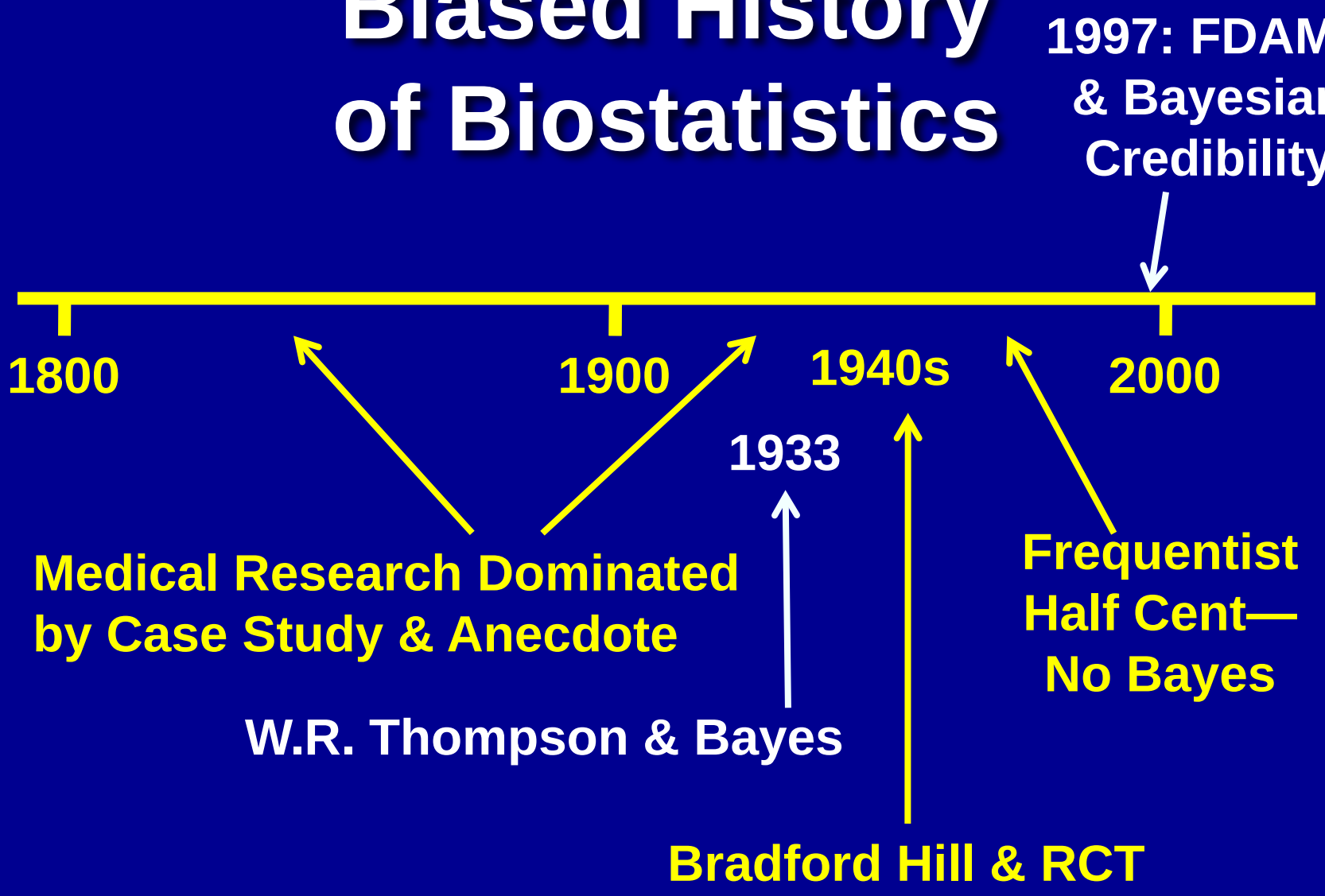
Bayesian Ideas Are (Slowly) Revolutionizing Medical Research —A Personal Perspective

Donald A. Berry
dberry@mdanderson.org

THE UNIVERSITY OF TEXAS
MD Anderson Cancer Center
Making Cancer History®

Biased History of Biostatistics



My exposures

- 1970: Bandit Problems
- 1980: 3M Riker Labs
- 1990: Duke, CALGB & breast cancer
- 1990s: FDA & Bayes
- 1997: Screening mammography in 40s
- 1999: M.D. Anderson Cancer Center
- 2015: Bandit Problems!

Top 5 Reasons for Bayes

1. On-line learning
2. Predictive probabilities
3. Hierarchical modeling
4. Modeling generally
5. Decision analysis

Genetics & Bayes (80s & 90s)

- DNA fingerprinting
- BRCAPRO
- Doping
- Paternity testing
- Determining ancestry of corn, cabernet sauvignon, etc. using SNPs

Appl. Statist. (1992)
41, *No.* 3, *pp.* 499–531

Statistical Inference in Crime Investigations using Deoxyribonucleic Acid Profiling

By D. A. BERRY,

Duke University, Durham, USA

I. W. EVETT† and R. PINCHIN

Home Office Forensic Science Service, Aldermaston, UK

[*Read before* The Royal Statistical Society on *Wednesday, November 13th, 1991, the President,* Professor T. M. F. Smith, *in the Chair*]

SUMMARY

Deoxyribonucleic acid profiling has attracted widespread publicity because of the impact that it is making on the investigation of crime. Whereas considerable effort has been expended on the refinement of the laboratory systems for carrying out the technique, the development of efficient numerical procedures for evaluating the evidence contained in the profiles has attracted little attention. We have developed a method, based on the Bayesian likelihood ratio, for evaluating fragment length data in a crime case where a profile from a suspect is to be compared with a profile from a sample taken from the scene of the crime. Our treatment takes account of correlation in fragment length measurement errors and avoids an independence assumption which is currently being made by practitioners. We describe experiments which demonstrate the superiority of the method over the conventional method which is based on simple hypothesis tests.

Keywords: Bayes; Bayes factor; Bivariate normal; Density estimation; Deoxyribonucleic acid; Forensic science; Likelihood ratio; Match-binning; Measurement error; Smoothing

BRCAPRO Validation, Sensitivity of Genetic Testing of *BRCA1/BRCA2*, and Prevalence of Other Breast Cancer Susceptibility Genes

Conclusion: BRCAPRO is an accurate counseling tool for determining the probability of carrying mutations of BRCA1 and BRCA2. Genetic testing for BRCA1 and BRCA2 is highly sensitive, missing an estimated 15% of mutations. In the populations studied, breast cancer susceptibility genes other than BRCA1 and BRCA2 either do not exist, are rare, or are associated with low disease penetrance.

for 301 individuals were made using BRCAPRO, a statistical model and software using Mendelian genetics and Bayesian updating. Model predictions were compared with the results of genetic testing.

