Il metodo scientifico: trial e nulla più?

Fabrizio Faggiano
Università del Piemonte Orientale
Associazione Italiana di Epidemiologia

Premessa

- Lo screening mammografico del tumore della mammella è attualmente oggetto di un conflitto fra gruppi di ricercatori
  - di elevato livello scientifico
  - dai toni spesso aspri
  - fondato sull'incertezza delle evidenze disponibili
    - sui benefici
    - e sui rischi

Il metodo scientifico

The **scientific method**

"a method or procedure that has characterized natural science since the 17th century, ***consisting in systematic observation, measurement, and experiment, and the formulation, testing, and modification of hypotheses***."

Oxford English Dictionary

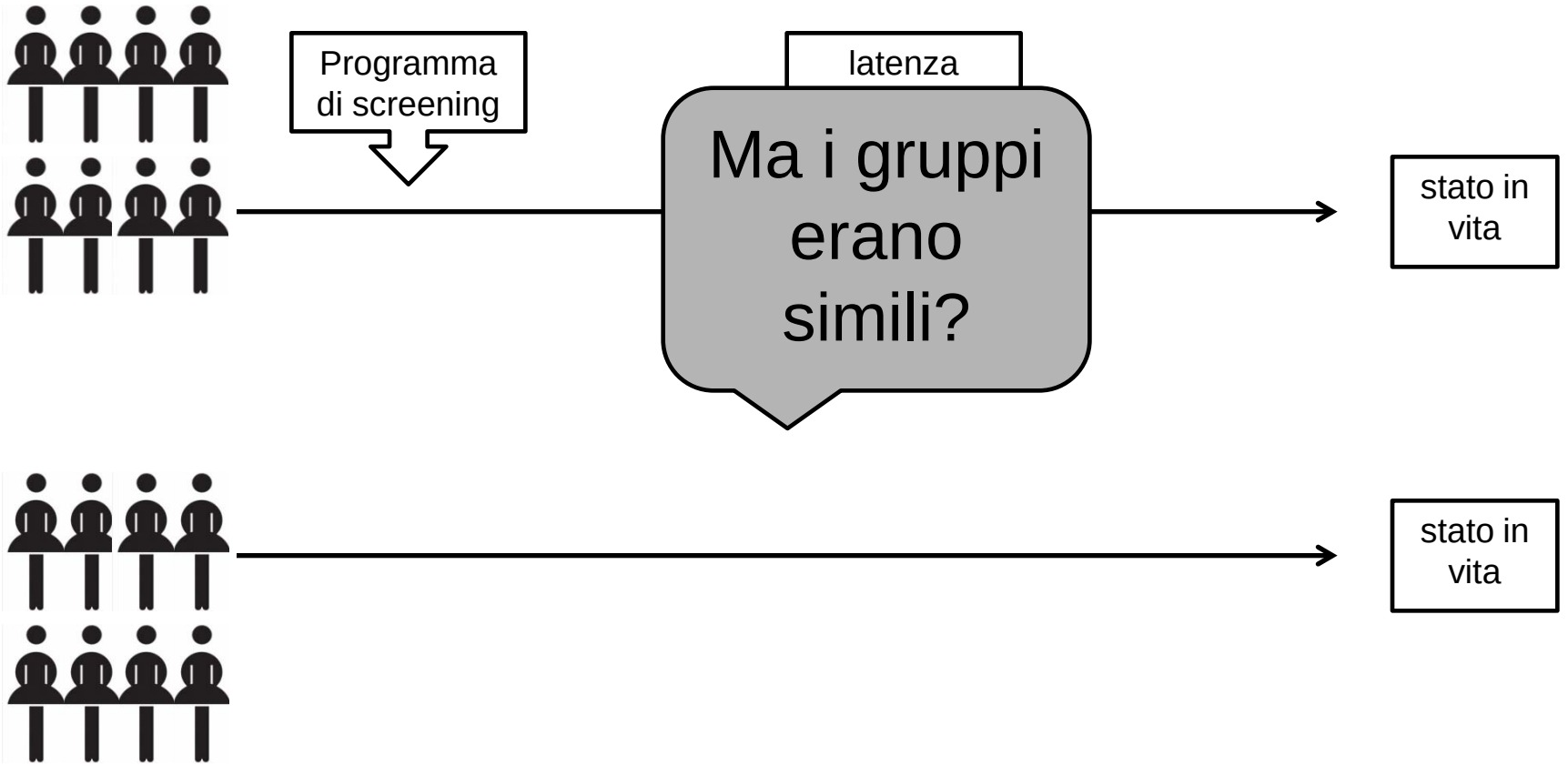
Osservazione

Ipotesi: la mammografia ripetuta ogni X anni dai 50 ai 64 riduce la mortalità per tumore della mammella

- Come verificarla?
- Con l'osservazione!

