The problem with personalized risk

Beverly Rockhill Levine, Ph.D.

Individuals, populations, and risk

Disease prevention

Individualistic perspective:

- Preventive interventions should be targeted/ tailored to each person's individual susceptibility

Societal /population perspective:

- Mass diseases and mass exposures require mass remedies--Geoffrey Rose

"Individual susceptibility" strategy: Advantages

- Motivation: intervention matched to perceived needs of individual
- Can improve benefit-risk ratio of prevention intervention
- Readily accommodated within US health care system, consumerist environment

"Individual susceptibility" strategy: Disadvantages

- Disease prevention can become medicalized —being "at risk" becomes medical condition
- May overlook roots of public health problem
- Limited by poor ability to predict individuals' futures

How to identify susceptible individuals?

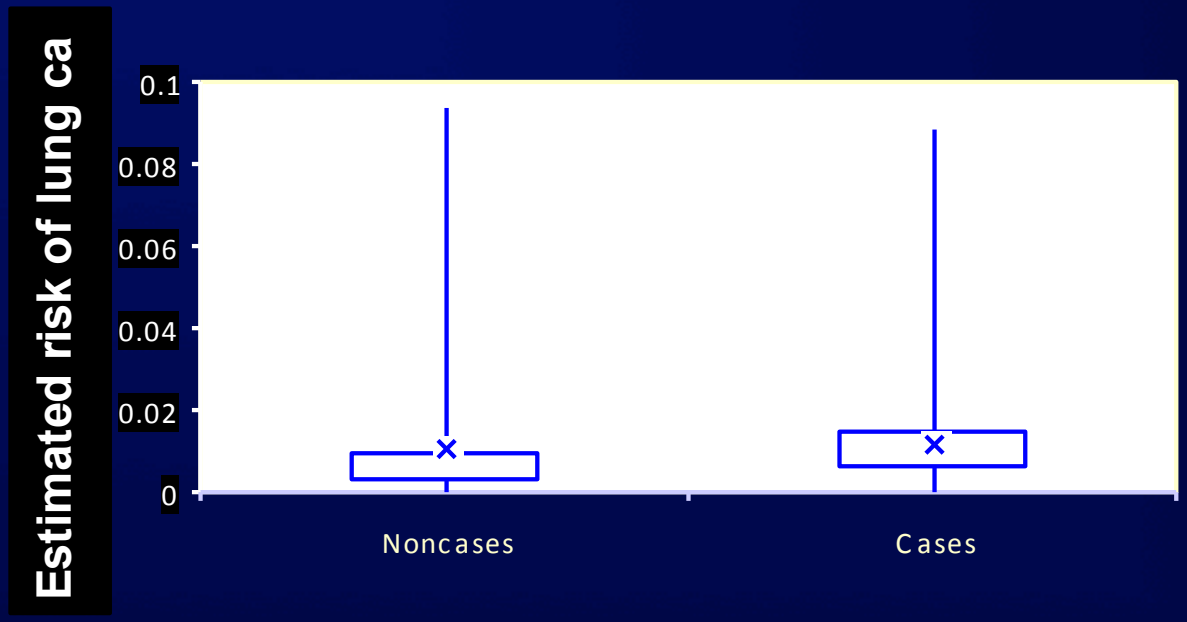
- Most identified (and "major") risk factors for noninfectious disease have modest associations with disease (RRs 1.2-3.0)
- Vast majority of these factors unnecessary, insufficient to cause disease—poor discriminators at individual level
- Bulk of disease cases will arise from mass of population with "average" risk factor values/ average predicted risks

A "best-case" scenario? Predicting risk of lung cancer

- 90,000 nurses (NHS) followed for 14 years
- Approximately 900 developed lung cancer over 14-year period
- Model predicting risk of lung cancer: age, smoking status defined in very fine strata
- RR comparing current, heavy, long-term smokers to never smokers ~30


Estimated risk of lung ca .
0.1
0.08
0.06
0.04
0.02
0
Noncases
Cases

Boxplots of 14-year estimated risk of lung cancer, by case status (NHS)



Assessing accuracy of risk prediction

- Calibration
- Discrimination

Assessing accuracy of risk prediction

- Calibration=goodness of fit; agreement between observed and expected number of events

- Example: if the average predicted risk for group of individuals over time interval is 0.10, and 10% of persons develop disease over that interval, model is well-calibrated

