Public Health Policy

Does screening for disease save lives in asymptomatic adults? Systematic review of meta-analyses and randomized trials

Natzus Saquib,1 Juliann Saquib1 and John PA Ioannidis1,2,3*

1Stanford Prevention Research Center, Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA 2Department of Health Research and Policy, Stanford University School of Medicine, Stanford, CA, USA and 3Department of Statistics, Stanford University School of Humanities and Sciences, Stanford, CA, USA

*Corresponding author. Stanford Prevention Research Center, Medical School Office Building, Room X306, 1265 Welch Rd, Stanford, CA 94305, USA. E-mail: jioannid@stanford.edu

Accepted 16 June 2014

Abstract

Background: Several popular screening tests, such as mammography and prostate-specific antigen, have met with wide controversy and/or have lost their endorsement recently. We systematically evaluated evidence from randomized controlled trials (RCTs) as to whether screening decreases mortality from diseases where death is a common outcome.

Methods: We searched three sources: United States Preventive Services Task Force (USPSTF), Cochrane Database of Systematic Reviews, and PubMed. We extracted recommendation status, category of evidence and RCT availability on mortality for screening tests for diseases on asymptomatic adults (excluding pregnant women and children) from USPSTF. We identified meta-analyses and individual RCTs on screening and mortality from Cochrane and PubMed.

Results: We selected 19 diseases (39 tests) out of 50 diseases/disorders for which USPSTF provides screening evaluation. Screening is recommended for 6 diseases (12 tests) out of the 19. We assessed 9 non-overlapping meta-analyses and 48 individual trials for these 19 diseases. Among the results of the meta-analyses, reductions where the 95% confidence intervals (CIs) excluded the null occurred for four disease-specific mortality estimates (ultrasound for abdominal aortic aneurysm in men; mammography for breast cancer; fecal occult blood test and flexible sigmoidoscopy for colorectal cancer) and for none of the all-cause mortality estimates. Among individual RCTs, reductions in disease-specific and all-cause mortality where the 95% CIs excluded the null occurred in 30% and 11% of the estimates, respectively.

Conclusions: Among currently available screening tests for diseases where death is a common outcome, reductions in disease-specific mortality are uncommon and reductions in all-cause mortality are very rare or non-existent.
Introduction

Screening for disease is a key component of modern health care. The rationale is simple and attractive—to detect diseases early in asymptomatic individuals and to treat them in order to reduce morbidity, mortality and the associated costs. However, the role of screening often comes into question. Some high-profile controversies have appeared lately in this regard. For example, for breast cancer, the United States Preventive Services Task Force (USPSTF) currently recommends against routine mammographic screening for women aged 40–49 years after retracting its previous recommendation in favour of mammography, as the data failed to show that benefit outweighed harm.1 The decision against screening drew sharp criticism from various interest groups including patients who overestimate the benefit of screening.2 Similarly, USPSTF now recommends against screening for prostate cancer in healthy men because harms from prostate specific antigen (PSA) screening exceed the benefit, trials do not show improvement in long-term survival3 and screening carries a high risk of over-diagnosis with adverse consequences. Again, heated debates have been generated around this change of recommendation, both in the scientific and the popular press.

Some screening tests were entrenched in clinical and public health practice before randomized controlled trials (RCTs) became widely used. As the screening agenda encompasses a large number of tests, and new ones are continuously proposed, it is useful to reassess the evidence supporting their use. Our research question is whether recommended screening tests, among asymptomatic adults, have evidence from RCTs on mortality outcomes for diseases where death is a common outcome. In particular, is there evidence of mortality reduction, either disease-specific or all-cause, from screening? To this end, we have compiled and examined systematically the evidence from individual RCTs and meta-analyses thereof for screening tests that have been proposed for detecting major diseases in adults who have no symptoms.

Methods

Eligibility criteria

We assessed the diseases/disorders in adults, which USPSTF grouped in different clinical categories and made screening recommendations. We focused on the ‘Cancer’ and ‘Heart and vascular diseases’ categories, as well as type 2 diabetes mellitus and chronic obstructive pulmonary disease, because mortality is a common outcome for these diseases. We did not include diseases/disorders where mortality is not a common outcome, and that included the following clinical categories: infectious diseases; mental health conditions and substance abuse; metabolic, nutritional and endocrine disorders (except type 2 diabetes); musculoskeletal disorders; injury and violence; vision and hearing disorders; obstetric and gynaecological conditions; and miscellaneous (except chronic obstructive pulmonary diseases).

For the included diseases, we compiled a list of screening tests and assessed which of them are recommended by USPSTF, and whether they have randomized evidence on mortality outcomes. We defined screening as using a specific test on an otherwise asymptomatic, non-diseased population in order to detect a certain disease. We only considered evidence that compared mortality between screening and no-screening control groups. We did not consider screening/testing in already diseased individuals (e.g. patients who have diabetes mellitus or already have some cancer diagnosis).

Search strategies and documentation of evidence

We compiled information from USPSTF, Cochrane Database of Systematic Reviews and PubMed. We documented current recommendations and the corresponding level of evidence from USPSTF. We gathered meta-analytic evidence on screening from Cochrane and PubMed. In addition, we collected from PubMed information about individual RCTs on screening which had not been included in a published meta-analysis.

In the USPSTF website, we reviewed the documentation of RCT evidence for screening for each disease in adults (last update: January 2014).

We searched the Cochrane Database of Systematic Reviews using the search term ‘screening’ in title, abstract or keywords. We documented all systematic reviews on screening tests that had at least one eligible RCT or meta-analysis of several RCTs with mortality outcomes (last update: January 2014).
We searched PubMed using the search terms ‘screen’ or ‘screening’ or ‘testing’ in the title and ‘death’ or ‘mortality’ or ‘survival’ in title, abstract or keywords. We ran two searches for articles published in English; one was limited to RCTs and the other to meta-analyses (using ‘type of publication’ limit) (last update: January 2014).

