BMJ

CrossMark
click for updates

RESEARCH

Quantification of harms in cancer screening trials: literature review

OPEN ACCESS

Bruno Heleno *PhD fellow*[1], Maria F Thomsen *registrar*[1], David S Rodrigues *consultant general practitioner*[2], Karsten J Jørgensen *senior researcher*[3], John Brodersen *associate professor*[1]

[1]Research Unit for General Practice and Section of General Practice, Department of Public Health, University of Copenhagen, Øster Farimagsgade 5, PO Box 2099, 1014 Copenhagen K, Denmark; [2]Family Medicine Department, Nova Medical School, Campo dos Mártires da Pátria 130, 1169-056 Lisbon, Portugal; [3]Nordic Cochrane Centre, Rigshospitalet, Department 7811, Blegdamsvej 9, 2100 Copenhagen, Denmark

Abstract

Objectives To assess how often harm is quantified in randomised trials of cancer screening.

Design Two authors independently extracted data on harms from randomised cancer screening trials. Binary outcomes were described as proportions and continuous outcomes with medians and interquartile ranges.

Data sources For cancer screening previously assessed in a Cochrane review, we identified trials from their reference lists and updated the search in CENTRAL. For cancer screening not assessed in a Cochrane review, we searched CENTRAL, Medline, and Embase.

Eligibility criteria for selecting studies Randomised trials that assessed the efficacy of cancer screening for reducing incidence of cancer, cancer specific mortality, and/or all cause mortality.

Data extraction Two reviewers independently assessed articles for eligibility. Two reviewers, who were blinded to the identity of the study's authors, assessed whether absolute numbers or incidence rates of outcomes related to harm were provided separately for the screening and control groups. The outcomes were false positive findings, overdiagnosis, negative psychosocial consequences, somatic complications, invasive follow-up procedures, all cause mortality, and withdrawals because of adverse events.

Results Out of 4590 articles assessed, 198 (57 trials, 10 screening technologies) matched the inclusion criteria. False positive findings were quantified in two of 57 trials (4%, 95% confidence interval 0% to 12%), overdiagnosis in four (7%, 2% to 18%), negative psychosocial consequences in five (9%, 3% to 20%), somatic complications in 11 (19%, 10% to 32%), use of invasive follow-up procedures in 27 (47%, 34% to 61%), all cause mortality in 34 (60%, 46% to 72%), and withdrawals because of adverse effects in one trial (2%, 0% to 11%). The median percentage of space in the results section that reported harms was 12% (interquartile range 2-19%).

Conclusions Cancer screening trials seldom quantify the harms of screening. Of the 57 cancer screening trials examined, the most important harms of screening—overdiagnosis and false positive findings—were quantified in only 7% and 4%, respectively.

Introduction

Cancer screening can lead to harm as well as benefit.[1-3] Harm related to screening can be somatic or psychosocial.[4-13] Harms result from the screening test itself, from investigations because of false positive findings, and from overdiagnosis with subsequent overtreatment.[3 5 12 13] Given the potential for serious harms in healthy individuals, screening should be offered only when the benefits are firmly documented and considered to outweigh the harms, which should be equally well quantified. The determination of benefit from screening requires assessment in randomised clinical trials, which are also capable of providing high quality evidence on harms.[14 15] In general, however, harms are poorly reported in randomised trials,[16-31] and there is some evidence that reporting of harms is worse in non-pharmacological trials than in trials assessing drugs.[22-24]

At least three additional arguments support the importance of reporting harms in randomised trials of cancer screening. Firstly, screening is offered to healthy individuals and is an intervention initiated by the healthcare system, not at the request from a patient to solve a health problem. Secondly, interventions for which the benefits are modest or uncertain merit detailed consideration of harms,[32] and systematic reviews of randomised

trials of screening have shown either modest[33-35] or no[36] reductions in cancer specific mortality. Thirdly, a benefit for some will come at the expense of harm to others.[37-39]

The minimum evidence required to assess the harms of screening includes the frequencies of false positive findings, overdiagnosis, and complications of diagnostic investigations and treatment.[13] In addition, withdrawals because of harms[19] and the use of invasive follow-up procedures can be considered as proxy measures of severe harms. We hypothesised that cancer screening trials would not consistently or sufficiently quantify the expected associated harms.

Methods

Eligibility criteria

We included trials that evaluated breast cancer screening with mammography, self examination, or clinical examination; colorectal cancer screening with sigmoidoscopy or colonoscopy, faecal occult blood testing, or virtual colonoscopy; liver cancer screening with ultrasonography, α fetoprotein, or a combination; lung cancer screening with chest radiography or low dose spiral computed tomography of chest; ovarian cancer screening with ultrasonography, serological markers, or a combination; oral cancer screening with visual inspection; prostate cancer screening with prostate specific antigen, digital rectal examination, or a combination; and testicular cancer screening with self examination or clinical examination.