Assessing accuracy of risk prediction

Discrimination: ability to separate individuals with different outcomes

Assessing accuracy of risk prediction

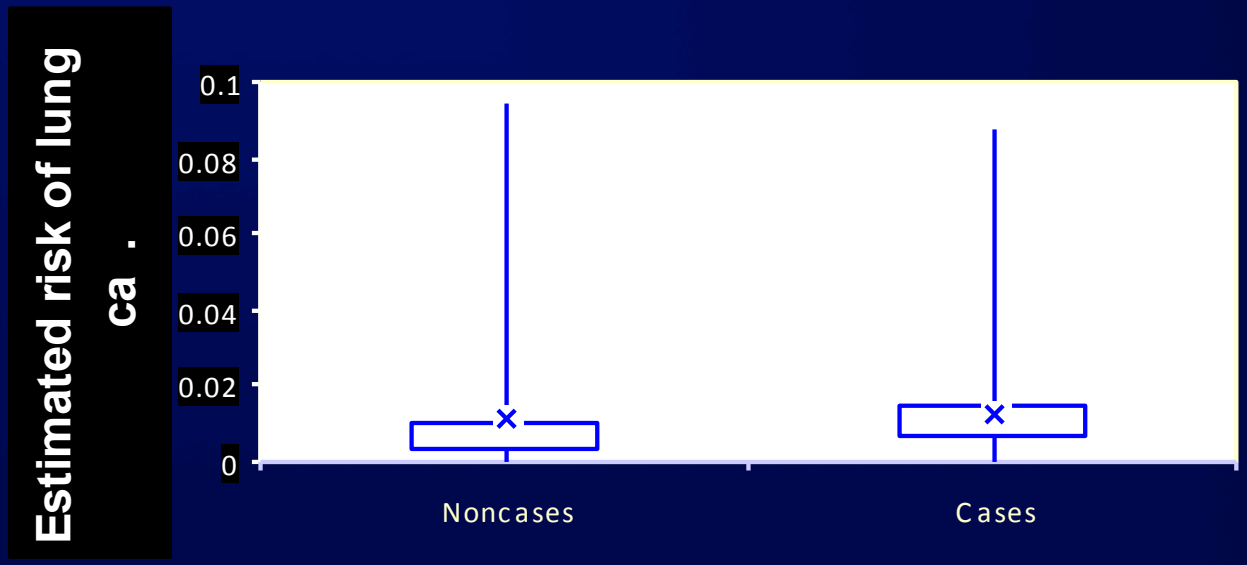
A model assigning everyone in population the same estimated risk of disease, say 0.10, is well-calibrated if 10% of population actually develops disease, but the model has no discriminatory ability.

Assessing accuracy of risk prediction

- A discriminating model produces wide distribution of estimated risks; these risks are consistently higher for persons who develop disease compared to those who do not.

- Discrimination can be assessed with concordance statistic (c-statistic)—ranges from 0.5 (coin flip) to 1.0

Boxplots of 14-year estimated risk of lung cancer, by case status (NHS)

