The Breast xxx (2016) 1–9

Contents lists available at ScienceDirect

The Breast

journal homepage: www.elsevier.com/brst

Original article

Overview of guidelines on breast screening: Why recommendations differ and what to do about it

Karsten Juhl Jørgensen [a, *], Mette Kalager [b, c, d], Alexandra Barratt [e], Cornelia Baines [f], Per-Henrik Zahl [g], John Brodersen [h], Russell P. Harris [i]

[a] The Nordic Cochrane Centre, Rigshospitalet Department 7811, Blegdamsvej 9, DK-2100 Copenhagen, Denmark
[b] University of Oslo, Institute of Health and Society, Department of Health Management and Health Economics, P.O. Box 1089, Blindern, 0318 Oslo, Norway
[c] Department of Transplantation Medicine and KG Jebsen Center for Colorectal Cancer Research, Oslo University Hospital, Oslo, Norway
[d] Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA
[e] School of Public Health, University of Sydney, NSW 2006, Australia
[f] Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
[g] Norwegian Institute of Public Health, P.O. Box 4404, Nydalen, N-0403 Oslo, Norway
[h] Research Unit for General Practice and Section of General Practice, Department of Public Health, University of Copenhagen, Copenhagen & Primary Health Care Research Unit, Zealand Region, Denmark
[i] University of North Carolina, Chapel Hill, NC, USA

ARTICLE INFO

Article history:
Received 25 May 2016
Received in revised form
4 August 2016
Accepted 6 August 2016
Available online xxx

Keywords:
Breast cancer screening
Guideline development
Cancer epidemiology

ABSTRACT

Updated guidelines on breast cancer screening have been published by several major organisations over the past five years. Recommendations vary regarding both age range, screening interval, and even on whether breast screening should be offered at all. The variation between recommendations reflects substantial differences in estimates of the major benefit (breast cancer mortality reduction) and the major harm (overdiagnosis). Estimates vary considerably among randomised trials, as well as observational studies: from no benefit to large reductions, and from no overdiagnosis to substantial levels. The estimates vary according to the methodology of the randomised trials, and the design of the observational studies. Guideline recommendations reflect the choice of evidence informing them. While there are well-developed tools to deal with randomised trials in guideline work, these are not always used, or they may not be followed as recommended. Further, results of trials performed decades ago may no longer be applicable. For observational studies, the framework for inclusion in guidelines is not similarly well-developed and there are methodological concerns specific to screening interventions, such as small effects in absolute terms. There is a need for agreement on a hierarchy of observational study designs to quantify the major benefit and harm of cancer screening. This review provides a summary of recent guidelines on breast cancer screening and their major strengths and weaknesses, as well as a short overview of the major strengths and limitations of observational study designs. There is a need for agreement on a hierarchy of observational study designs in this field.

Introduction

Breast screening has been debated intensely for many years. A systematic Cochrane review [1] and additional analyses published in Lancet [2] questioned its benefit because key trials were sub-optimally randomised, and quantified overdiagnosis for the first time. Furthermore, the generalisability of trials from the 1970s and 1980s has been questioned [3]. Since the Cochrane review, some observational studies have questioned the promised benefit and attempted to quantify overdiagnosis [4–9]. Other observational studies, some using statistical modelling, have claimed that screening is associated with clear benefit and little or no overdiagnosis [10–13]. Two research challenges thus exist: how to best evaluate the evidence from the randomised trials; and how to best conduct and evaluate observational studies. In the absence of a consensus for performing observational studies, the methodologies chosen have determined the estimates [14–16].

* Corresponding author.
E-mail address: kj@cochrane.dk (K.J. Jørgensen).

http://dx.doi.org/10.1016/j.breast.2016.08.002
0960-9776/

This paper summarises the key guidelines on mammography screening published within the past five years. Guidelines were defined as summaries of evidence that led to statements about screening policy, developed by panels of individuals from different disciplines. This distinguishes guidelines from systematic reviews; Cochrane systematic reviews, for example, are not intended to advise on policy [17]. Our summary includes the methods, results, and recommendations of key guidelines and highlights their strengths and weaknesses. A short discussion of methodological designs in observational studies is also provided.