Study selection: meta-analyses and individual trials

We screened, identified and organized the eligible meta-analyses by disease and the associated screening test (Cochrane, PubMed). When several meta-analyses were eligible on the same disease and screening test, we selected the most comprehensive ones (more trials, more long-term follow-up in included trials).

We screened and identified the eligible individual trials (PubMed). We organized the list first by screening test, then by trial name and then by year of publication. If there were more than one citation per trial, we selected the most recent publication. Simultaneously, we compiled a list of trials that were in the selected meta-analyses. We cross-checked the individual trials in PubMed with those in the selected meta-analyses to determine how many trials were in common. Finally, we compiled a list of individual trials—including those that were in common and those that were unique to either PubMed or a meta-analysis. Finally, if no meta-analysis was available, but multiple individual trials existed for a given screening test, we performed the meta-analysis ourselves using inverse variance synthesis with fixed effects.

Data extraction

From USPSTF, we documented the following for each disease: screening tests, recommendation statement, category of evidence, presence or absence of RCT evidence, and the specific population for whom the recommendation is applicable.

For each included meta-analysis of RCTs (Cochrane or PubMed) and single RCT (PubMed), we extracted the following: disease; screening intervention assessed; number of RCTs analysed; use of stratified analysis (yes, no) and, if so, types of strata; number of disease-specific deaths/total sample and disease-specific mortality risk estimates; and number of total deaths/total sample and all-cause mortality risk estimates. Data were extracted by two co-authors and any disagreement was resolved with discussion with the senior (third) author.

Presentation of mortality outcomes

Disease-specific mortality was defined as death attributed to the disease in question and all-cause mortality was defined as death from any cause; in both instances, the denominator was the total sample per randomized group and not those who were detected as diseased. We presented the treatment effect (risk estimates with 95% confidence interval) as they were reported in the original RCTs or meta-analyses.

Results

Evaluated screening tests in USPSTF

USPSTF provides evaluation of screening for 19 diseases where mortality is a common outcome (cancer n = 12, heart and vascular diseases n = 5, type 2 diabetes, chronic obstructive pulmonary disease) (Supplementary Table 1, available as Supplementary data at IJE online).

Overall, 39 different screening tests are addressed for these 19 diseases. Screening is recommended for 6 of the 19 diseases (for a total of 12 recommended tests out of 14 available tests for these 6 diseases). Randomized evidence with a mortality outcome is cited for only 5 diseases (breast, cervical and colorectal cancer, abdominal aortic aneurysm and type 2 diabetes) for 11 recommended tests among 13 assessed by USPSTF (Figure 1).

Randomized trials with data on mortality are not available for one disease (hypertension) where screening is recommended (one out of one test is recommended). Further, BRCA-gene mutation screening for breast cancer and colonoscopy for colorectal cancer do not have randomized trials on their effectiveness, but they are both currently recommended for adults with a family history.

Screening is not recommended for the remaining 13 diseases where there are 25 available tests; of those, randomized trials with data on mortality are available only for 7 tests on 4 diseases: lung, oral, ovarian and prostate cancer. For breast cancer, screening for BRCA and mammography are recommended but clinical and self-examination of the breast are not recommended. Randomized evidence exists for mammography, clinical and self-examination but not for BRCA (there is a trial on genetic counselling but not for the screening test per se).

![Figure 1. Flow diagram for randomized controlled trial (RCT) evidence from the United States Preventive Services Task Force (USPSTF).](image-url)
Evaluated meta-analyses on screening tests in Cochrane and PubMed

The search produced 595 items in Cochrane and 125 items in PubMed; of those, 59 and 85, respectively, were assessed in full text. In Cochrane and PubMed, 12 and 44 meta-analyses, respectively, met the eligibility criteria; these included 8 Cochrane reviews that had also been presented as journal articles, thus there were 48 different eligible meta-analyses. These 48 meta-analyses were clustered by test and disease to identify the latest, non-overlapping meta-analysis on each topic. Eventually, eight meta-analyses were selected covering eight screening tests for six diseases;\(^5\),\(^6\),\(^8\)–\(^12\) additionally we performed ourselves a meta-analysis of the trials’ data on screening with computer tomography (CT) for lung cancer, as there were several individual trials but no published meta-analyses (Figure 2).

Evaluated individual trials on screening tests in PubMed

The search produced 590 items; 83 records were evaluated further and 40 trials met the inclusion criteria. Of the 40, 28 trials had been included in at least one of the eight eligible meta-analyses mentioned above.\(^13\)–\(^40\) The other 12 trials\(^7\),\(^41\)–\(^51\) found in PubMed included mostly \((n=9)\) trials on topics for which there were no eligible previous meta-analyses; three trials\(^42\),\(^44\),\(^45\) were excluded from the respective meta-analysis because the follow-up time was less than 5 years. Another eight trials\(^52\)–\(^59\) that had been included in the eight eligible meta-analyses were not captured by the PubMed search for trials; these were not picked by our PubMed search because one was in Russian language, two were not tagged as randomized controlled trials by PubMed and five did not have the search terms in their titles. Therefore, a total of 48 eligible RCTs were considered (Figure 3).

Meta-analytic and individual trial evidence by disease

**Abdominal aortic aneurysm.** Eight meta-analyses were found; we used the meta-analysis by Takagi et al.\(^12\) that had included four trials (Chichester,\(^16\) MASS,\(^60\) Viborg,\(^13\) Western Australia\(^35\)) with the longest follow-up (\(\geq 10\) years).

