



Quantification of overdiagnosis in randomised trials of cancer screening: an overview and re-analysis of systematic reviews

Theis Voss^{a,*}, Mikela Krag^{a,1}, Frederik Martiny^{a,b,1}, Bruno Heleno^c, Karsten Juhl Jørgensen^{d,e}, John Brandt Brodersen^{a,f,g}

^a The Centre of General Practice in Copenhagen, Department of Public Health, University of Copenhagen, Øster Farimagsgade 5, Post box 2099, DK-1014 Copenhagen K, Denmark

^b Center for Social Medicine, Bispebjerg and Frederiksberg Hospital, Copenhagen, Denmark

^c CHRC, NOVA Medical School, Faculdade de Ciências Médicas, NMS, FCM, Universidade Nova de Lisboa, Lisbon, Portugal

^d Centre for Evidence-Based Medicine Odense (CEBMO) and Cochrane Denmark, Department of Clinical Research, University of Southern Denmark, JB Winsløvsvej 9b, 3rd Floor, 5000 Odense, Denmark

^e Open Patient data Exploratory Network (OPEN), Odense University Hospital, Odense, Denmark

^f The Research Unit for General Practice in Region Zealand, Øster Farimagsgade 5, Post box 2099, DK-1014 Copenhagen K, Denmark

^g Research Unit for General Practice, Department of Community Medicine, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø

ARTICLE INFO

Keywords:
Cancer
Screening
Overdiagnosis

ABSTRACT

The degree of overdiagnosis in common cancer screening trials is uncertain due to inadequate design of trials, varying definition and methods used to estimate overdiagnosis. Therefore, we aimed to quantify the risk of overdiagnosis for the most widely implemented cancer screening programmes and assess the implications of design limitations and biases in cancer screening trials on the estimates of overdiagnosis by conducting an overview and re-analysis of systematic reviews of cancer screening. We searched PubMed and the Cochrane Library from their inception dates to November 29, 2021. Eligible studies included systematic reviews of randomised trials comparing cancer screening interventions to no screening, which reported cancer incidence for both trial arms. We extracted data on study characteristics, cancer incidence and assessed the risk of bias using the Cochrane Collaboration's risk of bias tool. We included 19 trials described in 30 articles for review, reporting results for the following types of screening: mammography for breast cancer, chest X-ray or low-dose CT for lung cancer, alpha-fetoprotein and ultrasound for liver cancer, digital rectal examination, prostate-specific antigen, and transrectal ultrasound for prostate cancer, and CA-125 test and/or ultrasound for ovarian cancer. No trials on screening for melanoma were eligible. Only one trial (5%) had low risk in all bias domains, leading to a post-hoc meta-analysis, excluding trials with high risk of bias in critical domains, finding the extent of overdiagnosis ranged from 17% to 38% across cancer screening programmes. We conclude that there is a significant risk of overdiagnosis in the included randomised trials on cancer screening. We found that trials were generally not designed to estimate overdiagnosis and many trials had high risk of biases that may draw the estimates of overdiagnosis towards the null. In effect, the true extent of overdiagnosis due to cancer screening is likely underestimated.

1. Introduction

Overdiagnosis of cancer is the diagnosis of indolent neoplastic pathology that would never progress to cause symptoms and/or death during an individual's lifetime[1] and it is the most serious harm of cancer screening.[2–4] If cancer is detected, clinicians cannot know

which individuals that are overdiagnosed as it is not possible to know how the cancer would have evolved in the absence of screening. Therefore all patients are offered treatment or routine observation.[5,6] Overdiagnosed individuals are thus needlessly diagnosed, subsequently overtreated, and thereby harmed. For this reason, it is critical to know the extent of overdiagnosis in cancer screening to enable informed

* Corresponding author.

E-mail address: t.voss@sund.ku.dk (T. Voss).

¹ Shared 2nd authorship

decisions about screening, e.g. whether to participate as an individual or whether to provide a given screening programme on a national level such as prostate cancer screening.[7,8].

In theory, the most robust method to estimate overdiagnosis is to use data from RCTs with life-long follow-up of all participants and no contamination of the control group nor the intervention group, i.e. without screening of the two trial arms during and after the end of the study.[5,9] At the end of the active screening phase, an excess of cancers in the screened population is expected because screening should advance the time of diagnosis (lead time).[5] If there was no overdiagnosis, these excess cancers should be compensated over time because they would all progress to cancer that would be detected clinically after the active screening phase. Thus, a persistent excess in the cumulative incidence of cancers in the screened population after a sufficient follow-up time to account for lead time is high-quality evidence of overdiagnosis.[5, 8, 10].

The purpose of this overview and re-analysis of systematic reviews of RCTs of cancer screening was to assess the extent of design limitations and bias in the included RCTs for quantifying overdiagnosis and, if possible, to estimate the probability that cancer detected by screening was overdiagnosed for the most widespread cancer screening programmes. Many, if not all types of cancer screening, might lead to overdiagnosis. To our knowledge, we are the first to compile the evidence for overdiagnosis across screening for different cancers. For this paper, we chose to focus on the most widely offered cancer screening programmes.

2. Methods

This overview and re-analysis of systematic reviews (SR) were based on a protocol published prior to the conduct of the present study.[11].

2.1. Eligibility criteria

Systematic reviews (SR) of randomised trials were eligible if they:

- 1) investigated screening that aims to detect cancer earlier than it would appear clinically
- 2) compared a cancer screening intervention to no screening
- 3) reported cancer incidence for both screened and non-screened participants and the number of screen-detected cancers.
- 4) Were made by the Cochrane Collaboration, i.e. Cochrane reviews, and included randomised controlled trials only. For a detailed description of the reasons to only include Cochrane reviews and one USPSTF review, see supplementary files.

Systematic reviews of randomised trials were excluded if:

- 1) The included trial offered the screening test to the control group at the beginning or immediately at the end of the trial or used active comparators as a control group. When this happens, screening identifies indolent cancers in the control group thus diluting the true estimate of overdiagnosis.
- 2) The screening intervention study aimed to detect cancer precursors (e.g., screening for colorectal cancer or cervical cancer). Such screening technologies may decrease cancer incidence due to a primary preventive effect following screening. This precludes quantification of overdiagnosis using the cumulative incidence method described below and these types of screening were therefore excluded.[1]

We included SR irrespective of the risk of cancer in the study population, i.e. general population as well as high-risk populations. We included SR regardless of the risk of bias. There were no restrictions concerning the date of publication or language. From the eligible SR, we extracted data from each trial. If a trial was reported in multiple SR, we

selected the SR reporting the longest follow-up time to avoid or diminish lead time bias.

2.2. Search strategy

We searched the Cochrane Library of Systematic Reviews (February 2016) using the search terms ‘screening’ and ‘cancer’ in the title, abstract, or keywords. During the process of conducting this overview, we became aware of one non-Cochrane systematic review assessing ovarian cancer, which we decided to include in the overview since there was no Cochrane Review about this type of screening. We extracted the references to cancer screening trials included in reviews. We updated the literature searches of the individual reviews using the name of the trial and/or principal investigators using PubMed (Last search: November 2021).

2.3. Data collection and extraction

We searched the Cochrane Library of Systematic Reviews for relevant SRs. Identified studies were compiled in Endnote.[12] Here, two reviewers independently screened all titles against the eligible criteria. Disagreements were handled via discussion, potentially involving a third review author when disagreements could not be resolved. If the title or abstract did not provide sufficient information to determine eligibility, we assessed the trial on full-text level. For the included SR, and all relevant RCTs in the included SR, two authors independently looked through the reference list to identify potentially relevant studies that the search strategy had failed to identify.

Two authors independently extracted data from the included trials and entered them into a piloted data extraction form in Excel.[13] Disagreements were resolved through discussion with a third review author. The data extraction strategy is described in the protocol and included in the Supplementary Table A1.