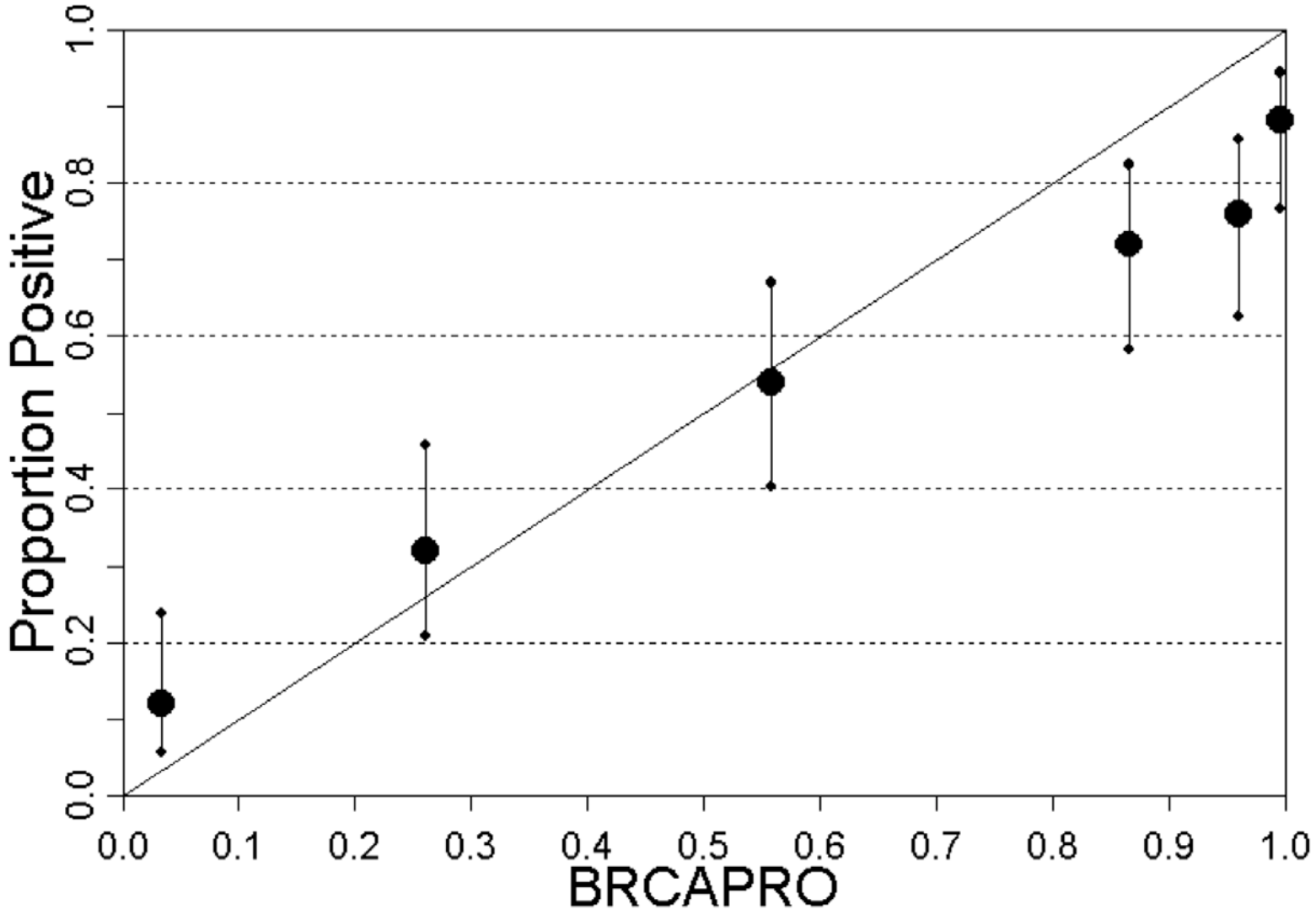
Results: Among the test individuals, 126 were Ashkenazi Jewish, three were male subjects, 243 had breast cancer, 49 had ovarian cancer, 34 were unaffected, and 139 tested positive for *BRCA1* mutations

BRCA1 and *BRCA2*. Genetic testing for *BRCA1* and *BRCA2* is highly sensitive, missing an estimated 15% of mutations. In the populations studied, breast cancer susceptibility genes other than *BRCA1* and *BRCA2* either do not exist, are rare, or are associated with low disease penetrance.

J Clin Oncol 20:2701-2712. © *2002 by American Society of Clinical Oncology.*

INDIVIDUALS WITH a family history of breast and/or ovarian cancer are at increased risk of carrying delete-

and second-degree relatives. (We use "carrier probability" to mean the probability of carrying a deleterious mutation of


Proportion Positive
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
BRCAPRO

NATURE|Vol 454|7 August 2008

The science of doping ... or lack thereof

The processes used to charge athletes with cheating are often based on flawed science and flawed logic, says **Donald A. Berry**.

Recently, the international Court of Arbitration for Sport upheld doping charges against cyclist Floyd Landis, stripping him of his title as winner of the 2006 Tour de France and suspending him from competition for two years. The court agreed with the majority opinion of a divided three-member US Anti-Doping Agency (USADA) arbitration panel and essentially placed a stamp of approval on a laboratory test indicating that Landis had taken synthetic testosterone. Although Landis asserts his innocence, his options for recourse have all but dried up.

Already, in the run-up to this year's Olympic Games, vast sums of time, money and media coverage have been spent on sports doping. Several doping experts have contended that tests aren't sensitive enough and let dozens of cheaters slip through the cracks. And some athletes are facing sanctions. US swimmer Jessica Hardy, upon testing positive for Clenbuterol,

P. DEJONG/AP

How scientically rigorous were the tests that disqualified Floyd Landis from the 2006 Tour de France?

NATURE|Vol 454|7 August 2008

The science of doping

The processes used to charge athletes with cheating are often based on flawed science and flawed logic, says **Donald A. Berry**.

Recently, the international Court of Arbitration for Sport upheld doping charges against cyclist Floyd Landis, stripping him of his title as winner of the 2006 Tour de France and suspending him from competition for two years. The court agreed with the majority opinion of a divided three-member US Anti-Doping Agency (USADA) arbitration panel and essentially placed a stamp of approval on a laboratory test indicating that Landis had taken synthetic testosterone. Although Landis asserts his innocence, his options for recourse have all but dried up.

Already, in the run-up to this year's Olympic Games, vast amounts of time, money and media coverage have been spent on sports doping. Several doping experts have contended that tests aren't sensitive enough and let dozens of cheaters slip through the cracks. And some athletes are facing sanctions. US swimmer Jessica Hardy, upon testing positive for clenbuterol, was held back from the Olympic team and faces

P. DEJONG/AP

Floyd Landis (centre) was disqualified after winning the 2006 Tour de France.

Prosecutor's fallacy

One factor at play in many cases that involve statistical reasoning, is what's known as the prosecutor's fallacy[1]. At its simplest level, it concludes guilt based on an observation that would be extremely rare if the person were innocent. Consider a blood test that perfectly matches a suspect to the perpetrator of a crime. Say, for example, the matching profile occurs in just 1 out of every 1,000 people. A naive prosecutor might try to convince a jury that the odds of guilt are 999:1, that is, the probability of guilt is 0.999. The correct way to determine odds comes from Bayes rule[2–4] and is equal to 999 times $P/(1-P)$ where P is the 'prior probability' of guilt. Prior probabil-

Google: Mark Buchanan "The Prosecutor's Fallacy"
Or: Peter Donnelly "TED Conference"

http://news.err.ee/sports/447c6353-50f1-46ee-83ce-c5470dada70f

After Veerpalu Ruling, All Eyes on WADA

SEE ALSO

- NFL Players 'Hail' Veerpalu Verdict
- Expert Calls Anti-Doping Agency Test Flawed

“The March 26 [2013] acquittal of Andrus Veerpalu in a high-profile doping case has attracted the attention of major sports organizations, and other athletes may follow in the Estonian skier's footsteps to dispute test results issued by the World Anti-Doping Agency.”

