senter for teknologi og samfunn

STS

Ann Rudinow Sætnan

TO SCREEN OR NOT TO SCREEN?

The Impact of Science on Two Medical Technology Controversies

STS-arbeidsnotat 17/92

ISSN 0802-3573-61

arbeidsnotat
working paper

senter for teknologi og samfunn
universitetet i trondheim
7055 dragvoll

tlf: +47 73 59 17 58
fax: +47 73 59 13 27

university of trondheim
centre for technology and society
n-7055 dragvoll, norway

TO SCREEN OR NOT TO SCREEN?
The impact of science on two medical technology controversies

1. "To screen or not to screen?" What sort of question is that?

The question "To screen or not to screen?" is a special class of question in medical technology. "Screening" refers to the application of diagnostic technologies to a whole population. This as opposed to clinical (or selective) applications where only patients presenting with symptoms or other special indications for a given technology are considered.

In many ways, screening decisions are like all other medical technology decisions. Patients may individually decide to comply or not comply with medical advice, to attend or not attend when invited to an examination. Health providers may offer or not offer a particular treatment or diagnostic procedure. But prior to these individual decisions, it is generally assumed that the procedure in question has been approved, recommended or required by some central authority. To some extent, medical technology "decisions" are the end result of atomized, more or less haphazard processes; but, to some extent, they are also identifiable bureaucratic events which are expected to be conducted on (as near as possible) a rational basis.

Screening decisions encompass all the issues common to questions of the choice of medical technologies in general, plus questions which arise out of the scale of the proposed screening program. In general, controversies over medical technologies encompass a set of sub-controversies: Does the technology "work?" If so, under what conditions? What are the specific indications for its use and benefits to be gained from using it? In what ways, to what extent, and under what conditions is it harmful? What are the costs and benefits of the technology -- not only in terms of money spent and saved, but also in terms of alternative uses of resources (manpower, skills, etc.), health risks, ethical trade-offs, pain endured ... The list of costs and benefits considered is potentially endless, but empirically bounded: Some questions are successfully introduced into the debate in a given instance, either by individuals and/or by central authorities; others, regardless of theoretical relevance or experience from other debates, are not.

The sub-issues in medical technology decision situations are the same whether the technology in question is being considered for clinical (selective) usage or on a population level as a screening or general preventive technology, but are magnified by the scaling up of the procedure. Costs are greatly expanded. Benefits also change in magnitude -- sometimes expanding, sometimes shrinking. As screening programs are directed at an entire population, screening decisions are likely to be more centralized than decisions concerning selective applications of technologies.[1] Due to these magnification and centralization effects, the discourse involved in reaching a screening decision is likely to be highly visible and thereby accessible for research purposes.

This paper is the first of several which will analyze two cases of screening technology decisions in Norway: ultrasound screening in pregnancy and mammography screening for breast cancer. Here, I will examine the role played by science in the two decision-making processes. Did science made clear recommendations in the two screening controversies, and, if so, were those recommendations determinant in the decision-making process?

This begs the question of why we should expect science to have any impact on (medical) technology controversies, and there are a number of reasons why we should NOT. One is the scope of the sub-issues potentially involved, many of which (such as the value of a life saved relative to costs) are commonly seen as outside the purview of science. Another is that the sub-issues are inextricably intertwined so that even those seen as within the purview of science cannot be decided by science alone. And even if one could isolate some 'strictly scientific' question from the whole, some argue that technology questions are settled differently than in science. Constant[2] proposes that in spite of some basic similarities (rigorous empirical testing, communities of practitioners), and aside from any consideration of the social, moral and economic issues involved, science and technology differ in two key areas: the hierarchical structure of practice, and satisficing modes.

Constant sees technological practice as more hierarchically organized than science. Where science bounds discourse within separate disciplines, technology decomposes whole projects into sub-projects and forces cross-disciplinary cooperation to solve problems encountered by the project as a whole. While both science and technology can be seen as satisficing -- seeking solutions that are "good enough" -- Constant proposes that their criteria differ: Science explores a vicarious environment; technology is confronted by a real one. Scientific anomalies can be explained and reinterpreted endlessly; technological artifacts either work or they don't (planes either fly or crash, bridges stand or fall). Addressing that situation, technologists accept simplifications and compensate by over-dimensioning; scientists, on the other hand, seek sophistication and precision.

Whether or not we see science and technology as separate practices, we do expect science to settle technological controversies all the time.[3] In medicine, science is legally mandated to take a decisive role. For instance, in many countries new drugs must, by means of a series of scientific experiments culminating in replicated randomized controlled clinical trials (RCTs), be shown to be safe and effective before they are allowed to be marketed. Other medical technologies are expected to conform to the same standard of proof as a matter of professional ethics.[4]

2. *How does science settle controversies?*

Sociological theories of science differ as to how science arrives at conclusions, but all see this as a process which takes some amount of time. Consensus does not arise immediately or automatically from a correctly conducted experiment. In this section I will present three theories on how science reaches consensus. These theories imply different indicators of whether consensus has been achieved.

2.1. *The Mertonian model*

According to the traditional Mertonian model[5], science reaches consensus on questions about its objects of study through adhering to two sets of norms: technical norms and social norms. **Technical norms** pre- and proscribe methods of research. Of the technical norms, that requiring *logical consistency* is universal to all fields of science, while *rules of empirical evidence* are specific to their respective fields and historical times. The status of RCTs as a "gold standard" for measuring the safety and effectiveness of medical technologies would be an example of a field-specific technical norm. **Social norms** (*communism, universalism, disinteredness, and organized skepticism*) prescribe a set of behaviours within the scientific community which serve both to delineate that community and to enable it to police adherance to technical norms.

Consensus, according to Merton's model, is the end result of a period of organized skepticism. Various rewards can serve as indicators of that result. Publication and citations especially are often used as indicators of the quality and recognition of scientists' work.[6] They may, however, also be negative sanctions within the reward system. "Organized skepticism" is not brought to a conclusion by referees alone. Acceptance for publication is only a first step in making work available for critical evaluation by a scientist's peers. Chastisement by those peers may also occur in the form of citations, and referees and editors may even publish faulty work explicitly to subject it to such chastisement.[7] Consensus, then, is best indicated by the relative numbers of positive and negative citations to published work, rather than by total numbers of publications or citations. This differentiation, though relevant for the Mertonian model, is more closely associated with the discourse model discussed below.

2.2. *An interest model*

Two groups of constructivist theories[8] -- discourse models and interest models -- are relevant to analyze the issue of consensus. They share some precepts but differ on others concerning measuring scientific consensus.

The two groups share the concept of "interpretive flexibility" -- that there is never only one logically defensible interpretation of a given set of observations. Thus, experimentation alone cannot be decisive; social negotiations are required for an interpretation to be accepted. Merton's model also requires social negotiations, but only to control that previously approved methods and logic have been applied. In constructivist models, the very meanings of the data are socially negotiated. Collins[9] calls the demonstration of the existence of this phenomenon the "first stage" of an empirical program of relativism. Collins' "second stage" is the description of mechanisms through which interpretive flexibility is limited, allowing consensus to arise. Discourse models can be seen as directed at this problem, with Actor-Network Theory as an exemplar of the group. As this theory is more recent, I will return to it later.

According to Collins, "stage three" of the empirical program of relativism explains the directions taken by different groups' interpretations in terms of the groups' interests. The interests claimed as influential are not merely the communal

interests of science as a whole or authors' personal interests in pursuing a career within the scientific community; they are particularist interests in pursuit of specific goals, including social and political goals which would be considered external to science according to the Mertonian norms of disinterestedness and universality.

In "Scientific judgment: The biometry-mendelism controversy," MacKenzie and Barnes[10] reject explanations based on different sets of technical norms or background texts, and propose a goal-oriented, interest-based explanation of the controversy. MacKenzie and Barnes' method is to identify communities on the basis of shared activities, goals, and truth claims. They then seek to demonstrate empirical and logical links between the respective communities' goals and truth claims. Inter-community controversies can continue indefinitely; but they can also be, or at least appear to be, resolved. Resolution can occur if controverted truth claims and methods come to be seen as serving the goal-interests of communities previously in conflict with them. Or it **appears** to occur when new communities with new goals find the claims and methods of several existing, conflicting communities useful.

2.3. A discourse model

In a sense, Actor Network Theory[11] steps back to the second stage of the empirical program of relativism by bracketing the question of interests. While accepting that interests do affect interpretations, Actor Network Theory does not assume a simple one-to-one relationship. Group identities are not given a priori, and causes (interests) are not readily deducible from effects (interpretations). Regardless of the variety of interests which may result in a given interpretation, however, the aggregate effect on consensus, arising from the working of texts upon texts, will be the same.

According to Actor Network Theory, scientists "invest" their credibility, or intellectual capital, in a cycle of resource acquisition. Submitting results to the scientific community is one of many investments which may be made in different "markets." In each market, the scientist seeks allies who will support the scientist's project in various ways. In the scientific community market, support from allied peers is offered in the form of acceptance of manuscripts by referees, positive citations from readers, offers of employment, apprenticeship by students, etc. This support, however, may not be a reward for good work; it is offered in the allies' self-interest in furthering their own work. From the point of view of a cited scientist, each accepted manuscript, each positive citation increases the credibility value of my work. From the point of view of a citing scientist, it may serve my interests to cite work which I don't find very credible (for instance because others do, or because a referee has advised me to do so). It may serve my interests to distance myself from work which I find highly credible (for instance if association with that work might disqualify me from a job, or might make my own work seem less original). I may even apply a modality inadvertently: I may not reflect on my choice of the phrase "Latour claims" as opposed to "Latour has shown," but the result is there nonetheless. If I write "Latour claims," his statement is described as a mere speech act and the content of his claim as an artifact of that act; "Latour has shown" points to phenomena outside Latour's speech act and the truth value of his claim is supported by my implied acceptance of the objective existance of those phenomena. Regardless

of my intention to reward or chastise, citation of another's work has the effect of transforming it from artifact to fact or fact to artifact. It is through this rhetorical process that science arrives at and announces conclusions.

Another phenomenon Latour points out is that the content of the statement itself is modified by its later users, even by those applying positive modalities as they help establish it as fact. The following example illustrates this point:

"Ultrasound may become the best screening test for open NTDs.[93,94 n12]

In form, this citation has a strong positive modality. Sources 93 and 94 are implied to be previously established facts supporting the author's new statement. The text referred to in note 94, however, does not mention NTDs (neural tube defects) and does not recommend ultrasound screening. In citing a text, a reader's co-authorship of the text becomes apparent. It is not the text itself, but the reader's own interpretation of the text which is cited -- a point which is especially clear when the text is so radically reinterpreted as in the sample above. Thus, according to Actor Network Theory, in searching for consensus in the science community we would need to track changes in modalities referring to a truth claim and in content attributed to that truth claim over time.

As previously stated, the aim of this paper is to examine **whether** medical science made clear recommendations in the two screening controversies, and, if so, whether those recommendations were decisive in settling the controversies. This does not require that I choose among the three models of **how** scientific controversies are resolved. Instead, I will apply methods from all three. While the three models are associated with different methods, they do overlap in that citations can serve as indicators in each model. For this indicator, the differences among the models lie in different interpretations and different utilization of the content and context of the citations.

In choosing citations as indicators of controversy and consensus, I am combining two traditions in social studies of science -- bibiometrics and controversy studies. Though not entirely new (as witness my sources for theory and method in this article), this combination is not common. Bibliometric methods are most commonly used to study scientific stratification (performance, status and impact using citations as a measure of rewards) and specialization (diversification and selection processes using shared citations as a measure of community membership).[13] Resting on the Mertonian model, this tradition brackets questions about the content of controversy and consensus since these must be settled within science communities according to technical and social norms. The study of controversies within the Mertonian model is generally the study of rewards and sanctions applied to enforce those norms. In the interest and discourse models, which do aim to study the content of scientific controversy and consensus, bibliometric methods play a minor role. More emphasis is placed on studies of rhetoric and on the institutional and historical contextualization of that rhetoric.[14] For my purposes, however, three styles of citation analysis will serve to answer my main question from the standpoint of each of the respective models. In the following section I will present the two controversies which will be examined, the source and citing texts which form the data base on each controversy, and the analysis techniques which I will apply.