2.4. Assessment of the risk of bias in included trials

We extracted the risk of bias assessments from the included Cochrane Systematic Reviews. They used the Cochrane Risk of Bias Tool version 1.0[14] that include the following six domains:

1. Selection bias: random sequence generation and allocation concealment
2. Performance bias: blinding of participants and personnel (not extracted)
3. Detection bias: blinding of outcome assessment
4. Attrition bias: Incomplete outcome data
5. Reporting bias: Selective reporting of outcome
6. Other possible sources of bias

We re-assessed the risk of bias when our updated search identified relevant articles from studies included in the SR that were published after the review. Trials that were not identified via Cochrane reviews, but through the USPSTF review on ovarian cancer screening, or via our updated searches for primary research articles, were independently assessed for risk of bias by two authors (TV and FM) using the Cochrane risk of bias tool 1.0.

Performance bias was not assessed. First, it is not possible to blind participants and clinicians for screening status at the time of diagnosis. Second, we judged that co-interventions would not affect the incidence of cancer and thus did not bias the overdiagnosis estimates. Thirdly, blinded cause of death assessment is a possible and important risk of bias in some screening trials, but it is not relevant to overdiagnosis.[9].

We assessed two additional biases that can affect the estimate of overdiagnosis (Table 1):

Table 1
Probable impact from types of biases on estimates of overdiagnosis.

Known impact	Bias direction on the overdiagnosis estimate
Contamination (screening of control group during or after the active screening phase)	Bias towards underestimation[15–17]
Lead time (follow-up after end of intervention)	Bias towards overestimation[15,17]
Possible impact	
Randomisation (random sequence generation)	Uneven distribution of “cancer risk” between the intervention group and the control group may bias in either direction.
Allocation concealment	As for randomisation bias.
Attrition (incomplete outcome data)	Could bias in either direction
Contamination of the screened group during or after the active screening phase	Bias towards overestimation
Reporting bias	Selective reporting might impact overdiagnosis estimates

Table 1 How bias might impact overdiagnosis estimates specifically.

1. Contamination of the control group after randomisation.[15] Contamination was defined as the reported amount of participants in the control group, who were exposed to the same screening technology as the screened group. We used the Cochrane review by Ilic et al. as a benchmark for our evaluation of the risk of bias from contamination (Supplementary Table B1).[16]
2. Inadequate consideration of lead time (too short post-intervention follow-up or screening offered to the control group at the end of the trial).[15] Here, we used studies on the natural growth rate for each type of cancer to determine mean lead time and this to determine threshold for bias assessment (Supplementary Table C1).

Other factors that influence estimates of overdiagnosis.

1. Different cancer risk at baseline between intervention and control groups (equal to selection bias included in the Cochrane’s Risk of Bias tool)
2. Participation rate over screening rounds. Participation was not considered a bias for the purpose of estimating overdiagnosis but a component of screening (Supplementary Table B1).
3. Number of, and the interval between, screening rounds
4. Continued screening, i.e. if participants continued receiving the offered screening modality on their own initiative after the end of screening.

For detailed descriptions of how we addressed the potential effect of the above-mentioned biases and study factors on the estimates of overdiagnosis, see [Supplementary Table B1](#).

2.5. Data management and statistical analysis

We defined overdiagnosis as the percentage of screen-detected cancers that were overdiagnosed following the definition from the UK Independent Panel review of breast cancer[18]. This definition estimates an individual’s risk of being overdiagnosed when diagnosed with cancer due to participating in screening.

When overdiagnosis is estimated as the difference in cumulative incidence between the two arms in a randomised trial, it includes any overdiagnosis in the initial screening round as well as subsequent screening rounds.[8] This measure of overdiagnosis can be interpreted as the average probability for all screening rounds that a screen-detected cancer is overdiagnosed. The precision of this average will depend on the proximity between the trial and real-life screening in terms of the number of rounds offered and screening intervals. To facilitate comparison across studies with the same target cancer and similar screening

modalities, we used a standard measure of overdiagnosis, [5]:

$$\left(\frac{\frac{\text{Cumulative cancer incidence screened population}}{N_{\text{screened}}} - \frac{\text{Cumulative cancer incidence control population}}{N_{\text{control}}}}{\frac{\text{Cumulative number of screen-detected cancers}}{N_{\text{screened}}}} \right) \times 100$$

The cumulative cancer incidence in the screened population was defined as all cancers detected in the population offered screening during and after the active phase.

The cumulative cancer incidence in the control population was defined as all cancers detected in the control population during and after the active phase.

The cumulative number of screen-detected cancers was defined as all cancers detected by screening in the population offered screening during the active phase.

We calculated standard deviations through bootstrapping using R [19] and used a normal approximation to compute 95% confidence intervals using Review Manager version 5.3.[20].

Heterogeneity was assessed using the statistical heterogeneity indicator I^2 to determine if it was reasonable to pool trials with similar screening modalities for the same target cancer, e.g. trials on low-dose CT (LDCT) scans for lung cancer screening. Causes of heterogeneity for trials assessing overdiagnosis for each type of cancer screening was assessed qualitatively according to key study characteristics.

We pooled overdiagnosis estimates for RCTs that assessed the same target cancer using the same screening technology. Results were summarised with a random-effects meta-analysis using the inverse-variance method, as we anticipated some variation due to the different timing, populations, and setting of individual trials. Data were analysed using Review Manager version 5.3.[20].

The planned primary meta-analysis was restricted to trials with a low risk of bias across all bias domains. However, only one included trial fulfilled this criterion. Therefore, we performed the following two post-hoc meta-analyses: One, estimating overdiagnosis using results from the most reliable trials, i.e. excluding trials with a high risk of bias for domains of particular relevance to overdiagnosis: random sequence generation, allocation concealment, contamination, and lead time. Two, estimating overdiagnosis using results from all included trials, regardless of their bias profile.

To investigate the impact of bias in the overdiagnosis estimates, we also performed the following sensitivity analyses:

- Cluster randomised versus individually randomised trials
- Excluding trials with a high risk of bias either due to poor allocation concealment and/or poor random sequence generation.
- Excluding trials with a high risk of contamination bias
- Excluding trials with a high risk of lead time bias (post-hoc)

3. Results

3.1. Study selection

We included results from 19 trials encompassed in 15 systematic reviews (SR) reported in 30 relevant articles. The trials investigated screening for five types of cancer using seven different screening technologies. The search was performed in 2016 and updated two times. In total, we identified 2694 articles, from which we included 19 trials (Table 2, Fig. 1 and Fig. 2).

A recent Cochrane review on screening for melanoma showed that no randomised trials have been completed[21], so this review was excluded.

3.2. Study characteristics

Across the 19 trials included for review, the smallest trial had 3206 participants (ITALUNG [22]), the largest trial had 202,546 participants

Table 2
Study characteristics.

Target cancer	Screening technology	Number of trials included
Breast cancer	Mammography (with/without clinical breast examination)	4
Lung cancer	Chest X-ray	1
Lung cancer	Low-Dose Computed Tomography	5
Liver cancer	Ultrasonography and alpha-foetoprotein	1
Prostate cancer	Biomarker Prostate specific antigen (with/without digital rectal examination and transrectal ultrasonography)	4
Ovarian cancer	Biomarker CA-125 (with/without transvaginal ultrasonography)	3
Ovarian cancer	Transvaginal ultrasonography	1

Table 2 Number of included trials sorted by target cancer and screening technology

(UKCTOCS[23]), and the median across trials was 26,602 participants (Stockholm [24]) (Table 2 and Table 3). Four trials investigated breast cancer; Six trials investigated lung cancer; One trial investigated liver cancer; Four trials investigated prostate cancer; Four trials investigated ovarian cancer (Table 2, Fig. 2).

The number of screening rounds across trials varied from one to 10, intervals between screening rounds varied from six months to 48 months, and the length of follow-up after the last screen varied from 12 months to more than 17 years of follow-up.

3.3. Risk of bias in the included studies

The risk of bias varied considerably between trials (Fig. 2). Only one trial (Malmö 1976) had a low risk of bias across all domains.

