HEDS Discussion Paper 08/08


ScHARR
School of Health and
Related Research

ScHARR

Health Economics and Decision Science Discussion Paper Series

No. 08/08

From evidence-based to decision-analytic medicine: A mammography case study.

Jason Madan[1]

[1] Health Economics and Decision Science, School of Health and Related Research, University of Sheffield

Corresponding author:
Jason Madan
Health Economics and Decision Science,
School of Health and Related Research
University of Sheffield
Regent Court
30 Regent Street
Sheffield
S1 4DA
UK
Tel: 0114 222 0827
Fax: 0114 222 4095
Email: j.madan@sheffield.ac.uk

Acknowledgements

The author is the recipient of a Researcher Development Award from the Department of Health National Co-ordinating Centre for Research Capacity Development. This award is funding work towards a PhD supervised by Dr. Jonathon Karnon, including the work presented here. The author would also like to thank Paul Tappenden and Dr. Paul Sutcliffe for their helpful comments. The views expressed are the author's own.

Abstract

The technology needed to implement mass screening by mammography existed well before the implementation of national screening programmes. This delay arose partly because of the complexities involved in conducting randomised controlled trials (RCTs) of screening programmes. These complexities not only extend the time needed to conduct trials of cancer screening, they reduce the external validity of the results. There is potential to improve the application of evidence-based medicine (EBM) to the evaluation of cancer screening programmes and other complex interventions through adding insights from Operational Research and Decision Theory. This would extend EBM to what might be called Decision-Analytic Medicine (DAM).

A brief history of x-ray mammography and its evaluation

The first use of x-rays to view tumours in the breast, in 1913, is credited to Albert Solomon, a German surgeon [1]. Interest in x-ray mammography as a diagnostic tool began to develop in the 1930s. By the late 1950s, Robert Egan and colleagues in the U.S.A. were able to present data illustrating the potential of x-ray mammography in detecting asymptomatic breast tumours.

A combination of technological ability and cultural acceptability led to growing calls, at least in the U.S.A, for mass screening programmes based on mammography [2]. By the early 1960s, the use of controlled clinical trials to evaluate new interventions was established. Philip Strax, a radiologist with the Health Insurance Plan (HIP) of New York, proposed carrying out a clinical trial of mammography on members of HIP. This trial, the first of mammography, was carried out between 1963 and 1968. Results from the trial began to appear in the early 1970s, and they suggested that mammography could reduce breast cancer mortality by around 30% [3].

The broadly positive results of the HIP trial were one factor in placing mass screening by mammography on the public health agenda in the 1970s. In the US, the diagnosis of prominent individuals (e.g. Betty Ford and Margaretta Rockefeller) with breast cancer raised awareness of the disease, and interest in diagnosis, significantly [4]. The promise shown by the HIP trial spurred improvements in technology such as the single emulsion film, introduced by DuPont in 1972 [5]. The American Cancer Society (ACS), responding positively to the findings of the HIP trial, inaugurated the Breast Cancer Detection Demonstration project (BCDDP) jointly with the National Cancer Institute. The BCDDP was essentially the pilot for a national mammography screening programme.

From the 70s onwards, further trials of mass screening were carried out in the UK, Sweden and Canada (Table 1.1). A consensus began to emerge that mass screening by mammography was of benefit for at least some age groups. From the mid-80s

onwards, a number of countries began to implement national screening programmes. In 1988, the International Breast Cancer Screening Database Project, subsequently renamed the International Breast Screening Network (IBSN), was established. The IBSN is a voluntary consortium of countries that have active population-based screening mammography programs. Currently, 27 countries participate in the IBSN. A 1995 survey of 22 countries carried out by the organisation [6] showed that 19 had begun organised mass screening programmes, and 9 had, or planned to have, national coverage by 2000.

Study	**Year Started**	**Age at entry (years)**	**Number of women**		**Number of breast cancer deaths after 13 years of follow-up**	
			Control	*Intervention*	*Control*	*Intervention*
New York HIP	1963	40-64	30,239	30,756	218	262
Swedish two-counties	1977	40-74	78,085	56,782	261	277
Malmo	1976	45-69	21,088	21,195	87	108
Edinburgh	1979	45-64	23,226	21,904	176	187
Stockholm	1981	40-64	40,318	19,943	66	45
Canadian NBSS	1980	40-59	44,925	44,910	212	213
Goteborg	1982	40-59	21,650	29,961	88	162
All Studies			259,531	225,451	1,108	1,254

Table 1.1 Randomised clinical trials of screening by mammography [5, 7]

Breast cancer is a major source of mortality and morbidity across the world. There is an understandably high level of awareness of its effects, and public support behind any intervention that counteracts these effects. By 1963, the technology of mammography was considered sufficiently promising to justify the first RCT of its effectiveness as a screening tool. However, 35 years later, 13 out of 22 countries surveyed did not have a mammography programme with national coverage in place.