3. Two data sets, three methods

I will examine two cases in which readily identifiable decision events occurred at the level of the Directorate of the national health service. In both cases, conflicting evidence from randomized controlled clinical trials (RCTs) had been presented at the time when these decisions were reached.

3.1. The ultrasound controversy

In the Fall of 1986, following a consensus conference,[15] the Directorate of the Norwegian national health service decided to recommend that an ultrasound scan be offered to all pregnant women at around the 17th week of pregnancy.[16] A survey of obstetrical practice conducted in preparation for the consensus conference showed that obstetrical units responsible for 68% of all deliveries in Norway already offered an ultrasound screening program and that 94% of all pregnant women delivering during the week of the survey had received at least one ultrasound scan. On the average, each woman had been scanned 2.45 times.[17] De facto, ultrasound screening in pregnancy was already a fully diffused technology in Norway at the time of the official decision.

As of 1986, results of four RCTs on the routine use of ultrasound in pregnancy had been published. Three of the RCTs (the London trial,[18] the Glasgow trial,[19] and the Trondheim, Norway trial[20]) reported no statistically significant health benefits from screening as opposed to selective use of ultrasound. One Norwegian study (the Ålesund trial[21]) reported a number of statistically significant benefits in the screened group: decreased hospitalization rates, fewer inductions for post-term pregnancy, fewer days of hospital stay for hyperbilirubinaemia, higher birthweights among twins, fewer deaths due to growth retardation, and generally reduced mortality and morbidity (statistical significance not reported). This study, however, had been published only as a letter to the editor of The Lancet and as a paper at the consensus conference held by the National Institutes of Health in 1984. The study had come under criticism for inadequate publishing and for discrepancies between the two reports.[22]

I will focus on the reports from two Norwegian trials,[23] as these were the focus of the Norwegian consensus conference. They have also been seen as particularly important by several review authors[24] since they followed the same design but arrived at very different conclusions. The two trials were, at least initially, a cooperative effort. Eik-Nes (a gynecologist and international expert in ultrasound diagnostics) trained the ultrasound operators at both hospitals and appears as a co-author also in the publications from the Trondheim trial. Bakketeig (an international expert in perinatal epidemiology) designed the protocol for the trials, although the authors of reports from the Ålesund trial have not listed him as co-author. Since both first authors were involved in both texts, I will refer to the two articles by their geographic sources as the Trondheim article (Bakketeig et al., 1984) and the Ålesund article (Eik-Nes et al., 1984) respectively.

Citing texts were identified using three sources. The main source was Science Citation Index for the years 1984-1991. Indeces from 1986-91 were available on CD-rom. For these years it was also possible to check for related articles (articles sharing

one or more references with an article already identified) and check these for references in which Eik-Nes's or Bakketeig's name had been misspelled. Finally I checked for the same misspellings in the hard-copy indeces for 1984 and 1985. A total of 55 articles (including one of the source articles) citing one or both of the source texts were found through the Science Citation Index. Of these, one article proved unobtainable.

The Science Citation Index does not index Tidsskrift for Den norske Lægeforening (TfDnL, the journal of the Norwegian medical association) nor the International Journal of Technology Assessment in Health Care (IJTAHC), journals relevant to Norwegian health policy debate and international health technology debate respectively. By checking the reference lists of all articles indexed under the keyword "ultrasound" in the index issues of TfDnL for the years 1984-91, one letter to the editor citing the source articles was identified. IJTAHC was founded in 1985 and comprises 30 issues so far. A search of the reference lists of all likely articles (judged by title) revealed one article referencing the two source texts.

These data are not a random sample from some larger "population" of articles. They are the entire universe of articles in indexed journals and in two other particularly relevant journals - all in all presumably the most influential set of relevant journals - in which our two source articles are cited. Statistics estimating the probability that correlations found in this universe are representative of some larger universe are not appropriate. Raw correlations are real correlations and raw distributions can serve as a measure of the robustness of these correlations.

3.2. The mammography controversy

Norway's second consensus conference was held February 1989 on screening for breast cancer with mammography. Again, the conference conclusions became official policy. This time the conclusion was not to recommend screening until further evidence confirmed benefit and clinical follow-up capacity had been developed.

In this case three RCTs had been published at the time of the consensus conference. Two -- the Health Insurance Plan of Greater New York (HIP) study, and the Swedish two-county (WE) study -- had reported finding significant reductions in mortality from breast cancer. In the HIP trial, however, the study group was offered a combined screening program of both mammography and clinical examination (palpation), making it difficult to estimate the individual contributions of the separate diagnostic techniques. Meanwhile, mammography technology had evolved substantially since the HIP study, improving image clarity and reducing radiation exposure and thereby making a test of mammography alone more attractive. The WE trial reported a significant reduction in breast cancer mortality after screening with mammography alone.[25]

A White Paper commissioned by the Norwegian health authorities[26] concluded that it was indisputable that a well-organized mammography screening program would reduce mortality from breast cancer in women over 50 years of age after a time, but that it was unclear what the magnitude of that reduction would be. The commission also concluded that clinical mammography in Norway was inadequate to receive the burden of follow-up tests a screening program would generate. They recommended that measures first be taken to ensure the availability and quality of

clinical mammography, and that a screening program for women 50-74 years of age be initiated following those measures.

As of 1989, these recommendations had not yet been acted upon. Although screening was offered by several private radiology clinics and clinical (selective) mammography was offered by most hospitals, only about 60 000 women had mammograms in 1987.[27] Screening a target population of women ages 50-74 would have resulted in approximately ten times that number of examinations.[28] The 60 000 examinations did not, however, represent one tenth of such a screening program, since they also included an unknown number of voluntary screenees under age 50 and another unknown number of selective examinations of women presenting with symptoms. Meanwhile, another RCT (the Malmö, Sweden study[29]) had reported an observed mortality reduction, but the difference between study and control groups had not achieved statistical significance during the time span set in the trial protocol. Although the researchers themselves recommended screening (on the basis that the study and control groups had been more "diluted" than anticipated by non-attendance and voluntary mammography respectively) these results encouraged others toward greater skepticism. Mammography screening was neither fait accompli nor undisputed recommendation when the consensus conference convened.

Later, a Canadian trial[30] and a trial in Stockholm, Sweden,[31] both published confirmational results. In 1992, results of a meta-analysis of the three Swedish trials were presented at a national conference.[32] Unsurprisingly (the WE-study being by far the largest of the three Swedish trials), the meta-analysis found significant survival benefit for screened women. Earlier that year, the Norwegian minister of health and social services announced that the government was prepared to contribute financing to a trial project to test organizational models for a mammography screening program.[33] That program is now under planning for possible implementation in three counties in 1994.

The Norwegian consensus conference on mammography screening for breast cancer focussed on three randomized controlled clinical trials which had been published by that time, and I choose to do the same. (The more recent decision to begin implementing screening is explicitly based on the Swedish meta-analysis, which is not yet published. Effects of the Stockholm and Canadian trials will be measured only indirectly as changes (if any) in consensus on the previous three trials.) The three are the Health Insurance Plan of Greater New York (HIP) study, the Swedish two-county (WE) study and the Malmö study. Each of these studies has resulted in a number of publications, not all of which I have been able to obtain. Using the "related articles" and "citations" features of Science Citation Index and the citing texts contents, I identified 32 articles used by later citers to refer to the HIP study. Six of these were cited more than two or three times with reference to the main conclusions of the study, and five of these again proved obtainable.[34] I have used these five as a basis for coding decisions.

Similarly, I identified 20 articles stemming from the WE study, seven of which[35] were the prime sources with reference to the screening issue in later citing articles, and eight from the Malmö study, one of which[36] was cited frequently for the study's main conclusions.

Due to the large number of potential source articles and the many authors' names and potential misspellings thereof, I made extensive use of the "related articles" feature and, when in doubt, included possible citing articles in the search

file. My initial search resulted in 122 possible citing articles from 1986, 111 from 1987, 144 from 1988, 168 from 1989, 180 from 1990, 155 from 1991, and 214 from the first six months of 1992. To cope with this large mass of texts, I decided not to search non-indexed journals and to limit my analysis to the 1986 and 1992 citations. Many citing texts turned out to refer to other sources by the HIP, WE and Malmö authors or by authors with similar names. This was especially the case for 1992, which was the first index I searched. Eliminating articles unobtainable through interlibrary loan and articles which turned out when obtained not to contain the searched references, the 1986 material was narrowed down to 89 articles and the 1992 material to 57 from the first six months.

In a sense, the mammography data have been sampled. Nonetheless, statistical tests of significance are inappropriate. The data represent, as in the ultrasound case, complete populations of citing articles, but from two points in continuous time. If these two points are taken to represent the whole span of that time, then the course of the development of consensus is taken to be linear. None of the theories being applied in the analysis assume linearity. Nor do they assume that any particular form of line or curve is more stable than another; a horizontal line (the same degree of consensus over many years) is as probable as a vertical one (sudden rejection or acceptance) or an evenly sloped one (gradual rejection or acceptance) or one that changes among all these shapes any number of times. A claim's position on a scale of acceptance at any given point in time is no basis for prediction of its position at some later point in time. Thus, there is no basis for testing for probable departure from linearity or for empirical departure from predicted degree of consensus. We must refrain from using the two points in time as indicators of the longer period, and see them instead as complete representations of those two points.

3.3 Three methods

I have analyzed the data in three steps. The first step refers to the *Mertonian model.* In this step I interpret citations, regardless of modality, as rewards. To differentiate rewards between positive and negative citations, I recorded the citing articles' main conclusions according to their agreement or disagreement with the conclusions of the source articles.

The second step of the analysis refers to the *interest model.* In this step I have followed up on hypotheses referring to authors' identities and interests, and arising from a closer reading of the texts. One hypothesis was that gender issues (both proposed screening programs are directed at women) might influence standpoints. To test this, I compared articles' main conclusions with their authors' gender (one or more female (co-)authors, no female authors). Another hypothesis was that there might be two medical research communities, one oriented towards clinical practice and one towards epidemiological issues. This was tested by linking main conclusions with supporting arguments and with citations to central clinical or epidemiological texts other than the source texts. This analysis was carried out only on the ultrasound material.

The third step refers to the *discourse model* and applies the modality scale presented in Figure 1. For this step, I read each citing text, identifying explicit and implicit references to the source texts and coding these according to the version of

the source text or texts they referred to and the modality applied in the reference. This process posed a number of problems. The first was how to decide what constitutes a single statement. It became immediately clear that there were no inflexible syntactic rules to follow. A paragraph might constitute a single statement; a sentence might break down into several statements. In retrospect I realized that the coding scheme and my active reading/rewriting of the texts were guiding my decisions: Each change of modality or referent text constituted a new statement. Thus my results are in part artifacts of my analytical instruments and personal perceptions, as in all science.(ref) Taking a more positive view, the analytical instrument also serves to discipline my personal reading of the texts, thus reducing the effect of ideosyncratic perception.

A second problem was determining which statements referred to the source texts. Frequently a citing author would footnote an initial reference and then continue for some paragraphs referring to the same text by means of an appropriate generic term or pronoun. Eventually - other references and long passages having intervened - the referent sometimes became unclear. I have only counted statements which I was absolutely certain referred to one or more source text(s).

The third problem was assigning referent text versions to the statements. References to articles claiming positive findings and stating positive conclusions (the Ålesund, HIP, and WE articles) were generally not a problem. References to these claims varied up to +/- 10% in the relative survival benefit attributed to the source claims, sometimes due to differences in the delineation of the precise populations referred to. I ignored that detail, aside from which all references referred correctly to the findings and conclusions of the source texts as positive. References to the two articles with non-significant results (the Trondheim and Malmö articles) sometimes claimed these source texts to have positive results. The Trondheim article points out some non-significant positive trends in the trial data, but does not claim these as confirmational results and concludes against screening until confirmation has been found. I have therefore coded all references to confirmational results attributed to the Trondheim trial as references to a radically reauthored version of that text, either to its partial or main conclusions depending on the precise formulation of the reference. The Malmö authors present their results as positive. The reservation that the results are non-significant appears only in a footnote. The non-significance of their results is explained in terms of "polution" of the test and control groups through non-compliance and voluntary screening respectively, and the authors conclude in favor of screening. References to the Malmö trial as confirmational I have coded as positive references to the main conclusions and/or (depending on their precise formulation) negative references to partial claims, both in their original form. References to the Malmö trial as non-confirmational I have coded as positive references to partial claims and/or negative references to main conclusions.

