



Late-Stage Outcomes as Surrogates for Mortality in Cancer Screening Trials: A Systematic Review and Meta-analysis

Matejka Rebolj¹, Adam R. Brentnall², Julia Geppert³, Nefeli Kouppa¹, Bethany Shinkins³, Karoline Freeman³, Chris Stanton³, Matthew J. Randell³, Samantha Johnson⁴, Robert A. Smith⁵, Peter Sasieni¹, Sam M. Janes⁶, Ruth Etzioni⁷, Stephen W. Duffy¹, and Sian Taylor-Phillips³

ABSTRACT

Late-stage cancer incidence has been proposed as a surrogate outcome for cancer-specific mortality in future screening trials. Two previous meta-analyses with 33 and 39 trials assessed trial-level surrogacy but provided inconsistent conclusions about the suitability of late-stage cancer endpoints replacing mortality. Our systematic review and meta-analysis (PROSPERO ID, CRD42023369320) investigated the association between the effect of cancer screening on the incidence of late-stage cancer and cancer-specific mortality. From 57 trials with 61 trial arm comparisons, correlation between late-stage incidence and mortality outcomes was 0.69 [95% confidence interval (CI), 0.47–0.84] for all cancers combined. Specifically, correlations were 0.58 (95% CI,

0.27–0.93) for bowel ($N = 11$ trials), 0.79 (95% CI, 0.49–0.94) for breast ($N = 13$), and 0.91 (95% CI, 0.84–0.96) for lung cancer ($N = 14$). Trial point estimates of the screening effect on mortality were within each trial's 95% CI late-stage incidence estimates in 56 of 61 (92%) trial arm comparisons and in 16 of 19 (84%) trial arm comparisons in which the entire 95% CI for screening effect on late-stage incidence was below one. Evidence suggests potential for late-stage cancer incidence as a key outcome in screening trials, but further research is needed to clarify when to measure late-stage outcomes, extrapolation for cancer types without trials, and the conditions when late-stage cancer does not accurately predict mortality.

Introduction

Guideline development methodology prioritizes randomized controlled trial ("trial") evidence on reduced cancer-specific mortality before recommending adoption of a new cancer screening program (1). Although a long-standing criterion for a high degree of confidence in the efficacy of a screening test, this can be a practical barrier to timely implementation of effective screening interventions because evaluating mortality in apparently healthy, asymptomatic screening cohorts is challenging, slow, and expensive. Randomized cancer screening trials powered for cancer-specific mortality need to be large, commonly in the tens (if not hundreds) of thousands of

individuals, and are typically conducted with a follow-up duration of one to two decades. This also poses a risk that by the time the results can be reported, the trialed screening technology may have become obsolete (2). Thus, there is interest in the possibility of using surrogate primary outcomes instead of cancer-specific mortality to facilitate faster and more efficient trials that ultimately support an earlier (planning of the) implementation of effective screening tests before direct evidence on mortality is available (3). Although some cancer screening trials designed to report surrogate primary outcomes are underway (4), there is no consensus on how or whether to utilize these endpoints as predictors of mortality, especially for the evaluation of a new screening test. Consequently, surrogate endpoints have largely not been adopted into decision-making by expert advisory groups responsible for national policy or guidelines.

The relationship between a reduction in late-stage diagnoses and mortality in cancer screening is complex and can differ across the spectrum of cancer locations and subtypes (5–7). For most cancer types, stage at diagnosis is prognostic of both morbidity and short- and long-term survival (8–11). Here, it is reasonable to expect that a stage shift from a late-stage to an early-stage diagnosis would be followed by a substantial reduction in cancer-specific mortality. Although nearly all cancers have better 5-year survival for earlier- versus late-stage disease, there is variation in the difference in prognosis across cancer types (12–15). In settings in which there is little difference in prognosis between earlier- versus late-stage disease groups, it is less likely that a stage shift would be followed by a substantial reduction in cancer-specific mortality. Conversely, earlier detection through screening might permit more successful treatment even though there was no stage shift. This might lead to a reduction in mortality but no effect would be predicted based simply on stage at diagnosis, whereas more granular evaluation of tumor characteristics within and across groups could be more informative. Whether late-stage incidence is a reasonable surrogate in the light of these effects requires in-depth study (16).

¹Centre for Cancer Screening, Prevention, and Early Detection, Wolfson Institute of Population Health, Queen Mary University of London, London, United Kingdom. ²Centre for Evaluation and Methods, Wolfson Institute of Population Health, Queen Mary University of London, London, United Kingdom. ³Warwick Screening, Warwick Applied Health, Warwick Medical School, University of Warwick, Coventry, United Kingdom. ⁴University of Warwick Library, University of Warwick, Coventry, United Kingdom. ⁵American Cancer Society Center for Early Cancer Detection Science, American Cancer Society, Atlanta, Georgia. ⁶Division of Medicine, Lungs for Living Research Centre, University College London, London, United Kingdom. ⁷Division of Public Health Sciences, Biostatistics Program, Fred Hutchinson Cancer, Seattle, Washington.

M. Rebolj and A.R. Brentnall contributed equally to this article.

J. Geppert and N. Kouppa contributed equally to this article.

Corresponding Author: Sian Taylor-Phillips, Warwick Applied Health, Warwick Medical School, University of Warwick, Coventry CV4 7AL, United Kingdom. E-mail: s.taylor-phillips@warwick.ac.uk

Cancer Epidemiol Biomarkers Prev 2025;34:1694–709

doi: 10.1158/1055-9965.EPI-25-0201

This open access article is distributed under the Creative Commons Attribution 4.0 International (CC BY 4.0) license.

©2025 The Authors; Published by the American Association for Cancer Research

Previous studies on this topic include meta-analytic evaluations of early indicators of a mortality reduction in breast screening, which found high trial-level correlations between the screening effect on late-stage breast cancer and breast cancer mortality in mammography trials (17, 18). Two recent meta-analytic evaluations included more cancer types but reached different conclusions about the suitability of late-stage outcomes as surrogates for mortality across multiple cancer types (Supplementary Methods and Tables, Section 1, Supplementary Tables S1 and S2; refs. 19, 20). Following their search strategies and inclusion criteria, these meta-analyses included 33 and 39 trials, respectively, for which one concluded that there was compelling evidence for the use of late-stage alternative endpoints in future cancer screening trials and the other concluded that late stage is not a suitable alternative endpoint for some cancer types.

In this article, we report a systematic literature review of cancer screening trials that reported both cancer-specific late-stage diagnoses and mortality outcomes and included any testing technology, cancer type, or trial size, following the inclusion criteria and other decisions made available in a pre-published protocol [PROSPERO ID: CRD42023369320; Supplementary Methods (Protocol)]. Many different criteria for surrogate endpoints have been developed. Here, we evaluate surrogacy by following the definition used to support the development of extensions to the CONSORT and SPIRIT checklists for studies reporting surrogate endpoints, that is, “An outcome that is used in clinical trials as a substitute for a direct measure of how a patient feels; functions; or survives. A surrogate outcome does not measure the clinical benefit of primary interest in and of itself; but rather is expected to predict that clinical benefit or harm based on epidemiologic; therapeutic; pathophysiologic; or other scientific evidence.” (21). Through this broad lens, we consider how well detection at an earlier stage at diagnosis through screening, measured through a reduction in the rate of late-stage cancer, can be seen to predict the clinical benefit of reduced cancer-specific mortality. We use meta-analysis to evaluate the association between a change in the relative incidence of late-stage cancer and a change in cancer-specific mortality at a trial level or the level of the class of screening test. We further consider whether one would be able to substitute evaluation of a screening test using a reduction in cancer-specific mortality with analysis using a reduction in a late-stage surrogate.

Materials and Methods

Literature search

The systematic search of the literature was undertaken in two stages. Both stages are explained in detail in Supplementary Methods and Tables, Section 2, Supplementary Tables S3–S11. Briefly, the aim of the first stage (“Search 1”) was to identify trials of primary screening for any human cancer, using any screening test, that reported the mortality outcomes, without size, time, or other restrictions except English language. These searches were undertaken in the MEDLINE, Embase, and Web of Science databases and included publications registered therein by the 21st of September 2022. The second stage (“Search 2”) was undertaken in the MEDLINE and Embase databases for publications registered until the period spanning from the 11th of January 2023 to the 28th of April 2023 (the exact date depended on the trial) and aimed to identify, for each trial, all relevant publications reporting outcomes that might be considered as surrogates for mortality (see below for the definitions) or any additional mortality outcomes not identified by

“Search 1”. If additional trials were identified at this stage, they were included in the review. “Search 1” was supplemented by contacting cancer screening experts, checking the lists of trials included in the most recent cancer screening reviews by independent bodies such as the International Agency for Research on Cancer and the U.S. Preventive Services Task Force, and by publications identified via “Search 2”. To supplement “Search 2,” we checked the introduction and discussion sections of articles included in the mapping and data extraction phases for any previously unidentified references. Because of funder deadlines, no additional data were requested from the trialists.

Study selection

Titles and abstracts of the publications found by the searches were screened independently by the same two reviewers in both stages (J. Geppert and M. Reboli), consulting all full-text publications considered potentially relevant by either reviewer. Full-text publications were assessed against the inclusion and exclusion criteria (Supplementary Methods and Tables, Section 2, Supplementary Tables S3 and S6) independently by five reviewers (J. Geppert, M. Reboli, K. Freeman, N. Kouppa, and C. Stinton), with disagreements resolved by a third reviewer (A.R. Brentnall, M. Reboli, and S. Taylor-Phillips).

In “Search 1,” individual or cluster randomized controlled trials were eligible for inclusion if they reported cancer-specific or all-cause mortality outcomes for all arms. In “Search 2,” publications were eligible if they reported outcomes that might be considered as surrogates (see below for definitions) or further mortality outcomes for trials identified in “Search 1.” In both stages, we excluded non-randomized studies; nonhuman studies, letters, reviews, editorials, and communications with insufficient information on methods and/or no numerical outcomes data; gray literature and conference abstracts; and articles not available in the English language. Reasons for exclusions at the full-text level are reported in Supplementary Methods and Tables, Section 2, Supplementary Tables S5 (“Search 1”) and S9 (“Search 2”).