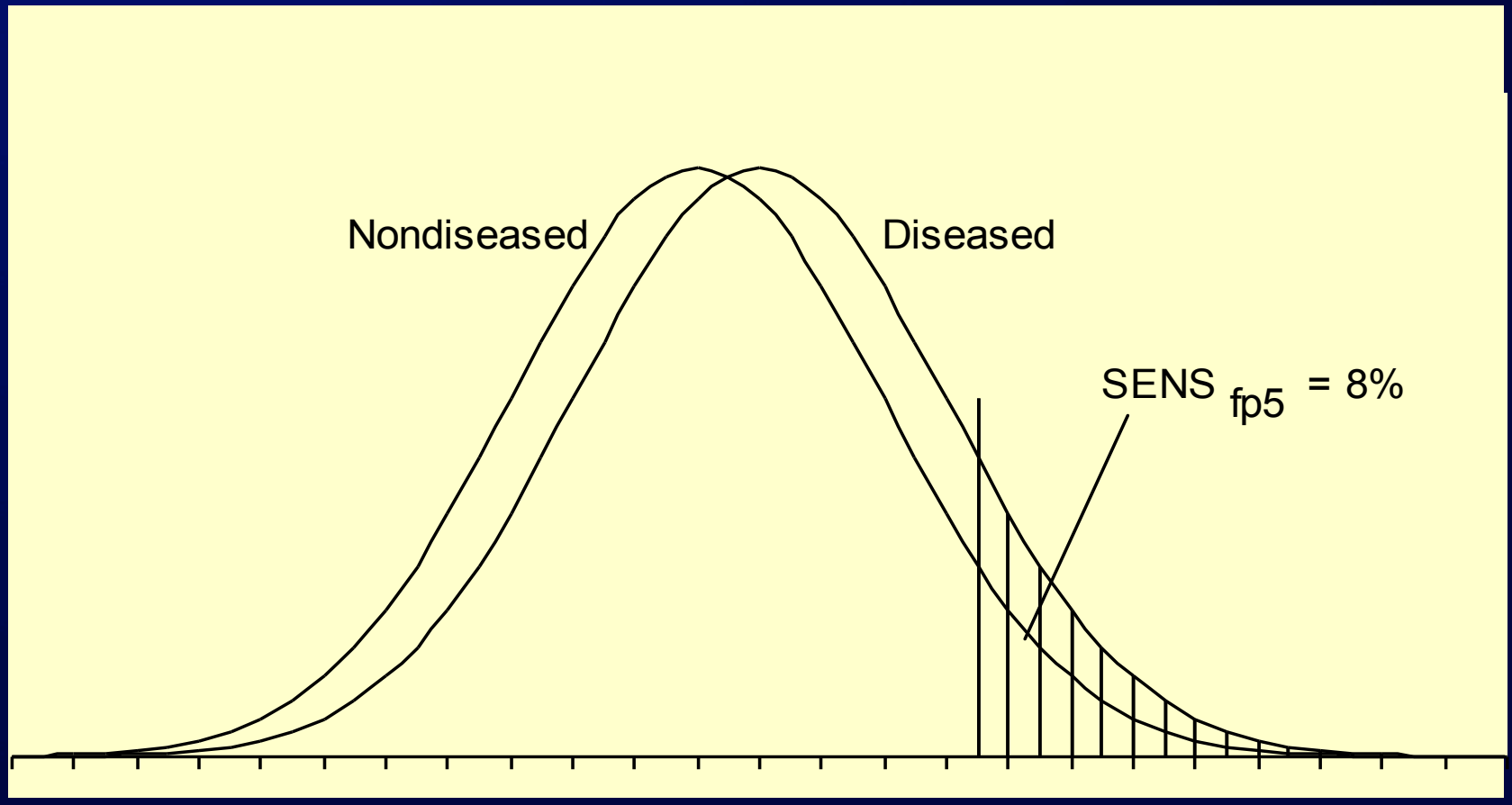


Concordance statistic = 0.60

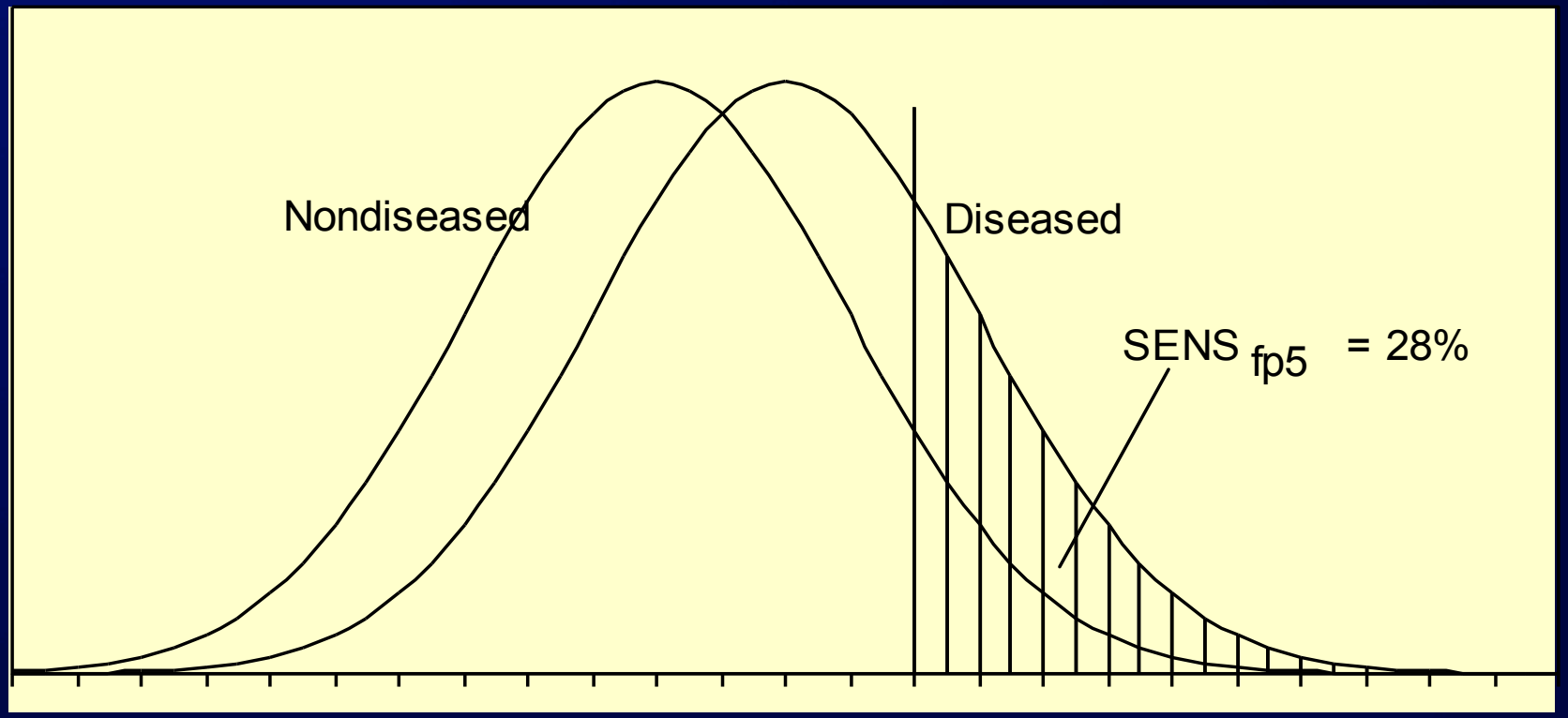
Risk factors as screening tools

A risk prediction tool must have a large associated relative risk (>>20) comparing the extremes of exposure or predicted risk in order to serve as useful screening tool (i.e., to discriminate well) at the individual level