---

**Figure 2.** Flow diagram of meta-analytic search results from Cochrane Library and PubMed. RCT, randomized controlled trial; BSE, breast self-examination; BC, breast cancer; FOBT, fecal occult blood test; CRC, colorectal cancer; PSA, prostate-specific antigen; PC, prostate cancer; LC, lung cancer; US, ultrasound; AAA, abdominal aortic aneurysm; OC, ovarian cancer; FS, flexible sigmoidoscopy.

*Meta-analytic evidence: we conducted the meta-analyses for screening with computerized tomography (CT) scan using data from DANTE, DLCST and MILD trial (see Table 2).*
In addition, MASS\textsuperscript{18} has published extended follow-up data (13 years) after the Takagi meta-analysis. The reason for excluding six meta-analyses was shorter follow-up.\textsuperscript{61–66} The final one was excluded because it evaluated a 30-day mortality following elective surgery for aortic aneurysm.\textsuperscript{67}

**Breast cancer.** Twelve meta-analyses of screening with mammography were found; Gotzsche \textit{et al.}\textsuperscript{5} had reviewed all eight trials (Canada 1980a,b,\textsuperscript{17,68} Edinburgh,\textsuperscript{14} Goteborg,\textsuperscript{15} Malmo,\textsuperscript{19} New York,\textsuperscript{26} Stockholm,\textsuperscript{20} Two-county,\textsuperscript{27} and UK age trial\textsuperscript{[24]} and reported the longest follow-up time (13 years). The other meta-analyses were excluded because they were earlier publications,\textsuperscript{69–71} had fewer trials\textsuperscript{72–76} or shorter follow-up\textsuperscript{77} or selected a particular age group\textsuperscript{78–82} or a sub-type of cancer.\textsuperscript{83} For the Two-county study, Tabar \textit{et al.} presented also disease-specific mortality estimates with longer follow-up (29 years);\textsuperscript{27} the trial’s estimates for all-cause mortality were extracted from Gotzsche \textit{et al.}\textsuperscript{5}

Only one meta-analysis\textsuperscript{6} was found for screening with breast self-examination with two trials (Russia\textsuperscript{56} and Shanghai\textsuperscript{59}) and only a single trial (Mumbai\textsuperscript{7}) for clinical breast examination.

**Cervical cancer.** Two single trials (Tamil Nadu\textsuperscript{46} and Maharashtra\textsuperscript{47} in India) were found on screening with visual inspection, human papilloma virus testing and cytological testing for cervical cancer.

**Colorectal cancer.** Four meta-analyses of screening with fecal occult blood (FOBT) test were found; Hewitson \textit{et al.}\textsuperscript{8} presented four trials (Funen,\textsuperscript{21} Goteborg,\textsuperscript{23} Minnesota\textsuperscript{84} and Nottingham\textsuperscript{25}) with the longest follow-up (11.7 to 18 years); the other three were excluded for including fewer trials\textsuperscript{85} or shorter follow-up time.\textsuperscript{86,87} After the Hewitson meta-analysis was published, the Minnesota study has published a 30-year follow-up.\textsuperscript{49} Two meta-analyses\textsuperscript{9,88} of screening with flexible sigmoidoscopy (single, multiple or in combination with FOBT) were also found; both included five trials: Telemark Polyp Study,\textsuperscript{54} NCCPS,\textsuperscript{31} UK trial,\textsuperscript{29} SCORE,\textsuperscript{40} and PLCO.\textsuperscript{38} One meta-analysis was excluded because it did not provide all-cause mortality estimates.\textsuperscript{88}

**Hepatocellular cancer.** Two reviews of screening with alpha-fetoprotein plus ultrasound were found; Wun \textit{et al.}\textsuperscript{89} included two trials (Toronto,\textsuperscript{90} and Shanghai,\textsuperscript{91}) and Aghoram \textit{et al.}\textsuperscript{92} included three trials (Toronto,\textsuperscript{90} Taiwan\textsuperscript{93} and Shanghai\textsuperscript{21,91}); neither review gave meta-analytical evidence. The Toronto and Taiwan trials evaluated comparative screening and therefore were not included in our evaluation of individual trials. The two reports from the Shanghai trial had discrepant results; the earlier report\textsuperscript{91} did not show benefit with screening [odds ratio (OR) 0.81, 95% confidence interval (CI) = 0.54, 1.22] but the later one\textsuperscript{51} showed benefit [relative risk (RR) = 0.63, 95% CI = 0.41, 0.98]. The later one has been included in the analysis as it had longer follow-up data.
<table>
<thead>
<tr>
<th>Disease</th>
<th>Screening</th>
<th>No. RCTs</th>
<th>Disease-specific death</th>
<th>All-cause death</th>
<th>Risk estimate (95% CI)</th>
<th>Risk estimate (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abdominal aortic aneurysm</td>
<td>Ultrasound</td>
<td>3</td>
<td>221/43211</td>
<td>405/43238</td>
<td>0.55 (0.35, 0.86)</td>
<td>0.98 (0.96, 1.00)</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>Mammography</td>
<td>3</td>
<td>404/119304</td>
<td>572/172649</td>
<td>0.90 (0.79, 1.02)</td>
<td>0.99 (0.95, 1.03)</td>
</tr>
<tr>
<td></td>
<td>Mammography</td>
<td>4</td>
<td>633/170048</td>
<td>746/136889</td>
<td>0.75 (0.67, 0.83)</td>
<td>0.99 (0.97, 1.01)</td>
</tr>
<tr>
<td></td>
<td>Breast self-exam</td>
<td>2</td>
<td>292/190691</td>
<td>295/197844</td>
<td>1.05 (0.90, 1.24)</td>
<td></td>
</tr>
<tr>
<td>Colorectal cancer</td>
<td>Fecal occult blood test (FOBT)</td>
<td>4</td>
<td>1476/172734</td>
<td>1592/156908</td>
<td>0.84 (0.78, 0.90)</td>
<td>1.00 (0.99, 1.01)</td>
</tr>
<tr>
<td></td>
<td>Flexible sigmoidoscopy</td>
<td>5</td>
<td>Not given</td>
<td>Not given</td>
<td>0.71 (0.61, 0.81)</td>
<td>0.99 (0.91, 1.07)</td>
</tr>
<tr>
<td>Lung cancer</td>
<td>Chest X-ray</td>
<td>4</td>
<td>710/42668</td>
<td>629/38635</td>
<td>1.11 (1.00, 1.23)</td>
<td>1.01 (0.94, 1.08)</td>
</tr>
<tr>
<td></td>
<td>Chest X-ray + cytology</td>
<td>2</td>
<td>256/10194</td>
<td>293/10233</td>
<td>0.88 (0.74, 1.03)</td>
<td>1.03 (0.91, 1.15)</td>
</tr>
<tr>
<td></td>
<td>CT scan</td>
<td>3</td>
<td>53/5704</td>
<td>38/4971</td>
<td>1.23 (0.81, 1.87)</td>
<td>1.34 (0.93, 1.93)</td>
</tr>
<tr>
<td>Ovarian cancer</td>
<td>CA-125</td>
<td>2</td>
<td>127/45211</td>
<td>118/45281</td>
<td>1.08 (0.84, 1.38)</td>
<td>1.00 (0.96, 1.06)</td>
</tr>
<tr>
<td>Prostate cancer</td>
<td>Prostate-specific-antigen</td>
<td>5</td>
<td>698/156157</td>
<td>1318/185185</td>
<td>1.00 (0.86, 1.17)</td>
<td>1.00 (0.96, 1.03)</td>
</tr>
</tbody>
</table>