The two most significant types of bias for overdiagnosis are contamination and lead time bias. For contamination, two of the 19 trials (11%) had a high risk of bias: both reported annual opportunistic screening rates above 30% in the control group.[44,49] Twelve of 19 trials (63%) had unclear risk of contamination bias: eight provided no data on contamination[22, 24–27, 30, 31, 38, 39, 43, 45, 51, 53]; one reported an annual contamination rate of 11%[35]; two reported low quality data (survey of off-protocol screening three years after the last trial round and a response rate of 38%)[23]; and one trial reported the implementation of a national screening programme shortly after the end of the trial.[32–34] Five trials (26%) had a low risk of bias from contamination.[29, 35, 37, 40–42] For lead time bias, two trials (11%) had a high risk of bias: in both, follow-up was too short to account for lead time.[32–34, 45, 53] Six trials (32%) had an unclear risk of lead time bias.[37–41] Eleven trials (58%) had an adequate follow-up time to account for lead time, i.e. low risk of lead time bias (Fig. 2, Supplementary Table C1).

The bias domain “other” included various biases not pertaining to the standard bias categories. One trial had unexplained data discrepancies among different publications of the trial.[43] Three trials used an unreliable measure of cancer incidence (self-report survey completed by participants).[35, 49, 51].

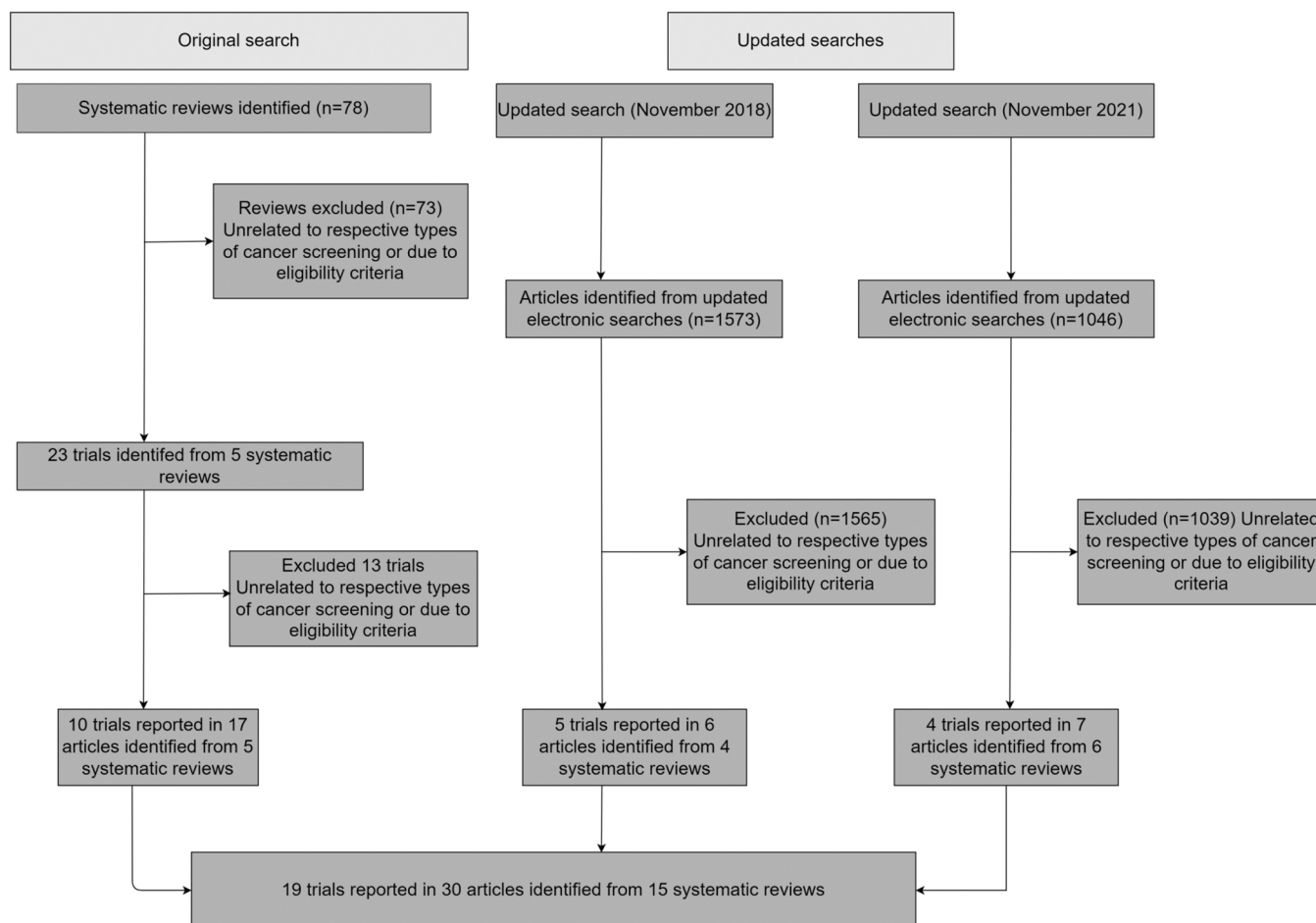


Fig. 1. Flow chart detailing the literature search and study selection.

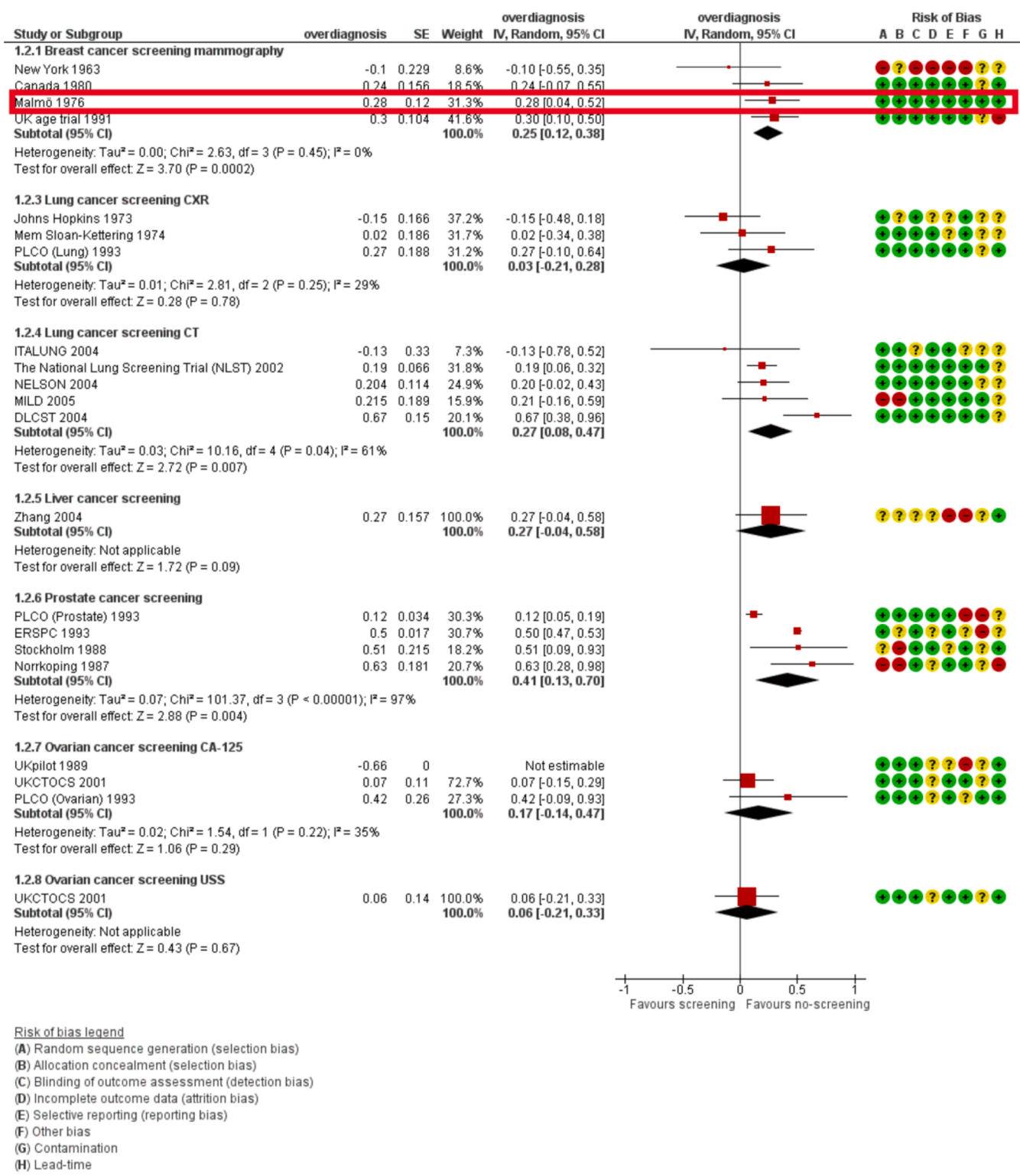


Fig. 2. Forest plot detailing all included trials. The red box highlights the only included trial with low risk of bias across all bias domains (Malmö 1976).