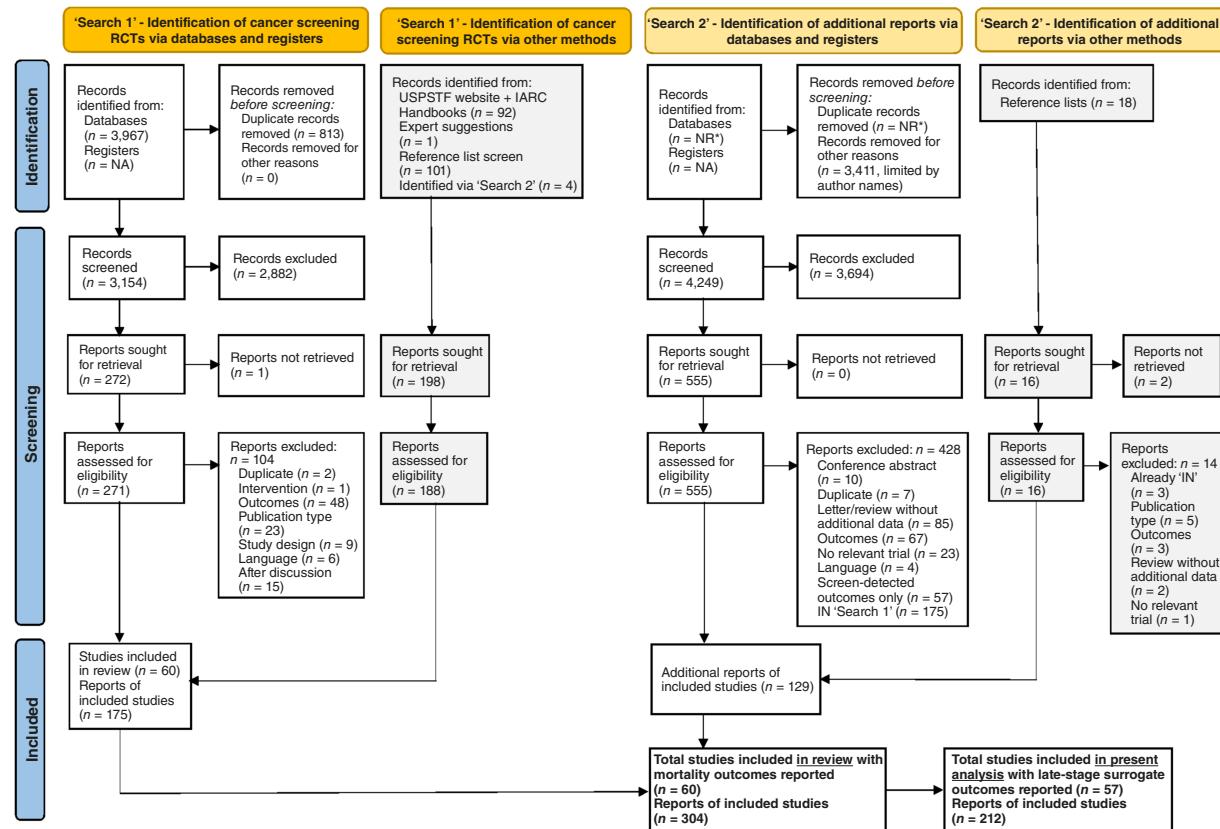
For each trial, all identified publications were mapped by the reported outcomes and timepoints, and publications that did not provide additional data were excluded from the data extraction. **Figure 1** shows the PRISMA flowchart for the two search stages combined. The designs and other characteristics of the included trials are summarized in **Table 1**; more details can be found in Supplementary Methods and Tables, Section 3, Supplementary Tables S12 and S13.

Data extraction

The relevant data were extracted into a piloted electronic data collection form by one researcher and checked by a second (J. Geppert, M. Reboli, N. Kouppa, C. Stinton, M.J. Randell, K. Freeman, and B. Shinkins). For consistency, the data for most trials were extracted by the same researcher (J. Geppert); for the remaining trials, J. Geppert performed the checks. Disagreements were resolved by a third reviewer (A.R. Brentnall, M. Reboli, and S. Taylor-Phillips). The extracted information included general information on the trial design and methods such as the eligibility criteria, study flow, population, intervention, and comparator and publication-specific statistical methods and results on the reported mortality and surrogate outcomes.

Outcomes

Our protocol included an extraction of the following outcomes: (i) cancer-specific mortality, (ii) the incidence of late-stage cancer,

**Figure 1.**

PRISMA flowchart for the systematic search of the literature. The figure describes the process of identifying, screening, and selecting the studies for the systematic review. * As half of the original MEDLINE search results (prior deduplication) were inadvertently overwritten, the original number retrieved is not available for "Search 2". Records: titles with abstracts. Reports: full-text articles. List of reasons for exclusion with explanation. Duplicate: the same abstract was included twice for full-text assessment i.e., it was identified from different databases and not picked up as a duplicate at the de-duplication stage. Intervention: the intervention was not invitation or no invitation to screening with a well-defined test, but study group subjects were urged to have annual multiphasic health checkups, with screening tests not being defined. Outcomes: Search 1: no relevant mortality outcomes reported for all arms (all cause or cancer specific). Search 2: no relevant comparative surrogate outcomes specified in our protocol were reported. Publication type: included letters, reviews, editorials, and communications with insufficient information on methods and/or no numerical outcomes data; gray literature and conference abstracts. Study design: not a randomized controlled trial. Language: the report was not published in English. After discussion: these are full texts for which the two reviewers (MR/JG) initially disagreed on the inclusion but agreed on the exclusion after a discussion. Conference abstract: only an abstract was reported. Letter/review without additional data: these are letters or reviews that reported sufficient information on methods but did not report any additional outcomes data that we have not already identified from other reports. No relevant trial: this report was not on a trial that could be included in the study. Screen-detected outcomes only: these reports only included outcomes from the intervention arm e.g., proportion of screen-detected cancers. IN "Search 1": reports identified in "Search 2" that were already included after "Search 1". Already "IN": reports that were already identified and included via "Search 1" or "Search 2" database searches. Review without additional data: previous reviews that reported sufficient information on methods but did not report any additional outcome data that we have not identified already from other reports. IARC, International Agency for Research on Cancer; NA, not applicable; NR, not reported; USPSTF, US Preventive Services Task Force.

and (iii) the proportion of cancers diagnosed at a late stage. Numerators of (ii) and (iii) included the cumulative numbers of late-stage target cancer diagnoses since randomization, by arm, up to the timepoint of the reporting by trialists. For the incidence of late-stage cancer, the denominator included person-years of follow-up in randomized individuals, by arm, up to the time of the reporting; if this was not available, we substituted it with the number of randomized individuals by arm. For the proportion of cancers diagnosed at a late stage, the denominator included the cumulative number of target cancers diagnosed at any stage, including cancers with unknown stage, up to the same timepoint since randomization.

For (i) and (ii), the endpoints extracted were the unadjusted rate ratio, relative risk (RR), or hazard ratio estimates comparing the

intervention versus the control arm and their 95% confidence intervals (CI) as reported in trial publications if available; otherwise, we used the extracted aggregated data and calculated rate ratios where person-years were available, or risk ratios if only numbers randomized were available, with Wald-type 95% CI. For (iii), we used the extracted aggregate data to estimate the relative proportions.

Definition of late-stage cancer

There is no standardized definition of late-stage cancer across cancer types or across the periods of time in which screening trials have been run. Therefore, we defined it using a variety of nomenclatures. The definition chosen for each trial was based on what was reported and (subjective) judgment on the most appropriate

definition based on the difference in expected prognosis and frequency between early and late stages. Late-stage definitions of cancer included (i) stage II+, stage IIB+, stage III+, or stage IV using a variety of staging systems; (ii) T4 (tumor invades adjacent structures) and/or N ≥ 1 (describing the number of nodes that contain cancer) and/or M1 (cancer has spread to other parts of the body) in the tumor–node–metastasis classification of malignant tumors; (iii) cancer-specific alternatives such as Dukes stage C or D for bowel cancer; or (iv) close cancer-specific approximations such as node positivity for breast cancer or commonly used alternative prognostic factors such as Gleason score 8+ for prostate cancer. Where the trialists did not categorize cancers into late and early stages, we attempted to follow a consistent definition across trials and cancer types, but this was not always possible. In particular, for one bowel and one cervical screening trial, data on cancer diagnoses were not provided by stage or another prognostic factor. However, screening tests in both trials targeted (early) preinvasive lesions with the aim of preventing invasive cancer. Therefore, we defined late-stage cancer as any invasive (bowel or cervical, respectively) cancer as screening with these tests is expected to reduce the cancer-specific incidence. The final definitions of late-stage cancer by trial, as used in our primary analyses, are given in **Table 2**.

Trial-specific timepoints for late-stage outcomes

The number of relevant publications varied substantially by trial, ranging from a single report on the mortality and a late-stage cancer outcome to multiple reports with assessments at various timepoints since randomization. Where late-stage cancer outcomes were reported at several timepoints, we defined the “primary” available late-stage outcome as (i) one that was measured the closest to the midway point between the end of the intervention period and the “main” mortality timepoint (defined below). This is because late-stage outcomes measured after the conclusion of the intervention in a trial and before the “main” mortality outcome could be informative for early decision-making about a population-based rollout. If no such report was identified, then an estimate reported (ii) at the same time as the “main” mortality outcome was considered as the “primary” available late-stage outcome or else (iii) at the latest timepoint when the screening intervention was still ongoing or soon thereafter. Thus, the defined “primary” late-stage outcomes for each trial are reported in **Table 2** and in Supplementary Methods and Tables, Section 3, Supplementary Tables S14 (relative incidence of late-stage diagnoses) and S15 (relative proportion of cancers diagnosed at late stage).

Trial-specific “main” timepoints for cancer-specific mortality outcomes

Where cancer-specific mortality was reported at more than one timepoint, we defined the “main” mortality timepoint as one at which primary results were to be calculated by trialists in the original publications, in the trial protocol, or in the statistical analysis plan and the power calculations or if necessary, the only or the latest available timepoint. The “main” mortality endpoints are reported by trial in **Table 2** and in Supplementary Methods and Tables, Section 3, Supplementary Tables S14 and S15.

Statistical analysis

We evaluated the association between the screening effect on late-stage incidence and proportion cancer outcomes and cancer-specific mortality using a fixed-effects linear model of log RR estimates weighted by the inverse variance of the screening effect on mortality [rather than late-stage incidence, because we felt that precision on

cancer-specific mortality better standardizes information at a trial level across different cancers than either surrogate examined, which are also affected by differing late-/early-stage incidence (in control) across cancer types]. We made no assumption on the position of the intercept in this regression model; a value close to the origin, however, would be indicative of no screening effect on mortality when there is no screening effect on the late-stage outcome. We calculated a weighted Pearson correlation coefficient and associated coefficient of determination (R^2) when at least three trials were available; 95% CIs used a non-parametric empirical bootstrap (50,000 resamples), with trial as the resampling unit.

We evaluated heterogeneity in multiple ways. We explored the relationship between late-stage cancer and cancer-specific mortality by the type of screening test for cancers for which sufficient data were available. To study the impact of the timing of the reporting on these associations, we plotted estimates of the relative incidence of late-stage cancer against the “main” relative mortality estimate when the former were reported for at least two timepoints from the same trial (Supplementary Methods and Tables, Section 4, Supplementary Table S16). Subgroup analyses focused on trials in which the late-stage outcomes were measured before the “main” mortality endpoint but after the screening period had ended. Sensitivity analysis used stage III+ across all cancers. Sensitivity of correlation coefficients to individual trial (as a unit of observation) was assessed by leave-one-out validation, in which each trial was left out once and the weighted correlation was performed on the other trials.

Exploratory analysis included the use of a single combined outcome for each combination of cancer type and class of screening test. The combined outcomes and the associated z-statistics were obtained from separate fixed-effect meta-analysis for late-stage and mortality outcomes. The aim of this analysis was to reduce noise from the trial-level analysis and help understand potential surrogacy across classes of screening tests. Another exploratory analysis evaluated the forecast calibration of late-stage incidence effects for cancer-specific mortality effects. A (naïve) density forecast treated 95% CIs for log RR as the 2.5% and 97.5% quantiles from a normal distribution. Calibration of probability integral transform values (the cumulative distribution function value of observed mortality given the forecast density) was evaluated through an Anderson–Darling goodness-of-fit test (22).

The systematic search of the literature was operationalized and documented using EndNote version 20 (Clarivate). The data were extracted into Excel files (Microsoft). All statistical analyses were undertaken with R version 4.2.1 [packages “weights” (23) to calculate the weighted correlations, “metafor” (24) to calculate effect measures and CIs for rate ratios and RRs from aggregate data when these were not reported by trialists, and “boot” (25) to calculate bootstrap CIs for summary statistics].

Data availability

All data analyzed in this report are available on GitHub at <https://github.com/brentnall/sums>

Results

Literature review

Study selection

For “Search 1,” we screened 3,154 unique records identified by either of the three bibliographical databases (**Fig. 1**; Supplementary Methods and Tables, Section 2). Of those, we assessed 271 full texts for eligibility. Another 188 full texts were assessed after they had

Table 1. Summary of the design and other characteristics of the 57 trials included in the systematic review and meta-analysis.

