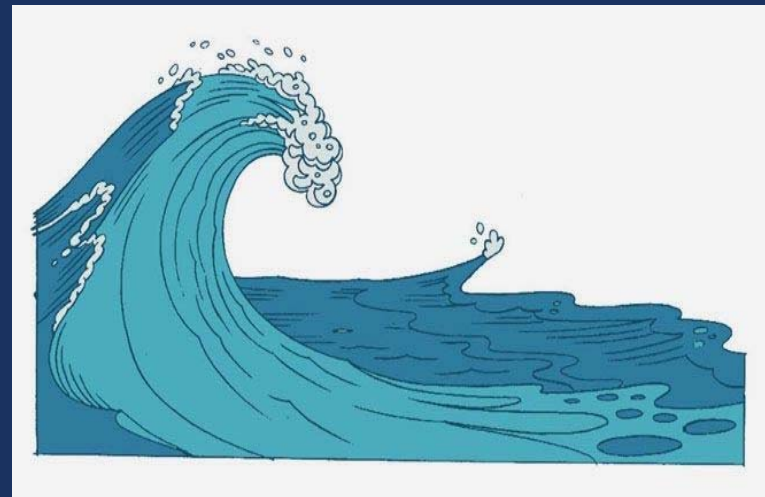

ESTIMATING OVERDIAGNOSIS FROM TRIALS AND POPULATIONS

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.⁴ The US rate of overdiagnosis has been estimated to be 22–31 percent of all breast cancers diagnosed.²



HEALTH CARE

© 2014 Dow Jones & Company. All Rights Reserved.

THE WALL STREET JOURNAL.

Monday, September 15, 2014 | R1

Follow
The Experts
An Online
Conversation
DETAILS, R2

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:

Prostate	Breast	Thyroid	Skin	Lung
60%	30%	90%	90%	18%

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

Ductal carcinoma in situ is an early, noninvasive form of breast cancer in which abnormal cells (the small dark spots) are confined to milk ducts. Experts think only about 20% of cases would eventually become invasive cancer, but virtually all are treated with surgery and radiation.

IT'S TIME TO RETHINK
EARLY CANCER DETECTION

BY MELINDA BECK

EARLY DETECTION HAS long been seen as a powerful weapon in the battle against cancer. But some experts now see it as double-edged sword.

While it's clear that early-stage cancers are more treatable than late-stage ones, some leading cancer

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.

Gleason 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.

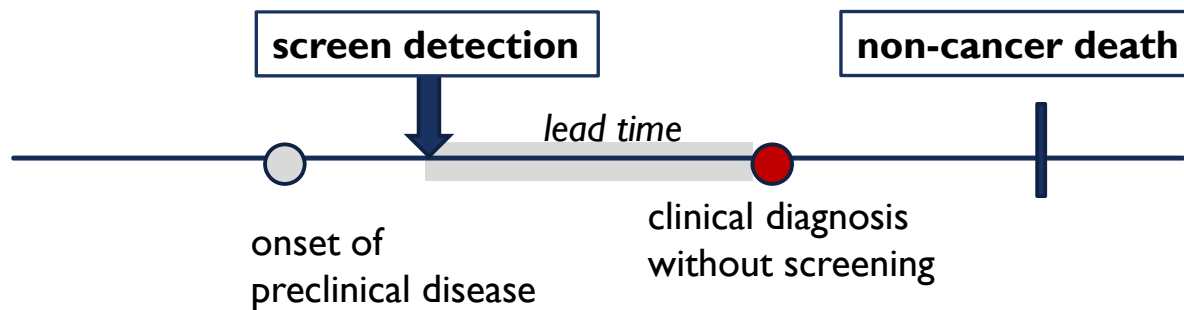
Overdiagnosis—the detection of tumors that aren't likely to cause harm—is now a hot topic in other cancers as well. A growing volume of studies estimate that as many as 20% of invasive breast cancers, 18%

TODAY'S PRESENTATION

- 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

WHAT IS OVERDIAGNOSIS?

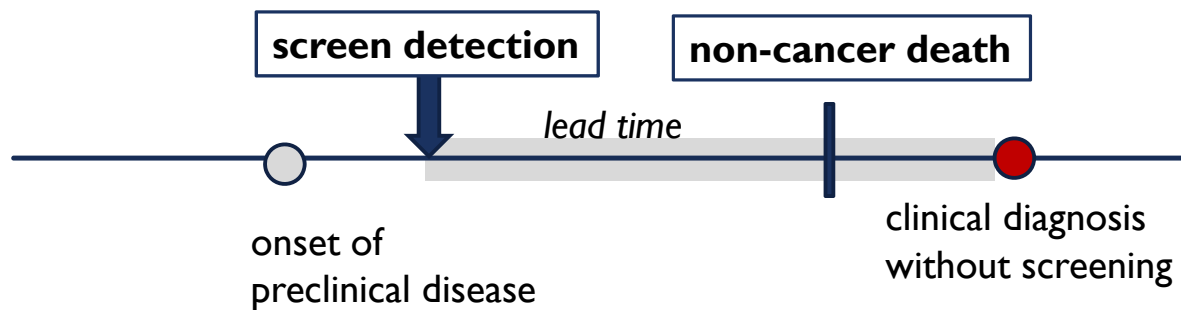
- Overdiagnosis occurs when a cancer is detected by screening but it would not have been detected in the absence of screening



NOT OVERDIAGNOSED

WHAT IS OVERDIAGNOSIS?

- Overdiagnosis occurs when a cancer is detected by screening but it would not have been detected in the absence of screening