The fourth problem was assigning modalities to the statements. Latour[37] describes a method for assigning and interpreting citation modalities. The method involves classifying individual statements[38] involving explicit or implicit citations according to their semantic structure and function. Positive modalities point toward the referenced objects of the cited text, reasserting the factual status of the text independent of the conditions of its production. In other words, the statement cited is taken to be objectively true. Negative modalities point toward the conditions of production of the cited text, realerting readers to the conditional, artifactual status of

Figure 1: An illustrated scale of citation modalities

Classification	Semantic structure	Example
Factual +++	Previously established fact silently assumed in production of new statements	none
Factual +++	Previously established fact used as support for new statement without citing source	none
Factual ++	Cited text used as support for new statement	"(U)ltrasound screening has been shown to reduce perinatal mortality and morbidity.[4,73]
Consensual +	New results taken to support those of cited text	"This study has confirmed that a single BPD measurement in the third trimester is a poor predictor of SGA.[28,74]
Neutral -/+	Citation points to conditions of production of cited text in neutral or supportive terms	"Allocation of subjects: This was done using formal randomization with opaque, sealed envelopes in the Norwegian trials.[75]
Non-consensual -	Citation claims new results contradict claims of cited text	"In contrast to the trials of Bakketeig et al[22] and Walden-strom et al,[23] a significant reduction in the number of inductions for post-dates pregnancy was not shown in this study."[76]
Artifactual --	Citation claims errors in production of source statement	"A possible source of bias of the Ålesund study was that fetuses with anatomic defects detected by screening were deleted from the screened group, resulting in some loss of randomization."[77]
Artifactual ---	No citation ever	none

the source text and thus weakening its status as objective fact. Given that a statement gets cited at all, it will go through a career of modalities moving it gradually towards status as fact or as artifact. Early in this career, even citations with supportive intent may be structured as negative modalities: Since the statement is still unestablished as fact, they may point to the quality of the research behind it in order to strengthen it. This is one indication that we have a scale, rather than a pair of modalities, along which a cited statement will be moved over time. During initial exploration of my data, I devised the scale which is illustrated in figure 1 below.

Note that both the strongest positive and the strongest negative modalities are no reference at all. One needs to trace the course of a statement's citation career over time in order to differentiate between the two, assuming that non-citation can be identified at all. Note too that if we interpret positive and negative modalities as rewards and sanctions respectively, this classification scale is also relevant to the normative-structuralist model.

It is important to note that these modalities, following the method of interpretation implied by Latour, do not necessarily reflect the intentions of the statements' authors. The point is whether readers of the statements, guided by conventional grammatical and lexical rules, are likely to have their attention directed towards the conditional and/or controversial status of the source texts or to their factual and/or consensual status. The citing author, however, may be unaware of this effect of the precise form of the citing statement. It is after all unlikely that these authors are aware of, not to mention concerned with, this type of analysis. A pair of statements may differ in only one word (for instance "proven" or "shown" as opposed to "reported," "suggested" or "claimed") and thereby fall into diametrically opposite categories. It is by no means certain that the citing authors have made a considered choice between the words "shown" and "reported" in terms of their positive and negative modalities respectively. The formation of modalities in a citing text may serve other purposes for the citing author than that of supporting or undermining the claims of the cited text.[39] Thus the directions in which user authors move cited texts, as measured by the modalities of the statements they apply to them, may not correspond to their own opinions of those texts. Nonetheless, keeping the above-emphasized point in mind, assigning modalities to the citing statements was the most straightforward of the steps involved in the analysis.[40]

4. Did science settle the ultrasound and mammography controversies?

4.1. Distribution of rewards

In the ultrasound material, thirty-four of the fifty-seven citing texts cite both the Trondheim (Bakketeig et al 1984) and Ålesund (Eik-Nes et al 1984) articles. Twelve texts refer only to the Trondheim article and nine (including the Trondheim article) refer only to the Ålesund article. The result slightly favours the Trondheim study, but not enough to be considered a consensus. Another indicator of citing authors' (and their accepting editors and referees') intention to reward sources is the articles' main conclusions. Articles citing only the Ålesund study tend to favour screening with ultrasound in pregnancy (which the Ålesund article recommends). No articles citing only the Trondheim study (which reports not having found any

significant benefit from screening) do the same. Opinion among those citing both studies and in the material as a whole is evenly divided between support and opposition to screening. This is shown in Table 1.

Table 1. Citing articles' standpoint on ultrasound screening by source article(s) cited

	Ålesund cited	both cited	Trondheim cited	Total
supports screening	5	12	0	17
indecisive	0	7	1	8
opposes screening	2	12	3	17
other agenda	2	3	9	14
Total	9	34	13	56

Both citations and conclusions indicate an ongoing controversy. Table 2 shows the distribution of articles' conclusions on screening by their year of publication.

Table 2. Citing articles' standpoint on ultrasound screening by year of publication.

	1984	1985	1986	1987	1988	1989	1990	1991/92	Total
supports screening	1	2	3	4	2	3	2	0	17
indecisive	0	2	1	3	1	0	0	1	8
opposes screening	2	3	1	0	2	3	6	0	17
other agenda	0	0	2	0	4	3	4	1	14
Total	3	7	7	7	9	9	12	2	56

In the mammography case, the large and increasing number of citing articles per year led me to expect that the texts would reveal a similarly lasting and perhaps more heated controversy. This expectation was further encouraged by some of the articles' titles and the contrast between

these and the first few texts I read. An example might be this text excerpt which adamantly claims closure in favour of screening:

> *As Strax and Shapiro, Tábar, et al., and Verbeek, et al., have shown, there can be no doubt that screening women over age 50 with even limited examination and periods for as long as three years between screens is effective, even in the short term, in reducing breast cancer mortality. Therefore, the key question concerning early detection has been answered clearly, unambiguously and beyond doubt.*[41]

in contrast with these titles expressing doubt and controversy:

> *Breast Cancer Screening -- A Different Look at the Evidence*[42]
> *Screening Mammography for Older Women -- A Case of Mixed Messages*[43]
> *Mammography is Fallible*[44]

As I read more and more of the actual texts, however, this expectation was put to shame. Table 3 shows the distributions of the citing articles' positions on mammography screening by year of publication and by source studies cited. In 1986, 74 of the 89 articles cite the HIP study. Only three of them (4%) take a negative standpoint to mammography screening. Six (8%) are indecisive and 16 (21%) have some other agenda. 54 articles cite the WE study; only 3 of them (6%) negative to mammography screening, 4 (7%) indecisive, and 7 (13%) having some other agenda. Only four articles cite the Malmö study, which had not yet published long-term follow-up results. In other words, as of 1986 there seems to be an overwhelming consensus. Twenty articles (22%) have some other agenda, for the most part one which assumes the benefit of mammography screening. Of the articles dealing directly with the screening controversy, 55 (75%) favor mammography screening while only 4 (6%) oppose it, 10 (14%) are indecisive.

Table 3. Citing articles' standpoint on mammography screening by year of publication.

	1986	1992
supports screening	57	33
indecisive	7	6
opposes screening	4	5
other agenda	21	13
Total	89	57

In 1988 the Malmö study group published their first follow-up results. Although they concluded in favor of screening, the reported survival advantage is not statistically significant. Did this have any effect on the consensus apparent in 1986? It would seem so, but not a dramatic one. In 1992, a larger minority of the articles cite the Malmö study (15 of 57) and, compared with articles citing only the other studies, a larger but still minority proportion of these fifteen are critical to screening (three, or 20%). Articles citing the HIP- and WE-studies, however, are still by far the majority and massively in favor of screening.

Taking a Mertonian standpoint, RCTs are the method most favoured according to the techn ical norms of medical science for evaluating medical technologies. I have argued that citations

and citing article's conclusions serve as rewards within the social norm system and thus as indicators of consensus. On this basis, I find that the medical science community did not reach consensus on the merits of ultrasound screening in pregnancy during the period 1984-early 1992. In spite of the fact that only one study, published only in the form of an unrefereed letter, claimed to have found statistically significant benefits from screening, opinion among citing articles was evenly divided throughout the period. In the question of mammography screening for breast cancer, the medical science community appears to have been in a state of consensus both in 1986 and early 1992 to the effect that mammography screening was beneficial. The one study which did not find statistically significant benefits has had little or no impact so far. Both cases could be interpreted as showing a tendency towards confirmational bias. Let us now see whether these conclusions are born out according to the other two models.

4.2. Groups and interests

While advocates of the Mertonian and discourse models claim to demonstrate whether and how scientific controversies are settled, advocates of the interest model claim to be able to discover **why** specific conclusions are arrived at. They claim to find this "why" in the goals and interests of those who reach any given conclusion, and to find those by examining scientists' standpoints, arguments, and affiliations.

In the citing articles themselves, I found two clues as to what factors might be affecting how the source articles were being received. One clue came from the citing authors' names, the other from the arguments presented in their texts.

Articles are listed in Science Citation Index by authors' last names and first initials. Having acquired and read the articles, I also had access to most authors' first names. In the ultrasound case, it seemed that articles with women among the authors tended to oppose screening, while articles with no women among the authors tended to support it. Perhaps a hidden gender agenda was at work in the medical academic debate on ultrasound screening in pregnancy. The technology, after all, is applied only to women and explicit gender issues were frequent in my interview material.[45] Some issues seen as explicitly gendered by my interview informants (for instance reliance on clinical data rather than women's reports of menstrual dates for the estimation of gestational age) were present but not explicitly gendered in the citation material. The hypothesis of gender interests affecting conclusions was plausible. First names were still missing, however, for authors of 13 of the 56 articles. When the genders of the authors of all but four articles had been traced, the correlation between gender and opinion had entirely disappeared. Regardless of their classification, the remaining four articles would not have recreated a gendered pattern.

The other clue was in the arguments offered for the authors' evaluations of the cited texts and/or for their own positions on screening. There seemed to be a correlation between opinions on screening (and on screening trial results) and two different sets of criteria for what constituted adequate scientific documentation of the efficacy and safety of a medical technology. I think of the two as "epidemiological" versus "clinical" points of view. The choice of terms was not only due to the fact that the first author of the Trondheim article is an epidemiologist while the first author of the Aalesund article is a clinical gynecologist. The epidemiological viewpoint corresponds to a program of making epidemiology (or the inter-disciplinary field of medical technology assessment) a medical meta-science. This program is presented by Cochrane[46] and by McKinlay[47] explicitly as a program to change the decision-making process in medical technology questions.[48] The existence of such a program can be further documented by the

contents of the International Journal of Technology Assessment in Health Care, for instance by such special issues as that on "Educating practitioners in the appropriate use of technology."[49]

The epidemiological viewpoint is recognizably Popperian and Mertonian. Until the null hypothesis has been tested through credible methods (RCTs), rejected with statistically significant results, and the results subjected to peer review, published, evaluated and replicated, the hypotheses could not be assumed to be true. The hypotheses are that the technology is beneficial and unharmful. The null hypotheses are that there is no difference between treated and untreated groups or that untreated groups fair better than treated groups. In the ultrasound case, three out of four RCTs had found no statistically significant benefits for the screened groups. The fourth trial was both uncorroberated, inadequately published, and presented with conflicting descriptions of protocol. Later, a fifth RCT[50] also reported no significant benefits, while a sixth[51] reported a reduction in perinatal mortality which was entirely attributable to the early abortion of fetuses diagnosed with serious malformations. To the epidemiologically oriented, one could not under the circumstances assume screening with ultrasound to be beneficial. As to the question of harm, harmlessness can never be proven. Although the RCTs had not reported finding any harmful effects, no long-term follow-up studies had yet been carried out and some possible short-term effects had not been studied.

