HOW BIG ARE THE REAL MORTALITY REDUCTIONS PRODUCED BY CANCER SCREENING?

WHY DO SO MANY TRIALS REPORT ONLY 20%?

James A. Hanley[1], Zhihui (Amy) Liu[1], Nandini Dendukuri[1,2], Erin Strumpf[1,3]

[1]Dept. of Epidemiology, Biostatistics & Occupational Health
[2]Dept. of Medicine, and Technology Assessment Unit
[3]Dept. of Economics

McGill University, Montréal, Québec, CANADA

(SSC-sponsored) JSM Invited Session
Montréal
Aug 7, 2013

Summary: the 3 points I wish to make

Summary: the 3 points I wish to make

- With their blindness to the delay until the reductions in mortality are expressed,

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

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 ***under-estimate*** the mortality reductions that ***would be produced by a sustained screening program***

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 ***under-estimate*** the mortality reductions that ***would be produced by a sustained screening program***
- P-value-driven stopping rules exacerbate the underestimation

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 ***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

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 ***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

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 ***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

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 ***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

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 ***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 / non-proportional hazards

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 ***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 / non-proportional hazards
  (v) focus on the parameters that describe impact of 1 round of screening

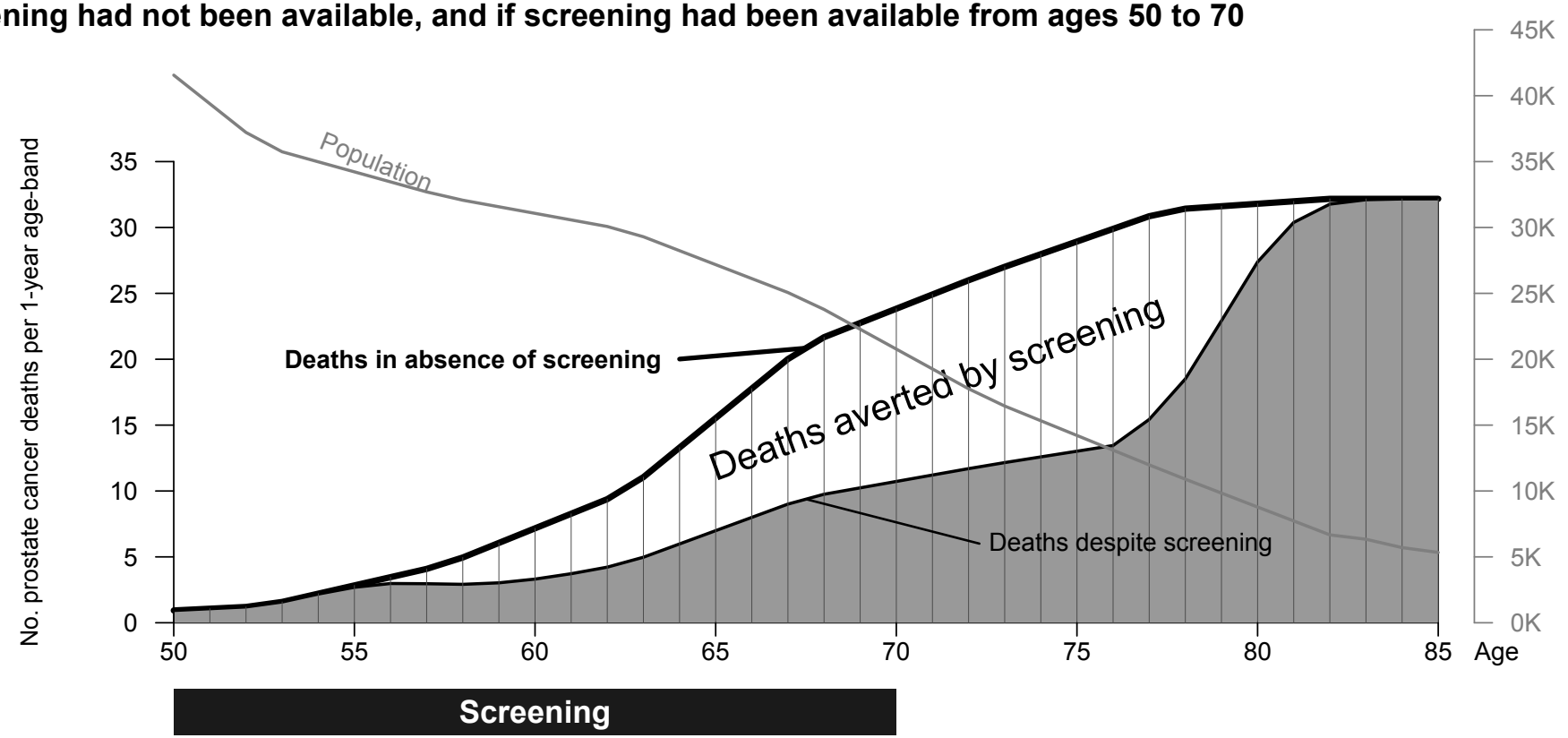
Outline

- The mortality reductions produced by a screening regimen: what payers want to know
- European Randomized Study of Screening for Prostate Cancer
  [and Göteborg portion of this study]
- Data-analysis practice in other cancer screening trials
- How to stop a screening RCT at a 20% mortality reduction? [Theorem]
- A way ahead?

What payers would like to know...

What payers would like to know...

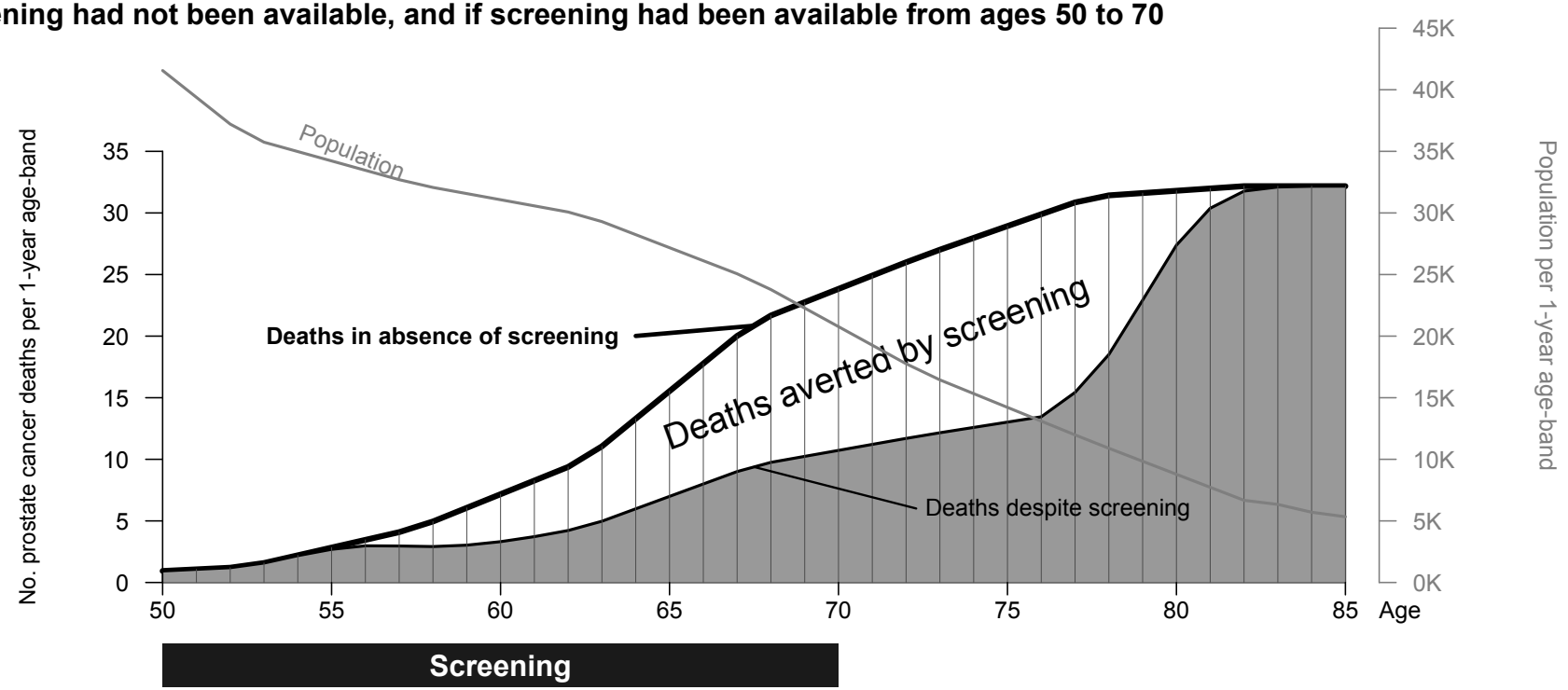
(a) Age-specific numbers of prostate cancer deaths in a steady state population with a given age-structure, if screening had not been available, and if screening had been available from ages 50 to 70



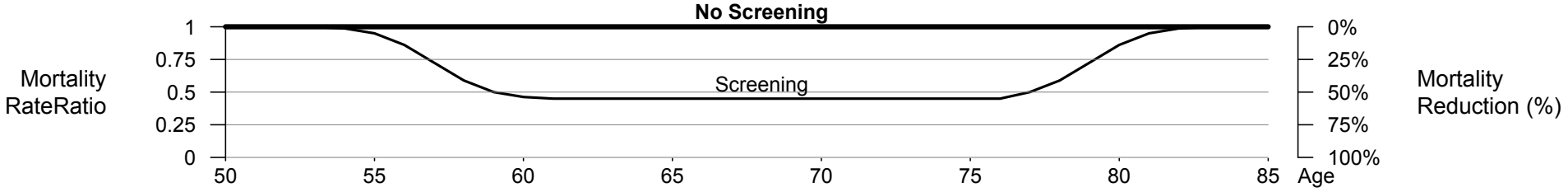
They could arrive at these numbers if they had...

They could arrive at these numbers if they had...

(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)

Screening & Prostate-Ca Mortality in Randomized European Study ("ERSPC" nejm2009.04)

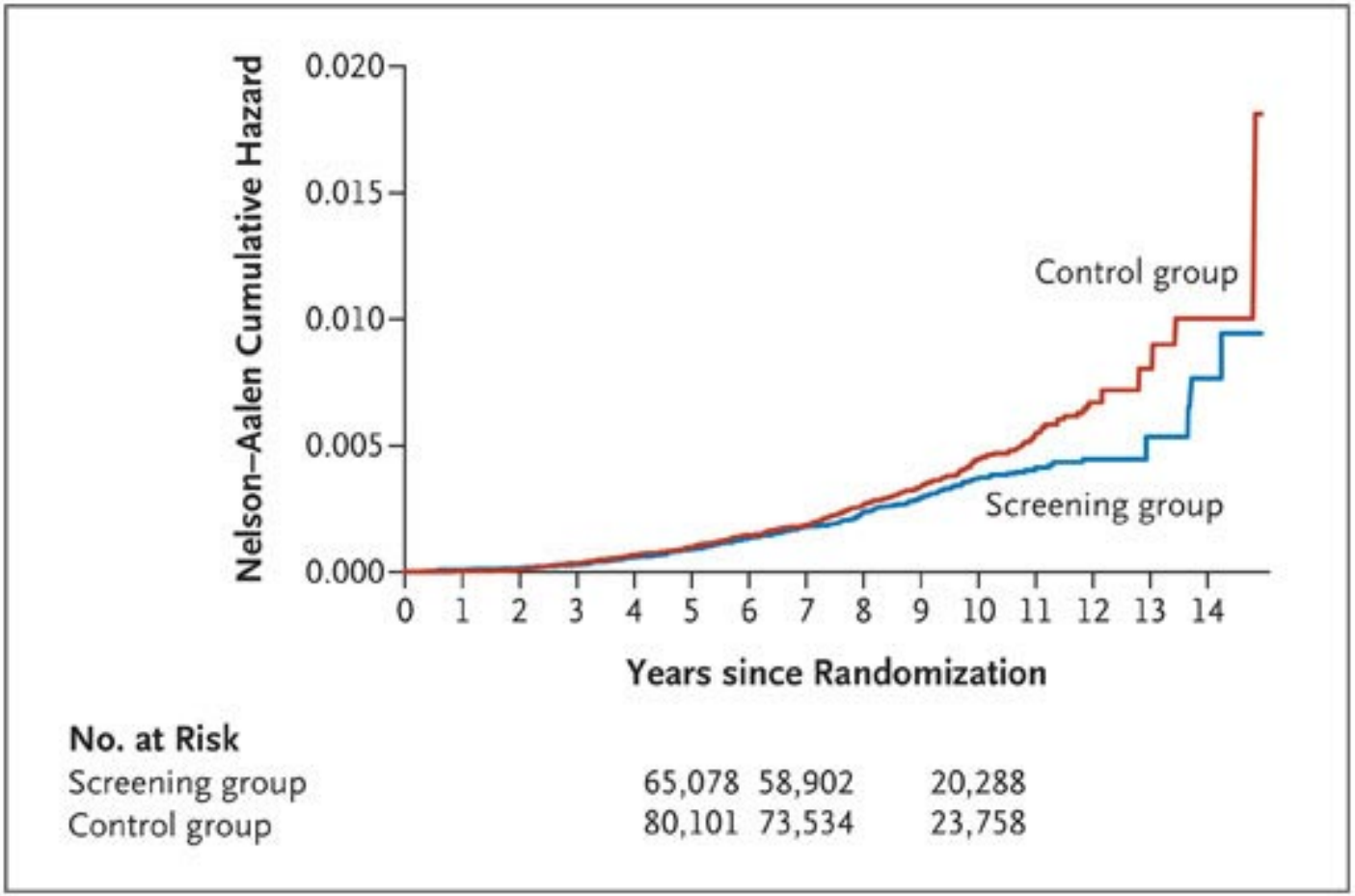
As of December 31, 2006, with an average follow-up time of 8.8 years, there were 214 prostate-cancer deaths in the screening group and 326 in the control group. (...) The adjusted **rate ratio** for death from prostate cancer in the screening group was **0.80** (95% CI, 0.65 to 0.98; P=0.04).

"PSA-based screening reduced the rate of death from prostate cancer by **20%**. "

Screening & Prostate-Ca Mortality in Randomized European Study ("ERSPC" nejm2009.04)

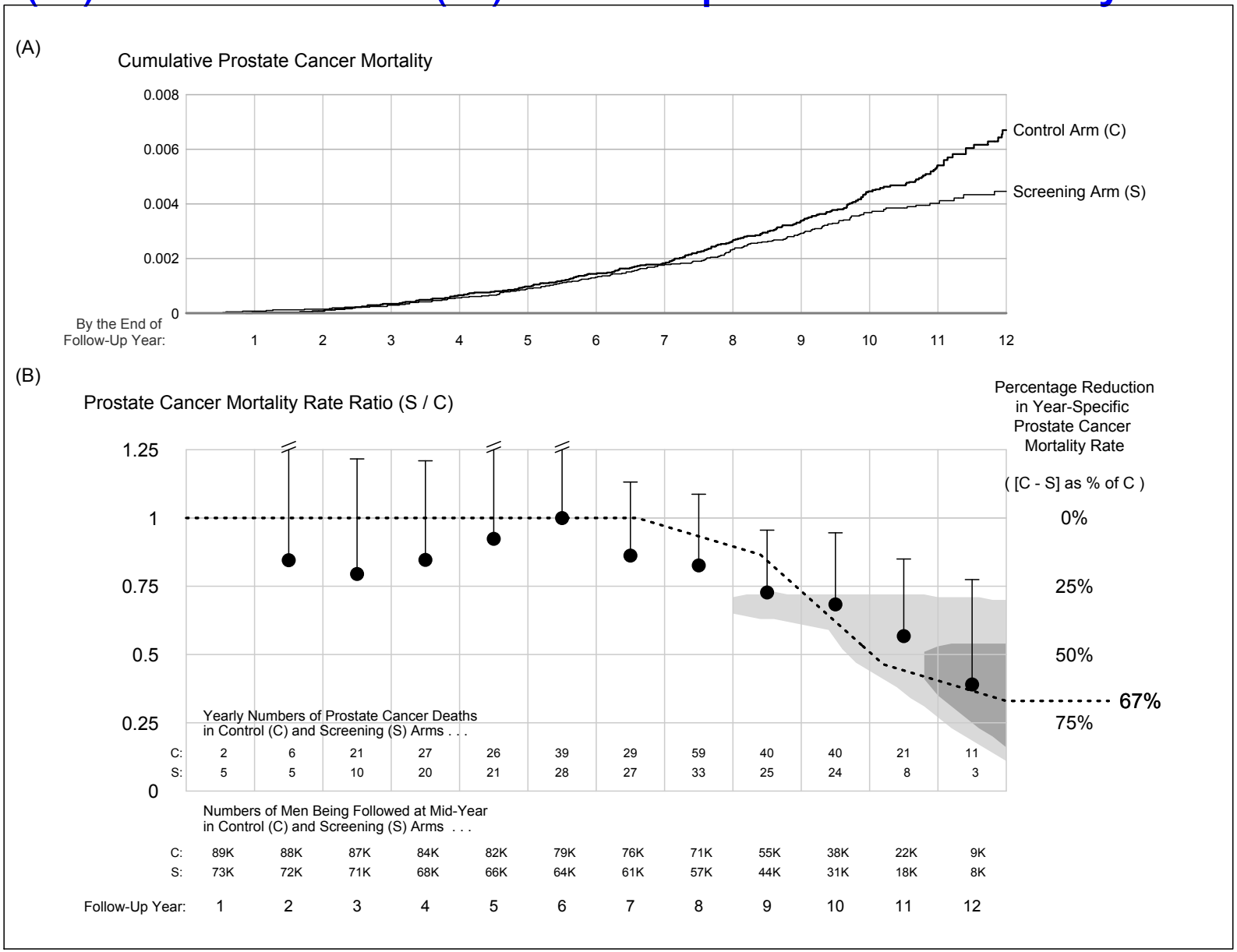
As of December 31, 2006, with an average follow-up time of 8.8 years, there were 214 prostate-cancer deaths in the screening group and 326 in the control group. (...) The adjusted **rate ratio** for death from prostate cancer in the screening group was **0.80** (95% CI, 0.65 to 0.98; P=0.04).

"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



Hanley, *J Medical Screening*, 2010.

Göteborg randomised population-based prostate-cancer screening trial

Göteborg randomised population-based prostate-cancer screening trial

Methods In December, 1994, 20 000 men born between 1930 and 1944, randomly sampled from the population register, were randomised by computer in a 1:1 ratio to either a screening group invited for PSA testing every 2 years (n=10 000) or to a control group not invited (n=10 000). Men in the screening group were invited up to the upper age limit (median 69, range 67–71 years) and only men with raised PSA concentrations were offered additional tests such as digital rectal examination and prostate biopsies. The primary endpoint was prostate-cancer specific mortality, analysed according to the intention-to-screen principle. The study is ongoing, with men who have not reached the upper age limit invited for PSA testing. This is the first planned report on cumulative prostate-cancer incidence and mortality calculated up to Dec 31, 2008. This study is registered as an International Standard Randomised Controlled Trial ISRCTN54449243.