Effetto dello screening sulla mortalità

Effetto dello screening sulla mortalità



Confondimento: il caso TOS

Postmenopausal Estrogen and Progestin Use and the Risk of Cardiovascular Disease

Grodstein Ann Int Medicine 2000

- 16 year results from the Nurses health study
- RR of CHD estrogen+progestin = 0.39 (0.19-0.78)
- RR of CHD estrogen alone = 0.60 (0.43-0.83).

Confondimento: il caso TOS

Risks and Benefits of Estrogen Plus Progestin in Healthy Postmenopausal Women

Principal Results From the Women's Health Initiative Randomized Controlled Trial

Writing Group for the Women's Health Initiative Investigators

Context Despite decades of accumulated observational evidence, the balance of risks and benefits for hormone use in healthy postmenopausal women remains uncertain.

Objective To assess the maior health benefits and risks of the most commonly used

- 16608 postmenopausal women aged 50-79 years

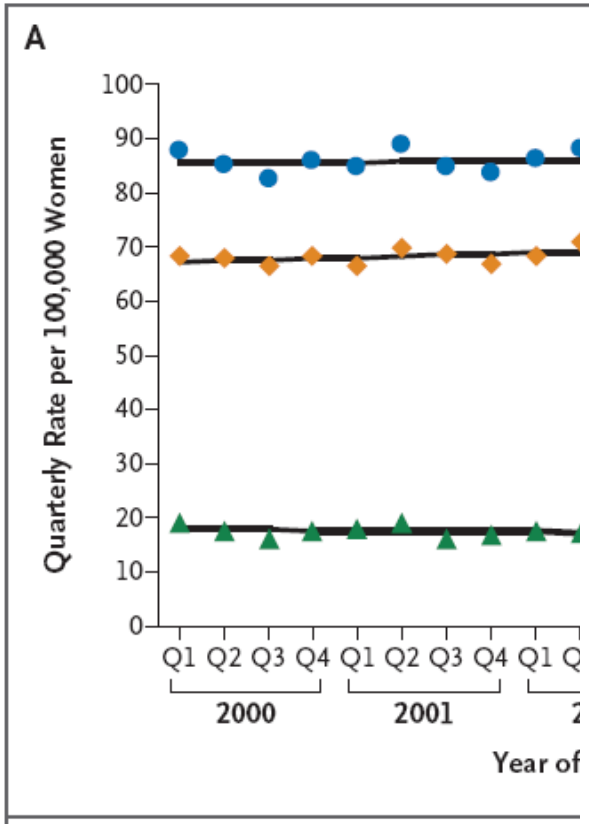
Results On May 31, 2002, after a mean of 5.2 years of follow-up, the data and safety monitoring board recommended stopping the trial of estrogen plus progestin vs placebo because the test statistic for invasive breast cancer exceeded the stopping boundary for this adverse effect and the global index statistic supported risks exceeding benefits. This report includes data on the major clinical outcomes through April 30, 2002. Estimated hazard ratios (HRs) (nominal 95% confidence intervals [CIs]) were as follows: CHD, 1.29 (1.02-1.63) with 286 cases; breast cancer, 1.26 (1.00-1.59) with 290 cases; stroke, 1.41 (1.07-1.85) with 212 cases; PE, 2.13 (1.39-3.25) with 101 cases; colorectal cancer, 0.63

The NEW ENGLAND JOURNAL of MEDICINE

SPECIAL REPORT

The Decrease in Breast-Cancer Incidence [illegible] States

Ph.D., Nadia Howlader, M.S.,
[illegible] Ph.D.



age-adjusted incidence was 8.6% (95% confidence interval [CI], 6.8 to 10.4). The decrease was evident only in women who were 50 years of age or older and was more evident in cancers that were estrogen-receptor–positive than in those that were estrogen-receptor–negative. The decrease in breast-cancer incidence seems to be temporally related to the first report of the Women's Health Initiative and the ensuing drop in the use of hormone-replacement therapy among postmenopausal women in the United States. The contributions