Cancer type	Trial acronym or name	Country	Calendar years of enrollment	Population type	Population size	Intervention arm(s)	Control arm
Bowel	Finnish Funen Gothenburg Minnesota	Finland Denmark Sweden United States	2004–2012 ^a 1985 1982–1990 1975–1978	General General General General	362,165 61,933 Cohorts 1–3: 68,308 46,551	FOBT FOBT FOBT (1) FOBT (annual) FOBT (2) FOBT (biennial) FS (±FOBT; 1:1)	Usual care Usual care Usual care Usual care
NORCCAP		Norway	1999–2000 (55–64 y); 2001 (50–54 y) 2009–2014	General	100,210 (55,736 aged 55–64 years)	FS	Usual care
NordICC		Poland, Norway, and Sweden (Netherlands)	Pilot: 1981–1983 Main: 1985–1991 1993–2001 1995–1999 1983	General	Poland, Norway, and Sweden: 85,179	Colonoscopy	Usual care
Nottingham		United Kingdom	Pilot: 1981–1983 Main: 1985–1991 1993–2001 1995–1999 1983	General	152,850	FOBT	Usual care
PLCO (bowel)		United States	1994–1999	General	154,887	FS	Usual care
SCORE		Italy	1980–1985	General	34,292	FS	Usual care
Telemark		Norway	1980–1985	General	799	FS (±colonoscopy after 13 years)	Usual care
UKFSS-1		United Kingdom	1978–1985	General	170,432	FS	Usual care
CNBSS-1		Canada	1982–1984	General	50,430	Mammography + CBE	CBE (1 round)
CNBSS-2		Canada	1982–1984	General	39,405	Mammography + CBE	CBE
Edinburgh		United Kingdom	1963–1964	General	54,671 (cohort 1: 44,288)	Mammography + CBE	Usual care
Gothenburg		Sweden	1976–1978	General	52,833	Mammography	Usual care
HIP NY		United States	1998–2002	General	~62,000	Mammography + CBE	Usual care
Malmo		Sweden	1983–1989	General	42,283	Mammography	Usual care
Mumbai		India	1989–1991	General	151,538	CBE	Usual care
Russia/WHO		Russia	1981–1983	General	267,400	BSE	Usual care
Shanghai		China	Initiated 2006	General	60,261	Mammography	Usual care
Stockholm		Sweden	1977–1980	General	115,290	CBE	Usual care
Trivandrum		India	(Koppaberg); 1978–1981	General	162,981 (134,867 aged 40–74 years)	Mammography	Usual care
Two-County		Sweden	(Ostergotland)	General	160,921	Mammography	Usual care
UK Age		United Kingdom	1990–1997	General	520,372	Papnet cytology	Conventional cytology
Finnish		Finland	1999–2003	General	151,538	VIA	Usual care
Mumbai		India	1998–2002	General	131,016	(1) VIA	Usual care
Osmanabad		India	Initiated 2000	General		(2) Conventional cytology	
Cervical						(3) HPV testing	
Liver	Tamil Nadu Qidong Shanghai	India China China	Initiated 1999	General	80,282	VIA	Usual care
Lung	Czech Study	Czech	1989–1992 1993–1995 1976–1977	High risk High risk High risk	5,581 19,200 6,345	AFP + ALT (3–6 rounds) AFP + US	AFP + ALT (1 round)
DANTE		Italy	2001–2006	High risk	2,811	CXR + sputum cytology	CXR + sputum cytology (less intensive schedule)
DLCST		Denmark	2004–2006	High risk	LDCT	LDCT	Clinical review (baseline): + CXR + sputum cytology
ITALUNG		Italy	2004–2006	High risk	CXR + sputum cytology	CXR	CXR
LSS		United States	1973–1978	High risk	LDCT	LDCT	Usual care
LUSI		Germany	2000	High risk	CXR	CXR	Usual care
MLP		United States	2007–2011	High risk	CXR + sputum cytology	CXR + sputum cytology	Usual care
MSK		United States	1971–1976	High risk	CXR + sputum cytology	CXR + sputum cytology	CXR
MILD		Italy	1974–1978	High risk	(1) LDCT (annual)	(1) LDCT (annual)	Usual care
NELSON		Belgium and Netherlands	2003–2006	High risk	(2) LDCT (biennial)	(2) LDCT	Usual care

(Continued on the following page)

Table 1. Summary of the design and other characteristics of the 57 trials included in the systematic review and meta-analysis. (Cont'd)

Cancer type	Trial acronym or name	Country	Calendar years of enrollment	Population type	Population size	Intervention arm(s)	Control arm
NLST PLCO (lung)	United States United States United Kingdom China	United States United States India United States United Kingdom	2002-2004 1993-2001 2011-2013 2009-2014 ^b	High risk General High risk General	53,454 154,887 4,055 122,074	LDCT CXR LDCT EBV serology + indirect mirror + lymphatic palpation Visual inspection CA125 ± TVU CA125 (+ TVU as 2nd tier)	CXR Usual care Usual care Usual care Usual care Usual care Usual care Usual care
NPC	China NPC						
Oral	Trivandrum	India	1995-2004	General	191,873		
Ovarian	PLCO (ovarian) UK Pilot	United States United Kingdom	1993-2001 1989	General General - volunteered for a previous prevalence screen	78,215 21,935		
Prostate	UKCTOCS	United Kingdom	2001-2005	General	202,638	(1) CA125 (+ TVU as 2nd tier) (2) TVU Blood PSA Blood PSA + DRE ± TRUS (depending on country)	Usual care
	CAP ERSPC	United Kingdom Belgium, Finland, France, Italy, Netherlands, Portugal, Spain, Sweden, and Switzerland	2001-2009 1991-2005, depending on country	General General	415,537 266,512 (62,243 aged 55-69 years, excluding France and Portugal)	Blood PSA Blood PSA + DRE ± TRUS (depending on country)	Usual care Usual care
	ERSPC Pilot 1	Netherlands	1991-1992	Low risk (PSA <10.0 ng/mL at baseline)	1134		
	Norrköping	Sweden	1987	General	9,026	DRE only first, later DRE + blood PSA	Usual care
	PLCO (prostate) DAquapendente MVTEP SOMIT	United States Italy France Italy	1993-2001 NR (likely 2010s) 2009-2012 1993-1997	General High risk High risk High risk	76,683 195 399 201	Blood PSA + DRE Torso CT + FOBT Limited screening + PET-CT Extensive screening including torso CT	Usual care Usual care Usual care Usual care

Abbreviations: AFP, α -fetoprotein; ALT, alanine aminotransferase; BSE, breast self-examination; CA125, cancer antigen 125; CBE, clinical breast examination; CXR, chest X-ray; DRE, digital rectal examination; EBV, Epstein-Barr virus; FOBT, fecal occult blood test; FS, flexible sigmoidoscopy; HPV, human papillomavirus; LDCT, low-dose CT; NPC, nasopharyngeal cancer; NR, not reported; TRUS, transrectal ultrasound; TVU, transvaginal ultrasound; US, ultrasound; VIA, visual inspection with acetic acid; WHO, World Health Organization.

Trial name abbreviations: CAP, Cluster Randomized Trial of PSA Testing for Prostate Cancer; CNBSS, Canadian National Breast Screening Study; DANTE, Detection and Screening of Early Lung Cancer by Novel Imaging Technology and Molecular Essays Trial; ERSPC, European Randomized Study of Screening for Prostate Cancer; HIP NY, Health Insurance Plan of Greater New York; ITALUNG, Italian Lung Cancer Screening Trial; LSS, Lung Screening Study; LUSI, German Lung Cancer Screening Trial; MILD, Multicentric Italian Lung Detection; MLP, Mayo Lung Project; MSK, Memorial Sloan Kettering Lung Study; MVTEP, Standard Diagnostic Procedures With or Without Fludeoxyglucose F 18 Positron Emission Tomography in Finding Cancer in Patients With a Blood Clot in a Vein; NELSON, Netherlands-Leuven Longkanker Screenings Onderzoek; NLST, National Lung Screening Trial; NORCCAP, Norwegian colorectal cancer prevention; NordICC, Nordic-European Initiative on Colorectal Cancer; PLCO, Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial; SCORE, Screening for Colon Rectum trial; SOMIT, Subsequent diagnosis Of Malignancy in patients presenting with idiopathic venous Thromboembolism; UKCTOS, UK Collaborative Trial of Ovarian Cancer Screening; UKFST, UK Flexible Sigmoidoscopy Screening Trial; UKLS, UK Lung Cancer Screening Trial.

^alater extended to 2014.

^bInterim analysis of the Zhongshan cohort recruited from 2009 to 2014. The whole trial included three towns in Zhongshan City and 13 towns in Sihui City, recruiting people from 2008 to 2015.

^cDifferent target cancers screened for in each trial. High-risk population was defined as patients with a recent diagnosis of venous thrombosis \pm pulmonary embolism.

Table 2. Summary of the “primary” late-stage and “main” mortality outcome estimates in the 57 trials.

