Decision-analysis models have been used to study a wide range of screening interventions, including mammography screening for breast cancer^{50–52}, sigmoidoscopy and colonoscopy for colorectal cancer^{53,54}, and PSA screening for prostate cancer^{55–57}. Although the results of decision analyses can vary widely due to differences in model structure or assumptions made about input parameters, they have become an important component of the body of evidence that must be assembled before screening policy decisions can be made.

Conclusions

We are now at a unique and unprecedented moment in the history of early-detection research. With recent developments in molecular technology, new models for early detection are rapidly becoming a reality. These promise to overcome many of the inadequacies

of approaches that are available at present. However, to successfully navigate the growing biomarker development pipeline and translate discoveries to the clinic requires a commitment to all phases of biomarker development. Randomized, controlled screening trials remain the ultimate test of screening benefits and costs, but timely evaluation presents a greater challenge than ever before because of the sheer number of new tests anticipated.

Coordinating efforts in early detection is important and is also likely to yield substantial increases in efficiency with respect to study design. In practice, individuals are screened for multiple cancers and, therefore, it is appropriate to consider the impact of multiple screening tests when used together. Similarly, genetic changes, and the tests based thereon, might be common to more than one cancer. Therefore, as more tests become available, it will be important to consider the global picture of how screening affects mortality, morbidity and costs.

As more biologically sensitive biomarkers are developed, the potential increases for overdiagnosis — the detection of early or non-progressive lesions that would not be detected in the absence of screening. Avoiding overdiagnosis requires a firm understanding of the natural history of the disease and a commitment to conducting the necessary evaluative studies before making the test widely available. If even a few biomarkers are found that can identify our most common cancers at early, curable stages, while distinguishing those that need treatment from those that do not, the payoff will be substantial improvements in survival and quality of life, and reductions in costs for the thousands of persons who would otherwise develop advanced cancer each year.