Publications reporting randomised trials were eligible if the trial compared a group of participants undergoing a cancer screening intervention with either no screening or an alternative screening intervention. Participants could be part of the general population or of a high risk population, such as heavy smokers. Trials had to assess the efficacy of cancer screening, defined as a reduction in the incidence of cancer, cancer specific mortality, or all cause mortality. Trials were included regardless of risk of bias. Individual articles were eligible if they provided data for both the screening and control groups, and if they did not pool data from randomised trials and observational studies. Finally, to be eligible, articles must have provided data for all participants, a random sample of all participants, or all participants enrolled in a single centre of a multicentre trial.

Search strategy for the identification of articles

We extracted the references to trials from Cochrane Systematic Reviews when these were available and performed an updated search in the Cochrane Central Register of Controlled Trials using the search terms described in each Cochrane review to find articles published since the review (appendix 1).

When no Cochrane Systematic Review was available, we sought clinical trial reports in the Cochrane Central Register of Controlled Trials. Our search strategies used a combination of controlled vocabulary (MeSH terms) and free text terms. They had three dimensions: terms related to cancer, terms related to the screening technology, and the term "screening" and its synonyms (appendix 1). We planned not to have language restriction, but because of lack of resources we were unable to translate 12 articles identified for assessment of eligibility (six in Mandarin and six in Russian). Our last search was in May 2012.

It subsequently became clear that our search strategy missed some potentially relevant articles. We amended the protocol and designed new searches including either the name of trials known to us or the name of the principal investigators of the trials. These new searches were performed in Medline (1946-14 Aug 2012), Medline In-process and other non-indexed citations (to 14 Aug 2012), and Embase (1974-14 Aug 2012) (appendix 1).

We did not contact study authors nor did we perform searches of grey literature, such as doctoral theses or conference proceedings, as this would not reflect the information that is readily available in the literature.

Data collection and analysis

Selection of studies

BH and DSR independently scanned the titles and abstracts from reference lists of Cochrane Systematic Reviews and from the electronic searches. When the title or abstract did not provide sufficient data to rule out eligibility, the full text was obtained. Disagreements were solved through consensus.

Data extraction and management

All articles were collected in a digital file format (pdf). Two weeks before data extraction, BH concealed information about authors, affiliations, date of publication, journal, and references with the stamp function in Adobe Acrobat 9 Pro (version 9.5.2). The pdf files were encrypted with a password, which restricted changes in the file security settings. BH and MFT independently extracted the data using standardised forms and blinded to each other's results. Both authors extracted the data from the encrypted pdf files that concealed author identification, year of publication, and the name of the journal, but not trial identification. Disagreements were solved through consensus.

When results from a single trial were reported in multiple publications, data were collected in separate forms for each publication, but our unit of analysis was at trial level. If a single publication provided data from two or more trials, information for each trial was collected in separate forms.

Harm data

We included seven types of harms related to cancer screening: overdiagnosis, false positive findings, somatic complications caused by screening or follow-up procedures, negative psychosocial consequences caused by screening test or follow-up procedures, the additional number of participants subjected to invasive procedures, all cause mortality (which might increase if harms include, for example, invasive follow-up procedures or substantial overdiagnosis and overtreatment), and withdrawals because of adverse events. For the qualitative assessment of harms, we required that two criteria were met before considering that an outcome had been reported: the absolute numbers or incidence rates had to be provided and the outcome must have been explicitly mentioned. We accepted the trial authors' definition of the outcome and did not assess whether that definition was appropriate. For example, in sigmoidoscopy trials we considered that false positives had been reported if false positives were mentioned, regardless of they were defined as a screening test with a positive result but no cancer, a screening test with positive result but no advanced adenoma or cancer, or a screening test with a positive result but no polyp.

We also extracted a crude quantitative measure of harm reporting. We marked the results section of each publication using a 1 cm^2 grid. Thereafter, we measured the space devoted to the results section and the space devoted to reporting any of the harms mentioned above. The quantitative measure of harm reporting was the percentage of space devoted to harms out of the total space in the results section. Similar quantitative

measures have been used in previous reviews of reporting of harm in randomised trials.[19 23 24 30]

Other publication parameters

We extracted information on the type of screening technology, the year of recruitment of the first participant, whether disease specific mortality or incidence had been quantified, target population (general population or high risk group), geographical location, type of control group (unscreened group or alternative screening technology), and whether participants were individually or cluster randomised. If publications from the same trial mentioned different dates for the year of recruitment of the first participant, we chose the earliest date.

Data at trial level

We pooled data from all articles from the same trial. If at least one article from the trial did so, we considered that a trial had quantified a specific harm. We applied the same criterion to data on incidence of cancer and mortality. When trials were reported in more than one article, the space devoted to the results section was defined as the sum of the space of the results sections in each article. Likewise, the space devoted to harms was the sum of the space devoted to harms in each of the articles reporting that trial.