The clinical program is less readily documented. It is implied -- but also implied to be obsolete, irrational and unscientific -- by the proponents of the epidemiological program. I have found only one written source which presents the clinical viewpoint as legitimated by clinical circumstances.[52] Yet, in informal conversations, all health services researchers I asked for references on this point agreed that clinical circumstances do legitimate a non-epidemiological point of view. This because clinicians are confronted with individual patients whom they seek to diagnose, cure, and/or comfort on an individual basis. A technology which may not be demonstrably beneficial at a population level, may still be beneficial in certain individual cases. Thus, clinicians have an interest in maintaining access to the largest possible number of alternative technologies. This might lead them to credit a different sort of evidence than do the epidemiologically oriented. Rare benefits might not make for statistical significance, but could nonetheless be convincingly shown to result from the technology in question. If such benefits had been found and harm had not, then harm could be assumed to be less probable than benefit. This was how the evidential status on screening with ultrasound in pregnancy was presented by those I called clinically oriented.

The following excerpts from articles skeptical towards ultrasound screening cover several of the points I consider typical of the epidemiological viewpoint. They emphasize the value of a randomized trial, insist on the duty to publish, reject all but statistically significant results, and remind that safety cannot be assumed. They bring in the issue of costs, and position themselves as on morally superior ground to that occupied by "lay" doctors:

> *The only randomized trial to suggest that routine ultrasonography has any beneficial effects on substantive outcomes of pregnancy is that conducted in Ålesund, Norway, by Eik-Nes and colleagues in 1979 and 1980. This potentially important trial has never been fully reported in a scientific journal. The reports that are available are contradictory in a very important respect. (...) This inconsistency between the only two reports of this trial is clearly of great importance in any attempt to answer the still inadequately addressed question of whether routine ultrasonography is preferable to selective ultrasonography (as opposed to withholding ultrasonography completely). Again, substantial health service resources are involved, and furthermore, there is no basis for assuming that routine ultrasonography is innocuous.*[53]

Since clinical trials have failed to show a benefit of routine ultrasound testing, the safety of fetal exposure to ultrasound has not been definitively established, and it is likely to cost more to screen than not to screen, then it is legitimate to ask why routine ultrasound testing is a controversial issue. Why do some support its use?[54]

In contrast, these excerpts from articles supporting ultrasound screening use arguments which typify the clinical standpoint. They emphasize anecdotal evidence that ultrasound screening provides diagnosis, comfort, and/or cure even though this may not be demonstrable statistically. They reject RCTs as irrelevant and impractical and elevate benefits morally over any question of costs:

Serendipitous findings were especially important in finding fetal anomalies. Routine ultrasound provided the obstetrician with reassurance in normal pregnancies.[55]

Randomized controlled trials have not and are unlikely to answer the question of cost-benefit. It is difficult to cost the benefits, such as the savings that are associated with the abortion of an anomalous fetus. Such an analysis will vary depending on factors such as the type of equipment, operator expertise, and relative costs in different countrieas at different times. Published studies become dated rapidly as equipment and training improve, and more markers of severe disease (for example, subcutaneous occipital skin thickening in Downs' syndrome) and new diagnostic criteria (for example, the head signs of spina bifida) are described.[56]

According to the interest model, the epidemiological/clinical dichotomy in medicine can be explained in terms of goal orientations inherent to different career situations, but also exists as an ideological choice available independent of career situation. Thus it would have been both difficult and inconclusive to trace all the authors' and co-authors' places and categories of employment at the time the articles in the data set were written. Instead, I carried out a limited co-citation analysis. I chose two authors as indicators of an epidemiological orientation: Ian Chalmers and Stephen B. Thacker, both internationally known epidemiologists who had published on the subject of ultrasound screening in pregnancy and/or on the evaluation of research publications. I chose two other authors as indicators of a clinical orientation: Stuart Campbell and M Hansmann, both internationally known gynecologists who had published on ultrasound screening in pregnancy as well as on techniques for ultrasound scanning and analysis of ultrasound images. After consulting with a gynecologist research colleague, I added the names of three more clinical gynecologists with similar reputations and publication records: Per-Håkan Persson, G Gensser and Sturla Eik-Nes. In the latter instance I checked for references to publications other than those from the Ålesund study.

Table 4 shows the distribution of opinions on screening among papers citing Chalmers and/or Thacker, papers citing Campbell, Hansmann, Persson, Gensser and/or Eik-Nes, papers citing from both groups and papers citing from neither group. The table indicates several things to me. First, I take the fact that most articles which cite from neither group have some other agenda as support for the choice of indicator references. It would seem that we are looking for the key references for writers engaged in the ultrasound screening debate. Next, I see the table as supporting the hypothesised effects of an epidemiological versus a clinical point of view. None of the articles citing the indicator epidemiologists and not the indicator clinicians supports screening. The majority of articles citing only from among the clinicians support screening. Articles citing among both groups are evenly divided. Finally, I see the table as hinting at another hypothesis, namely that the epidemiological orientation has a weaker standing within medicine.

It seems that even those opposing screening are under some inclination or obligation to cite more clinical evidence than the rct's, while neither they nor those supporting screening are equally inclined or obliged to cite sources on epidemiological method or epidemiologists' evaluations of the clinical evidence.

Table 4. Citing articles' standpoint on ultrasound screening by other indicator citations.

	cites among indicator epidemiologists	cites among both indicator groups	cites among indicator clinicians	cites from neither indicator group	Total
supports screening	0	5	10	2	17
indecisive	2	2	2	2	8
opposes screening	5	4	7	1	17
other agenda	1	2	3	8	14
Total	8	13	22	13	56

The massive consensus in the mammography material makes it difficult to test for a split in interpretations of the source texts between epidemiologically and clinically oriented citing authors. Only nine out of 146 citing articles take a negative position on mammography screening. There is no point in dividing up the two groups according to co-citations (or by any other criteria for that matter) to search for differential trends. A qualitative approach does, however, indicate that a similar epidemiological/clinical split may be operative. Three of the nine critical articles focus on a close examination of the statistics of the source texts. As illustrated in the excerpts below, this focus emphasizes the implications of those statistics on a population basis:

> *As for the absolute risk reduction, women and health care purchasers should be made aware that this can range from 0.02% to 0,0001%.*[57]

> *Advocates of breast cancer screening may accurately quote the HIP and SNBH [another acronym for WE] studies as showing a 25% to 30% relative reduction in mortality rates for breast cancer in older women, but in any analysis of cost/harm/benefit, it is essential to consider the* ***absolute*** *[emphasis in original] changes in mortality rates (Tables IV and V). The absolute reduction in breast cancer mortality rate was 0.144% and 0.049% in the HIP and SNBH studies, respectively. This may also be stated differently: 1 in 694 (HIP) or 1 in 2041 (SNBH) women screened derived actual benefit in terms of increased survival with breast cancer. If this is designated as the* ***actual*** *[emphasis in original] benefit rate, then the ratio of harm to benefit for breast cancer screening ranges from 21:1 to 62:1.*[58]

Only one of the screening-supportive citing articles is similarly concentrated on statistics. This article[59] makes the point that while there may be no statistically significant difference in outcomes between the screened and unscreened groups under age 50 in the HIP study, neither is there any statistically significant difference between the screened group under 50 and the screened group aged 50 and over, for which a survival advantage has been proven. On this basis, the author concludes that the HIP study does contain evidence for a benefit from screening for women under 50. This maneuver, which discounts the role of the control group, is not likely to win acceptance from professors of epidemiology.Aside from these four who focus on statistical manipulations, all other citing authors simply take the source texts' claims of relative survival advantage at face value and accept or reject these claims on the basis of test designs and/or reproducibility -- an approach which cannot be exclusively attributed to one or the other of the two paradigms.

Summing up so far, the interest model offers, at least as a hypothesis, an explanation for the tendency towards confirmational bias found in the previous section. This bias could be due to the epidemiological/clinical split and the dominance of the clinical standpoint in the medical litterature as a whole. In the mammography case, this split could be neutralized by the statistically significant confirmational findings in the majority of the randomized trials, findings which would lead both the clinically and epidemiologically oriented to confirmational conclusions. While the hypothesis seems reasonable given available evidence, the material in this study is not sufficient for a thorough test.

4.3. Artifacts and facts

According to the discourse model, citing authors' convictions and intentions are not necessarily relevant. Regardless of authors' intent, texts turn the claims of other texts into fact or artifact. According to Latour, we should expect each source articles to have a career over the years, starting with neutral (technically negative) modalities of reference to the whole or parts of the article. If the article is successful and comes to be seen as establishing fact, references will tend more and more towards positive modalities - for the whole article if its main argument is seen as factual, for single statements if only parts are seen as factual. If the article is unsuccessful, references will tend more and more towards negative modalities.

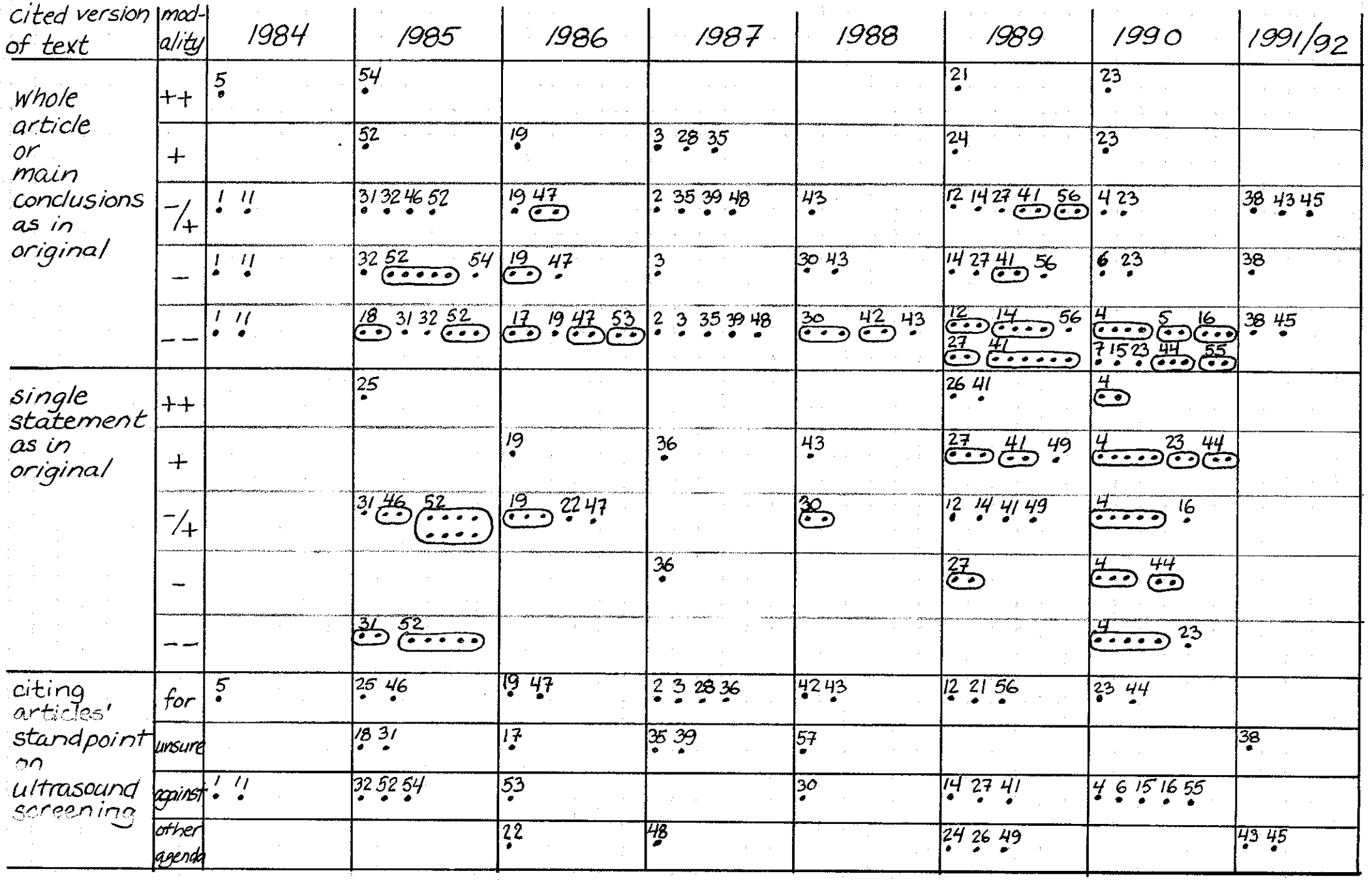
Figures 2 through 6 map the distribution of citing statements' modalities for each of the source texts. For the ultrasound source texts, modalities are shown for the periods 1984 through Spring 1992. For the mammography source texts, modalities are shown for 1986 and the first six months of 1992. Each dot marks a single statement according to year of publication, modality and version of referent text. Dots grouped together by a ring are statements from the same citing article and falling in the same category. Numbers above the dots refer to a numbered bibliography which can be obtained from me on request. For comparison, the articles' conclusions re screening are noted at the bottom of each table. This makes it possible to explore the maps in some detail, but they are also designed to be read in terms of overall graphic impressions: On graphs of successful articles and/or claims, dots should drift towards the upper right-hand corner. Unsuccessful articles and/or claims should have graphs with few marks and/or show a drift towards the lower right-hand corner.