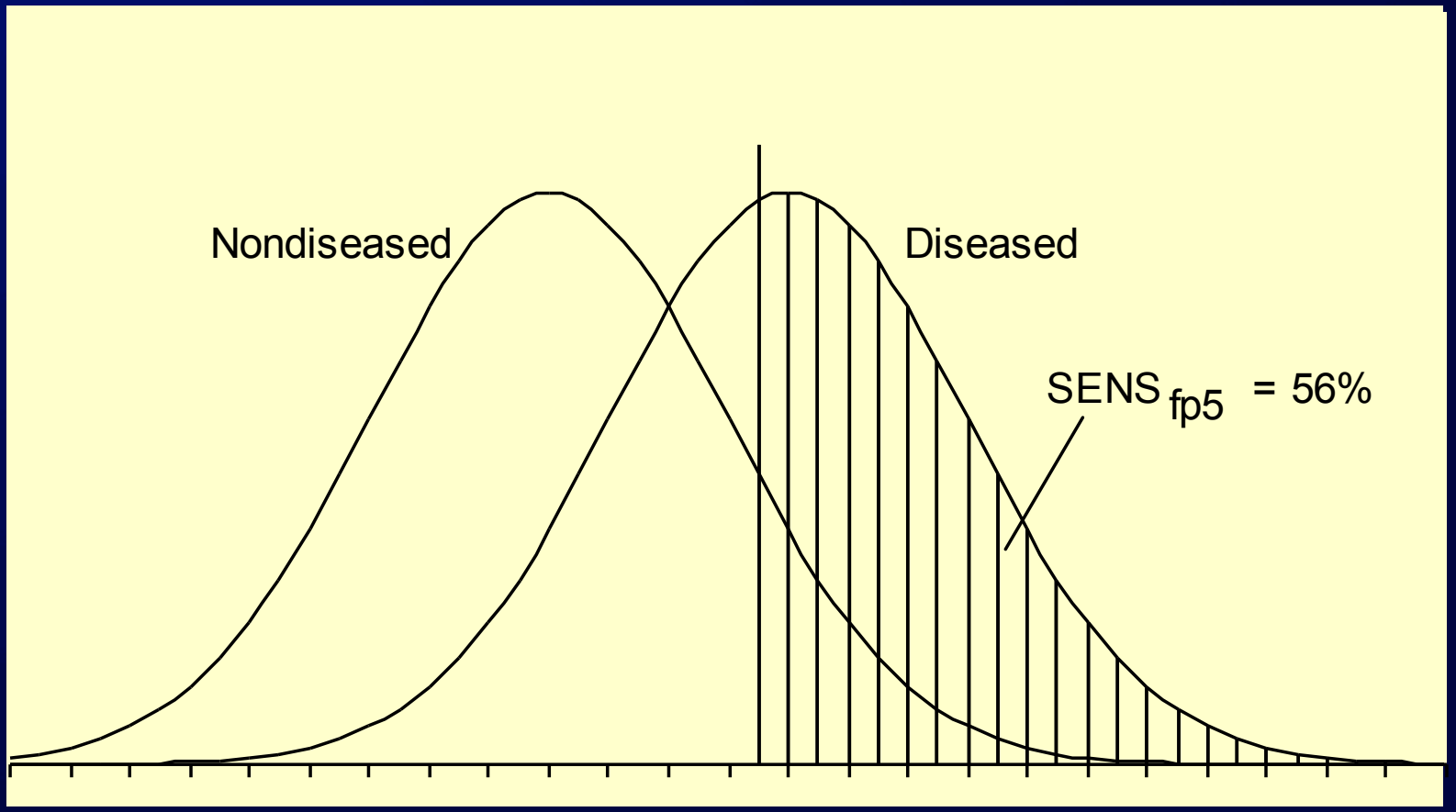
$RR_{q1\text{-}5} = 2$



$RR_{q1\text{-}5} = 20$



$RR_{q1\text{-}5} = 200$



A contemporary scenario: Breast cancer prevention

- Chemoprevention likely an important part of future breast cancer prevention strategies

- FDA guidelines for tamoxifen and raloxifene prophylaxis stated in terms of estimated five-year risk of breast cancer, based on Gail et al. model

Gail et al. model of breast cancer risk ("Risk Disk")

- Developed in 1989; used to estimated expected incidence in Breast Cancer Prevention Trial

- Risk factors in model: age, age at menarche, age at first birth/nulliparity, # of affected first-degree relatives, history of benign breast biopsy/hyperplasia

age-0.74948 + 0.09401 (AGEMEN) + 0.52926(NBIOPS)
+ 0.21863(AGEFLB) + 0.95830(NUMREL)
+ 0.01081 (AGECAT) - 0.28804(NBIOPS
X AGECAT) - 0.19081 (AGEFLB X NUMREL).

Gail et al. model of breast cancer risk ("Risk Disk")

- Risk Disk was widely distributed to clinicians, members of lay public, now available as online Risk Calculator—for use in clinical decision-making about tamoxifen