Prespecified analyses

Analyses consisted of a descriptive assessment of included variables. We used proportions and exact confidence intervals for binary outcomes and medians and interquartile ranges for continuous outcomes. In the protocol, we hypothesised that the date of enrolment of participants could be an explanatory variable for quantification of harm and that this could be tested with regression models. Because of a lack of data, however, this could not be done. All statistical analyses were performed in R version 3.0.1.

Additional analyses

In the original protocol (appendix 2), we specified that we would include only articles that reported data from all trial arms. While BH and DSR were assessing the articles for eligibility, we noted that several articles contained relevant information on harms only for the screened participants. Hence, BH reassessed all articles related to the included trials and identified those that reported data on harms only in the intervention group. These articles were included in an unplanned subsidiary analysis to test the robustness of the findings of the main analysis. As 12 articles were not translated, we preformed another unplanned sensitivity analysis assuming that these articles had quantified harms.

We tabulated harm quantification according to the screening technology, geographical location, and type of control group (unscreened group or alternative screening technology). Given the small number of trials, we did not perform stratified statistical analysis. In some trials screening led to a reduced incidence of cancer (thus, if overdiagnosis existed it would be impossible to detect), and in other trials the study design might have been inappropriate to assess it (short follow-up or use of another screening intervention as a control group), so we performed an analysis that excluded both these groups of trials from the denominators.

The protocol for this literature review and its amendment are available in appendix 2.

Results

Out of 4590 titles identified, we found 63 trials that aimed to assess the effect of cancer screening on cancer specific or all cause mortality, or both. Of these, only 57 had published results in at least one article that matched our eligibility criteria (figure⇓ and appendix 3). Of the six remaining trials, one trial was not completed because of low compliance, one trial had not started enrolling participants, and four trials are not yet completed. These four trials have reported results for the screened group but not for the control group. The 57 trials assessed 10 different screening interventions and enrolled 3 419 036 participants. We found no trials on testicular cancer screening or colorectal cancer screening with virtual colonoscopy.

Some of the 57 trials were reported in several articles. We found 198 articles that included data on both the screened and the control groups and used these in our main analyses. We also found 44 articles that reported data for the screened groups but not for the control groups. Our analyses were replicated in the combined 242 articles to assess whether harms had been quantified in at least the screened groups.

Table 1 shows the proportion of trials that quantified each individual outcome in at least one of the eligible articles describing that trial⇓ (the individual assessment of the trials and respective characteristics are available from the authors). Overall, cancer specific mortality and cancer specific incidence were quantified more often than harm related outcomes. While the former two were quantified in more than 80% of the trials, false positive findings were quantified for only two trials (4%, 95% confidence interval 0% to 12%), and overdiagnosis was quantified in four trials (7%, 2% to 18%). Only one trial (2%, 0% to 11%) quantified the number of withdrawals because of adverse effects. The median percentage of space in the results section devoted to harms data was 12% (interquartile range 2-19%). Table 2 shows quantification of harm stratified by type of screening.⇓ Quantification of harm stratified by geographical location of the trial and type of control group is shown in tables S1 and S2 in appendix 4.

We performed several sensitivity analyses with less strict criteria. When we also considered the 44 articles that reported data from only the screened groups, the proportion of trials that quantified some of the harms increased (table 1⇓). For example, false positive findings were now quantified in 32% (95% confidence interval 20% to 45%) of trials, while overdiagnosis remained quantified in only 7% of trials. The results of the sensitivity analyses—which assumed the non-translated articles had quantified all the outcomes—were similar to those of the main analyses (table S3 in appendix 4). We also restricted the analysis of overdiagnosis to the trials which, after extended follow-up, found a higher incidence of cancer in screened participants than in the unscreened control group. In this subset, overdiagnosis was quantified in two out of 12 trials (17%, 2% to 48%).

Discussion

Summary of main results

The most important harms of screening—overdiagnosis and false positive findings—were quantified in only a minority of trials. Out of 57 cancer screening trials, 7% quantified overdiagnosis[40-43] and 4% quantified false positive results.[44 45] Only one trial reported the number of withdrawals because of harmful events (2%),[46] and the median amount of space devoted to reporting of harms in the results section of the trial reports was 12%. Consequently, cancer screening trials rarely report

what is considered the minimum amount of evidence required to quantify the harms of screening. In contrast, the effect of cancer screening on cancer specific mortality was reported in 89% of trials. It is therefore often difficult or impossible to weigh benefits against harms in cancer screening.