Could, and should, such screening programmes have been implemented earlier? This question will be a key motivation for the research presented in this thesis.

There is a broad consensus on the existence of benefit from screening by mammography. However, there has been an ongoing debate on how far that benefit goes, and some dissenters have taken a more sceptical view on mammography than others. In the next section, I take a brief look at this debate.

The debate over mass screening by mammography

The BCDDP was an uncontrolled pilot rather than a controlled study, and there were those who argued that its implementation was premature. One of the most prominent critics of the speed with which mass screening by mammography was being implemented in the USA was John Bailar. Bailar was a clinician, biostatistician, and the NCI Deputy Director for Cancer Control. He argued that the HIP had not conclusively demonstrated the benefits of screening mammography, citing the possibility of lead time and length bias in the results [8].

Bailar also cited reasons why screening might not be as efficacious as its proponents hoped. These reasons lie at the heart of many sceptical positions on cancer screening programmes. The first reason was that many of the lesions detected by screening would be slow-growing and clinically unthreatening. Detecting them would therefore lead to over-treatment. The second reason was that screening could also cause harm. This harm includes physical effects such as increased incidence of cancer from radiation exposure, and mental harm e.g. the needless anxiety generated by false positives.

Whilst Bailar, amongst others, argued that the benefits of mammography screening were unproven in general, he was particularly sceptical about its value in the under 50s. Mammography is generally less effective in the pre-menopausal breast, as it tends to be denser. When stratified into age cohorts, the results of the HIP trial did not demonstrate a statistically significant survival improvement in the under 50s.

However, a number of issues continued to be debated. One was the frequency of screening. In the UK, for example, the national screening programme operates once every three years. In the US, screening is recommended every 1-2 years. [http://www.ahrq.gov/clinic/3rduspstf/breastcancer/]. Also, the debate over screening the 40-49 age group continues to this day. The Canadian Task Force on Preventative Care, amongst others, argues that there is insufficient evidence to justify screening in

this age group [http://www.ctfphc.org/]. In the US, however, the ACS currently recommends that women over 40 receive annual mammographies [9].

The boundaries of consensus were shifted by the publication, in 2000, of a systematic (Cochrane) review and analysis of mammography RCTs [10], [11]. The authors of the review argued that there were methodological flaws in the design of several of these RCTs, which could lead to biases in the results. For example, they argued that errors in recording the cause of death of participants could have systematically favoured screening. Crucially, they argued that the trials which had shown the least benefit from screening were also the trials where the methodology was strongest. They concluded that existing trials did not provide enough evidence to demonstrate conclusively that screening by mammography was effective.

The publication of this Cochrane review lead to a resurgence in the mammography debate, both in the academic community and in the wider media [12]. Many have responded to counter the criticisms raised by the review [13], [14], [15], [16], [17], whilst the authors have continued to defend their analysis, [18], [19], [20], supported by others [21]. The original Cochrane review has been recently updated [7]. This version moderates its conclusions, stating that it is likely that screening reduces breast cancer mortality, and that a reasonable estimate for the relative reduction is 15%. However, the authors go on to argue that this effect is small in absolute terms, and that (echoing Bailar) screening also leads to overdiagnosis and overtreatment. They state that

'It is thus not clear whether screening does more good than harm' [7].

There has been an extensive debate over the analysis presented in the Cochrane review, and it is not my aim to join in this debate. Instead, I would like to explore the following question. Despite 40 years of investigation, involving half a million women followed for up to 15 years, why have RCTs not been able to provide enough evidence for policy-makers to settle debates over the effectiveness of screening by mammography?