“Truth and justice of our Andrus Veerpalu”

“We identify with our heroes; they become a part of our self-image. … there are some beautiful messages in the vast numbers on ‘We Believe’ page. … these people [are] unwilling to count themselves cheats. So, their moral beliefs were expressed by faith in our hero.”

http://news.postimees.ee/1183132/editorial-truth-and-justice-of-our-andrus-veerpalu/

“Thus, Estonian mathematician [biostatistician] Krista Fischer with her academic colleagues from Tartu and US has defeated an organisation [WADA] claiming to be able, in an academically clean manner, to stand for clean sports.”

http://news.postimees.ee/1183132/editorial-truth-and-justice-of-our-andrus-veerpalu/

An Extreme in Junk Science

- Establishing “decision limits” of HGH isoform ratio
- Determining comparison: 106 “non-dopers”
- How selected?
- False-positive rate: 1 in 10,000!

From dope to corn …

Assessing Probability of Ancestry Using Simple Sequence Repeat Profiles: Applications to Maize Hybrids and Inbreds

Donald A. Berry,*[,1] Jon D. Seltzer,[†] Chongqing Xie,[‡] Deanne L. Wright[‡] and J. Stephen C. Smith[‡]

**The University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030, [†]Third Wave Technologies, Inc., Madison, Wisconsin 53719 and [‡]Pioneer Hi-Bred International, Inc., Johnston, Iowa 50131*

Manuscript received July 24, 2001
Accepted for publication March 11, 2002

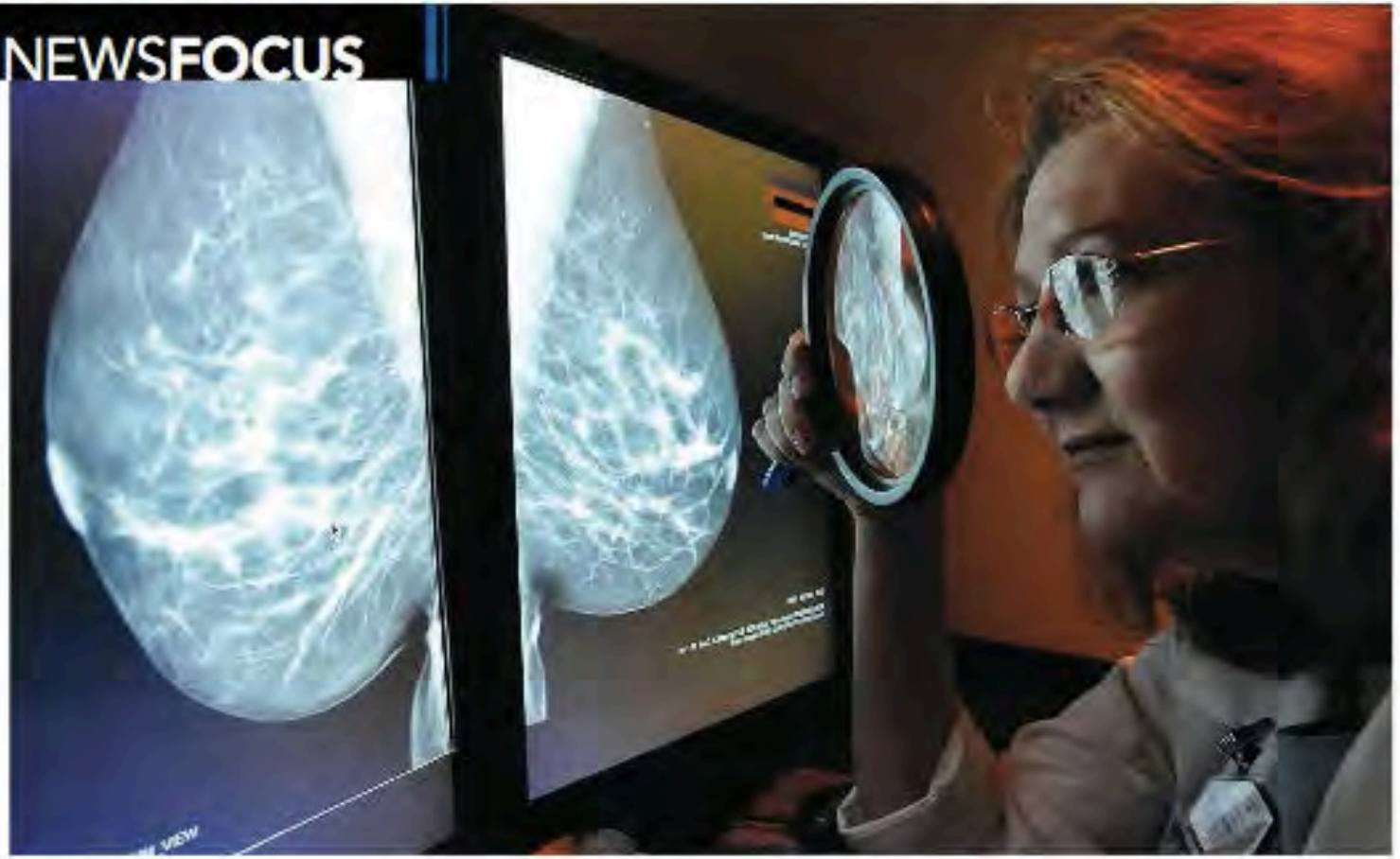
ABSTRACT

Determination of parentage is fundamental to the study of biology and to applications such as the identification of pedigrees. Limitations to studies of parentage have stemmed from the use of an insufficient number of hypervariable loci and mismatches of alleles that can be caused by mutation or by laboratory error and that can generate false exclusions. Furthermore, most studies of parentage have been limited to comparisons of small numbers of specific parent-progeny triplets thereby precluding large-scale surveys of candidates where there may be no prior knowledge of parentage. We present an algorithm that can determine probability of parentage in circumstances where there is no prior knowledge of pedigree and that is robust in the face of missing data or mistyped data. We present data from 54 maize hybrids and 586 maize inbreds that were profiled using 195 SSR loci including simulations of additional levels of missing and mistyped data to demonstrate the utility and flexibility of this algorithm.

DETERMINATION of parentage is fundamental to the study of reproductive and behavioral biology. The increasing availability of highly discriminant genetic markers for many diverse species provides the potential to uniquely characterize individuals at numerous loci and to unambiguously resolve parentage where genealogical relationships are unknown, in error, or in dispute (Gotz and Thaller 1998; Primmer *et al.* 2000; White *et al.* 2000).

Most studies of pedigree have utilized exclusion analysis where the molecular marker genotypes of either one or a restricted number of potential triplets of offspring and putative parents are compared. Often the identity of the mother is not in question; the maternal profile

Analysis was applied to soybeans, cattle, sunflowers, cabernet sauvignon, etc.