*aMen.*  
*bOptimal trial.*  
*cSub-optimal trial.*  
*dOdds ratio.*  
*eRelative risk or rate ratio.*  
*fHazard ratio.*  
*gWe conducted the meta-analyses for screening with CT scan using data from DANTE, DLCST, and MILD trial (see Table 2).*
<table>
<thead>
<tr>
<th>Disease</th>
<th>Screening test</th>
<th>Individual trial</th>
<th>Disease-specific death</th>
<th>All-cause death</th>
<th>Disease-specific death</th>
<th>All-cause death</th>
<th>Risk estimate (95% CI)</th>
<th>Risk estimate (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Abdominal aortic aneurysm</strong></td>
<td>Ultrasound</td>
<td>Viborg 13</td>
<td>14/633</td>
<td>51/6306</td>
<td>218/633</td>
<td>2234/6306</td>
<td>0.27 (0.15, 0.49) d</td>
<td>0.97 (0.91, 1.03) d</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Western Australia 35,a</td>
<td>18/19352</td>
<td>25/19352</td>
<td>2232/19352</td>
<td>2571/19352</td>
<td>0.72 (0.39, 1.32) b</td>
<td>0.98 (includes 1.00)c</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MASI 18</td>
<td>224/33883</td>
<td>381/33887</td>
<td>13858/33883</td>
<td>14134/33887</td>
<td>0.58 (0.49, 0.69)d</td>
<td>0.97 (0.95, 0.99)d</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chichester 16</td>
<td>19/6333</td>
<td>55/6306</td>
<td>2931/6333</td>
<td>2964/6306</td>
<td>0.34 (0.20, 0.57) d</td>
<td>0.98 (0.93, 1.03)d</td>
</tr>
<tr>
<td><strong>Breast cancer</strong></td>
<td>Mammography</td>
<td>Canada 1980 37,a,d</td>
<td>105/25214</td>
<td>108/25216</td>
<td>413/25214</td>
<td>413/25216</td>
<td>0.97 (0.74, 1.27)c</td>
<td>1.00 (0.87, 1.14)c</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Western Australia 35,a</td>
<td>156/28628</td>
<td>167/26026</td>
<td>734/19711</td>
<td>690/19694</td>
<td>1.02 (0.78, 1.33) c</td>
<td>1.06 (0.96, 1.18)c</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Malmo 19</td>
<td>87/20695</td>
<td>10820783</td>
<td>253721088</td>
<td>259321195</td>
<td>0.75 (0.58, 0.97) c</td>
<td>0.89 (0.83, 0.95)c</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Stockholm 26</td>
<td>218/31000</td>
<td>262/31000</td>
<td>2062/30239</td>
<td>2116/30765</td>
<td>0.82 (0.66, 1.04) c</td>
<td>0.99 (0.94, 1.05)c</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Two-country 27,a,d</td>
<td>66/40318</td>
<td>45/19943</td>
<td>1768/39139</td>
<td>1036/20978</td>
<td>0.73 (0.50, 1.06) c</td>
<td>0.91 (0.85, 0.99)c</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Copenhagen 21</td>
<td>351/77080</td>
<td>367/55985</td>
<td>6034/38568</td>
<td>4829/38942</td>
<td>0.73 (0.59, 0.89) d</td>
<td>1.03 (0.99, 1.08)c</td>
</tr>
<tr>
<td></td>
<td></td>
<td>UK Age trial 24</td>
<td>105/53884</td>
<td>251/106956</td>
<td>1975/106956</td>
<td>960/53884</td>
<td>0.83 (0.66, 1.04) f</td>
<td>0.96 (0.89, 1.04)c</td>
</tr>
<tr>
<td><strong>Cervical cancer</strong></td>
<td>Visual inspection</td>
<td>Tamil Nadu, India 46</td>
<td>83/31343</td>
<td>92/30958</td>
<td>1303/31343</td>
<td>977/30958</td>
<td>0.65 (0.47, 0.89) d</td>
<td>0.87 (0.78, 0.96)c</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Visual inspection Maharashtra India 47,f</td>
<td>135/132979</td>
<td>13/133085</td>
<td>6/34074</td>
<td>6/34142</td>
<td>0.86 (0.60, 1.25)f</td>
<td>0.92 (0.82, 1.00) f</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Human papilloma virus</td>
<td>34/34126</td>
<td>64/31484</td>
<td>0.52 (0.33, 0.83) d</td>
<td>0.89 (0.62, 1.27)d</td>
<td>0.87 (0.78, 0.96)c</td>
<td>1.00 (0.82, 1.00) f</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cytology</td>
<td>54/32058</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cardiovascular disease</strong></td>
<td>Echocardiography</td>
<td>Not given</td>
<td>250/3272</td>
<td>299/3589</td>
<td>880/3272</td>
<td>989/3589</td>
<td>0.91 (0.77, 1.08) d</td>
<td>0.97 (0.89, 1.06)d</td>
</tr>
<tr>