3.4. Estimates of overdiagnosis in the included studies

Across all trials and all types of cancer screening programmes, estimates of overdiagnosis ranged from -66-67%. In trials of breast cancer screening with mammography, estimates ranged from -10-30%; in lung cancer with LDCT overdiagnosis ranged from -13-67%; in prostate cancer from 12% to 63%; in ovarian cancer with CA-125 from

-66-42%. Only one trial on liver cancer screening and one on lung cancer screening with CXR were included and both found that 27% of screen-detected lung or liver cancers were overdiagnosed, respectively (Table 4 and Fig. 2).

Table 3
Incidence of cancer.

Trial name or location, and year	Group size (Control/intervention)	Control group cumulative incidence	Intervention group cumulative incidence (at end of follow-up)	N screen-detected	Estimation of over-diagnosis (Overdiagnosis/1000 screen-detected cancers)
Breast cancer screening with mammography					
Canada 1980 [25–27]	44,910/ 44,925	3133	3250	484	240/1000
Malmö 1976 cohort 1 and 2 * [28,29]	13,133/ 13,107	698	780	298	280/1000
New York 1983 [30,31]	31,092/ 31,092	439	426	132	- 98/1000
UK age trial 1991 # [32–34]	53,883/ 106,953	821	482	229	299/1000
Lung cancer screening with CXR					
PLCO (lung) 1993 [35,36]	77,453/ 77,444	1719	1801	308	267/1000
Lung cancer screening with LDCT					
ITALUNG 2004 [22]	1593/ 1613	71	67	38	-129/1000
DLCST 2004 [37]	2052/ 2052	53	96	64	672/1000
NELSON [38,39]	6612/ 6583	304	344	203	203/1000
MILD [40,41]	1723/ 2376	60	98	71	215/1000
LUSI [42]	2023/ 2029	74	90	63	250/1000
Liver cancer screening					
Zhang 2004 [43]	9443/ 9373	67	86	71	275/1000
Prostate cancer screening					
ERSPC 1993 [44]	89,352/ 72,891	7732	8444	4957	431/1000
Norrköping 1987 [45,46]	7532/ 1494	292	85	43	630/1000
PLCO (prostate) 1993 [47–49]	38,345/ 38,340	5287	5574	2417	119/1000
Stockholm 1988 [24]	24,202/ 2400	1972	229	65	515/1000
Ovarian cancer screening					
PLCO (ovarian) 1993 [35,50]	34,305/ 34,252	231	265	82	419/1000
UKCTOCS 2001 (MMS) [23]	101,299/ 50,624	1016	522	212	67/1000
UKCTOCS 2001 (USS) [23]	101,299/ 50,623	1016	517	164	56/1000
UK pilot 1989 [51]	10,977/ 10,958	20	16	6	-661/1000

PLCO=Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial

NLST=National Lung Screening Trial

ERSPC=European randomised Study of Screening for Prostate Cancer

ITALUNG=Italian Lung Cancer Trial

DLCST=Danish Lung Cancer Screening Trial

NELSON= Dutch-Belgian Lung Cancer Screening Trial (Nederlands-Leuvens Longkanker Screenings Onderzoek)

MILD= Multicentric Italian Lung Detection Trial

UKCTOCS=United Kingdom Collaborative Trial of Ovarian Cancer Screening

MMS=Multimodal screening

USS=Ultrasonography screening

*Cohort 3 was excluded from this analysis.

#Two articles [34,52] reported more recent cumulative cancer incidences but not more recent number of screen-detected cases.

Table 3 All included trials, sorted by target cancer and screening technology, showing the number of included participants in each group, cancer incidence in each group, incidence of screen-detected (target cancer diagnosed by the intended screening technology).

3.5. Synthesis of results

One trial (5%) had low risk of bias in all bias domains. In our primary meta-analysis, we estimated that 28% (95% CI 4–52%) of screen-detected breast cancers were overdiagnosed using data from the Malmö trial of breast cancer screening. This trial had a three-percentage point higher rate of overdiagnosis compared to the meta-analysis based

on all included trials (Table 4, Fig. 2, Supplementary Figure A1). [28, 29].

Our post hoc meta-analysis of the most reliable trials, i.e. excluding trials with a high risk of bias for the domains random sequence generation, allocation concealment, contamination, or lead time included data from 12 trials reporting results for six types of cancer screening. On average 27% (95% CI 8–45%) of mammography-detected breast cancers

Table 4
Estimates of overdiagnosis.

Target cancer (screening technology)	All trials, regardless of risk of bias (95% CI)	Trials without high risk of bias in critical domains (95% CI)	Unclear/low risk of Selection bias (95% CI)	Unclear/low risk of Contamination (95% CI)	Unclear/low risk of Lead time (95% CI)
N of trials	19	12	15	17	17
Breast cancer (mammography)	25% (12–38%)	27% (8–45%)	28% (14–42%)	25% (12–38%)	21% (2–39%)
Lung cancer (CXR)	27% (–10% to 64%)	27% (–10% to 64%)	27% (–10% to 64%)	27% (–10% to 64%)	27% (–10% to 64%)
Lung cancer (LDCT)	29% (7–52%)	31% (2–59%)	31% (2–59%)	29% (7–52%)	29% (7–52%)
Liver cancer (ultrasonography and AFP)	27% (–4% to 58%)	27% (–4% to 58%)	27% (–4% to 58%)	27% (–4% to 58%)	27% (–4% to 58%)
Prostate cancer (PSA)	38% (15–62%)	N/A	28% (–3% to 58%)	58% (31–85%)	32% (6–59%)
Ovarian cancer (CA-125)	17% (–14% to 48%)	17% (–14% to 48%)	17% (–14% to 48%)	17% (–14% to 48%)	17% (–14% to 48%)
Ovarian cancer (USS)	6% (–27% to 39%)	6% (–27% to 39%)	6% (–27% to 39%)	6% (–27% to 39%)	6% (–27% to 39%)
GRADE assessment	Very low	Moderate	Low	Low	Low

Table 4 The estimates of overdiagnosis for each target cancer regardless of risk of bias, when only including trials with the lowest risk of bias relevant for estimating overdiagnosis (the post-hoc meta-analysis). Furthermore, when excluding the trials with high risk of selection bias, contamination bias, or lead time bias.

and 30% (95% CI 2–59%) of LDCT-detected lung cancers were overdiagnosed. For the other four types of screening, results were not significant. We estimated that on average 27% (95% CI –10 to 64%) of CXR-detected lung cancers, 27% (95% CI –4% to 58%) of screen-detected liver cancers, and 17% (95% CI –14% to 48%) of CA-125-detected ovarian cancers are overdiagnosed. For ovarian cancer screening, screening with ultrasound only, findings were not significant concerning the rate of overdiagnosis with an overdiagnosis estimate of 6% (95% CI –21% to 33%) (Table 4, Fig. 2, Supplementary Figure B1).