OVERDIAGNOSED

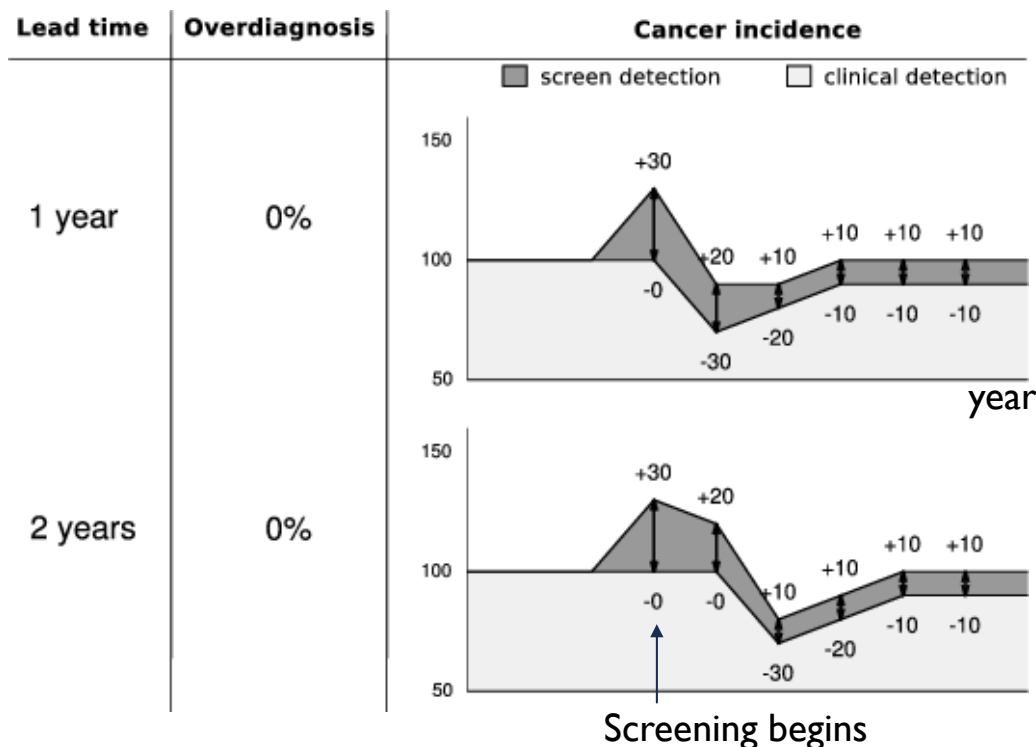
OVERDIAGNOSIS AS AN ICEBERG

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

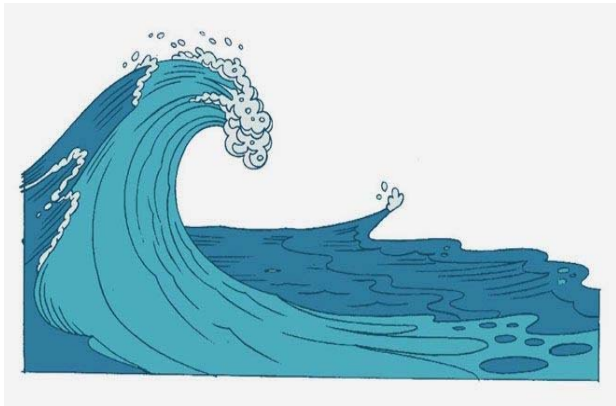


- Incidence pattern after screening starts:
 - Incidence excesses (+) followed by corresponding deficits (-)
 - Excesses: screening pulls cases from the future
 - Deficits: cases screen detected no longer in prevalent pool
- Note: Bump in incidence observed even if there is no overdiagnosis!

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

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

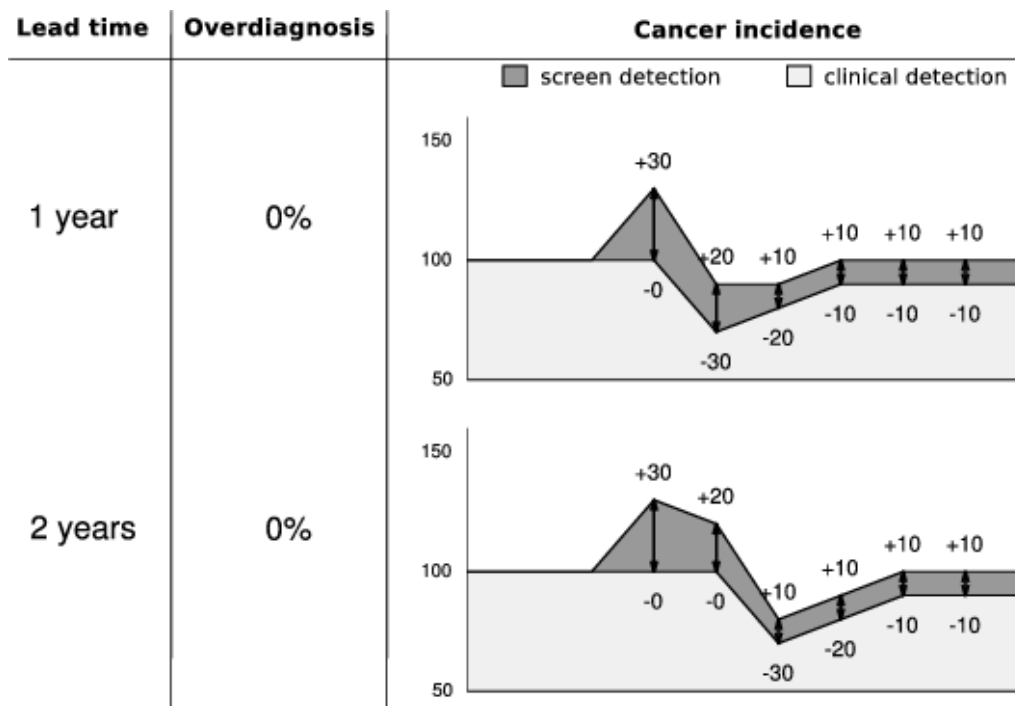


SEND QUESTIONS TO PREVENTION@MAIL.NIH.GOV USE @NIHPREVENTS & #NIHMTG ON TWITTER

PUBLISHED ESTIMATES VARY WIDELY

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

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

I. CONTINUED SCREEN TRIAL

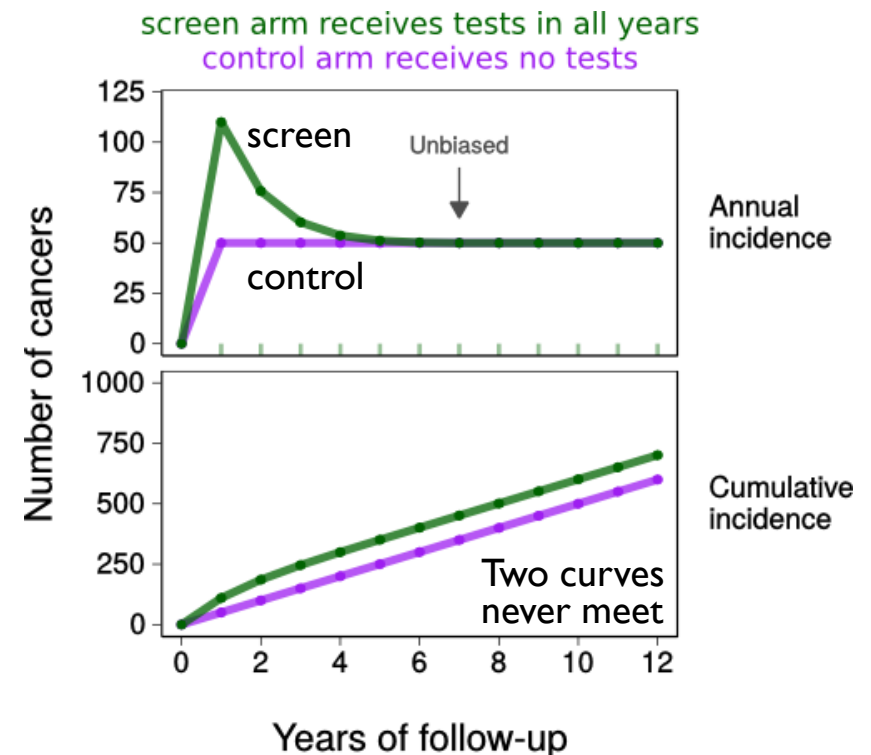
Hypothetical setting:

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

Practice of Epidemiology

Conditions for Valid Empirical Estimates of Cancer Overdiagnosis in Randomized Trials and Population Studies

Roman Gulati*, Eric J. Feuer, and Ruth Etzioni



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

Overdiagnosis calculator



Home

Trial setting

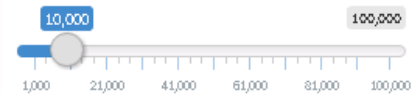
Population setting

Help

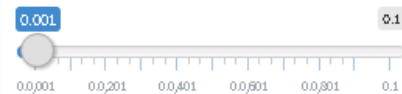
In the **Trial setting** equal numbers of individuals are randomized to a screen or control arm. Latent disease onset occurs at a constant rate. The preclinical duration follows a uniform distribution. The control arm receives no screen tests, so control arm incidence matches latent onset in this arm. 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.