A first graphic impression from scanning the graphs of the ultrasound articles is that the Trondheim article has been somewhat better received than the Ålesund article. Where the Trondheim article is debated, the Ålesund article seems for the most part to be rejected. This does not correlate entirely with the citing authors' own standpoints on the screening question. It

Figure 2. Modalities of references to Bakketeig et al. (1984), Jan 1984 - Mar 1992

cited version of text	modality	1984	1985	1986	1987	1988	1989	1990	1991/92
whole article or main conclusions as in original	++		18 32 52	17 19	39	30 42	14	4 15 16	
	+	11	32 52 54	19 47	48	30 43	14 27 41 56	15 16 23 44 55	38
	-/+		31 32 52	9 19 47	2 35 39 48	33 43	14 27 41 56	4 10 23 40 44	38 43 45
	-		52		3 20		41	23 55	
	--	11	18 31	9 19 47	2 3	30 37 42 43	12 14 27 56 41	4 16 23 44 55	38
single statement as in original	++								
	+			19	36	43	14 27 49 57		
	-/+		52	9 19 22		36 34	14 41 49		
	-				36				
	--		52				57		
citing articles' standpoint on ultrasound screening	for		25	19 47	2 3 36	42 43	12	23 44	
	unsure		18 31	17	35 39		57		38
	against	11	32 52 54	9	20	30	14 27 41	4 10 15 16 55	
	other agenda			8 22	48	29 33 34 37	25 49	40 50 51	43 45

Figure 3. Modalities of references to Eik-Nes et al. (1984), Jan 1984 - Mar 1992



seems to be an effect of the citations serving different purposes in the citing texts. Many screening advocates seem to have one of two reasons for claiming the Ålesund trial to have been too small, unreproduced and therefore inconclusive: either to justify their own trials as a quest to finalize a tentative finding, or because the same arguments they apply to weaken the negative trials also apply to the positive one.

Furthermore, much of the difference in trends between the two articles can be attributed to a few thorough review articles. Due to their thoroughness, each of these contributes a large number of statements about the source texts. Review articles tended to have far more citing statements than other types, and review articles tended to oppose screening.[60]

This raises the question of which is more influential on readers' conceptions of fact versus artifact.[61] Are we as readers more swayed by our own critical readings of original texts? by secondary authors' conclusions and intentions? by the sum of secondary authors' individual statements? by some secondary authors more than others? Do review articles have high status due to the honorific and authoritative position of those invited to write them? or due to the thoroughness of their analysis and the sheer number of statements they make about the articles they review? or don't they have high status at all? In other words, is opinion on the Trondheim and Ålesund articles almost evenly divided, as indicated by the the citing articles' conclusions? Or has the Trondheim article been slightly more successful, as indicated by the number of citing articles? Or have the claims of the Trondheim article, though still disputed, been considerably more successful than those of the Ålesund article, as indicated by the modality analysis?

Another impression which runs contrary to one of comparative success of the Trondheim article relative to the Ålesund article is the frequency of positive finds attributed to the Trondheim article which its authors do not claim. Eight articles (one each in 1985, 86, 87 and 88 and two in 1989 and 90) make one or more such re-authored references. There are fifteen such statements in all -- ten in positive modalities, three neutral and two negative. One of the eight articles containing this type of reference concludes against ultrasound screening, three have other agendas, two are indecisive, and two favor screening.

This phenomenon of referring to re-authored versions of the source text was practically absent in the other cases.[62] Could this be a sign of the well-known confirmational bias in science journals? Could it be that not only are articles claiming positive results more likely to be submitted, accepted, and cited,[63] but that readers' reauthorship of articles showing negative results further contributes to this bias? And if so, is this a matter of the science community rewarding (in this case, even non-significant) confirmational results, in conflict with the norm of skepticism, but otherwise in keeping with the normative behavioural structures of the community? Or are these reinterpretations somehow a product of earlier readings of the original text?

Finally, it would seem that what success either of the ultrasound articles has enjoyed may be dabbing off in recent years. This impression comes from the accumulation of references to single statements and partial conclusions in 1989 and 1990, followed by a drop in references altogether in 1991-92. Could it be that both studies are seen as no longer crucial to the debate on screening with ultrasound in pregnancy?

In the mammography case as well, articles taking a position in favor of the claims of the source texts may nonetheless contain negative modalities towards those claims for other reasons. The opposite is also true: opposing texts sometimes contain positive modalities. For the main claims of the studies, there are only a handful of negative modalities, while statements with positive modalities are so numerous, especially towards the HIP and WE trials, as to stretch the capacity of the graphs to contain them. Modalities toward partial claims show a more negative trend in the HIP case and a less emphatically positive trend in the WE case, but there are few

Figure 4. Modalities of references to HIP-study, 1986 and Jan - Jun 1992

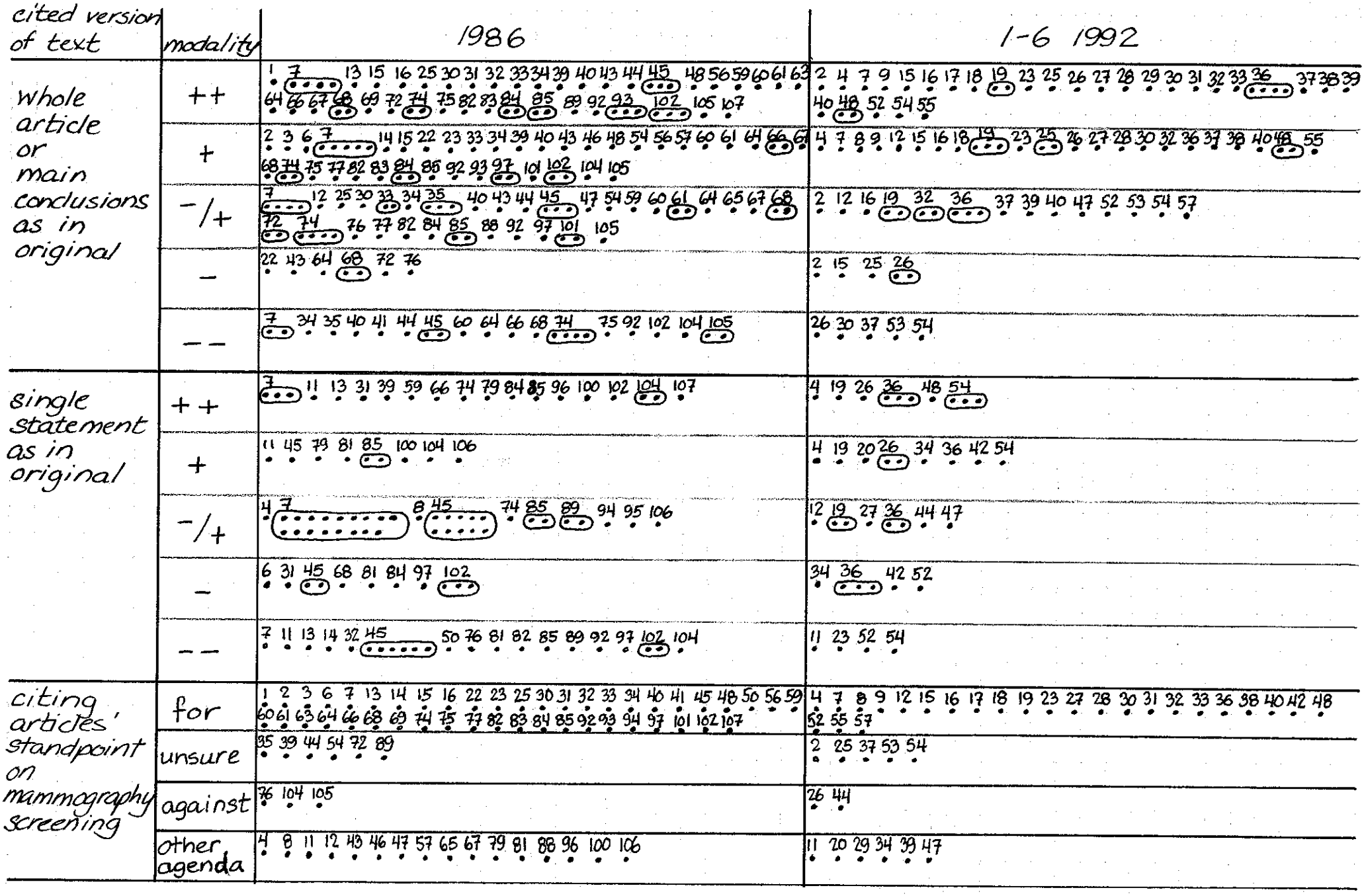


Figure 5. Modalities of references to WE-study, 1986 and Jan-Jun 1992

cited version of text	modality	1986	1-6 1992
whole article or main conclusions as in original	++	6 10 24 34 35 39 40 43 48 49 50 51 52 59 61 64 66 74 75 78 85 87 91 93 97 98 102 103 105	4 6 7 10 15 16 18 19 23 25 26 27 28 30 37 38 40 41 46 48 50 51 52 54 55 57
	+	2 3 6 10 14 15 16 19 22 23 24 26 33 34 35 39 40 43 46 48 49 54 61 64 66 74 75 77 82 85 87 92 93 97 101 102 104 105	4 5 7 8 9 10 12 15 16 18 19 21 23 24 25 26 27 28 30 32 36 37 38 40 41 45 48 51 53 57
	-/+	19 20 30 34 35 40 48 51 52 57 61 64 74 77 85 98 101 102 105	2 3 12 16 18 19 32 33 36 44 51 52 53 54 57
	-	38 43 64	15 25 26 35 44
	--	34 40 64 74 75 102 104 105	1 26 30 35 36 37 44 53 56
single statement as in original	++	2 3 17 39 53 85 87 91 94	4 14 26 32 49 54 56
	+	10 35 50 85 102	4 12 13 19 26 36 43 54
	-/+	9 15 35 85 94 95 97	12 16 18 19 22 27 43 49
	-	6 94	36 46
	--	9 14 50 82 97 102 104	1 3 19 21 23 26 37 44 54
citing articles' standpoint on ultrasound screening	for	2 3 6 10 14 15 16 22 23 24 26 30 33 34 40 48 49 50 51 52 60 61 64 66 74 75 77 78 82 85 87 91 92 93 94 95 97 98 101 102 103	3 4 5 6 7 8 9 10 12 15 16 18 19 22 23 27 28 30 32 33 36 38 40 45 46 48 50 51 52 55 57
	unsure	20 35 39 54	2 14 21 25 35 37 53 54
	against	38 104 105	1 26 41 44 56
	other agenda	9 17 19 43 46 53 57	13 24 43 49

Figure 6. Modalities of references to Malmö study, 1986 and Jan-Jun 1992

cited version of text	modality	1986	1-6 1992
whole article or main conclusions as in original	++		4 10 14 19 25 56
	+		4 10 14 19 36 44
	-/+	35 74	2 19 36 37 57
	-		14 25 35
	--		35
single statement as in original	++	35 102	4 14 19 30 37 56
	+	5 35	4 19 25 30 36
	-/+	6 94	19
	-	35	19 25 36
	- -		19 37
citing articles' standpoint on mammography screening	for	6 74 94 102	4 10 19 30 36 57
	unsure	35	2 14 25 35 37
	against		26 44 56
	other agenda	5	

such statements compared with statements referring to the main conclusions. It would seem that these studies have been the object of mild criticisms directed at minor points of method and analysis. There are no instances, as I read the texts, of the claims of the source texts being radically rewritten.