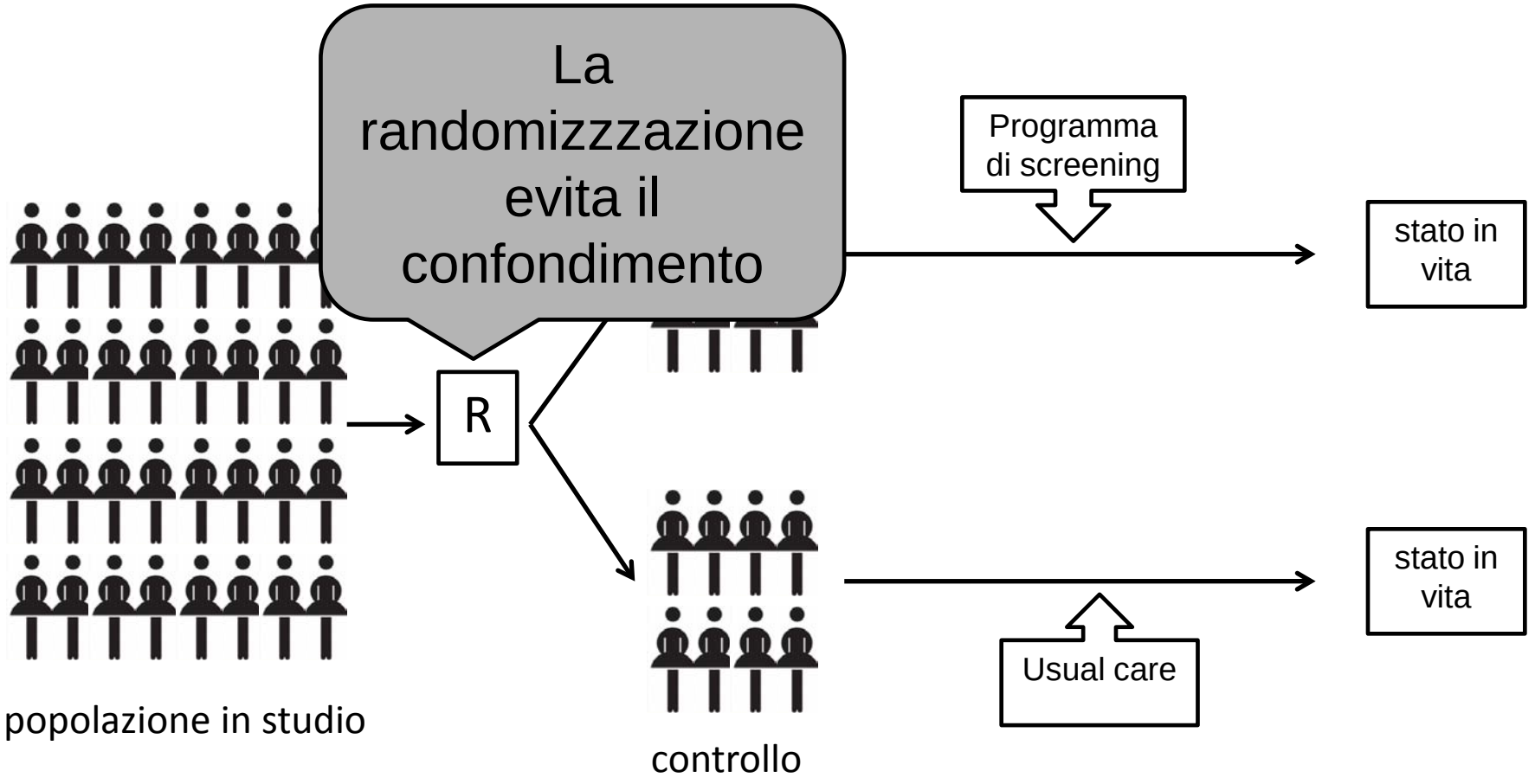
Il confondimento

- Il confondente è ***una concausa dell'evento in studio*** che è statisticamente associata con il fattore determinante
- La conoscenza dei possibili confondenti permette la produzione di stime aggiustate da modelli statistici multivariati
- ma ***non si può mai escludere un confondimento residuo***

E allora?