Input parameters

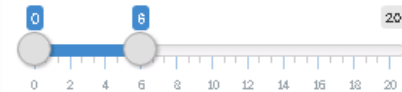
Arm size:



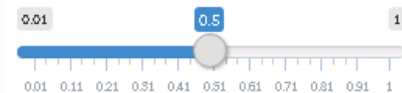
Annual rate of onset:



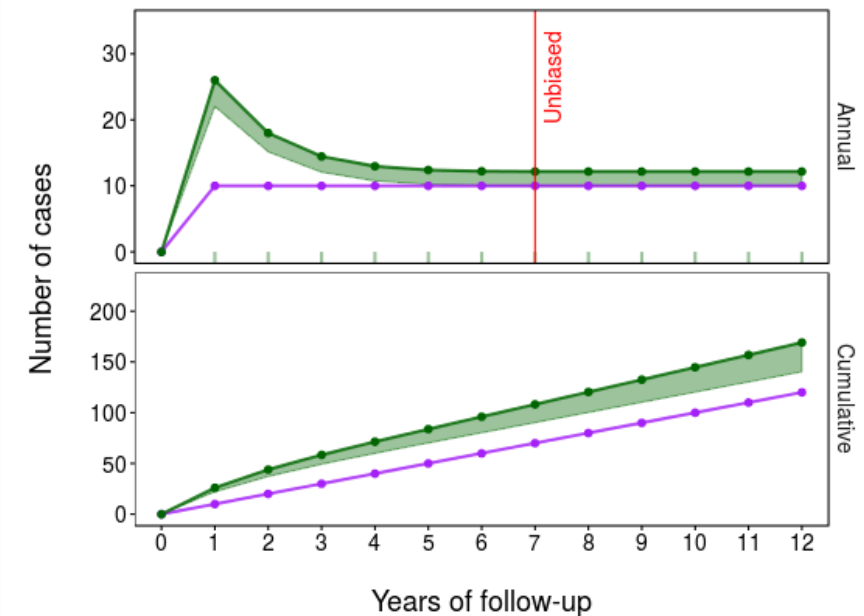
Range of preclinical durations:



Episode sensitivity:



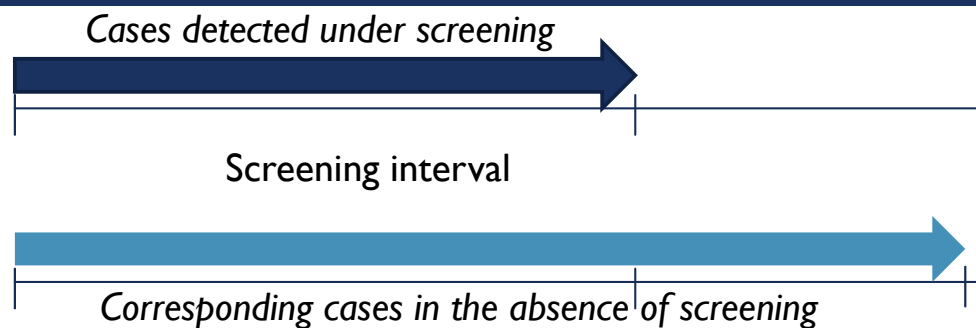
Disease incidence



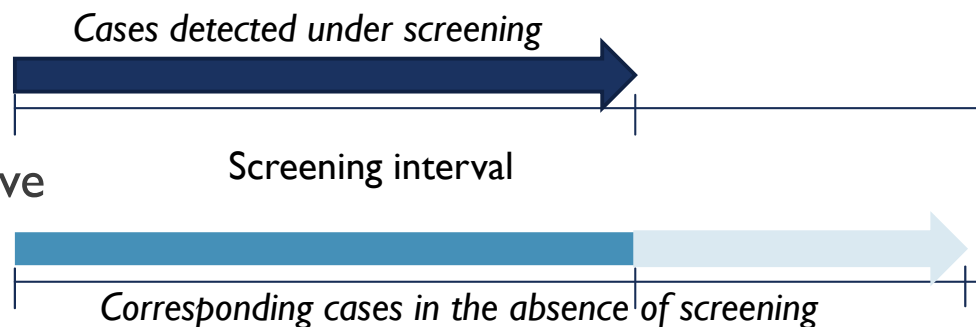
<https://rgulati.shinyapps.io/calculator/>

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!

GETTING EXCESS INCIDENCE RIGHT – CLINICAL TRIALS

II. STOP SCREEN TRIAL

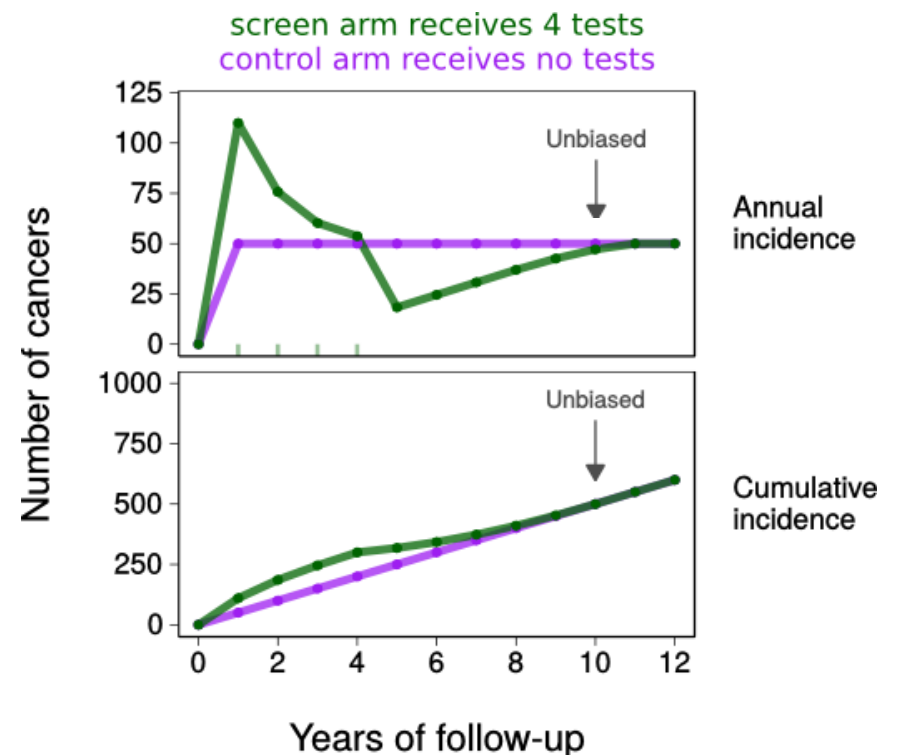
Hypothetical setting:

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

Practice of Epidemiology

Conditions for Valid Empirical Estimates of Cancer Overdiagnosis in Randomized Trials and Population Studies