I have no evidence for concluding which of the indicators of the acceptance of the source texts -- modalities or citing authors' conclusions -- has the greater influence on secondary readers' acceptance of those texts. In this case, however, it does not seem important to find such evidence. For while the two indicators may not match well within an individual citing text, the overall impressions from the modality analysis and the rewards analysis are the same. In the ultrasound case, both indicated an ongoing controversy, but of different degrees. The modality analysis showed a trend towards closure in favour of the Trondheim trial conclusions, but compared with the signs of consensus in the mammography case, that trend was not a strong one. In the mammography case, both analyses indicate an overwhelming consensus.

In the ultrasound case, I found evidence of confirmational bias in the citing authors' readings of the source texts. This bias also shows up in the mammography case, but in a different and perhaps more conventional way. In the ultrasound case, one sign of the confirmational bias was the number of radical reinterpretations of the one study which claimed not to have found statistically significant benefits to the screened group. In the mammography case, no source articles making such a negative claim were found, although one source article bases its positive claims on non-significant results. This is in line with the widely acknowledged confirmational publication bias. A confirmational citation bias may be evidenced by the small number of citations received by this article and the large proportion of those citations which support the article's main claims and ignore the lack of statistical significance.

5. What impact science?

5.1. What impact science?

The main question addressed in this article has been what impact science had on the two screening decisions. The main answer is: Not much. Regardless of which model of science and consensus we apply, that answer remains the same.

In the ultasound case, we found that medical science as a whole has not reached consensus, although the Mertonian and discourse models shows a slight trend against screening. Nevertheless, ultrasound screening in pregnancy has been standard practice since at least 1986 and official policy since 1987. The interest model provides a possible explanation for both the lack of consensus and for implementation in spite of that lack. The explanation hypothesizes a dichotomy between epidemiological and clinical interests, one which echoes Constant's dichotomy between science and technology issues. According to Constant, science explores a vicarious environment in pursuit of precision and sophistication of theory and method. Similarly, the epidemiological approach addresses relatively abstract issues of population level benefits and costs, and insists on certain methodological standards to reach reliable answers. According to Constant, technology issues are practical and confront reality directly, accepting simplification and compensating through over-dimensioning in favour of immediate solutions. Similarly, the clinical approach is directed at immediate patient:provider interactions. Rather than sophisticated measures of population-level costs and benefits, clinicians are concerned with individual outcomes. The rare life saved may be statistically irrelevant, but as an individual outcome represents a 100% improvement. The acceptance of new technologies based on anecdotal

evidence of rare benefits can be seen as a form of over-dimensioning -- here directed at medical treatment rather than technological development as a project. Interestingly, in medicine the two communities seem to share the same journal fora as well as the same professional training.

Evidence for an epidemiological/clinical dichotomy was both the arguments offered in support or opposition of screening, the citation of epidemiological vs. clinical experts, and an apparent confirmational interpretation and citation bias. The confirmational citation bias was evidenced in the mammography case by a tendency to ignore the one study which did not report statistically significant benefit to the screened group and to ignore that lack of statistical significance if citing the report at all. In the ultrasound case, the bias is evidenced by a number of articles which reverse the claims of the negative study when citing it. These reversals occurred most frequently in articles concluding in favour of screening.

Since implementation of screening is a clinical issue, and assuming that the clinical orientation is even more dominant among practicing clinicians than among publishing medical authors, this might explain why ultrasound screening has spread so quickly in spite of ongoing controversy in the medical journal literature. The mammography case, however, becomes even more paradoxical in light of those assumptions.

In the mammography case, a consensus for screening was confirmed according to all three models of science. It was apparent in citation patterns, in citing articles' conclusions, and in modality patterns. It could also be accounted for in terms of the epidemiological/clinical dichotomy since the supporting evidence was both individual cases of cured cancers and statistically significant confirmational results from published randomized controlled trials. Consensus for screening was clear both in 1986 and 1992. Yet mammography screening is neither established practice nor official policy, although policy has recently taken a cautious turn towards implementation.

So science has not had a decisive voice in these two medical technology decisions. Can we learn anything more about which model of science is most appropriate? And can our model of science tell us anything more about why science has had so little impact in these cases?

5.2. Models vs. models

Although the analysis was not designed to test the models, there are two ways in which it is appropriate to compare them after applying them to the same data sets and questions. One is to reflect on which model provided the most fruitful methods; the other to ask which theory is best able to account for the conclusions.

First, some methodological reservations against raw citation-count analysis, regardless of interpretive framework, seem born out by the texts and the modality analysis of them. Citing texts do often contain more than one reference to the source text, references which often differ in modalities and serve various rhetorical purposes. Modality analysis (whether interpreted within a Mertonian or discourse framework) accounts for the first two points, but the last is not accounted for in modality analysis either.

Ignoring the various rhetorical purposes of citations can be defended on the basis offered for modality analysis within the discourse model -- that authors' intentions are not at issue, that it is the effects of their use of the cited texts we are studying. To my knowledge, the effects of citations have not yet been studied. In Latour's presentation of the method, it is axiomatic that usage of a text modifies its content and credibility; but how can we know this to be true? In the present study I find some possible evidence:

One explanation for the re-writing of the Trondheim study's conclusions in several citing texts might be that earlier citing texts have established the Trondheim authors as the more credible source, but not the Trondheim study claims. The article by Thacker[64] evaluated the trial protocols and publications from the four rct's on ultrasound screening in pregnancy. Each trial was ranked on a scale from 0-100% for quality of methods and of reporting. The Ålesund trial received only 27%, mainly for insufficient publication. The Trondheim trial received 78% - top marks among the four. In the same vein, two articles by Chalmers et al.[65] chastized the Ålesund trial group as guilty of "scientific misconduct" for not having fully published their results. Neither of these articles, however, took a stand on the truth or untruth of the claims of the respective trials. We may hypothesize that this helped establish the Trondheim study as the source to cite, while leaving unscathed the belief (coloured perhaps by clinical and confirmational bias) that ultrasound screening had been proven beneficial. We may hypothesize three scenarios for how the citations of rewritten claims came about. One would be that citing authors, having read the Trondheim study through a filter of faith in ultrasound screening, saw only those of its non-significant results which confirmed that faith. This scenario fits neatly with the interests model. The second is that authors favouring screening were instructed by referees (who in turn acted on the factual status conferred by the Thacker article) to confer with the Trondheim study, and then simply added a reference to that study in their revisions, without actually reading the source text. This scenario fits with the discourse model but also incorporates the rewards system postulated by the Mertonian model. The third scenario is a similar combination with the addition of the interests models. In this scenario, authors instructed to add a reference read the source more or less perfunctorily, and reinterpreted it in accordance with their interests.

We may hypothesize these scenarios, but we may never find proof. The latter two scenarios correspond to one of the response alternatives in a recent study by Leydesdorff and Amsterdamska[66] of authors' motivations to cite. Although they still suspect that this scenario probably is acted out at least occasionaly, not a single informant gave that response.[67]

Thus, the reasoning behind modality analysis received considerable support in this study, but we cannot judge whether it gave a more precise picture of the consensus we were out to measure. On the whole, the impressions given by the analyses of modalities, citation counts and citing articles' positions were reassuringly similar.

A feature shared by the modality and interests analyses is that they force us to read the texts closely. This was fruitful in that it generated reasonable hypotheses as to values which might influence the target population's interpretations and acceptance of the competing truth claims. These hypotheses remained speculative, however, if we relied solely on the texts themselves and their authors' identities to test them. This seems inevitable for two reasons: One, it is a common trait in the genre of scientific articles to exclude references to personalities, values, interests, etc. Only those interests which have been constructed as relevant for and legitimate within scientific debate are likely to be apparent in the texts.[68] This might explain why feminist interests, though explicit in my interview material, could not be confirmed in the citation material, while competition between scientific specializations is found in both. Two, we may assume that the cognitive processes behind the production of texts are so complex and diverse as to preclude simple deductions of interests from the end results of those processes.[69] However, by combining modality analysis with citation and co-citation analysis it was possible to evaute these new hypotheses from several angles and find substantial support for one.

The three models I have applied provide me not only with methods of data selection and analysis, but also with interpretive frameworks for the results as a whole. We have seen that each of the three methods brought us to similar conclusions -- that there was no consensus in the ultrasound controversy and a consensus for screening in the mammography case, and that the

practical decisions to screen or not to screen were not in keeping with those consensual statuses. Now let us look to how the three theories explain consensus or lack of it and lack of impact.

The Mertonian model is normative. After a period of critical evaluation, a scientific community ought to reach consensus. If, as in the ultrasound case, controversy continues indefinitely, we must either be dealing with the effects of counter-norms,[70] or some of the scientists are breaking the norms. In either case, it is up to the community itself to work out which norms apply and what claims to accept. As social scientists outside that community, we can go no further via the Mertonian model than to conclude whether controversy exists. Nor does the Mertonian model offer any explanations for the lack of impact on practice that we have found in these cases.

The interest model has offered us a plausible explanation for both the ongoing controversy in the ultrasound case and consensus in the mammography case. The explanation is the possible existence of a dichotomy between clinical science and epidemiological science. The two medical science communities would have different values and different interpretations of evidence due to their commitments to different practice constituencies outside science: a constituency of medical practitioners vs. one of health politicians and administrators respectively. According to the interest model, powerful constituencies outside science shape science rather than vice versa. Thus, the interest model can also explain the implementation of ultrasound screening in the face of scientific controversy. The explanation would be that the clinical constituency was more powerful than the political/administrative constituency in shaping that area of practice. But the interest model leaves us puzzled by the lack of implementation of mammography screening in the face of scientific consensus favouring it. Is there some even more powerful group opposing mammography screening? And if so, why hasn't that group excercized any noticeable influence on the science(s) which have addressed the question?

The discourse model, Actor Network Theory, has the advantage of incorporating both of the other models: Actor Network Theory does not deny the existence of norms within the scientific community. It interprets adherence to those norms in a framework of pursuing ones career interests as a scientist, and interprets scientists' dependence on alliances outside the scientific community in the same framework. Actor Network Theory has the further advantage of offering possible explanations for science's lack of impact on entrepreneurship. Although the model is developed to explain how science succeeds in being influential,[71] it has also been applied to cases of failed innovations.[72] Explanations offered by the model might be scientists' inattention to the need for alliances, scientists' misconceptions as to the interests of potential allies, the solidity of previously established networks which the innovation would have had to replace, or the irreconciliability of the interests of allies needed for the innovation. The latter explanation reminds us that scientists and engineers are not the only active, interested parties in innovation situations. Other parties are pursuing programs of their own and will seek science and/or engineering allies only if it suits their own purposes. In our present context, we could sum this up by stating: Science is neither a necessary nor an adequate basis for practical innovation.

It remains then for us to map out which parties were active in promoting and/or opposing the two screening technologies. In incorporating the results found via all three models and in pointing out a direction for future research, Actor Network Theory has not necessarily proven itself truer than the other two models applied, but has proven itself the most fruitful.

NOTES

1. For a discussion of drug regulation as another example medical technology decisions in general, see H.J.H.W. Bodewitz, H. Buurma, and G.H. de Vries, 'Regulatory Science and the Social Management of Trust in Medicine' in W.E. Bijker, T.P. Hughes, and T. Pinch, The Social Construction of Technological Systems, Cambridge/London (1987): MIT Press, 242-259. For a discussion of the special considerations involved in screening technology decisions, see J.P. Koplan and F.F. Gutzwiller, 'Some Observations on the Assessment of Preventive Technologies', International Journal of Technology Assessment in Health Care, Vol 7 (Summer 1991), 361-364.