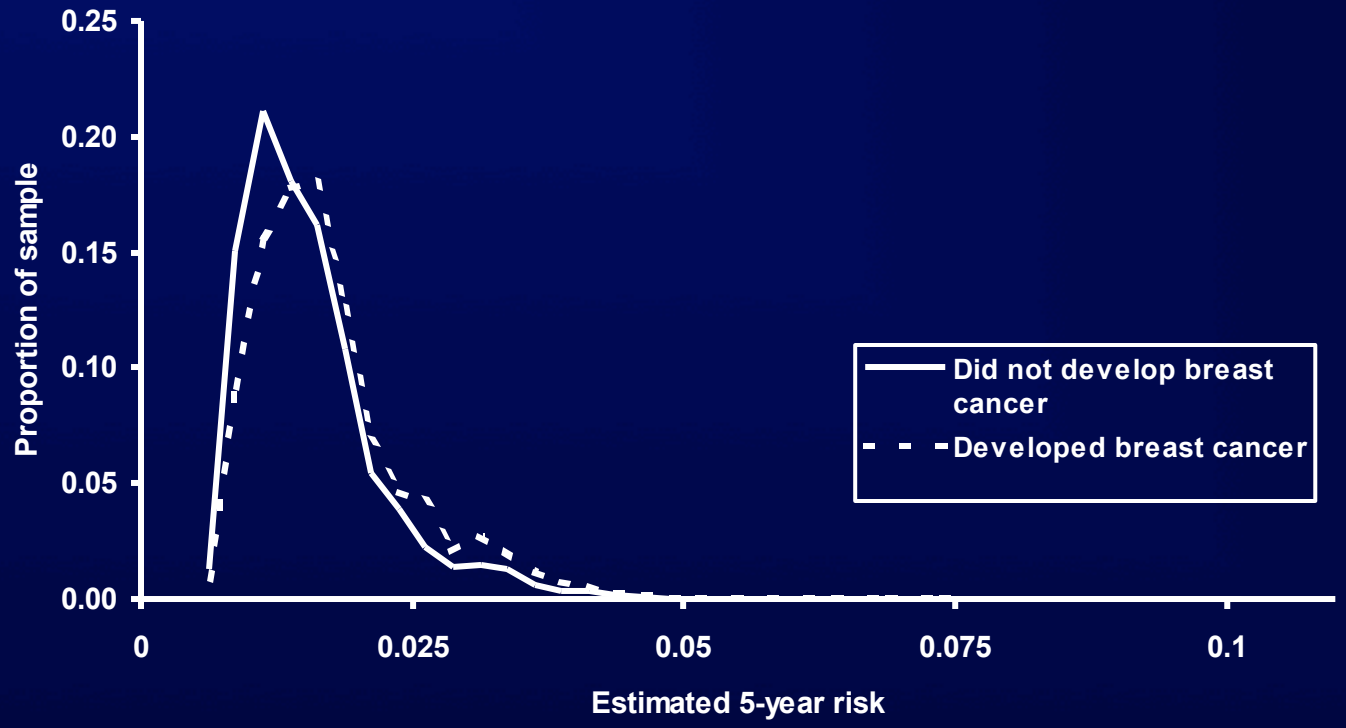
- http://www.cancer.gov/bcrisktool/

Gail et al. model validation in Nurses' Health Study

- Cohort of 82,109 white women aged 45-71 years
- Observed 1,354 cases of breast cancer over five-year period

- Expected/observed ratios:
  - total sample: 0.94 (0.89-0.99)
  - high-risk subsample: 1.03 (0.96-1.12)

Estimated five-year risk of breast cancer, according to breast cancer status at end of follow-up



Discriminatory accuracy of Gail et al. model

- Concordance statistic: 0.58 (0.56-0.60)
- RR comparing top to bottom decile of estimated 5-year risk: 2.83 (2.19-3.65)
- 44% of women who developed breast cancer had estimated risk >=1.67% (sens)
- 66% of women who remained disease-free had estimated risk <1.67% (spec)

New Gail Model findings

- Concordance statistic: 0.58 (0.56-0.60)

Age- only breast cancer risk prediction model

- Concordance statistic: 0.54

The line is drawn
in breast cancer risk

1.67% draws the line for breast cancer risk. NOLVADEX® reduces breast cancer incidence in many high-risk women.

More than a number,
a medical indicator.
It helps determine when a woman
is at high risk for breast cancer.

≥1.67%

General conclusions

- Failure to acknowledge that good predictors of incidence (i.e., of averages) have poor individual-level discriminatory accuracy has led to unrealistic expectations about discovery of "causes" and about opportunities for "personalized" prevention/ intervention

- Ability of information on genes, gene/environment interaction, to greatly improve discrimination of risk prediction models not yet apparent

Implications with respect to "individualized" prevention and risk communication

Risk communication for prevention

- Strategy relies, largely, on well-intended but frequently inaccurate persuasion to "change" ("false positives") in addition to well-intended but frequently inaccurate reassurance ("false negatives").

Risk communication for prevention

- In most circumstances, rational individual provided with all relevant numeric epidemiologic information should NOT change behavior/ exposure (based on quantitative data alone)
  - E.g., CRC risk and diet change—"high risk" of 15/10,000 in 5 years for 60-yo American; "increase in fruit and veg consumption can lower risk by 30%...."
  - "NNT" to prevent 1 case=1/(abs risk difference)
  - =2000, in this example

Risk communication for prevention

-
- A rational “numerate” individual could (should?) justifiably *dismiss* most (not all) epidemiologic information as personally unconvincing or irrelevant—at least on quantitative grounds

Clinical perspective vs. public health perspective

- Knowledge of rare, but "highly penetrant," risk factors (e.g., BRCA1) important in clinic; indeed, "individualized prevention" likely an important part of medicine ever since physicians began to recognize that individuals with certain traits, trajectories were at observedly higher risk of certain outcomes than others

Clinical perspective vs. public health perspective

- In public health, success is measured in declines in incidence/mortality over time.
- While a risk factor with a low c-statistic (e.g., BRCA1/2) can be highly informative clinically, low c-stat means substantial proportion of individuals in general population who end up with disease will NOT be identifiable through high predicted risks, thus will not be targeted through "individualized prevention" programs

General conclusions

- Poor ability to single out small minority of individuals who will develop disease with almost any statistical technique means prevention strategy, even "individualized prevention," will have to affect many, to meaningfully reduce disease burden

- *"Mass remedies" needed—which are socially acceptable?*

Ovarian Cancer Screening Dec 2015

Ovarian cancer screening and mortality in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial

Ian J Jacobs, Usha Menon*, Andy Ryan, Aleksandra Gentry-Maharaj, Matthew Burnell, Jatinderpal K Kalsi, Nazar N Amso, Sophia Apostolidou, Elizabeth Benjamin, Derek Cruickshank, Danielle N Crump, Susan K Davies, Anne Dawnay, Stephen Dobbs, Gwendolen Fletcher, Jeremy Ford, Keith Godfrey, Richard Gunu, Mariam Habib, Rachel Hallett, Jonathan Herod, Howard Jenkins, Chloe Karpinskyj, Simon Leeson, Sara J Lewis, William R Liston, Alberto Lopes, Tim Mould, John Murdoch, David Oram, Dustin J Rabideau, Karina Reynolds, Ian Scott, Mourad W Seif, Aarti Sharma, Naveena Singh, Julie Taylor, Fiona Warburton, Martin Widschwendter, Karin Williamson, Robert Woolas, Lesley Fallowfield, Alistair J McGuire, Stuart Campbell, Mahesh Parmar†, Steven J Skates†*

Summary

Background Ovarian cancer has a poor prognosis, with just 40% of patients surviving 5 years. We designed this trial to establish the effect of early detection by screening on ovarian cancer mortality.