- National Cancer Institute. *Surveillance Epidemiology and End Results Program* [online], (cited 19 Nov 2002), <http://seer.cancer.gov/> (2002).
- Winawer, S. J. *et al.* Prevention of colorectal cancer: guidelines based on new data. WHO Collaborating Center for the Prevention of Colorectal Cancer. *Bull. World Health Organ.* **73**, 7–10 (1995).
- Bell, R., Petticrew, M. & Sheldon, T. The performance of screening tests for ovarian cancer: results of a systematic review. *Br. J. Obst. Gynaecol.* **105**, 1136–1147 (1998).
- Henschke, C. Early Lung Cancer Action Project: overall design and findings from baseline screening. *Cancer* **89**, 274–282 (2000).
- Henschke, C. *et al.* Early Lung Cancer Action Project: initial findings on repeat screening. *Cancer* **92**, 153–159 (2001).
- Mahadevia, P. J. *et al.* Lung cancer screening with helical computed tomography in older adult smokers: a decision and cost-effectiveness analysis. *JAMA* **289**, 313–322 (2003).
- Grann, V. R. & Neugut, A. I. Lung cancer screening at any price? *JAMA* **289**, 357–358 (2003).
- Etzioni, R. *et al.* Overdiagnosis due to prostate-specific antigen screening: lessons from US prostate cancer incidence trends. *J. Natl Cancer Inst.* **94**, 981–990 (2002).
- Petricoin, E. F. *et al.* Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* **359**, 572–577 (2002). **This article illustrates the potential utility of serum proteomics in developing sensitive and specific cancer screening tests.**
- Dhanasekaran, S. M. *et al.* Delineation of prognostic biomarkers in prostate cancer. *Nature* **412**, 822–826 (2001).
- Pepe, M. S. *et al.* Phases of biomarker development for early detection of cancer. *J. Natl Cancer Inst.* **93**, 1054–1061 (2001). **First paper to propose the classification of biomarker studies into five phases.**
- Pepe, M. S., Longton, G. M., Anderson, G. L. & Schummer, M. *Selecting Differentially Expressed Genes from Microarray Experiments UW Biostatistics Working Paper Series*. Working Paper 184 [online], (cited 24 Jan 2003), <http://www.bepress.com/uwbiostat/paper184/> (2003).
- Tusher, V. G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl Acad. Sci. USA* **98**, 5116–5121 (2001).
- Yeung, K. Y., Haynor, D. R. & Ruzzo, W. L. Validating clustering for gene expression data. *Bioinformatics* **17**, 309–318 (2001).
- Yeung, K. Y., Fraley, C., Murua, A., Raftery, A. E. & Ruzzo, W. L. Model-based clustering and data transformations for gene expression data. *Bioinformatics* **17**, 977–987 (2001).
- Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference and Prediction* (Springer-Verlag, New York, 2001).
- Gann, P. H., Hennekens, C. H. & Stampfer, M. J. A prospective evaluation of plasma prostate-specific antigen for detection of prostatic cancer. *JAMA* **273**, 289–294 (1995).
- McIntosh, M., Urban, N. & Karlan, B. Generating longitudinal screening algorithms using novel biomarkers for disease. *Cancer Epidemiol. Biomarkers Prev.* **11**, 159–166 (2002).
- Carter, H. B. & Pearson, J. D. Prostate-specific antigen velocity and repeated measures of prostate-specific antigen. *Urol. Clin. N. Am.* **24**, 333–338 (1997).
- Polascik, T. J., Oesterling, J. E. & Partin, A. W. Prostate specific antigen: a decade of discovery — what we have learned and where we are going. *J. Urol.* **162**, 293–306 (1999).
- Skates, C. J., Pauler, D. K. & Jacobs, I. Screening based on the risk of cancer calculation from Bayesian hierarchical change-point models of longitudinal markers. *J. Am. Stat. Assoc.* **96**, 429–439 (2001).
- Carter, H. B. *et al.* Estimation of prostatic growth using serial prostate-specific antigen measurements in men with and without prostate disease. *Cancer Res.* **52**, 3323–3328 (1992).
- Woods, W. G. *et al.* Screening of infants and mortality due to neuroblastoma. *N. Engl. J. Med.* **346**, 1041–1046 (2002).
- Farrow, D. C. & Vaughan, T. L. Determinants of survival following the diagnosis of esophageal adenocarcinoma (United States). *Cancer Causes Control* **7**, 322–327 (1996).
- Conio, M. *et al.* Secular trends in the epidemiology and outcome of Barrett's oesophagus in Olmsted County, Minnesota. *Gut* **48**, 304–309 (2001).
- Drewitz, D. J., Sampliner, R. E. & Garewal, H. S. The incidence of adenocarcinoma in Barrett's esophagus: a prospective study of 170 patients followed for 8 years. *Am. J. Gastroenterol.* **92**, 212–215 (1997).
- Schnell, T. G. *et al.* Long-term nonsurgical management of Barrett's esophagus with high-grade dysplasia. *Gastroenterology* **120**, 1607–1619 (2001).
- Reid, B., Blount, P. & Rabinovitch, P. Biomarkers in Barrett's esophagus: a guideline for clinicians. *Gastrointest. Clin. N. Am.* (in the press).
- Alikhan, M. *et al.* Variable pathologic interpretation of columnar lined esophagus by general pathologists in community practice. *Gastrointest. Endosc.* **50**, 23–26 (1999).
- Ofman, J. J. *et al.* The quality of care in Barrett's esophagus: endoscopist and pathologist practices. *Am. J. Gastroenterol.* **96**, 876–881 (2001).
- Macdonald, C. E., Wicks, A. C. & Playford, R. J. Final results from 10 year cohort of patients undergoing surveillance for Barrett's oesophagus: observational study. *BMJ* **321**, 1252–1255 (2000).
- Connor, R. J., Prorok, P. C. & Weed, D. L. The case-control design and the assessment of the efficacy of cancer screening. *J. Clin. Epidemiol.* **44**, 1215–1221 (1991).