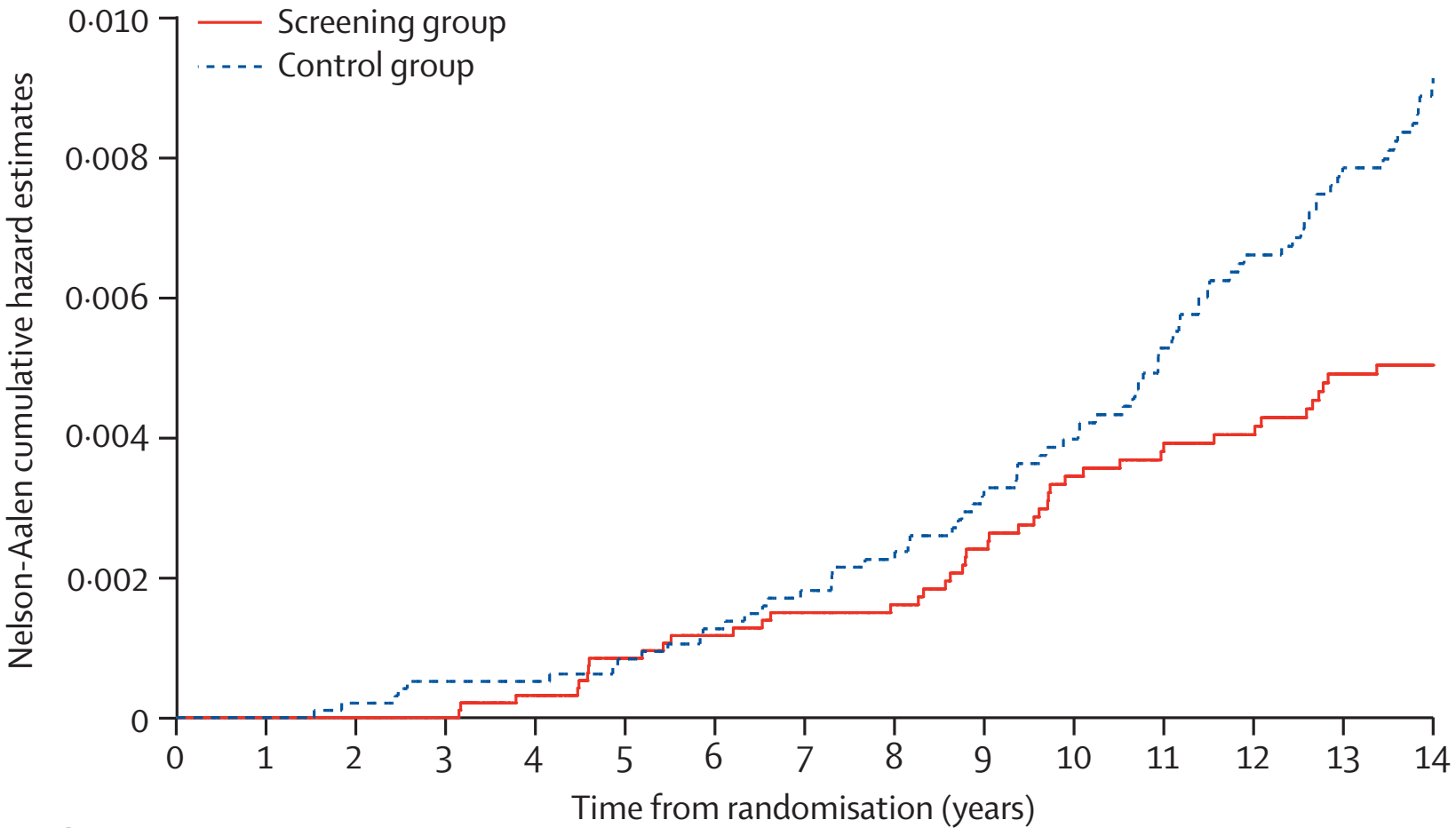
Findings In each group, 48 men were excluded from the analysis because of death or emigration before the randomisation date, or prevalent prostate cancer. In men randomised to screening, 7578 (76%) of 9952 attended at least once. During a median follow-up of 14 years, 1138 men in the screening group and 718 in the control group were diagnosed with prostate cancer, resulting in a cumulative prostate-cancer incidence of 12·7% in the screening group and 8·2% in the control group (hazard ratio 1·64; 95% CI 1·50–1·80; p<0·0001). The absolute cumulative risk reduction of death from prostate cancer at 14 years was 0·40% (95% CI 0·17–0·64), from 0·90% in the control group to 0·50% in the screening group. The rate ratio for death from prostate cancer was 0·56 (95% CI 0·39–0·82; p=0·002) in the screening compared with the control group. The rate ratio of death from prostate cancer for attendees compared with the control group was 0·44 (95% CI 0·28–0·68; p=0·0002). Overall, 293 (95% CI 177–799) men needed to be invited for screening and 12 to be diagnosed to prevent one prostate cancer death.

Interpretation This study shows that prostate cancer mortality was reduced almost by half over 14 years. However, the risk of over-diagnosis is substantial and the number needed to treat is at least as high as in breast-cancer screening programmes. The benefit of prostate-cancer screening compares favourably to other cancer screening programs.

Hugosson et al. Lancet Oncol. 2010 Aug;11(8):725-32. Epub 2010 Jul 2.

Mortality Results

Mortality Results



Number at risk

Screening group	9952	9333	8585	7746
Control group	9952	9345	8580	7755

***Figure 3*: Cumulative risk of death from prostate cancer using Nelson-Aalen cumulative hazard estimates**

YEARLY numbers of Pr Ca Deaths in Control and Screening groups

YEARLY numbers of Pr Ca Deaths in Control and Screening groups

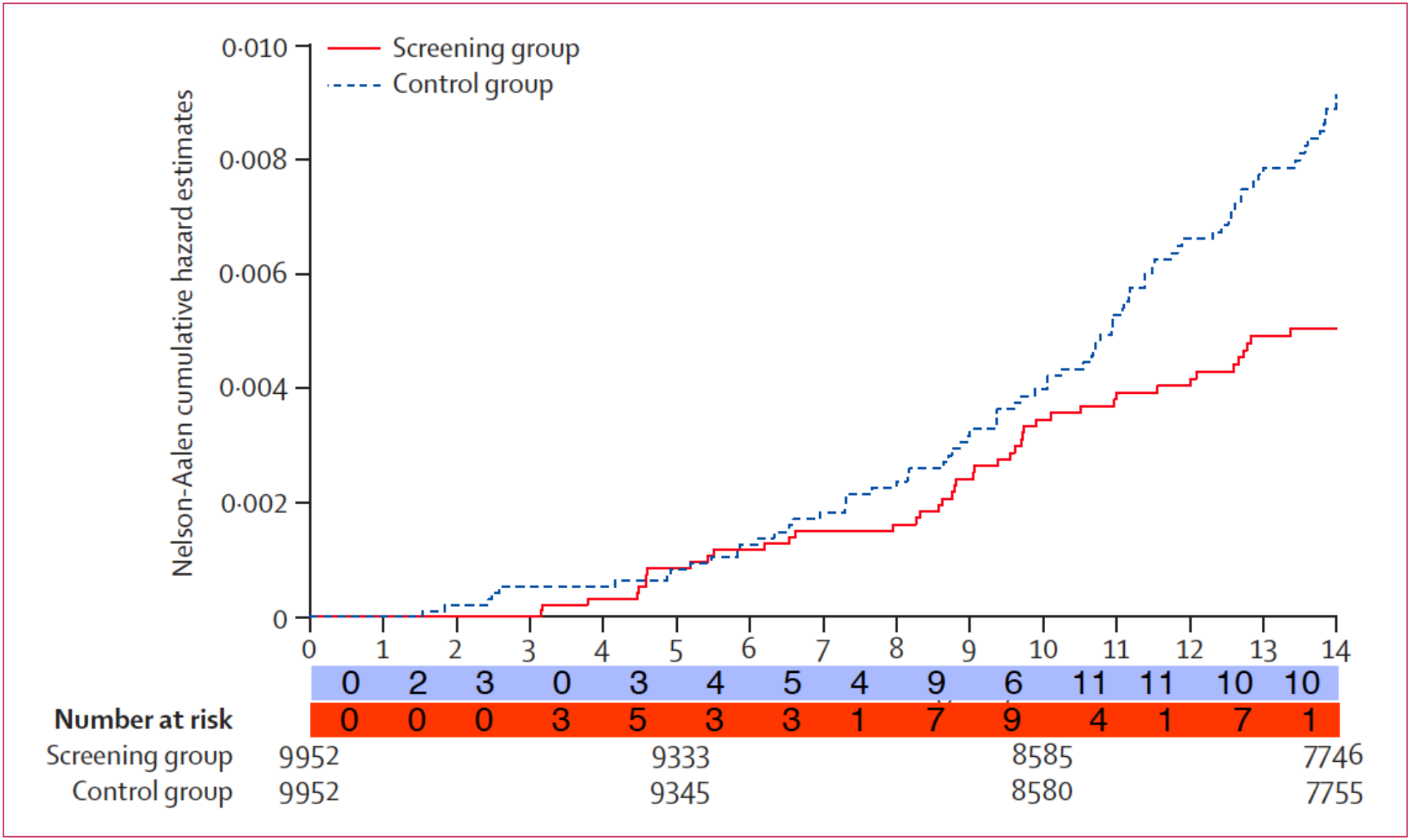
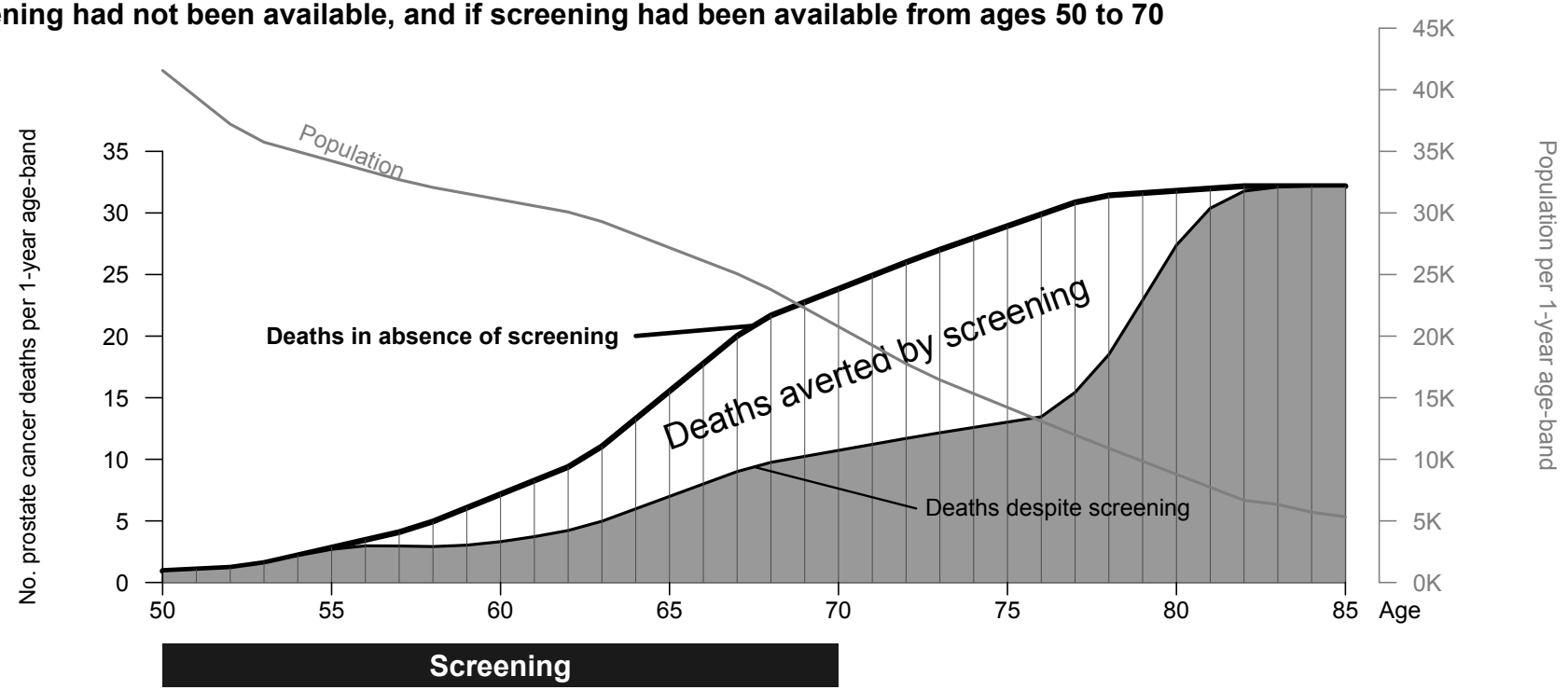


Figure 3: **Cumulative risk of death from prostate cancer using Nelson-Aalen cumulative hazard estimates**

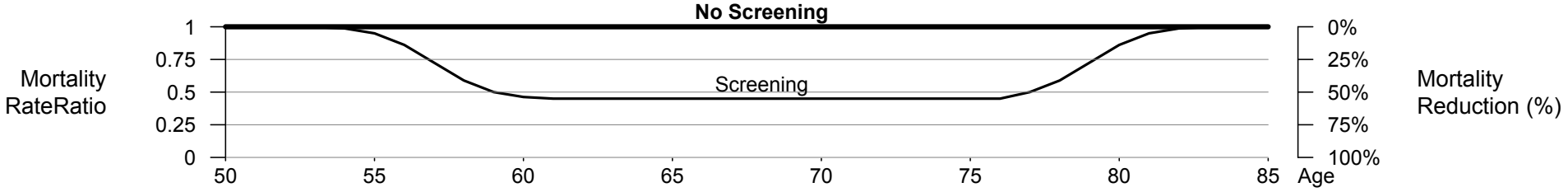
They could arrive at these numbers if they had...

They could arrive at these numbers if they had...

(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



BREAST CANCER

BREAST CANCER

EVERY TRIAL & META-ANALYSIS:

and (nejm2010) REPORT on NORWAY NATIONAL SCREENING PROGRAM:

REDUCTION UNDER-ESTIMATED

BREAST CANCER

EVERY TRIAL & META-ANALYSIS:

and (nejm2010) REPORT on NORWAY NATIONAL SCREENING PROGRAM:

REDUCTION UNDER-ESTIMATED

- Miettinen et al., *Lancet* 2002.

BREAST CANCER

EVERY TRIAL & META-ANALYSIS:

and (nejm2010) REPORT on NORWAY NATIONAL SCREENING PROGRAM:

REDUCTION UNDER-ESTIMATED

- Miettinen et al., *Lancet* 2002.
- Hanley, *Epidemiologic Reviews* 2011.

BREAST CANCER

EVERY TRIAL & META-ANALYSIS:

and (nejm2010) REPORT on NORWAY NATIONAL SCREENING PROGRAM:

REDUCTION UNDER-ESTIMATED

- Miettinen et al., *Lancet* 2002.
- Hanley, *Epidemiologic Reviews* 2011.
- Hanley, Liu, Strumpf, Dendukuri, McGregor.
  "No.s of breast cancer deaths averted by mammography screening".
  (Response to Canadian Task Force on Preventive Health Care)
  ... manuscript under review at Canadian J Public Health
- Hanley JA, Z Liu Z, McGregor M. The [ratio of] benefits [to] harms of breast cancer screening. Letter re the Report The Independent UK Panel on Breast Cancer Screening (Lancet Nov 17, 2012)

LUNG CANCER

Deaths from lung cancer in the NLST, with corresponding relative deficit in CT arm

Deaths from lung cancer in the NLST, with corresponding relative deficit in CT arm

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

Follow-up Year:	1	2	3	4	5	6	7	ALL
Screens	↑	↑	↑					
X-ray Arm:								442
CT Arm:								354
Relative Deficit:								20%

Deaths from lung cancer in the NLST, with corresponding relative deficit in CT arm

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

Follow-up Year:	1	2	3	4	5	6	7	ALL
Screens	↑	↑	↑					
X-ray Arm:								442
CT Arm:								354
Relative Deficit:								20%

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%		

Deaths from lung cancer in the NLST, with corresponding relative deficit in CT arm

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

Follow-up Year:	1	2	3	4	5	6	7	ALL
Screens	↑	↑	↑					
X-ray Arm:								442
CT Arm:								354
Relative Deficit:								20%

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

20% MORTALITY REDUCTION

A UNIVERSAL CONSTANT IN SCREENING TRIALS?

Reductions in ‘event rates’: 3 ‘prevention’ studies

Reductions in ‘event rates’: 3 ‘prevention’ studies

- HPV 6,11,16,18 infection:
  - Quadrivalent human papillomavirus (HPV) vaccine

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*

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*

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...

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*) ?

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...

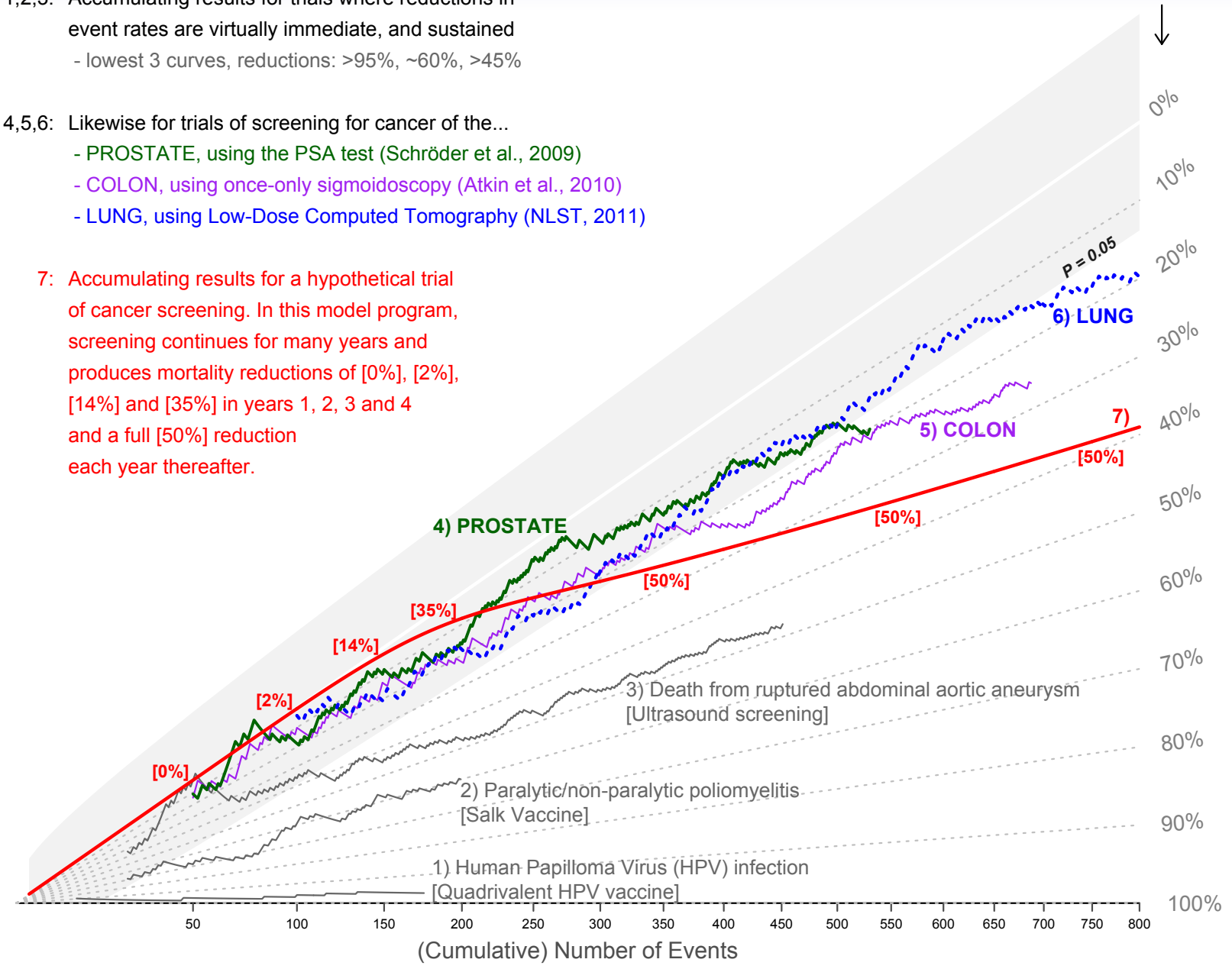
- all events up to that point in f-up time? (*1 'average' rate*) ?
- when in f-up time events occurred (*'time-specific'* rates) ?