Use of proteomic patterns in serum to identify ovarian cancer

Methods Proteomic spectra were generated by mass spectroscopy (surface-enhanced laser desorption and ionisation). A preliminary "training" set of spectra derived from analysis of serum from 50 unaffected women and 50 patients with ovarian cancer were analysed by an iterative searching algorithm that identified a proteomic pattern that completely discriminated cancer from non-cancer. The discovered pattern was then used to classify an independent set of 116 masked serum samples: 50 from women with ovarian cancer, and 66 from unaffected women or those with non-malignant disorders.

Findings The algorithm identified a cluster pattern that, in the training set, completely segregated cancer from non-cancer. The discriminatory pattern correctly identified all 50 ovarian cancer cases in the masked set, including all 18 stage I cases. Of the 66 cases of non-malignant disease, 63 were recognised as not cancer. This result yielded a sensitivity of 100% (95% CI 93–100), specificity of 95% (87–99), and positive predictive value of 94% (84–99).

Numbers counterintuitive!

Ovarian cancer screening example:

sens. 1.0,

spec 0.95,

prev 50/100,000 (US women aged 50+ yrs)

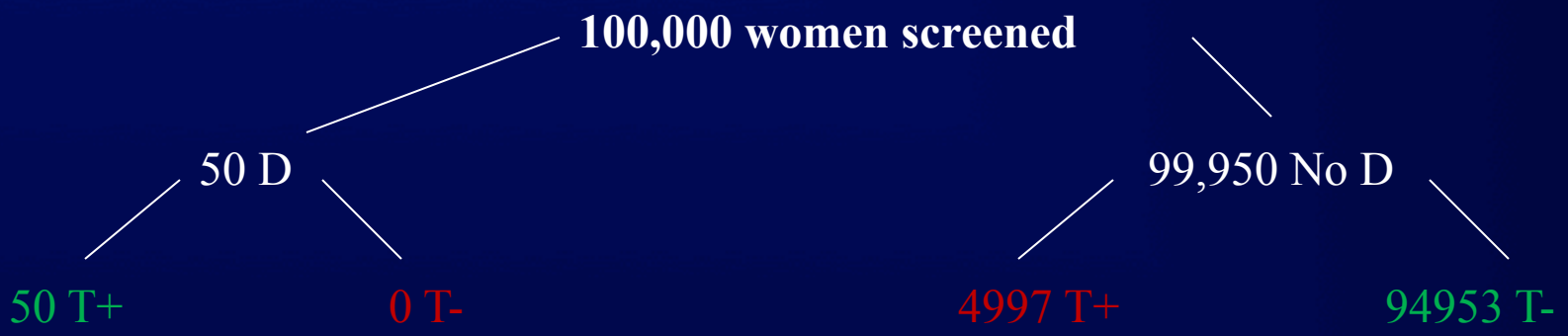
Numbers counterintuitive

Ovarian cancer screening example:

sens. 1.0,

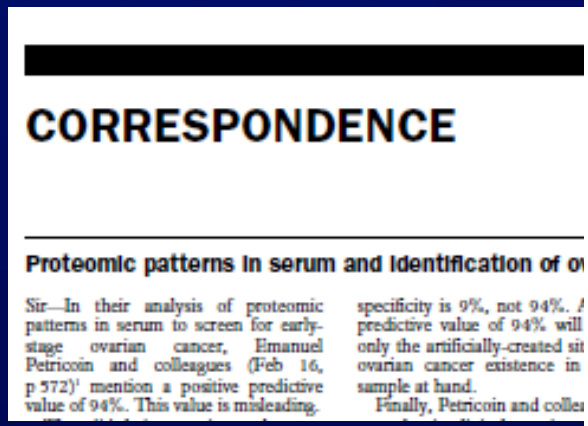
spec 0.95,

prev 50/100,000 (US women aged 50+ yrs)