Roman Gulati*, Eric J. Feuer, and Ruth Etzioni

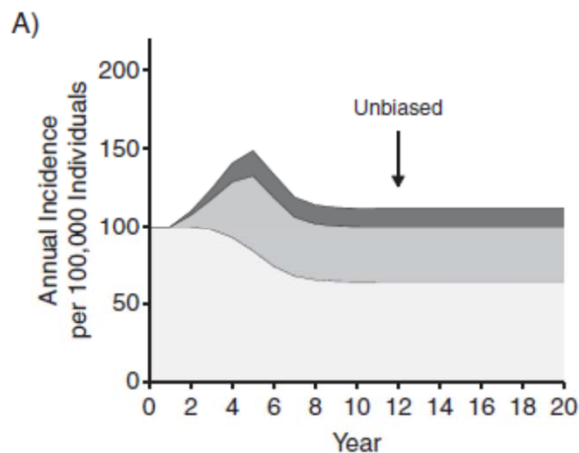


SEND QUESTIONS TO PREVENTION@MAIL.NIH.GOV USE @NIHPREVENTS & #NIHMTG ON TWITTER

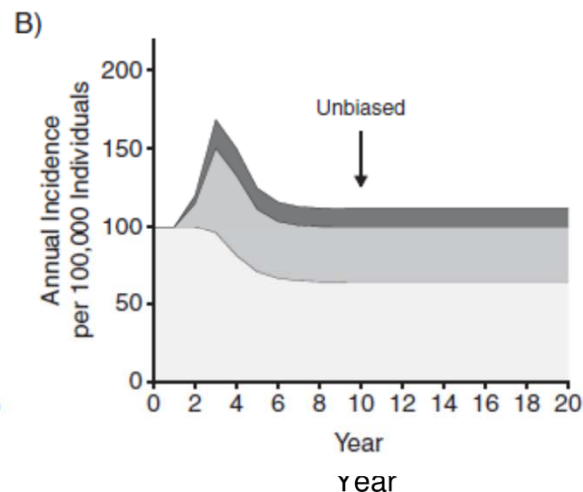
POPULATION STUDIES

- 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

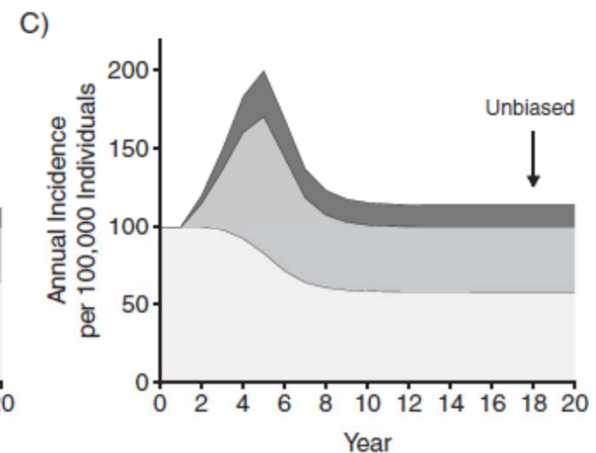
A) Cumulative uptake (years 2-6): 5%,15%,30%,45%,50%
Max preclinical detectable period: 6 years



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



Overdiagnosis calculator



Home

Trial setting

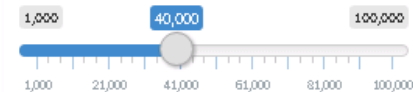
Population setting

Help

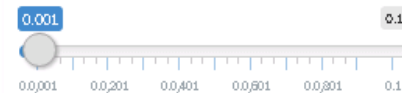
In the **Population setting** latent disease onset occurs at a constant rate and the preclinical duration follows a uniform distribution. Initially, before screening starts, the model projects disease incidence in steady state, so that diagnosis without screening matches latent onset in the population. 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

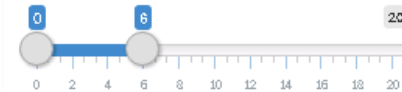
Population size:



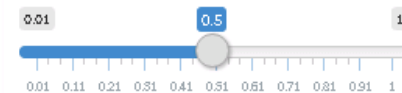
Annual rate of onset:



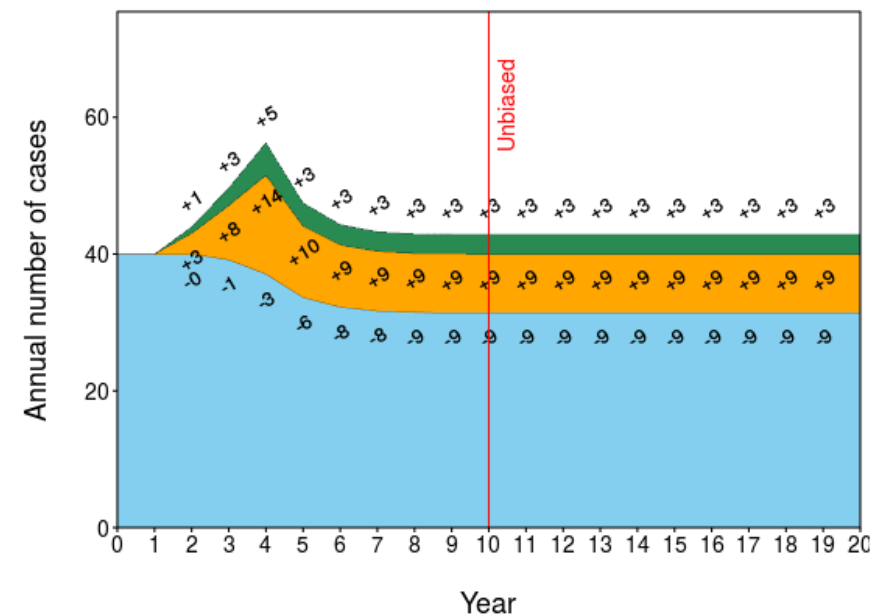
Range of preclinical durations:



Episode sensitivity:



Disease incidence



<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

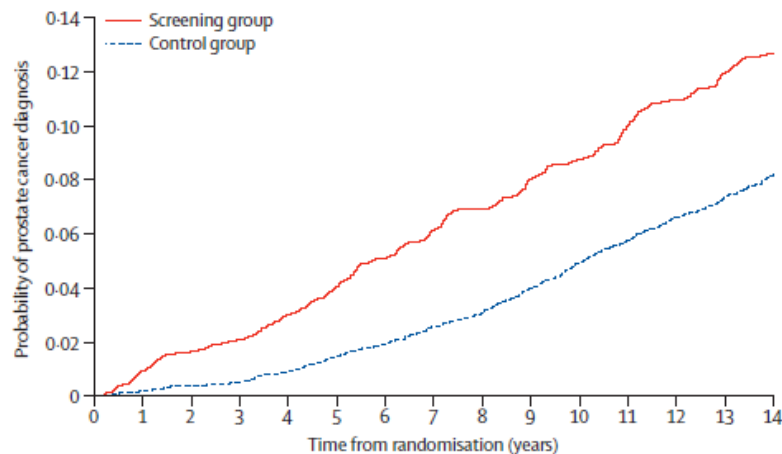
SEND QUESTIONS TO PREVENTION@MAIL.NIH.GOV USE @NIHPREVENTS & #NIHMTG ON TWITTER

EUROPEAN RANDOMIZED STUDY OF SCREENING FOR PROSTATE CANCER

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



Year of publication	Median follow-up, years	Overdiagnosis among screen detections
2009	9	58%
2012	11	55%
2014	13	49%