Percentage Reduction in Average Event Rate, if data are analyzed after indicated no. of events

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.



PLANS

Data and Methods, Parameters, their Use

Data and Methods, Parameters, their Use

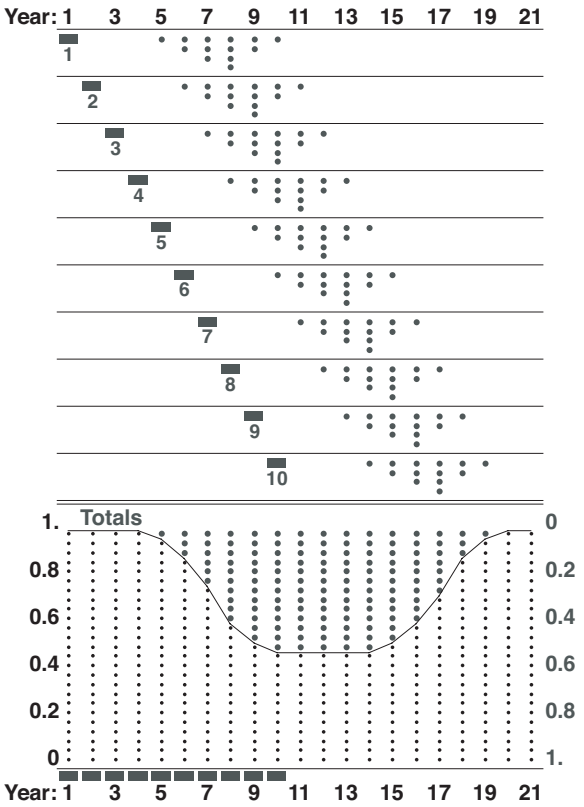
- **Data**: completed RCTs of screening for prostate, breast, colon and lung ca; population-based screening programs.

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:

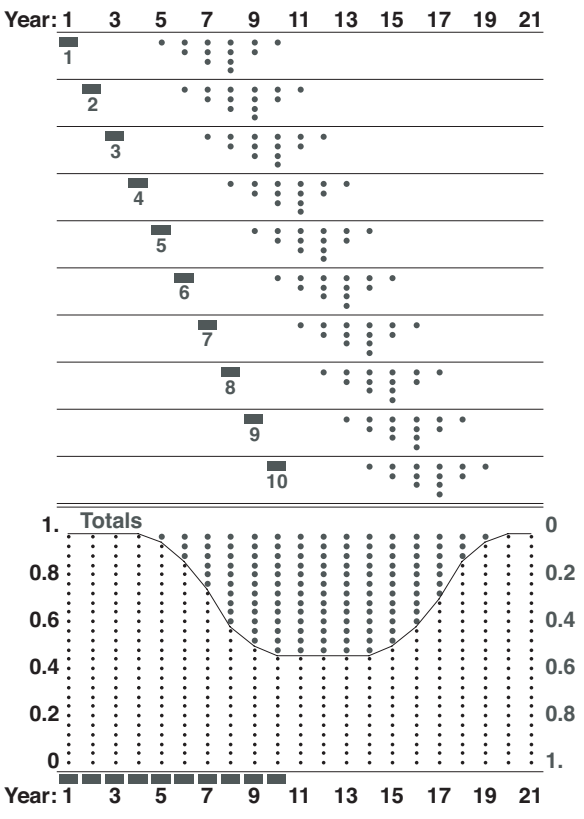
Data and Methods, Parameters, their Use

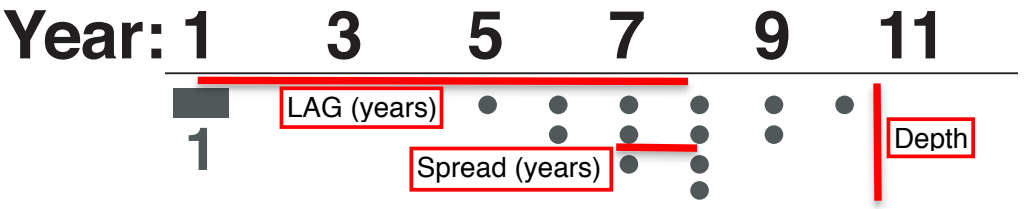
- **Data**: completed RCTs of screening for prostate, breast, colon and lung ca; population-based screening programs.
- **3 Parameters** (‘*deliverables*’) and how they will be fitted:



Data and Methods, Parameters, their Use

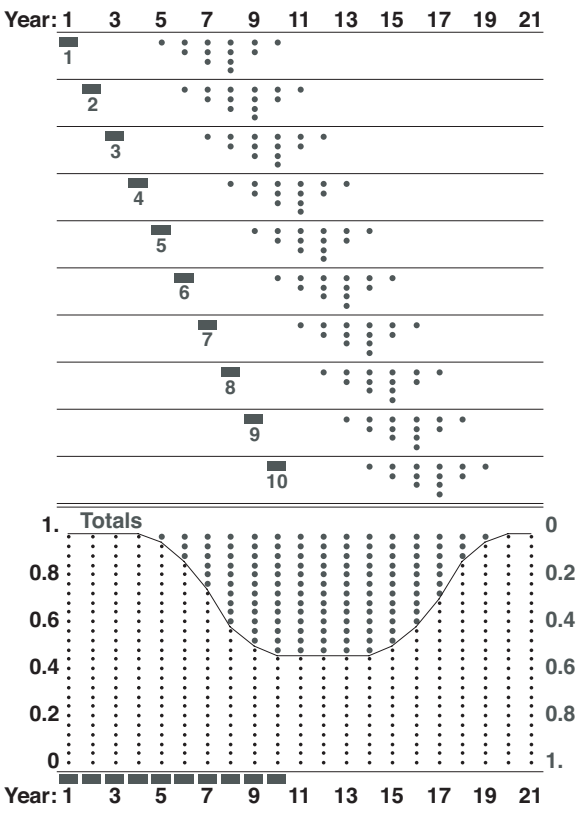
- **Data**: completed RCTs of screening for prostate, breast, colon and lung ca; population-based screening programs.
- **3 Parameters** (‘*deliverables*’) and how they will be fitted:

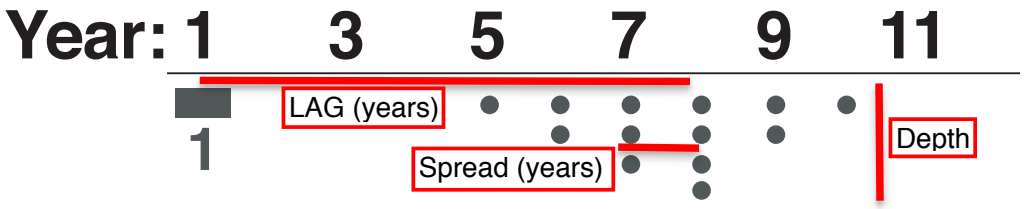




Data and Methods, Parameters, their Use

- **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:

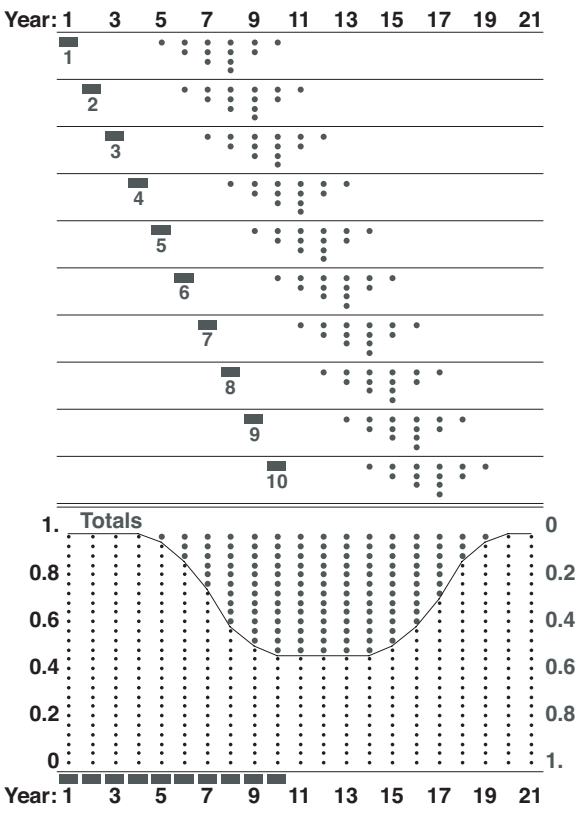


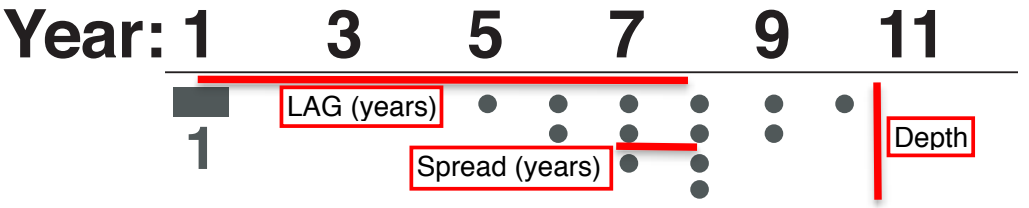


y = years since screening commenced

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:





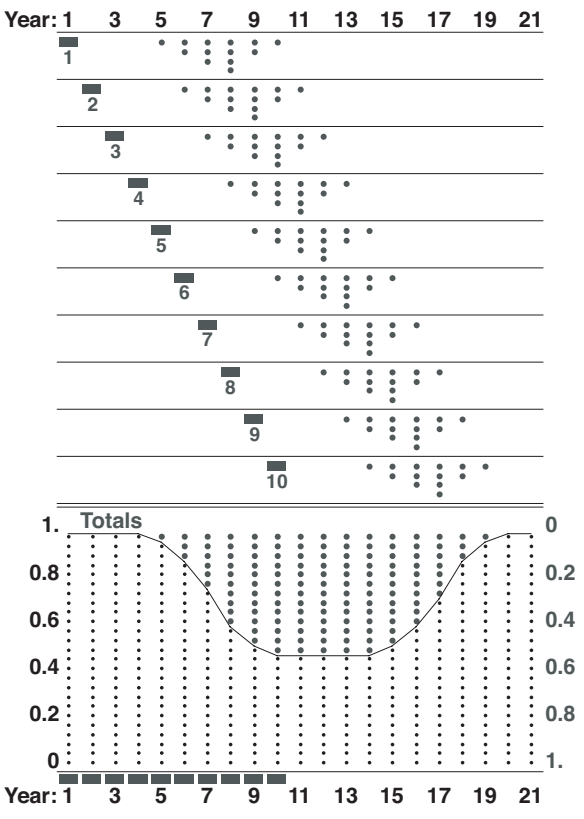
y = years since screening commenced

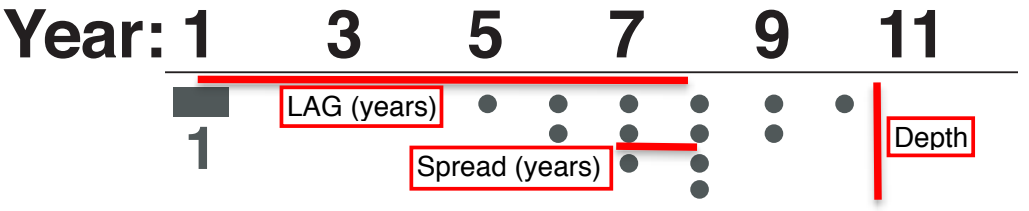
- Rate ratio in Year y, Age a in Study s :

RateRatio(y, a, s) =

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:





y = years since screening commenced

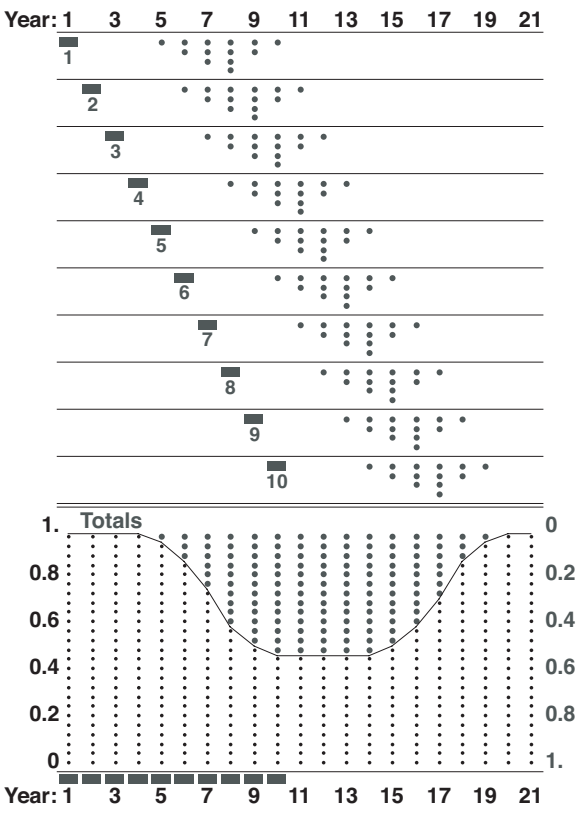
- Rate ratio in Year y, Age a in Study s :

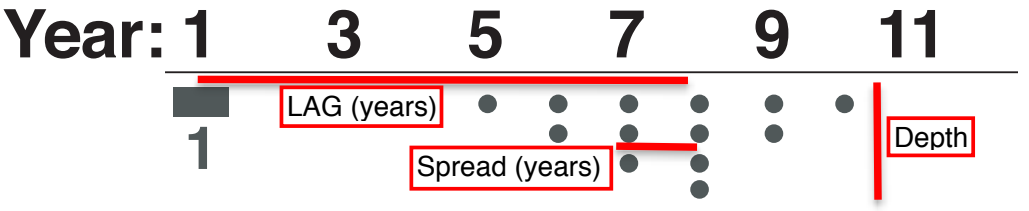
RateRatio(y, a, s) =

sum of reductions from all previous rounds of screening in study s

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:





y = years since screening commenced

- Rate ratio in Year y, Age a in Study s :

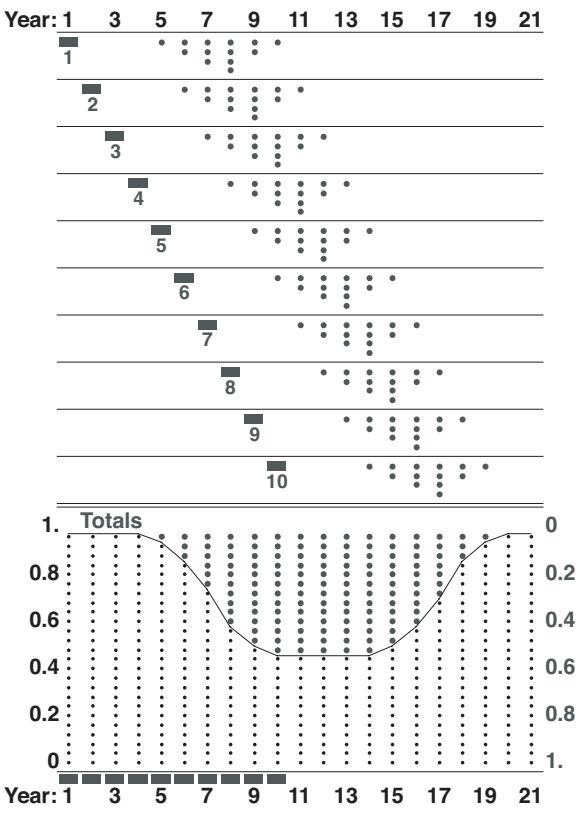
RateRatio(y, a, s) =

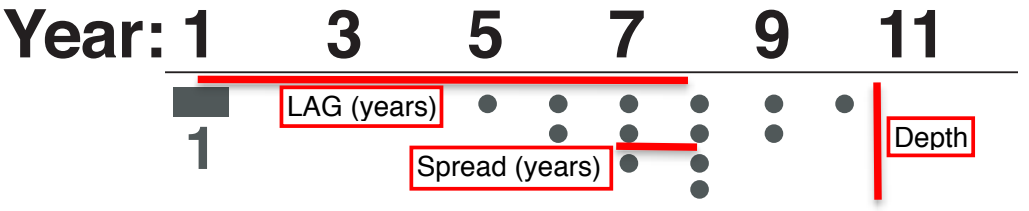
sum of reductions from all previous rounds of screening in study s

- Design matrix: 1 row per y-a-s ‘cell’

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:





y = years since screening commenced

- Rate ratio in Year y, Age a in Study s :

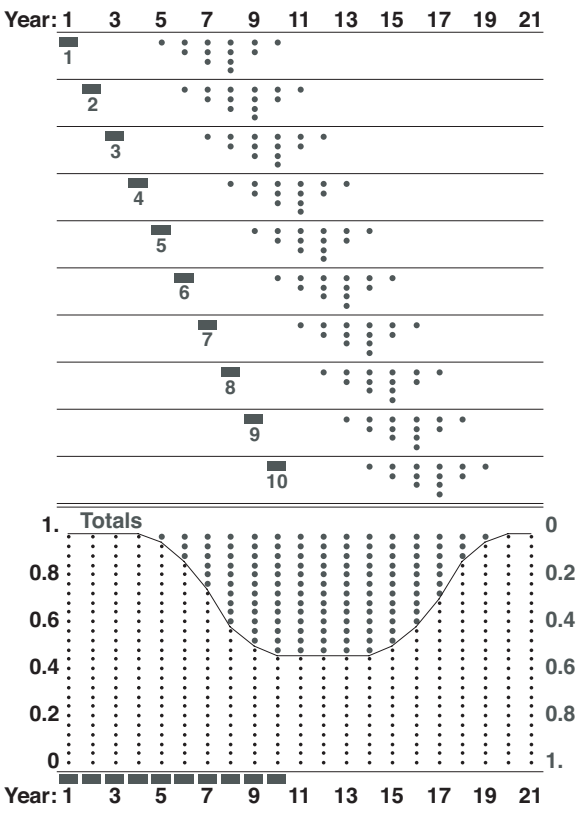
RateRatio(y, a, s) =

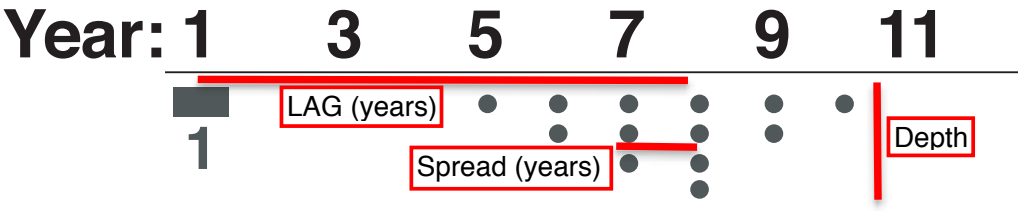
sum of reductions from all previous rounds of screening in study s

- Design matrix: 1 row per y-a-s ‘cell’
- $\frac{\text{No. deaths in screening arm}}{\text{No. deaths in 2 arms combined}}$ in each ‘cell’

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:





y = years since screening commenced

- Rate ratio in Year y, Age a in Study s :