Brawling Over Mammography

A scientific study of the benefits and harms of screening women in their 40s got buried by politics

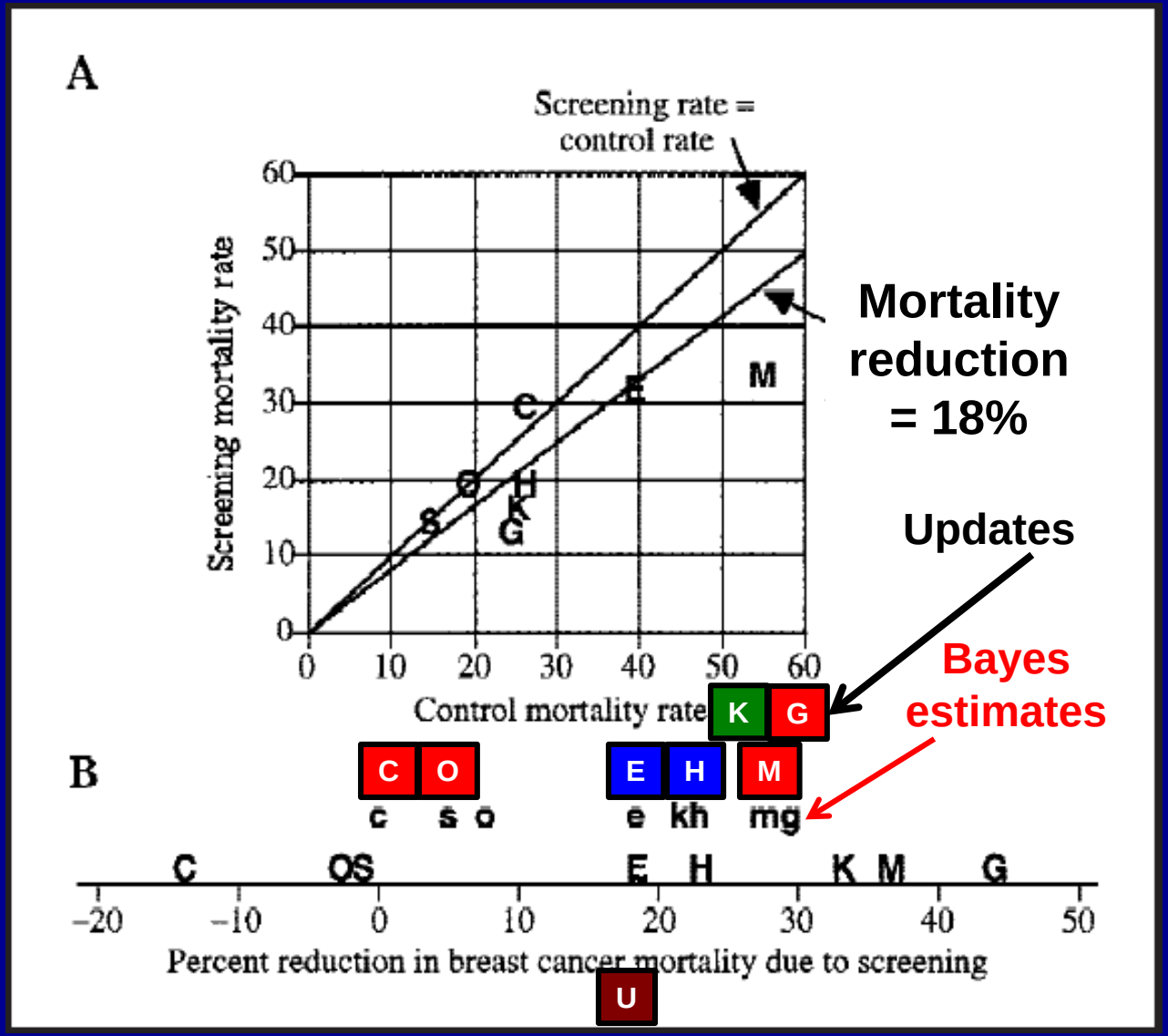
Researchers who had worked on the USPSTF guidelines were disappointed that their analysis was being dismissed out of hand. "Politics got in the way of the science and the best public health practice," says Jeanne Mandelblatt, an M.D.-epidemiologist at Georgetown University in Washington, D.C., and first author of an analysis for USPSTF by six groups that compared models to find the best screening strategy. "It was very unfortunate," adds Heidi Nelson,

Back up 12 years to 1997
U.S. NIH Consensus Conference: Mammography for women in 40s

“The Panel concludes that the data currently available do not warrant a universal recommendation for mammography for all women in their forties. Each woman should decide for herself whether to undergo mammography.”

Bayesian metaanalysis of randomized trials in 1997

Berry JNCI 1998 … with 2009 updates



Berry JNCI 1998

"A way to understand risks is to relate them with risks that are familiar. For example, the estimated average of 5 days of life lost if a woman in her early forties delays mammography for 10 years is similar to that for not wearing a seat belt over 20 years of typical automobile travel, of riding a bicycle for 15 hours without a helmet (or 50 hours if wearing a helmet), and of gaining two ounces of body weight (and keeping them on) *(41).*"

In spite of the evidence & the expert panel, the U.S. Senate …

- Voted 98-0 dictating that mammography would be effective (!) for women in their 40s
- Withheld NIH's budget until NCI agreed to recommend screening for women in their 40s [update]

2002 Cochran Review

U.S. Senate again jumps into the fray

Then, in 2005 …

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Effect of Screening and Adjuvant Therapy on Mortality from Breast Cancer

Donald A. Berry, Ph.D., Kathleen A. Cronin, Ph.D., Sylvia K. Plevritis, Ph.D., Dennis G. Fryback, Ph.D., Lauren Clarke, M.S., Marvin Zelen, Ph.D., Jeanne S. Mandelblatt, Ph.D., Andrei Y. Yakovlev, Ph.D., J. Dik F. Habbema, Ph.D., and Eric J. Feuer, Ph.D., for the Cancer Intervention and Surveillance Modeling Network (CISNET) Collaborators*

ABSTRACT

BACKGROUND

We used modeling techniques to assess the relative and absolute contributions of screening mammography and adjuvant treatment to the reduction in breast-cancer

of treatment with respect to the rate of death from breast cancer.