PPV=50/(50+4997) =<1%

Correspondence 2002



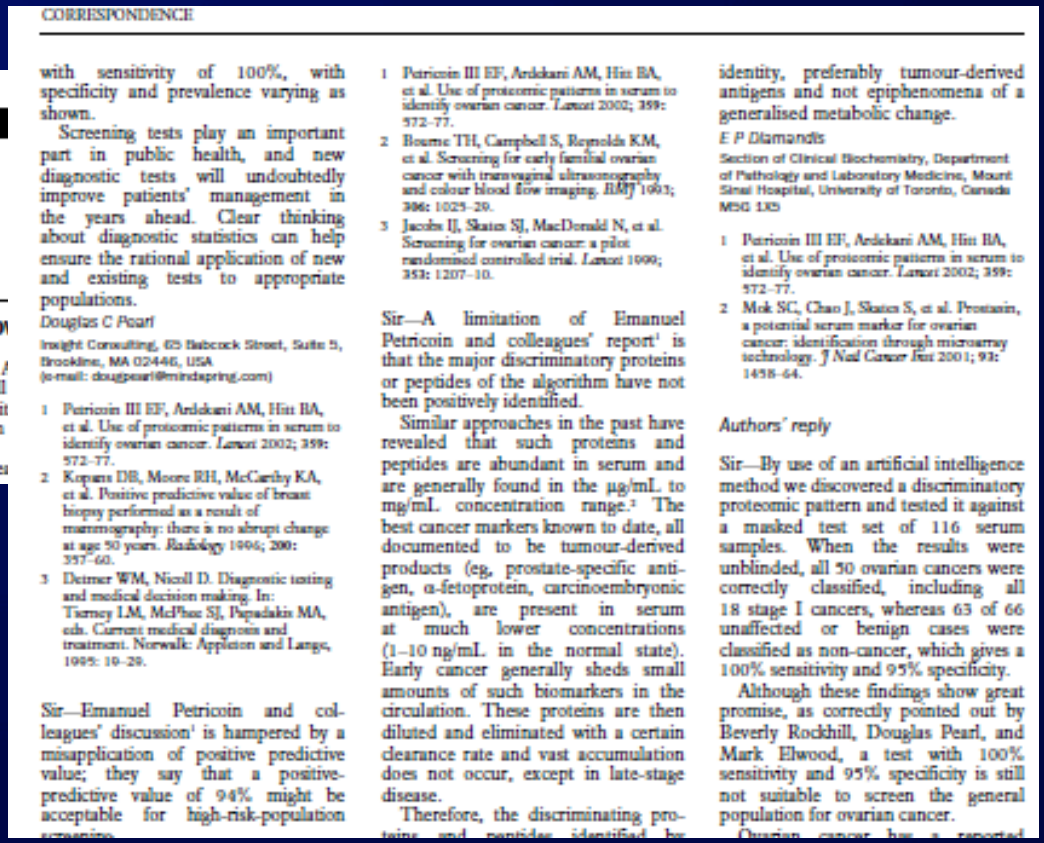
CORRESPONDENCE

Proteomic patterns in serum and identification of ov

Sir—In their analysis of proteomic patterns in serum to screen for early-stage ovarian cancer, Emanuel Petricoin and colleagues (Feb 16, p 572)[1] mention a positive predictive value of 94%. This value is misleading.

specificity is 9%, not 94%. A predictive value of 94% will only the artificially-created sit ovarian cancer existence in sample at hand.

Finally, Petricoin and colle



CORRESPONDENCE

with sensitivity of 100%, with specificity and prevalence varying as shown.

Screening tests play an important part in public health, and new diagnostic tests will undoubtedly improve patients' management in the years ahead. Clear thinking about diagnostic statistics can help ensure the rational application of new and existing tests to appropriate populations.

Douglas C Pearl

Insight Consulting, 65 Babcock Street, Suite 5, Brookline, MA 02446, USA
(e-mail: dougpearl@mindspring.com)

1 Petricoin III EF, Ardekani AM, Hitt BA, et al. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* 2002; **359**: 572–77.
2 Kopans DB, Moore RH, McCarthy KA, et al. Positive predictive value of breast biopsy performed as a result of mammography: there is no abrupt change at age 50 years. *Radiology* 1996; **200**: 357–60.
3 Detmer WM, Nicoll D. Diagnostic testing and medical decision making. In: Tierney LM, McPhee SJ, Papadakis MA, eds. Current medical diagnosis and treatment. Norwalk: Appleton and Lange, 1995: 19–29.

Sir—Emanuel Petricoin and colleagues' discussion[1] is hampered by a misapplication of positive predictive value; they say that a positive-predictive value of 94% might be acceptable for high-risk-population

1 Petricoin III EF, Ardekani AM, Hitt BA, et al. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* 2002; **359**: 572–77.
2 Bourne TH, Campbell S, Reynolds KM, et al. Screening for early familial ovarian cancer with transvaginal ultrasonography and colour blood flow imaging. *BMJ* 1993; **306**: 1025–29.
3 Jacobs IJ, Skates SJ, MacDonald N, et al. Screening for ovarian cancer: a pilot randomised controlled trial. *Lancet* 1999; **353**: 1207–10.

Sir—A limitation of Emanuel Petricoin and colleagues' report[1] is that the major discriminatory proteins or peptides of the algorithm have not been positively identified.

Similar approaches in the past have revealed that such proteins and peptides are abundant in serum and are generally found in the μg/mL to mg/mL concentration range.[2] The best cancer markers known to date, all documented to be tumour-derived products (eg, prostate-specific antigen, α-fetoprotein, carcinoembryonic antigen), are present in serum at much lower concentrations (1–10 ng/mL in the normal state). Early cancer generally sheds small amounts of such biomarkers in the circulation. These proteins are then diluted and eliminated with a certain clearance rate and vast accumulation does not occur, except in late-stage disease.

Therefore, the discriminating pro-

identity, preferably tumour-derived antigens and not epiphenomena of a generalised metabolic change.

E P Diamandis

Section of Clinical Biochemistry, Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, University of Toronto, Canada M5G 1X5

1 Petricoin III EF, Ardekani AM, Hitt BA, et al. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* 2002; **359**: 572–77.
2 Mok SC, Chao J, Skates S, et al. Prostasin, a potential serum marker for ovarian cancer: identification through microarray technology. *J Natl Cancer Inst* 2001; **93**: 1458–64.

