HOW BIG ARE THE REAL MORTALITY REDUCTIONS PRODUCED BY CANCER SCREENING?
WHY DO SO MANY TRIALS REPORT ONLY 20%?

James A. Hanley¹
Zhihui (Amy) Liu¹, Nandini Dendukuri¹,², Erin Strumpf¹,³

¹Dept. of Epidemiology, Biostatistics & Occupational Health
²Dept. of Medicine, and Technology Assessment Unit
³Dept. of Economics
McGill University, Montréal, Québec, CANADA

32nd Annual Conference of the International Society for Clinical Biostatistics
Ottawa, Canada
Aug 24, 2011
Summary: the 3 points I wish to make

• With their blindness to the delay until the reductions in mortality are expressed, the prevailing design and data-analysis of cancer screening trials *underline estimate* the mortality reductions that *would be produced by a sustained screening program*

• P-value-driven stopping rules exacerbate the underestimation

• We *might* be able to avoid such misleading numbers if we
  (i) recognize the issue, and avoid the standard RCT paradigm
  (ii) run trials with sufficient rounds of screening and sufficient follow-up
  (iii) spend major portion of career waiting to measure real reductions
  (iv) analyze the data using time-specificity
  (v) focus on the *parameters that describe impact of 1 round of screening*
Outline

- The mortality reductions produced by a screening regimen: what payers want to know

- European Randomized Study of Screening for Prostate Cancer

- Data-analysis practice in other cancer screening trials

- How to stop a screening RCT at a 20% mortality reduction? [Theorem]

- A way ahead?
(a) Age-specific numbers of prostate cancer deaths in a steady state population with a given age-structure, if screening had not been available, and if screening had been available from ages 50 to 70.
(a) Age-specific numbers of prostate cancer deaths in a steady state population with a given age-structure, if screening had not been available, and if screening had been available from ages 50 to 70.

(b) The corresponding age-specific prostate cancer mortality rate ratios.
Can they obtain these (or asymptote) from published reports?
Screening & Prostate-Ca Mortality in Randomized European Study  (“ERSPC” nejm2009.04)

As of December 31, 2006, with an average follow-up time of 8.8 years, there were 214 prostate-cancer deaths in the screening group and 326 in the control group. (...) The adjusted rate ratio for death from prostate cancer in the screening group was **0.80** (95% CI, 0.65 to 0.98; P=0.04).

“PSA-based screening reduced the rate of death from prostate cancer by **20%**.”
RE-ANALYSIS OF ERSPC DATA
using
year-specific prostate cancer mortality ratios
(A) Overall vs. (B) Year-specific mortality ratios

BREAST CANCER

EVERY TRIAL:
and (nejm2010) REPORT on NORWAY NATIONAL SCREENING PROGRAM:

REDUCTION UNDER-ESTIMATED

LUNG CANCER
Mayo Lung Project (chest x-ray & sputum cytology)

• Enrollment: 1971-1976; negative on ‘prevalence’ screen; screening every 4 mo. for 6 years (vs., on enrollment, recommendation to receive annual chest x-ray & sputum cytology).


  Would 24-year follow up "allow for a reduction in lung cancer mortality to be observed?"

• ALL lung cancer deaths, from those in year...
  • 1, before impact could become evident, to
  • 24, 18 years after last screen.
Deaths from lung cancer in the NLST, with corresponding relative deficit in CT arm

What was reported (NEJM Aug 4, 2011) ...

<table>
<thead>
<tr>
<th>Follow-up Year:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>ALL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screens</td>
<td>↑</td>
<td>↑</td>
<td>↑</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X-ray Arm:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>442</td>
</tr>
<tr>
<td>CT Arm:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>354</td>
</tr>
<tr>
<td>Relative Deficit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20%</td>
</tr>
</tbody>
</table>

Year-specific data extracted from graph in that report ...

| X-ray Arm: | 37 | 68 | 82 | 95 | 84 | 73 | 4 |
| CT Arm:    | 31 | 57 | 67 | 84 | 72 | 42 | 3 |
| Relative Deficit | 16% | 16% | 18% | 12% | 14% | 42% |

Further year-specific numbers essential to measure impact of 3 rounds of screening.
20% MORTALITY REDUCTION

A UNIVERSAL CONSTANT IN SCREENING TRIALS?
Reductions in ‘event rates’: 3 ‘prevention’ studies

- HPV 6,11,16,18 infection:
  - *Quadrivalent human papillomavirus (HPV) vaccine*
- Paralytic or non-paralytic poliomyelitis:
  - *Salk Vaccine*
- Death from ruptured abdominal aneurym:
  - *Ultrasound screening*

**QUESTION**: Shape of $\downarrow (t)$ function, i.e., % Reduction in Rate as function of follow-up time, if rates based on...