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

CANADIAN NATIONAL BREAST SCREENING STUDY

- Cumulative excess incidence
- Stop-screen trial

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



 CNBSS Includes years after trial screens

CANADIAN NATIONAL BREAST SCREENING STUDY

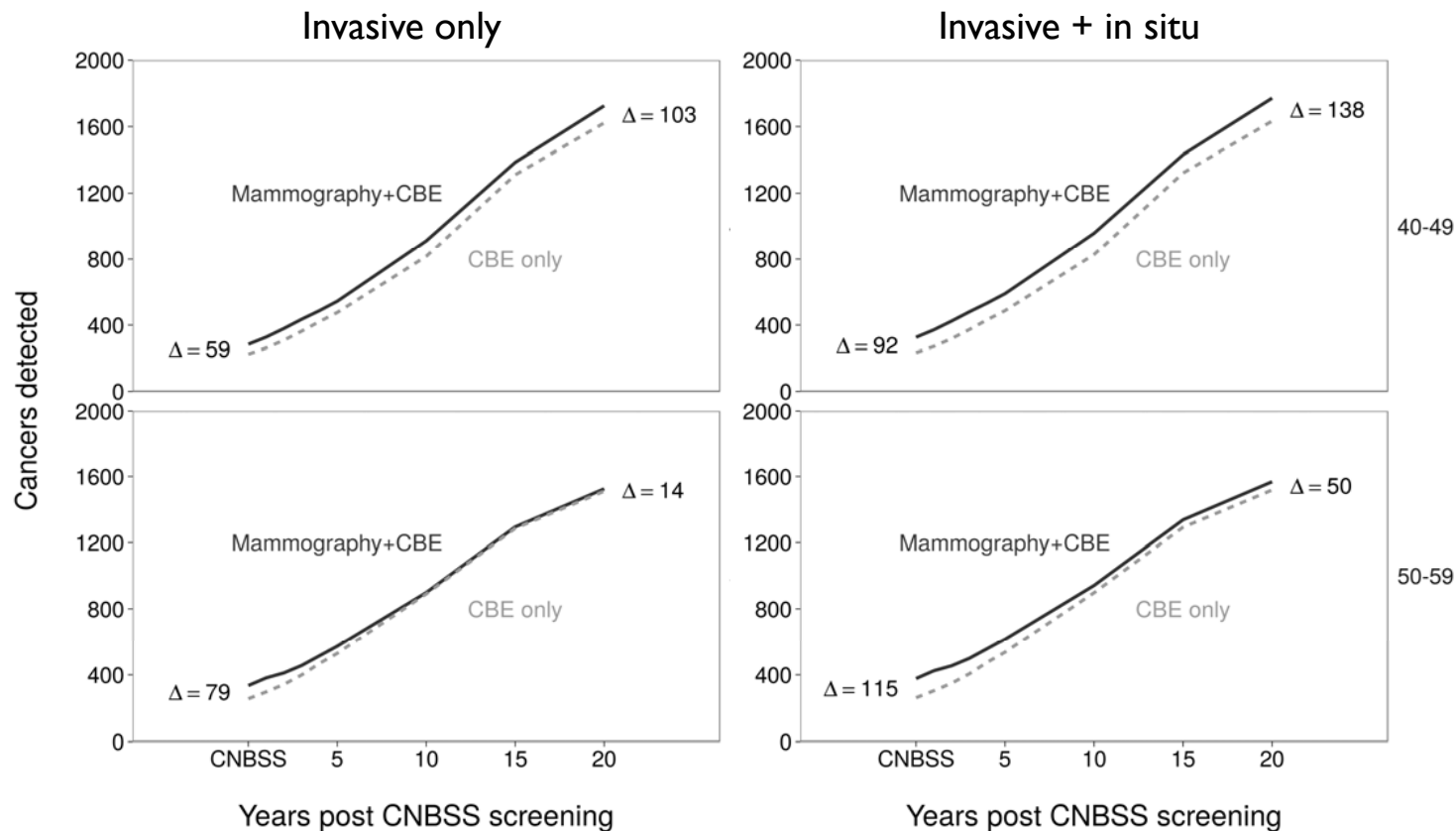
Table 3

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

Most provinces started screening programs soon after trial screens ended

Province	CNBSS participants		Years participants		Year provincial program commenced
	Number	Percent	Entered CNBSS	Completed CNBSS	
Alberta	9536	10.6	1984–85	1988	1990
British Columbia	9365	10.4	1983–85	1987–88	1988
Manitoba	10094	11.2	1981–84	1985–88	1995
Nova Scotia	6782	7.6	1983–85	1987–88	1991
Ontario	31972	35.6	1980–85	1984–88	1990
Quebec	22095	24.6	1980–85	1984–88	1998
Total	89835	100.0	1980–85	1984–88	–

CANADIAN NATIONAL BREAST SCREENING STUDY



More screening in
mammography arm
after trial screens?

PROSTATE CANCER INCIDENCE IN THE US POPULATION

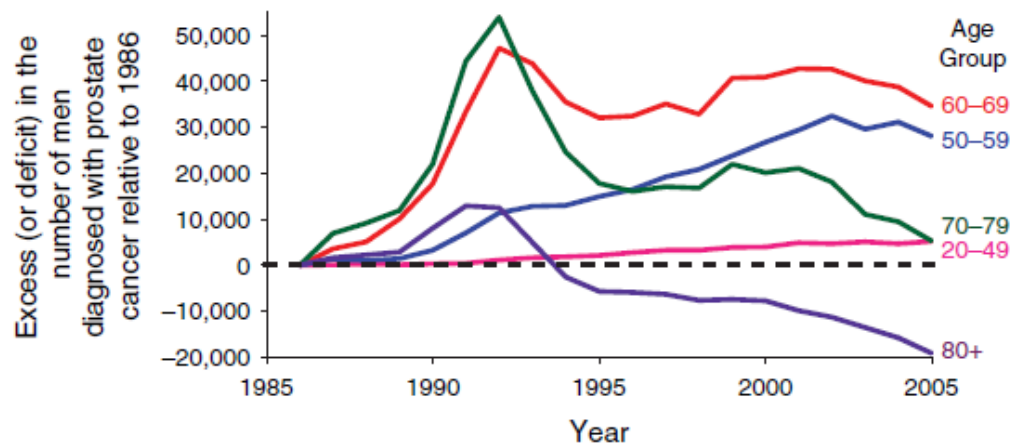


Figure 2. Excess (or deficit) in the number of men diagnosed with prostate cancer relative to 1986.

Prostate Cancer Diagnosis and Treatment After the Introduction of Prostate-Specific Antigen Screening: 1986–2005

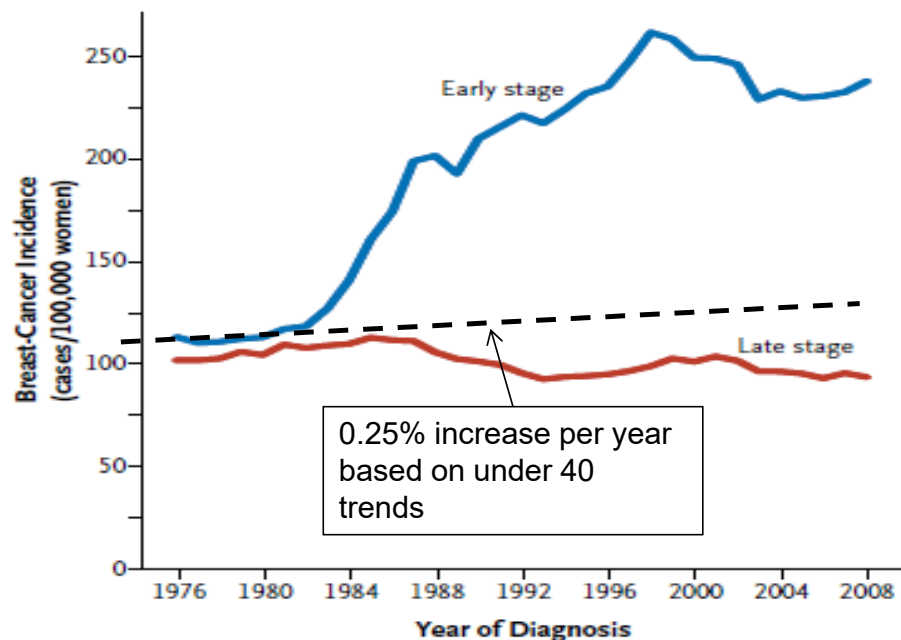
H. Gilbert Welch, Peter C. Albertsen

Since 1986, an estimated additional 1,305,600 men were diagnosed with prostate cancer

- Cumulative 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



The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Effect of Three Decades of Screening Mammography on Breast-Cancer Incidence

