COCHRANE'S CORNER

Systematic reviews highlight the complex balance between good and harm from screening studies

L. Moja • I. Moschetti • A. Liberati

G.F. Gensini • R. Gusinu

L. Moja (✉) • I. Moschetti • A. Liberati
Italian Cochrane Centre, Istituto Mario Negri
Via Eritrea 62
I-20157, Milano, Italy
e-mail: moja@marionegri.it

G.F. Gensini
Department of Critical Care Medicine and Surgery, University of
Florence and Azienda Ospedaliero-Universitaria Careggi,
Florence, Italy

R. Gusinu
DAI Cardiologico e dei Vasi, Azienda Ospedaliero-Universitaria
Careggi, Florence, Italy

Olsen O, Gotzsche PC (2001)
Screening for breast cancer with mammography.
Cochrane Database Syst Rev (4):CD001877

Background Mammographic screening for breast cancer is controversial, as reflected in greatly varying national policies.

Objectives The objective was to assess the effect of screening for breast cancer with mammography on mortality and morbidity.

Search strategy MEDLINE (16 May 2000), The Cochrane Breast Cancer Group's trial register (24 Jan 2000) and reference lists. Letters, abstracts and unpublished trials. Authors were contacted.

Selection criteria Randomised trials comparing mammographic screening with no mammographic screening.

Data collection and analysis Data were extracted by both authors independently.

Main results Seven completed and eligible trials involving half a million women were identified. The two best trials provided medium-quality data and, when combined, yield a relative risk for overall mortality of 1.00 (95% CI 0.96-1.05) after 13 years. However, the trials are underpowered for all-cause mortality, and confidence intervals include a possible worthwhile effect as well as a possible detrimental effect. If data from all eligible trials (excluding flawed studies) are considered then the relative risk for overall mortality after 13 years is 1.01 (95% CI 0.99-1.03). The best trials failed to show a significant reduction in breast cancer mortality with a relative risk of 0.97 (95% CI 0.82-1.14). If data from all eligible trials (excluding flawed studies) are considered then the relative risk for breast cancer mortality after 13 years is 0.80 (95% CI 0.71-0.89). However, breast cancer mortality is considered to be an unreliable outcome and biased in favour of screening. Flaws are due to differential exclusion of women with breast cancer from analysis and differential misclassification of cause of death.

Reviewer's conclusions The currently available reliable evidence does not show a survival benefit of mass screening for breast cancer (and the evidence is inconclusive for breast cancer mortality). Women, clinicians and policy makers should consider these findings carefully when they decide whether or not to attend or support screening programs.

The methodologist's point of view

L. Moja, I. Moschetti, A. Liberati

We selected this review from The Cochrane Library for four main reasons. First, it addresses an issue of great public health relevance, which, in many ways, has been often referred to as 'a model case' for an evidence-based assessment of a population intervention. Second, it indicates that the same 'evidence base' can be interpreted in different ways depending on the "quality/methodology filter" adopted. Third, it shows that the aim of being 'fully explicit and reproducible' is not an easy one. Finally, it shows that the relationship between research evidence and health policy is more complex than usually anticipated.

A key date to dispute about screening for breast cancer

In 2000 a systematic review by Gotzsche and Olsen published in *Lancet* concluded that 'Screening for breast cancer with mammography is unjustified' [1]. The statement was sharp, direct and emotional. This was the conclusion of one of the most controversial reviews developed within The Cochrane Collaboration.

Looking at outcomes, numbers and trial subgroups

The results in the first *Lancet* review version reported that, when only the two trials considered adequately randomised were meta-analysed, the combined relative risk

(RR) was 1.04 (95% CI 0.84–1.27), with no effect of screening on breast cancer-specific mortality. No effect on total mortality emerged (0.99 (0.94–1.05)), though the studies were underpowered for this outcome. Gotzsche and Olsen stressed the point that all-cause (total) mortality should be preferred to disease-specific as primary outcome, because it is more reliable. Cancer-specific mortality (i.e., breast cancer deaths) is difficult to assess consistently and tends to favour breast screening. The pooled RR for breast cancer mortality for the other trials, considered by the authors to be methodologically flawed, was 0.75 (0.67–0.83), a result significantly different (p=0.005) from that for the two unbiased trials. In the review the authors refused to calculate the total pooled RR as they felt that the two groups of trials should not be combined.