Interpretation of the results

We found few trials that met our criteria for minimal harm reporting, which suggests poor reporting of harms. An alternative explanation would be that our assessment criteria focus on irrelevant outcomes or that they do not capture the important aspects of harm reporting. Their relevance, however, is supported in several concept papers and editorials about the harms of screening.[4-13] The exception is withdrawals, which is an unusual concept in screening literature and was taken from previous reviews of harm reporting.[16 19] We included this outcome because it reflects the ultimate decision of the participant, the physician, or both, to discontinue an intervention.[16] The importance of withdrawals, however, is controversial,[32] and trialists might have considered it irrelevant. When the decision to withdraw is made by a clinician, it is possible to recognise this from the participant's case report form; but when the decision is taken by the participant it might be difficult to distinguish it from other causes of loss to follow-up. Additionally, in some trials there was no direct contact with the control group and their information was collected from registries. These participants could not withdraw as they were unaware that they were part of a trial. In these trials it would be inappropriate to require withdrawal data from the controls. In summary, there are arguments against the relevance of withdrawals as a surrogate of harm; however, even when we excluded this outcome, the general pattern of poor reporting of harm persisted.

Three aspects of the criteria used to appraise the harm outcomes could be discussed. The first is whether these outcomes can be assessed for all trials. It is possible to argue that overdiagnosis cannot be assessed in all trials as this ideally requires a persistent increase in incidence after a long follow-up.[41 47] We chose to present overdiagnosis by including all 57 trials in the denominator as it had been specified in our review protocol. For completeness, we also restricted the analysis to trials with long term follow-up, an increased incidence of cancer in the screened group, and an unscreened control group. Although the proportion of trials reporting overdiagnosis is higher in this subset (17% compared with 7% in the main analysis), it is still unacceptably low and makes no change to our conclusion. We found no reason to exclude trials from the denominators of the other analyses.

The second point is whether it is relevant to collect data from the unscreened arms for all harm outcomes. Reporting harm outcomes for the intervention and control groups is a central recommendation in the guidelines for reporting randomised trials,[48 49] and we therefore assessed whether each screening harm was reported in both groups of the trials. Symptoms and incidental findings in unscreened participants can lead to invasive procedures, adverse psychosocial consequences, complications of diagnosis and treatment, and mortality from other causes. Hence, these four outcomes should be reported for the control groups to make it possible to assess the level of any surplus harm in the screened group. In contrast, it can be argued that it is necessary to undergo a screening test to experience a false positive result. If we accept this premise, participants in unscreened control groups cannot experience false positive findings and it would be adequate to report false positives only for the screened group when the comparator is

no screening. This would mean that 18 trials (32%) would have reported false positive findings adequately, instead of the two (4%) that report them for both arms. Even if we allow for a less strict criterion, the vast majority of trials do not provide data on false positive findings. It is also possible to argue that it is not meaningful to report overdiagnosis for the screened and the unscreened groups because it can be quantified only indirectly (that is, comparing the incidence between groups). We considered, however, that overdiagnosis was reported when incidence data had been provided for both arms and the term "overdiagnosis" or equivalent had been used.

Thirdly, we could have included data presented outside of the results section. Some trials presented data on harm in the discussion section but not in the results section: five trials presented numbers for overdiagnosis,[50-54] three trials presented the number of false positive results in the screened group,[55-57] and one presented the number of invasive procedures in the screened group.[57] In the case of overdiagnosis, the trials presented the difference in incidence of cancer in the results section and interpreted this as overdiagnosis in the discussion section. Addition of these to the numerator would mean that of 57 trials, nine (16%, 95% confidence interval 7% to 28%) reported overdiagnosis. When we considered the smaller subset of 12 trials with long term follow-up, increased incidence of cancer in the screened group, and an unscreened control group, overdiagnosis was reported in four trials (33%, 10% to 65%). Thus, in the most optimistic scenario only a third of the trials report the most serious harm of screening.

Strengths and weaknesses

We identified trial articles from either Cochrane Reviews or electronic searches in three different databases. It is therefore unlikely that we missed any important article for the trials included in this review. Our searches also identified trials of screening interventions that we were not aware of before this study (for example, cervical cancer screening with visual inspection or human papillomavirus typing). We have not included these in this review, which is therefore not fully comprehensive.

We tried to avoid underestimation of quantification of harm in four ways. Firstly, we assessed only whether any quantification of harms was present, not whether the harms were adequately reported or correctly defined. Secondly, we considered that harms had been reported even when the quality of the data on harms was lower than that for benefits (that is, when harms data were reported for only a subset of the included participants). Thirdly, when data on harms were provided in a table or a figure, we included the entire table or figure area in the numerator of our estimate of space devoted to harms. Data on harms often accounted for a small number of lines in a table. Finally, and unlike previous surveys of harm reporting,[16-25 27] we extracted data from multiple articles reporting on a single trial.

Most of the concept papers and editorials about harms of screening were published within the past decade.[4 7-12] Also, the data monitoring committee of the PLCO trial has recently recommended early stopping of the prostate cancer screening part because of concerns about harms.[58] This suggests more concern about screening harms in recent years. Because of the small number of trials, however, we could not assess whether harm reporting has improved in recent years, as originally planned in the review protocol.