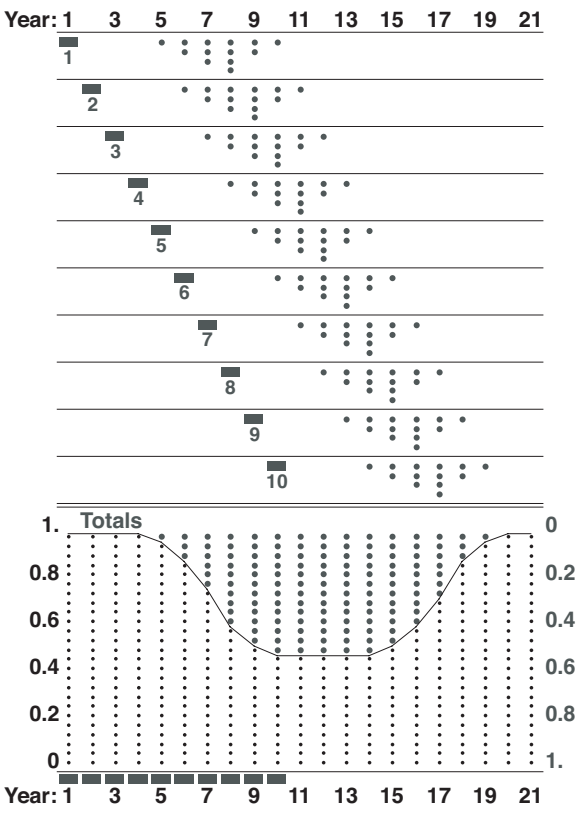
RateRatio(y, a, s) =

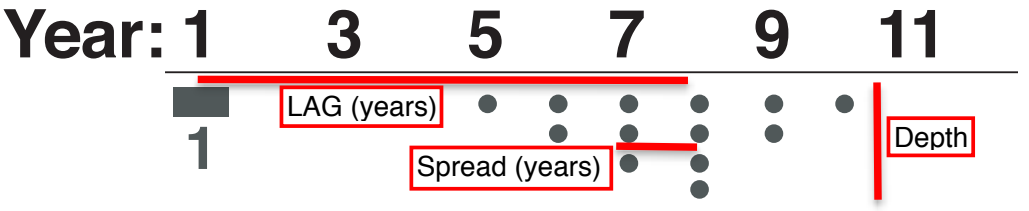
sum of reductions from all previous rounds of screening in study s

- Design matrix: 1 row per y-a-s ‘cell’
- $\frac{\text{No. deaths in screening arm}}{\text{No. deaths in 2 arms combined}}$ in each ‘cell’
- Fit by Max. Likelihood (binomial model)

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:





y = years since screening commenced

- Rate ratio in Year y, Age a in Study s :

RateRatio(y, a, s) =

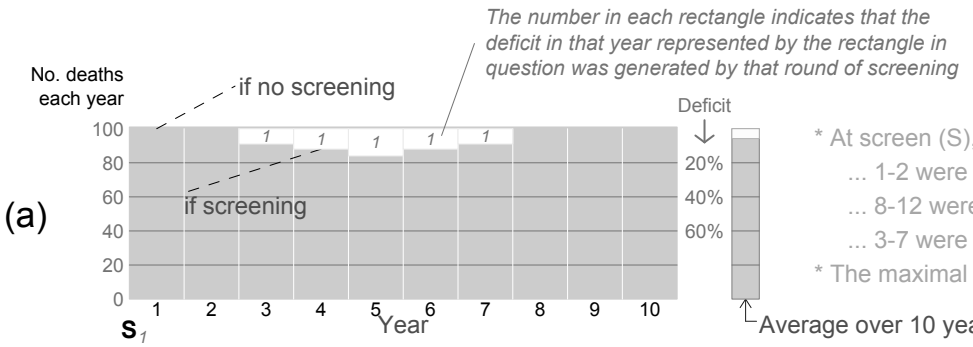
sum of reductions from all previous rounds of screening in study s

- Design matrix: 1 row per y-a-s ‘cell’
- $\frac{\text{No. deaths in screening arm}}{\text{No. deaths in 2 arms combined}}$ in each ‘cell’
- Fit by Max. Likelihood (binomial model)

- **USE**: project mort. reductions due to a screening regimen

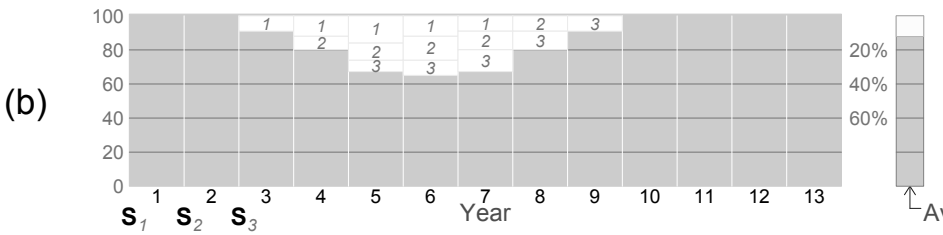
Mortality deficits produced by 1 or more rounds of screening

Mortality deficits produced by 1 or more rounds of screening

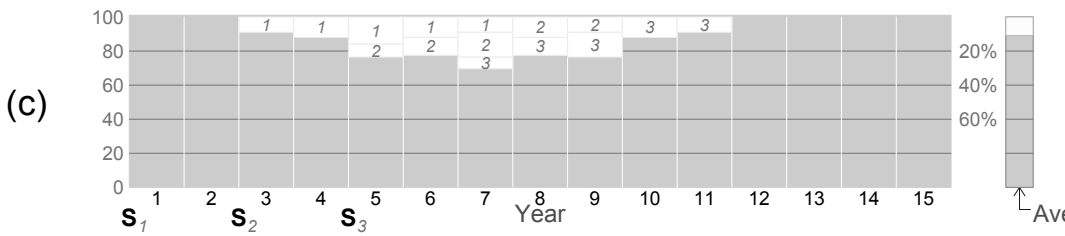


Description/remarks

* At screen (S), most/some cancers that -- S absent -- would have proved fatal in years
  ... 1-2 were probably screen-detectable, but already incurable [null mortality deficit];
  ... 8-12 were probably curable, but still undetectable [null mortality deficit];
  ... 3-7 were probably screen-detectable, and curable [some mortality deficits].
* The maximal mortality deficit produced by 1 round of screening is in year 5.



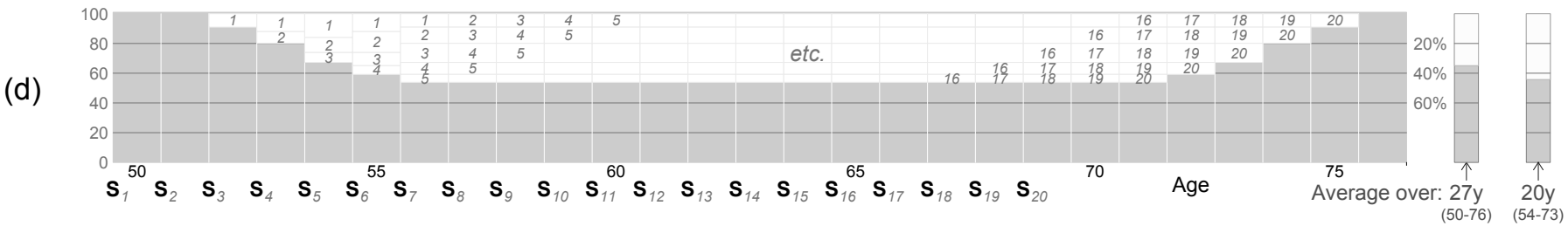
* 3 rounds of screening, 1 year apart, produce greater mortality deficits over a more prolonged time-window.
  [numbers in rectangles indicate deficits generated by successive rounds]
* The maximal mortality deficit, in year 6, is SMALLER THAN those produced by SUSTAINED screening: see asymptote in (d) below.



* 3 rounds of screening, 2 years apart, produce deficits shallower than those in (b), but extending over a wider time-window.

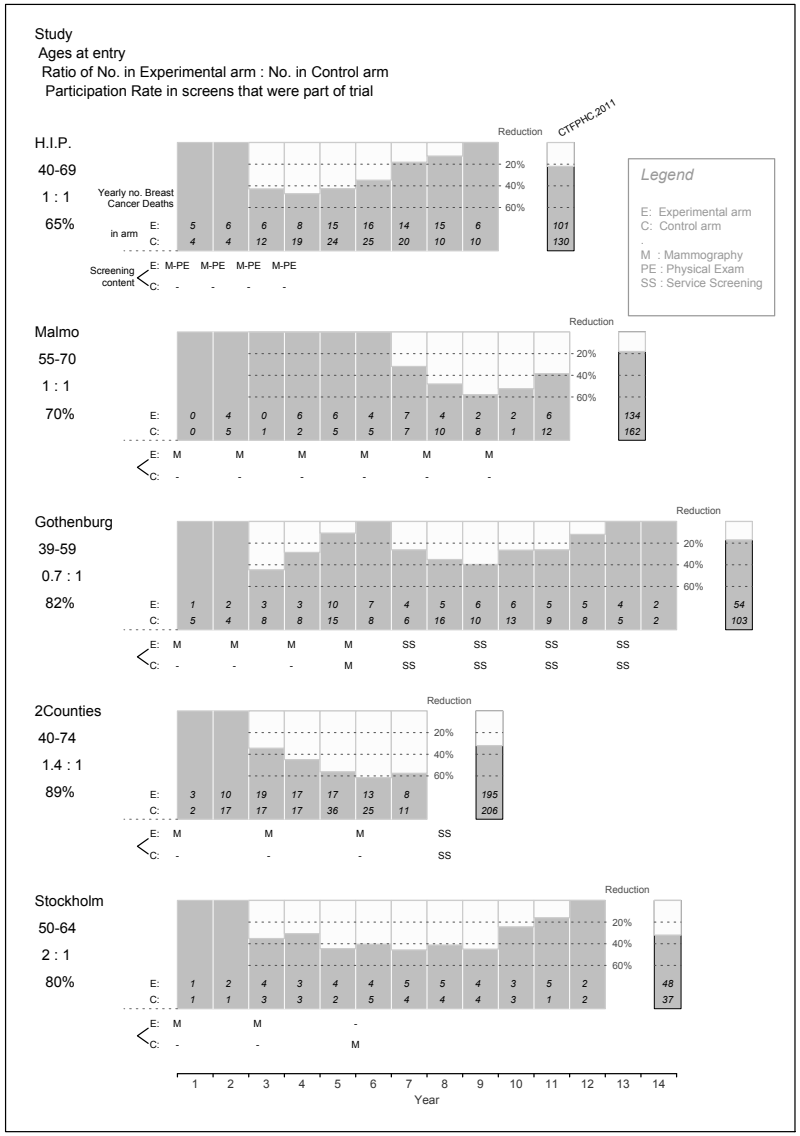
* The maximal mortality deficit, in year 7, would be achieved again in years 9, 11, ... if there were more than 3 such rounds.

Mortality deficits produced by once-a-year screening for 20 years, beginning when women reach 50, and ending when they reach 69

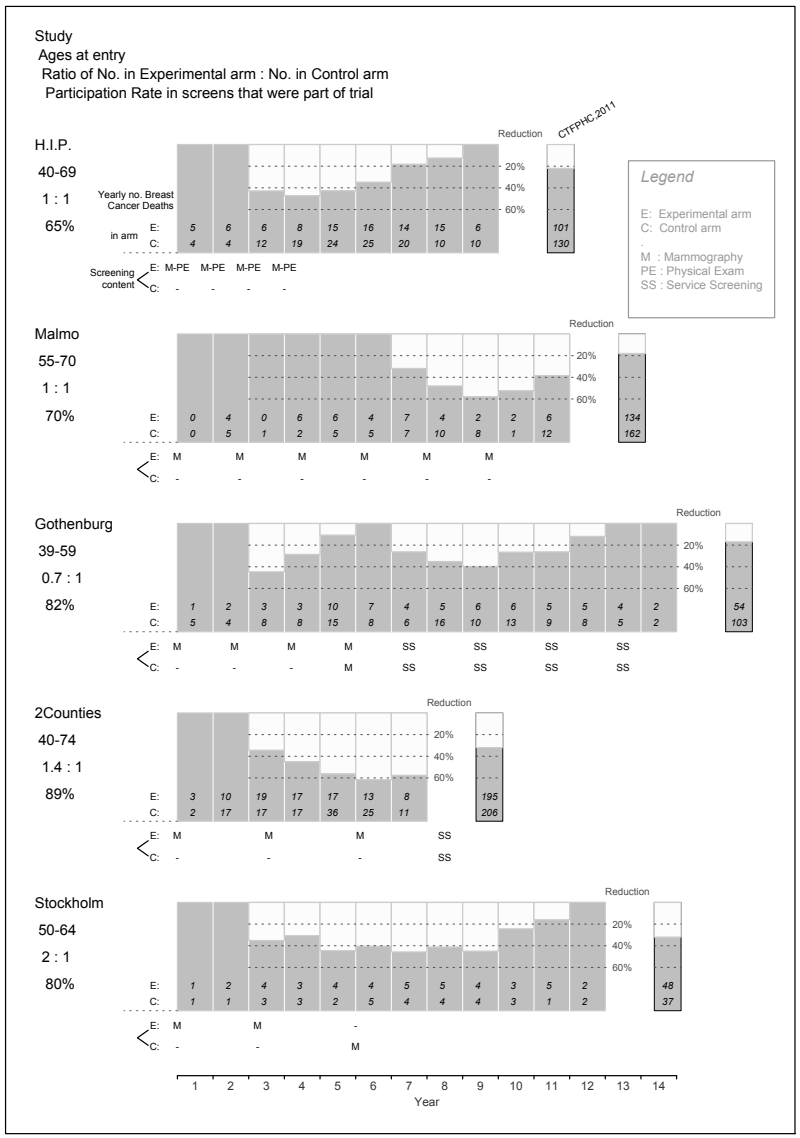


Observed breast cancer mortality deficits in 5 Mammography Trials

Observed breast cancer mortality deficits in 5 Mammography Trials

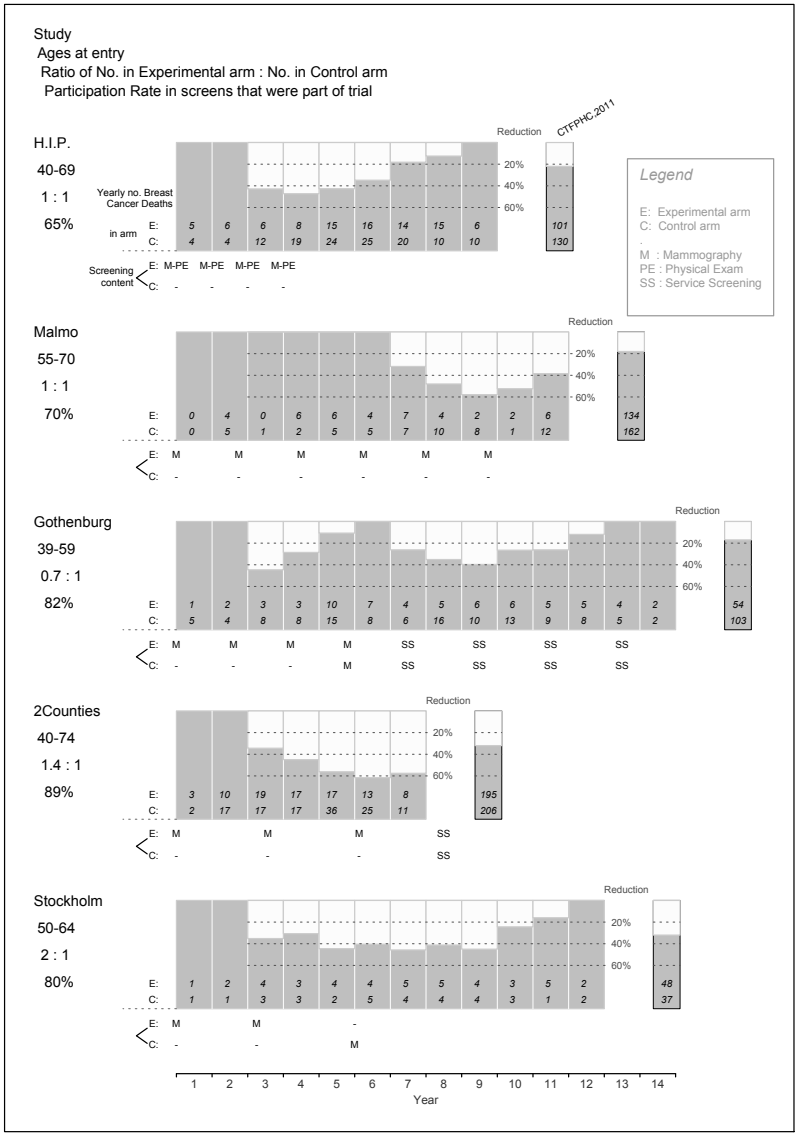


Observed breast cancer mortality deficits in 5 Mammography Trials



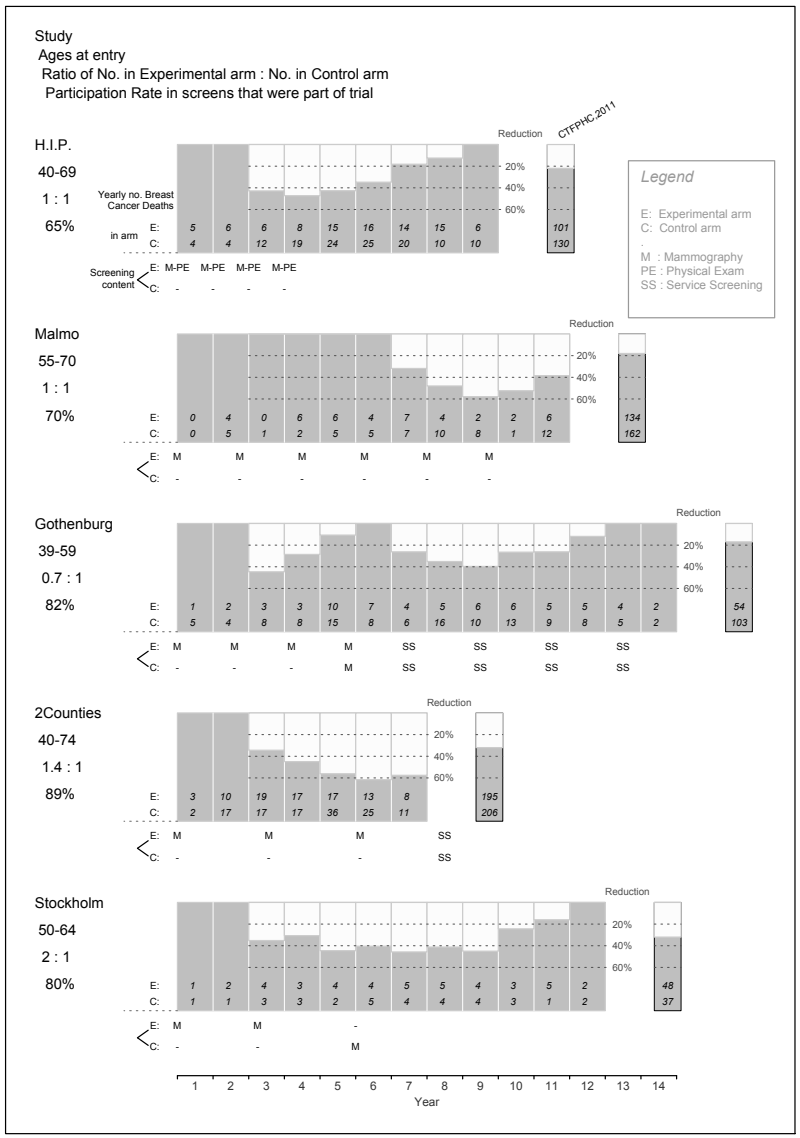
- Canadian Task Force guidelines are based on data-analyses that ignore some essential principles of cancer screening. The analyses underestimate the reductions in breast cancer mortality that would be seen in the 50-80 age range if women were screened regularly from when they reach age 50 until 69.