<td><strong>Colorectal cancer</strong></td>
<td>Fecal occult blood test (FOBT)</td>
<td>Funen 21</td>
<td>362/30967</td>
<td>431/30966</td>
<td>12205/30967</td>
<td>12248/30966</td>
<td>0.84 (0.73, 0.96) b</td>
<td>0.99 (0.96, 1.03)b</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Goteborg 23</td>
<td>252/34144</td>
<td>300/34164</td>
<td>10591/34144</td>
<td>10432/34164</td>
<td>0.84 (0.71, 0.99) b</td>
<td>1.02 (0.99, 1.06)b</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Minnesota – annual 49</td>
<td>200/15570</td>
<td>295/15394</td>
<td>11072/15570</td>
<td>10944/15394</td>
<td>0.68 (0.56, 0.82) d</td>
<td>1.00 (0.99, 1.01)c</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Minnesota – biennial 49</td>
<td>237/15587</td>
<td>295/15394</td>
<td>11004/15587</td>
<td>10944/15394</td>
<td>0.78 (0.65, 0.93) c</td>
<td>0.99 (0.98, 1.01) c</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nottingham 25</td>
<td>593/76466</td>
<td>684/76384</td>
<td>20421/76466</td>
<td>20363/76384</td>
<td>0.87 (0.77, 0.97) b</td>
<td>1.00 (0.98, 1.03)b</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Norwegian NCCPS 31</td>
<td>24/13633</td>
<td>99/41092</td>
<td>Not given</td>
<td>Not given</td>
<td>0.73 (0.47, 1.13) d</td>
<td>1.02 (0.98, 1.07)d</td>
</tr>
<tr>
<td></td>
<td>Flexible sigmoidoscopy</td>
<td>Telemark 54</td>
<td>1/400</td>
<td>4/399</td>
<td>62/400</td>
<td>40/399</td>
<td>0.25 (0.03, 2.23)c</td>
<td>1.55 (1.04, 2.30)c</td>
</tr>
<tr>
<td></td>
<td>with or without FOBT</td>
<td>UK trial 29</td>
<td>189/57099</td>
<td>538/112939</td>
<td>6773/57099</td>
<td>13768/112939</td>
<td>0.69 (0.59, 0.82) d</td>
<td>0.97 (0.94, 1.00)d</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SCORE 18</td>
<td>65/17136</td>
<td>83/17136</td>
<td>1202/17136</td>
<td>1233/17136</td>
<td>0.78 (0.56, 1.08) c</td>
<td>0.97 (0.90, 1.05)c</td>
</tr>
<tr>
<td></td>
<td>Multiple flexible sigmoidoscopy</td>
<td>PLCO 18</td>
<td>252/77445</td>
<td>34/77455</td>
<td></td>
<td></td>
<td>0.74 (0.63, 0.87)c</td>
<td></td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Disease</th>
<th>Screening test</th>
<th>Individual trial</th>
<th>Disease-specific death</th>
<th>All-cause death</th>
<th>Disease-specific death</th>
<th>All-cause death</th>
<th>Risk estimate (95% CI)</th>
<th>Risk estimate (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type 2 diabetes</td>
<td>Fasting blood glucose + HbA1c</td>
<td>ADDITION-Cambridge&lt;sup&gt;50&lt;/sup&gt;</td>
<td>Not given</td>
<td>Not given</td>
<td>1532/16047</td>
<td>37774/137</td>
<td>1.26 (0.75, 2.10)</td>
<td>1.06 (0.90, 1.25)</td>
</tr>
<tr>
<td>Hepatocellular carcinoma</td>
<td>Alpha fetoprotein + ultrasound</td>
<td>Shanghai&lt;sup&gt;51&lt;/sup&gt;</td>
<td>32/9373</td>
<td>54/9443</td>
<td>1.01 (0.81, 1.26)</td>
<td>1.01 (0.85, 1.21)</td>
<td>0.63 (0.41, 0.98)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Alpha fetoprotein</td>
<td>Qidong, China&lt;sup&gt;41&lt;/sup&gt;</td>
<td>218/3712</td>
<td>109/1869</td>
<td>1.14 (0.96, 1.36)</td>
<td>1.16 (1.00, 1.35)</td>
<td>1.10 (0.81, 1.51)</td>
<td></td>
</tr>
<tr>
<td>Lung cancer</td>
<td>Chest X-ray</td>
<td>Czech&lt;sup&gt;22&lt;/sup&gt;</td>
<td>247/3171</td>
<td>216/3174</td>
<td>1.13 (0.74, 1.72)</td>
<td>0.98 (0.88, 1.09)</td>
<td>1.01 (0.85, 1.21)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kaiser&lt;sup&gt;53&lt;/sup&gt;</td>
<td>44/5156</td>
<td>42/5557</td>
<td>1.11 (0.95, 1.28)</td>
<td>1.03 (0.93, 1.14)</td>
<td>1.01 (0.85, 1.21)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mayo Lung&lt;sup&gt;57&lt;/sup&gt;</td>
<td>337/4618</td>
<td>303/4593</td>
<td>1.03 (0.74, 1.42)</td>
<td></td>
<td>1.01 (0.85, 1.21)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>North London&lt;sup&gt;52&lt;/sup&gt;</td>
<td>82/29723</td>
<td>68/25311</td>
<td>1.01 (0.85, 1.21)</td>
<td></td>