Cancer type	Trial acronym or name	Definition of late stage	Timing of surrogate (years since entry)	Timing of mortality (years since entry)	Relative late-stage cancer incidence, RR (95% CI)	Relative proportion of late-stage cancer RR (95% CI)	Cancer-specific mortality, RR (95% CI)	Weights in the main analysis
Bowel	Finnish Funen	N ≥ 1 Dukes C or distant spread ^c	Range, 0–7 Me 9.1	Mid 4.5, range 0.0–8.3 Me 9.1	1.07 (0.90–1.27) ^a 0.84 (0.70–1.01)	0.89 (0.78–1.01) ^a 0.84 (0.74–0.96)	1.04 (0.84–1.28) ^b 0.82 (0.68–0.99)	1.99 2.41
	Gothenburg Minnesota	Dukes D Dukes D	Me 15.5 Me 11.8, max 13	Me 15.5 Me 11.8, max 13	0.95 (0.76–1.18) 0.50 (0.33–0.76) (annual vs. UC)	0.99 (0.81–1.20) 0.57 (0.38–0.84) (annual vs. UC)	0.84 (0.71–0.99) 0.67 (0.51–0.89) (annual vs. UC)	3.17 Annual: 1.09 Biennial: 1.33
	NORCCAP	Dukes C or distant spread	Mid 7, max 8	Mid 15	0.90 (0.70–1.15) ^d	0.88 (0.76–1.02) ^d	0.82 (0.66–1.02) ^d	180
	NordICC	Dukes C or D	Mid 10.0	Mid 10.0	0.80 (0.65–1.00)	0.97 (0.82–1.14)	0.91 (0.64–1.16)	1.33
	Nottingham PLCO (bowel)	Dukes C or D	Mid 7.8	Mid 7.8	0.91 (0.80–1.04)	0.88 (0.80–0.97)	0.74 (0.74–0.98)	4.23
	SCORE	Stage III+	Mid 11.9, max 13.0	Mid 12.1, max 13.0	0.71 (0.62–0.81)	0.90 (0.81–1.00)	0.74 (0.63–0.87)	3.27
	Telemark	Dukes C or D	Mid 10.5	Mid 11.4	0.73 (0.57–0.94)	0.90 (0.75–1.07)	0.78 (0.56–1.08)	0.81
	UKFSST	Invasive cancer ^e	Up to 11	Up to 13	0.50 (0.05–5.48)	1.00 (0.18–5.46)	0.33 (0.03–3.18)	0.02
	CNBSS-1	N ≥ 1	Me 10.8	Mid 11.2	0.77 (0.70–0.84)	No data on stage	0.68 (0.59–0.80)	3.24
	CNBSS-2	N ≥ 1	Max 7	Max 7	1.55 (1.13–2.11)	1.27 (0.97–1.66)	1.36 (0.84–2.21)	0.36
	Edinburgh	Stage III+	Me 6.8	Me 7	1.09 (0.82–1.45)	0.94 (0.74–1.20)	0.97 (0.62–1.52)	0.42
	Gothenburg	N ≥ 1	Me 6 (screening period)	Max 14; cancers diagnosed in screening period	0.63 (0.46–0.87) ^f	0.46 (0.35–0.61) ^f	0.84 (0.60–1.16) ^f	0.81
	HIP NY	N ≥ 1	5	10	0.80 (0.61–0.95)	0.80 (0.65–1.10)	0.78 (0.57–1.06)	0.90
	Nämålo	Stage II+	Me 8.8	Me 8.8	0.83 (0.68–1.00)	0.67 (0.58–0.76)	0.96 (0.68–1.35)	1.86
	Numbaï	Stage III+	Mid 18	Mid 18	0.81 (0.68–0.97)	0.83 (0.72–0.96)	0.85 (0.70–1.02)	0.74
	Russia/WHO	N ≥ 1	Up to 10	Up to 13; cancers diagnosed up to 1994	1.20 (1.01–1.43)	0.97 (0.86–1.09)	1.07 (0.86–1.34)	2.57
	Shanghai	N ≥ 1	Me 10	Me 10	0.92 (0.79–1.07)	0.95 (0.84–1.08)	0.74 (0.50–1.10)	1.79
	Stockholm	Stage II+	Up to 6	Me 12, cancers diagnosed 1981–1986	0.88 (0.68–1.12)	0.93 (0.78–1.12)	0.74 (0.50–1.33)	1.42
	Trivandrum Two-County UK Age	Stage III+	Me 12	Me 12	1.18 (0.92–1.52)	0.90 (0.74–1.10)	1.03 (0.75–1.41)	0.85
		Stage II+	Me 6	Me 6	0.79 (0.69–0.91) ^g	0.61 (0.55–0.67) ^g	0.72 (0.54–0.98) ^g	0.96
		N ≥ 1	Me 10.6	Mid 17.7; cancers diagnosed in screening period	0.90 (0.78–1.05)	0.89 (0.79–1.00)	0.88 (0.74–1.04)	3.07
	Cervical	Invasive cancer ^e	Me 6.3	Me 12	1.00 (0.76–1.29)	No data on stage	1.11 (0.62–1.92)	0.29
	Finnish Mumbai Osmannabad	Stage ≥IIB+ Stage II+	Me 8.0, max 12 Me 8	Me 8.0, max 12 Me 8	0.84 (0.63–1.12) 0.97 (0.72–1.32) (VIA vs. UC)	0.86 (0.72–1.04) 0.79 (0.65–0.95) (VIA vs. UC)	0.69 (0.50–0.94) 0.81 (0.57–1.16) (VIA vs. UC)	0.89
					0.70 (0.50–0.98) (cytology vs. UC)	0.55 (0.43–0.69) (cytology vs. UC)	0.83 (0.58–1.20) (cytology vs. UC)	VIA: 0.67
					0.44 (0.30–0.64) (HPV vs. UC)	0.44 (0.33–0.59) (HPV vs. UC)	0.49 (0.32–0.74) (HPV vs. UC)	HPV: 0.50
					0.59 (0.44–0.79) (HPV vs. UC)	1.01 (0.86–1.20)	0.59 (0.44–0.79)	0.97
					0.54 (0.37–0.81) (HPV vs. UC)	0.48 (0.35–0.67)	1.01 (0.81–1.29)	1.62
					0.53 (0.32–0.88) (HPV vs. UC)	0.63 (0.41–0.98)	0.41 (0.27–0.61)	0.44
	Liver	Stage II ^h	Me 5.2	Me 6, max 7	0.54 (0.37–0.81) (HPV vs. UC)	1.15 (0.78–1.71)	0.87 (0.67–1.15)	2.77
	Gidong Shanghai	Stage III ^h	Me 4.1–4.3	Me 5.2	0.53 (0.32–0.88) (HPV vs. UC)	1.14 (0.96–1.36)	1.00 (0.69–1.44)	0.63
	Czech Study	Stage II ⁱ	6	Me 4.1–4.3	0.94 (0.64–1.38) (HPV vs. UC)	0.69 (0.54–0.89)	1.03 (0.66–1.61)	0.43
	DANTE	Stage II+	Mid 8.35	Mid 5.2	1.12 (0.74–1.70)	0.59 (0.46–0.77)	0.70 (0.47–1.03)	0.57
	DLCST	Stage III+	Mid 9.8	Mid 4.1–4.3	0.81 (0.60–1.11)	0.86 (0.68–1.09)	0.91 (0.77–1.06)	1.68
	ITALUNG	Stage III+	Mid 9.3	Mid 5.2	0.76 (0.48–1.19)			
	Johns Hopkins	Stage II+	Me 7.2, max 9	Me 7.2, max 9	0.86 (0.68–1.09)			

(Continued on the following page)

Table 2. Summary of the “primary” late-stage and “main” mortality outcome estimates in the 57 trials. (Cont’d)

Cancer type	Trial acronym or name	Definition of late stage	Timing of surrogate (years since entry)	Timing of mortality (years since entry)	Relative late-stage cancer incidence, RR (95% CI)	Relative proportion of late-stage cancer RR (95% CI)	Cancer-specific mortality, RR (95% CI)	Weights in the main analysis
LSS		Stage III+	1	Md 5.2	1.78 (0.79-4.01)	0.89 (0.48-1.64)	1.24 (0.74-2.08)	0.32
LUSI		Stage II+	7	Md 8.9, max 11	0.55 (0.34-0.88)	0.39 (0.29-0.54)	0.72 (0.45-1.16)	0.38
MLP		Stage III+	Me 9 (range, 7.5-11.0)	Me 9 (range, 7.5-11.0)	0.98 (0.75-1.27)	0.76 (0.64-0.90)	1.06 (0.82-1.36)	1.35
MSK		Stage III+	Me 7.2, max 9	Me 7.2, max 9	1.01 (0.77-1.33)	0.98 (0.82-1.18)	0.95 (0.73-1.25)	1.14
MILD		Stage II+	Me 9.4-9.6	Me 9.7	0.74 (0.50-1.0)	0.64 (0.50-0.81)	0.70 (0.45-1.09)	0.45
NELSON		Stage III+	Me 9.5, max 10	Me 9.5, max 10	(LDCT combined vs. UC)	(LDCT combined vs. UC)	(LDCT combined vs. UC)	
NLST		Stage III+	Md 6.5	Me 6.5	0.71 (0.58-0.87) ^j	0.63 (0.55-0.72) ^j	0.76 (0.61-0.94) ^j	1.90
PLCO (lung)		Stage III+	Incidence: 7 (Me 6.7)	Md 11.9, max 13	0.79 (0.70-0.99)	0.70 (0.64-0.77)	0.84 (0.75-0.95)	5.66
UKLS		Stage III+	Proportion: Md 11.9, max 13	Me 11.9, max 13	0.94 (0.84-1.05)	0.92 (0.87-0.98)	0.99 (0.91-1.07)	15.47
China NPC		Stage III+	4	Md 7.3	0.44 (0.23-0.87)	0.33 (0.19-0.57)	0.65 (0.41-1.02)	0.42
Oral		Stage III+	Md 6.0	Md 6.0	0.78 (0.57-1.07)	0.70 (0.58-0.85)	0.81 (0.45-1.45)	0.25
Trivandrum		Stage III+	Me 4.4-4.9, max 9	Me 4.4-4.9, max 9	0.89 (0.68-1.16)	0.76 (0.64-0.91)	0.79 (0.51-1.22)	0.45
Ovarian		Stage III+	Md 12.4, max 13	Md 12.4, max 13	1.20 (0.96-1.5)	0.99 (0.89-1.10)	1.18 (0.91-1.54)	1.21
PLCO (ovarian)		Stage III+	Max 8 (1990-1997)	Max 8 (1990-1997)	0.61 (0.29-1.33)	0.76 (0.53-1.10)	0.50 (0.19-1.28)	0.10
UK Pilot		Stage III+	Md 10.8	Md 11.1	0.83 (0.70-0.99)	0.77 (0.69-0.87)	0.85 (0.70-1.03)	MMS: 2.31
UKCTOCS				(MMS vs. UC)	(MMS vs. UC)	(MMS vs. UC)	(MMS vs. UC)	TVU: 2.38
Prostate	CAP	Gleason 8-10	Max 6	Md 10, max 15	0.91 (0.77-1.08)	0.91 (0.82-1.10)	0.89 (0.73-1.07)	
	ERSPC	Gleason 8-10	Md 11.0	Me 10.5	(TVU vs. UC)	(TVU vs. UC)	(TVU vs. UC)	
	ERSPC Pilot 1	Gleason $\geq 3 + 4$	Md 19	Md 19	0.82 (0.74-0.9) ^k	0.63 (0.57-0.70)	0.96 (0.86-1.08)	6.62
		T3-4, N1, or Nx/M1	Max 13	Max 22, cancers diagnosed up to 1999	0.86 (0.50-1.48)	0.66 (0.43-1.01)	0.48 (0.17-1.36)	0.08
		NI and/or M1		Me 11.1, max 13	0.87 (0.62-1.23)	0.59 (0.46-0.76)	1.16 (0.78-1.73)	0.54
	PLCO (prostate)	Gleason 8-10	Max 10	Me 11.1, max 13	0.85 (0.73-0.99)	0.73 (0.63-0.85)	0.79 (0.68-0.9) ^k	4.05
Multiple ^l	D'Aquapendente	Stage IV “Advanced”	2	2	0.74 (0.17-3.23)	0.62 (0.18-2.16)	0.49 (0.09-2.64)	0.03
	MVTEP		2	2	0.71 (0.23-2.2)	0.77 (0.33-1.79)	0.40 (0.08-2.04)	0.03
	SOMIT	NI and/or M1	2	2	0.64 (0.22-1.90)	0.45 (0.21-0.96)	0.52 (0.10-2.75)	0.03

Abbreviations: HPV, human papillomavirus; LDCT, low-dose CT; M1, distant metastasis present; Max, maximum follow-up time; Me, mean follow-up time; MMS, multimodal screening; MX, metastasis cannot be measured; N1, cancer has spread to nearby lymph nodes; N1, N2, N3, number of nearby lymph nodes that have cancer; NPC, nasopharyngeal cancer; RR, relative risk or rate ratio; T3, tumor size or area; cancer has broken through the capsule (covering) of the prostate gland; T4, tumor size or area: cancer has spread into other body organs nearby; TVU, transvaginal ultrasound; UC, usual care; V/A, visual inspection with acetic acid; WHO, World Health Organization.
^a321,311 individuals randomized from 2004 to 2011.
^b362,165 individuals randomized from 2004 to 2012.
^c“Advanced” colorectal cancer was defined as stage C, distant spread, and no classification.
^d55,736 men and women aged 55-64 years.
^eNo stage data reported. For screening tests that can prevent cancer, the incidence of invasive cancer was used as late-stage surrogate.
^f44,288 women in cohort 1.
^g134,867 women aged 40 to 74 years.
^hChinese hepatocellular cancer staging system.
ⁱClinical-diagnostic staging system with occult, I, II, and III stages.
^j13,195 men.
^k162,243 men aged 55 to 69 years (excluding France and Portugal).
^lDifferent target cancers screened for in each trial.

been identified from supplementary sources. In total, we identified 60 cancer screening trials that reported mortality outcomes for all arms. The trial-specific searches in bibliographical databases during “Search 2” identified 4,249 records that required screening; of these, we assessed 555 full texts for eligibility. Another 16 full texts were identified from supplementary sources and assessed for eligibility. Altogether, “Search 1” and “Search 2” combined identified 57 trials including 61 trial arm comparisons that reported both the mortality outcome and a late-stage outcome covered in 212 publications. All references for the included original trial publications are available in Supplementary Methods and Tables, Section 3.