Observed breast cancer mortality deficits in 5 Mammography Trials



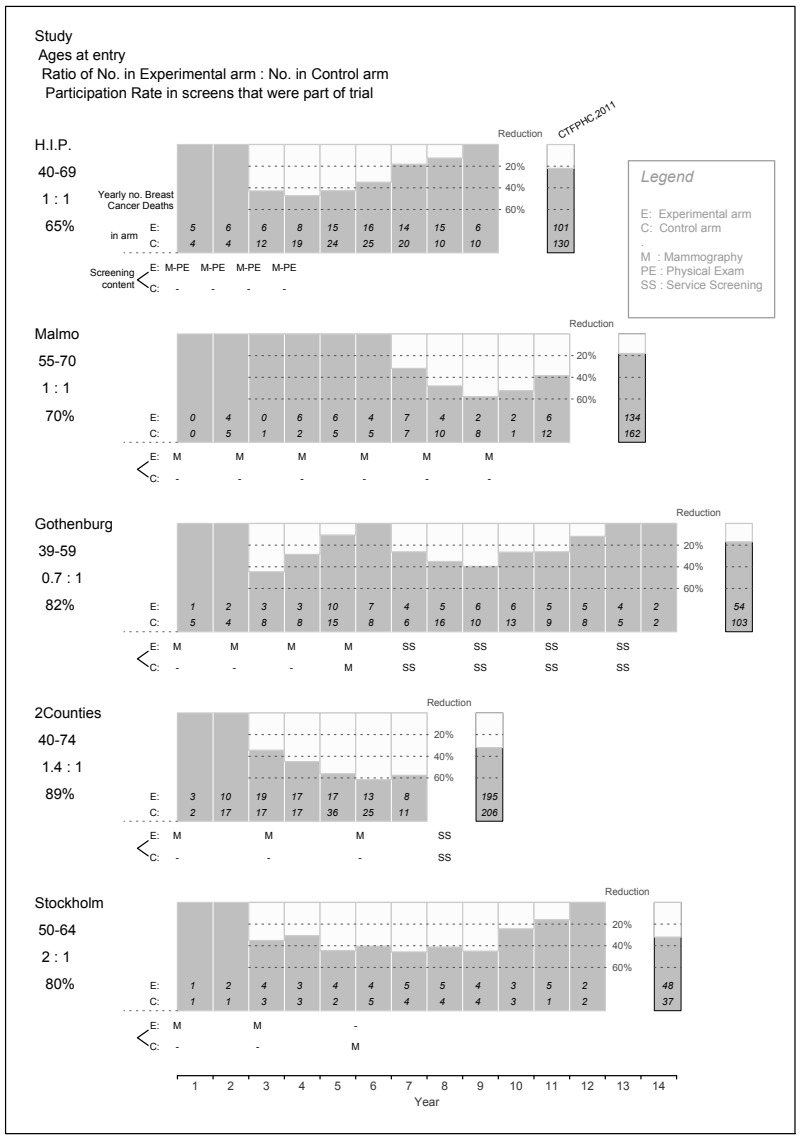
- Canadian Task Force guidelines are based on data-analyses that ignore some essential principles of cancer screening. The analyses underestimate the reductions in breast cancer mortality that would be seen in the 50-80 age range if women were screened regularly from when they reach age 50 until 69.
- We use year-specific data from the trials used by the Task Force, to quantify the magnitude and timing of the mortality reductions in relation to the no. & timing of the rounds of screening. We use the nadirs of the rate ratio curves as conservative estimates of what the reduction would be in a sustained program.

Observed breast cancer mortality deficits in 5 Mammography Trials



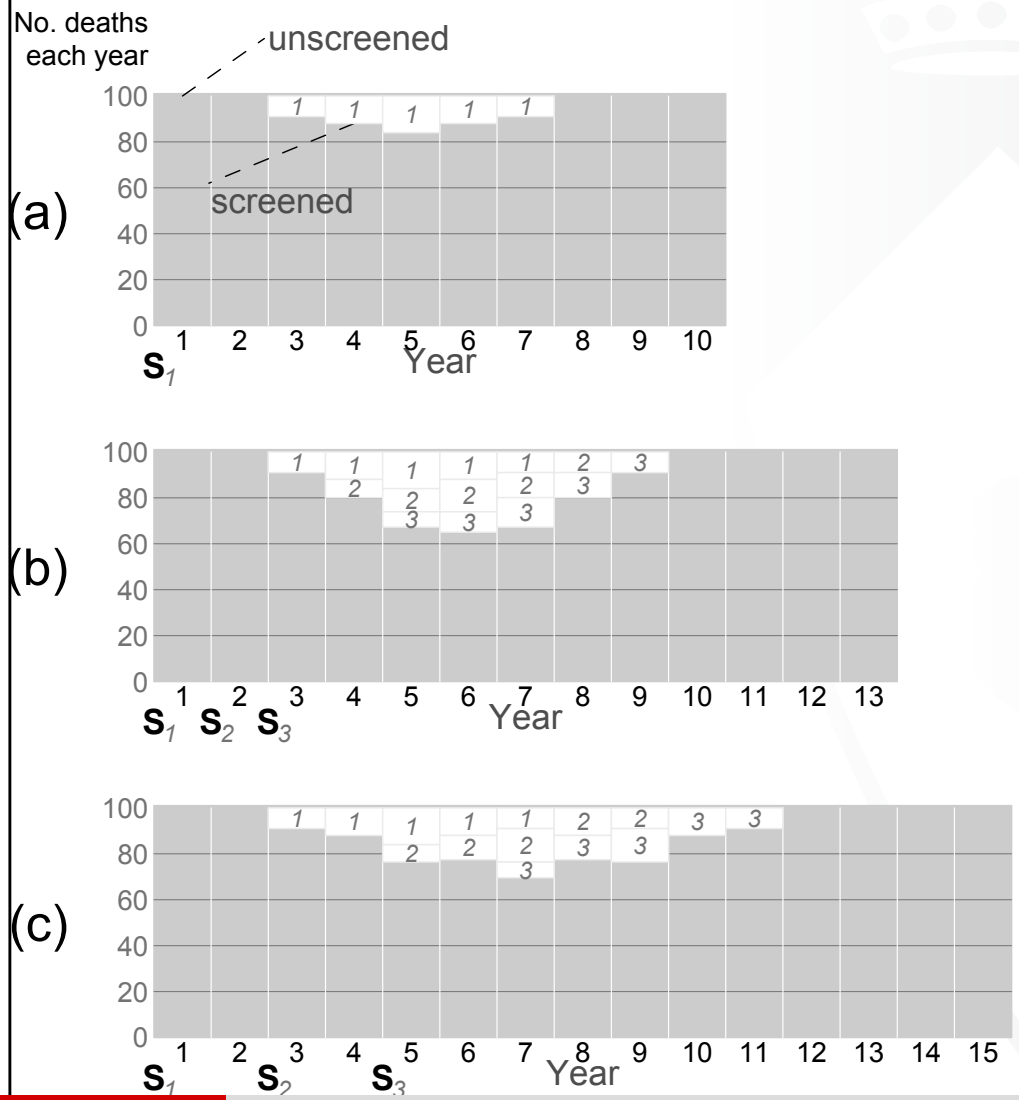
- Canadian Task Force guidelines are based on data-analyses that ignore some essential principles of cancer screening. The analyses underestimate the reductions in breast cancer mortality that would be seen in the 50-80 age range if women were screened regularly from when they reach age 50 until 69.
- We use year-specific data from the trials used by the Task Force, to quantify the magnitude and timing of the mortality reductions in relation to the no. & timing of the rounds of screening. We use the nadirs of the rate ratio curves as conservative estimates of what the reduction would be in a sustained program.
- Based on the 5 studies with adequate participation, 20 years of screening, 50–69, would be followed by 20 years (55–74) in which the breast cancer mortality reduction in these years would be $\geq$ 40%, with smaller deficits in other years. Fewer than 200 women would need to participate in such a program in order to avert a breast cancer death in the age range 50-80.

Observed breast cancer mortality deficits in 5 Mammography Trials

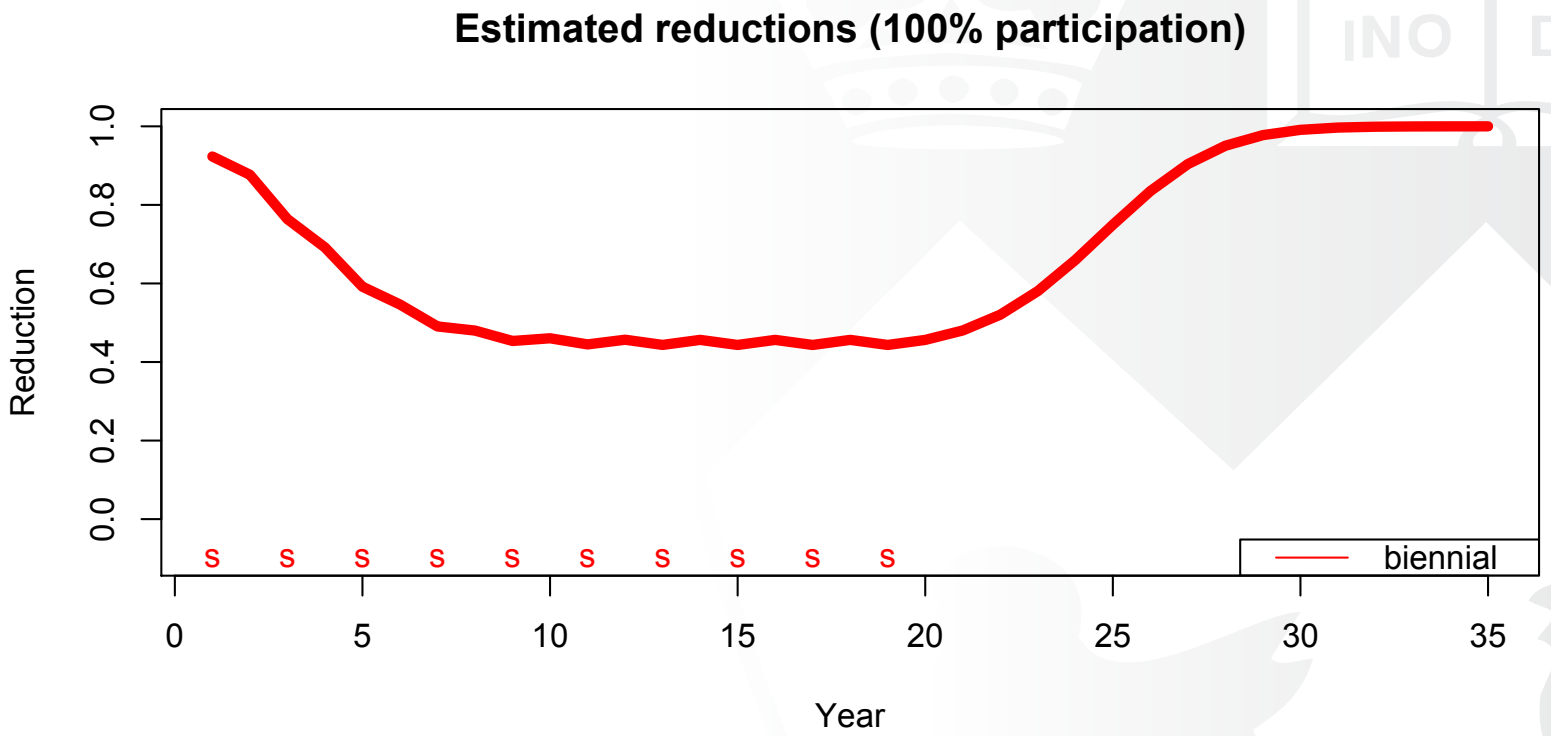


- Canadian Task Force guidelines are based on data-analyses that ignore some essential principles of cancer screening. The analyses underestimate the reductions in breast cancer mortality that would be seen in the 50-80 age range if women were screened regularly from when they reach age 50 until 69.
- We use year-specific data from the trials used by the Task Force, to quantify the magnitude and timing of the mortality reductions in relation to the no. & timing of the rounds of screening. We use the nadirs of the rate ratio curves as conservative estimates of what the reduction would be in a sustained program.
- Based on the 5 studies with adequate participation, 20 years of screening, 50–69, would be followed by 20 years (55–74) in which the breast cancer mortality reduction in these years would be $\geq$ 40%, with smaller deficits in other years. Fewer than 200 women would need to participate in such a program in order to avert a breast cancer death in the age range 50-80.
- The mortality reductions in these five studies are at least double the “average” figure of 21% used by the Task Force, while the number of women who, from age 50, would need to participate in a 20 year-screening program to avert one breast cancer death is a fraction of the 720 calculated by the Task Force.

A round-by-round approach?

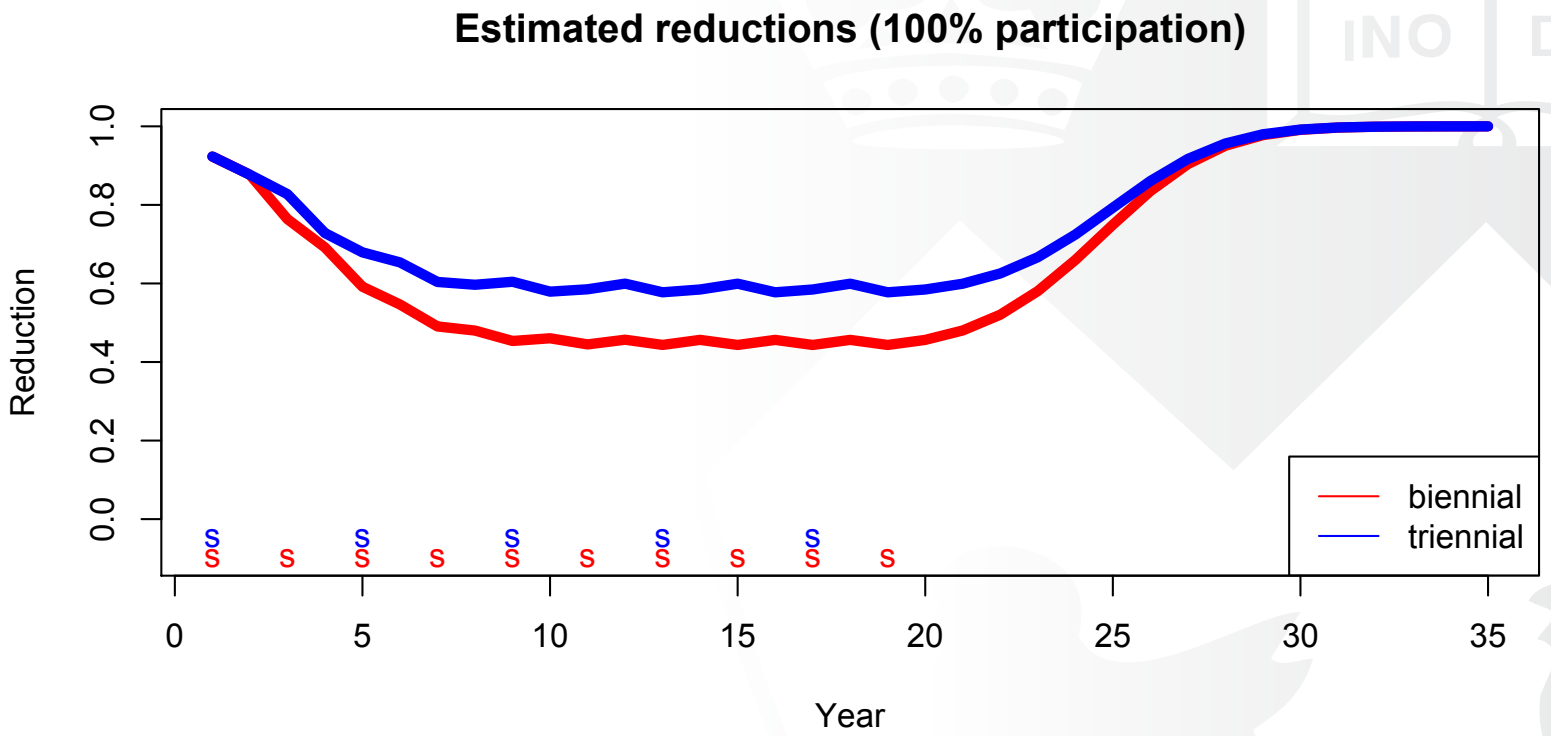


A round-by-round approach (SSC 2012)



Starting in year 7, a sustained yearly reduction of over **50%** in an **biennial** program,

A round-by-round approach (SSC 2012)



Starting in year 7, a sustained yearly reduction of over **50%** in an **biennial** program, or **40%** in an **triennial** program.