<td>1.01 (0.85, 1.21)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>John Hopkins&lt;sup&gt;56&lt;/sup&gt;&lt;sup&gt;f&lt;/sup&gt;</td>
<td>141/5226</td>
<td>173/5161</td>
<td>0.80 (0.65, 1.00)</td>
<td></td>
<td>1.01 (0.85, 1.21)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Memorial Sloan-Kettering&lt;sup&gt;54&lt;/sup&gt;&lt;sup&gt;f&lt;/sup&gt;</td>
<td>115/4968</td>
<td>120/5072</td>
<td>0.98 (0.76, 1.26)</td>
<td></td>
<td>1.01 (0.85, 1.21)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PLCO&lt;sup&gt;56&lt;/sup&gt;&lt;sup&gt;f&lt;/sup&gt;</td>
<td>1213/77445</td>
<td>1230/77456</td>
<td>0.99 (0.87, 1.22)</td>
<td></td>
<td>1.01 (0.85, 1.21)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CT scan</td>
<td>DANTE&lt;sup&gt;42&lt;/sup&gt;&lt;sup&gt;f&lt;/sup&gt;</td>
<td>20/1276</td>
<td>20/1196</td>
<td>0.94 (0.50, 1.73)</td>
<td></td>
<td>0.96 (0.64, 1.43)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>DLCST&lt;sup&gt;45&lt;/sup&gt;</td>
<td>15/2052</td>
<td>11/2052</td>
<td>1.36 (0.63, 2.96)</td>
<td></td>
<td>1.45 (0.98, 2.14)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>MILD&lt;sup&gt;44&lt;/sup&gt;</td>
<td>18/2376</td>
<td>7/1723</td>
<td>1.64 (0.67, 4.01)</td>
<td></td>
<td>1.40 (0.82, 2.38)</td>
<td></td>
</tr>
<tr>
<td>Oral cancer</td>
<td>Visual exam</td>
<td>Kerala, India – overall&lt;sup&gt;48&lt;/sup&gt;</td>
<td>138/96517</td>
<td>154/95356</td>
<td>0.88 (0.69, 1.12)</td>
<td></td>
<td>1.01 (0.85, 1.21)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kerala, India – high risk&lt;sup&gt;48&lt;/sup&gt;&lt;sup&gt;f&lt;/sup&gt;</td>
<td>129/4579</td>
<td>147/3951</td>
<td>0.76 (0.60, 0.97)</td>
<td></td>
<td>1.01 (0.85, 1.21)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kerala, India – low risk&lt;sup&gt;48&lt;/sup&gt;&lt;sup&gt;e&lt;/sup&gt;</td>
<td>95/50726</td>
<td>7/56205</td>
<td>1.36 (0.57, 3.26)</td>
<td></td>
<td>1.01 (0.85, 1.21)</td>
<td></td>
</tr>
<tr>
<td>Ovarian cancer</td>
<td>CA – 125</td>
<td>PLCO&lt;sup&gt;50&lt;/sup&gt;</td>
<td>118/34253</td>
<td>100/34304</td>
<td>1.18 (0.82, 1.71)</td>
<td></td>
<td>1.01 (0.85, 1.21)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>UK&lt;sup&gt;55&lt;/sup&gt;</td>
<td>9/10958</td>
<td>18/10977</td>
<td>0.50 (0.22, 1.11)</td>
<td></td>
<td>1.01 (0.85, 1.21)</td>
<td></td>
</tr>
<tr>
<td>Prostate cancer</td>
<td>Prostate-specific-antigen</td>
<td>ERSPC&lt;sup&gt;103&lt;/sup&gt;</td>
<td>261/82816</td>
<td>363/99184</td>
<td>0.86 (0.73, 1.01)</td>
<td></td>
<td>1.00 (0.98, 1.02)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Norkopping&lt;sup&gt;37&lt;/sup&gt;</td>
<td>30/1494</td>
<td>130/7532</td>
<td>1.16 (0.78, 1.73)</td>
<td></td>
<td>1.00 (0.98, 1.02)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PLCO&lt;sup&gt;53&lt;/sup&gt;</td>
<td>158/38340</td>
<td>143/38345</td>
<td>1.09 (0.87, 1.36)</td>
<td></td>
<td>1.00 (0.98, 1.02)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Quebec&lt;sup&gt;12&lt;/sup&gt;</td>
<td>153/31133</td>
<td>75/13333</td>
<td>1.01 (0.76, 1.33)</td>
<td></td>
<td>1.00 (0.98, 1.02)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Stockholm&lt;sup&gt;55&lt;/sup&gt;</td>
<td>53/2374</td>
<td>506/24772</td>
<td>1.09 (0.83, 1.45)</td>
<td></td>
<td>1.00 (0.95, 1.05)</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>The estimate for all-cause mortality in the Western Australia study is based on the ratio of the age standardized mortality data provided in Table 5 of Norman et al.<sup>35</sup> Although Lindholt et al.<sup>62</sup> show a decrease in all-cause mortality with longer follow-up (odds ratio = 0.91 (0.88, 0.95), this seems to be an artefact of age imbalance in the two arms; and Takagi et al.<sup>12</sup> show no difference between groups in the 65 to 74 age stratum, which is well age-balanced (hazard ratio = 0.99 (0.96, 1.02)).