33. Cronin, K. A., Weed, D. L., Connor, R. J. & Prorok, P. C. Case-control studies of cancer screening: theory and practice. *J. Natl Cancer Inst.* **90**, 498–504 (1998).
34. Weiss, N. S. Case-control studies of the efficacy of screening for cancer: can we earn them some respect. *J. Med. Screen.* **4**, 57–59 (1997).
- References 32–34 discuss the use of case-control studies for evaluating cancer screening with special attention to avoiding bias.**
35. Etzioni, R. *et al.* Cancer surveillance series: interpreting trends in prostate cancer. Part III: Quantifying the link between population prostate-specific antigen testing and recent declines in prostate cancer mortality. *J. Natl Cancer Inst.* **91**, 1033–1039 (1999).
36. Feuer, E. J., Mariotto, A. & Merrill, R. Modeling the impact of the decline in distant stage disease on prostate carcinoma mortality rates. *Cancer* **95**, 870–880 (2002).
37. Schatzkin, A. & Gall, M. The promise and peril of surrogate end points in cancer research. *Nature Rev. Cancer* **2**, 19–27 (2002).
38. Hirsch, F. R. *et al.* IV international conference on prevention and early detection of lung cancer, Reykjavik, Iceland, August 9–12, 2001. *Lung Cancer* **37**, 325–344 (2002).
39. Friedman, D. R. & Dubin, N. Case-control evaluation of breast cancer screening efficacy. *Am. J. Epidemiol.* **133**, 974–984 (1991).
40. Newcomb, P. A. *et al.* Breast self-examination in relation to the occurrence of advanced breast cancer. *J. Natl Cancer Inst.* **83**, 260–265 (1991).
41. Hurley, S. F. & Kaldor, J. M. The benefits and risks of mammographic screening for breast cancer. *Epidemiol. Rev.* **14**, 101–130 (1992).
42. Palli, D. *et al.* A case-control study of the efficacy of a non-randomized breast cancer screening program in Florence (Italy). *Int. J. Cancer* **38**, 501–504 (1986).
43. Selby, J. V., Friedman, G. D., Quesenberry, C. P. Jr & Weiss, N. S. A case-control study of screening sigmoidoscopy and mortality from colorectal cancer. *N. Engl. J. Med.* **326**, 653–657 (1992).
44. Beemsterboer, P. M. *et al.* Prostate specific antigen testing and digital rectal examination before and during a randomized trial of screening for prostate cancer: European randomized study of screening for prostate cancer, Rotterdam. *J. Urol.* **164**, 1216–1220 (2000).
45. Gohagan, J. K., Prorok, P. C., Hayes, R. B. & Kramer, B. S. The Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial of the National Cancer Institute: history, organization, and status. *Control Clin. Trials* **21**, 251S–272S (2000).
46. Bartsch, G. *et al.* Prostate cancer mortality after introduction of prostate-specific antigen mass screening in the Federal State of Tyrol, Austria. *Urology* **58**, 417–424 (2001).
47. Lu-Yao, G. *et al.* Natural experiment examining impact of aggressive screening and treatment on prostate cancer mortality in two fixed cohorts from Seattle area and Connecticut. *BMJ* **325**, 740 (2002).
48. Oliver, S. E., May, M. T. & Gunnell, D. International trends in prostate-cancer mortality in the 'PSA ERA'. *Int. J. Cancer* **92**, 893–898 (2001).
- References 46–48 are examples of recent population studies of PSA screening.**
49. Harris, R. & Lohr, K. N. Screening for prostate cancer: an update of the evidence for the US Preventive Services Task Force. *Ann. Intern. Med.* **137**, 917–929 (2002).
50. Brown, M. L. & Fintor, L. Cost-effectiveness of breast cancer screening: preliminary results of a systematic review of the literature. *Breast Cancer Res. Treat.* **25**, 113–118 (1993).