- all events up to that point in f-up time? (1 ‘average’ rate)?
- when in f-up time events occurred (‘time-specific’ rates)?
1,2,3: Accumulating results for trials where reductions in event rates are virtually immediate, and sustained
- lowest 3 curves, reductions: >95%, ~60%, >45%

4,5,6: Likewise for trials of screening for cancer of the...
- PROSTATE, using the PSA test (Schröder et al., 2009)
- COLON, using once-only sigmoidoscopy (Atkin et al., 2010)
- LUNG, using Low-Dose Computed Tomography (NLST, 2011)

7: Accumulating results for a hypothetical trial of cancer screening. In this model program, screening continues for many years and produces mortality reductions of [0%], [2%], [14%] and [35%] in years 1, 2, 3 and 4 and a full [50%] reduction each year thereafter.
1,2,3: Accumulating results for trials where reductions in event rates are virtually immediate, and sustained - lowest 3 curves, reductions: >95%, ~60%, >45%

4,5,6: Likewise for trials of screening for cancer of the...
- PROSTATE, using the PSA test (Schröder et al., 2009)
- COLON, using once-only sigmoidoscopy (Atkin et al., 2010)
- LUNG, using Low-Dose Computed Tomography (NLST, 2011)

7: Accumulating results for a hypothetical trial of cancer screening. In this model program, screening continues for many years and produces mortality reductions of [0%], [2%], [14%] and [35%] in years 1, 2, 3 and 4 and a full [50%] reduction each year thereafter.

If intervention continues over time to deflect the same % of events, an estimate of the % reduction, based on the total number events in more (person)-time will be more precise

Mortality reductions from cancer screening manifest distally. Enrolling and following more people for short length of time yields a more precise UNDERestimate.

The seemingly-universal 20% reduction is an artifact of prevailing data-analysis methods and stopping rules.

If use all data from time screening commences, the first % reduction which was statistically different from zero does not answer the question of interest to payers.
PLANS
Data and Methods, Parameters, their Use

- **Data**: completed RCTs of screening for prostate, breast, colon and lung ca; population-based screening programs.

- **3 Parameters** (‘deliverables’) and how they will be fitted:

  \[
  \text{Rate Ratio}(y, a, s) = \sum \text{of reductions from all previous rounds of screening in study } s
  \]

  \[\text{Design matrix: 1 row per } y-a-s \text{ ‘cell’} \]

  - \(\text{No. deaths in screening arm}\) in each ‘cell’
  - \(\text{No. deaths in 2 arms combined}\)
  - Fit by Max. Likelihood (binomial model)

- **USE**: project mort. reductions due to a screening regimen
Screening in Chronic Disease

Second Edition

ALAN S. MORRISON

Acknowledgments

Mammographic screening: no reliable supporting evidence?

Olli S Miettinen, Claudia I Henschke, Mark W Pasmanter, James P Smith, Daniel M Libby, David F Yankelevitz

Much confusion is being generated by the conclusion of a recent review that “there is no reliable evidence that screening for breast cancer reduces mortality.” In that review, however, there was no appreciation of the appropriate mortality-related measure of screening’s usefulness; and correspondingly, there was no estimation of the magnitude of this measure. We take this measure to be the proportional reduction in case-fatality rate, and studied its magnitude on the basis of the only valid and otherwise suitable trial. We found reliable evidence of fatality reduction, apparently substantial in magnitude.

Lancet 2002; 359: 404–06

Screening for breast cancer in women aged 40-49 years.
Montreal: CETS Report no. 22, 1993. 91p. Available at:
http://www.aetmis.gouv.qc.ca/en/

J. Caro and M. McGregor
Why do statisticians commonly limit their inquiries to Averages?

F. Galton, Natural Inheritance, 1889.

“It is difficult to understand why statisticians commonly limit their inquiries to Averages, and do not revel in more comprehensive views.

Their souls seem as dull to the charm of variety as that of the native of one of our flat English counties, whose retrospect of Switzerland was that, if its mountains could be thrown into its lakes, two nuisances would be got rid of at once.”
With their blindness to the delay until the reductions in mortality are expressed, the prevailing design and data-analysis of cancer screening trials *under-estimate* the mortality reductions that *would be produced by a sustained screening program*

P-value-driven stopping rules exacerbate the underestimation

We *might* be able to avoid such misleading numbers if we
(i) recognize the issue, and avoid the standard RCT paradigm
(ii) run trials with sufficient rounds of screening and sufficient follow-up
(iii) spend major portion of career waiting to measure real reductions
(iv) analyze the data using time-specificity
(v) focus on the *parameters that describe impact of 1 round of screening*
FUNDING, CO-ORDINATES, DOWNLOADS

Natural Sciences and Engineering Research Council of Canada

Le Fonds québécois de la recherche sur la nature et les technologies

Canadian Institutes of Health Research (2011-2014)

James.Hanley@McGill.CA

http://www.biostat.mcgill.ca/hanley

→ reprints / talks

McGill Biostatistics

http://www.mcgill.ca/epi-biostat-occh/grad/biostatistics/
Some References


2. * Hanley JA. Mortality reductions produced by sustained prostate cancer screening have been underestimated. Journal of Medical Screening. *J Medical Screening* 2010;17:147-151.