Meta-analysis of all included trials in the overview, regardless of risk of bias, showed that on average, 25% (95% CI 12–38%) of mammography-detected breast cancers, 27% (95% CI –10% to 64%) of CXR-detected lung cancers, 29% (95% CI 7–52%) of LDCT-detected lung cancers, 27% (95% CI –4% to 58%) of US-detected liver cancers, 38% (95% CI 14–62%) of PSA-detected prostate cancers, 17% (95% CI –14% to 48%) of CA-125-detected ovarian cancers, and 6% (95% CI –27% to 39%) of US-detected ovarian cancers were overdiagnosed (Fig. 2).

It was not possible to assess the effect on the overdiagnosis estimates from cluster randomisation, as only one trial utilised cluster randomisation.[43].

Four out of 19 RCTs (21%) were at high risk of bias due to either inadequate random sequence generation or allocation concealment.[24, 30, 31, 40, 41, 45, 53] One of the remaining 15 RCTs had unclear risk of bias for these domains.[43] From these 15 RCTs with the lowest risk of bias from random sequence generation and allocation concealment estimates of overdiagnosis were higher in screen-detected breast cancer and screen-detected lung cancer using LDCT when compared to estimates from all trials, regardless of risk of bias in other bias domains. The estimates were lower in screen-detected prostate cancer, and the same in screen-detected lung cancer using CXR, liver cancer, and ovarian cancer (Table 4, Fig. 2, Supplementary Figure C1).

Seventeen out of 19 RCTs had low or unclear risk of contamination bias. Estimates of overdiagnosis from these 17 RCTs were higher in screen-detected prostate cancer compared to all trials, regardless of risk of bias in other bias domains. Estimates of overdiagnosis in screen-detected breast, lung, liver, and ovarian cancer were the same. (Table 4, Fig. 2, Supplementary Figure D1).

Seventeen out of 19 RCTs had low or unclear risk of lead time bias. Estimates of overdiagnosis from these 17 RCTs were lower in screen-detected breast cancers and prostate cancers. Estimates were the same in screen-detected lung cancer, liver cancer, and ovarian cancer (Table 4, Fig. 2, Supplementary Figure E1).

4. Discussion

4.1. Summary of main results

Only one out of 19 RCTs (5%) on cancer screening had low risk of bias across all seven bias domains for estimating overdiagnosis. In our post-hoc meta-analysis of the most reliable trials, i.e., excluding trials with high risk of bias from random sequence generation, allocation concealment, contamination, or lead time, we found that 27% (95% CI 8–45%) of mammography-detected breast cancers; 31% (95% CI 2–59%) of LDCT-detected lung cancers; 27% (95% CI –4% to 58%) of screen-detected liver cancers, and 17% (95% CI –14% to 48%) of CA-125-detected ovarian cancers, were overdiagnosed.

Many trials were at risk of bias due to poor randomisation, contamination of the control group, or inadequate consideration of lead time, i.e. insufficient follow-up time to account for slow-growing cancers. Confidence in the estimates of overdiagnosis was further downgraded due to imprecision of the pooled estimate and due to inconsistency (heterogeneity) between trials (Fig. 2, Supplementary Table A1).

Two of the 19 RCTs reported that screening reduced the cumulative incidence of cancer. This cannot be correct, since these trials assessed screening tests that detect invasive cancer and not precursor lesions which, if treated, could reduce the incidence of cancer. One of these trials were at high risk of selection bias and had unclear allocation concealment (Fig. 2, Supplementary Table D1). [30] The other trial had a negative point estimate that was close to zero.[22] However, there is no apparent bias that explains why this trial found a reduced incidence of lung cancer but is presumably due to a combination of random chance and bias towards the null.

4.2. Strengths and weaknesses

We would like to emphasize the following strengths of our overview: First, the overview included trials from Cochrane Systematic Reviews, which are acknowledged for their exhaustive literature searches and structured assessment of risk of bias, and one USPSTF systematic review, also having high methodological standards.[54] Our search strategy is updated and we screened the reference list of included trials, both increasing the chances that we present a comprehensive and up-to-date overview. We had two authors independently assessing biases relevant for overdiagnosis which are not included in the Cochrane Risk of Bias Tool v. 1.0: contamination of the control population and inadequate follow-up after the active phase (lead time bias).

Two types of limitations threaten the validity of our findings,

limitations regarding the evidence at hand and methodological choices on the conduct of our overview. First, the following limitations regarding the evidence at hand are worth mentioning: First, the design, methodological quality, and reporting from the included studies was not optimal to estimate overdiagnosis. The included cancer screening trials were primarily designed to assess cancer mortality, not overdiagnosis. Some design choices limited the ability of the trials to provide trustworthy estimates of overdiagnosis, e.g., conducting baseline screens before enrolment, offering an alternative screening method, or screening all participants at the end of trial. We observed a large statistical heterogeneity in many of the meta-analyses, which is likely both due to between-study methodological differences, i.e. biases and uncertainty about outcome measurement, and clinical diversity, i.e., differences across studies concerning study participants, screening technologies, intervals, rounds etc. This heterogeneity between studies in meta-analyses should be considered when interpreting the results.

Second, some external circumstances, such as the implementation of national screening programmes shortly after the trial completion, led to contamination of the control group, i.e. a control group that cannot be considered “unscreened”. For example, many countries have introduced screening programmes with mammography. Because we had no knowledge of the contamination due to a nationwide screening programme, the trials on screening for breast cancer with mammography were assessed based on the reported data. Therefore, bias assessments on contamination in the breast cancer screening trials comes with reservations, i.e. the true extent of contamination is likely higher, biasing the estimate of overdiagnosis towards the null. The quantification of overdiagnosis requires an assessment of the cumulative incidence of cancer several years after the last screening round to account for lead time. Yet, a single round of screening in the control group has an immediate effect on cancer incidence, greatly reducing the contrast between screened and control groups. Of the trials which had a more straightforward parallel design, many were at high risk of bias due to poor sequence generation or allocation concealment procedures.

Confidence in our results is further threatened by four methodological choices. First, we relied on the original bias assessments in the Cochrane Reviews, which were carried out by different teams of systematic reviewers with potential for inter-rater variability in bias assessments. However, bias assessments in Cochrane reviews follow stringent criteria for assessment expected to reduce inter-rater variability. Also, performing the bias assessments ourselves would also be subject to inter-rater variability and is against Cochrane recommendations.[55].

Second, we chose to include RCTs from Cochrane SRs and updated the literature searches of the included reviews using the name of the RCT and/or principal investigators. This methodological choice had the drawback that new trials and SRs by new authors would not be identified in our updated search. However, the Cochrane SRs are generally acknowledged for their exhaustive work making it likely that we have included most relevant trials in the area. Furthermore, we updated the search to include the most recent data and eligible new trials, as previously described.

Third, the possible effect of contamination is uncertain. In our study, we have defined contamination as the provision of screening intervention(s) to the control group in any way (e.g. opportunistic use of LDCT for patients in the control group). We chose to use the article by Ilic et al. on prostate cancer screening with PSA as a benchmark for our evaluation of risk of bias from contamination.[16] However, we were unable to obtain information from Ilic and colleagues regarding specific criteria used to judge the degree of contamination bias. Thus, to assess contamination we chose to set a somewhat arbitrary threshold for high risk of bias when 30% or more of the control group received screening of any given intensity (Supplementary Table B1). We rated trials as unclear risk of bias if there were no information on contamination.

