ESTIMATING OVERDIAGNOSIS FROM TRIALS AND POPULATIONS

OVERCOMING CHALLENGES, AVOIDING MISTAKES
National Expenditure For False-Positive Mammograms And Breast Cancer Overdiagnoses Estimated At $4 Billion A Year

A recent randomized controlled trial of screening mammography in Canada reported an overdiagnosis rate of 22 percent for all screen-detected invasive breast cancer. The US rate of overdiagnosis has been estimated to be 22–31 percent of all breast cancers diagnosed.
Fatal Retraction
Not all cancers are lethal—despite the fear the name evokes. Although doctors often can't tell for certain which individual tumors are destined to be deadly, a growing number of studies suggest that many found at early stages may be so slow-growing they are unlikely to be fatal. Some recent estimates of this 'overdiagnosis' rate in common cancers:

<table>
<thead>
<tr>
<th></th>
<th>Prostate</th>
<th>Breast</th>
<th>Thyroid</th>
<th>Skin</th>
<th>Lung</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate</td>
<td>60%</td>
<td>30%</td>
<td>90%</td>
<td>90%</td>
<td>18%</td>
</tr>
</tbody>
</table>

Sources: American Cancer Society (Prostate); New England Journal of Medicine (Breast); The BMJ (Thyroid); American Academy of Dermatology (Skin); JAMA Internal Medicine (Lung); The Wall Street Journal

IT'S TIME TO RETHINK
EARLY CANCER DETECTION

A growing number of experts argue that zealous screening too often leads to overtreatment. They call for changing the way we even talk about the disease.

Glennon score of 6 or below "benign lesions"—although others note that that would mean half of the men treated for prostate cancer in the past 20 years didn't have cancer after all.

Overdiagnosis—the detection of tumors that aren't likely to cause harm—is now a hot topic in other cancers as well. A growing volume of studies estimate...
What is overdiagnosis?

Two ways of estimating the frequency of overdiagnosis
- Excess incidence
- Modeling

Excess incidence
- Conditions for valid estimates
- Some examples of published studies

The modeling approach
- Conditions for valid estimates
- Some examples of published studies

Summary – the questions that you, as consumers of overdiagnosis studies, should be asking
Overdiagnosis occurs when a cancer is detected by screening but it would not have been detected in the absence of screening.
WHAT IS OVERDIAGNOSIS?

- Overdiagnosis occurs when a cancer is detected by screening but it would not have been detected in the absence of screening.
Overdiagnosis as an Iceberg
What Lies Beneath

- Overdiagnosis depends on
  - Unobserved lead time
  - Risk of other-cause death

- Overdiagnosis occurs when
  - Lead time is longer than time to other-cause death

- Overdiagnosis is more likely when
  - Patients are older
  - Disease is slow-growing or non-progressive
OVERDIAGNOSIS AS A WAVE
OBSERVABLE CONSEQUENCES FOR DISEASE INCIDENCE

Incidence pattern after screening starts:
- Incidence excesses (+) followed by corresponding deficits (-)
- Excesses: screening pulls cases from the future
- Deficits: cases screen detected no longer in prevalent pool

Note: Bump in incidence observed even if there is no overdiagnosis!
TWO APPROACHES TO ESTIMATING OVERDIAGNOSIS
SYMPTOM VERSUS CAUSE

- Excess incidence
  - Empirically based
  - Calculate incidence with screening minus incidence without screening

- Modeling approach
  - Learn about latent disease process
  - Calculate lead time and derive estimate of overdiagnosis frequency