Strengths and weaknesses in relation to other studies

Several authors have stated that adequate evidence on the harms of cancer screening is lacking.[5 12 13 59] We identified only one previous study that attempted to quantify the problem. This was a review of research articles about mammography screening including study designs other than randomised trials. Thirty eight per cent of the included articles did not report any harms; 23% mentioned harms but presented them as unimportant; and 39% acknowledged the existence of harms.[60] Our results extend this finding to other types of cancer screening.

Several literature reviews have assessed harms reported in randomised trials of other types of medical interventions.[16-31] In these reviews, the proportion of trials reporting absolute numbers for various harms ranged from 41% to 88%,[19 23 24 27 29] withdrawals because of harms in 25-94%,[19 21 23 24 26 27 30] and the median space devoted to harms in the results section ranged from 0% to 14%.[19 20 23 30] Although cancer screening trials assess a preventive activity targeted at healthy individuals, our results show that reporting of harm is no better than what was found in reviews of therapeutic interventions.

Meaning of the study

The trials we reviewed included large numbers of participants, followed them for long periods of time, required many resources, and provided valuable information about the impact of screening on cancer specific mortality. However, we found that the harms were poorly reported. Healthcare decision makers, healthcare practitioners, and, ultimately, patients therefore cannot make informed choices about cancer screening. This is problematic as many cancer screening programmes have important associated harms.

While we acknowledge that collecting data on harms will complicate cancer screening trials, this is not a sound argument against the strong ethical obligation to collect such data. If trialists do not report certain outcomes because they consider that the harms will be either rare or irrelevant when compared with the potential decrease in mortality, such information will not be available for people who judge these outcomes differently. We think that future screening trials should collect and report the expected harms of screening (false positives, overdiagnosis and overtreatment, psychosocial consequences, somatic complications, and all cause mortality). Adequate reporting of harm requires data from the control group as these provide a reference level and help to interpret harms data from the screened group.

Implications for future research

The CONSORT statement,[48 49] which aims to improve reporting of clinical trials, has an extension specifically devoted to reporting of harm.[16] Although most of the examples in this extension come from pharmacological trials, the extension is also applicable to screening trials. There are some topics, however, where direct application of the CONSORT statement to cancer screening trials seems difficult. How can withdrawals because of harmful events be distinguished from other sources of loss to follow-up in screening trials? Are there specific harms in cancer screening where data from the intervention group is enough? Can scales be used to grade screening harms for severity? A discussion of these questions could help to standardise harm reporting in randomised trials of cancer screening and will hopefully lead to more complete evidence that allows informed decisions.

1　Brawley OW, Kramer BS. Cancer screening in theory and in practice. *J Clin Oncol* 2005;23:293-300.
2　Raffle A, Gray M. Screening: evidence and practice. Oxford University Press, 2007.
3　Croswell JM, Ransohoff DF, Kramer BS. Principles of cancer screening: lessons from history and study design issues. *Semin Oncol* 2010;37:202-15.
4　Holland W, Stewart S. Screening in disease prevention: what works? 1st ed. Radcliffe, 2005.
5　Marshall KG. Prevention. How much harm? How much benefit? 3. Physical, psychological and social harm. *CMAJ* 1996;155:169-76.
6　Barratt A, Irwig L. Users' guides to the medical literature: XVII. how to use guidelines and recommendations about screening. *JAMA* 1999;281:2029-34.
7　Woolf SH. Potential of screening to reduce the burden of cancer. In: Curry SJ, Byers T, Hewitt ME, National Cancer Policy Board (US), eds. Fulfilling the potential of cancer prevention and early detection. National Academies Press, 2003:156-223.
8　Alibhai SMH. Cancer screening: the importance of outcome measures. *Crit Rev Oncol Hematol* 2006;57:215-24.
9　Irwig L, McCaffery K, Salkeld G, Bossuyt P. Informed choice for screening: implications for evaluation. *BMJ* 2006;332:1148-50.
10　Brodersen J, Jørgensen KJ, Gøtzsche PC. The benefits and harms of screening for cancer with a focus on breast screening. *Pol Arch Med Wewn* 2010;120:89-94.
11　Newman TB, Kohn MA. Screening tests. Evidence-based diagnosis. Cambridge University Press, 2009:116-37.
12　Dans LF, Silvestre MAA, Dans AL. Trade-off between benefit and harm is crucial in health screening recommendations. Part I: general principles. *J Clin Epidemiol* 2011;64:231-9.
13　Harris R, Sawaya GF, Moyer VA, Calonge N. Reconsidering the criteria for evaluating proposed screening programs: reflections from 4 current and former members of the U.S. Preventive Services Task Force. *Epidemiol Rev* 2011;33:20-35.
14　Papanikolaou PN, Ioannidis JPA. Availability of large-scale evidence on specific harms from systematic reviews of randomized trials. *Am J Med* 2004;117:582-9.
15　Chou R, Aronson N, Atkins D, Ismaila AS, Santaguida P, Smith DH, et al. AHRQ Series Paper 4: Assessing harms when comparing medical interventions: AHRQ and the Effective Health-Care Program. *J Clin Epidemiol* 2010;63:502-12.
16　Ioannidis JPA, Evans SJW, Gøtzsche PC, O'Neill RT, Altman DG, Schulz K, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med* 2004;141:781-8.
17　Ioannidis JP, Contopoulos-Ioannidis DG. Reporting of safety data from randomised trials. *Lancet* 1998;352:1752-3.
18　Edwards JE, McQuay HJ, Moore RA, Collins SL. Reporting of adverse effects in clinical trials should be improved: lessons from acute postoperative pain. *J Pain Symptom Manage* 1999;18:427-37.
19　Ioannidis JP, Lau J. Completeness of safety reporting in randomized trials: an evaluation of 7 medical areas. *JAMA* 2001;285:437-43.
20　Loke YK, Derry S. Reporting of adverse drug reactions in randomised controlled trials—a systematic survey. *BMC Clin Pharmacol* 2001;1:3.
21　Bryant J, Loveman E, Cave C, Chase D, Milne R. Endocrinology trial design: adverse event reporting in randomised controlled trials of recombinant human GH in GH-deficient adults. *J Endocrinol* 2002;175:545-52.
22　Martin RCG 2nd, Brennan MF, Jaques DP. Quality of complication reporting in the surgical literature. *Ann Surg* 2002;235:803-13.
23　Papanikolaou PN, Churchill R, Wahlbeck K, Ioannidis JPA. Safety reporting in randomized trials of mental health interventions. *Am J Psychiatry* 2004;161:1692-7.
24　Ethgen M, Boutron I, Baron G, Giraudeau B, Sibilia J, Ravaud P. Reporting of harm in randomized, controlled trials of nonpharmacologic treatment for rheumatic disease. *Ann Intern Med* 2005;143:20-5.