Fourth, concerning lead time bias, we exclusively assumed that a trial had a low risk of lead time bias if the follow-up time was longer than

the highest model-based estimate of mean lead time (Supplementary Table B1 and Supplementary Table C1). This is a conservative choice since statistical models tend to include overdiagnosed cases when calculating lead time and for some individual cancers, lead time may be much longer than the mean.[56] The latter could be accommodated by using maximum lead time instead of the mean lead time, thereby minimising the bias from lead time toward overestimation.[57,58] However, that would require much longer follow-up and likely increase bias from attrition and contamination, which both tend to increase over time. For ovarian cancer, we were unable to find any model-based estimates of mean lead time for screening for ovarian cancer with US and thus chose to apply lead time from CA-125. We acknowledge that this is a simplification of the heterogeneity of the continuum of cancer.[7] Notwithstanding the simplification, it has the merit of allowing more transparent bias judgements for each included study. Scientific discussion of the underlying assumptions of these criteria can serve to make more accurate judgements of biases affecting estimates of overdiagnosis in the future.

Another concern is whether studies were similar enough to justify meta-analyses, i.e., combining the results from trials with similar screening modalities for the same target cancer within each screening programme. The same type of cancer screening might lead to vastly different results in different settings due to differences in the baseline risk of cancer, how the screening programme is implemented, e.g. number of screening rounds, the interval between rounds, varying participation rates, and more. In some cases, it might be argued that both the circumstances around the screening programme and the implementation of the screening programme itself are so heterogeneous, that it is not appropriate to combine trials in meta-analyses.[59] However, it is common practice to combine mortality estimates across trials in different settings in SRs. Therefore, we judged that it is also justified to pool data from similar trials in terms of target cancer and screening modalities to estimate overdiagnosis.

As outlined in the methods section, overdiagnosis estimates are especially affected by the two bias domains 1) contamination and 2) lead time but also by contextual factors such as different baseline risk of cancer, participation rate, the number of screening rounds, and the interval between them. We estimated the risk of overdiagnosis across RCTs targeting the same type of cancer with similar screening technology but with potentially widely varying participation rates, number of screening rounds, and intervals between rounds. Because we could not account for these factors, they might confound our sensitivity analyses on the potential association between contamination bias and lead time and the estimates of overdiagnosis. These five factors, therefore, should be considered when evaluating overdiagnosis estimates in cancer screening, ideally in RCTs designed to measure overdiagnosis or via monitoring data. In this overview, we chose to test the effect on overdiagnosis estimates from contamination and lead time.

4.3. Comparison of findings to similar studies

Cochrane systematic reviews of cancer screening do not always quantify overdiagnosis.[60] Indeed, it is not always possible given the available trials, and for some types of cancer screening, overdiagnosis of cancers is less of a concern. However, several of the included reviews discuss overdiagnosis as significant harm of screening and report estimates of overdiagnosis derived from individual primary studies,[61,62] and the Cochrane review on mammography screening found about 30% overdiagnosis.[9] In comparison, we exclude trials where the control group was screened during or at the end of the active phase and estimate overdiagnosis in individual trials. For example, the NLST trial was excluded because the trial compared screening with LDCT to screening with CXR (active comparator), making it impossible to reliably estimate the degree of overdiagnosis. The USPSTF considers overdiagnosis in its balance of benefits and harms of screening and reports estimates of overdiagnosis extracted from primary studies in its recommendation

statements.[63–65] This overview, however, adds individual overdiagnosis estimates in cancer screening and, by adding bias assessments particularly relevant for estimating overdiagnosis, emphasize the lack of high quality evidence especially in regards to the harms of screening.

4.4. Implications for future research

To facilitate more trustworthy and accurate estimates of overdiagnosis in cancer screening, future methodological research is needed to investigate the influence of contamination as currently there is no evidence supporting a specific threshold and as any effect is likely to be a continuum. Likewise, identifying valid and precise lead time estimates for different types of cancer screening is warranted. Additionally, research on the influence of different baseline risks of cancer between the intervention and the control groups, number of screening rounds, interval between screening rounds, and participation rate is needed.

Furthermore, we suggest that future trials of cancer screening aimed at detecting invasive cancers at a localised stage should be designed to estimate overdiagnosis, i.e. they should follow participants for enough time after the active phase to account for the lead time while assessing the extent of the contamination. Finally, they should report outcomes such as cancer incidence, participation rates, average follow-up time since the last screening round, the time between screening rounds, number of screening rounds, and contamination, including diagnostic use of the investigated screening technology after the trials if relevant. However, that is not always possible because of early reporting of results from a given trial, underlining the importance of reporting results after a sufficiently long follow-up.

During review of full-text articles in our update of the search strategy, we had to exclude many trials due to inadequate reporting of cancer incidence rates to allow estimation of overdiagnosis, e.g. number of screen-detected cancers not reported[52], or no cancer incidence number for the control group reported [66]. This points to the poor reporting of harms in general, although such guidance is available.[2, 67, 68]. Overall, further research is warranted even in screening technologies already implemented or currently being considered for implementation in many countries.

4.5. Implications for practice

Overdiagnosis is the most serious harm of cancer screening. Yet, we found that many trials of screening for various types of cancer were not adequately designed to estimate its extent. Many screening programmes have been implemented following preliminary beneficial results. However, the harms of screening, like overdiagnosis, takes many years to be adequately estimated. This overview highlights the need for continued evaluation (such as by the USPSTF) of both current and future cancer screening programmes, to consider any potential harms that might necessitate modifications or even discontinuation of a screening programme.

5. Conclusion

RCTs are the most reliable design to quantify overdiagnosis if they are designed to do so; however, our overview shows that confidence in the estimates of overdiagnosis in cancer screening RCTs is moderate to very low. We found that 9 of 19 (47%) included RCTs had a high risk of bias and most trials had high or unclear risk of bias in multiple bias domains, which theoretically act to bias the estimate of overdiagnosis. Estimates of overdiagnosis were often inconsistent and comes with reservations (e.g., old trials, risk of bias, issues with trial designs, and apparently unexplained heterogeneity between trials) but they are as good as it currently gets. Two screening technologies (lung cancer with LDCT and breast cancer with mammography) showed significant overdiagnosis of 30% and 27%, respectively. Furthermore, in screening for prostate cancer with PSA, the estimate suggests 38% of screen-detected

prostate cancers were overdiagnosed even with multiple high risks of bias in the included RCTs, biasing towards underestimation. For the ovarian cancer screening programmes, our best estimates are that 17% of the CA-125-screened ovarian cancers and 6% of transvaginal ultrasound-screened ovarian cancers may be overdiagnosed.

Funding

None of the authors had financial support for the submitted work.

Ethical approval

Not required.

Conflict of Interest Statement

Competing interests: All authors have completed the ICMJE uniform disclosure form at <http://www.icmje.org/conflicts-of-interest/> and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work. One of the authors was also an author on the Cochrane review of breast screening.

Transparency

The lead author affirms that this manuscript is an honest, accurate and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from this study as planned have been explained.

CRediT authorship contribution statement

TV and FM drafted the protocol, and BH and JB provided comments. TV and FM assessed references for eligibility and extracted and analysed the data. TV drafted the manuscript and MAK, BH, KJJ, FM and JB contributed to revisions with important intellectual content. All authors had full access to all data (including statistical reports and tables) in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. TV is guarantor.

Data Availability

Files with extracted data and analyses are available from the authors.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.canep.2023.102352](https://doi.org/10.1016/j.canep.2023.102352).

References

- [1] J. Brodersen, L.M. Schwartz, S. Woloshin, Overdiagnosis: how cancer screening can turn indolent pathology into illness, *Apmis* 122 (8) (2014) 683–689.
- [2] B. Heleno, M.F. Thomsen, D.S. Rodrigues, K.J. Jørgensen, J. Brodersen, Quantification of harms in cancer screening trials: literature review, *Bmj* 347 (2013) f5334.
- [3] K.J. Jørgensen, P.-H. Zahl, P.C. Gotzsche, Overdiagnosis in organised mammography screening in Denmark. A comparative study, *BMC Women's Health* 9 (1) (2009) 36.
- [4] Screening for Breast Cancer: U.S. Preventive Services Task Force Recommendation Statement, *Ann. Intern. Med.* 164 (4) (2016) 279–296.
- [5] H.G. Welch, W.C. Black, Overdiagnosis in cancer, *J. Natl. Cancer Inst.* 102 (9) (2010) 605–613.
- [6] H.G. Welch, S. Woloshin, L.M. Schwartz, L. Gordis, P.C. Gotzsche, R. Harris, et al., Overstating the evidence for lung cancer screening: the International Early Lung Cancer Action Program (I-ELCAP) study, *Arch. Intern Med* 167 (21) (2007) 2289–2295.