The Difficulties of Screening Trials

"Randomised screening trials are bothersome … Still, such long-term large-scale randomised screening trials are crucial, and there is no second-best option". [22]

The evidence based medicine (EBM) movement can be seen as an attempt to formalise the process of evaluation of medical treatments, and bring this process in line with other forms of scientific enquiry [23]. A key part of the approach is the idea of hierarchies of evidence. The quality of a type of evidence relates to the scope for bias to skew the results. There are a wide range of effects that can generate misleading outcomes, including selection bias, placebo effects, and misreporting of outcomes. An advocate of EBM might argue that there is a natural tendency to place excessive weight on personal experiences and anecdote, and a further tendency to sometimes interpret results of studies in line with hopes and expectations. The application of EBM is an attempt to counter these tendencies.

At the top of the hierarchy lies the randomised, double-blind, controlled trial (or, even better, a systematic review of such trials) http://www.cebm.net/levels_of_evidence.asp#levels. RCT evidence is highly regarded because this study design, if implemented well, eliminates most sources of systematic bias. For this reason, it is often referred to as the 'gold standard' of medical evidence. There are situations in which conducting RCTs is impractical or unethical. Sackett argues that:

> "Evidence based medicine is not restricted to randomised trials and meta-analyses… if no randomised trial has been carried out for our patient's predicament, we must follow the trail to the next best external evidence and work from there." [24].

Nevertheless, there is a view, implicit in the hierarchy of evidence, and often eluded to in the literature, that the efficacy of an intervention cannot be considered settled

until RCT evidence is available, and that such evidence should be obtained if at all possible.

For a number of reasons, some interventions are difficult to evaluate by RCT. Mass population screening for cancer is an example of such a challenging area [22]. An important reason for this is that the vast majority of participants will be disease-free, and so will not provide much information for the researcher. The result is that the RCT will need to be large to detect the affect of the screening tool. Power calculations are often performed to estimate the size of trial required to detect the expected effect…

The aim of cancer screening is to reduce mortality from the disease. However, it can take considerable time for any such benefit to become apparent in the results of a trial. There can be a substantial delay between diagnosis and outcome for patients on both arms. Also, several rounds of screening, several years apart, may be needed before the benefits present themselves. This is the reason why cancer screening trials can often take a decade or more to complete.

The combination of enormous size and lengthy follow-up leads to practical problems in conducting population cancer screening trials. Two of these problems are contamination and non-compliance [5]. Contamination occurs when participants in the control group seek out screening outside the study. Non-compliance occurs when participants in the screening arm fail to attend one or more of their allotted screenings. Both contamination and non-compliance dilute the ability of a trial to detect the clinical benefit of screening. Therefore, if these are likely to be significant issues, the trial will need to be even larger, and the duration even longer, to detect the benefits of screening.

The length of follow-up needed in trials of mass screening for cancer creates further dilemmas. There is a possibility that the screening technology will change over the duration of the trial. With mammography, there have been continual improvements in the technology (e.g. extended cycle processing [5]) and in the implementation (computer-assisted reading, double-view). Where there have been improvements, this may not be a crucial issue, as investigators can argue that actual benefits will be at

least as large as those identified in trials. However, if an entirely new screening technology is introduced, trials based on existing methods may provide limited information of value to decision-makers.

There is a conceptual problem with the design of screening RCTs which is exacerbated by the time needed to carry them out. In isolation, screening provides no therapeutic benefits. It is in the impact of early detection on the choice and outcome of treatment where any such benefits lie. Therefore changes in the treatments available, or in their effectiveness, will affect the benefit of early detection. Participants of the trials listed in table 1.1 will have received treatment between the early 1960s and the mid 1990s. As we shall subsequently see, treatment options have changed radically during that time, and substantially since. It is not intuitively obvious whether such changes will, in balance, benefit screen-detected cases more, less, or equally than clinically detected patients.

It is usually not feasible to blind the doctor or the patient as to whether or not a trial participant has received screening. This can affect the validity of trial results in a number of ways. It is possible that knowing whether a patient has a symptomatic or screen-detected may influence treatment decisions unduly. Also, the process of undergoing screening may have a psychological impact on the participant, leading to changes in health-related behaviour. The participant may change her lifestyle in ways that reduces risk, and may also be more alert to the symptoms of breast cancer if it were to present between screenings.