What is known about this topic

Cancer screening programmes require detailed consideration of harms as they target healthy people

Harms from screening include overdiagnosis and overtreatment, false positive findings, additional invasive procedures, negative psychosocial consequences, and somatic complications

It is unknown whether trials that assess cancer screening routinely quantify harms

What this study adds

The most important harms of screening, overdiagnosis and false positive findings were quantified in only 7% and 4% of 57 cancer screening trials, respectively

There was variation in the proportion of trials quantifying other outcomes (from 9% quantifying psychological consequences to 60% quantifying all cause mortality)

25 Lee P, Fischer H, Rochon P, Gill S, Herrmann N, Bell C, et al. Published randomized controlled trials of drug therapy for dementia often lack complete data on harm. *J Clin Epidemiol* 2008;61:1152-60.

26 Chowers MY, Gottesman BS, Leibovici L, Pielmeier U, Andreassen S, Paul M. Reporting of adverse events in randomized controlled trials of highly active antiretroviral therapy: systematic review. *J Antimicrob Chemother* 2009;64:239-50.

27 Pitrou I, Boutron I, Ahmad N, Ravaud P. Reporting of safety results in published reports of randomized controlled trials. *Arch Intern Med* 2009;169:1756-61.

28 Breau RH, Gaboury I, Scales Jr CD, Fesperman SF, Watterson JD, Dahm P. Reporting of harm in randomized controlled trials published in the urological literature. *J Urol* 2010;183:1693-7.

29 De Vries TW, Roon EN van. Low quality of reporting adverse drug reactions in paediatric randomised controlled trials. *Arch Dis Child* 2010;95:1023-6.

30 Haidich A-B, Birtsou C, Dardavessis T, Tirodimos I, Arvanitidou M. The quality of safety reporting in trials is still suboptimal: survey of major general medical journals. *J Clin Epidemiol* 2011;64:124-35.

31 Vera-Badillo FE, Shapiro R, Ocana A, Amir E, Tannock IF. Bias in reporting of end points of efficacy and toxicity in randomized, clinical trials for women with breast cancer. *Ann Oncol* 2013;24:1238-44.

32 Loke YK, Price D, Herxheimer A. Adverse effects. In: Higgins J, Green S, eds. Cochrane handbook for systematic reviews of interventions [Internet]. Cochrane Collaboration, 2011. www.cochrane-handbook.org

33 Gøtzsche PC, Jørgensen KJ. Screening for breast cancer with mammography. Cochrane Database of Systematic Reviews. John Wiley, 2013.

34 Hewitson P, Glasziou PP, Irwig L, Towler B, Watson E. Screening for colorectal cancer using the faecal occult blood test, Hemoccult. Cochrane Database of Systematic Reviews. John Wiley, 2007.

35 Bach PB, Mirkin JN, Oliver TK, Azzoli CG, Berry DA, Brawley OW, et al. Benefits and harms of CT screening for lung cancer: a systematic review. *JAMA* 2012;307:2418-29.

36 Wun YT, Dickinson JA. Alpha-fetoprotein and/or liver ultrasonography for liver cancer screening in patients with chronic hepatitis B. Cochrane Database of Systematic Reviews. John Wiley, 2003.