The Cochrane–Lancet dispute

A different version of the same review was also published in the Cochrane Library and presented less negative conclusions, claiming that the available evidence was compatible with both a beneficial and detrimental effect of screening [2]. This was the result of a long and complex peer review process that took place within the Cochrane Breast Cancer Review Group, which strongly challenged the conclusion, revising it against the authors' wishes [3]. The *Lancet* supported Gotzsche and Olsen's original view against screening and stated that the Cochrane editors' attempt to negotiate the 'spin' that authors placed on their work diminished the value of the Cochrane review [4, 5]. Same methods and results sections, different emphasis in the authors' conclusions; different impact on screening policies. The conflict raised by the interpretation even eroded the friendship of the investigators themselves. Olsen, who designed how the trials' quality was related to the results, the core of the review, is no longer part of the authorship.

The complex balance

There are at least two reasons why this controversial Cochrane review is worth discussing. First, it clarified that the benefit from breast cancer screening is small. How small? It corresponds to an absolute risk reduction in breast cancer mortality of 0.05%: for every 2000 women invited for screening over 10 years (not one), one will have her life prolonged (only one). The relatively low numbers of events means that misclassification or biased exclusion of a few deaths could change the statistical significance of the trial results. More importantly, they may change the direction. Furthermore, one could also start looking at harm from screening; in their origi-

nal review Gotzsche and Olsen highlighted anxiety, labelling effect, false-positive results, unnecessary biopsies and the cascade of treatments to treat cancers of uncertain clinical significance coming from screening. Ten healthy women will be diagnosed as cancer patients and will be treated because of false-positive findings. That the effect of breast cancer screening is modest was also recognised by another systematic review developed by the US Preventive Task Force, which hypothesised an absolute risk reduction of 0.1 [6]. The expectation of harms is relevant compared with the benefits. Informed consent for women contemplating whether to attend a breast screening programme is now a fundamental element of good practice: the information should consider that the possible mortality benefit is counter-balanced by extra surgical risks, particularly for those women younger than 50. As this age threshold is somewhat arbitrary, all women should be advised about possible risks. In summary, most of the mammography trials have methodological limitations. Eliminating the information coming from these trials completely erases the mortality benefit. And it is hard to imagine that new randomised trials will be performed in the future. Keeping this information in the meta-analyses gives a small (SMALL) advantage to screening.

The bottom line

Meta-analyses developed through systematic reviews are not an atlas that offers readers several equally reliable routes through the terrain mapped out by the authors. Rather, meta-analyses depend on the stringency of the methodological quality assessment filter that is applied. By examining this quality assessment more closely, we come to see the authors' intention and the means by which they convey this intention [7]. Assessing the quality of evidence is a complex task and this assessment should consider not only elements of the study design but also aspects of study conduct, epidemiological relevance and consistency. All readers should examine the consistency and transparency of authors' conclusions against the way they have set forth the quality filter of their own analysis. In the end it is this transparency that make systematic reviews better than opinion-based narrative ones.

The clinician's point of view

G.F. Gensini, R. Gusinu
Screening for cancer remains a very emotional and debated issue of great public health relevance in the contemporary medical practice. This review from the Cochrane

Library shows that the analysis of published data results in multiple opinions based on a limited amount of reliable data.

In the same way the relationship between research-based evidence and health policy is more complex than usually anticipated and recommendations and decisions regarding cancer screening should be based on highly reliable data, and not on assumptions or speculations.

The assessment of the balance between benefits and disadvantages from screening studies is particularly complex in this specific area due to the importance of the methodological quality filter applied to the studies.

This systematic review represents an excellent model of conditions characterised by great public health relevance, by a number of clinical trials available, but severely affected by the uneven quality of trials.

This may lead to markedly different conclusions in relation to the filter applied to the quality of the evidence produced by trials, stressing the importance of a thorough evaluation of the stability and strength of medical evidence, as very clearly described in a recent paper [8].

Reference

1. Gotzsche PC, Olsen O (2000) Is screening for breast cancer with mammography justifiable? Lancet 355:129–134
2. Olsen O, Gotzsche PC (2001) Screening for breast cancer with mammography. Cochrane Database Syst Rev (4):CD001877
3. Langhorne P, Gotzsche PC, Goodare H, Dimmer C, Page K (2002) Screening mammography: setting the record straight. Lancet 359(9304):439–440; author reply 440–442
4. Horton R (2001) Screening mammography – an overview revisited. Lancet 358:1284–1285
5. Olsen O, Gotzsche PC (2001) Cochrane review on screening for breast cancer with mammography. Lancet 358:1340–1342
6. Humphrey LL, Helfand M, Chan BK, Woolf SH (2002) Breast cancer screening: a summary of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med 137:347–360
7. Moja LP, Telaro E, D'Amico R et al (2005) Assessment of methodological quality of primary studies by systematic reviews: results of the metaquality cross sectional study. BMJ 330:1053
8. Treadwell JR, Treagear SJ, Reston JT, Turkelson CM (2006) A system for rating the stability and strength of medical evidence BMC Med Res Methodol 6:52