2. E. W. Constant II, 'Communities and Hierarchies: Structure in the Practice of Science and Technology' in R. Laudan (ed.), The Nature of Technological Knowledge. Are Models of Scientific Change Relevant? (Dordrecht/Boston/Lancaster: D. Reidel Publishing Co. (1984), 27-46.

3. A dozen of the many examples of the ways science is looked to to settle technology controversies are discussed in D. Nelkin (ed.), Controversy: Politics of Technical Decisions, Beverly Hills/New Delhi/London (1979): Sage publications.

4. For an overview of drug regulation policies see Bodewitz, Buurma, and de Vries (op. cit.). See also L. Lasagna and L. Werkö (eds.), 'Special Section: The Evaluation of Drugs: An International Perspective', International Journal of Technology Assessment in Health Care, Vol 2 (1986), 615-708. One of many statements arguing for similar policies for evaluation and control of other medical technologies is the following from Charles U. Lowe of the Office for Medical Applications of Research, National Institutes of Health: 'Are there some medical technologies in general use that are unsafe or ineffective? Are there still other drugs, devices, and medical or surgical procedures that have not been widely accepted even though well validated? These are the kinds of questions, raised in recent years, that have led to a reassessment of the process by which technologies are transferred from research and development into practice.' (C.U. Lowe, 'The Consensus Development Programme: Technology Assessment at the National Institute of Health', British Medical Journal, (June 28, 1980), 1583.)

5. R.K. Merton, Social Theory and Social Structure, (New York: Free Press (1968), originally published in 1957). Also H. Zuckerman, 'The Sociology of Science' in N.J. Smelser (ed.), Handbook of Sociology, Newbury Park/Beverly Hills/London/New Delhi: Sage Publications (1988), 511-574.

6. L. Leydesdorff and O. Amsterdamska, 'Dimensions in Citation Analysis', Science, Technology, & Human Values, Vol 15 (Summer 1990), 305-335. Also Zuckerman, op. cit.

7. For instance in Science, Technology, & Human Values, Vol 15, No 1, editor Susan E. Cozzens writes:
'Publications can serve many functions. If [article in question] had not been written by an outsider to the science studies community; if it had not come to me with a history of conflict and controversy; and if it had not raised important ethical questions that had not yet been widely discussed, at least not as our own problem, in the research community served by Science, Technology, & Human Values, then it would not have been published here. It has been. (...) To those familiar with U.S. standards for informed consent, the problem here will be obvious: the participants in this research should have been asked whether they wanted to participate. (...) The members of this community who reviewed Epstein's manuscript for Science, Technology, & Human Values were split on the question of whether Epstein's research was ethical. (...) When members of our own research community disagree to this extent about an ethical issue, it needs airing and wider discussion. For this reason, I am publishing Epstein's article in ST&HV.'

8. A number of approaches have been lumped together, sometimes against their authors' wills, under the rubric "constructivism." The approach called Social Construction of Technology (a title introduced in T.J. Pinch and W.E. Bijker, 'The Social Construction of Facts and Artifacts:

Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other', in w.e. Bijker, T.P. Hughes, and T. Pinch, op. cit., 17-50) is the only one with "construction" in its name. Most of these approaches deal primarily with technology studies, although Actor Network Theory (see for instance B. Latour, Science in Action: How to Follow Scientists and Engineers through Society, Milton Keynes (1987): Open University Press.) is also a theory of science. For a presentation of what the approaches frequently grouped as constructivist share, see W.E. Bijker and J. Law, 'General Introduction', in W.E. Bijker and J. Law (eds.), Shaping Technology/Building Society, Cambridge/London: MIT Press, 1992, 1-14.

Social Construction of Technology borrows theoretical elements from the Empirical Program of Relativism. (For a presentation of this program, see H.M. Collins, An Empirical Relativist Programme in the Sociology of Scientific Knowledge, in K.D. Knorr-Cetina and M. Mulkay (eds.), Science Observed. Perspectives on the Social Study of Science, London/New Delhi/Beverly Hills (1983): Sage Publications, 85-113; and also Social Studies of Science, Vol 11 (1981), 3-158, which is a special issue on the relativist program, guest edited by Collins.) There are many senses in which one might be relativist (ethical, epistemological, ontological, methodological), and neither constructivist nor self-declared relativist approaches are all of these. For instance, the interest model presented here could be called 'social realist' since 'facts' in the natural sciences are seen as based on flexible interpretations of data and interpretations based in turn on scientists' objectively real social interests. Actor Network Theory bridges so-called relativism and realism by including non-social actors in the networks it maps. And few if any of the authors using constructivist or relativist approaches see their work as precluding taking moral stands. In fact, some (W.E. Bijker, 'Do Not Despair: There Is Life after Constructivism', Science, Technology, & Human Values, Vol 18 (Winter 1993), 113-138; A. Clarke and T. Montini, 'The Many Faces of RU486: Tales of Situated Knowledges and Technological Contestations', Science, Technology, & Human Values, Vol 18 (Winter 1993), 42-78.) see constructivist research as a tool for ethical evaluation of technologies and for political activism.

9. H.M. Collins, 'Stages in the Empirical Programme of Relativism', Social Studies of Science, Vol 11 (1981), 3-10.

10. D. MacKenzie and B. Barnes, 'Scientific judgment: The biometry-mendelism controversy', in B Barnes & S Shapin (eds.), Natural order. Historical studies of scientific culture, Sage Publications: Beverly Hills & London (1979), 191-210.

11. Latour, op. cit.

12. R.J. Lemire, 'Neural tube defects', Journal of the American Medical Association. Vol 259 (1988), 559.

13. Leydesdorff and Amsterdamska, op. cit.; Zuckerman, op. cit.

14. Zuckerman, op.cit.; M. Mulkay, J. Potter, and S. Yearley, 'Why an Analysis of Scientific Discourse is Needed', in Knorr-Cetina and Mulkay (eds.), op. cit., 171-203; G.N. Gilbert and M. Mulkay, Opening Pandora's Box: A Sociological Analysis of Scientists' Discourse, Cambridge (1984): Cambridge University Press.

15. Consensus conferences are a forum developed by the National Institutes of Health, Office of Medical Applications of Research to offer science-based advice on health policy and medical practice. The consensus conference model borrows elements from jury trials, scientific meetings, and town meetings or public hearings. As practiced in Norway, a preparatory committee has selected a panel of experts and lay members, prepared a set of questions for the panel to address, provided the panel with relevant scientific literature, and invited a set of expert witnesses. The conference lasts three days. During the first two days, the witnesses present papers and are questioned by the panel. The public is invited to ask questions and offer evidence at the end of the second day. Then the panel works through the night formulating their answers (the consensus statement) to the questions set by the preparatory committee. The statement is presented at an open press conference on the third day. For a description of the history of the NIH model, see S.

Perry and J.T. Kalberer, 'The NIH Consensus-Development Program and the Assessment of Health-Care Technologies', The New England Journal of Medicine, Vol 303 (July 17, 1980), 169-172. See also I. Jacoby, 'The Consensus Development Program of the National Institutes of Health. Current Practices and Historical Perspectives', International Journal of Technology Assessment in Health Care, Vol 1 (1985), 420-432. For a description variations on the consensus conference model as practiced in various countries, see E.A. McGlynn, J. Kosecoff, and R.H. Brook, 'Format and Conduct of Consensus Development Conferences. Multination Comparison', International Journal of Technology Assessment in Health Care, Vol 6 (1990), 450-469.

16. The conference was held 27-29 August, 1986. (For an English translation of the statement, see Department of Social Affairs and Norwegian Institute of Hospital Research, 'Ultrasound in Pregnancy: Consensus Statement, 1986', International Journal of Technology Assessment in Health Care, Vol 3 (1987), 463-470.) Accompanied by a letter dated 9 September, 1986, the Directorate of Health sent copies of the consensus statement to all county health authorities "for (their) information." The letter also announces that "The Directorate of Health will begin its work of following up the consensus statement in the immediate future."

17. B. Backe, 'Ultralydundersøkelser i Norge. Dagens Utbredelse', in B. Backe and H. Buhaug (eds.), Bruk av Ultralyd i Svangerskapet. Konsensuskonferansen 17-19/8-1986, Trondheim: NIS-rapport 8 (1986), 35-37.

18. M.J. Bennett, G. Little, J. Dewhurst, and G. Chamberlain, 'Predictive Value of Ultrasound Measurements in Early Pregnancy: A Randomised Controlled Trial', British Medical Journal, Vol 89 (1982), 338-341.

19. J.P. Neilson, S.P. Munjanja, and C.R. Whitfield, 'Screening for Small for Dates Fetuses: A Controlled Trial', British Medical Journal, Vol 289 (November 3, 1984), 1179-1182.

20. L.S. Bakketeig, G. Jacobsen, C.J. Brodtkorb, B.C. Eriksen, S.H. Eik-Nes, M.K. Ulstein, P. Balstad & N.P. Jörgensen, 'Randomised controlled trial of ultrasonographic screening in pregnancy', Lancet, No 8396 (1984), 207-211.

21. S.H. Eik-Nes, O. Økland, J.C. Aure & M. Ulstein, 'Ultrasound screening in pregnancy: A randomized controlled trial', Letter, Lancet, No 8390 (1984), 1347.

22. S.B. Thacker, 'Quality of controlled clinical trials. The case of imaging ultrasound in obstetrics: a review', British Journal of Obstetrics and Gynaecology, Vol 92 (1985), 437-444.

23. Bakketeig et al., op. cit. and Eik-Nes et al., op cit.

24. For instance: G. Breart & V. Ringa, 'Routine or selective ultrasound scanning', Bailièrres Clinical Obstetrics and Gynaecology, Vol 4 (1990), 45-63; J.W. Goldkrand, D.S. Benjamin & D.M. Cantor, 'Role of ultrasound in obstetric management', Journal of Clinical Ultrasound. Vol 14 (1986), 589-594; T.C.M. Li, R.A. Greenes, M. Weisberg, D. Millan, M. Flatley & L. Goldman, 'Data assessing the usefulness of screening obstetrical ultrasonography for detecting fetal and placental abnormalities in uncomplicated pregnancy', Medical Decision Making, Vol 8 (1988), 48-54; V. Ringa, B. Blondel & G. Breart, 'Ultrasound in obstetrics: Do the published evaluative studies justify its routine use?', International Journal of Epidemiology, Vol 18 (1989), 489-497; and, Thacker, op. cit.

25. L. Tabár, A. Gad, L.H. Holmberg, U. Ljungquist, C.J.G. Fagerberg, L. Baldetorp, O. Gröntoft, B. Lundström, J.C. Månson, G. Eklund, N.E. Day & F. Pettersson, 'Reduction in mortality from breast cancer after mass screening with mammography. Randomised trial from the breast cancer screening working group of the Swedish National Board of Health and Welfare', Lancet (Apr 13, 1985), 829-832.

26. NOU, 1987:7, Mammografiscreening i Norge.

27. B. Backe, 'Hvorfor en konsensuskonferanse om mammografiscreening?', in B. Backe (red.), Konsensuskonferansen om Mammografiscreening. 8.-10. februar 1989, Trondheim: NIS rapport 2 (1989), 19-21.

28. The female population of Norway ages 50-74 was 517 226 as of December 31, 1987. My rough estimate of 600 000 examinations arising from a screening program targeting that group is based on nearly complete attendance plus one or more repeat examinations for suspicious finds in 10% of the screenees, and an additional small proportion of mammographies in younger or older women on clinical indications or by their own request.

29. I. Andersson, K. Aspegren, L. Janzon, T. Landberg, K. Lindholm, F. Linell, O. Ljungberg, J. Ranstam, and B. Sigfússon, 'Mammographic screening and mortality from breast cancer: the Malmö mammographic screening trial', British Medical Journal, Vol 297 (1988), 943-948.

30. A.B. Miller, C.J. Baines, T. To, and C. Wall, 'Canadian National Breast Screening Study. 1. Breast-Cancer Detection and Death Rates Among Women Aged 40 to 49 Years' Canadian Medical Association Journal, Vol 147 (1992), 1459-1476; and A.B. Miller, C.J. Baines, T. To, and C. Wall, 'Canadian National Breast Screening Study. 2. Breast-Cancer Detection and Death Rates Among Women Aged 50 to 59 Years', Canadian Medical Association Journal, Vol 147 (1992), 1477-1488.