SEND QUESTIONS TO PREVENTION@MAIL.NIH.GOV USE @NIHPREVENTS & #NIHMTG ON TWITTER
PUBLISHED ESTIMATES VARY WIDELY

<table>
<thead>
<tr>
<th>Author</th>
<th>Study years</th>
<th>DCIS?</th>
<th>Estimate</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morrell, 2010</td>
<td>1999–2001</td>
<td>No</td>
<td>30–42%</td>
<td>Excess cases/cases expected without screening</td>
</tr>
<tr>
<td>Gøtzsche, 2011</td>
<td>Multiple</td>
<td>Yes</td>
<td>30%</td>
<td>Excess cases/cases expected without screening</td>
</tr>
<tr>
<td>Kalager, 2012</td>
<td>1996–2005</td>
<td>No</td>
<td>15–25%</td>
<td>Excess cases/cases expected without screening</td>
</tr>
<tr>
<td>Bleyer, 2012</td>
<td>1976–2008</td>
<td>Yes</td>
<td>31%</td>
<td>Excess cases/detected cases</td>
</tr>
<tr>
<td>Paci, 2006</td>
<td>1986–2001</td>
<td>Yes</td>
<td>4.6%</td>
<td>Cases overdiagnosed/cases expected without screening</td>
</tr>
<tr>
<td>Olsen, 2006</td>
<td>1991–1995</td>
<td>No</td>
<td>4.8%</td>
<td>Cases overdiagnosed/detected cases</td>
</tr>
<tr>
<td>de Gelder, 2011</td>
<td>1990–2006</td>
<td>Yes</td>
<td>8.9%</td>
<td>Cases overdiagnosed/Screen-detected cases</td>
</tr>
</tbody>
</table>
GETTING EXCESS INCIDENCE RIGHT

- **Timing**
- **Metric**
  - Annual excess incidence
  - Cumulative excess incidence
  - Denominator issues
- **Counterfactual**
  - Clinical trials (control group)
  - Population studies
GETTING EXCESS INCIDENCE RIGHT – CLINICAL TRIALS

1. CONTINUED SCREEN TRIAL

Hypothetical setting:

- Constant preclinical incidence
- Maximum preclinical period = 6 y
- Constant test sensitivity
- No overdiagnosis

**Screen arm receives tests in all years**

**Control arm receives no tests**

**Two curves never meet**

**Send questions to prevention@mail.nih.gov**

**Use @nihprevents & #nihmtg on twitter**
In the Trial setting equal numbers of individuals are randomized to a screen or control arm. Latent disease onset occurs at a constant rate. The preclinical duration follows a uniform distribution. The control arm receives no screen tests, so control arm incidence matches latent onset in this arm. The screen arm begins screening in year 1. The empirical difference between screen and control arm incidence can provide an unbiased estimate of the number of overdiagnoses cases under 2 conditions. (1) The difference is based on cumulative incidence if the screen arm stops screening and on annual incidence if the screen arm continues screening. (2) The difference is calculated after screening stabilizes plus the maximum preclinical duration.
The problem with cumulative excess incidence

- What we know

- What we observe

In the continued-screen setting cumulative excess incidence will be greater than zero even if NO overdiagnosis!
Hypothetical setting:
- Constant preclinical incidence
- Maximum preclinical period = 6 y
- Constant test sensitivity
- No overdiagnosis
POPULATION STUDIES

- Background incidence generally not available – no control group
- As in clinical trials – cumulative excess incidence is persistently biased
- Annual excess incidence – wait until screening stabilizes plus max preclin duration

A) Cumulative uptake (years 2-6): 5%, 15%, 30%, 45%, 50%
Max preclinical detectable period: 6 years

B) Cumulative uptake (years 2-4): 10%, 40%, 50%
Max preclinical detectable period: 6 years

C) Cumulative uptake (years 2-6): 5%, 15%, 30%, 45%, 50%
Max preclinical detectable period: 12 years
In the Population setting latent disease onset occurs at a constant rate and the preclinical duration follows a uniform distribution. Initially, before screening starts, the model projects disease incidence in steady state, so that diagnosis without screening matches latent onset in the population. Annual screening begins in segments of the population at specified starting years. The empirical difference between annual incidence with and without screening provides an unbiased estimate of overdiagnosis after screening stabilizes plus the maximum preclinical duration.

Input parameters

<table>
<thead>
<tr>
<th>Population size:</th>
<th>Disease incidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual rate of onset:</td>
<td></td>
</tr>
<tr>
<td>Range of preclinical durations:</td>
<td></td>
</tr>
<tr>
<td>Episode sensitivity:</td>
<td></td>
</tr>
</tbody>
</table>

https://rgulati.shinyapps.io/calculator/
CONDITIONS FOR VALID EXCESS INCIDENCE ESTIMATES OF OVERDIAGNOSIS

**Cumulative excess incidence**
- Continued-screen trials and population settings: persistently biased
- Stop-screen trials: wait until end of screening interval plus maximum preclinical duration