- [7] J.L. Carter, R.J. Coletti, R.P. Harris, Quantifying and monitoring overdiagnosis in cancer screening: a systematic review of methods, *Bmj* 350 (2015) g7773.
- [8] C. Biesheuvel, A. Barratt, K. Howard, N. Houssami, L. Irwig, Effects of study methods and biases on estimates of invasive breast cancer overdiagnosis with mammography screening: a systematic review, *Lancet Oncol.* 8 (12) (2007) 1129–1138.
- [9] P.C. Gotzsche, K.J. Jørgensen, Screening for breast cancer with mammography, *Cochrane Database Syst. Rev.* 2013 (6) (2013), Cd001877.
- [10] R. de Gelder, E.A. Heijnsdijk, N.T. van Ravesteyn, J. Fracheboud, G. Draisma, H. J. de Koning, Interpreting overdiagnosis estimates in population-based mammography screening, *Epidemiol. Rev.* 33 (1) (2011) 111–121.
- [11] T. Voss MK, B. Heleno, K.J. Jørgensen, F. Martiny, J. Brodersen. Quantification of overdiagnosis in randomised trials of cancer screening: an overview and re-analysis of systematic reviews. [crdyorkcak/prospero/](https://doi.org/10.1093/cdyork/cak017). 2019.
- [12] Team TE Endnote. EndNote X9 ed. Philadelphia, PA: Clarivate; 2013.
- [13] Corporation M.. Microsoft Excel. Microsoft Exel, Microsoft 365. 2018 ed: Microsoft Corporation; 2018.
- [14] Higgins J.P.T. AD, Sterne JAC. Assessing risk of bias in included studies: The Cochrane Collaboration; 2011 [updated updated March 2011. 5.1.0 [Available from: www.cochrane-handbook.org.
- [15] J. Brodersen, How to conduct research on overdiagnosis. A keynote paper from the EGPRN May 2016, Tel Aviv, Eur. J. Gen. Pr. 23 (1) (2017) 78–82.
- [16] D. Ilic, M. Djulbegovic, J.H. Jung, E.C. Hwang, Q. Zhou, A. Cleves, et al., Prostate cancer screening with prostate-specific antigen (PSA) test: a systematic review and meta-analysis, *BMJ* 362 (2018) k3519.
- [17] J. Brodersen, Overdiagnosis: An Unrecognised and Growing Worldwide Problem in Healthcare, *Zdr. Varst.* 56 (3) (2017) 147–149.
- [18] M.G. Marmot, D.G. Altman, D.A. Cameron, J.A. Dewar, S.G. Thompson, M. Wilcox, The benefits and harms of breast cancer screening: an independent review, *Br. J. Cancer* 108 (11) (2013) 2205–2240.
- [19] Team RC. R: A Language and Environment for Statistical Computing. 2013.
- [20] Review Manager (RevMan) 5.3 ed. Copenhagen: The Nordic Cochrane Centre. The Cochrane Collaboration; 2014.
- [21] M. Johansson, J. Brodersen, P.C. Gotzsche, K.J. Jørgensen, Screening for reducing morbidity and mortality in malignant melanoma, *Cochrane Database Syst. Rev.* 6 (6) (2019), Cd012352.
- [22] E. Paci, D. Puliti, A. Lopes Pegna, L. Carrozzi, G. Picozzi, F. Falaschi, et al., Mortality, survival and incidence rates in the ITALUNG randomised lung cancer screening trial, *Thorax* 72 (9) (2017) 825–831.
- [23] U. Menon, A. Gentry-Maharaj, M. Burnell, N. Singh, A. Ryan, C. Karpinskyj, et al., Ovarian cancer population screening and mortality after long-term follow-up in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial, *Lancet* 397 (10290) (2021) 2182–2193.
- [24] A. Kjellman, O. Akre, U. Norming, M. Törnblom, O. Gustafsson, 15-year followup of a population based prostate cancer screening study, *J. Urol.* 181 (4) (2009) 1615–1621.
- [25] A.B. Miller, T. To, C.J. Baines, C. Wall, Canadian National Breast Screening Study-2: 13-year results of a randomized trial in women aged 50–59 years, *J. Natl. Cancer Inst.* 92 (18) (2000) 1490–1499.
- [26] Miller A.B., To T., Baines C.J., 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(5 Part 1):305–12.
- [27] A.B. Miller, C. Wall, C.J. Baines, P. Sun, T. To, S.A. Narod, Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomised screening trial. *BMJ, Br. Med. J.* 348 (2014) g366.
- [28] I. Andersson, L. Janzon, Reduced breast cancer mortality in women under age 50: updated results from the Malmö Mammographic Screening Program, *J. Natl. Cancer Inst. Monogr.* 22 (1997) 63–67.
- [29] S. Zackrisson, I. Andersson, L. Janzon, J. Manjer, J.P. Garne, Rate of over-diagnosis of breast cancer 15 years after end of Malmö mammographic screening trial: follow-up study, *Bmj* 332 (7543) (2006) 689–692.
- [30] S. Shapiro, Evidence on screening for breast cancer from a randomized trial, *Cancer* 39 (6) (1977) 2772–2782.
- [31] S. Shapiro, Periodic Screening for Breast Cancer: The HIP Randomized Controlled Trial, *JNCI Monogr.* 1997 (22) (1997) 27–30.
- [32] S. Moss, I. Thomas, A. Evans, B. Thomas, L. Johns, Randomised controlled trial of mammographic screening in women from age 40: results of screening in the first 10 years, *Br. J. Cancer* 92 (5) (2005) 949–954.
- [33] S.M. Moss, H. Cuckle, A. Evans, L. Johns, M. Waller, L. Bobrow, Effect of mammographic screening from age 40 years on breast cancer mortality at 10 years' follow-up: a randomised controlled trial, *Lancet* 368 (9552) (2006) 2053–2060.
- [34] S.M. Moss, C. Wale, R. Smith, A. Evans, H. Cuckle, S.W. Duffy, Effect of mammographic screening from age 40 years on breast cancer mortality in the UK Age trial at 17 years' follow-up: a randomised controlled trial, *Lancet Oncol.* 16 (9) (2015) 1123–1132.
- [35] P.C. Prorok, P. Wright, T.R. Riley, B.S. Kramer, C.D. Berg, J.K. Gohagan, Overall and Multiphasic Findings of the Prostate, Lung, Colorectal and Ovarian (PLCO) Randomized Cancer Screening Trial. *Rev Recent Clin Trials* 13 (4) (2018) 257–273.
- [36] Oken M.M., Hocking W.G., Kvale P.A., Andriole G.L., Buys S.S., Church T.R., et al. Screening by chest radiograph and lung cancer mortality: the Prostate, Lung, Colorectal, and Ovarian (PLCO) randomized trial Differences in risk for type 1 and type 2 ovarian cancer in a large cancer screening trial. *JAMA.* 2011;306(17): 1865–73.
- [37] Z. Saghir, A. Dirksen, H. Ashraf, K.S. Bach, J. Brodersen, P.F. Clementsen, et al., CT screening for lung cancer brings forward early disease. The randomised Danish Lung Cancer Screening Trial: status after five annual screening rounds with low-dose CT, *Thorax* 67 (4) (2012) 296–301.
- [38] H.J. de Koning, C.M. van der Aalst, P.A. de Jong, E.T. Scholten, K. Nackaerts, M. A. Heuvelmans, et al., Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial, *N. Engl. J. Med* 382 (6) (2020) 503–513.
- [39] U. Yousaf-Khan, C. van der Aalst, P.A. de Jong, M. Heuvelmans, E. Scholten, J. W. Lammers, et al., Final screening round of the NELSON lung cancer screening trial: the effect of a 2.5-year screening interval, *Thorax* 72 (1) (2017) 48–56.
- [40] U. Pastorino, M. Rossi, V. Rosato, A. Marchiano, N. Sverzellati, C. Morosi, et al., Annual or biennial CT screening versus observation in heavy smokers: 5-year results of the MILD trial, *Eur. J. Cancer Prev.* 21 (3) (2012) 308–315.
- [41] U. Pastorino, M. Silva, S. Sestini, F. Sabia, M. Boeri, A. Cantarutti, et al., Prolonged lung cancer screening reduced 10-year mortality in the MILD trial: new confirmation of lung cancer screening efficacy, *Ann. Oncol.* 30 (7) (2019) 1162–1169.
- [42] S. González Maldonado, E. Motsch, A. Trotter, H.U. Kauczor, C.P. Heussel, S. Hermann, et al., Overdiagnosis in lung cancer screening: Estimates from the German Lung Cancer Screening Intervention Trial, *Int J. Cancer* 148 (5) (2021) 1097–1105.
- [43] B.H. Zhang, B.H. Yang, Z.Y. Tang, Randomized controlled trial of screening for hepatocellular carcinoma, *J. Cancer Res Clin. Oncol.* 130 (7) (2004) 417–422.
- [44] J. Hugosson, M.J. Roobol, M. Månsson, T.L.J. Tammela, M. Zappa, V. Nelen, et al., A 16-yr Follow-up of the European Randomized study of Screening for Prostate Cancer, *Eur. Urol.* 76 (1) (2019) 43–51.
- [45] G. Sandblom, E. Varenhorst, O. Löfman, J. Rosell, P. Carlsson, Clinical consequences of screening for prostate cancer: 15 years follow-up of a randomised controlled trial in Sweden, *Eur. Urol.* 46 (6) (2004) 717–723.
- [46] G. Sandblom, E. Varenhorst, J. Rosell, O. Löfman, P. Carlsson, Randomised prostate cancer screening trial: 20 year follow-up, *Bmj* 342 (2011) d1539.
- [47] G.L. Andriole, E.D. Crawford, R.L. Grubb 3rd, S.S. Buys, D. Chia, T.R. Church, et al., Mortality results from a randomized prostate-cancer screening trial, *N. Engl. J. Med* 360 (13) (2009) 1310–1319.
- [48] G.L. Andriole, E.D. Crawford, R.L. Grubb 3rd, S.S. Buys, D. Chia, T.R. Church, 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.* 104 (2) (2012) 125–132.
- [49] P.F. Pinsky, E. Miller, P. Prorok, R. Grubb, E.D. Crawford, G. Andriole, Extended follow-up for prostate cancer incidence and mortality among participants in the Prostate, Lung, Colorectal and Ovarian randomized cancer screening trial, *BJU Int* 123 (5) (2019) 854–860.
- [50] S.M. Temkin, E.A. Miller, G. Samimi, C.D. Berg, P. Pinsky, L. Minasian, Outcomes from ovarian cancer screening in the PLCO trial: Histologic heterogeneity impacts detection, overdiagnosis and survival. in: *European journal of cancer (Oxford, England: 1990)*, 87, 2017, pp. 182–188.
- [51] L.J. Jacobs, S.J. Skates, N. MacDonald, U. Menon, A.N. Rosenthal, A.P. Davies, et al., Screening for ovarian cancer: a pilot randomised controlled trial, *Lancet* 353 (9160) (1999) 1207–1210.
- [52] S.W. Duffy, D. Vulkan, H. Cuckle, D. Parmar, S. Sheikh, R.A. Smith, et al., Effect of mammographic screening from age 40 years on breast cancer mortality (UK Age trial): final results of a randomised, controlled trial, *Lancet Oncol.* 21 (9) (2020) 1165–1172.
- [53] G. Sandblom, E. Varenhorst, J. Rosell, O. Löfman, P. Carlsson, Randomised prostate cancer screening trial: 20 year follow-up, *BMJ* 342 (2011) d1539.
- [54] Force USPST. [Available from: <https://www.uspreventiveservicestaskforce.org/uspstf/about-uspstf/methods-and-processes/procedure-manual/procedure-manual-section-4-evidence-review-development>.
- [55] Pollock M.F.R., Becker L.A., Pieper D., Hartling L. Chapter V: Overviews of Reviews. Available from www.training.cochrane.org/handbook.2022.
- [56] P.H. Zahl, K.J. Jørgensen, P.C. Gotzsche, Overestimated lead times in cancer screening has led to substantial underestimation of overdiagnosis, *Br. J. Cancer* 109 (7) (2013) 2014–2019.
- [57] S.W. Duffy, D. Parmar, Overdiagnosis in breast cancer screening: the importance of length of observation period and lead time, *Breast Cancer Res* 15 (3) (2013) R41.
- [58] R. Gulati, E.J. Feuer, R. Etzioni, Conditions for Valid Empirical Estimates of Cancer Overdiagnosis in Randomized Trials and Population Studies, *Am. J. Epidemiol.* 184 (2) (2016) 140–147.
- [59] Deeks J.J.H.J., Altman D.G. (editors). Chapter 10: Analysing data and undertaking meta-analyses. Chapter 10: Analysing data and undertaking meta-analyses. Available from www.training.cochrane.org/handbook.2022.
- [60] M. Johansson, F. Borys, H. Peterson, G. Bilamour, M. Bruschetti, K.J. Jørgensen, Addressing harms of screening - A review of outcomes in Cochrane reviews and suggestions for next steps, *J. Clin. Epidemiol.* 129 (2021) 68–73.
- [61] R. Manser, A. Lethaby, L.B. Irving, C. Stone, G. Byrnes, M.J. Abramson, et al., Screening for lung cancer, *Cochrane Database Syst. Rev.* 2013 (6) (2013), Cd001991.
- [62] D. Ilic, M.M. Neuberger, M. Djulbegovic, P. Dahm, Screening for prostate cancer, *Cochrane Database Syst. Rev.* (1) (2013).
- [63] Siu A.L. Screening for Breast Cancer: U.S. Preventive Services Task Force Recommendation Statement. *Annals of Internal Medicine*; 2016. [Accessed October 2016] Available from <https://www.uspreventiveservicestaskforce.org/Page/Document/UpdateSummaryFinal/breast-cancer-screening1?ds=1&s=breast%20cancer%20screening>.
- [64] Moyer V.A. Screening for Lung Cancer: U.S. Preventive Services Task Force Recommendation Statement. *Annals of Internal Medicine*; 2013. [Accessed October 2016] Available from: <https://www.uspreventiveservicestaskforce.org/Page/>