RESULTS

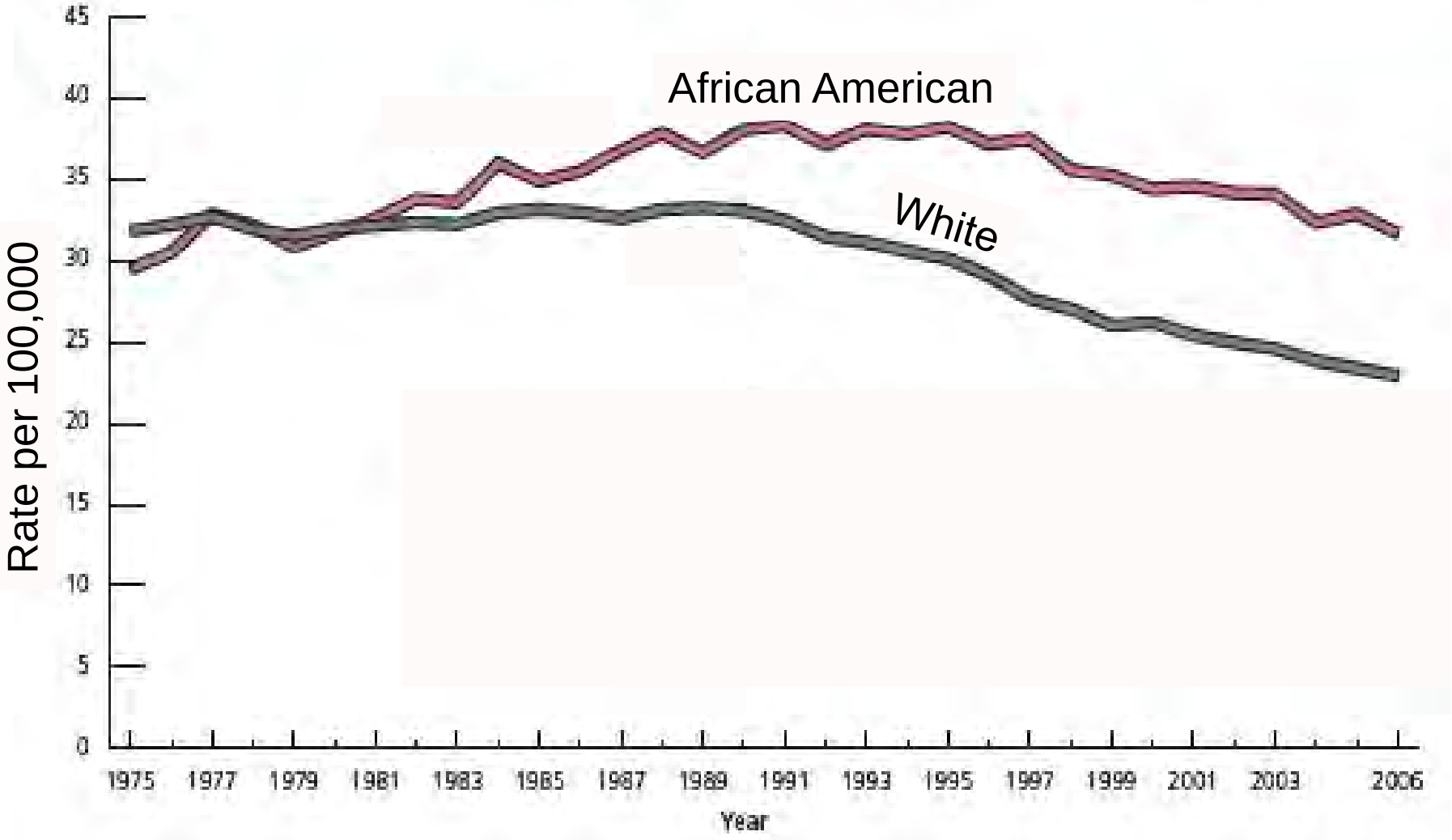
The proportion of the total reduction in the rate of death from breast cancer attributed to screening varied in the seven models from 28 to 65 percent (median, 46 percent),

From M.D. Anderson Cancer Center, Houston (D.A.B.); the National Cancer Institute, Beth
Univ
versi
(D.G
Wash
tute,
sity,
sity
and
Rotterdam, the Netherlands (J.D.F.H.). Address reprint requests to Dr. Berry at the Department of Biostatistics and Applied Mathematics, M.D. Anderson Cancer Center, Unit 447, 1515 Holcombe Blvd., Houston, TX 77030, or at dberry@mdanderson.org.

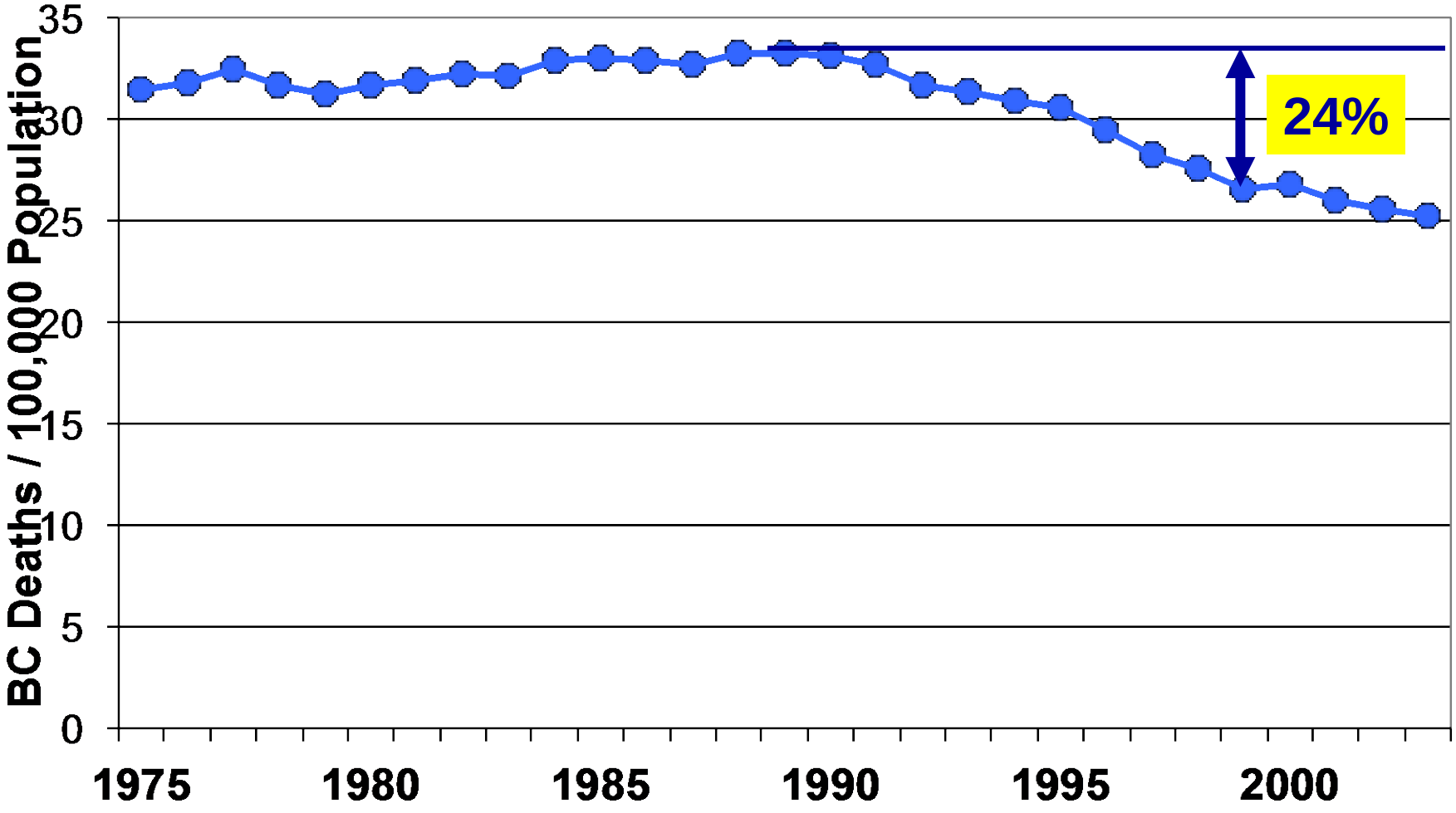
CNN: Statistical Blitz Helps Pin Down Mammography Benefits


CISNET
CANCER INTERVENTION AND
SURVEILLANCE MODELING NETWORK

Trends in Female Breast Cancer Death Rates, U.S., 1975-2006



U.S. Breast Cancer Mortality
(Age Standardized to U.S. Population in 2000)



