Introduction to Statistical Methods for Clinical Trials

Amy Liu

Cancer Care Ontario

amy.liu@cancercare.on.ca

October 30, 2014
Early Detection Trials for Cancer
PSA test should be abandoned as screen for prostate cancer, task force says

Recommendation ‘undoubtedly will lead to more prostate cancer deaths,’ urologist counters


The blood test mostly commonly used to screen men for prostate cancer should be dropped, because it can result in more harm than good, says a Canadian task force.

The prostate-specific antigen, or PSA, test measures inflammation that can be elevated for many reasons other than cancer, such as normal enlargement of the prostate with age or an infection.
Screening

- Screening (Miettinen 2011):
  
  Pursuit of early (preclinical) detection of (i.e., of early rule-in diagnosis about) a particular illness.

- The aim of early detection is to enhance curability through early treatment.

- Examples:
  - screening of athletes to prevent sudden cardiac death;
  - ultrasound scans for abdominal aortic aneurysm;
  - pap smear to detect potentially precancerous lesions;
  - mammography screening for breast cancer;
  - faecal occult blood test for colorectal cancer.
Measuring the benefits of cancer screening

- The benefits of cancer screening is typically studied by means of a randomized trial.
- Asymptomatic persons are randomly assigned to receive either a number of screening examinations or usual care, and then are followed up for cancer-specific deaths.
- Cancer specific mortality reduction is considered as the definitive measure of the benefits. (Why?)
- In particular, cancer incidence would not be a meaningful outcome in a screening trial.
- Screening itself is not an intervention; any mortality benefits are due to successful early treatment (vs. late treatment following clinical diagnosis based on symptoms).
A schematic motivating the mortality reduction

$$1 - \frac{D_1}{D_0}$$
as the measure of benefit

$\text{c0} = \text{diagnosed cancers}$

$\text{c1} = \text{diagnosed cancers}$

$\text{f} = \text{fatal cancers if untreated}$

$\text{f} = \text{fatal cancers if untreated}$

$\text{d0} = \text{fatal cancers despite therapy}$

$\text{d1} = \text{fatal cancers despite therapy}$
The benefits of early vs. late treatment could be studied directly by randomizing screen-detected individuals into early treatment and late treatment arms. This would not be feasible in practice. (Why?)

A randomized screening trial could be powered in a usual way to detect a particular cancer mortality risk difference or risk ratio, using either a fixed follow-up period or stopping rule based on the number of deaths observed.

However, unlike the participants in therapeutic trials, the screenees are asymptomatic.

Any cancer-specific death is a rare outcome.

Because of this a very large number of participants is typically needed for adequate power.
Delayed effect and the length of the follow-up period

- Cancer screening is a textbook example of non-proportional hazards.
- Because the screenees are asymptomatic, any mortality benefits of cancer screening can only manifest after a delay.
- A cancer death that was avoided due to screening induced early treatments would have occurred years later since the time of the screening examination.
- Furthermore, the mortality benefits gradually disappear after the screening is discontinued.
- Continuing the follow-up after the benefits have disappeared can only dilute the result, if measured using average mortality reduction over the entire follow-up period.
A schematic illustrating the time-specific impact

(a) Yearly numbers of cancer deaths in a cohort of 50-year old individuals, without and with a 20-year screening program

(b) The corresponding cancer mortality rate ratio curve
Number of screening rounds

- While a screening program for a population in an eligible age range might involve annual testing for 20 years, a randomized screening trial usually involves only a few rounds of screening.

- This is mostly because of cost considerations; the screened cohort is large, and continuing the testing for years would be prohibitively expensive.

- As an example of a typical randomized screening trial, let us consider the design of the US National Lung Screening Trial.
Example: US National Lung Screening Trial

NLST design and projected timeline

CT Arm
CXR Arm
1:1
High-Risk Subjects
time
Final: October 2010

http://radiology.rsna.org/content/early/2010/10/28/radiol.10091808.full
NLST Study Base

The graph illustrates the population sizes for the CT arm and the X-ray arm over different follow-up years. The CT arm shows a significant drop in population size as the follow-up year increases, while the X-ray arm maintains a relatively stable population size.
Case series

The image shows a scatter plot with two arms: CT arm and X-ray arm. The x-axis represents the follow-up year, ranging from 0 to 8. The y-axis represents the population, ranging from 0 to 25,000. The CT arm is represented by red dots, and the X-ray arm is represented by black dots. The number of individuals in each arm decreases over time.
Numbers of lung cancer deaths in the two arms