Many of these issues with the assessment of cancer screening programmes can be described as problems with external validity. External validity describes the extent to which an experimental result applies in the real world. In medicine, this means the extent to which the results of study reflect the consequences of introducing an intervention into clinical practice. Lack of consideration of external validity is the most frequent criticism by clinicians of RCTs [25]. There are several reasons why this is generally so. Ethnic differences and differences in health care systems may reduce the external validity of a study carried out in one country to the population of another. RCTs are often carried out by expert proponents of a new technology in idealised settings, so that results obtained in routine practice are not as favourable. To reduce

confounding, RCT protocols often place tight restrictions on the characteristics of participants. This leaves unclear how the results might generalise to a more heterogeneous population. Rothwell lists 39 factors that can reduce the external validity of a trial to a given clinical setting.

Most of these factors undermining the external validity of RCTs in general apply specifically to trials of cancer screening programmes. In this section, I have put forward further reasons, specific to this type of intervention, why the results of an RCT might not reflect the outcomes of real-world implementation. Nevertheless, many would support the assertion of De Koning given at the beginning of this section; that RCTs are the only valid and conclusive source of evidence on which to base clinical policy, despite their occasional shortcomings. To challenge this view, it is necessary to look at the intellectual framework, or paradigm, underpinning it. This is the topic of the next section.

Evidence Based Medicine as an application of the Scientific Method

The term 'evidence-based medicine' has been criticised as too vague a description of the aims of its proponents [26]. Any component practitioners would surely base their medical decisions on evidence. The key to EBM lies in the weight it puts on the quality of evidence in driving clinical decisions. To understand this fully, compare EBM to a stylised alternative which we might call 'judgement-based medicine' (JBM). In JBM, the clinician makes decisions based on his personal training and experience. JBM is still evidence-based, but relies on the expert and subjective judgement of the practitioner to interpret that evidence. In particular, doctors have the freedom to derive and evaluate their own conclusions from the evidence available to them.

It could be argued that JBM is a crude, but not a fanciful, description of the approach to medicine pre-EBM. The problem JBM creates has been described already – it leaves space for several sources of bias to creep in, leading to sub-optimal treatment choices. EBM arose as a direct response to this problem, and that response involves the formal, systematic and objective collection of evidence to eliminate bias. The randomised, double-blind, placebo-controlled clinical trial is the study design strongest at eliminating bias, and so it is given the highest position in the EBM hierarchy of evidence.

EBM can therefore be seen as an attempt to compensate for the weaknesses of subjective human judgement with the objectivity of evidence obtained from rigourous experiment. This does not mean that it eliminates the need for expert judgement. Sackett describes EBM as 'integrating individual clinical expertise and the best external evidence' [27]. However, EBM does represent a bound to clinical freedom. In particular, it places a responsibility on the medical profession to choose treatments that have a formally approved evidence base, to justify why an evidence base does not apply to a particular patient, or to create that evidence base when it does not exist.

EBM, therefore, aims to ground the practice of medicine within the empirical scientific method. This method requires that theories be tested against carefully structured observations of the natural world. There are many characterisations of the scientific method, but it can be seen as a process with four components:

- Observation: This stage involves study of the natural world in a search for patterns. For example, we might observe that a high proportion of patients with heart disease have furred arteries.
- Hypothesis: From our observations, we try and draw general principles and conclusions. For example, our study of patients with heart disease might lead to the hypothesis that elevated blood cholesterol leads to furred arteries and increases the risk of a heart attack.
- Prediction: Combining observations and hypotheses can lead to predictions. If we observe that a drug lowers cholesterol in laboratory conditions, we might combine this with the hypothesis created above to conclude that the drug will reduce the risk of a heart attack.
- Testing: This key stage involves testing our prediction in a setting carefully controlled so that we can be sure any observations relate to the prediction we are testing.

These four stages are an idealised formalisation of the process by which scientific knowledge advances.

EBM is clearly an attempt to ensure that this last stage is carried out appropriately. It therefore equates clinical decision-making with the evaluation of a scientific hypothesis. The question ‘should patient X receive treatment A’ becomes a hypothesis, and the toolkit of hypothesis testing is applied – the creation of a null hypothesis, controlled hypothesis testing, and the analysis of results using frequentist tests of statistical significance. Seen in this light, clinical debate focuses on the question of whether the evidence is of sufficient quality and quantity to support the hypothesis. This can be seen clearly in the debate on mammography described above, which is largely a debate about the quality of the trials.