Age-standardised mortality rates, breast cancer, females, UK, 1971-2008

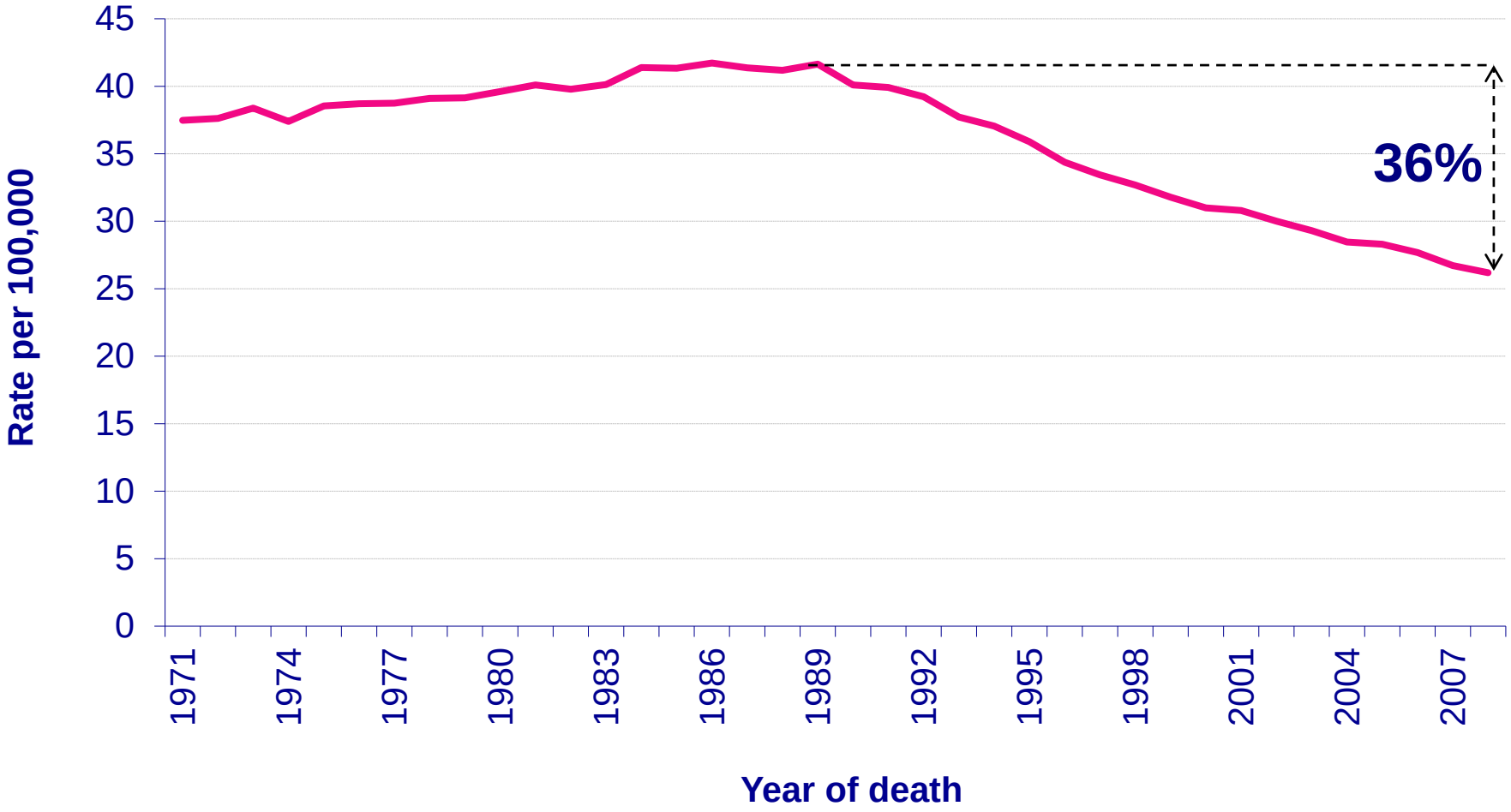
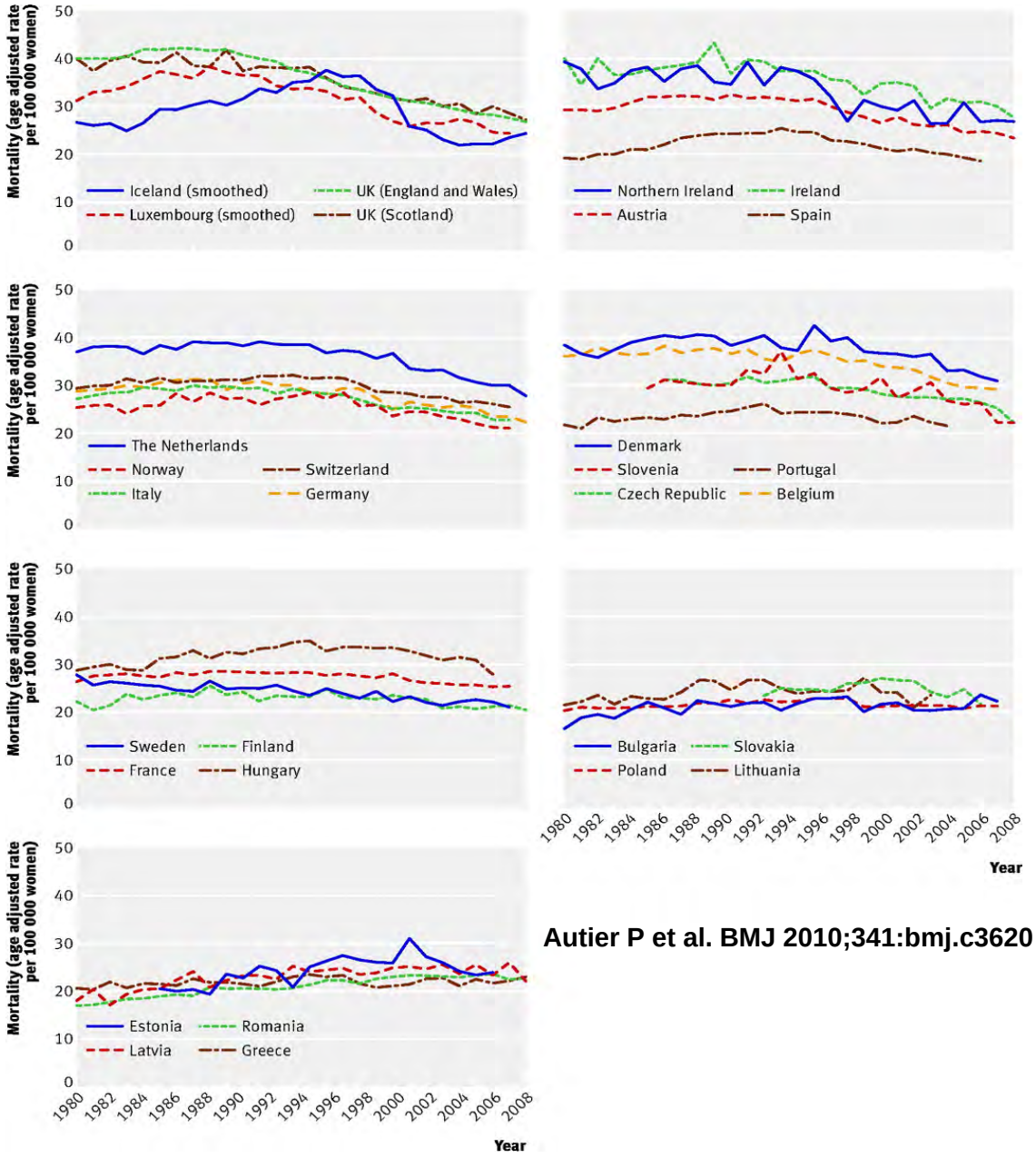