- In teoria l'unico modo per controllare per ogni fattore di confondimento è ***l'analisi controfattuale***:
  - Confrontare cioè un individuo esposto al fattore in studio ***con lo stesso individuo NON esposto***
- ovviamente ciò non è possibile *(forse nei crossover randomised trials)*
- il disegno che meglio approssima le caratteristiche dell'osservazione controfattuale è il RCT

Lo studio randomizzato - RCT



La validità di un RCT

CRITERIO	BIAS
random	selection bias
allocation concealment	
blinding	
incomplete outcome data	
selective reporting	reporting bias

in popolazioni sane?

valutazione use di morte

poco rilevante negli studi su BCS

poco rilevante in studi di grandi dimensioni

La validità di un RCT

indicatore di qualità più importante

random sequence generation	selection bias
allocation concealment	selection bias
blinding	performance bias selection bias
incomplete outcome data	attrition bias
selective reporting	reporting bias

E gli studi osservazionali?

nella valuta

La Cochrane accusata di dogmatismo negli scorsi anni!

- si tratta di studi con u[illegible]
- usualmente in presenza di studi randomizzati di qualità, non sono presi in considerazione

Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials

Gordon C S Smith, Jill P Pell

Abstract

Objectives To determine whether parachutes are effective in preventing major trauma related to gravitational challenge.
Design Systematic review of randomised controlled trials.
Data sources: Medline, Web of Science, Embase, and the Cochrane Library databases; appropriate internet sites and citation lists

accepted intervention was a fabric device, secured by strings to a harness worn by the participant and released (either automatically or manually) during free fall with the purpose of limiting the rate of descent. We excluded studies that had no control group.

Definition of outcomes
The major outcomes studied were death or major trauma, defined as an injury severity score greater than 15.[6]

data. We think that everyone might benefit if the most radical protagonists of evidence based medicine organised and participated in a double blind, randomised, placebo controlled, crossover trial of the parachute.

data. We think that everyone might benefit if the most radical protagonists of evidence based medicine organised and participated in a double blind, randomised, placebo controlled, crossover trial of the parachute.

Results

Our search strategy did not find any randomised controlled trials of the parachute.

Department of Obstetrics and Gynaecology, Cambridge University, Cambridge CB2 2QQ
Gordon C S Smith *professor*

Department of Public Health, Greater Glasgow

E gli studi osservazionali?

- Il ruolo degli studi osservazionali è molto cambiato in questi anni
- la Cochrane collaboration (*che non li ha mai esclusi come fonte ausiliaria di dati*), ne sta definendo i criteri di validità
- Ma il vero cambiamento è stato il GRADE (GRADE working group, BMJ, 2008)

RATING QUALITY OF EVIDENCE AND STRENGTH OF RECOMMENDATIONS

GRADE: an emerging consensus on rating quality of evidence and strength of recommendations

Guidelines are inconsistent in how they rate the quality of evidence and the strength of recommendations. This article explores the advantages of the GRADE system, which is increasingly being adopted by organisations worldwide

Guideline developers around the world are inconsistent in how they rate quality of evidence and grade strength of recommendations. As a result, guideline users face challenges in understanding the messages that grading systems try to communicate. Since 2006 the *BMJ* has requested in its "Instructions to Authors" on bmj.com that authors should preferably use the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system for grading evidence when submitting a clinical guidelines article. What was behind this decision?

In this first in a series of five articles we will explain why many organisations use formal systems to grade

Gordon H Guyatt professor, Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, ON, Canada L8N 3Z5

Andrew D Oxman researcher, Norwegian Knowledge Centre for the Health Services, PO Box 7004, St Olavs Plass, 0130 Oslo, Norway

Gunn E Vist researcher, Norwegian Knowledge Centre for the Health Services, PO Box 7004, St Olavs Plass, 0130 Oslo, Norway

Regina Kunz associate professor, Basel Institute of Clinical Epidemiology, University Hospital

advantages and disadvantages but also by their confidence in these estimates. The cartoon depicting the weather forecaster's uncertainty captures the difference between an assessment of the likelihood of an outcome and the confidence in that assessment (figure). The usefulness of an estimate of the magnitude of intervention effects depends on our confidence in that estimate.

Expert clinicians and organisations offering recommendations to the clinical community have often erred as a result of not taking sufficient account of the quality of evidence.[2] For a decade, organisations recommended that clinicians encourage postmenopausal women to use hormone replacement therapy.[3] Many primary care phy-

Randomizzati e osservazionali contribuiscono insieme all'evidenza

Study Design	Quality of Evidence	Lower if	Higher if
Randomised trial →	High	Risk of bias -1 Serious -2 Very serious Inconsistency -1 Serious -2 Very serious Indirectness -1 Serious -2 Very serious Imprecision -1 Serious -2 Very serious Publication bias -1 Likely -2 Very likely	Large effect +1 Large +2 Very large Dose response +1 Evidence of a gradient All plausible confounding +1 Would reduce a demonstrated effect or +1 Would suggest a spurious effect when results show no effect
	Moderate		
Observational study →	Low		
	Very low		