Summary: my 3 points again

Summary: my 3 points again

- With their blindness to the delay until the reductions in mortality are expressed,

Summary: my 3 points again

- With their blindness to the delay until the reductions in mortality are expressed, the prevailing design and data-analysis of cancer screening trials

Summary: my 3 points again

- With their blindness to the delay until the reductions in mortality are expressed, the prevailing design and data-analysis of cancer screening trials

Summary: my 3 points again

- 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***

Summary: my 3 points again

- 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

Summary: my 3 points again

- 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

Summary: my 3 points again

- 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

Summary: my 3 points again

- 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

Summary: my 3 points again

- 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

Summary: my 3 points again

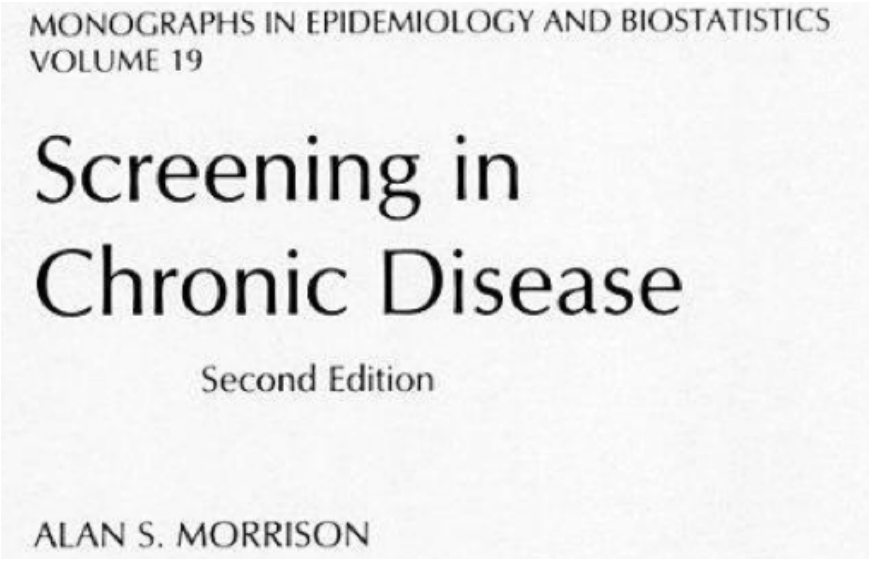
- With their blindness to the delay until the reductions in mortality are expressed, the prevailing design and data-analysis of cancer screening trials ***under-estimate*** the mortality reductions that ***would be produced by a sustained screening program***
- P-value-driven stopping rules exacerbate 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 / non-proportional hazards

Summary: my 3 points again

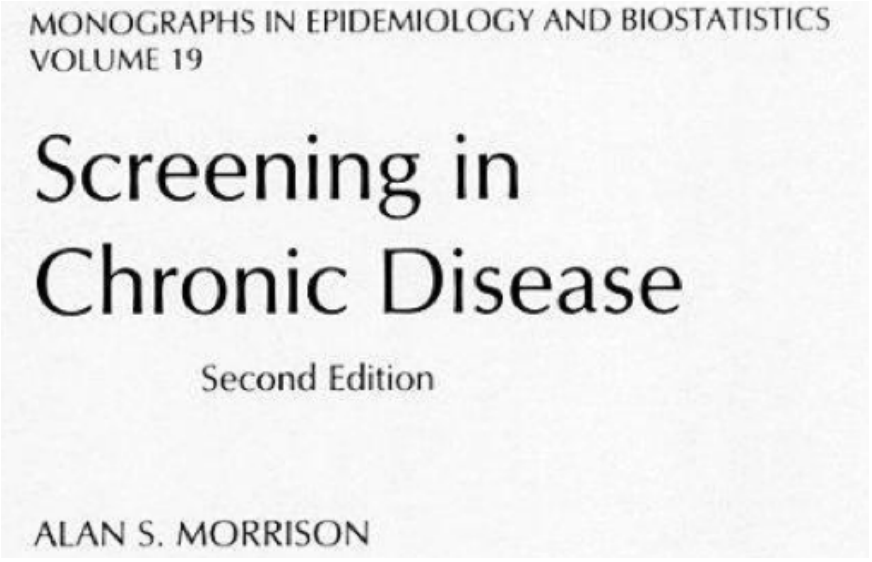
- With their blindness to the delay until the reductions in mortality are expressed, the prevailing design and data-analysis of cancer screening trials ***under-estimate*** the mortality reductions that ***would be produced by a sustained screening program***
- P-value-driven stopping rules exacerbate 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 / non-proportional hazards
  (v) focus on the parameters that describe impact of 1 round of screening

Acknowledgments

Acknowledgments



Acknowledgments

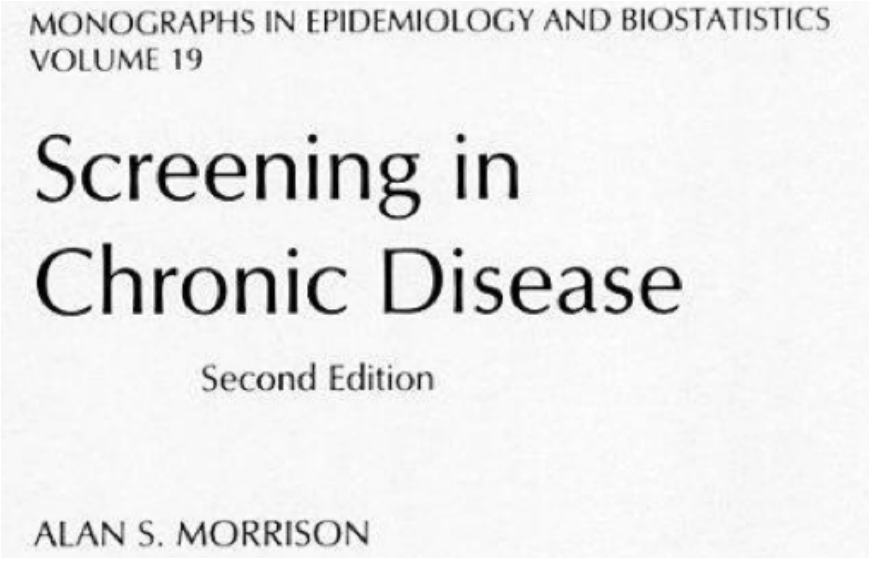


.....

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

Acknowledgments



Mammographic screening: no reliable supporting evidence?

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

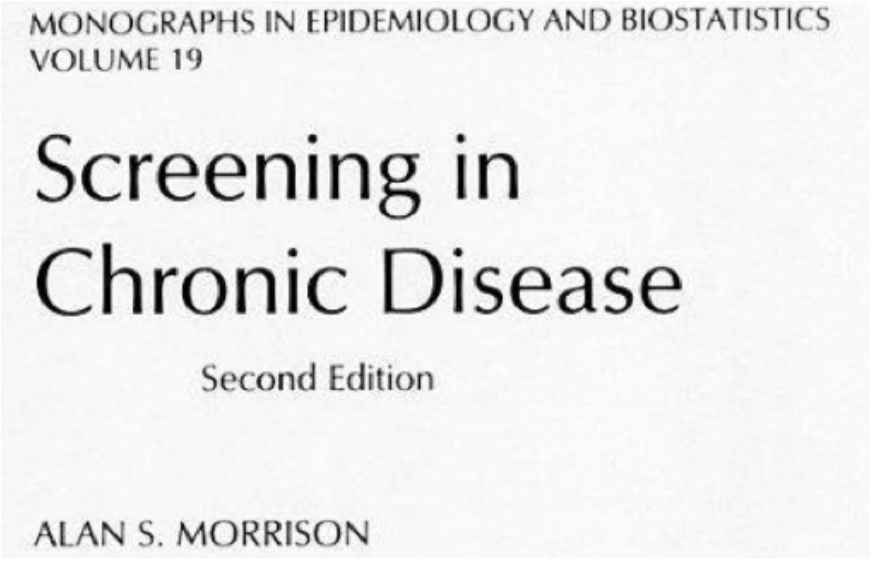
Much confusion is being generated by the conclusion of a recent review that "there is no reliable evidence that screening for breast cancer reduces mortality." In that review, however, there was no appreciation of the appropriate mortality-related measure of screening's usefulness; and correspondingly, 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.

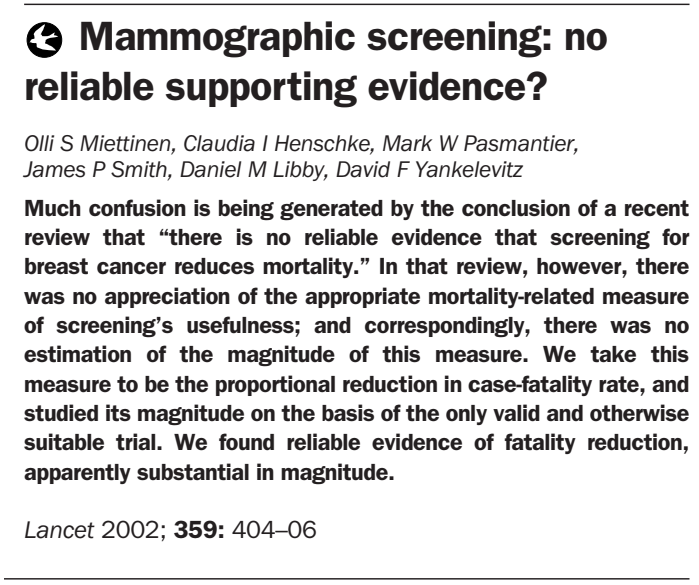
Lancet 2002; **359:** 404–06

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

Acknowledgments





Mammographic screening: no reliable supporting evidence?

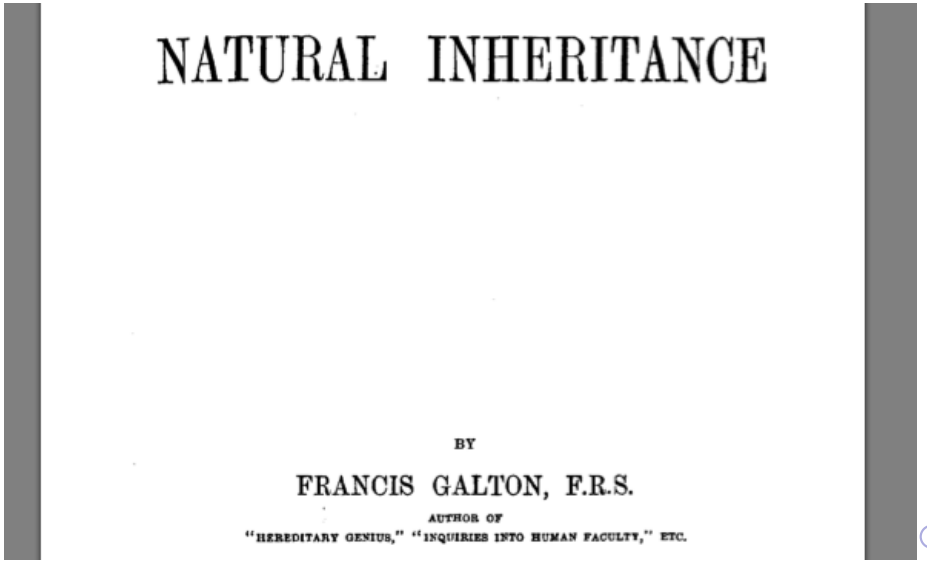
Olli S Miettinen, Claudia I Henschke, Mark W Pasmantier, James P Smith, Daniel M Libby, David F Yankelevitz

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



NATURAL INHERITANCE

BY

FRANCIS GALTON, F.R.S.

AUTHOR OF

"HEREDITARY GENIUS," "INQUIRIES INTO HUMAN FACULTY," ETC.

Why do statisticians commonly limit their inquiries to Averages?

Why do statisticians commonly limit their inquiries to Averages?

F. Galton, Natural Inheritance, 1889.

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.

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,

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.”

FUNDING, CO-ORDINATES, DOWNLOADS

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)

.....

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

→ **r e p r i n t s / talks**

Some References

Some References

1. Schröder FH, Hugosson J, Roobol MJ, et al. Screening and prostate-cancer mortality in a randomized European study. *N Engl J Med* 2009;360:1320-1328.
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.
3. *Hanley JA. Measuring Mortality reductions in cancer screening studies. *Epidemiologic Reviews* 2011. Advance Access published May 30, 2011.
4. *Hanley JA. CANNeCTIN Clinical Trials Methodology Seminar Series. Videoconference April 9, 2010. Slides: http://www.cannectin.ca/ . Video: Archived Events, http://webcast.otn.ca/
5. Thompson SG, Ashton HA, Gao L, Scott RAP on behalf of the Multicentre Aneurysm Screening Study Group. Screening men for abdominal aortic aneurysm: 10 year mortality and cost effectiveness results from the randomised Multicentre Aneurysm Screening Study. *BMJ* 2009;338:b2307 doi:10.1136/bmj.b2307.
6. *Hanley JA. Analysis of Mortality Data From Cancer Screening Studies: Looking in the Right Window. *Epidemiology* 2005; 16: 786-790.
7. Miettinen OS, Henschke CI, Pasmantier MW, et al. Mammographic screening: no reliable supporting evidence? *Lancet* 2002;359:404-406.
8. Miettinen OS, Henschke CI, Pasmantier MW, et al. Mammographic screening: no reliable supporting evidence? Available at: http://image.thelancet.com/extras/1093web.pdf. Accessed July 6, 2005.
9. The National Lung Screening Trial Research Team. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *N Engl J Med* 2011; 365:395-409.
10. Hanley JA, Liu Z, Strumpf E, Dendukuri N, McGregor M. Numbers of breast cancer deaths averted by mammography screening". under review at Canadian Journal of Public Health

* `http:www.biostat.mcgill.ca/hanley/` (reprints/talks)

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

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).

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).
- JNCI 2000: "Lung Cancer Mortality in the Mayo Lung Project: Impact of Extended Follow-up"

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).
- JNCI 2000: "Lung Cancer Mortality in the Mayo Lung Project: Impact of Extended Follow-up"

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

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).
- JNCI 2000: "Lung Cancer Mortality in the Mayo Lung Project: Impact of Extended Follow-up"

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

- ALL lung cancer deaths, from those in year...

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).
- JNCI 2000: "Lung Cancer Mortality in the Mayo Lung Project: Impact of Extended Follow-up"

  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,

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).
- JNCI 2000: "Lung Cancer Mortality in the Mayo Lung Project: Impact of Extended Follow-up"

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

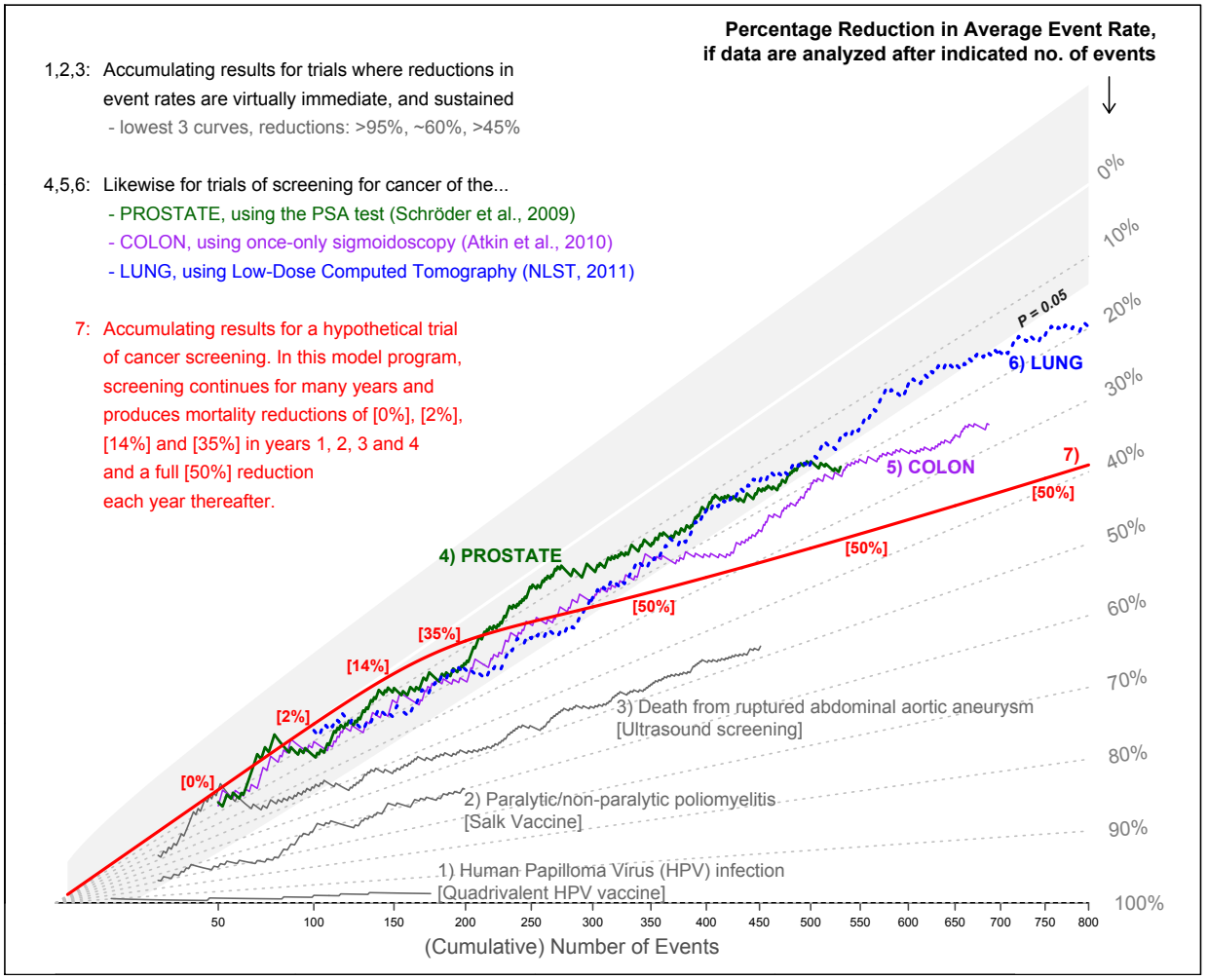
- ALL lung cancer deaths, from those in year...
  - 1, before impact could become evident,
    to

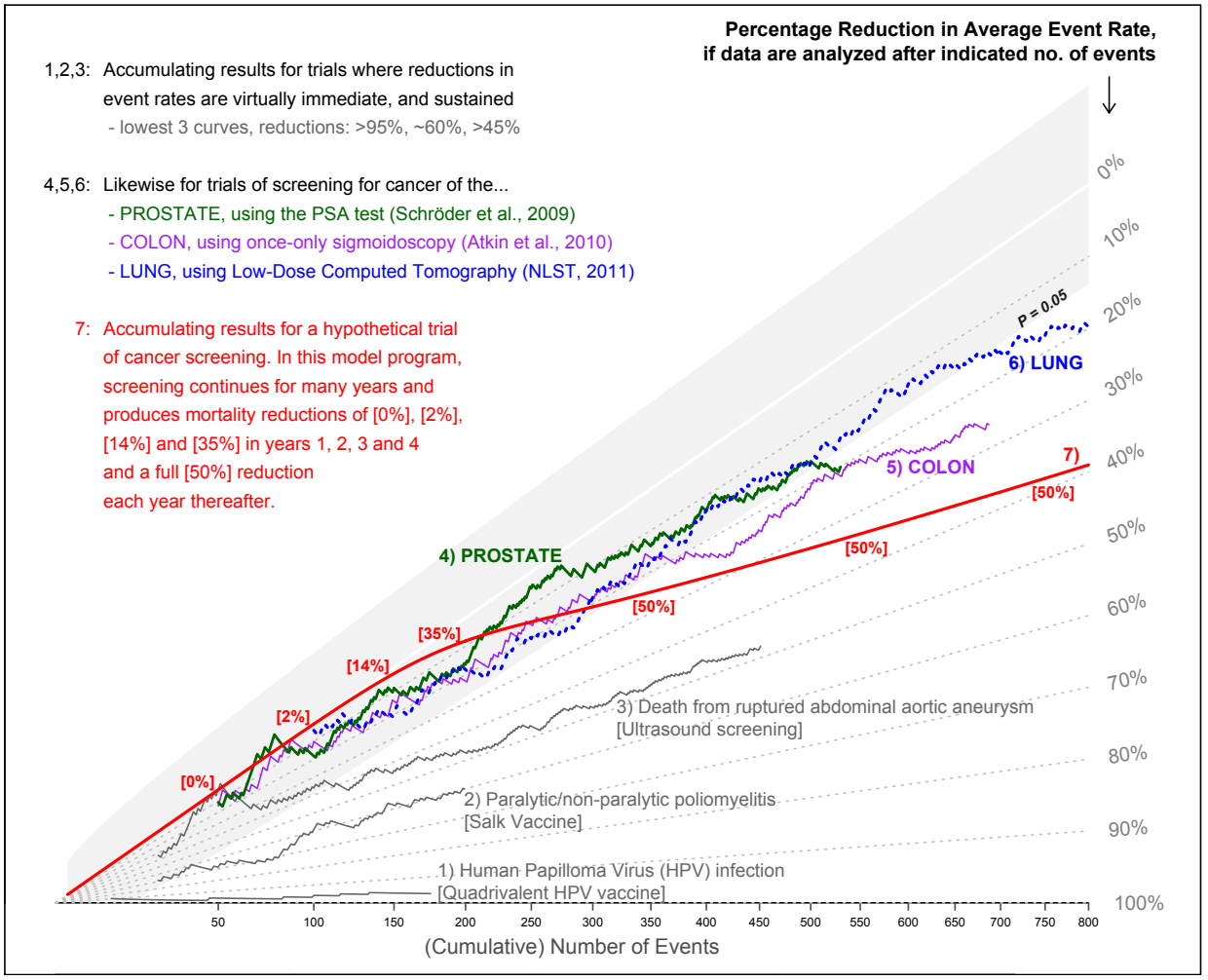
Mayo Lung Project (chest x-ray & sputum cytology)