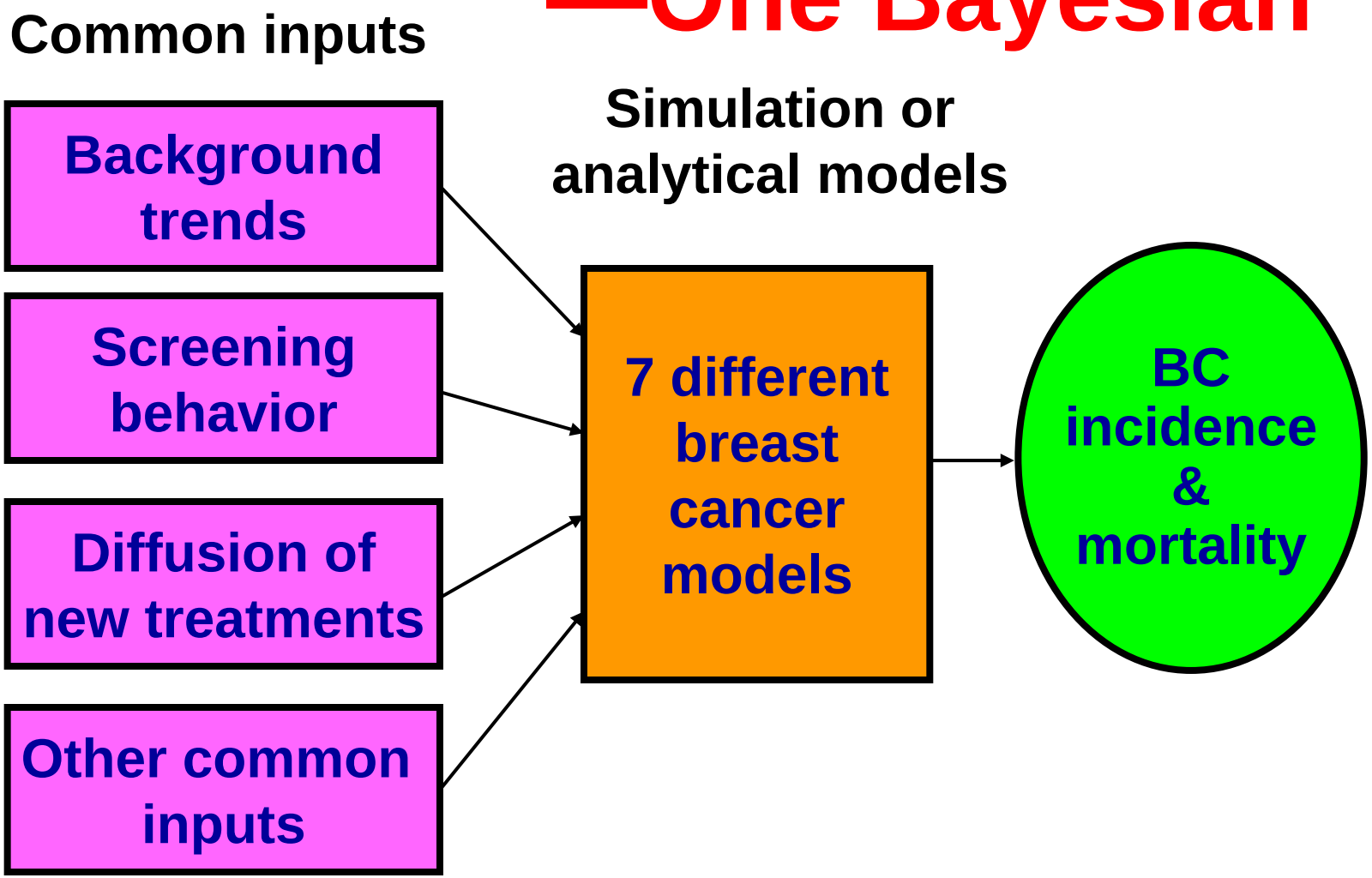
Fig 1 Temporal trends in breast cancer mortality in European countries.