Utilità degli studi osservazionali

- oltre a surrogare l'evidenza di efficacia di RCT
  - per esposizioni non randomizzabili
  - quando randomizzare può essere non etico o non accettato (es: *studio VEdeTTE per trattamenti TD*)
- per outcome rari o distanti nel tempo (es: *mortalità nell'efficacia della terapia metadonica*)
- per misurare differenze di effetto in sottogruppi di popolazione (es*: One million women per la TOS*)
- per misurare l'impatto di programmi a livello di popolazione

Utilità per l'assessement dello screening mammografico

EFFETTO SULLA MORTALITÀ

incidence-based mortality studies

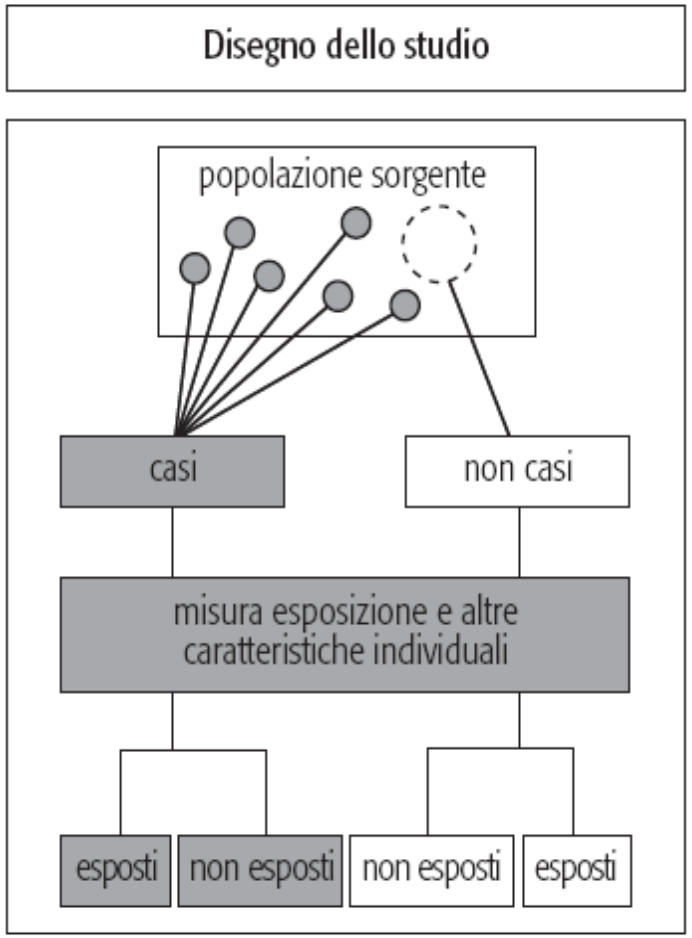
- donne con dx di ca mammario precedentemente invitate allo screening
- confrontate con:
  - donne non invitate
  - anni pre-screening
  - combinazione dei due
- risultati: comparabili con screening
- bias: diverse assunzioni producono risultati diversi

Utilità per l'assessement dello screening mammografico

EFFETTO SULLA MORTALITÀ

studi caso-controllo

- casi: decedute x K mammario
- ctrl: vive, appaiate x età (residenza)
- risultati: stessa direzione ma > RCT
- bias:
  - healthy screened effect (SES..)
  - trattamento più efficace