![Graph showing numbers of lung cancer deaths in the two arms. The graph compares X-ray arm and CT arm over a follow-up period of 8 years. The CT arm shows significantly lower lung cancer deaths compared to the X-ray arm.]
Result: mortality rate ratio

- Note the modest numbers of lung cancer deaths despite more than 53,000 smokers followed up.
- Empirical lung cancer mortality rate ratio:
  \[
  \hat{\theta} = \frac{D_1/Y_1}{D_0/Y_0} = \frac{469/171}{552/170} \approx 0.844
  \]
- 95% confidence interval:
  \[
  \left\{ \hat{\theta} \times e^{-1.96 \times \sqrt{\frac{1}{D_1} + \frac{1}{D_0}}}, \hat{\theta} \times e^{1.96 \times \sqrt{\frac{1}{D_1} + \frac{1}{D_0}}} \right\} \approx \{0.746, 0.955\}.
  \]
Result: mortality reduction

- Lung cancer mortality reduction (one minus risk ratio):

\[
1 - RR = \frac{D_1/N_1}{D_0/N_0} \approx 1 - \frac{D_1}{D_0} = 1 - \frac{469}{552} \approx 0.15.
\]

- Screening was discontinued after 3 years; the impact on mortality in the cohort had already faded by the last year of follow-up.
**Posted:** 06/29/2011

NCI Press Release

**NIH-funded study shows 20 percent reduction in lung cancer mortality with low-dose CT compared to chest X-ray:**
Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening

The National Lung Screening Trial Research Team*

ABSTRACT

BACKGROUND

The aggressive and heterogeneous nature of lung cancer has thwarted efforts to reduce mortality from this cancer through the use of screening. The advent of low-dose helical computed tomography (CT) altered the landscape of lung-cancer screening, with studies indicating that low-dose CT detects many tumors at early stages. The National Lung Screening Trial (NLST) was conducted to determine whether screening with low-dose CT could reduce mortality from lung cancer.

METHODS

From August 2002 through April 2004, we enrolled 53,454 persons at high risk for lung cancer at 33 U.S. medical centers. Participants were randomly assigned to undergo three annual screenings with either low-dose CT (26,722 participants) or single-view posteroanterior chest radiography (26,732). Data were collected on cases of lung cancer and deaths from lung cancer that occurred through December 31, 2009.

The members of the writing team (who are listed in the Appendix) assume responsibility for the integrity of the article. Address reprint requests to Dr. Christine D. Berg at the Early Detection Research Group, Division of Cancer Prevention, National Cancer Institute, 6130 Executive Blvd., Suite 3112, Bethesda, MD 20892-7346, or at bergc@mail.nih.gov.

*A complete list of members of the National Lung Screening Trial research team is provided in the Supplementary Appendix, available at NEJM.org.

This article (10.1056/NEJMoal102873) was published on June 29, 2011, at NEJM.org.
What was reported?

Lung cancer deaths in the NLST before January 15th, 2009:

<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>Total</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>38</td>
<td>70</td>
<td>83</td>
<td>91</td>
<td>88</td>
<td>74</td>
<td>4</td>
<td>448</td>
</tr>
<tr>
<td>CT Arm:</td>
<td>31</td>
<td>57</td>
<td>67</td>
<td>84</td>
<td>72</td>
<td>45</td>
<td>3</td>
<td>359</td>
</tr>
<tr>
<td>Reduction (%)</td>
<td>18</td>
<td>19</td>
<td>19</td>
<td>8</td>
<td>18</td>
<td>39</td>
<td>25</td>
<td>20</td>
</tr>
</tbody>
</table>

\[1 - \frac{359}{448} = 20\%\]
## All follow-up data

### Year-specific data including deaths before and after the cutoff

<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>Total</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>38</td>
<td>70</td>
<td>83</td>
<td>91</td>
<td>89</td>
<td>116</td>
<td>65</td>
<td>552</td>
</tr>
<tr>
<td>CT Arm:</td>
<td>31</td>
<td>57</td>
<td>67</td>
<td>84</td>
<td>73</td>
<td>85</td>
<td>70</td>
<td>467</td>
</tr>
<tr>
<td>Reduction (%)</td>
<td>18</td>
<td>19</td>
<td>19</td>
<td>8</td>
<td>18</td>
<td>27</td>
<td>-8</td>
<td>15</td>
</tr>
</tbody>
</table>

\[
1 - \frac{467}{552} = 15\%
\]
NLST abstract in NEJM

- **Background**: The NLST was conducted to determine whether screening with low-dose CT could reduce mortality from lung cancer.