The power of the scientific method is beyond dispute, as is particularly evident in the technologies of modern medicine. The aim of grounding medical practice in this method is therefore understandably attractive. However, medicine is more than a science; it is the application of science to an end. That end is the best use of medical resources (time and materials) in improving health. The ultimate aim of medical evidence is to guide treatment choices rather than establish scientific truths.

Why might it be necessary to differentiate between these two aims that would seem to go hand-in-hand? The answer lies in the earlier discussion about the external validity of an RCT. Where external validity is an issue, there is a knowledge gap between the results of the trial and the answer to the question of how to treat a particular patient or group of patients. The results of the RCT help answer the latter question, but they may not provide the answer. We have seen that this is a particular issue in assessing cancer screening programmes. As well as the standard sources of external validity issues, we have the complication of accounting for changing treatment patterns and screening technologies over time. There is also the issue of predicting the effectiveness for varying age cohorts and screening cycle frequencies.

Where the problem of external validity is significant, as it appears to be in cancer screening, the solution might be to rely on clinical expertise to bridge the ‘knowledge gap’. Yet this returns us to the issues of subjectivity EBM is designed to combat. In particular, it can lead to the difficult situation where experts disagree. With screening by mammography, there is a divergence in clinical opinion and national guidelines over the same RCT evidence base. In the next section, I sketch out an alternative conceptual framework that could potentially advance the debate.

Decision-analytic medicine and the role of modelling

What sort of problem is posed by the question – 'should we implement a mammography-based mass screening programme for breast cancer'? The EBM process interprets this as a scientific hypothesis that needs to be tested under controlled conditions; i.e. within a large-scale RCT. This interpretation struggles with the fact that the value of screening is affected by many factors, including the treatment options available, the target population, quality of service, and frequency of screening. This raises the issue of external validity in applying the result of a given trial – there is a 'knowledge gap' between the result of the RCT and the answer to the question posed above. We could tackle this issue by carrying multiple RCTs with varying protocols. This will not eliminate the problem, though – the time needed to carry out a definitive RCT raises the possibility that the results are obsolete by the time they become available.

Instead, we can characterise the question as a problem of decision-making under uncertainty. The scientific method which EBM seeks to apply to medicine can be seen as a formalised process for the valid acquisition of knowledge. Similarly, we can talk about a formal methodology to validate the making of difficult decisions – decision methodology. There have been many and varied attempts to formalise such a methodology. However, there are general principles and concepts that tend reappear in different decision methodologies. We can illustrate these principles by applying them to the specific problem faced here – the evaluation of a proposed breast cancer screening programme.

Any decision involves a choice between a set of mutually exclusive options. Formal approaches to decision-making require that these options are made explicit. This may not be a trivial exercise, as decisions can often be broken down into combinations of choices. Our decision problem seemingly has two options – to screen or not to screen. If we choose to screen, however, this involves further decisions about who to screen, and how often. If we choose not to screen, there is an option to carry out further research, in order to revisit the decision in the future.

Once we have identified the options, we need to analyse the consequences of choosing each one. It is important to identify all the relevant consequences of an option. To get a complete picture involves thinking about consequences in the broadest terms, and also examining carefully the link between immediate and indirect outcomes. With screening, the obvious consequence is the saving or prolonging of lives. However, there are other consequences – possible over-treatment, anxiety from false positive diagnoses, reassurance from participation, increased pressure on the time of specialist staff, and so on. Also, improved survival for cancer patients is not a direct consequence of screening. The direct consequence is the early detection of cancer in some patients. The relative impact on survival will depend on how early detection changes treatment choices and outcomes. It will also depend on the relationship between disease progression, onset of symptoms, and prognosis.

Closely related to the analysis of consequences is the gathering of relevant evidence. A structured analysis of consequences will determine what evidence is relevant to the decision problem. Also, through gathering evidence, we may gain new insights into the consequences of the options being evaluated. Evidence is defined here in a broad sense. For our decision problem, it will include evidence from trials of mammography, but it will also include trials of breast cancer treatments, observational studies, health care databases, and scientific studies of the biological processes involved in breast cancer. The quality of a piece of evidence will influence the weight it plays in decision making, but it will not necessarily exclude it from the process.