- 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).
- JNCI 2000: "Lung Cancer Mortality in the Mayo Lung Project: Impact of Extended Follow-up"

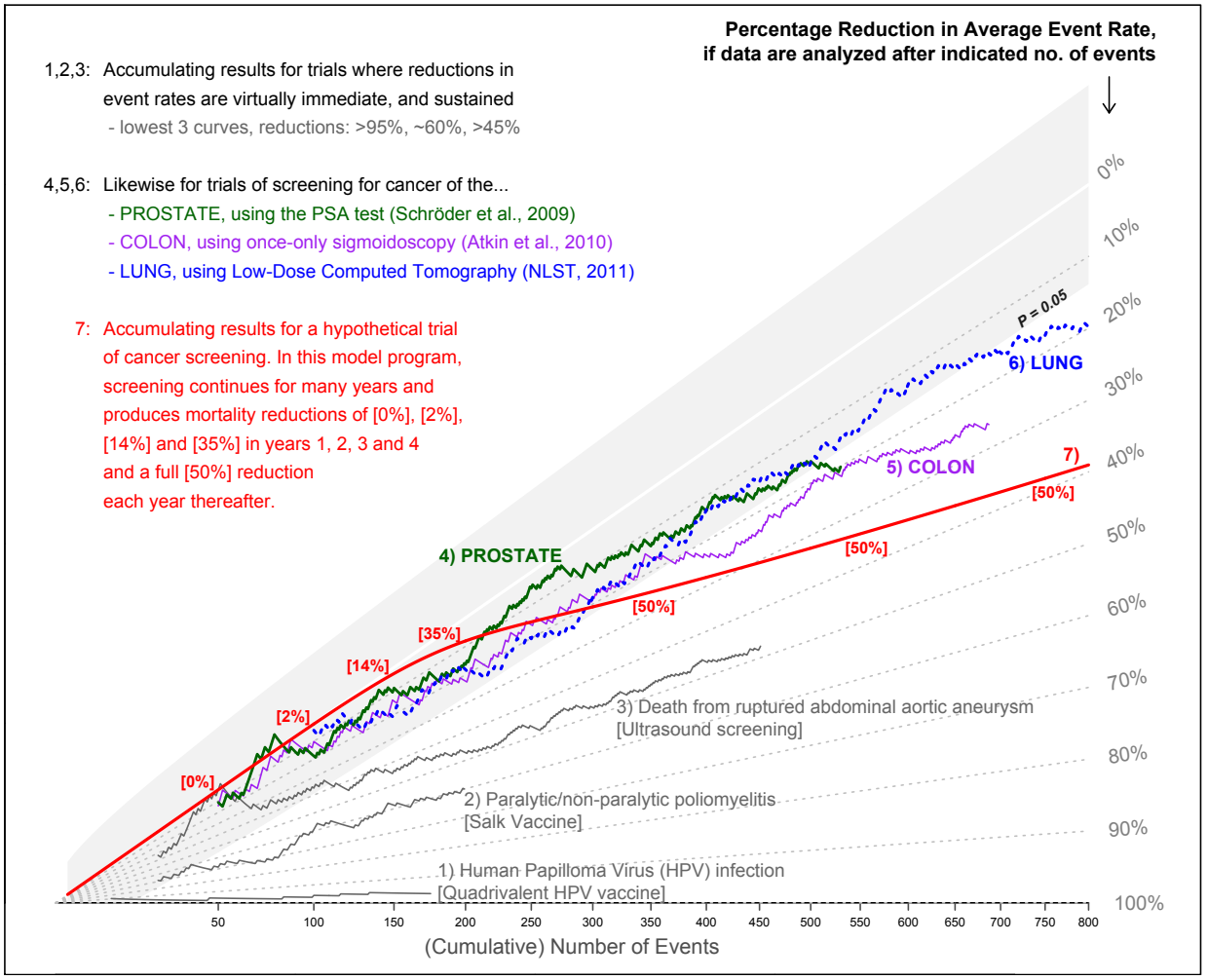
  Would 24-year follow up "allow for a reduction in lung cancer mortality to be observed?"

- ALL lung cancer deaths, from those in year...
  - 1, before impact could become evident,
    to
  - 24, 18 years after last screen.


Percentage Reduction in Average Event Rate,
if data are analyzed after indicated no. of events
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.
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
P = 0.05
6) LUNG
5) COLON
7)
[50%]
[50%]
[50%]
4) PROSTATE
[35%]
[14%]
[2%]
[0%]
3) Death from ruptured abdominal aortic aneurysm
[Ultrasound screening]
2) Paralytic/non-paralytic poliomyelitis
[Salk Vaccine]
1) Human Papilloma Virus (HPV) infection
[Quadrivalent HPV vaccine]
50
100
150
200
250
300
350
400
450
500
550
600
650
700
750
800
(Cumulative) Number of Events

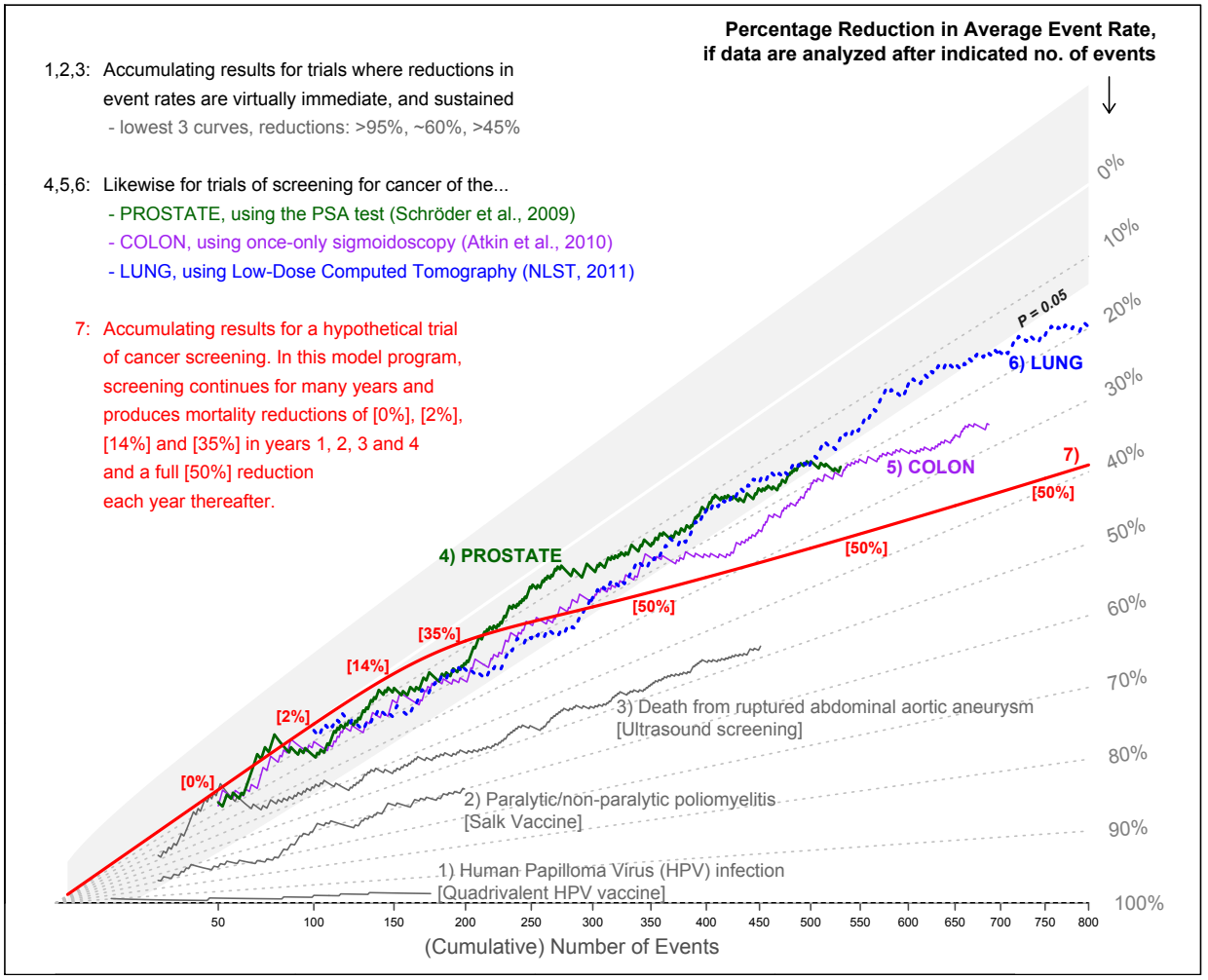


If **intervention continues over time to deflect the same % of events**, an estimate of the % reduction, based on the total number events in more (person)-time will be more precise



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

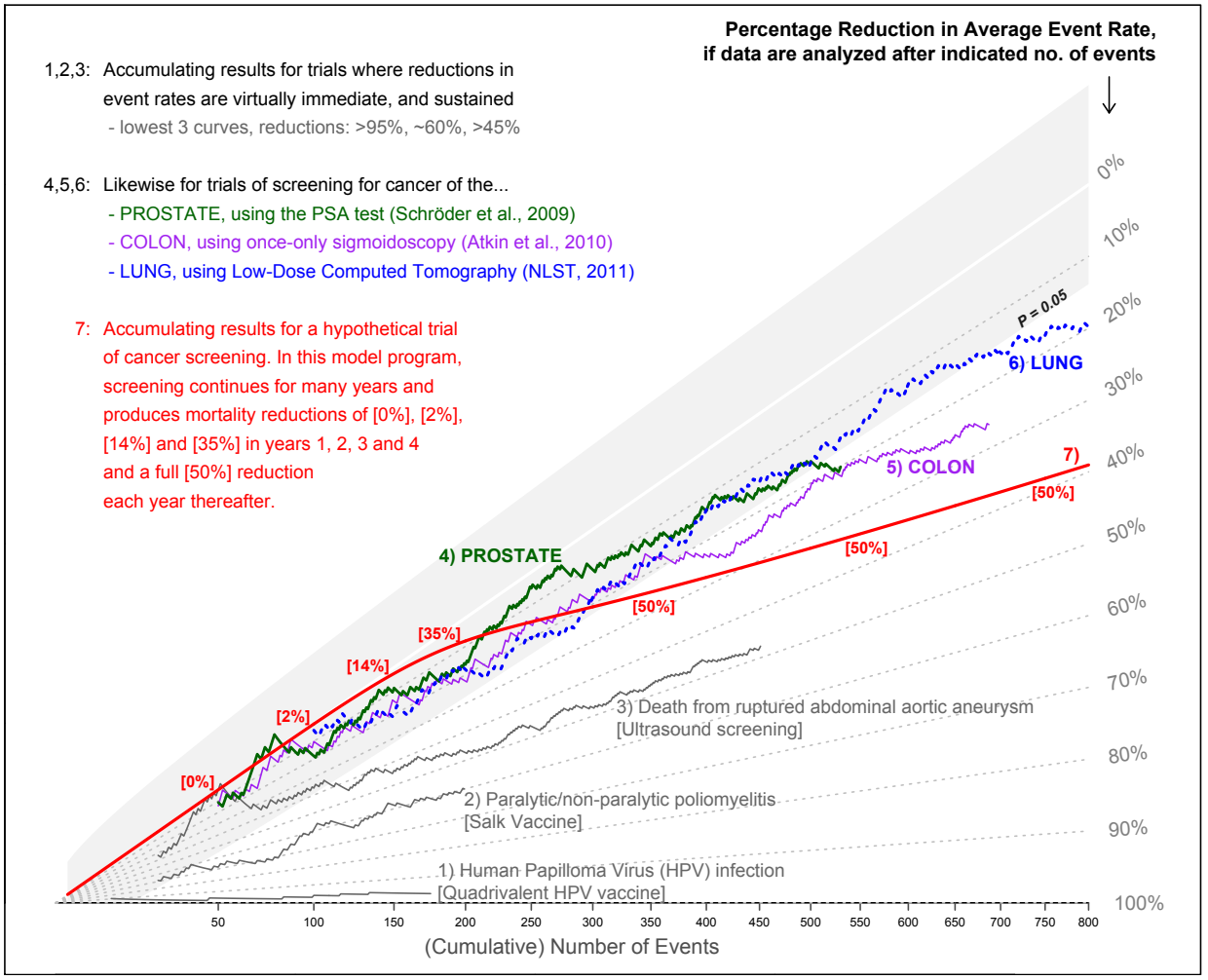
Mortality reductions from cancer screening manifest distally. Enrolling and following more people for short length of time yields a more precise UNDERestimate.



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 **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

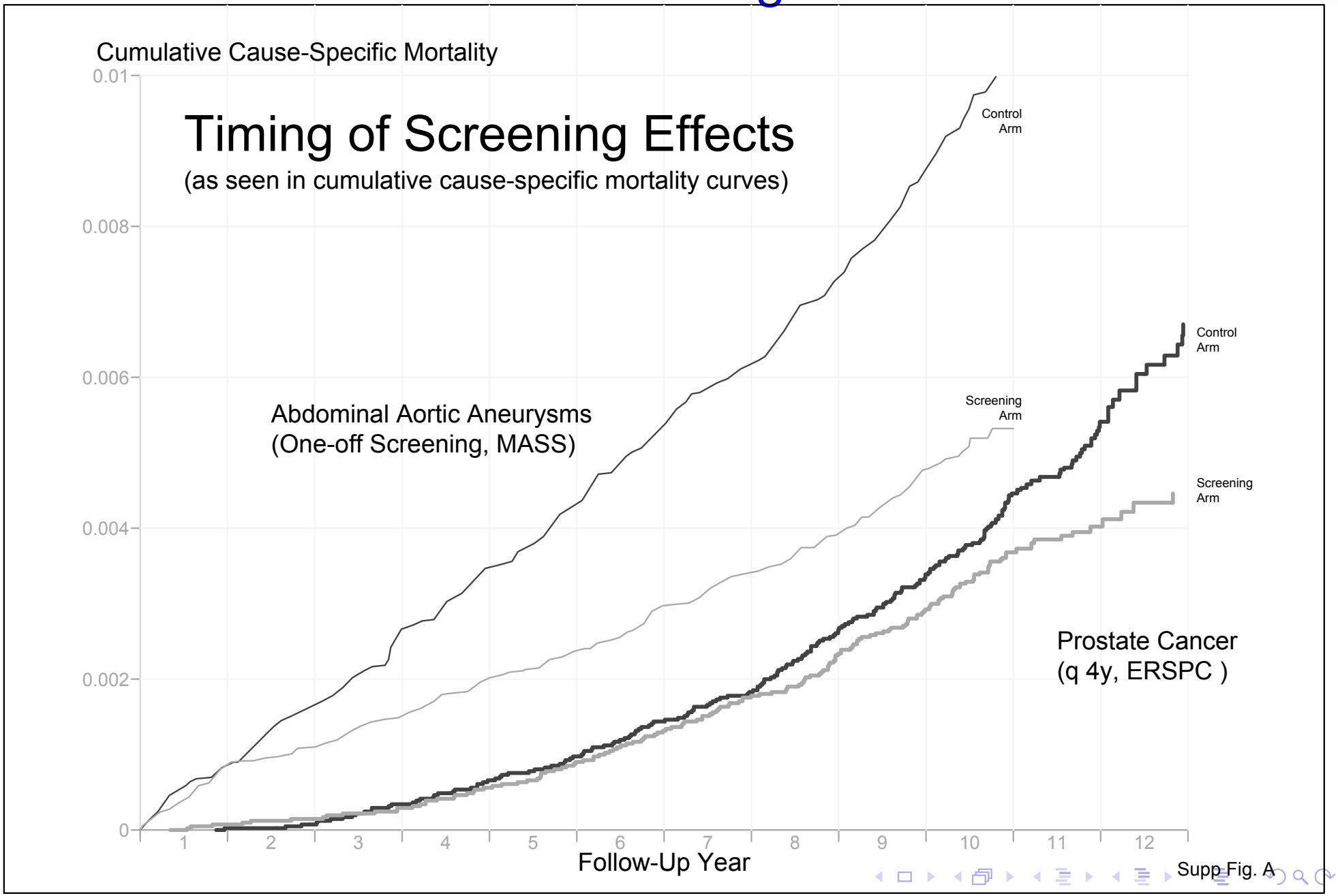
Mortality reductions from cancer screening manifest distally. Enrolling and following more people for short length of time yields a more precise UNDERestimate.

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.

The loneliness of the long-distance trialist

The loneliness of the long-distance trialist

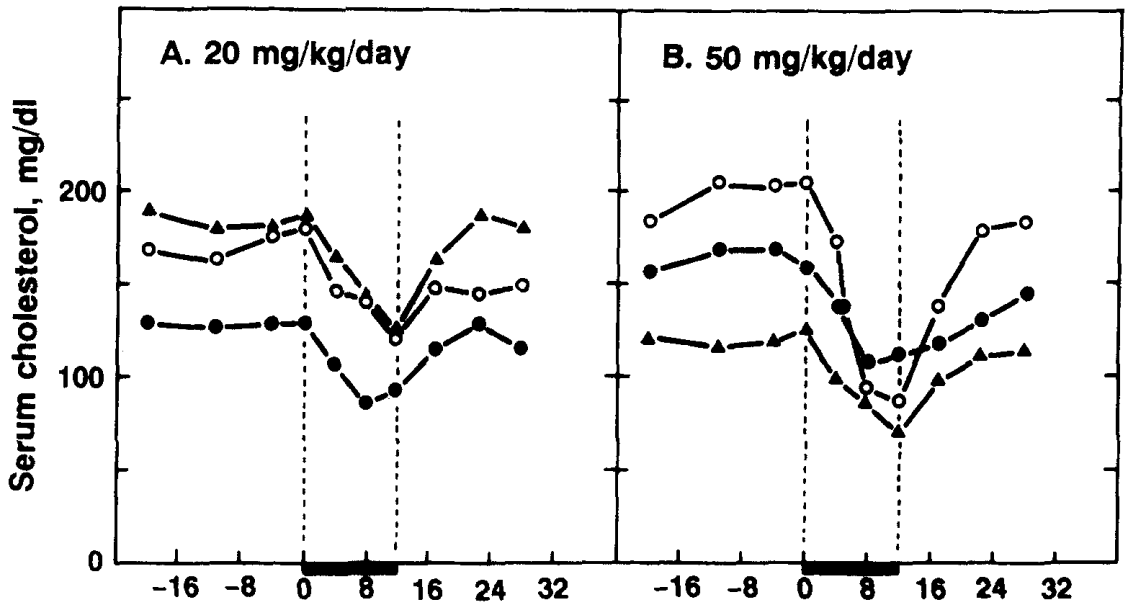


Supp Fig. A

Timing of cholesterol reductions produced by statins

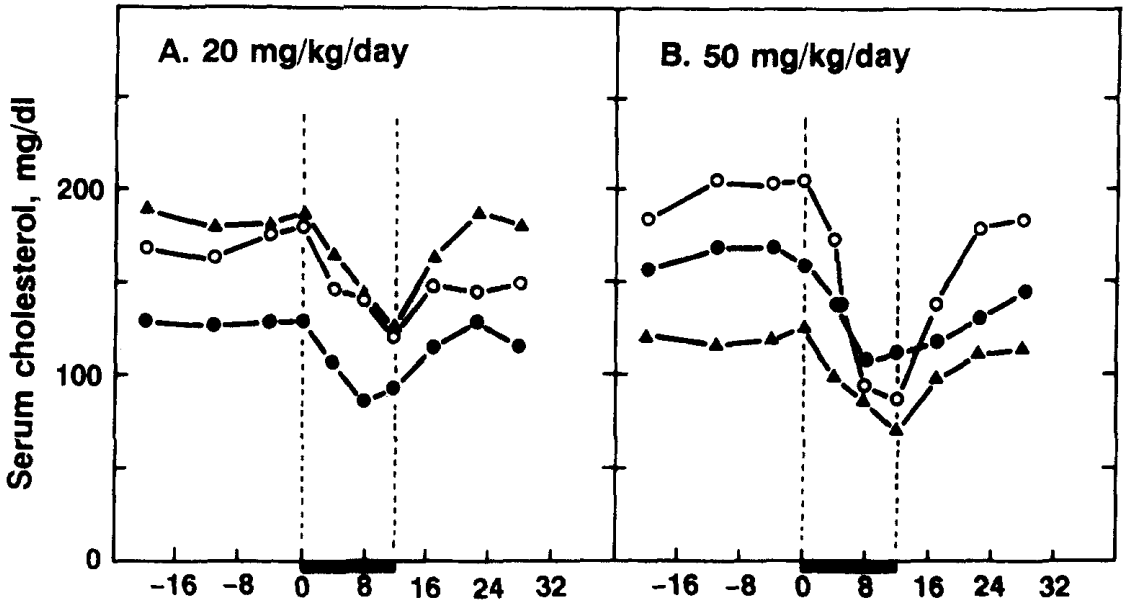
Timing of cholesterol reductions produced by statins