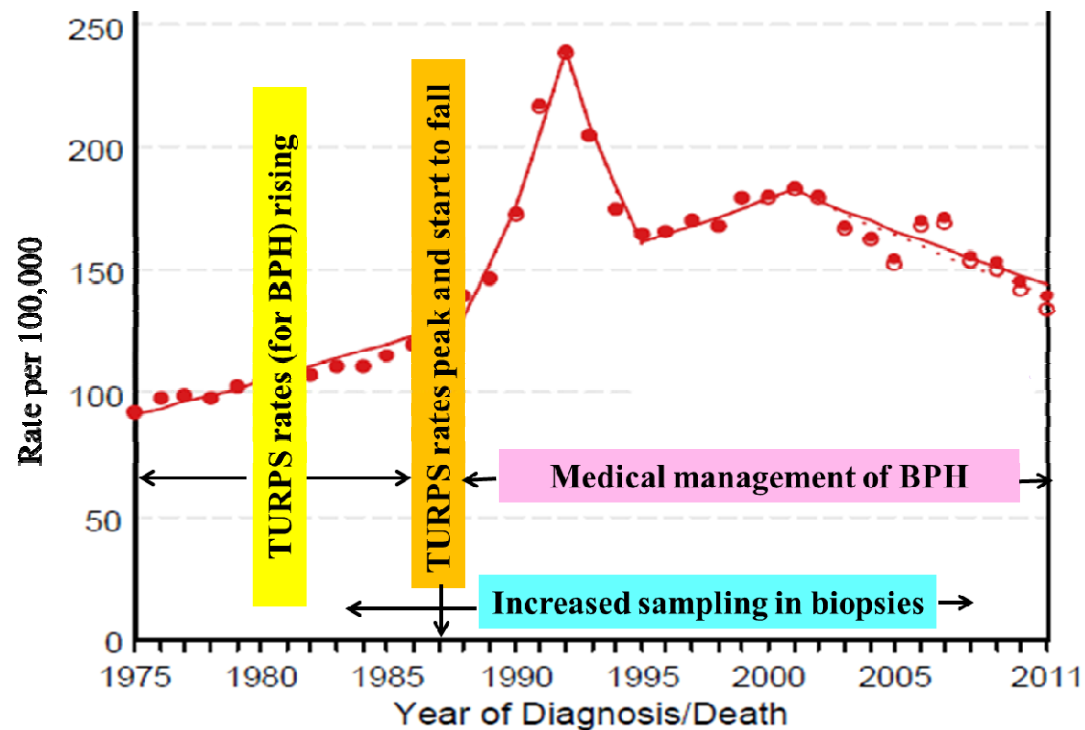
Archie Bleyer, M.D., and H. Gilbert Welch, M.D., M.P.H.

31% of detected cancers in 2008 overdiagnosed

- Annual excess incidence
- Background incidence imputed based on incidence trends in women under 40

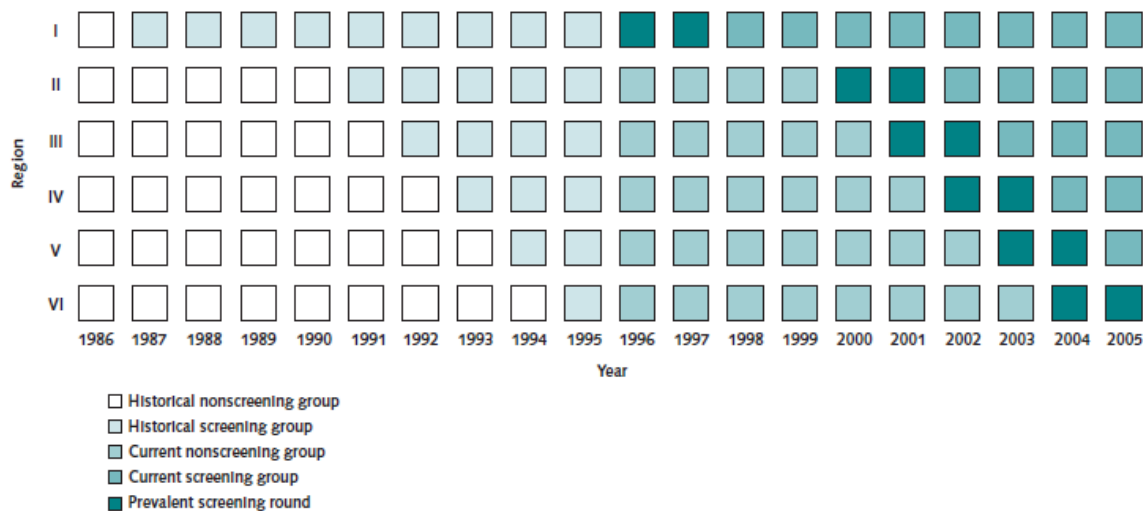
FIGURING OUT BACKGROUND INCIDENCE CAN BE HARD!

Prostate Cancer Incidence: White Males (SEER 9 Registries)



BREAST CANCER INCIDENCE IN NORWAY

Figure 1. Implementation of the different groups in the breast cancer screening program in different regions from 1986 to 2005.



Annals of Internal Medicine

ORIGINAL RESEARCH

Overdiagnosis of Invasive Breast Cancer Due to Mammography Screening: Results From the Norwegian Screening Program

Mette Kalager, MD; Hans-Olov Adami, MD, PhD; Michael Bretthauer, MD, PhD; and Rulla M. Tamimi, ScD

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?

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

By Yu Shen and Marvin Zelen

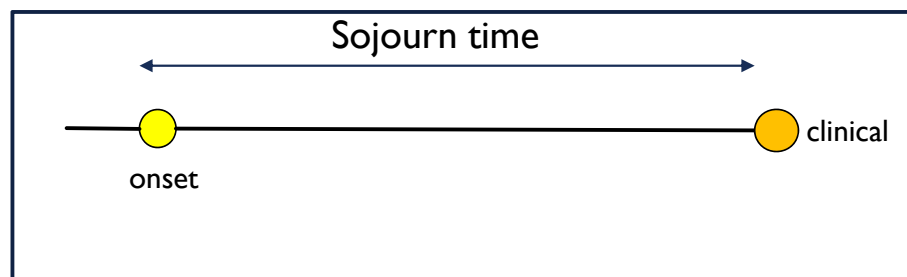
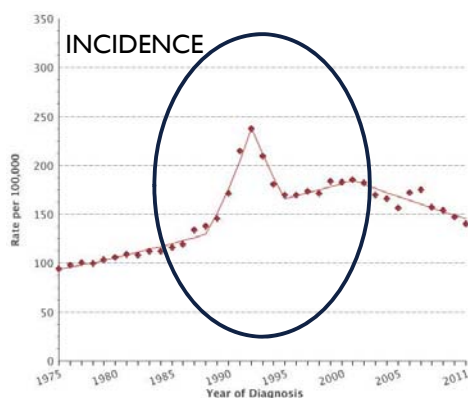
JCO 2001

	MST	
	Years	SD
HIP	2.5	1.2
Sweden		
Malmö	5.5	2.1
Stockholm ⁺		
40-49	2.1	1.3
50-59	2.6	.61
Edinburgh	4.3	.37
Canada 1	1.9	1.2
Canada 2	3.1	.94

Consensus mean 3.3 (~40 months)

GOING BEYOND THE DATA USING MODELING TO LEARN ABOUT OVERDIAGNOSIS

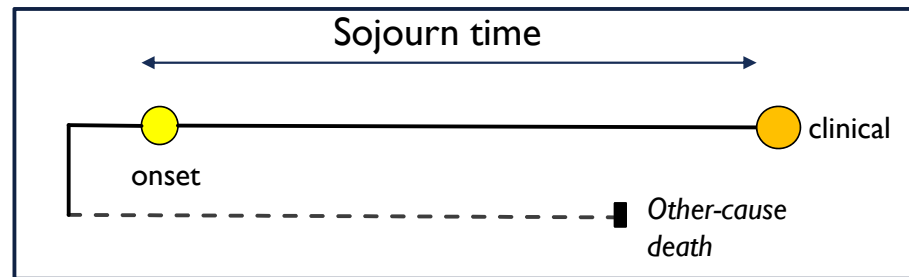
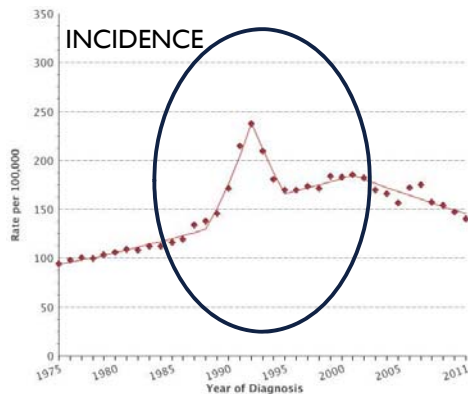
- I. Go beyond observed data to learn about underlying disease process
 - Given data on screening uptake
 - Use incidence before and after screening to learn about disease natural history