Having gathered evidence and analysed all the consequences of each option in our decision, we should be in a position to choose between them. There are several issues that arise at this stage. There may be a range of consequences that are quite different in nature, and different options may favour different outcomes. With screening, for example, consequences include extended life spans, pain and discomfort from medical interventions, psychological well-being, and lost productivity. We could approach this problem by choosing a key consequence to prioritise. However, the usual approach in decision methodology is to establish a common metric that can be used to quantify disparate consequences. Money, utility and QALYs are examples of such metrics.

Then, it is possible to aggregate the positive and negative consequences of each option.

An issue that almost always arises in a decision problem is uncertainty. Due to gaps or weaknesses in the evidence, and the general unpredictability of the real world, there will be a degree of uncertainty in our predictions of outcomes. An element of decision methodology is an approach to balancing risk and reward. The most well established approach for this is expected utility maximisation [28].

The approach described above might seem a description of a self-evident process that is already included in current practice. Does it add to, or change, medical decision-making in any way? One aspect in which it does is that it involves a more formal and structured approach than is currently used. Key to this is the use of mathematical models. These models can be used to describe the decision itself, or aspects of the problem such as the natural history of a disease. The benefit of this is that it highlights complex aspects of the decision process, such as the dependence of screening on treatment. Also, a model provides a structure to synthesis a broad range of evidence. The value of this is discussed below.

Another important affect of the approach described here lies in the treatment of information and uncertainty. The scientific method on which EBM is based is designed to drive out uncertainty by performing well-designed experiments. Hence EBM is interpreted to involve establishing the effectiveness of an intervention beyond reasonable doubt (defined through significance levels in statistical hypothesis testing) before it is introduced into practice. This ignores the feasibility, practical difficulties and costs of obtained enough information to reduce uncertainty to the required level. In the case of cancer screening programmes, as we have seen, these are likely to be substantial.

However, we can apply the decision methodology described above to the sub-problem of decision uncertainty. This approach is known as ‘value-of-information’ analysis [29]. It involves recognising that we are rarely in a state of complete ignorance prior to a proposed experiment. The choice of whether or not to conduct a trial can be seen as a decision problem. The outcome of conducting the trial will either be that it changes

the original decision over implementing a treatment, or that it doesn't. The expected impact of the new (or better) study in influencing the decision can be weighed against its costs. Note that any delay to implementing an intervention that turns out to be beneficial whilst we wait for conclusive proof is a cost of obtaining that proof. Chilcott et al discuss in depth the application of this approach to the prioritisation of clinical trials [30].

This approach reflects a key conceptual difference between EBM and decision-analytic medicine (DAM) – the response to uncertainty. EBM seeks to eliminate uncertainty, whereas DAM seeks to quantify it. DAM relies on modelling to do this. Through a well-constructed model, a broad evidence base can be synthesised to get the best possible representation of current knowledge. This is crucial to predicting the likely outcomes of further experiments, and estimating the marginal enhancement to our knowledge base. This process is most valuable where gathering strong evidence is expensive and time-consuming, as is often the case with cancer screening programmes.

DAM represents an approach that strengthens the link between medical decision-making and its goal – making the best decisions for patients. It represents a more sophisticated approach to the uncertainty of medicine than the arbitrary thresholds of significance used in conventional hypothesis testing. However, its use outside the field of health economics is limited. This may be because its approach to uncertainty is seen as unethical – that it is wrong for clinicians to offer treatment if they are not completely sure of its value. My discussion so far presents the counterargument that there can still be a significant degree of uncertainty for patients, or sub-groups of patients, after RCT evidence has been obtained. DAM makes this uncertainty an explicit part of the decision-making process.

Modelling is a key part of DAM. We represent our understanding of the decision problem through building a model. This allows us to interpret and synthesise all the available evidence. It does this because, in building the model, we extract from the problem its key elements. We can then translate these elements, and the relationships between them, into a formal mathematical structure. This process of abstraction and formalisation is an essential part of any attempt to understand the world. However, it

creates a problem which goes some way to explain why the DAM-type approach is largely restricted to health economics. The choice of which elements to include in the model, and how to represent them, is a subjective one. If it is done poorly, the analysis derived from the model is suspect. The challenge for those who wish to see a greater role for decision science in evidence-based medicine is to develop a formal methodology to ensure a close relationship between the model and the clinical problem.

Reference List

(1) Gold RH, Bassett LW, Widoff BE. Highlights from the history of mammography. RadioGraphics 1990; 10(6):1111-1131.