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

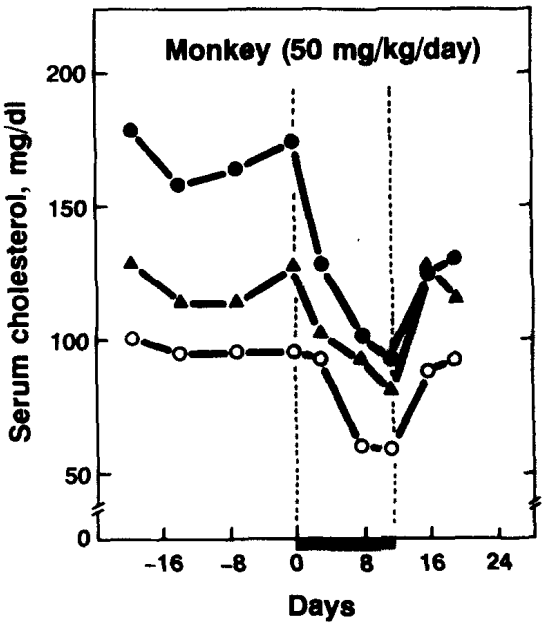


Timing of cholesterol reductions produced by statins

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



3 monkeys at 50

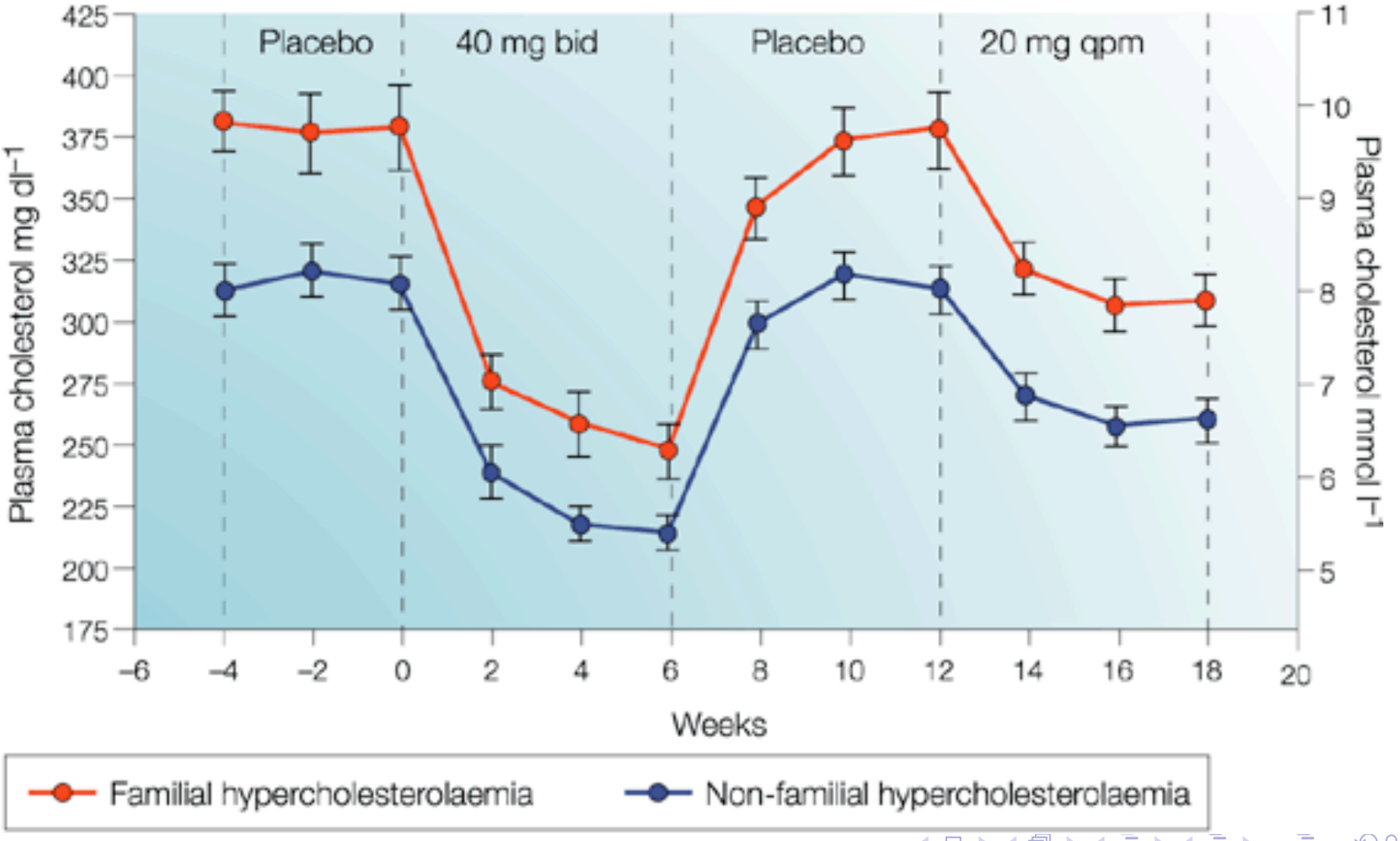


Timing of cholesterol reductions produced by statins

Humans

Timing of cholesterol reductions produced by statins

Humans



Cumulative vs. Year-specific Mortality...

Cumulative vs. Year-specific Mortality...

in 100,000 men

(average age at entry: 62 years)

Cumulative vs. Year-specific Mortality...

in 100,000 men

(average age at entry: 62 years)

if screened using PSA test

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,

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

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

Cumulative vs. Year-specific Mortality...

in 100,000 men

(average age at entry: 62 years)

if screened using PSA test

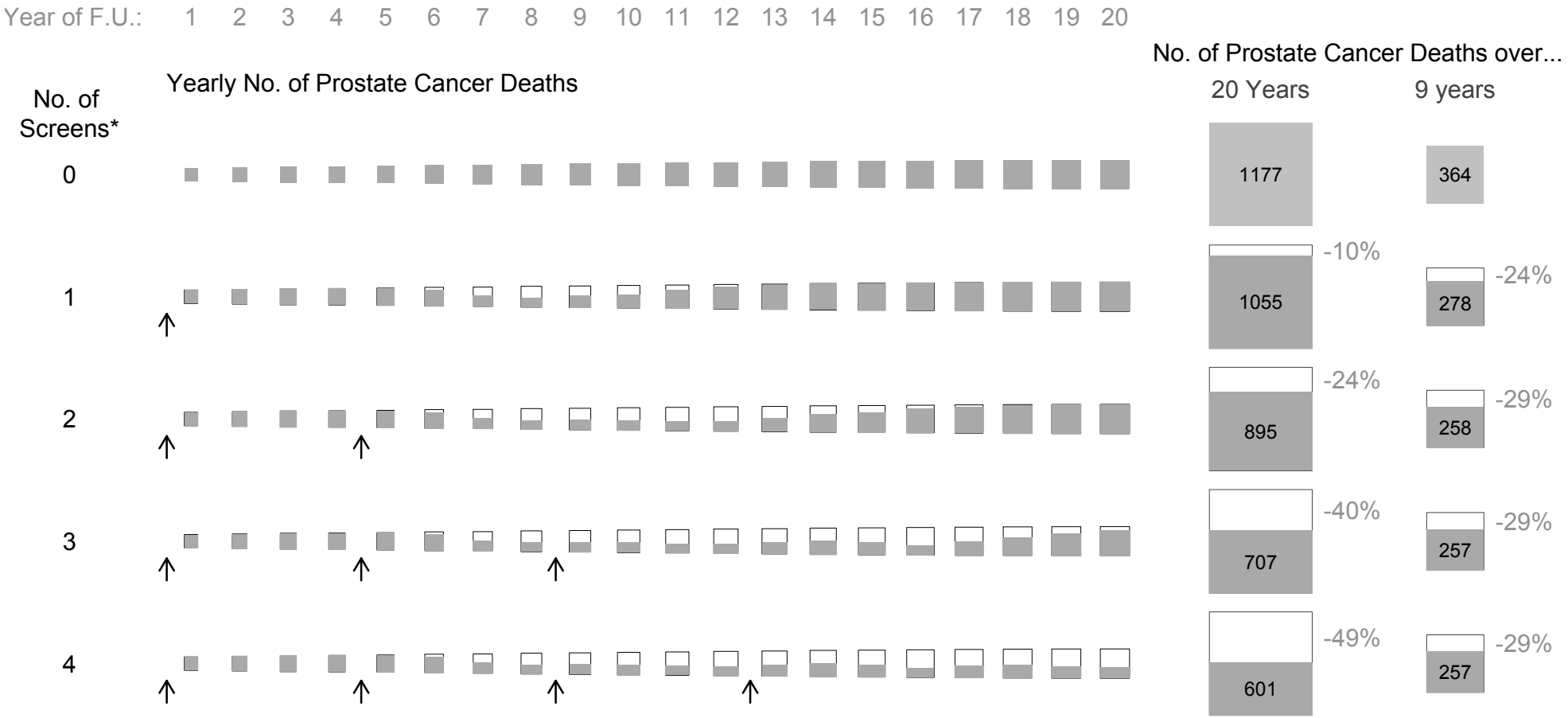
0, 1, 2, 3, or 4 times,

tests 4 years apart

and followed for (9) 20 years

HYPOTHETICAL DATA

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



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

(B) Year-specific Rate Ratios & Percent Reductions [HYPOTHETICAL]

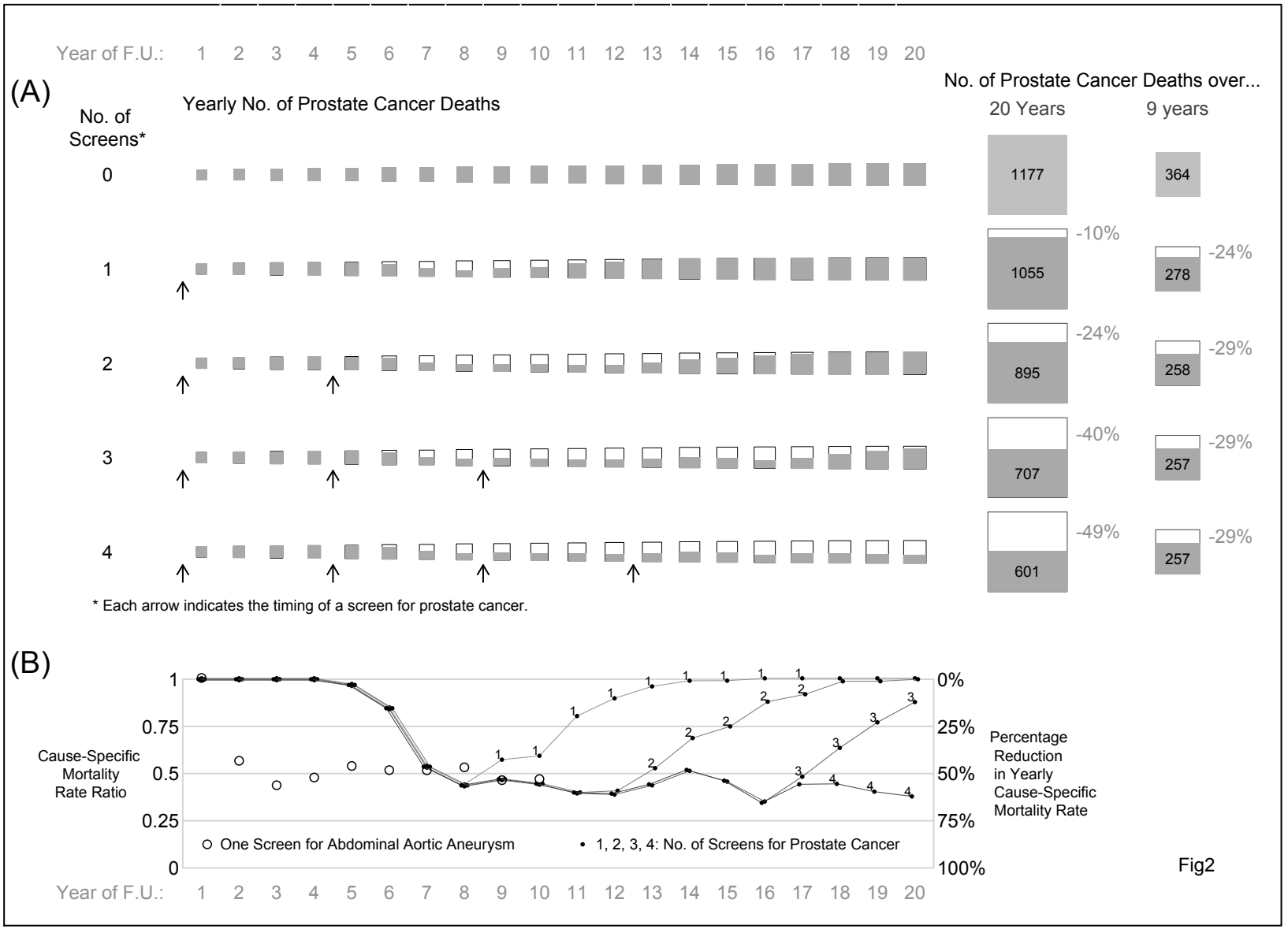


Fig2

Norway - 'before-after' study

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 SEPTEMBER 23, 2010 VOL. 363 NO. 13

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

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.

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.

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.

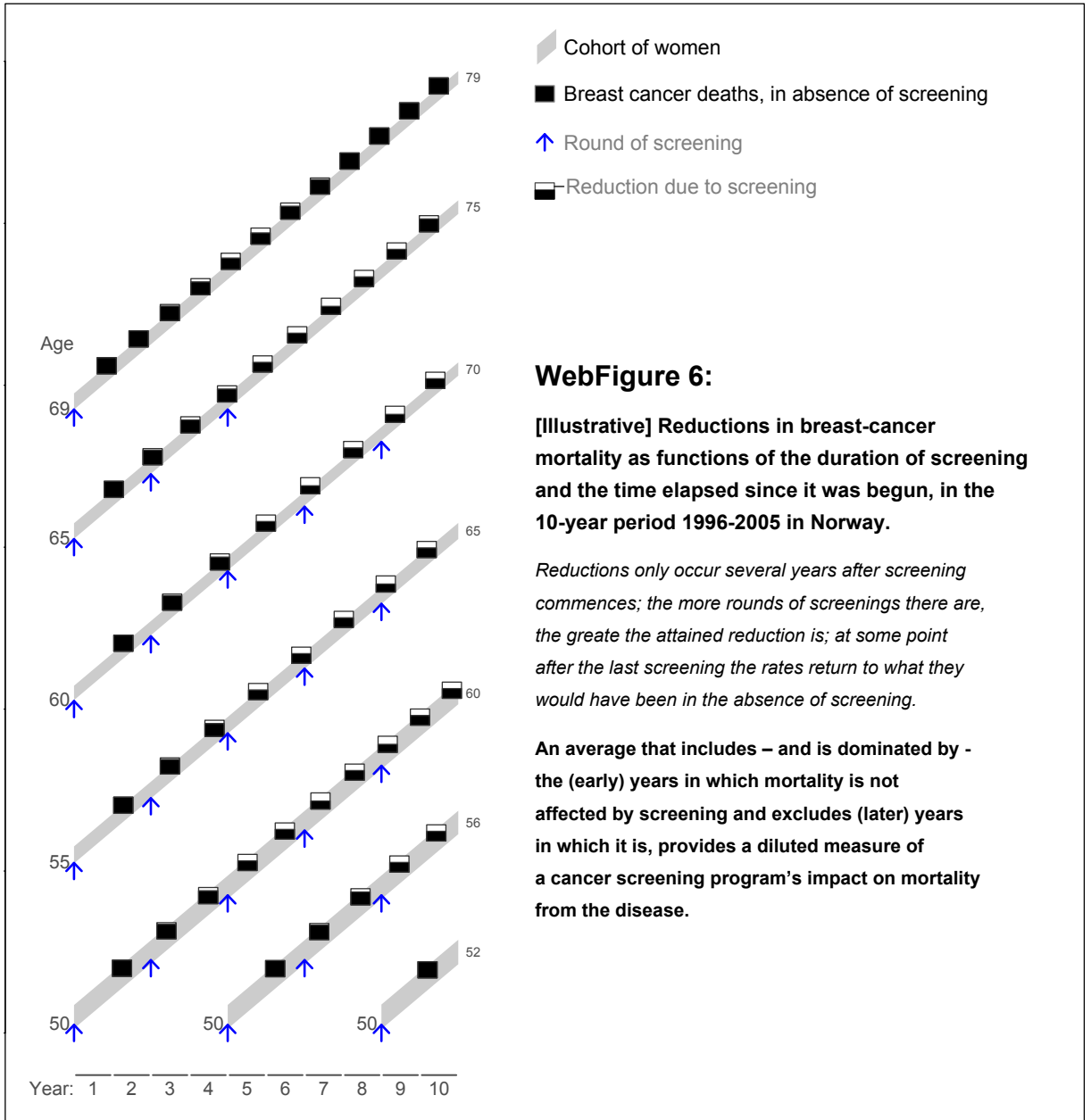
Epidemiologic Reviews, 2011

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 greate 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

emphasis on time-specificity

- Year-specific* mortality rate ratios

emphasis on time-specificity

- Year-specific* mortality rate ratios
- Moving averages* to reduce the statistical noise (deaths in moving 3-year intervals)

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)

National Lung Screening Trial (NLST)

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

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:

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

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:

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]

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,

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)]

ACR Imaging Network: Press Release

ACR Imaging Network: Press Release

Table 3: Interim Analysis of Primary Endpoint Reported on October 20, 2010

Trial Arm	Person years (py)	Lung cancer deaths	Lung cancer mortality per 100,000 py	Reduction in lung cancer mortality (%)	Value of test statistic	Efficacy boundary
LDCT	144,097.6	354	245.7	20.3	–3.21	–2.02
CXR	143,363.5	442	308.3			

ACR Imaging Network: Press Release

Table 3: Interim Analysis of Primary Endpoint Reported on October 20, 2010

Trial Arm	Person years (py)	Lung cancer deaths	Lung cancer mortality per 100,000 py	Reduction in lung cancer mortality (%)	Value of test statistic	Efficacy boundary
LDCT	144,097.6	354	245.7	20.3	–3.21	–2.02
CXR	143,363.5	442	308.3			

"Deficit": 88