* http:www.biostat.mcgill.ca/hanley/ (reprints/talks)
The loneliness of the long-distance trialist

Timing of Screening Effects
(as seen in cumulative cause-specific mortality curves)

Abdominal Aortic Aneurysms
(One-off Screening, MASS)

Prostate Cancer
(q 4y, ERSPC )
Timing of cholesterol reductions produced by statins

3 dogs at 20 mg/kg/day; 3 at 50 mg/kg/day

3 monkeys at 50
Timing of cholesterol reductions produced by statins

Humans
Cumulative vs. Year-specific Mortality...

in 100,000 men
(average age at entry: 62 years)

if screened using PSA test

0, 1, 2, 3, or 4 times,

tests 4 years apart

and followed for (9) 20 years

HYPOTHETICAL DATA
Cumulative & Year-specific results, if screen 0, 1, ..., 4 times, q 4y

<table>
<thead>
<tr>
<th>Year of F.U.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Screens*</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Each arrow indicates the timing of a screen for prostate cancer.
(B) Year-specific Rate Ratios & Percent Reductions

(A) Yearly No. of Prostate Cancer Deaths

* Each arrow indicates the timing of a screen for prostate cancer.

(B) Percentage Reduction in Yearly Cause-Specific Mortality Rate

Fig2
Effect of Screening Mammography on Breast-Cancer Mortality in Norway

Mette Kalager, M.D., Marvin Zelen, Ph.D., Frøydis Langmark, M.D., and Hans-Olov Adami, M.D., Ph.D.

Screening program was started in 1996 and expanded geographically during the subsequent 9 years.

Women between the ages of 50 and 69 years were offered screening mammography every 2 years.
Results & Conclusions

The rate of death was reduced by 7.2 deaths per 100,000 person-years in the screening group as compared with the historical screening group (rate ratio, 0.72; and by 4.8 deaths per 100,000 person-years in the nonscreening group as compared with the historical nonscreening group (rate ratio, 0.82; for a relative reduction in mortality of 10% in the screening group. Thus, the difference in the reduction in mortality between the current and historical groups that could be attributed to screening alone was 2.4 deaths per 100,000 person-years, or a third of the total reduction of 7.2 deaths. The availability of screening mammography was associated with a reduction in the rate of death from breast cancer, but the screening itself accounted for only about a third of the total reduction.
Time-insensitivity: not exclusive to RCT reports

Paraphrase of (refused) letter by JH to NEJM re 2010 analysis of data from Norway

Kalager Zelen Langmark Adami.

Epidemiologic Reviews, 2011

WebFigure 6:

[Illustrative] Reductions in breast-cancer mortality as functions of the duration of screening and the time elapsed since it was begun, in the 10-year period 1996-2005 in Norway.

Reductions only occur several years after screening commences; the more rounds of screenings there are, the greater the attained reduction is; at some point after the last screening the rates return to what they would have been in the absence of screening.

An average that includes – and is dominated by - the (early) years in which mortality is not affected by screening and excludes (later) years in which it is, provides a diluted measure of a cancer screening program’s impact on mortality from the disease.
emphasis on time-specificity

- Year-specific* mortality rate ratios
- Moving averages* to reduce the statistical noise (deaths in moving 3-year intervals)
- Smooth curve for rate ratio function (data bins 0.2 y wide).

* cf. Miettinen et al. 2002
National Lung Screening Trial (NLST)

- Enrollment: August 2002 - March-2004
- 3 annual screens: low-dose helical CT (vs. standard chest X-ray).

Primary scientific goal:

*to determine whether three annual screenings with low-dose helical computerized tomography (LDCT) reduces [sic] mortality from lung cancer*

- Press Releases, November 2010:

  Screening of people at high-risk for lung cancer with low dose CT significantly reduces lung cancer death: 20% fewer lung cancer deaths [ACR]

  An interim analysis of the study’s primary endpoint, reported to the DSMB on October 20, 2010, revealed a deficit of lung cancer deaths in the LDCT arm, and the deficit exceeded that expected by chance, even allowing for the multiple analyses conducted during the course of the trial. Data presented at previous meetings of the DSMB did not meet the requirements for statistical significance with respect to the primary endpoint. [NCI(US)]
Table 3: Interim Analysis of Primary Endpoint Reported on October 20, 2010

<table>
<thead>
<tr>
<th>Trial Arm</th>
<th>Person years (py)</th>
<th>Lung cancer deaths</th>
<th>Lung cancer mortality per 100,000 py</th>
<th>Reduction in lung cancer mortality (%)</th>
<th>Value of test statistic</th>
<th>Efficacy boundary</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDCT</td>
<td>144,097.6</td>
<td>354</td>
<td>245.7</td>
<td>20.3</td>
<td>-3.21</td>
<td>-2.02</td>
</tr>
<tr>
<td>CXR</td>
<td>143,363.5</td>
<td>442</td>
<td>308.3</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

“Deficit”: 88