Study characteristics

Fifty-four trials evaluated screening tests for nine individual cancer types (**Table 1**). An additional three trials offered screening for multiple cancer types combined to a well-defined high-risk patient group (unprovoked venous thrombosis with or without pulmonary embolism), using a variety of methods. Overall, trials undertaken in average-risk individuals were larger than trials recruiting from high-risk populations. We noted a wide variation between the trials in factors such as epochs and settings in which they were undertaken, inclusion criteria, design, or screening tests (Supplementary Methods and Tables, Section 3; Supplementary Tables S12 and S13). Likewise, the definition of late-stage cancer depended on the cancer type and trial as did the timing (counted since randomization) of the reporting of the outcomes (**Table 2**; Supplementary Methods and Tables, Section 3, Supplementary Tables S14 and S15). Almost all trials reported their data at multiple timepoints (for late-stage cancer, see Supplementary Methods and Tables, Section 4, Supplementary Table S16).

Meta-analysis: primary analysis

Trial-level surrogacy meta-analysis: relative incidence of late-stage cancer between trial arms

Our primary analysis was of trial-level surrogacy using data from 57 trials. Only 14 of 57 trials reported 15 trial arm comparisons with late-stage incidence at an earlier timepoint than mortality but after the screening intervention had ended. Across the 57 trials, correlation between the relative incidence of late-stage cancer and relative mortality between trial arms was 0.69 (95% CI, 0.47–0.84; R^2 : 0.47; **Fig. 2**; Supplementary Methods and Tables, Section 4, Supplementary Table S17) for all cancers combined. The point estimate was unchanged in an unweighted analysis and when we excluded the two trials in which the late-stage outcome was defined as invasive cancer. The intercept was close to the origin (log RR: −0.04; 95% CI, −0.09 to −0.01). These values differed somewhat between cancer types. For cancer types with the largest number of trials, the correlation was 0.58 (95% CI, 0.27–0.93; R^2 : 0.34) for bowel cancer ($N = 11$), 0.79 (95% CI, 0.49–0.94; R^2 : 0.62) for breast cancer ($N = 13$), and 0.91 (95% CI, 0.84–0.96; R^2 : 0.83) for lung cancer ($N = 14$; **Fig. 3A–C**). For these three cancers, **Fig. 3** also categorizes each trial depending on the type of the screening test, and clustering by test type can be observed. Detailed results for all cancer types with at least three trials are presented in Supplementary Methods and Tables, Section 4, Supplementary Table S17 and in Supplementary Figure S1.

The estimate of the screening effect on the late-stage incidence changed depending on the time of the reporting (**Fig. 4**; Supplementary Methods and Tables, Section 4, Supplementary Table S16; and Supplementary Figure S2). In most examples, the relative incidence of late-stage cancer (screening arm relative to comparator

arm) decreased between the end of the intervention phase and the timepoint at which the mortality endpoint was measured.

Trial-level surrogacy meta-analysis: relative proportion of cancers diagnosed at late stage

Data for the relative proportion of cancers diagnosed at late stage were available from 55 trials. As expected, correlation between this late-stage outcome and the “main” relative mortality outcome was weaker than for the relative incidence of late-stage diagnoses. The estimate of the correlation coefficient for all cancers combined was 0.30 (95% CI, 0.00–0.54; R^2 : 0.09; **Fig. 2B**; Supplementary Methods and Tables, Section 4, Supplementary Table S18). Also here, the intercept was close to the origin (log RR: −0.07; 95% CI, −0.15 to −0.01). For bowel cancer ($N = 10$), the correlation was 0.21 (95% CI, −0.25 to 0.48; R^2 : 0.05), for breast cancer ($N = 13$) 0.53 (95% CI, 0.16–0.87; R^2 : 0.28), and for lung cancer ($N = 14$) 0.67 (95% CI, 0.27–0.85; R^2 : 0.44; **Fig. 3D–F**). Detailed results for all cancer types with at least three trials are presented in Supplementary Methods and Tables, Section 4, Supplementary Table S18 and in Supplementary Figure S3.

Meta-analysis: sensitivity and subgroup analyses

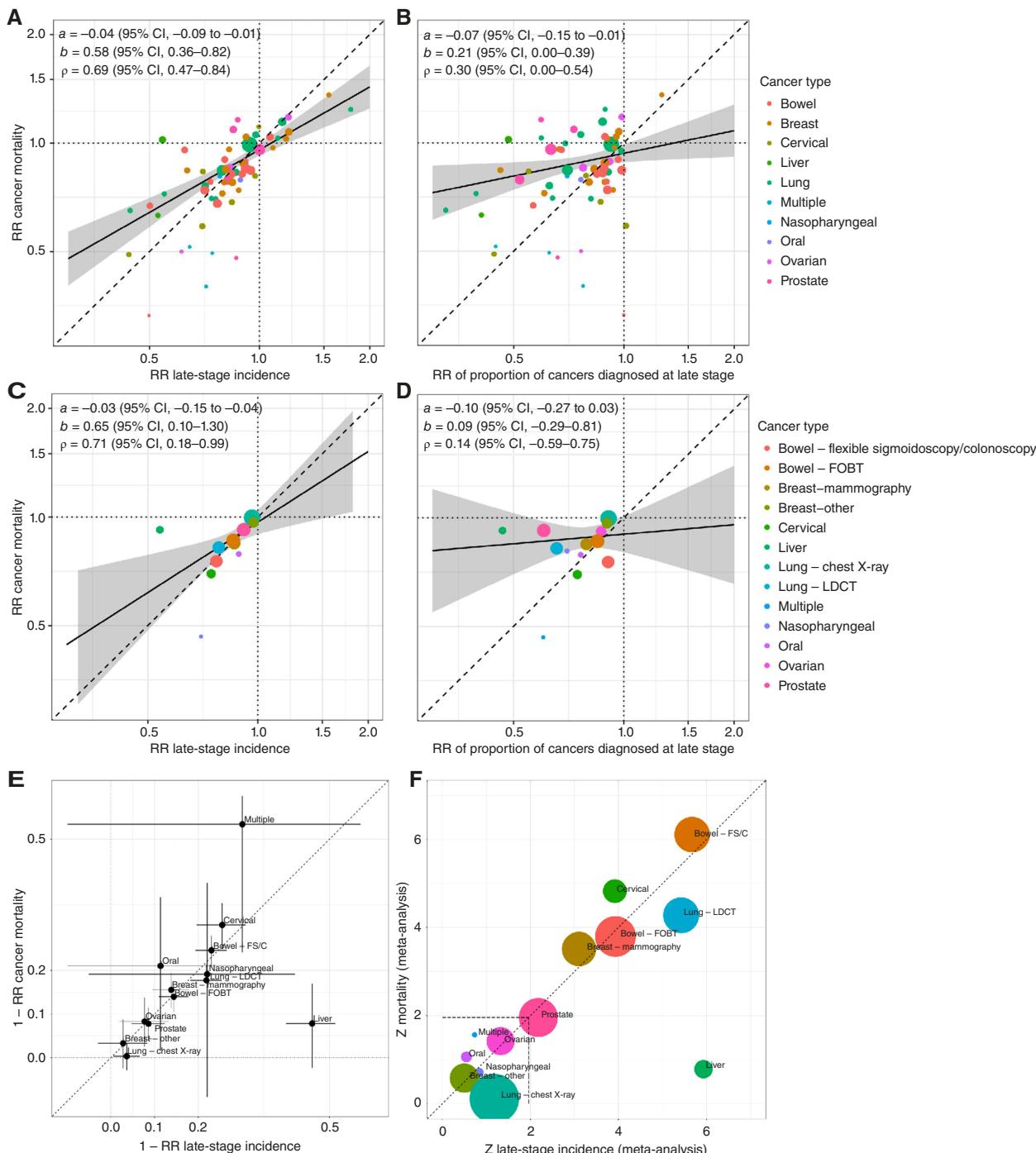
Data were available for fewer trials for pre-planned sensitivity and subgroup analyses. However, for both late-stage outcomes, the observed patterns remained similar as in the primary analysis. For the relative incidence of late-stage cancer, the correlation was 0.74 (95% CI, 0.48–0.90; R^2 : 0.54), including only trial arm comparisons that reported late-stage outcomes before the main mortality outcome ($N = 15$), and 0.67 (95% CI, 0.33–0.89; R^2 : 0.45) when late-stage cancer was defined as stages III+ ($N = 30$; Supplementary Methods and Tables, Section 4, Supplementary Table S17). In leave-one-out analysis, the weighted correlation coefficient across all trials ranged between 0.66 and 0.78 with an IQR of 0.68 to 0.69 (results not tabulated). The estimate was very sensitive to the Qidong liver trial results (exclusion of this trial led to correlation 0.78, 95% CI, 0.69–0.85).

Exploratory analyses

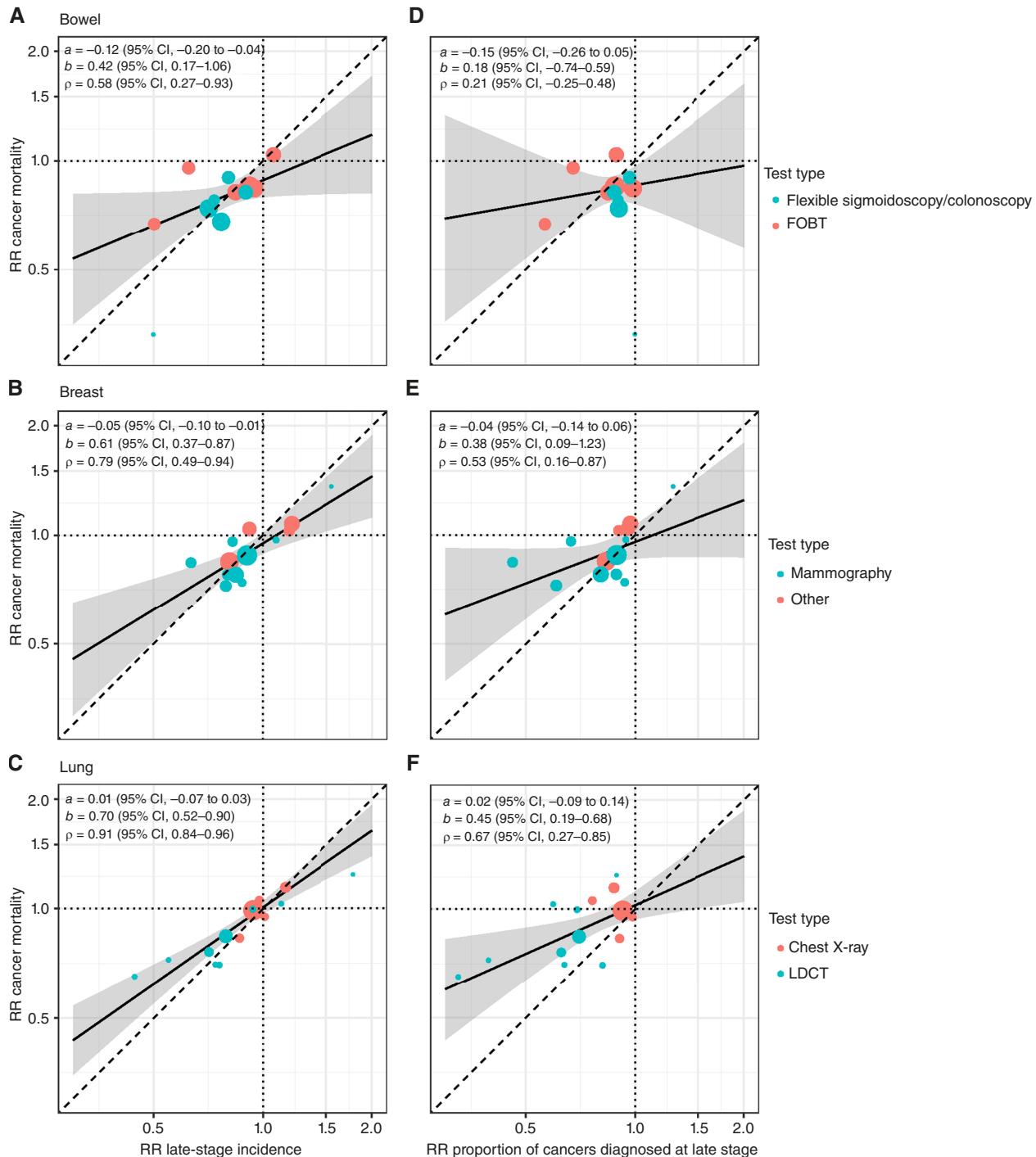
Association between the effect of screening on late-stage cancer incidence and on mortality at screening test level

Across all 57 included trials, there were on average 12% fewer target cancer deaths (95% CI, 9%–14%) in those randomized to a screening intervention compared with a control group. For comparison, there was a mean 13% reduction (95% CI, 11%–16%) in the late-stage cancer incidence (57 trials) and a mean 21% (95% CI, 19%–24%) reduction in the proportion of cancers diagnosed at late stage (55 trials).