(2) Lerner BH. "To see today with the eyes of tomorrow": A history of screening mammography. Canadian Bulletin of Medical History 2003; 20(2):299-321.

(3) Shapiro S, Strax P, Venet L. Periodic breast cancer screening in reducing mortality from breast cancer. Journal of the American Medical Association 1971; 215(11):1777-1785.

(4) Newcomb PA, Lantz PM. Recent trends in breast cancer incidence, mortality, and mammography. Breast Cancer Research and Treatment 1993; 28(2):97-106.

(5) Harris KM, Vogel VG. Breast cancer screening. Cancer and Metastasis Reviews 1997; 16(3-4):231-262.

(6) Shapiro S, Coleman EA, Broeders M, Codd M, de Koning H, Fracheboud J et al. Breast cancer screening programmes in 22 countries: Current policies, administration and guidelines. International Journal of Epidemiology 1998; 27(5):735-742.

(7) Gotzsche PC, Nielsen M. Screening for breast cancer with mammography. Cochrane Database of Systematic Reviews 2006; 2006(4).

(8) Bailar III JC. Mammography: a contrary view. Annals of Internal Medicine 1976; 84(1):77-84.

(9) Wells J. Mammography and the politics of randomised controlled trials. British Medical Journal 1998; 317(7167):1224-1229.

(10) Olsen O, Gotzsche PC. Cochrane review on screening for breast cancer with mammography. Lancet 2001; 358(9290):1340-1342.

(11) Gotzsche PC, Olsen O. Is screening for breast cancer with mammography justifiable? Lancet 2000; 355(9198):129-134.

(12) Gelmon KA, Olivotto I. The mammography screening debate: time to move on. Lancet 2002; 359(9310):904-905.

(13) Miller AB. Commentary: A defence of the Health Insurance Plan (HIP) study and the Canadian National Breast Screening Study (CNBSS). International Journal of Epidemiology 2004; 33(1):64-65.

(14) Duffy SW, Tabar L. Screening mammography re-evaluated. Lancet 2000; 355(9205):747-748.

(15) Duffy SW. Interpretation of the breast screening trials: a commentary on the recent paper by Gotzsche and Olsen. Breast 2001; 10(3):209-212.

(16) Miettinen OS, Henschke CI, Pasmantier MW, Smith JP, Libby DM, Yankelevitz DF. Mammographic screening: no reliable supporting evidence? Lancet 2002; 359(9304):404-405.

(17) Miller AB. Screening for breast cancer with mammography. Lancet 2001; 358(9299):2164.

(18) Gotzsche PC. Screening mammography: setting the record straight - Reply. Lancet 2002; 359(9304):440-441.

(19) Gotzsche PC. On the benefits and harms of screening for breast cancer. International Journal of Epidemiology 2004; 33(1):56-64.

(20) Olsen O. Beyond randomized controlled trials - Organized mammographic screening substantially reduces breast carcinoma mortality. Cancer 2002; 94(2):578-579.

(21) Horton R. Screening mammography - an overview revisited. Lancet 2001; 358(9290):1284-1285.

(22) de Koning HJ. Mammographic screening: evidence from randomised controlled trials. Annals of Oncology 2003; 14(8):1185-1189.

(23) Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence based medicine: What it is and what it isn't. It's about integrating individual clinical expertise and the best external evidence. British Medical Journal 1996; 312(7023):71-72.

(24) Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence based medicine: What it is and what it isn't. It's about integrating individual clinical expertise and the best external evidence. British Medical Journal 1996; 312(7023):71-72.

(25) Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". Lancet 2005; 365(9453):82-93.

(26) Sehon SR, Stanley DE. A philosophical analysis of the evidence-based medicine debate. BMC Health Services Research 2003; 3:1-10.

(27) Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence based medicine: What it is and what it isn't. It's about integrating individual clinical expertise and the best external evidence. British Medical Journal 1996; 312(7023):71-72.

(28) Clemen R. *Making Hard Decisions: An Introduction to Decision Analysis*. Belmont CA: Wiley; 1996.

(29) Briggs AH, Sculpher M, Claxton K. Decision modelling for health economic evaluation. Oxford: OUP; 2006.

(30) Chilcott J, Brennan A, Booth A, Karnon J, Tappenden P. The role of modelling in prioritising and planning clinical trials. [Review] [133 refs]. Health Technology Assessment (Winchester, England) 2003; 7(23):iii-125.