- Explores the reasons for variability in estimates of the cost-effectiveness of screening interventions.**
51. Rosenquist, C. J. & Lindfors, K. K. Screening mammography beginning at age 40 years: a reappraisal of cost-effectiveness. *Cancer* **82**, 2235–2240 (1998).
52. Salzmann, P., Kerlikowske, K. & Phillips, K. Cost-effectiveness of extending screening mammography guidelines to include women 40 to 49 years of age. *Ann. Intern. Med.* **127**, 955–965 (1997).
- References 51 and 52 show how different results of cost-effectiveness studies of the same intervention can be.**
53. Pignone, M., Saha, S., Hoerger, T. & Mandelblatt, J. Cost-effectiveness analyses of colorectal cancer screening: a systematic review for the US Preventive Services Task Force. *Ann. Intern. Med.* **137**, 96–104 (2002).
54. van der Maas, P. J. *et al.* The cost-effectiveness of breast cancer screening. *Int. J. Cancer* **43**, 1055–1060 (1989).
55. US Congress Office of Technology Assessment. *Costs and Effectiveness of Prostate Cancer Screening in Elderly Men* (US Government Printing Office, Washington DC, 1995).
56. Etzioni, R., Cha, R. & Cowen, M. E. Serial prostate specific antigen screening for prostate cancer: a computer model evaluates competing strategies. *J. Urol.* **162**, 741–748 (1999).
57. Ross, K. S., Carter, H. B., Pearson, J. D. & Guess, H. A. Comparative efficiency of prostate-specific antigen screening strategies for prostate cancer detection. *JAMA* **284**, 1399–1405 (2000).
58. Boyes, D. A. The value of a pap smear program and suggestions for its implementation. *Cancer* **48**, 613–621 (1981).
59. Christopherson, W. M., Parker, J. E. & Mendez, W. M. Cervix cancer death rates and mass cytologic screening. *Cancer* **26**, 808–811 (1970).
60. Schiffman, M. H., Brinton, L. A., Devesa, S. S. & Fraumeni, J. F. Jr in *Cancer Epidemiology and Prevention* 2nd edn (eds Schottenfeld, D. & Fraumeni, J. F. Jr) 1090–1116 (Oxford Univ. Press, New York, 1996).
61. Parkin, D. M., Pisani, P. & Ferlay, J. Estimates of the worldwide frequency of eighteen major cancers in 1985. *Int. J. Cancer* **54**, 594–606 (1993).
62. Cuzick, J. Time to consider HPV testing in cervical screening. *Ann. Oncol.* **12**, 1511–1514 (2001).
63. Stoler, M. H. HPV for cervical cancer screening: is the era of the molecular pap smear upon us? *J. Histochem. Cytochem.* **49**, 1197–1198 (2001).
64. Cuzick, J. *et al.* Human papillomavirus testing in primary cervical screening. *Lancet* **345**, 1533–1536 (1995).
65. Kulasingam, S. L. *et al.* Evaluation of human papillomavirus testing in primary screening for cervical abnormalities: comparison of sensitivity, specificity, and frequency of referral. *JAMA* **288**, 1749–1757 (2002).
66. Ratnam, S., Franco, E. L. & Ferenczy, A. Human papillomavirus testing for primary screening of cervical cancer precursors. *Cancer Epidemiol. Biomarkers Prev.* **9**, 945–951 (2000).
67. Woolas, R. P. *et al.* Combinations of multiple serum markers are superior to individual assays for discriminating malignant from benign pelvic masses. *Gynecol. Oncol.* **59**, 111–116 (1995).
68. Woolas, R. P. *et al.* Elevation of multiple serum markers in patients with stage I ovarian cancer. *J. Natl Cancer Inst.* **85**, 1748–1751 (1993).
69. Bates, J. *et al.* Clinical value of CYFRA 21.1, carcinoembryonic antigen, neuron-specific enolase, tissue polypeptide specific antigen and tissue polypeptide antigen in the diagnosis of lung cancer. *Eur. Resp. J.* **10**, 2535–2538 (1997).