<sup>b</sup>Odds ratio.

<sup>c</sup>Relative risk or rate ratio.

<sup>d</sup>Hazard ratio.

<sup>e</sup>Not specified.

<sup>f</sup>High-risk participants (smokers and alcohol users).

<sup>g</sup>Low-risk participants.
One additional single trial (Qidong, China) of screening with only alpha-fetoprotein was also found.

**Lung cancer.** Three meta-analyses of screening with chest X-ray were found; Manser *et al.* presented seven trials of screening with chest X-ray (Czech, Erfurt County, Kaiser, Mayo lung, North London, Johns Hopkins, and Memorial Sloan-Kettering) and has been selected for analysis. However, the Erfurt County study was excluded from the individual trial evidence (Table 2) because of its non-randomized design. The other two meta-analyses on chest X-ray were excluded because they were earlier publications and contained non-randomized data. Data from PLCO (chest X-ray) was not included in any of the meta-analyses but we presented its estimates in the individual trial evidence. There were four trials (DLCST, MILD, DANTE and NLST) on computer tomography (CT) scan, but no available meta-analyses. We excluded NLST as it evaluated comparative screening (CT scan vs chest X-ray). We recorded the estimates from the other three trials and conducted our own meta-analysis.

**Oral cancer.** One review of screening with visual examination was found. It contained only one trial from Kerala, India. The estimate with longer follow-up data from that trial was presented in the individual trial evidence.

**Ovarian cancer.** One meta-analysis on screening with CA-125 was found; it contained two individual trials (PLCO and UK).

**Prostate cancer.** Six meta-analyses of screening with prostate specific antigen (PSA) were found. Four were by Ilic *et al.* and contained the same data from five trials (ERSPC, Norrkoping, PLCO, Quebec and Stockholm); we used the estimates from the most recent publication. The other two were not used because they included site-specific data (e.g. French ERSPC and Gothenburg are part of original ERSPC) or used non-randomized data. Of the individual trials, we also included updated estimates for Norrkoping and PLCO.

**Cardiovascular disease.** One individual trial of screening with echocardiography was found.

**Type 2 diabetes mellitus.** One individual trial of screening with fasting blood glucose and haemoglobin A1c (HbA1c) was found.

Synopsis of RCT evidence (meta-analytic and individual) for mortality

**Meta-analytic evidence**

As shown in Table 1, meta-analyses of randomized trials were available for nine screening tests on six diseases. The 95% CIs excluded the null in 4 out of 11 available estimates (36%) of disease-specific mortality, but in none out of 10 available estimates for all-cause mortality. Disease-specific mortality was reduced with ultrasound for abdominal aortic aneurysm in men; mammography for breast cancer; and fecal occult blood test and flexible sigmoidoscopy for colorectal cancer. The range of relative risk reduction in these four cases was between 16% and 45%. Relative risk estimates for all-cause mortality were all very close to 1.00 (range 0.98–1.03).

**Individual trial evidence**

As shown in Table 2, we compiled evidence from 48 randomized trials on 19 screening tests for 11 diseases. The 95% CIs excluded the null in 16 out of 54 reported estimates (30%) (some trials reported more than one estimate, e.g. in different subgroups) for disease-specific mortality and for 4 out of 36 reported estimates (11%) for all-cause mortality. The range of relative risk reduction in the 16 cases with improved disease-specific mortality was between 13% and 73% (median 29%) and in the four cases of improved all-cause mortality it was between 3% and 13% (median 10%).

Disease-specific mortality was reduced with ultrasound for abdominal aortic aneurysm in the Viborg, MASS and Chichester trials; with mammography for breast cancer in the Goteborg and Two-county trials; with visual inspection for cervical cancer in the Tamil Nadu and Maharashtra trials; with FOBT for colorectal cancer in the Funen, Goteborg, Minnesota and Nottingham trials; with flexible sigmoidoscopy for colorectal cancer in the UK trial and PLCO; with alpha-fetoprotein and ultrasound for hepatocellular cancer in the Shanghai trial; and with visual examination for oral cancer in the Kerala trial. Overall, seven tests for six diseases had at least one RCT with a disease-specific mortality benefit; of those, three diseases had also been documented in meta-analyses.

All-cause mortality was reduced with ultrasound in abdominal aortic aneurysm in MASS; with mammography in breast cancer in Goteborg and Stockholm; and with visual examination for cervical cancer in Tamil Nadu. Mammography and ultrasound for aortic aneurysm had no all-cause mortality benefits in the respective meta-analyses including all the relevant trials. Visual examination for cervical cancer had also been assessed in
another trial that did not report results on all-cause mortality.47

Discussion
Our comprehensive overview shows that there are currently at least 48 RCTs and 9 non-overlapping meta-analyses that have evaluated the impact of any screening test vs no screening on mortality in asymptomatic adults for diseases where mortality is a common outcome. Documented reductions in disease-specific mortality in randomized trials of screening are uncommon. Reduction in all-cause mortality is even more uncommon in single trials and has not been documented in the latest available meta-analysis of multiple trials for any of the examined topics. This overview offers to researchers, policy makers and healthcare providers a synthesis of RCT evidence on the potential benefits of screening on mortality, and is timely in the wake of recent controversies around breast and prostate cancer screening.