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

GOING BEYOND THE DATA USING MODELING TO LEARN ABOUT OVERDIAGNOSIS

- I. 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



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

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

Strategy and Age Group	
	All Overdiagnosed Cases, %†‡
Biennial	
50-74	12 (8-22)
45-74	12 (8-22)
40-74	13 (9-24)
Hybrid	
45-74	13 (8-25)
40-74	14 (9-27)
Annual	
50-74	15 (8-36)
45-74	17 (9-38)
40-74	18 (9-39)

(as a proportion of all cases detected)

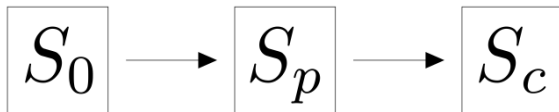
- Population study
 - A. Background incidence imputed based on age-period-cohort model (increasing trend)
 - B. Retrospective reconstruction of screening patterns
 - C. Identifiability? Model-dependent

THE IDENTIFIABILITY PROBLEM

CAN THE MODEL BE LEARNED FROM THE DATA?



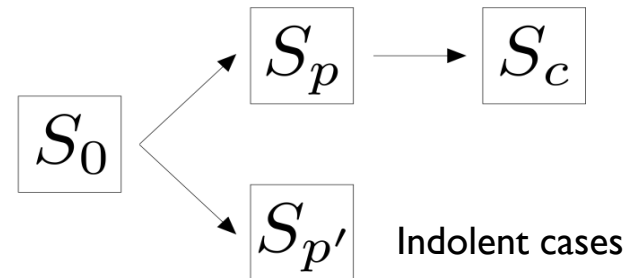
A) Progressive disease only



- 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

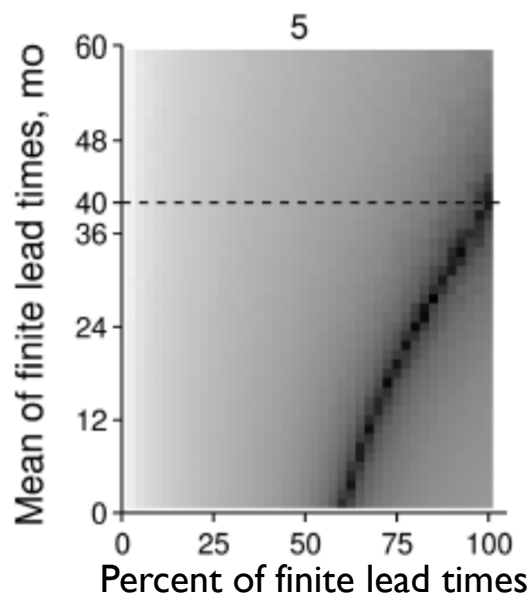
B) Mixture of progressive and indolent disease



- 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



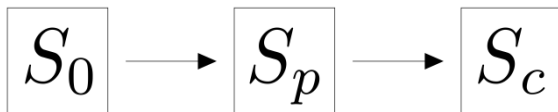
- Exponential mean 40 months
- Mixture of 75% exponential with mean 18 months, 25%(effectively) infinite
- Mixture of 95% exponential with mean 26 months, 5% infinite

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

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



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

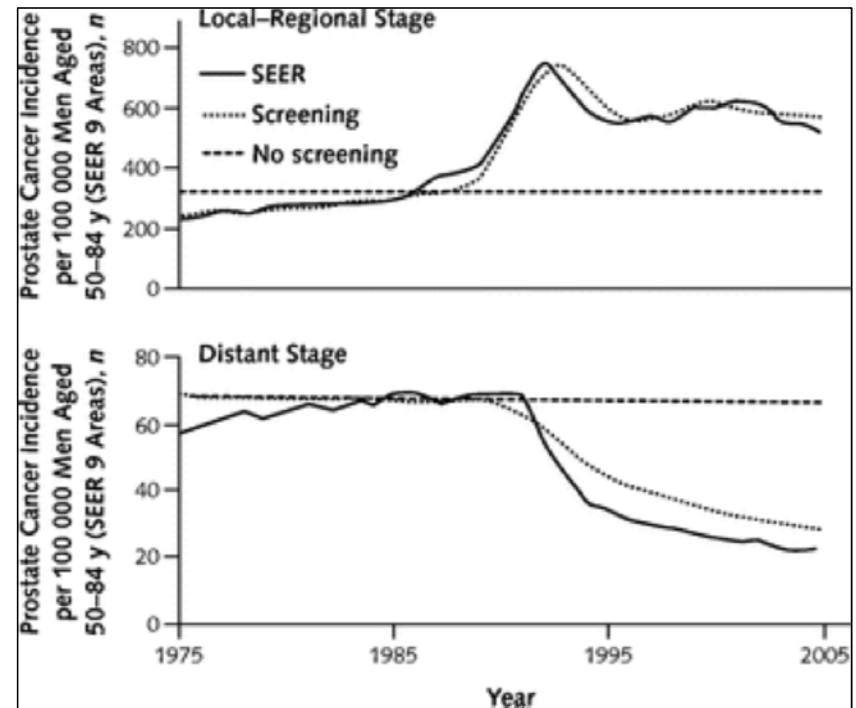
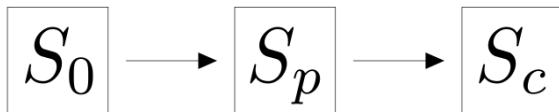
By Yu Shen and Marvin Zelen

	MST	
	Years	SD
HIP	2.5	1.2
Sweden		
Malmö	5.5	2.1
Stockholm ⁺		
40-49	2.1	1.3
50-59	2.6	.61
Edinburgh	4.3	.37
Canada 1	1.9	1.2
Canada 2	3.1	.94