- **Methods**: three annual screenings

- **Results**: There were 247 deaths from lung cancer per 100,000 person-years in the low-dose CT group and 309 deaths per 100,000 person-years in the radiography group, representing a relative reduction in mortality from lung cancer with low-dose CT screening of **20.0%** (95% CI, 6.8 to 26.7; P=0.004).

- **Conclusions**: Screening with the use of low-dose CT reduces mortality from lung cancer.
How much did it cost?

- > 53,000 participants,
- > 33 centres,
- > 8 years of follow-up,
- > $ 250 MILLION.
Hypothesis testing vs. measuring of benefits

- Most screening trials have been carried out for the purpose of hypothesis-testing (zero vs. non-zero reduction).

- e.g. NLST objective: “compare whether screening with low-dose helical CT scan vs chest x-ray reduces lung cancer-specific mortality in participants who are at high risk for developing lung cancer.”

- Sometimes results are announced when the accumulated mortality reduction first becomes statistically significantly different from zero.

- e.g. the European Randomized Study of Screening for Prostate (ERSPC) protocol: “the study will have to be discontinued if a significant difference is reached between the screening and control arm with regard to the main endpoint of the study: prostate cancer mortality.”

- Such a stopping rule may hide the extent of the mortality benefit.
The cancer screening controversy

- Screening for cancer has been controversial.
- Many countries have devoted a lot of resources to screening programs for cancers over the last 40-50 years.
- Despite many long and costly randomized screening trials involving large numbers of participants, it is unclear how large the benefits are.
- Screening trials in cancer are “notoriously difficult to run, and notoriously susceptible to errors” (Mukherhee 2010).
- Mammography screening for breast cancer has been particularly controversial.
Mammography screening trials

There are 8 large randomized mammography screening trials over the past 50 years since 1963:

<table>
<thead>
<tr>
<th>Study (Reference)</th>
<th>Age Range (y)</th>
<th>Median Follow-up</th>
<th>Breast Cancer Deaths/Total Women Screened Group</th>
<th>Control Group</th>
<th>Breast Cancer Death Rate per 1000 Women Screened Group</th>
<th>Control Group</th>
<th>Relative Risk for Death from Breast Cancer (95% CI)</th>
<th>Absolute Risk Reduction per 1000 Women</th>
<th>Number Needed to Invite to Screening†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stockholm alone</td>
<td>40–64</td>
<td>13.8</td>
<td>82/39 139</td>
<td>50/20 978</td>
<td>2.10</td>
<td>2.38</td>
<td>0.91 (0.65–1.27)</td>
<td>0.288</td>
<td>3468</td>
</tr>
<tr>
<td>Gothenburg (23)</td>
<td>39–59</td>
<td>12.8</td>
<td>62/20 724</td>
<td>113/29 200</td>
<td>2.99</td>
<td>3.87</td>
<td>0.76 (0.56–1.04)</td>
<td>0.078</td>
<td>1139</td>
</tr>
<tr>
<td>Malmö (23)</td>
<td>45–70</td>
<td>17.1</td>
<td>161/21 088</td>
<td>198/21 195</td>
<td>7.63</td>
<td>9.35</td>
<td>0.82 (0.67–1.00)</td>
<td>0.172</td>
<td>584</td>
</tr>
<tr>
<td>Swedish Two-County Trial (26)</td>
<td>40–74</td>
<td>17</td>
<td>319/77 080</td>
<td>333/55 985</td>
<td>4.14</td>
<td>5.95</td>
<td>0.68 (0.59–0.80)</td>
<td>1.809</td>
<td>553</td>
</tr>
<tr>
<td>Mammography plus CBE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CNBSS-1 (22)</td>
<td>40–49</td>
<td>13</td>
<td>105/25 214</td>
<td>108/25 216</td>
<td>4.16</td>
<td>4.28</td>
<td>0.97 (0.74–1.27)</td>
<td>0.12</td>
<td>–</td>
</tr>
<tr>
<td>CNBSS-2 (20)</td>
<td>50–59</td>
<td>13</td>
<td>107/19 711</td>
<td>105/19 694</td>
<td>5.43</td>
<td>5.33</td>
<td>1.02 (0.78–1.33)</td>
<td>0.097</td>
<td>–</td>
</tr>
<tr>
<td>HIP (19)</td>
<td>40–64</td>
<td>16</td>
<td>232/30 239</td>
<td>281/30 256</td>
<td>5.46</td>
<td>6.89</td>
<td>0.79</td>
<td>1.438</td>
<td>883</td>
</tr>
<tr>
<td>Edinburgh (18)</td>
<td>45–64</td>
<td>13</td>
<td>156/22 926</td>
<td>167/21 342</td>
<td>6.80</td>
<td>7.82</td>
<td>0.79 (0.60–1.02)</td>
<td>1.020</td>
<td>980</td>
</tr>
</tbody>
</table>