Dealing with evidence from randomised trials and observational studies in guidelines

The Grading of Recommendations, Assessment, Development and Evaluation (GRADE) approach is a widely used method for guideline development [18]. The Cochrane Risk of Bias tool can be used to assess the methodological rigor of randomised trials [17], an empirically based "checklist" of key methodological components (Table 1). Many of these quality criteria are straight-forward; for example individual randomisation is more likely than cluster-randomisation to produce comparable groups. GRADE further includes an assessment of the reliability of the effect estimates across trials, by considering factors that may reduce confidence in the evidence such as indirectness (the external validity of trial results); heterogeneity (variation between trial results); and imprecision (width of confidence intervals). GRADE recommends analysing trials of high and low reliability separately in sensitivity analyses, and that estimates from trials of high reliability are preferred when these differ substantially from those from low reliability trials [19].

GRADE allows observational studies to be included in the body of evidence, e.g. high quality cohort studies with a low probability of confounding (or where confounders would reduce the observed effect); a very large effect; and/or with a dose-response relationship. If these issues are satisfied, confidence in estimated effects may increase [19]. However, observational studies of breast screening are unlikely to fulfil these criteria. There are substantial possible confounders such as self-selection bias and improved therapy; the benefit is small in absolute terms; and there is no dose-effect relationship [20–22]. The small screening effect compared to therapeutic interventions means that biases in screening trials are more likely to create or erase screening effects. As observational studies are more susceptible to bias than randomised trials, their estimates of effect are likely to vary substantially among studies.

Despite this, it is essential to quantify the effects of breast screening today in the most reliable way possible as the premises for mammography screening have changed importantly with increased awareness and improved treatment [16].

Canadian Task Force on Preventive Health Care (2011)

The Canadian Task Force on Preventive Health Care is an independent panel of clinicians and methodologists. Their recommendations were based on a systematic review, building on a review from the United States Preventive Services Task Force. The strength of the evidence and recommendations were determined using GRADE [20]. Observational studies were not used to assess benefits but were considered for quantifying harms, as these are generally not well-reported in the trials [23]. The Task Force found that blinding and concealment of allocation was not clear in five of the trials and that only three could be "considered truly randomised". Apparently, no sensitivity analyses were performed to test if effect estimates were robust to differences in the reliability of the trials.

Analyses and recommendations were presented by age group. Average risk women below age 50 years were given a weak recommendation against screening based on moderate-quality evidence, reflecting that the undesirable effects probably outweigh the desirable ones but that important uncertainties exist. To avoid 1 death from breast cancer, 2108 women should be screened biennially for 11 years at the cost of 690 false positives, 75 of which would lead to an unnecessary biopsy, and about 10 women would be overdiagnosed and overtreated.

Average risk women aged 50–69 years were given a weak recommendation in favour of screening every 2–3 years based on moderate-quality evidence. To avoid 1 breast cancer death, 721 women should be screened biennially for 11 years at a cost of 204 false positives and 26 unnecessary biopsies, whereas 4 would be overdiagnosed. An overdiagnosis estimate appeared only in an Appendix [24]. The Task Force noted that women who do not greatly value the small benefit and worry about false positives and overdiagnosis may decline screening. For average risk women aged 70–74 years, the recommendation was similar, but based on weak evidence.

Independent UK Panel on breast screening (2012)

The Independent UK Panel [25] explicitly adopted efficacy data from the Cochrane review (Fig. 1), but preferred a random effects rather than a fixed effect meta-analysis [26], with negligible effect on the result. As in the Cochrane review, the Edinburgh trial [27] was excluded due to baseline differences caused by cluster-randomisation. The UK Panel considered the remaining trials to be reliable and a sensitivity analysis was not presented.

The Panel considered comparisons of all-cause mortality in the trials irrelevant due to the small expected effect but acknowledged substantial uncertainty arising from trial limitations and age.

Table 1
The Cochrane risk of bias tool.

Risk of bias item	Description
Random sequence generation	Is the method used to generate the random sequence generation robust (i.e. centrally done by computer) or modifiable (i.e. toss of coin) and likely to result in comparable groups?
Allocation concealment	Can the investigator who determines if an individual can be enrolled foresee the allocated group prior to making a decision?
Blinding of participants and personnel (performance bias)	Were the participants and administrators of the intervention aware or able to guess the allocated intervention?
Blinding of outcome assessor (detection bias)	For example; blinded cause of death assessment.
Incomplete outcome data (attrition bias)	For example; drop-out rates.
Selective reporting (reporting bias)	Were pre-planned outcomes likely missing?
Other bias	For example; early stopping of the trial, substantial baseline-imbalances or use of cluster-randomisation.

Please cite this article in press as: Jørgensen KJ, et al., Overview of guidelines on breast screening: Why recommendations differ and what to do about it, The Breast (2016), http://dx.doi.org/10.1016/j.breast.2016.08.002