PROSTATE CANCER NATURAL HISTORY FROM A POPULATION

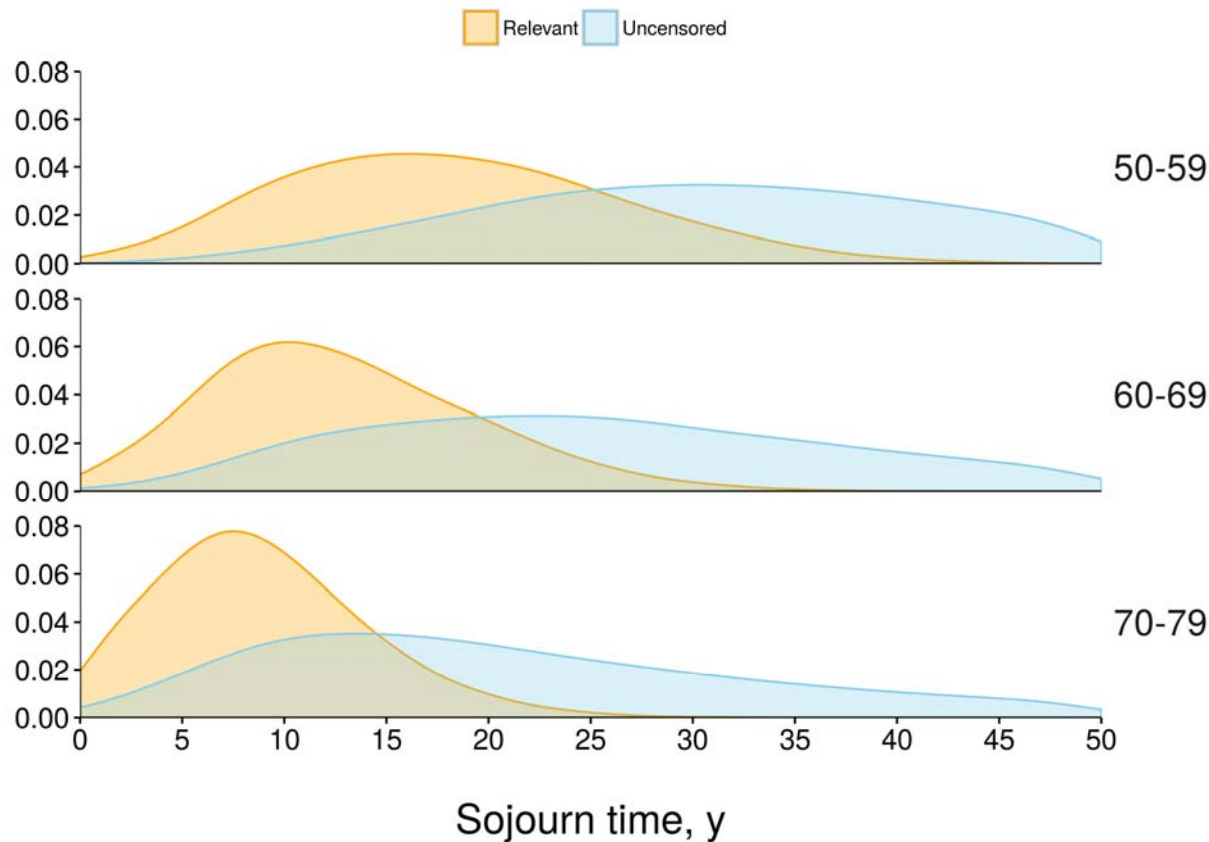
- A. Assume incidence in the absence of screening would have remained constant at pre-PSA rates
- B. Aggregate screening histories retrospectively constructed from NHIS and SEER-Medicare
- C. Progressive disease assumption – risk of progression to advanced or symptomatic disease depends on PSA growth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

A) Progressive disease only



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

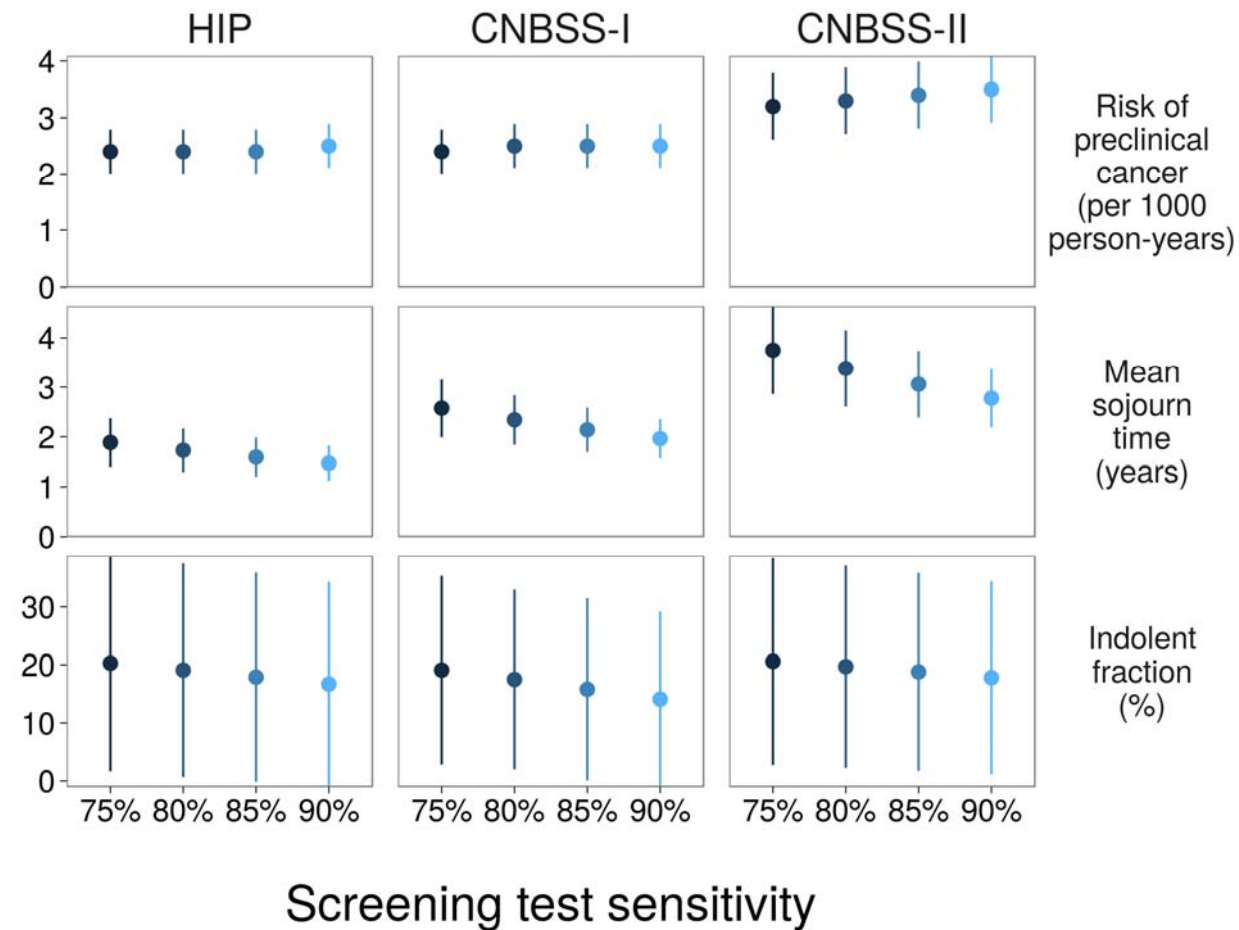
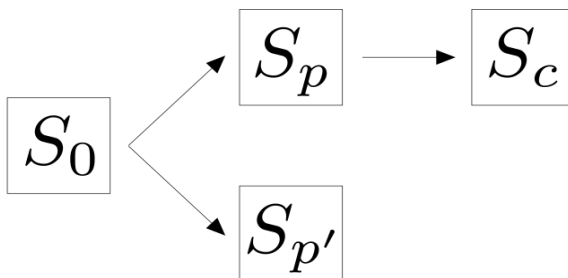
While a mixture model is not explicitly assumed, the model structure builds in heterogeneity

CAN WE ESTIMATE A MIXTURE MODEL?

Shen et al 2016

- 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



IDENTIFYING IDENTIFIABILITY (OR LACK THEREOF) CAN BE HARD

Annals of Internal Medicine

ORIGINAL RESEARCH

Collaborative Modeling of the Benefits and Harms Associated With Different U.S. Breast Cancer Screening Strategies

Strategy and Age Group	All Overdiagnosed Cases, %†‡
Biennial	
50-74	12 (8-22)
45-74	12 (8-22)
40-74	13 (9-24)
Hybrid	
45-74	13 (8-25)
40-74	14 (9-27)
Annual	
50-74	15 (8-36)
45-74	17 (9-38)
40-74	18 (9-39)

- 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



FRED HUTCH
CURES START HERE®

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

QUESTIONS?

Send questions to prevention@mail.nih.gov

Or

Use [@NIHprevents](https://twitter.com/NIHprevents) & [#NIH MtG](https://twitter.com/NIHprevents) on Twitter