* CBE = clinical breast examination; CNBSS = Canadian National Breast Screening Study; HIP = Health Insurance Plan of Greater New York.
† Number needed to invite to screening to prevent one death from breast cancer 13–20 years after randomization.
Numerous systematic reviews and meta-analyses of the same trial data

Breast Cancer Screening: A Summary of the Evidence for the U.S. Preventive Services Task Force

Linda L. Humphrey, MD, MPH; Mark Helfand, MD, MS; Benjamin K.S. Chan, MS; and Steven H. Woolf, MD, MPH

Purpose: To synthesize new data on breast cancer screening for the U.S. Preventive Services Task Force.

Data Sources: MEDLINE; the Cochrane Controlled Trials Registry; and reference lists of reviews, editorials, and original studies.

Study Selection: Eight randomized, controlled trials of mammography and 2 trials evaluating breast self-examination were included. One hundred fifty-four publications of the results of these trials, as well as selected articles about the test characteristics and harms associated with screening, were examined.

Data Extraction: Predefined criteria were used to assess the quality of each study. Meta-analyses using a Bayesian random-effects model were conducted to provide summary relative risk estimates and credible intervals (CrIs) for the effectiveness of screening with mammography in reducing death from breast cancer.

Data Synthesis: For studies of fair quality or better, the summary relative risk was 0.84 (95% CrI, 0.77 to 0.91) and the number needed to screen to prevent one death from breast cancer after approximately 14 years of observation was 1224 (CrI, 665 to 2564). Among women younger than 50 years of age, the summary relative risk associated with mammography was 0.85 (CrI, 0.73 to 0.99) and the number needed to screen to prevent one death from breast cancer after 14 years of observation was 1792 (CrI, 764 to 10,540). For clinical breast examination and breast self-examination, evidence from randomized trials is inconclusive.

Conclusions: In the randomized, controlled trials, mammography reduced breast cancer mortality rates among women 40 to 74 years of age. Greater absolute risk reduction was seen among older women. Because these results incorporate several rounds of screening, the actual number of mammograms needed to prevent one death from breast cancer is higher. In addition, each screening has associated risks and costs.


For author affiliations, see end of text.

See related article on pp 344-346 and editorial comments on pp 361-362 and pp 363-365.
Mammography screening for breast cancer has been particularly controversial.

- Some claim that the benefits were large – the reduction in breast cancer mortality is 25 - 33% (Day 2000).
- Others argue that the reduction is much more modest – only 10% and thus suggested discontinuing mammography screening (Gotzsche 2011).
Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomised screening trial

Anthony B Miller professor emeritus¹, Claus Wall data manager¹, Cornelia J Baines professor emerita¹, Ping Sun statistician², Teresa To senior scientist³, Steven A Narod professor¹²

¹Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario M5Y 3M7, Canada; ²Women’s College Research Institute, Women’s College Hospital, Toronto, Ontario M5G 1N9, Canada; ³Child Health Evaluative Services, The Hospital for Sick Children, Toronto, Ontario, Canada

Abstract

Objective To compare breast cancer incidence and mortality up to 25 years in women aged 40-59 who did or did not undergo mammography screening.

Design Follow-up of randomised screening trial by centre coordinators, the study’s central office, and linkage to cancer registries and vital

Conclusion Annual mammography in women aged 40-59 does not reduce mortality from breast cancer beyond that of physical examination or usual care when adjuvant therapy for breast cancer is freely available. Overall, 22% (105/484) of screen detected invasive breast cancers were over-diagnosed, representing one over-diagnosed breast cancer for every 424 women who received mammography screening in the trial.
CNBSS result

- Design: 90,000 women randomized to mammography or no mammography, followed up for breast cancer mortality over 25 years.