37 Principles of screening. In: Levin B, Prorok PC, Schottenfeld D, Fraumeni JF, eds. Cancer Epidemiology and Prevention. Oxford University Press, 2006.

38 De Koning HJ. Assessment of nationwide cancer-screening programmes. *Lancet* 2000;355:80-1.

39 Raffle AE, Alden B, Quinn M, Babb PJ, Brett MT. Outcomes of screening to prevent cancer: analysis of cumulative incidence of cervical abnormality and modelling of cases and deaths prevented. *BMJ* 2003;326:901.

40 Robinson MH, Hardcastle JD, Moss SM, Amar SS, Chamberlain JO, Armitage NC, et al. The risks of screening: data from the Nottingham randomised controlled trial of faecal occult blood screening for colorectal cancer. *Gut* 1999;45:588-92.

41 Zackrisson S, Andersson I, Janzon L, Manjer J, Garne JP. Rate of over-diagnosis of breast cancer 15 years after end of Malmö mammographic screening trial: follow-up study. *BMJ* 2006;332:689-92.

42 Andersson I, Janzon L. Reduced breast cancer mortality in women under age 50: updated results from the Malmö Mammographic Screening Program. *J Natl Cancer Inst Monographs* 1997;22:63-7.

43 Andriole GL, Crawford ED, Grubb RL 3rd, Buys SS, Chia D, Church TR, et al. Prostate cancer screening in the randomized prostate, lung, colorectal, and ovarian cancer screening trial: mortality results after 13 years of follow-up. *J Natl Cancer Inst* 2012;104:125-32.

44 Semiglazov VF, Moiseyenko VM, Bavli JL, Migmanova NS, Seleznyov NK, Popova RT, et al. The role of breast self-examination in early breast cancer detection (results of the 5-years USSR/WHO randomized study in Leningrad). *Eur J Epidemiol* 1992;8:498-502.

45 National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011;365:395-409.

46 Varenhorst E, Carlsson P, Capik E, Löfman O, Pedersen KV. Repeated screening for carcinoma of the prostate by digital rectal examination in a randomly selected population. *Acta Oncol* 1992;31:815-21.

47 Welch HG, Black WC. Overdiagnosis in cancer. *J Natl Cancer Inst* 2010;102s:605-13.

48 Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c332.

49 Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, et al. CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c869.

50 Marcus PM, Bergstralh EJ, Zweig MH, Harris A, Offord KP, Fontana RS. Extended lung cancer incidence follow-up in the Mayo Lung Project and overdiagnosis. *J Natl Cancer Inst* 2006;98:748-56.

51 Miller AB, To T, Baines CJ, Wall C. The Canadian National Breast Screening Study-1: breast cancer mortality after 11 to 16 years of follow-up. A randomized screening trial of mammography in women age 40 to 49 years. *Ann Intern Med* 2002;137:305-12.

52 Buys SS, Partridge E, Black A, Johnson CC, Lamerato L, Isaacs C, et al. Effect of screening on ovarian cancer mortality: the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Randomized Controlled Trial. *JAMA* 2011;305:2295-303.

53 Oken MM, Hocking WG, Kvale PA, Andriole GL, Buys SS, Church TR, et al. Screening by chest radiograph and lung cancer mortality. *JAMA* 2011;306:1865-73.

54 Sandblom G, Varenhorst E, Löfman O, Rosell J, Carlsson P. Clinical consequences of screening for prostate cancer: 15 years follow-up of a randomised controlled trial in Sweden. *Eur Urol* 2004;46:717-724.

55 Hardcastle JD, Armitage NC, Chamberlain J, Balfour TW, Amar SS. A control trial of faecal occult blood screening for colorectal cancer: 2-year results. *Br J Surg* 1985;72 suppl:S69-71.

56 Tabár L, Akerlund E, Gad A. Five-year experience with single-view mammography randomized controlled screening in Sweden. *Recent Results Cancer Res* 1984;90:105-13.

57 Moss SM, Cuckle H, Evans A, Johns L, Waller M, Bobrow L. Effect of mammographic screening from age 40 years on breast cancer mortality at 10 years' follow-up: a randomised controlled trial. *Lancet* 2006;368:2053-60.

58 Andriole GL, Crawford ED, Grubb RL, Buys SS, Chia D, Church TR, et al. Mortality results from a randomized prostate-cancer screening trial. *N Engl J Med* 2009;360:1310-9.

59 Harris RP, Helfand M, Woolf SH, Lohr KN, Mulrow CD, Teutsch SM, et al. Current methods of the US Preventive Services Task Force: a review of the process. *Am J Prev Med* 2001;20(3 s):21-35.

60 Jørgensen KJ, Klahn A, Gøtzsche PC. Are benefits and harms in mammography screening given equal attention in scientific articles? A cross-sectional study. *BMC Med* 2007;5:12.