70. Beduschi, M. C. & Oesterling, J. E. Percent free prostate-specific antigen: the next frontier in prostate-specific antigen testing. *Urology* **51**, 98–109 (1998).
71. Brawer, M. K. Prostate-specific antigen: current status. *CA Cancer J. Clin.* **49**, 264–281 (1999).
72. Catalona, W. J. *et al.* Use of the percentage of free prostate-specific antigen to enhance differentiation of prostate cancer from benign prostatic disease: a prospective multicenter clinical trial. *JAMA* **279**, 1542–1547 (1998).
73. Gann, P. H., Ma, J., Catalona, W. J. & Stampfer, M. J. Strategies combining total and percent free prostate specific antigen for detecting prostate cancer: a prospective evaluation. *J. Urol.* **167**, 2427–2434 (2002).
74. Zhang, Z. *et al.* Combination of multiple serum markers using an artificial neural network to improve specificity in discriminating malignant from benign pelvic masses. *Gynecol. Oncol.* **73**, 56–61 (1999).
75. McIntosh, M. W. & Pepe, M. M. Combining several screening tests: optimality of the risk score. *Biometrics* **58**, 657–664 (2002).
76. Baker, S. G. Identifying combinations of cancer markers for further study as triggers of early intervention. *Biometrics* **56**, 1082–1087 (2000).
77. Etzioni, R., Kooperberg, C., Pepe, M. S., Smith, R. & Gann, P. H. Combining biomarkers to detect disease with application to prostate cancer. *Biostatistics* (in the press).
- References 74–77 present new statistical methods for identifying optimal combinations of biomarkers.**
78. Optenberg, S. A. *et al.* Development of a decision-making tool to predict risk of prostate cancer: the Cancer of the Prostate Risk Index (CAPRI) test. *Urology* **50**, 665–672 (1997).
79. Urban, N., Drescher, C., Clarke, L. & Kiviat, N. in *Ovarian Cancer 5* (eds Sharp, F., Blackett, T., Berek, J. & Bast, R.) 217–229 (Isis Medical Media Ltd., Oxford, 1998).
80. Urban, N. *et al.* in *Ovarian Cancer* (ed. Jacobs, I.) 199–208 (Oxford Univ. Press, 2002).
81. Urban, N. Screening for ovarian cancer. *BMJ* **319**, 1317–1318 (1999).
82. American Cancer Society. *Costs of Cancer, 2002* [online], (cited 25 Jan 2003), <http://www.cancer.org/docroot/MIT/content/MIT_3_2X_Costs_of_Cancer.asp> (2002).
83. Gold, M. E., Siegel, J. E., Russell, L. B. & Weinstein, M. C. *Cost-Effectiveness in Health and Medicine* (Oxford Univ. Press, New York, 1996).
84. US Congress Office of Technology Assessment. *Breast Cancer Screening for Medicare Beneficiaries; Effectiveness, Costs to Medicare and Medical Resources Required* (US Government Printing Office, Washington DC, 1987).
85. Shibata, A. & Whittemore, A. S. Re: Prostate cancer incidence and mortality in the United States and the United Kingdom. *J. Natl Cancer Inst.* **93**, 1109–1110 (2001).
86. Crocetti, E., Ciatto, S. & Zappa, M. Prostate cancer: different incidence but not mortality trends within two areas of Tuscany, Italy. *J. Natl Cancer Inst.* **93**, 876–877 (2001).

Acknowledgements

Supported in part by cooperative agreement from the National Cancer Institute. We thank R. Smith of the American Cancer Society and the anonymous referees for helpful comments on an earlier version of this article.

Online links

DATABASES

The following terms in this article are linked online to: **Cancer.gov**: http://www.cancer.gov/cancer_information/ breast cancer | cervical cancer | colorectal cancer | lung cancer | oesophageal cancer | ovarian cancer | prostate cancer **LocusLink**: <http://www.ncbi.nlm.gov/LocusLink/ERBB2> | PSA Access to this interactive links box is free online.