- Finding: cumulative breast cancer mortality reduction was

\[
1 - RR = 1 - \frac{500/44,925}{505/44,910} \approx 1 - \frac{500}{505} \approx 1\%.
\]
Responses to the CNBSS report

- American College of Radiology and Society of Breast Imaging:
  
  *an incredibly misleading analysis based on the deeply flawed and widely discredited study.*

- New York Times:
  
  *[o]ne of the largest and most meticulous studies of mammography ever done, involving 90,000 women and lasting a quarter-century, has added powerful new doubts about the value of the screening test for women of any age.*
From BMJ “Rapid Responses” (1)

- A gynaecologist writes:

My sister, age 49, was diagnosed with breast cancer in the UK on the basis of a mammogram. It was 5 cms. When I asked her had she felt it, she said no, and that no-one had ever advised her to check her breasts. My sister was very lucky to be node negative, but after chemo and a local excision she did need a mastectomy. She received excellent care in the NHS. However had she been doing monthly breast self-examination I believe she would have picked up her “lump” when it was below 2 cms – perhaps 6 months - 1 year before – and she would not have needed the mastectomy.

- Why detecting a smaller tumour earlier is not good enough to prove the value of screening?
Lead time (Mukherjee 2010)

Hope:
(screened)

Prudence:
(unscreened)
Lead time (Mukherjee 2010)

Hope: (screened)
Prudence: (unscreened)


survival
Lead time (Mukherjee 2010)

Lead time (Mukherjee 2010)

- Hope: (screened)  
  - Onset of screen dx (1994)  
  - Survival until death (2000)

- Prudence: (unscreened)  
  - Onset of clinical dx (1998)  
  - Lead time  
  - Survival until death (2000)
From BMJ “Rapid Responses” (2)

- A physician writes:

  Ten year disease-free survival for my patients with mammographically detected cancers is 92 percent compared to 82 percent if the cancer was detected on clinical examination. My results are not exceptional.

- His observations may not be exceptional, but they do not provide evidence on the effectiveness of screening. (Why?)
Common biases in evaluating the benefits of screening

- Comparing survival of women detected by mammograms versus those detected by clinical examination is exactly what one should not do in order to prevent
  1. lead-time,
  2. length,
  3. and overdiagnosis biases.

- The screen-detected cancers are
  1. detected earlier (irrespective of whether the subsequent early treatments are effective),
  2. exclude the fast growing interval cancers,
  3. and include cancers that might not have proven fatal in the absence of early detection.
From BMJ “Rapid Responses” (3)

- A physician writes:
  
  *I did not see any reference to breast size, shape [...] as possible confounding issues that nullified by these data.*

- This is a randomized trial, and if the randomization was done properly, all such factors should be balanced between the two arms and therefore there should be no confounding issues.
Failed randomization?

- A particular criticism aimed at CNBSS was that the randomization was compromised.
- This is because the participating women first underwent a physical examination, and after that were assigned to either screening or control groups.
- If something suspect was found in the physical examination, was the examiner able to influence entering the participant into the screening group?
From BMJ “Rapid Responses” (4)

- A radiologist writes:
  
  *It is indisputable that this happened since there was a statistically significant excess of women with advanced breast cancers who were assigned to the screening arm compared to those assigned to the control arm. This guaranteed that there would be more early deaths among the screened women than the control women and this is what occurred in the NBSS. Shifting women from the control arm to the screening arm would increase the cancers in the screening arm and reduce the cancers in the control arm which would also account for what they claim is “overdiagnosis”.*

- Can you test whether randomization is violated?
From BMJ “Rapid Responses” (5)

- Authors of the CNBSS report respond:
  
  *The Canadian study encompassed more than 50 center-years of operation - and Dr. Tabar proposes that one coordinator in a short span of time was able to corrupt the results from 90,000 women by improper randomization of five, ten, even a hundred? of her so-called friends. Absurd.*

- While almost 90,000 women were randomized, the cumulative numbers of breast cancer deaths were only 500 and 505 in the non-screened and screened groups, respectively, after 25 years of follow-up.

- It does not take much to alter the result, if symptomatic women (presumably some of them with otherwise fatal cancers) were moved from one arm to the other.
From BMJ “Rapid Responses” (6)

- A surgeon writes:

  Most importantly, a woman who does not die from breast cancer does not mean she does not strive with it; breast cancer remains a very hard and consuming personal, psychological, familial and social adventure for millions of women worldwide. Early detection of a non-palpable breast cancer promptly leads to appropriate management in order to fight the disease at an early stage, and offers optimal care and quality of life.

- Early detection could also lead to overdiagnosis and overtreatment.
Summary

- The most objective measure of benefit of cancer screening is the mortality reduction between the screened and unscreened groups in a randomized screening trial.

- However, this benefit is not constant over time; when screening is discontinued, the mortality benefit gradually disappears.

- Comparison of survival after screen-detected vs. clinically detected cancers would be distorted by lead time bias.

- Comparison of survival between screened vs. non-screened individuals non-experimentally would be confounded. (Why?)

- Randomized screening trials are free of these problems.
References