Utilità per l'assessement dello screening mammografico

EFFETTO SULLA MO

studi di serie storic

effetto di cause diverse

– migliori trattamenti

– screening

– consapevolezza

– dopo il 2001, calo TOS

impossibili da distinguere

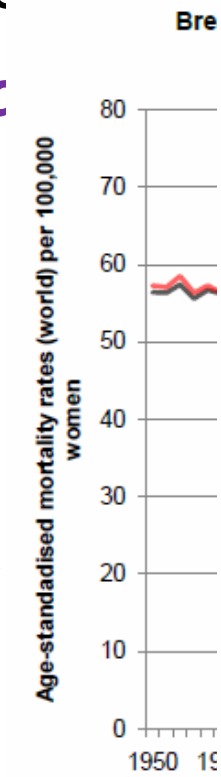


Figure 2.2 Breast
Mortality database
Accessed October

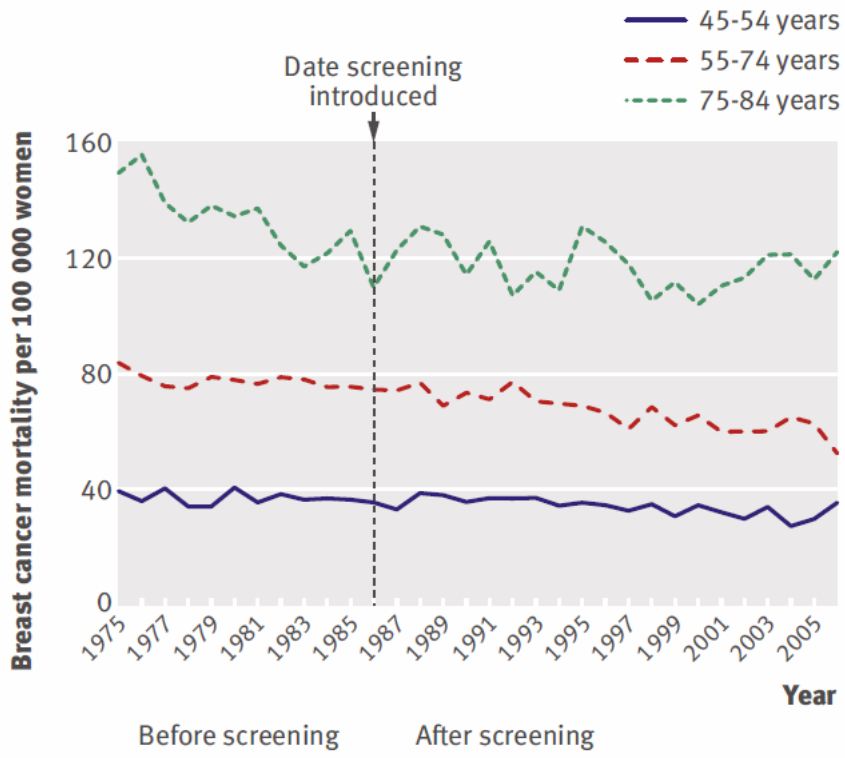


Fig 3 | Unadjusted breast cancer mortality rates in Sweden for screened and non-screened age groups. Data from Statistics Sweden (http://www.scb.se/default____2154.aspx)