**Annual (point) excess incidence**
- Continued-screen trials: unbiased at end of maximum preclinical duration
- Stop-screen trials: unbiased at end of screening interval plus max preclin duration
- Population setting: unbiased at end of screening stabilization plus max preclin duration

- In all cases: take note of denominator used and verify background trend is reasonable
- Also note work done to remedy some of the known biases in excess incidence when a restricted age range is screened

SEND QUESTIONS TO PREVENTION@MAIL.NIH.GOV  USE @NIHPREVENTS & #NIHMTG ON TWITTER
Screening and Prostate-Cancer Mortality in a Randomized European Study

Fritz H. Schröder, M.D., Jonas Hugosson, M.D., Monique J. Roobol, Ph.D.,

- Cumulative excess incidence
- Continued-screen trial

<table>
<thead>
<tr>
<th>Year of publication</th>
<th>Median follow-up, years</th>
<th>Overdiagnosis among screen detections</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>9</td>
<td>58%</td>
</tr>
<tr>
<td>2012</td>
<td>11</td>
<td>55%</td>
</tr>
<tr>
<td>2014</td>
<td>13</td>
<td>49%</td>
</tr>
</tbody>
</table>

SEND QUESTIONS TO PREVENTION@MAIL.NIH.GOV  USE @NIHPREVENTS & #NIHMTG ON TWITTER
# Cumulative Excess Incidence of Invasive Cancers

<table>
<thead>
<tr>
<th>Trial arm</th>
<th>N</th>
<th>Cumulative incidence of invasive cancers</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Years 1-5</td>
<td>Years 1-10</td>
<td>Years 1-25</td>
<td></td>
</tr>
<tr>
<td>Mammography+CBE</td>
<td>44,925</td>
<td>666</td>
<td>1180</td>
<td>3250</td>
<td></td>
</tr>
<tr>
<td>CBE only</td>
<td>44,910</td>
<td>524</td>
<td>1080</td>
<td>3133</td>
<td></td>
</tr>
<tr>
<td>Excess cancers in mammography arm</td>
<td>142</td>
<td>100</td>
<td>117</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Excess among 484 screen detections</td>
<td>29%</td>
<td>21%</td>
<td>24%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Cumulative excess incidence
- Stop-screen trial

CWBSS Includes years after trial screens

Miller et al, BMJ, 2014
Most provinces started screening programs soon after trial screens ended.

### Table 3
CNBSS participants by province, years participated, and year organized provincial screening programs commenced

<table>
<thead>
<tr>
<th>Province</th>
<th>CNBSS participants</th>
<th>Years participants</th>
<th>Year provincial program commenced</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Percent</td>
<td>Entered CNBSS</td>
</tr>
<tr>
<td>Total</td>
<td>89835</td>
<td>100.0</td>
<td>1980–85</td>
</tr>
</tbody>
</table>

More screening in mammography arm after trial screens?
Since 1986, an estimated additional 1,305,600 men were diagnosed with prostate cancer.

- Cummulative excess incidence
- Background incidence imputed based on incidence in years prior to screening
BREAST CANCER INCIDENCE IN THE US POPULATION

Women aged 40 and older

0.25% increase per year based on under 40 trends

31% of detected cancers in 2008 overdiagnosed

- Annual excess incidence
- Background incidence imputed based on incidence trends in women under 40
Prostate Cancer Incidence: White Males (SEER 9 Registries)
BREAST CANCER INCIDENCE IN NORWAY

15-20% overdiagnosis relative to incidence expected in absence of screening

- Cumulative excess incidence after 1st yr
- Background incidence imputed based on counties not implementing screening

SEND QUESTIONS TO PREVENTION@MAIL.NIH.GOV  USE @NIHPREVENTS & #NIHMTG ON TWITTER
WHAT IS THE MAXIMUM PRECLINICAL DURATION FOR INVASIVE BREAST CANCER?