Of the handful of trials that have reported survival benefits from screening, it is likely that in a few of them the benefit is substantially overestimated. For example, visual inspection of the cervix with acetic acid (cervical cancer screening) offered a 13% estimated relative risk reduction for all-cause death in one trial 46 conducted in rural India. Women in the screened group received other interventions apart from screening, such as correction of anaemia and measurement of blood pressure. Hence this large difference in total mortality, if true, was likely the result of multiple interventions and not the screening alone (cervical cancer does not account for 13% of all deaths even in rural India). Similarly, a mortality reduction shown in the Shanghai trial for screening in hepatocellular cancer51 is in question. The earlier paper91 from that trial did not report a risk estimate but only reported percent survival; a subsequent Cochrane review 89 used the survival data to calculate a risk estimate with 95% CIs that did not exclude the null. In the same way, the original publication 35 from the Western Australia study for screening in abdominal aortic aneurysm did not report a relative risk estimate for all-cause mortality; a subsequent meta-analysis 62 calculated a mortality reduction with 95% CIs that excluded the null, but this did not take into account the substantial age imbalance that existed between the study groups; and another more recent meta-analysis 12 that realized this caveat had 95% CIs that did not exclude the null.

There are many potential underlying reasons for the overall poor performance of screening in reducing mortality: the screening test may lack sufficient sensitivity and specificity to capture the disease early in its process; there are no markedly effective treatment options for the disease; treatments are available but the risk-benefit ratio of the whole screening and treatment process is unfavourable; or competing causes of death do not allow us to see a net benefit. Often, these reasons may coexist. Whether screening saves lives can only be reliably proven with RCTs.108 However, even for newly proposed tests, we suspect that their adoption in practice may evade RCT testing. A very large number of tests continuously become available due to technological advancement.109 One may be tempted to claim a survival benefit of screening based on observational cohorts showing improved survival rates,110 but these are prone to lead-time and other types of bias. Even RCTs can be biased sometimes, as has been discussed and hotly debated in the controversy over mammography.71

Some limitations should be acknowledged in our overview. First, we synthesized randomized evidence, but did not include data from other research designs, such as cohort and case-control studies. However, as we stated above, non-randomized studies have serious limitations. Non-randomized studies may provide useful suggestions and insights, but typically these would be less definitive, unless the effect is very robust and large, and most screening tests do not seem to have large effects on mortality. Second, one should acknowledge that given the many competing causes of death, it is very difficult to document reductions in all-cause mortality, unless the disease of interest is a leading cause of death and extremely large RCTs are performed. This is the reason why we also addressed comprehensively all the available data on disease-specific mortality. Third, we used broad search terms in PubMed and in Cochrane to maximize the capture of relevant trials and meta-analyses. It is possible that a few trials may have been missed, but it is unlikely that we have missed major trials that had found mortality benefits. As a quality check, we also matched our search results with the USPSTF documents. We found that we had not missed any trials that USPSTF has cited, whereas we have detected several additional recent trials that USPSTF did not cite (not unexpected since the USPSTF updates the evidence periodically). Finally, we did not include evidence on the effectiveness of one screening test against another (i.e. comparative screening). Nevertheless, it is difficult to interpret a trial that shows that a screening test is better than an older comparator, when it is unknown whether the older comparator does more good than harm.

To avoid uncertainty and a continuing conundrum in the world of screening for disease, we need to choose the appropriate study design and outcome, depending on the disease, to evaluate the effectiveness of screening tests. We argue that for diseases where short- and medium-term mortality are a relatively common outcomes, RCT should be the default evaluation tool and disease-specific and
all-cause mortality should be routinely considered as main outcomes. Our overview suggests that even then, all-cause mortality may hardly ever be improved. One may argue that a reduction in disease-specific mortality may sometimes be beneficial even in the absence of a reduction in all-cause mortality. Such an inference would have to consider the relative perception of different types of death by patients (e.g. death by cancer vs death by other cause), and it may entail also some subjectivity. For diseases where mortality outcomes are potentially important but only in the very long term, one has to consider whether the use of other, intermediate outcomes and/or other quasi-experimental designs that may be performed relatively quickly with very large sample sizes (e.g. before and after the introduction of a test) are meaningful alternatives to very long-term RCTs or may add more bias and confusion in a field that has already seen many hot debates. Screening may still be highly effective (and thus justifiable) for a variety of other clinical outcomes, besides mortality. However, our overview suggests that expectations of major benefits in mortality from screening need to be cautiously tempered.

Supplementary Data
Supplementary data are available at IJE online.

Conflict of interest: None declared.

References
Commentary on N Saquib et al. Does screening for disease save lives in asymptomatic adults? Systematic review of 5 meta-analyses and randomized trials

Paul G Shekelle

RAND Corporation, 1776 Main Street, Santa Monica, CA 90407, USA. E-mail: shekelle@rand.org

In this issue, Drs Saquib, Saquib and Ioannidis perform a valuable service by reviewing the evidence that screening for various diseases save lives. The authors examined the randomized controlled trials (RCTs) and meta-analyses of trials of the various screening strategies, and then assessed the outcomes of disease-specific mortality and all-cause mortality. They found that evidence of an effect on disease-specific mortality was relatively uncommon, and that evidence of an effect on all-cause mortality was essentially non-existent. The authors conclude that the effects of screening on mortality are likely to be modest at best, and that future evaluations of screening tests for diseases where short- and medium-term mortality are common, RCTs should be the default evaluation tool and disease-specific and all-cause mortality should be the main outcomes.

This raises the larger question of what should be the evidence upon which to base decisions about the appropriateness of screening tests, which by definition are...