Jorgensen BMJ 2010

Evidenza di effetto sulla mortalità

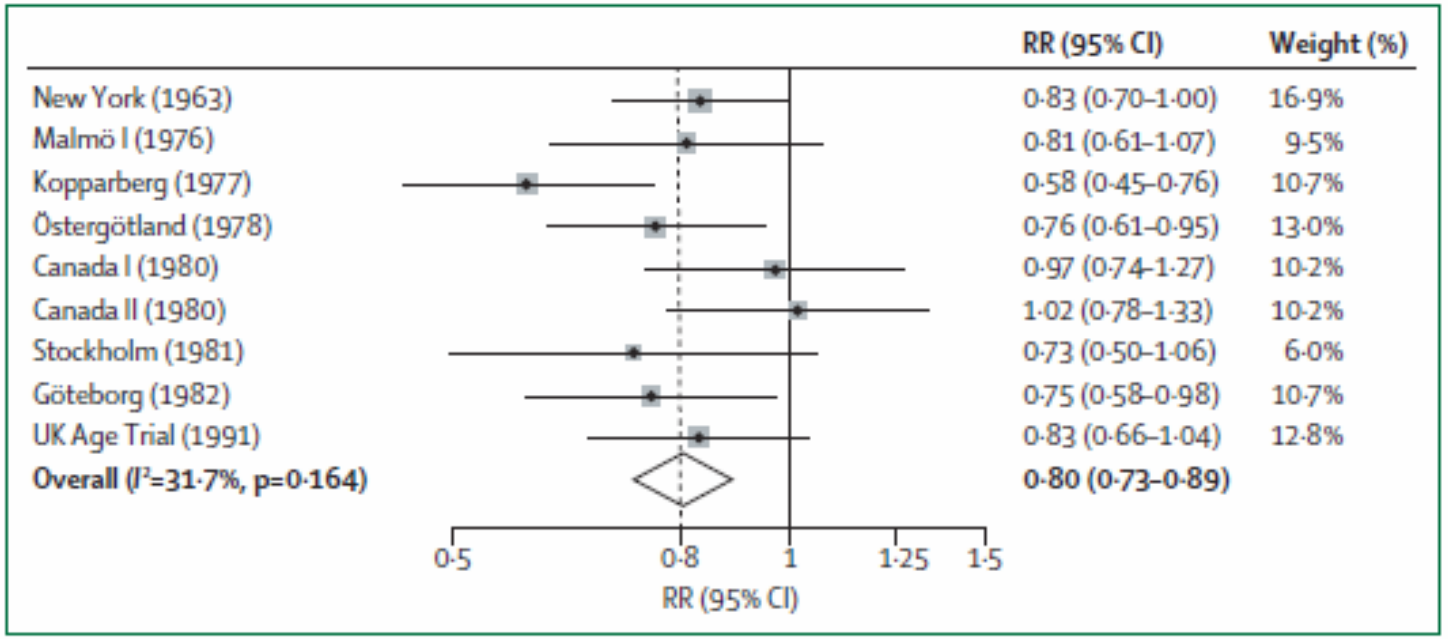


Figure 1: **Meta-analysis of breast cancer mortality after 13 years of follow-up in breast cancer screening trials**
Adapted from the Cochrane Review.[5] RR=relative risk. Malmö II is excluded because follow-up of about 13 years was not available; the Swedish Two County (Kopparberg and Östergötland) and Canada I and II trials are split into their component parts; the Edinburgh trial is excluded because of severe imbalances between randomised groups. Weights are from random-effects analysis.

UK Independent panel 2012

Utilità per l'assessement dello screening mammografico

STIMA SOVRADIAGNOSI

da serie storiche

- incidenza tumori prima e dopo l'i (o offerta di screening)
- in fasce di età interessate dallo sc
- vs altre fasce di età

- assunzioni diverse producono s

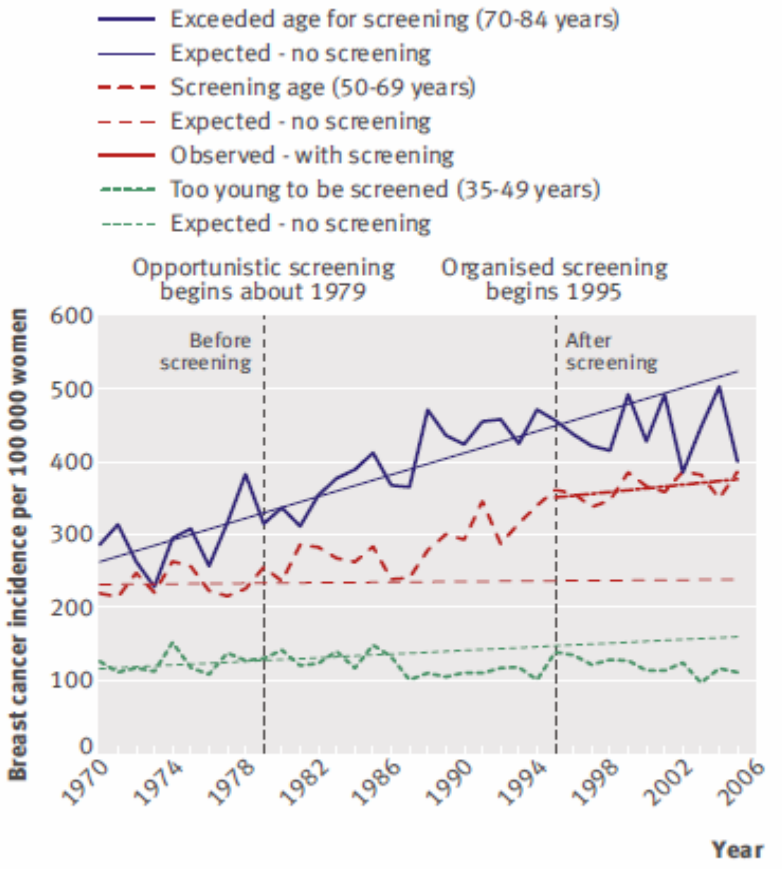


Fig 3| Incidence of invasive breast cancer and carcinoma in situ per 100 000 women in Manitoba, Canada

Jorgensen BMJ 2009

Stima della sovradiagnosi

STIMA SOVRADIAGNOSI

da RCT

- a

	A	B	C	D
Malmö I ages 55–69 years	11·7% (82/698)	10·5% (82/780)	18·7% (82/438)	29·1% (82/282)
Canada I	14·1% (82/581)	12·4% (82/663)	22·7% (82/361)	29·4% (82/279)
Canada II	10·7% (67/626)	9·7% (67/693)	16·0% (67/420)	19·8% (67/338)

Numbers of excess cancers are expressed as a percentage of different denominators. A=excess cancers as a proportion of cancers diagnosed over whole follow-up period in unscreened women. B=excess cancers as a proportion of cancers diagnosed over whole follow-up period in women invited for screening. C=excess cancers as a proportion of cancers diagnosed during screening period in women invited for screening. D=excess cancers as a proportion of cancers detected by screening in women invited for screening.