<table>
<thead>
<tr>
<th>MST</th>
<th>Years</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIP</td>
<td>2.5</td>
<td>1.2</td>
</tr>
<tr>
<td>Sweden</td>
<td>5.5</td>
<td>2.1</td>
</tr>
<tr>
<td>Malmö</td>
<td>2.1</td>
<td>1.3</td>
</tr>
<tr>
<td>Stockholm +</td>
<td>2.6</td>
<td>0.61</td>
</tr>
<tr>
<td>40-49</td>
<td>4.3</td>
<td>0.37</td>
</tr>
<tr>
<td>50-59</td>
<td>1.9</td>
<td>1.2</td>
</tr>
<tr>
<td>Edinburgh</td>
<td>3.1</td>
<td>0.94</td>
</tr>
<tr>
<td>Canada 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canada 2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Consensus mean 3.3 (~40 months)
GOING BEYOND THE DATA
USING MODELING TO LEARN ABOUT OVERDIAGNOSIS

1. Go beyond observed data to learn about underlying disease process
   • Given data on screening uptake
   • Use incidence before and after screening to learn about disease natural history

SEND QUESTIONS TO PREVENTION@MAIL.NIH.GOV  USE @NIHPREVENTS & #NIHMTG ON TWITTER
GOING BEYOND THE DATA USING MODELING TO LEARN ABOUT OVERDIAGNOSIS

1. Go beyond observed data to learn about underlying disease process
   • Given data on screening uptake
   • Use incidence before and after screening to learn about disease natural history

   ![Incidence Graph](image)

   - Infer overdiagnosis based on the estimated natural history (lead time)
     • Overdiagnosis occurs when other-cause death happens before the data of clinical diagnosis

   - Sojourn time
     - Onset
     - Clinical
     - Other-cause death
PREREQUISITES FOR A USEFUL MODEL

A. Need data on disease incidence with and without screening
   • Screening trials: control group provides the counterfactual incidence
   • Population studies: may need to guesstimate a counterfactual

B. Need information on screening patterns that produced the incidence
   • Screening trials: have individual-level data on screening and mode of diagnosis
   • Population studies: typically have to reconstruct screening trends; individual-level data generally not available

C. Need a model that is identifiable (estimable) from the data

SEND QUESTIONS TO PREVENTION@MAIL.NIH.GOV  USE @NIHPREVENTS & #NIHMTG ON TWITTER
Collaborative Modeling of the Benefits and Harms Associated With Different U.S. Breast Cancer Screening Strategies

<table>
<thead>
<tr>
<th>Strategy and Age Group</th>
<th>All Overdiagnosed Cases, %†‡</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Biennial</strong></td>
<td></td>
</tr>
<tr>
<td>50-74</td>
<td>12 (8-22)</td>
</tr>
<tr>
<td>45-74</td>
<td>12 (8-22)</td>
</tr>
<tr>
<td>40-74</td>
<td>13 (9-24)</td>
</tr>
<tr>
<td><strong>Hybrid</strong></td>
<td></td>
</tr>
<tr>
<td>45-74</td>
<td>13 (8-25)</td>
</tr>
<tr>
<td>40-74</td>
<td>14 (9-27)</td>
</tr>
<tr>
<td><strong>Annual</strong></td>
<td></td>
</tr>
<tr>
<td>50-74</td>
<td>15 (8-36)</td>
</tr>
<tr>
<td>45-74</td>
<td>17 (9-38)</td>
</tr>
<tr>
<td>40-74</td>
<td>18 (9-39)</td>
</tr>
</tbody>
</table>

(as a proportion of all cases detected)
THE IDENTIFIABILITY PROBLEM
CAN THE MODEL BE LEARNED FROM THE DATA?

A) Progressive disease only

\[
S_0 \rightarrow S_p \rightarrow S_c
\]

- Three parameters:
  - Risk of onset
  - Risk of progression to clinical dx
  - Screening test sensitivity

Can be learned from incidence with and without screening given screening patterns

B) Mixture of progressive and indolent disease

\[
S_0 \rightarrow S_p \rightarrow S_c
\]

- Four parameters:
  - Risk of onset
  - Risk of being indolent
  - If not: Risk of progression to clinical dx
  - Screening test sensitivity
A SIMPLE EXPERIMENT OF IDENTIFIABILITY

In a survival analysis dataset with data censored at 5 years, the following underlying models are all consistent with the data:

- Exponential mean 40 months
- Mixture of 75% exponential with mean 18 months, 25%(effectively) infinite
- Mixture of 95% exponential with mean 26 months, 5% infinite

All will yield a mean of 40 months under an exponential model. Different models are equally consistent with the same data.