- Document/UpdateSummaryFinal/lung-cancer-screening?ds=1&s=lung%20cancer%20screening. [
- [65] Moyer V.A.. Screening for Prostate Cancer: U.S. Preventive Services Task Force Recommendation Statement. *Annals of Internal Medicine*; 2012. [Accessed October 2016] Available from <https://www.uspreventiveservicestaskforce.org/Page/Document/UpdateSummaryFinal/prostate-cancer-screening?ds=1&s=prostat%20cancer%20screening>.
- [66] M.W. Marcus, S.W. Duffy, A. Devaraj, B.A. Green, M. Oudkerk, D. Baldwin, et al., Probability of cancer in lung nodules using sequential volumetric screening up to 12 months: the UKLS trial, *Thorax* 74 (8) (2019) 761–767.
- [67] A. Hodgkinson, J.J. Kirkham, C. Tudur-Smith, C. Gamble, Reporting of harms data in RCTs: a systematic review of empirical assessments against the CONSORT harms extension, *BMJ Open* 3 (9) (2013), e003436.
- [68] J.P. Ioannidis, S.J. Evans, P.C. Gotzsche, R.T. O'Neill, D.G. Altman, K. Schulz, et al., Better reporting of harms in randomized trials: an extension of the CONSORT statement, *Ann. Intern Med* 141 (10) (2004) 781–788.