31. J. Frisell, G. Eklund, L. Hellstrom, E. Lindbrink, L.E. Rutqvist, and A. Somell, 'Randomized Study of Mammography Screening -- Preliminary Report on Mortality in the Stockholm Trial', Breast Cancer Research and Treatment, Vol 18 (1991), 49-56.

32. Results were presented at Läkarstämman (an annual medical conference) in Stockholm, November 26, 1992.

33. In advance of her lecture at an annual cancer conference, September 25, 1991, the Minister of Social Services was requested by the Norwegian Cancer Association to address certain questions, among them mammography screening. In the lecture, she announced that she was willing to consider implementing a mammography screening program, although there were still some unanswered questions concerning logistics and quality assurance. In a letter dated December 16, 1991, the Norwegian Cancer Association offered joint financing for the trial project. Government joint financing for the project was included in the national budget proposal for 1993. In an interview with me in December 1992, the coordinator of the project (which was already under planning in anticipation of final budget approval) informed me that the results of the meta-analysis were already known in 1991 and had been leaked to the relevant ministers in the other Nordic countries to allow them to prepare any changes in policy.

34. S. Shapiro, 'Evidence on screening for breast cancer from a randomized trial', Cancer, Vol. 39 (1977), 2772-2782; P. Strax, 'Strategy (motivation) for detection of early breast cancer', Cancer, Vol 27 (1980), 1563-1568; S. Shapiro, W. Venet, P. Strax, L. Venet & R. Roesner, 'Ten- to fourteen-year effect of screening on breast cancer mortality', Journal of the National Cancer Institute, Vol 69 (1982), 349-355; S. Shapiro, W. Venet, P. Strax, L. Venet, and R. Roeser, 'Selection, follow-up and analysis in the Health Insurance Plan (HIP) study', in Selection, follow-up and analysis in prospective studies: A workshop, National Cancer Institute Monography 67 (1985), 65-74; S. Shapiro, W. Venet, P. Strax and L. Venet, 'Periodic screening for breast cancer: the Health Insurance Plan project and its sequelae, 1963-1986', Baltimore (1988): Johns Hopkins University Press

35. L. Tabar, C.J.G. Fagerberg, A. Gad, L. Baldetorp, L.H. Holmberg, O. Gröntoft, U. Ljungquist, B. Lundström, J.C. Månson, G. Eklund, N.E. Day, and F. Pettersson, 'Reduction in mortality from breast cancer after screening with mammography', Lancet (1985), 829-832; L. Tabar, G. Fagerberg, N.E. Day, and L. Holmberg, 'What is the optimal interval between mammographic screening examinations: An analysis based on the latest results of the Swedish two-county breast cancer screening trial', British Journal of Cancer, Vol 55 (1987), 547-551; L.

Tabar, S.W. Duffy, and U.B. Krusemo, 'Detection method, tumor size and node metastases in breast cancers diagnosed during a trial of breast cancer screening', European Journal of Cancer and Clinical Oncology, Vol 23 (1987), 959-962; L. Tabar, C.J.G. Fagergerg, S.W. Duffy, and N.E. Day, 'The Swedish two county trial of mammographic screening for breast cancer: recent results and calculation of benefit', Journal of Epidemiology and Community Health, Vol 43 (1989), 107-114; L. Tabar & P.B. Dean, 'The present state of screening for breast cancer', Seminars in Surgical Oncology, Vol 5 (1989), 94-101; L. Tabar, 'Control of breast cancer through screening mammography', Radiology, Vol 174 (1990), 655-656; L. Tabar, C.J.G. Fagerberg, S.W. Duffy, N.E. Day, A. Gad and O. Gröntoft, 'Update of the Swedish two-county program of mammographic screening for breast cancer', Radiological Clinics of North America, Vol 30 (1992) 187-210.

36. Andersson et al., op. cit.

37. Latour, op. cit.

38. I read Latour's description of modalities as necessarily refering to individual statements. In practice, however, both Latour and others applying his method have worked with whole citing articles as if these constituted a single statement about the source text.

39. For a discussion of which purposes, see Leydesdorff and Amsterdamska, op. cit.

40. Readers who wish to test my coding judgment against their own are welcome to contact me for a copy of the complete reference list.

41. M. Moskowitz, 'Do the results of the Swedish trial, the Dutch case control study, and the Cincinnati breast cancer detection demonstration project tell us anything of importance about the natural history of breast cancer?' in Colin & Gordenne (eds.), Evaluation du risque de cancer mammaire, Brussels (1985), Pierre Mardaga, 119-120.

42. C.J. Wright, 'Breast-Cancer Screening - A Different Look at the Evidence', Surgery, Vol 100 (1986), 594-598.

43. A.B. Nattinger and J.S. Goodwin, 'Screening Mammography for Older Women. A case of mixed messages', Archives of Internal Medicine, Vol 152 (1992), 922-925.

44. G. Ramsey-Stewart, 'Mammography is fallible', The Medical Journal of Australia, Vol 156 (1992), 67.

45. For instance, one critic of ultrasound screening said, in an interview with me: 'It's a feminist issue too. That things get thrust on us. And that we make these men out to be our saviours, as if they had saved us, when it's all just so much humbug! My God! It's infuriating, isn't it? Here are these women -- in deep gratitude, thinking they are so much more secure now they have ultrasound screening -- and it's all just a deception! That infuriates me. And mammography screening is another case of the same. If it turns out to have significant influence on peoples' lives and health, then obviously I'll just lie right down and accept it, and then I think the feminist aspects of it and the dependency on the health system fade into the background. If it's valid and not just humbug. Then I'll yield completely. I'm not a fanatic feminist who can't tolerate the male dominance in the health system. But as long as it isn't proven, then I'm skeptical.' More on the subject of gender will be appearing in A. Sætnan, 'Ultrasonic Tales: Gendered Controversies in the Construction of Ultrasound', (forthcoming).

46. A.L. Cochrane, Effectiveness and Efficiency. Random Reflections on Health Services, The Nuffield Provincial Hospitals Trust, (1972).

47. J.B. McKinlay, 'From "Promising Report" to "Standard Procedure": Seven Stages in the Career of a Medical Innovation', Milbank memorial Fund Quarterly/Health and Society, Vol 59 (1981), 374-411.

48. In the introduction to his book Effectiveness and efficiency, Cochrane (op. cit.) writes: 'I had been convinced for some time about the final form in which any analysis of the over-all result of the various activities in the NHS should be expressed. If we are ever going to get the 'optimum' results from our national expenditure on the NHS we must finally be able to express the results in the form of the benefit and the cost to the population of a particular type of activity, and the increased benefit that could be obtained if more money were made available.' He concludes his chapter on "Evaluation of evidence" with: 'The main job of medical administrators is to make choices between alternatives. To enable them to make the correct choices they must have accurate comparable data about the benefit and cost of the alternatives. These can really only be obtained by an adequately costed RCT.'

McKinlay (op. cit.) emphasizes the point with italics and capital letters: 'Government Should Not Support through Public Funding for General Public Use Any Service, or Technology, the Effectiveness of Which Has Not Been, or Cannot Be, Demonstrated.' By 'demonstrated,' he too means 'preferably and where appropriate by RCTs.'

49. International Journal of Technology Assessment in Health Care, Vol 3 (1987), 3-99.

50. U. Waldenström, S. Lindeberg, S. Nilsson, Y. Sjodin, O. Axelsson, G. Eklund, and O. Fall, 'Effects of Routine One-Stage Ultrasound Screening in Pregnancy: A Randomised Controlled Trial', Lancet, (1988), 585-588.

51. A. Saari-Kemppainen, O. Karjalainen, P. Ylöstalo and O.P. Heinonen, 'Ultrasound Screening and Perinatal Mortality: Controlled Trial of Systematic One-Stage Screening in Pregnancy', Lancet, Vol 336 (1990), 387-391.

52. R.B. Deber, 'Translating Technology Assessment into Policy. Conceptual Issues and Tough Choices', International Journal of Technology Assessment in Health Care, Vol 8 (1992), 131-137.

53. I. Chalmers, 'Underreporting Research Is Scientific Misconduct', Journal of the American Medical Association, Vol 263 (March 9, 1990), 1406.

54. B.G. Ewigman, 'An Opposing View', The Journal of Family Practice, Vol 29 (1989), 663.

55. J.W. Goldkrand, D.S. Benjamin and D.M. Cantor, 'Role of Ultrasound in Obstetric Management', Journal of Clinical Ultrasound, Vol 14 (1986), 589.

56. L.Ch. DeCrespigny, P Warren and B. Buttery, 'Should All Pregnant Women Be Offered an Ultrasound Examination?', The Medical Journal of Australia, Vol 151 (December 4/18, 1989), 614.

57. P. Alfonsi and V. Cutrupi, 'Natural History of Breast Cancer', Lancet, Vol 339 (1992), 810.

58. Wright, op. cit.

59. Moskowitz, op. cit.

60. Examples of review articles with particularly high numbers of statements and a tendency to favour the Trondheim study over the Ålesund study are: G. Breart and V. Ringa, 'Routine of Selective Ultrasound Scanning', Baillièrres Clinical Obstetrics and Gynaecology, Vol 4 (1990), 45-63; C. Köck, 'Ultraschallscreening in der Schwangerschaft -- eine kritische Literaturanalyse', Wiener klinische Wochenschrift, Vol 101 (1989), 341-345; V. Ringa, B. Blondel, and G. Breart, 'Ultrasound in Obstetrics: Do the Published Evaluative Studies Justify Its Routine Use?', International Journal of Epidemiology, Vol 18 (1989), 489-497; and S.B. Thacker, 'Quality of

Controlled Clinical Trials. The Case of Imaging Ultrasound in Obstetrics: A Review', British Journal of Obstetrics and Gynaecology, Vol 92 (1985), 437-444.

61. I will not go into the possibility that texts lead completely autonomous lives of their own, influencing on the status of truth claims independent of all cognition.

62. In the only other instance of reauthorship, a citing author attributed the delay in publication from the Ålesund trial as due to that trial undergoing some sort of extension.

63. U. Ravnskov, 'Frequency of Citation and Outcome of Cholesterol Lowering Trials', British Medical Journal, Vol 305 (1992), 15-19.

64. Thacker, op. cit.

65. Chalmers, 1990 op. cit. and I. Chalmers, M. Adams, K. Dickersin, J. Hetherington, W. Tamow-Mordi, C. Meinert, S. Tonascia and T.C. Chalmers, 'A Cohort Study of Summary Reports of Controlled Trials', Journal of the American Medical Association, Vol 263 (March 9, 1990), 1401-1405.

66. Leydesdorff and Amsterdamska, op. cit.

67. O. Amsterdamska, personal communication August 1992.

68. See also K.H. Sørensen, 'Towards a Feminized Technology? Gendered Values in the Construction of Technology', Social Studies of Science, Vol 22 (February 1992), 5-31.

69. J. Law and R.J. Williams, 'Putting Facts Together: A Study of Scientific Persuasion', Social Studies of Science, Vol 12 (1982), 535-558, demonstrates the depth of interview and observational material needed to work out the intentions which lie behind citations and modalities in a published text.

70. I.I. Mitroff, 'Norms and Counter-Norms in a Select Group of the Apollo Moon Scientists: A Case Study of the Ambivalence of Scientists', American Sociological Review, Vol 39 (August 1974), 579-595.

71. One of the stronger formulations of this goal is found in B. Latour, 'The Impact of Science Studies on Political Philosophy', Science, Technology, & Human Values, Vol 16 (Winter 1991), 3-19.

72. For instance in J. Law and M. Callon, 'The Life and Death of an Aircraft: A Network Analysis of Technical Change' in W.E. Bijker and J. Law (eds.), op. cit., 21-52. Another example is A.R. Sætnan, 'Rigid Politics and Technological Flexibility - The Anatomy of a Failed Hospital Innovation', Science, Technology, & Human Values, Vol 16 (Autumn 1991), 419-447.