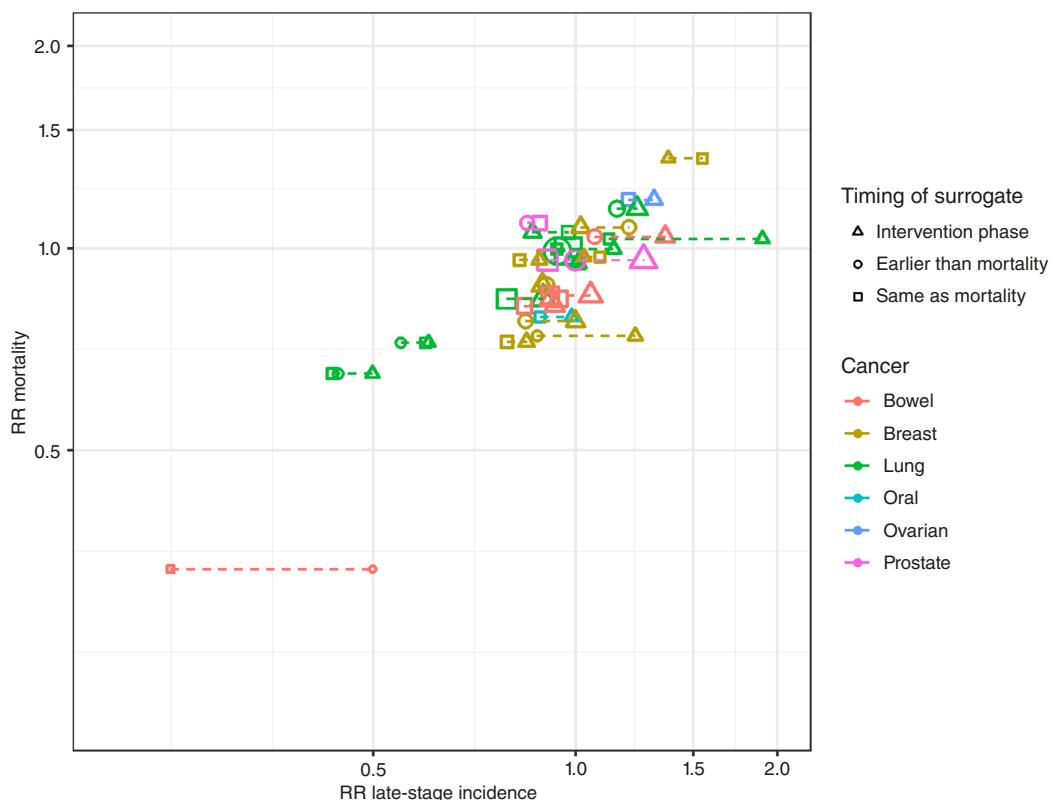
Analysis shown in **Fig. 2C** identified that the association between the screening effect on late-stage cancer incidence and on cancer-specific mortality overall, which was dominated by breast, colorectal, and lung cancer screening trials, was also observed across different classes of cancer screening tests. These associations were, as expected, weaker for the relative proportions of cancers diagnosed at a late stage (**Fig. 2D**). However, liver cancer again emerged as an outlier. The two included liver cancer screening trials (combined) found only a small and nonsignificant effect on liver cancer mortality but a significant and substantial effect on late-stage liver cancer incidence (**Fig. 2E**). This was driven by the Qidong trial. A similar pattern emerged from the analysis of the z-statistics associated with each meta-analysis (**Fig. 2F**). The plot shows that screening tests that have been implemented in national programs have strong meta-analytic evidence of benefit from both late-

**Figure 2.**

Bubble plots to evaluate the strength of association between the trial-level effect of screening on cancer mortality and the effect on either incidence of late-stage cancer or the proportion of cancers diagnosed at late stage. These are the primary results from the meta-analysis. **A**, Incidence of late-stage cancer. **B**, Proportion of cancers diagnosed at late stage. Each data point in **A** and **B** is a trial arm comparison. **C**, Incidence of late-stage cancer. **D**, Proportion of cancers diagnosed at late stage. Each data point in **C** and **D** is from a class of screening test. The same type of plot is used for **A–D**. **E**, Incidence of late-stage cancer. The scales ($1 - RR$) from **C** are inverted and SEs are added to better show uncertainty. **F**, Associated z-statistics from **E**. Each data point in **E** and **F** is from a class of screening test. In all plots, the bubble sizes are proportional to the inverse variance of the screening effect on the mortality outcome. a , the intercept from the regression analysis; b , the regression coefficient from the regression analysis; p , the estimated correlation coefficients. Dotted lines: axes of no screening effect on late-stage and mortality outcomes. Dashed line: diagonal of equal screening effect on late-stage and mortality outcomes. FOBT, fecal occult blood test; FS/C, flexible sigmoidoscopy/colonoscopy; LDCT, low-dose CT.

**Figure 3.**

Bubble plots to evaluate the strength of association between the trial-level effect of screening on cancer mortality and the effect on either incidence of late-stage cancer or the proportion of cancers that are late stage, separately for bowel, breast, and lung cancer screening trials. The figure provides more details on the associations between the screening effect on mortality and the screening effect on the late-stage incidence for cancers with the largest numbers of trials and additionally stratifies by the type of the screening test within each cancer type. Plots A–C are for incidence of late-stage cancer and D–F for proportion of cancers that are late stage. a, the intercept from the regression analysis; b, the regression coefficient from the regression analysis; ρ , the estimated correlation coefficient. FOBT, fecal occult blood test; LDCT, low-dose CT.

**Figure 4.**

Exploratory analysis of the effect of the timing of the reporting of late-stage incidence on the association with the mortality outcome. The figure includes trials that reported their late-stage outcomes at least twice [at intervention phase (triangles), earlier than mortality (circles), and/or at mortality (squares)]. Each trial is then represented by two or up to three points connected by a straight line drawn at the main estimate of the screening effect on cancer-specific mortality; this line shows how the estimates of the screening effect on the late-stage incidence within a specific trial changed over time. Size of the points corresponds to precision of the mortality result (inverse variance of the estimate). Colors denote types of cancer (please see the legend).

stage incidence and cancer-specific mortality endpoints (e.g., mammography for breast cancer and low-dose CT for lung cancer screening). Some of those tests that have not been recommended have strong evidence of lack of substantial benefit on both cancer incidence and mortality endpoints (e.g., cancer antigen 125 for ovarian cancer and chest X-ray for lung cancer screening). Thus, late-stage cancer incidence would have provided similar inference across several classes of cancer screening test trialed to date, except for liver cancer screening.

Using the 95% CI of the effect on the late-stage incidence as a forecast to predict the effect on cancer-specific mortality in a trial

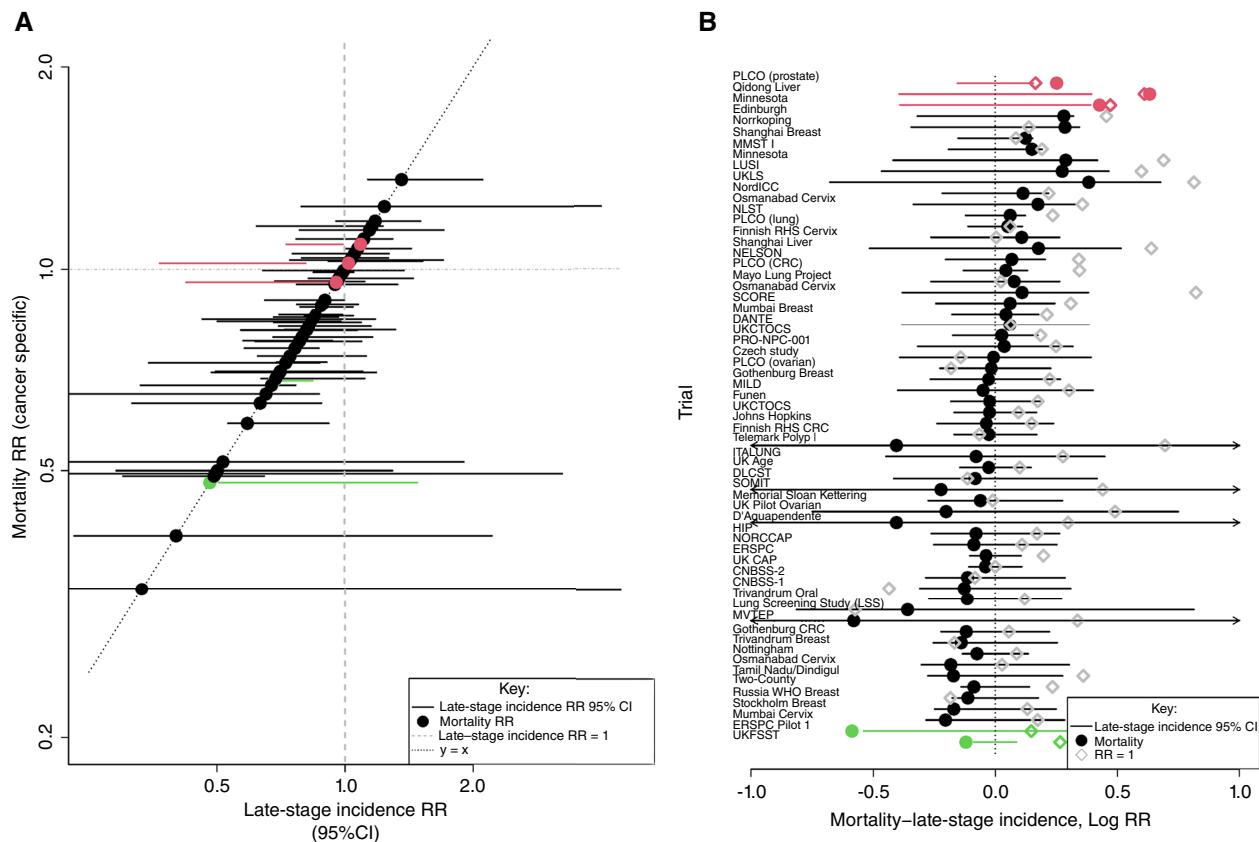
Figure 5 shows that the 95% CI of the RR for the incidence of late-stage cancer included the observed cancer-specific mortality RR (at the “main” timepoint) in 56 of 61 (92%) of trial comparisons; mortality was greater than the 95% CI range in three of 61 trial comparisons and less than the 95% CI range in two of 61 trial comparisons. This pattern was observed across all forecast intervals, and forecast calibration could not be rejected ($P = 0.35$). However, we also note that CIs were wide in several trials because of small sample size, limiting information. Among these, when the 95% CI of the RR for the incidence of late-stage cancer was less than 1 (suggesting benefit of the intervention), it included the observed mortality RR in 16 of 19 (84%) of trial comparisons. The clearest “outlier” from the five trials (with mortality observed

outside the 95% CI from late-stage incidence) was the Qidong liver trial. With regard to the other four trials, the mortality outcome in the prostate part of the US Prostate, Lung, Colorectal, and Ovarian Cancer Screening trial was within the 95% CI range with extended follow-up (26); the Minnesota bowel screening trial (biennial fecal occult blood test) would have been within the range had we chosen Dukes stage C + D as the surrogate (not Dukes D only); in the European Randomised Study of Screening for Prostate Cancer trial pilot, we used Gleason 7+ rather than our preferred 8+ because it was not reported; and the UK Flexible Sigmoidoscopy Screening Trial showed a larger effect on mortality than predicted based on the 95% CI range for incidence, as might be anticipated for screening tests that also prevent cancer (not tabulated). Similar results were observed in the subgroup of 15 trial arm comparisons, with late-stage incidence reported at our ideal “midpoint” time-point (Supplementary Methods and Tables, Section 4; and Supplementary Figure S4).