Tables

Table 1| Number of trials of cancer screening that quantified cancer mortality, incidence, and harms

	Trials that quantify data for screened and control groups*		Trials that quantify data at least in screened group†	
	Absolute number	Percentage of trials (95% CI)	Absolute number	Percentage of trials (95% CI)
Total No of trials	57	—	57	—
General outcomes:				
Cancer specific mortality	47	82 (70 to 91)	47	82 (70 to 91)
Cancer specific incidence	51	89 (78 to 96)	53	93 (83 to 98)
Harm outcomes:				
Withdrawals because of adverse events	1	2 (0 to 9)	3	5 (1 to 15)
Numerical estimate for overdiagnosis	4	7 (2 to 17)	4	7 (2 to 17)
Numerical estimate for false positive findings	2	4 (0 to 12)	18	32 (20 to 45)
Numerical estimate for negative psychosocial consequences	5	9 (3 to 19)	8	14 (6 to 26)
Numerical estimate for somatic complications	11	19 (10 to 32)	20	35 (23 to 49)
Numerical estimate for invasive procedures	27	47 (34 to 61)	46	81 (68 to 90)
All cause mortality	34	60 (46 to 72)	34	60 (46 to 72)

*198 articles provided data for both intervention and control groups.

†242 articles provided data for at least intervention group (198 for both groups and 44 for screened group alone).

Table 2| **Number of trials and number of articles for different screening technologies. Figures are numbers of trials that provided data for each outcome for both intervention and control group**

| | Breast | | Colon | | Liver | Lung | | | Ovary | Prostate |
	BSE/BCE	Mam	FOBT	Sigm	US/AF	CXR	CCT	Oral VI	US/CA125	PSA/DRE
Total No of trials	5	14	6	7	3	8	5	1	3	5
No of articles with data on all trial arms*	10	61	31	13	4	28	13	4	3	30
General outcomes:										
Cancer specific mortality	3	12	5	6	2	8	3	1	2	5
Cancer specific incidence	5	12	6	7	2	7	4	1	3	4
Harm outcomes:										
Withdrawals because of adverse events	0	0	0	0	0	0	0	0	0	1
Numerical estimate for overdiagnosis	0	2	1	0	0	0	0	0	0	1
Numerical estimate for false positive findings	1	0	1	1	0	0	2	0	0	0
Numerical estimate for negative psychosocial consequences	0	0	1	0	0	2	2	0	0	0
Numerical estimate for somatic complications	0	1	2	1	0	3	2	0	1	1
Numerical estimate for invasive procedures	3	6	3	0	1	7	3	0	1	3
All cause mortality	2	9	4	5	1	5	4	0	1	3
Median (IQR) space devoted to harm (%)	5 (0-7)	7 (2-13)	10 (2-18)	4 (1-16)	14 (7-17)	15 (13-19)	26 (25-30)	3	14(7-20)	17 (1-18)

BSE/BCE=breast self examination/breast clinical examination; Mam=mammography; FOBT=faecal occult blood test; Sigm=sigmoidoscopy or colonoscopy; US/AF=ultrasonography/α fetoprotein; CXR=chest radiography; CCT=chest computed tomography; VI=visual inspection; US/CA125=ultrasonography/cancer antigen 125; PSA=prostate specific antigen; DRE=digital rectal examination; IQR=interquartile range.

*Data for each trial could be provided in one or more publications. References to included articles are presented in appendix 3.

Figure
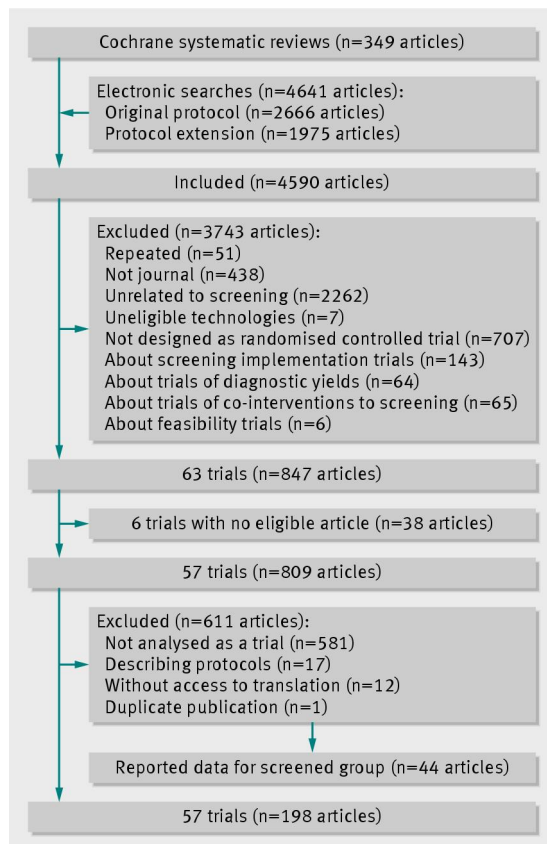
Flow diagram of articles screened, assessed for eligibility, and included in analysis