Send questions to prevention@mail.nih.gov. Use @nihprevents & #nihmtg on Twitter.
BREAST CANCER NATURAL HISTORY FROM A TRIAL

A. Counterfactual incidence from a control group
B. Individual level screening histories
C. Progressive disease assumption – exponential sojourn time assumed while screening test sensitivity is estimated

A) Progressive disease only

\[ S_0 \rightarrow S_p \rightarrow S_c \]

Screening Sensitivity and Sojourn Time From Breast Cancer Early Detection Clinical Trials: Mammograms and Physical Examinations

By Yu Shen and Marvin Zelen

<table>
<thead>
<tr>
<th></th>
<th>MST</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIP</td>
<td>2.5</td>
<td>1.2</td>
</tr>
<tr>
<td>Sweden</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malmö</td>
<td>5.5</td>
<td>2.1</td>
</tr>
<tr>
<td>Stockholm+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40-49</td>
<td>2.1</td>
<td>1.3</td>
</tr>
<tr>
<td>50-59</td>
<td>2.6</td>
<td>.61</td>
</tr>
<tr>
<td>Edinburgh</td>
<td>4.3</td>
<td>.37</td>
</tr>
<tr>
<td>Canada 1</td>
<td>1.9</td>
<td>1.2</td>
</tr>
<tr>
<td>Canada 2</td>
<td>3.1</td>
<td>.94</td>
</tr>
</tbody>
</table>

SEND QUESTIONS TO PREVENTION@MAIL.NIH.GOV USE @NIHPREVENTS & #NIHMTG ON TWITTER

JCO 2001
A. Assume incidence in the absence of screening would have remained constant at pre-PSA rates
B. Aggregate screening histories retrospectively constructed from NHIS and SEER-Medicare
C. Progressive disease assumption – risk of progression to advanced or symptomatic disease depends on PSA growth rate which varies across men based on data from the PCPT trial

Since 1986, an estimated additional 680,300 men were diagnosed with prostate cancer
A FLEXIBLE PROGRESSIVE DISEASE MODEL YIELDS HETEROGENEITY IN SOJOURN TIMES

Distributions of sojourn times from a population model of prostate cancer

- **Relevant**: diagnosed within lifetime
- **Uncensored**: indolent until death

Sojourn times for relevant cancers are shorter in older men to ensure diagnosis before death

While a mixture model is not explicitly assumed, the model structure builds in heterogeneity
CANNOT WE ESTIMATE A MIXTURE MODEL?

A. Counterfactual incidence from a control group – constant over interval analyzed

B. Individual level screening histories

C. Model allows for non-progressive disease but for identifiability needs to assume known test sensitivity

B) Mixture of progressive and indolent disease

\[ S_0 \xrightarrow{S_p} S_{p'} \xrightarrow{S_c} \]

Shen et al. 2016

Screening test sensitivity
IDENTIFYING IDENTIFIABILITY (OR LACK THEREOF) CAN BE HARD

Population study
A. Background incidence imputed based on age-period-cohort model (increasing trend)
B. Retrospective reconstruction of screening patterns
C. Each model has a different structure and method for estimating parameters
TAKE-HOME MESSAGES

Overdiagnosis is complex – must ask key questions about each estimation approach

Empirical approach – excess incidence
• Design – stop screen or continued-screen?
• Estimate – cumulative or point excess incidence? Denominator?
• Timing – has enough time elapsed?
• Counterfactual – Is a fitting counterfactual available?

Modeling approach
• Screening patterns – are these properly informed by available data?
• Counterfactual – what is the counterfactual in a population setting?
• Identifiability – how is the model constructed to permit identifiability?

SEND QUESTIONS TO PREVENTION@MAIL.NIH.GOV USE @NIHPREVENTS & #NIHMTG ON TWITTER
ACKNOWLEDGMENTS

- Roman Gulati
- Lurdes Inoue
- Yu Shen (MD Anderson)
- Eric Feuer (NCI)
- CISNET support

SEND QUESTIONS TO PREVENTION@MAIL.NIH.GOV USE @NIHPREVENTS & #NIHMTG ON TWITTER
QUESTIONS?

Send questions to prevention@mail.nih.gov

Or

Use @NIHprevents & #NIHMrG on Twitter