Discussion

Principal findings

Data from 57 randomized controlled trials in cancer screening showed a clear trial-level association between the screening effects on the incidence of late-stage cancer and the effects on cancer-specific

**Figure 5.**

Exploratory evaluation of the utility of the 95% CI RR for late-stage incidence as a “forecast” of the final trial result for cancer-specific mortality. **A**, The observed RR for mortality (point estimates) versus the “predicted” 95% CI for the RR based on the observed incidence of late-stage cancer in the intervention versus the control arm. **B**, The data from **A** are transformed so that observed – predicted mortality is shown by trial, ordered by where on the prediction the observed mortality fell (mortality at top end of range at top and bottom end of range at bottom of the plot). Estimates marked red are from trials in which the observed mortality exceeds predicted range based on the late-stage outcome; those marked green are trials in which the observed mortality was lower than predicted; and observed mortality RR was included within the 95% CI range of the late-stage outcome for all the other trials shown in black. Wide confidence intervals are due to inclusion of smaller trials.

mortality. Despite substantial differences between the trials in their designs and contexts in which they were undertaken, this overall association was observed across cancer types, particularly for cancer types for which most trials have been run (lung, breast, and bowel). Exploratory analysis suggested that for the majority of trials, the RR associated with cancer mortality was within the 95% CI of the RR for late-stage incidence. There was also consistency in evaluation of the utility of screening tests based on the surrogate or mortality in a screening test meta-analysis context. However, most trials did not report the incidence of late-stage cancer in advance of mortality. Furthermore, liver cancer trials were an outlier in which the incidence of late-stage cancer was a poor surrogate for mortality because (in these trials) there was not enough difference in survival by stage at detection and even stage I cancers had abysmal survival.

We did not find evidence to support the use of the relative proportion of cancers diagnosed at a late stage as a surrogate for cancer mortality. This may be partly due to susceptibility to cancers detected early which may progress very slowly; a high number of such cancers in the denominator would reduce the proportion of cancers that are late stage even though their treatment would not

improve cancer-specific mortality within a typical observation period. Furthermore, the relative proportion of cancers diagnosed at a late stage does not respect randomization since the denominator is a post-randomization event. We and others (16) therefore do not recommend using this measure as a trial endpoint (neither primary nor secondary) in cancer screening trials.

Strengths and weaknesses of the study

A key strength of our study is that we worked according to a pre-published protocol and a prespecified statistical analysis plan, and our study is the most comprehensive systematic review in this area (Supplementary Methods and Tables, Section 1). However, our analysis shares several limitations with all previous meta-analyses (17–20). As far as outcome surrogacy assessment is concerned, the quality of the reported data varied between trials: only 14 of 57 trials reported late-stage disease outcomes consistently and at our “primary” timepoint between the end of their intervention phases and the main mortality endpoint for which a validated surrogate would have the greatest utility for policy making; 42 of 57 trials did not report data on (differences in) the offered cancer treatment or treatment adherence by arm; and 40 of 57 trials either reported substantial

proportions of individuals with missing data on the stage at diagnosis, causes of deaths, or differences therein between the arms or the outcomes or did not report enough detail to determine the completeness of the data on these items (Supplementary Methods and Tables, Section 5, Supplementary Table S19). These issues, compounded by the differences in trial designs, may have influenced the observed relationships between the late-stage and mortality outcomes. Finally, the main analysis focuses on correlation coefficients which have a number of weaknesses for evaluating surrogate endpoints. In particular, stronger correlation is expected for cancer types with a greater differential in survival between early and late stages, irrespective of how much of the screening effect is mediated by the surrogate (6).

Comparison with the literature and policy implications

Two previous meta-analyses by Feng and colleagues (19) and Dai and colleagues (20) addressed the issue of outcome surrogacy but included fewer trials (39 and 33, respectively). Overall, estimated correlation between late-stage cancer incidence and cancer-specific mortality reported in the earlier publications is broadly consistent with our findings (Supplementary Methods and Tables, Section 1, Supplementary Table S1). This is unsurprising as many of the same trials feature in all analyses, and the number of screening trials conducted is modest; for a detailed description of the overlap between the three meta-analyses in terms of the included trials, see Supplementary Methods and Tables, Section 1, Supplementary Table S2. However, the two previous meta-analyses reached opposite conclusions about the suitability of trial outcome surrogacy, largely based on the data from the same trials. Some of this is likely to be due to subjective nature of data interpretation and not the data (or facts) themselves (27). But to an extent, differences in the conclusions are likely also related to review eligibility criteria, definition of late stage, differences in statistical methodology (e.g., weighted vs. unweighted analysis and methods for CIs), and choice of summary measures. In addition, there are differences in the choice of the timing of mortality effect, with Dai and colleagues choosing the latest timepoint possible, suggesting that (in general if not always) it is more likely to represent the true screening effect on mortality; Feng and colleagues used earlier mortality endpoints, but secondary analyses using later mortality results were more in agreement. They may also be related to the inclusion of smaller trials that cause considerable noise despite contributing to the totality of the evidence (Fig. 2) and the variation in the observed screening effects over time within each trial (Fig. 4). Interpretation in both studies focused on the point estimates of correlation coefficients and R^2 values rather than their CIs.

One of the most visible differences between the available meta-analyses is the definition of cancer-specific late-stage outcomes. Feng and colleagues (19), for example, attempted to define late-stage cancer as stage III+ across all cancer types and concluded against using this as a surrogate in breast screening trials. We might have drawn the same conclusion based on our sensitivity analysis (Supplementary Methods and Tables, Section 4, Supplementary Table S17). In contrast, all other meta-analyses that covered breast cancer screening trials (17, 18, 20) and our primary analysis used broader definitions of late-stage breast cancer including stage II+ and its equivalents and provided considerably more supportive evidence of outcome correlations. Stage III+ only represents a small proportion of diagnosed cases and will be more common in non-attenders in the screening arm (11, 20), so the observed reductions in breast cancer mortality in trials (28) are likely to have, at least in part, been derived from a reduction in stage II cases. Similar examples showing the impact of the definition of late-

stage diagnoses can be found for other cancers. In the context of screening for ovarian cancer, the trialists behind the UK Collaborative Trial of Ovarian Cancer Screening commented on the significant reduction in stage IV cases (*post hoc* analysis) and a nonsignificant reduction in mortality (primary endpoint) due to ovarian cancer and concluded that late stage is not a suitable surrogate endpoint in this case (29). However, all three multi-cancer meta-analyses considered stages III + IV as late-stage cancer, which seems to account for a considerable fraction of the observed mortality endpoints across the three ovarian screening trials included in the analyses.

These examples all reflect the lack of an established framework for the evaluation of outcome surrogacy in cancer screening and show that a concerted effort to further develop and optimize methodologic designs and validate a framework to appraise the risk of bias in the studies contributing the evidence would be beneficial (16). The development of both has been advanced for decades in the field of (cancer) treatment (30–34). However, in this setting, there are many more trials run for each class of treatment than are available for cancer screening. Therefore, it is likely that it will always be difficult to apply the methods from this field directly to cancer screening. Indeed, our review considered all cancer screening randomized controlled trials to report cancer-specific mortality since the 1960s and still included only 57 trials. In this context, making more efficient use of individual data from trials could be beneficial, including through widening access to individual-level trial data. Whether individual-level or aggregate-level data are undertaken, however, analyses will need to be tailored to the period within disease progression that offers the greatest opportunity to favorably affect screening outcomes as opposed to a common-stage endpoint across all cancers.

The topic of mortality outcome surrogacy in cancer screening has received renewed attention because of the emergence of new multi-cancer early detection (MCED) tests. In particular, a randomized controlled clinical trial of 140,000 participants is currently ongoing in England (NHS-Galleri trial, NCT05611632) to evaluate a cell-free DNA test (35). In this trial, the primary endpoint is a reduction in the number of cancers (of any type) diagnosed at stage III and IV. Most of the discussion in our article has been on screening tests designed to reduce mortality for individual cancer types and relates to a large extent to imaging tests which may detect a different stage distribution of cancer to MCEDs. Primary analysis in trials for MCEDs will use aggregate endpoints across multiple cancers. Our meta-analysis results do not rule out that late-stage cancer incidence might be considered as a surrogate outcome when combining multiple cancer types (Figs. 2, 3, and 5). However, there are caveats. The natural history of many cancers is complex with many subtypes of different prognosis and treatment effectiveness by stage, making it more challenging to predict when late-stage cancer is a sufficient surrogate for mortality. In our data, there were no sufficiently large trials of MCEDs, in particular using cell-free DNA technology. Biologically, positive screening results on cell-free DNA tests are expected to be associated with poorer cancer prognosis given the stage, and data support this (36). Therefore, there is a risk that a reduction in advanced cancer incidence does not translate into a reduction in mortality (37). Studies to evaluate cancer-specific survival by stage and MCED test positivity could provide insights into this issue, as well as trials in which modeling supports an evaluation of advanced stage and mortality reductions (38).

Lastly, the Chinese Qidong liver screening trial was an outlier with important implications for evidence on the surrogacy of late-stage cancer incidence for cancers with abysmal prognosis

across all stages. It highlights that the relative incidence of late-stage cancer might not be a good surrogate for long-term mortality benefits for cancers in which there is little difference in survival between early and late stages. On this note, the trialists in the Qidong study also considered a Poisson regression model using their individual-level data, not available to us. They estimated that the rate ratio for death in the screening group was 0.83 (95% CI, 0.68–1.03) relative to the control group. These results are consistent with a small benefit on mortality that ultimately was not sustained because of the very poor prognosis across all stage groups in the trial population. One may also note that in this trial, 2-year survival was under 10% (in both arms), so the gain of using advanced stage, even if it worked, would be small. At the time that one could have used advanced stage as an outcome, there would already have been many deaths from individuals diagnosed with early-stage cancers. In general, this example shows that there might be some types of cancers detected by an MCED test for which the stage shift gained through screening is inconsequential for an improved outcome. Such cases should not deter researchers from investigating, for example, a refined surrogate that could be a useful early endpoint for a reduction in the incidence of advanced disease for most cancer types detectable by an MCED test.

Conclusion

A reduction in the incidence of late-stage cancer diagnoses was correlated with a reduction in cancer-specific mortality across the majority of previous randomized controlled trials of cancer screening. In other words, a trial-level relationship between late-stage cancer incidence and mortality was evident. However, the evidence was based on a correlation of a limited range of cancer types and screening tests and limited data on late-stage incidence effects at an earlier timepoint than mortality results (but after the screening period had ended). Therefore, although the evidence supports further investigations into utilizing tumor characteristics as surrogates in future evaluations of new cancer screening tests, and acting now to develop methodologic guidance for including surrogates into future study designs, the interpretation of these data do not allow us to conclude that the incidence of late-stage disease can reliably replace disease-specific mortality for all cancers. Given the slow pace of starting and completing cancer screening trials, it would be prudent for trialists to act now to collect relevant stage and other prognostic tumor features so that these investigations can advance and the potential would exist to act on these findings if they were sufficiently persuasive.