Table 4: **Estimates of overdiagnosis in randomised trials without systematic end-of-trial screening of the control group, according to four calculation methods**

- Le
  - m
  - pi
  - –
  - –
- M
  - pc

	Overall RR (95% CI)
This review	
13-year follow-up in trials reported in the Cochrane Review[5] random-effects meta-analysis	0·80 (0·73–0·89)
Cochrane review[5]	
Fixed-effect meta-analysis of the above trials	0·81 (0·74–0·87)
As above, but excluding women <50 years	0·77 (0·69–0·86)
Trials considered adequately randomised (Canada, Malmö, and UK Age trial) had RR 0·90 (95% CI 0·79–1·02); trials deemed suboptimally randomised gave RR 0·75 (0·67–0·83). As a compromise between these two estimates, the authors concluded that an RR of 0·85 was plausible	0·85
US Task Force[9]	
RR 0·86 (95% CI 0·75–0·99) for women aged 50–59 years, and RR 0·68 (0·54–0·87) for those aged 60–69 years. These estimates have an inverse-variance weighted average RR of 0·81	0·81
Canadian Task Force[4]	
Routinely screening for breast cancer with mammography every 2–3 years for women aged 50–69 years was rated as a weak recommendation based on moderate-quality evidence according to GRADE criteria[11]	0·79 (0·68–0·90)
Duffy et al, 2012[10]	
Review of all trials and age groups	0·79 (0·73–0·86)

RR=relative risk.

Table 2: **Estimates of RR in a comparison of invited women versus control women in the trials of breast cancer screening**

UK Independent panel 2012

Una sintesi: Trial e nulla più?

- Il metodo scientifico (e la EBM negli ultimi anni) è andato molto oltre il trial, verso un concetto molto più complesso di

confidence in the effect estimates

- a cui contribuiscono tutte le osservazioni scientifiche, “pesate” per i loro bias
- Ma spesso le evidenze producono stime di effetto incerte, in particolare nelle aree delicate in cui i benefici non hanno grandi dimensioni e i danni (*harms*) sono misurabili.

Altri criteri di giudizio – il GRADE

Determinants of strength of recommendation

Factor	Comment
Balance between desirable and undesirable effects	The larger the difference between the desirable and undesirable effects, the higher the likelihood that a strong recommendation is warranted. The narrower the gradient, the higher the likelihood that a weak recommendation is warranted
Quality of evidence	The higher the quality of evidence, the higher the likelihood that a strong recommendation is warranted
Values and preferences	The more values and preferences vary, or the greater the uncertainty in values and preferences, the higher the likelihood that a weak recommendation is warranted
Costs (resource allocation)	The higher the costs of an intervention—that is, the greater the resources consumed—the lower the likelihood that a strong recommendation is warranted

Quando anche i moderni framework decisionali non bastano…

- Il GRADE non è stato sufficiente da solo a sciogliere il conflitto sullo screening mammografico.

E allora?

- Il ricorso ad un panel indipendente, composto da scienziati stimati in tutti gli ambienti, con grandi competenze metodologiche, e "*Public health advocacy*" è stata una soluzione originale e di alto profilo.

E quando i Panel non bastano?

It is doubtful, however, that the independent panel changed the minds of the principal proscreening and antiscreening groups in the debate over screening.[6] Positions are too entrenched. But the evidence on breast screening is more extensive than in many other areas relevant to population health. If this is not enough for an independent group, coming fresh to the debate, to reach a reasonable judgment, then evidence-based policy is a good deal more difficult than many would believe.

Statement GISMA

- Lo statement GISMA si assume questa sfida. In particolare indicando le aree in cui concentrare gli sforzi
- la comunicazione
- la ricerca
  - aumentare la performance del test al test
    - strumenti
    - criteri di selezione di popolazione a rischio
  - fattori prognostici di progressione

Conclusioni

E se il GISMA è riuscito a mobilitare un così grande numero di professionisti e ricercatori per discutere di questo tema, evidentemente è per vincere la sfida