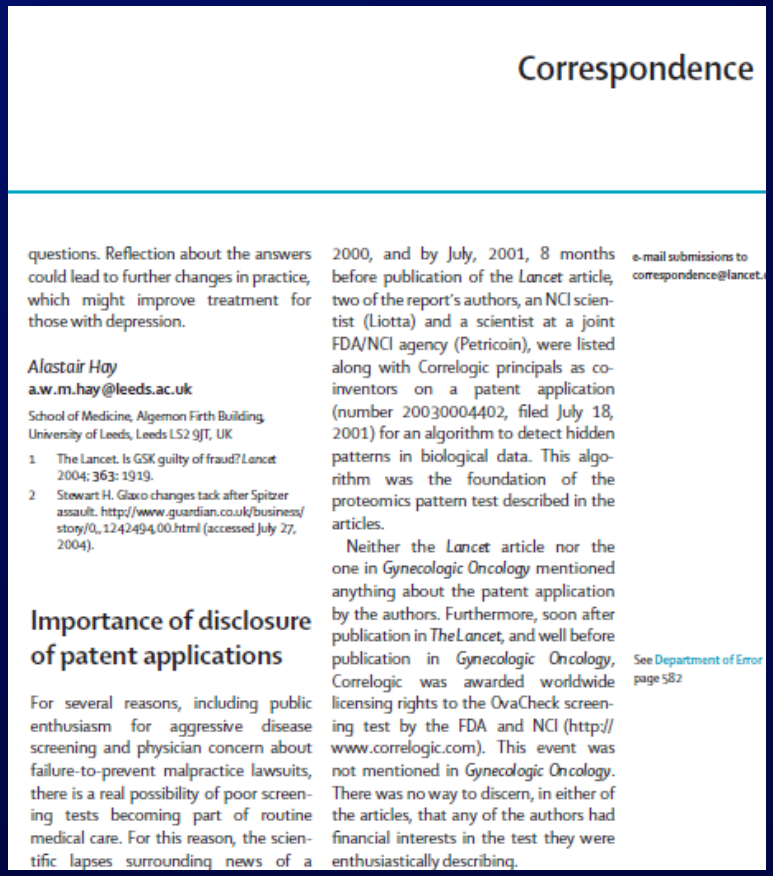
Authors' reply

Sir—By use of an artificial intelligence method we discovered a discriminatory proteomic pattern and tested it against a masked test set of 116 serum samples. When the results were unblinded, all 50 ovarian cancers were correctly classified, including all 18 stage I cancers, whereas 63 of 66 unaffected or benign cases were classified as non-cancer, which gives a 100% sensitivity and 95% specificity.

Although these findings show great promise, as correctly pointed out by Beverly Rockhill, Douglas Pearl, and Mark Elwood, a test with 100% sensitivity and 95% specificity is still not suitable to screen the general population for ovarian cancer.

Ovarian cancer has a reported

Correspondence 2004



questions. Reflection about the answers could lead to further changes in practice, which might improve treatment for those with depression.

Alastair Hay
a.w.m.hay@leeds.ac.uk

School of Medicine, Algernon Firth Building, University of Leeds, Leeds LS2 9JT, UK

1 The Lancet. Is GSK guilty of fraud? *Lancet* 2004; 363: 1919.
2 Stewart H. Glaxo changes tack after Spitzer assault. http://www.guardian.co.uk/business/story/0,,1242494,00.html (accessed July 27, 2004).

Importance of disclosure of patent applications

For several reasons, including public enthusiasm for aggressive disease screening and physician concern about failure-to-prevent malpractice lawsuits, there is a real possibility of poor screening tests becoming part of routine medical care. For this reason, the scientific lapses surrounding news of a 2000, and by July, 2001, 8 months before publication of the *Lancet* article, two of the report's authors, an NCI scientist (Liotta) and a scientist at a joint FDA/NCI agency (Petricoin), were listed along with Correlogic principals as co-inventors on a patent application (number 20030004402, filed July 18, 2001) for an algorithm to detect hidden patterns in biological data. This algorithm was the foundation of the proteomics pattern test described in the articles.

Neither the *Lancet* article nor the one in *Gynecologic Oncology* mentioned anything about the patent application by the authors. Furthermore, soon after publication in *The Lancet*, and well before publication in *Gynecologic Oncology*, Correlogic was awarded worldwide licensing rights to the OvaCheck screening test by the FDA and NCI (http://www.correlogic.com). This event was not mentioned in *Gynecologic Oncology*. There was no way to discern, in either of the articles, that any of the authors had financial interests in the test they were enthusiastically describing.

e-mail submissions to correspondence@lancet.

See Department of Error page 582

Ovarian Cancer Screening Dec 2015

Ovarian cancer screening and mortality in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial

Ian J Jacobs, Usha Menon*, Andy Ryan, Aleksandra Gentry-Maharaj, Matthew Burnell, Jatinderpal K Kalsi, Nazar N Amso, Sophia Apostolidou, Elizabeth Benjamin, Derek Cruickshank, Danielle N Crump, Susan K Davies, Anne Dawnay, Stephen Dobbs, Gwendolen Fletcher, Jeremy Ford, Keith Godfrey, Richard Gunu, Mariam Habib, Rachel Hallett, Jonathan Herod, Howard Jenkins, Chloe Karpinskyj, Simon Leeson, Sara J Lewis, William R Liston, Alberto Lopes, Tim Mould, John Murdoch, David Oram, Dustin J Rabideau, Karina Reynolds, Ian Scott, Mourad W Seif, Aarti Sharma, Naveena Singh, Julie Taylor, Fiona Warburton, Martin Widschwendter, Karin Williamson, Robert Woolas, Lesley Fallowfield, Alistair J McGuire, Stuart Campbell, Mahesh Parmar†, Steven J Skates†*

Summary

Background Ovarian cancer has a poor prognosis, with just 40% of patients surviving 5 years. We designed this trial to establish the effect of early detection by screening on ovarian cancer mortality.