Authors' Disclosures

M. Rebolj reports grants from Public Health England and UK National Screening Committee and personal fees and other support from Hologic outside the submitted work, as well as other support from various expert groups (member) providing advice to the English Cervical Screening Programme and human papillomavirus test manufacturers (meetings). A.R. Brentnall reports other support from the UK National Screening Committee Research and

Methodology subgroup (member). J. Geppert reports grants from Cancer Research UK and National Institute for Health and Care Research during the conduct of the study, as well as grants from National Institute for Health and Care Research outside the submitted work. N. Kouppa reports grants from Cancer Research UK during the conduct of the study. B. Shinkins reports grants from Cancer Research UK during the conduct of the study, as well as other support from the UK National Screening Committee (member). K. Freeman reports grants from Cancer Research UK during the conduct of the study, as well as grants from National Institute for Health and Care Research (ATHENA-M, WM-ESG, and ENCIPHER) outside the submitted work. C. Stinton reports grants from Cancer Research UK during the conduct of the study, as well as other support from the UK National Screening Committee's Adult Reference Group (member). S. Johnson reports grants from Cancer Research UK during the conduct of the study. R.A. Smith reports support from the Center for Early Cancer Detection Science at the American Cancer Society (director), from the American Cancer Society MCED Initiative (lead) and from the Steering Committee for the National Cancer Institute's Tomosynthesis Mammographic Imaging Screening Trial (member). P. Sasieni reports grants from Cancer Research UK during the conduct of the study, as well as personal fees from GRAIL outside the submitted work. S.M. Janes reports grants from GRAIL Inc. outside the submitted work. R. Etzioni reports grants from NCI during the conduct of the study. S. Duffy reports occasional support from GRAIL (consultant) and Roche (consultant). S. Taylor-Phillips reports grants from Cancer Research UK and National Institute for Health and Care Research Professorship during the conduct of the study, as well as other support from the UK National Screening Committee (member) and from the UK National Screening Committee Research and Methodology subgroup (chair). No disclosures were reported by the other author.

Acknowledgments

The authors wish to thank a patient and public representative, Ms Eleanor Cozens, for her contributions to drafts of the funding applications and her comments on the draft paper. The project was supported by (i) Cancer Research UK, reference EDDCTR-2022/100001 awarded to Dr Matejka Rebolj and Prof S. Taylor-Phillips and references C8162/A27047 and CTUQQR-Dec22/100005 awarded to Prof P. Sasieni. (ii) National Institute for Health and Care Research (NIHR), reference NIHR302434 awarded to Prof S. Taylor-Phillips. A.R. Brentnall, S. Johnson, B. Shinkins, R.A. Smith, P. Sasieni, and S. Duffy: none to be declared. M. Rebolj: Cancer Research UK (reference: C8162/A27047). J. Geppert: Cancer Research UK (reference: EDDCTR-2022/100001) and NIHR (reference: NIHR302434). N. Kouppa: Cancer Research UK (references: EDDCTR-2022/100001 and CTUQQR-Dec22/100005). K. Freeman, C. Stinton, M.J. Randell, and S. Taylor-Phillips: NIHR (reference: NIHR302434). S.M. Janes: Cancer Research UK (EDDCPGM/100002). R. Etzioni: NCI (reference: R35CA274442). The funders reviewed the grant application but had no editorial role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the article for publication. The researchers worked independently from funders. All authors, external and internal, had full access to all of the data (including statistical reports and tables) in the study and can take responsibility for the integrity of the data and the accuracy of the data analysis is also required.

Note

Supplementary data for this article are available at *Cancer Epidemiology, Biomarkers & Prevention* Online (<http://cebp.aacrjournals.org/>).

Received February 12, 2025; revised May 15, 2025; accepted July 18, 2025; posted first July 22, 2025.

References

- Etzioni R, Gulati R, Patriotis C, Rutter C, Zheng Y, Srivastava S, et al. Revisiting the standard blueprint for biomarker development to address emerging cancer early detection technologies. *J Natl Cancer Inst* 2024;116:189–93.
- Raoof S, Lee RJ, Jajoo K, Mancias JD, Rebbeck TR, Skates SJ. Multi-cancer early detection technologies: a review informed by past cancer screening studies. *Cancer Epidemiol Biomarkers Prev* 2022;31:1139–45.
- Cuzick J, Cafferty FH, Edwards R, Møller H, Duffy SW. Surrogate endpoints for cancer screening trials: general principles and an illustration using the UK Flexible Sigmoidoscopy Screening Trial. *J Med Screen* 2007;14:178–85.
- Pisano ED. Tomosynthesis for breast cancer screening: the important contribution the tomosynthesis mammographic imaging screening trial will make to our knowledge of breast cancer screening. *J Breast Imaging* 2019;1:23–4.

5. Owens L, Gogebakan KC, Menon U, Gulati R, Weiss NS, Etzioni R. Short-term endpoints for cancer screening trials: does tumor subtype matter? *Cancer Epidemiol Biomarkers Prev* 2023;32:741–3.
6. Owens L, Gulati R, Etzioni R. Stage shift as an endpoint in cancer screening trials: implications for evaluating multicancer early detection tests. *Cancer Epidemiol Biomarkers Prev* 2022;31:1298–304.
7. Elkin EB, Hudis C, Begg CB, Schrag D. The effect of changes in tumor size on breast carcinoma survival in the U.S.: 1975–1999. *Cancer* 2005;104:1149–57.
8. Cancer Research UK. Survival for breast cancer. [cited 2024 Aug 5]. Available from: <https://www.cancerresearchuk.org/about-cancer/breast-cancer/survival>.
9. Cancer Research UK. Survival for bowel cancer. [cited 2024 Aug 5]. Available from: <https://www.cancerresearchuk.org/about-cancer/bowel-cancer/survival>.
10. National Cancer Institute. Surveillance, Epidemiology, and End Results Program. Cancer stat facts: colorectal cancer. Bethesda (MD): National Cancer Institute. [cited 2024 Aug 5]. Available from: <https://seer.cancer.gov/statfacts/html/colorect.html>.
11. National Cancer Institute. Surveillance, Epidemiology, and End Results Program. Cancer stat facts: female breast cancer. Bethesda (MD): National Cancer Institute. [cited 2024 Aug 5]. Available from: <https://seer.cancer.gov/statfacts/html/breast.html>.
12. Cancer Research UK. Survival for primary liver cancer. [cited 2024 Aug 5]. Available from: <https://www.cancerresearchuk.org/about-cancer/liver-cancer/survival>.
13. Cancer Research UK. Survival for pancreatic cancer. [cited 2024 Aug 5]. Available from: <https://www.cancerresearchuk.org/about-cancer/pancreatic-cancer/survival>.
14. National Cancer Institute. Surveillance, Epidemiology, and End Results Program. Cancer stat facts: pancreatic cancer. Bethesda (MD): National Cancer Institute. [cited 2024 Aug 5]. Available from: <https://seer.cancer.gov/statfacts/html/pancreas.html>.
15. National Cancer Institute. Surveillance, Epidemiology, and End Results Program. Cancer stat facts: liver intrahepatic bile duct cancer. Bethesda (MD): National Cancer Institute. [cited 2024 Aug 5]. Available from: <https://seer.cancer.gov/statfacts/html/livibd.html>.
16. Webb AB, Berg CD, Castle PE, Crosby D, Etzioni R, Kessler LG, et al. Considerations for using potential surrogate endpoints in cancer screening trials. *Lancet Oncol* 2024;25:e183–92.
17. Tabár L, Yen AM-F, Wu WY-Y, Chen SL-S, Chiu SY-H, Fann JC-Y, et al. Insights from the breast cancer screening trials: how screening affects the natural history of breast cancer and implications for evaluating service screening programs. *Breast J* 2015;21:13–20.
18. Autier P, Héry C, Haukka J, Boniol M, Byrnes G. Advanced breast cancer and breast cancer mortality in randomized controlled trials on mammography screening. *J Clin Oncol* 2009;27:5919–23.
19. Feng X, Zahed H, Onwuka J, Callister MEJ, Johansson M, Etzioni R, et al. Cancer stage compared with mortality as end points in randomized clinical trials of cancer screening: a systematic review and meta-analysis. *JAMA* 2024;331:1910–7.
20. Dai JY, Georg Luebeck E, Chang ET, Clarke CA, Hubbell EA, Zhang N, et al. Strong association between reduction of late-stage cancers and reduction of cancer-specific mortality in meta-regression of randomized screening trials across multiple cancer types. *J Med Screen* 2024;31:211–22.
21. Ciani O, Manyara AM, Davies P, Stewart D, Weir CJ, Young AE, et al. A framework for the definition and interpretation of the use of surrogate endpoints in interventional trials. *EClinicalMedicine* 2023;65:102283.
22. Anderson TW, Darling DA. Asymptotic theory of certain “goodness of fit” criteria based on stochastic processes. *Ann Math Statist* 1952;23:193–212.
23. Pasek J, Tahk A, Culter G, Schwemmle M. Weights: weighting and weighted statistics. [cited 2024 Aug 19]. Available from: <https://cran.r-project.org/web/packages/weights/index.html>.
24. Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Softw* 2010;36:1–48.
25. Davison AC, Hinkley DV. Bootstrap Methods and Their Applications. Cambridge University Press; 1997.
26. Pinsky PF, Miller E, Prorok P, Grubb R, Crawford ED, Andriole G. Extended follow-up for prostate cancer incidence and mortality among participants in the Prostate, Lung, Colorectal and Ovarian randomized cancer screening trial. *BJU Int* 2019;123:854–60.
27. Silberzahn R, Uhlmann EL, Martin DP, Anselmi P, Aust F, Awtrey E, et al. Many analysts, one data set: making transparent how variations in analytic choices affect results. *Adv Methods Practices Psychol Sci* 2018;1:337–56.
28. Marmot MG, Altman DG, Cameron DA, Dewar JA, Thompson SG, Wilcox M. The benefits and harms of breast cancer screening: an independent review. *Br J Cancer* 2013;108:2205–40.
29. Menon U, Gentry-Maharaj A, Burnell M, Singh N, Ryan A, Karpinskyj C, 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* 2021;397:2182–93.
30. Buyse M, Molenberghs G, Burzykowski T, Renard D, Geys H. The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics* 2000;1:49–67.
31. Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. *Pharm Stat* 2006;5:173–86.
32. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med* 1989;8:431–40.
33. Freedman LS, Graubard BI, Schatzkin A. Statistical validation of intermediate endpoints for chronic diseases. *Stat Med* 1992;11:167–78.
34. Baker SG. A simple meta-analytic approach for using a binary surrogate endpoint to predict the effect of intervention on true endpoint. *Biostatistics* 2006;7:58–70.
35. Neal RD, Johnson P, Clarke CA, Hamilton SA, Zhang N, Kumar H, et al. Cell-free DNA-based multi-cancer early detection test in an asymptomatic screening population (NHS-galleri): design of a pragmatic, prospective randomised controlled trial. *Cancers (Basel)* 2022;14:4818.
36. Chen X, Dong Z, Hubbell E, Kurtzman KN, Oxnard GR, Venn O, et al. Prognostic significance of blood-based multi-cancer detection in plasma cell-free DNA. *Clin Cancer Res* 2021;27:4221–9.
37. Callister MEJ, Crosbie EJ, Crosbie PAJ, Robbins HA. Evaluating multi-cancer early detection tests: an argument for the outcome of recurrence-updated stage. *Br J Cancer* 2023;129:1209–11.
38. Sasieni P, Smittenaar R, Hubbell E, Broggio J, Neal RD, Swanton C. Modelled mortality benefits of multi-cancer early detection screening in England. *Br J Cancer* 2023;129:72–80.