1980-2008

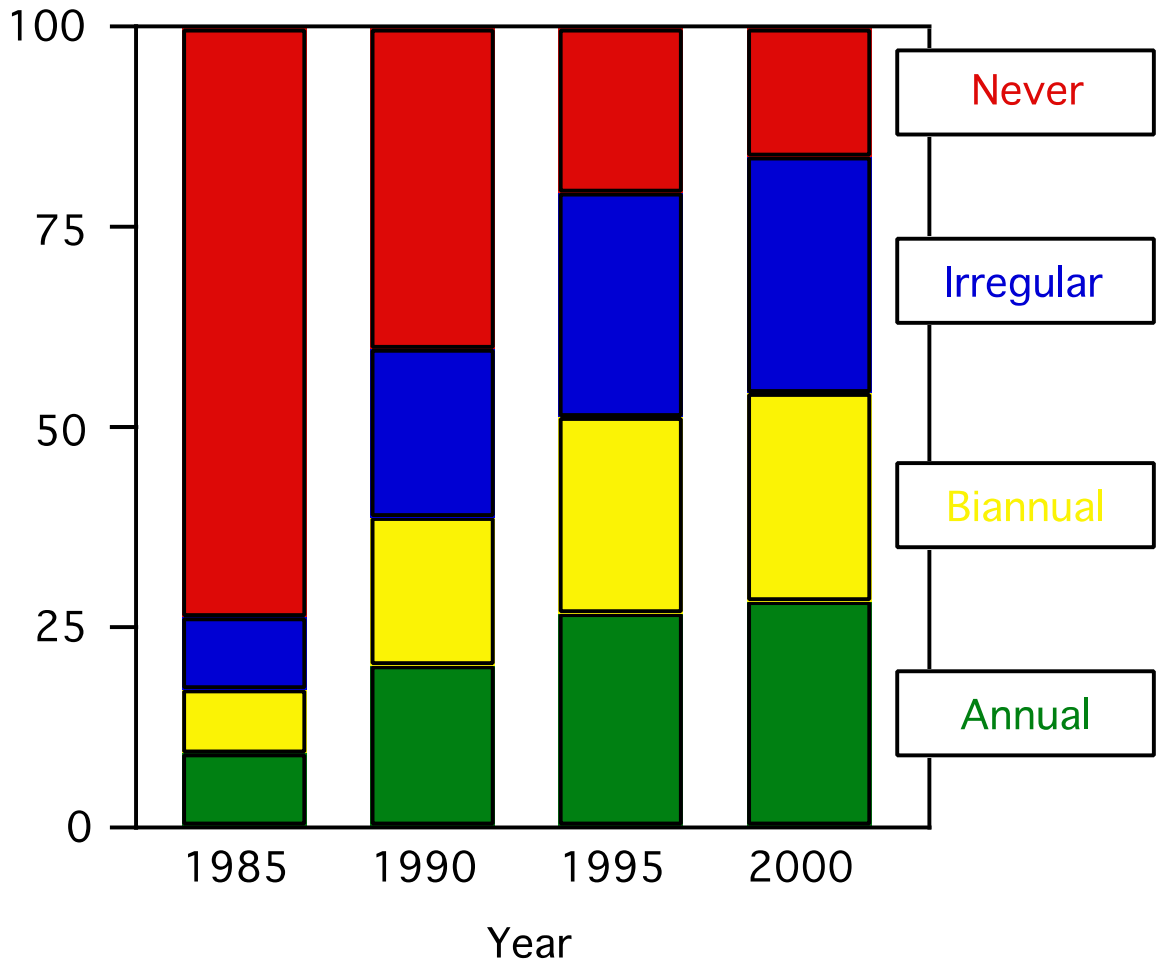


Autier P et al. BMJ 2010;341:bmj.c3620

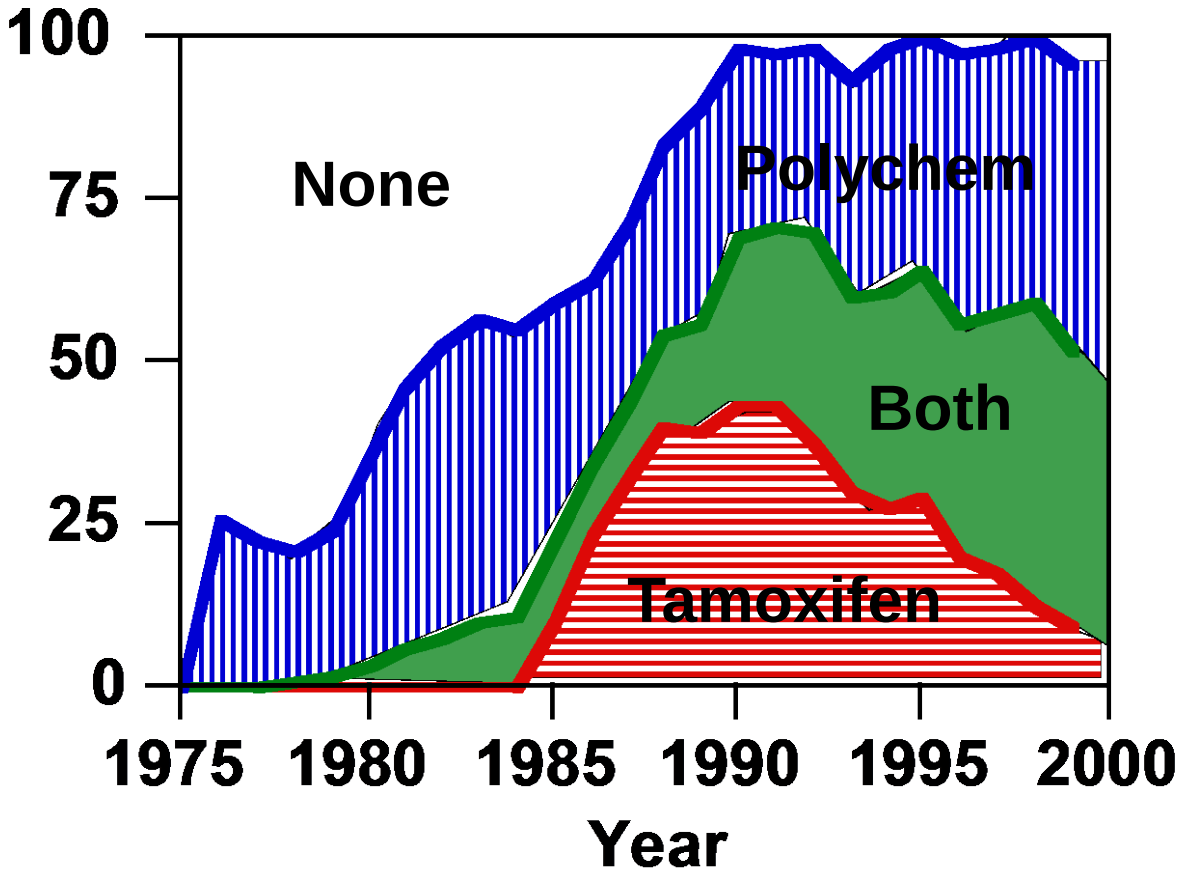
CISNET Population Models —One Bayesian



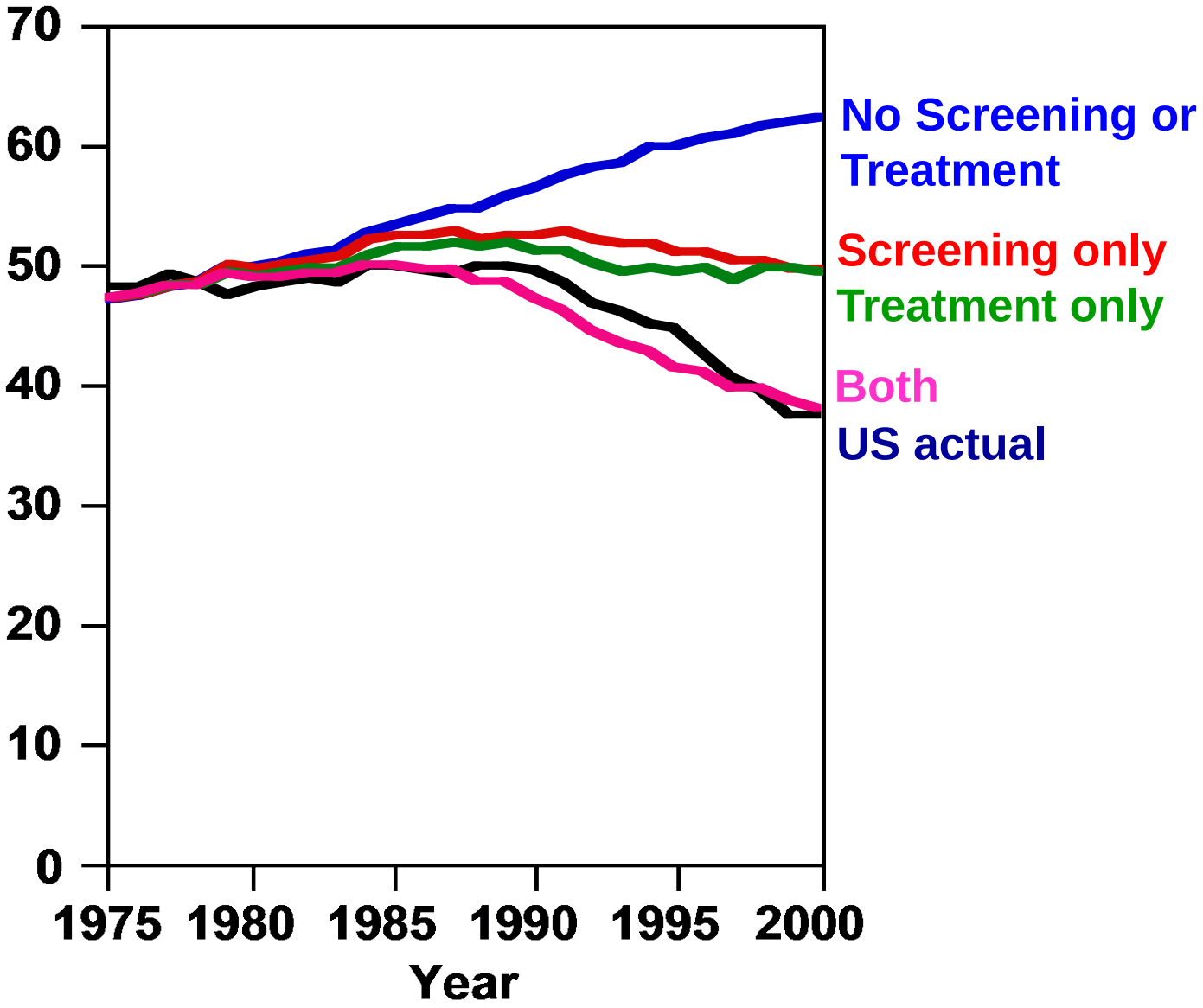
Mammography Screening Over Time, Women ages 40-79



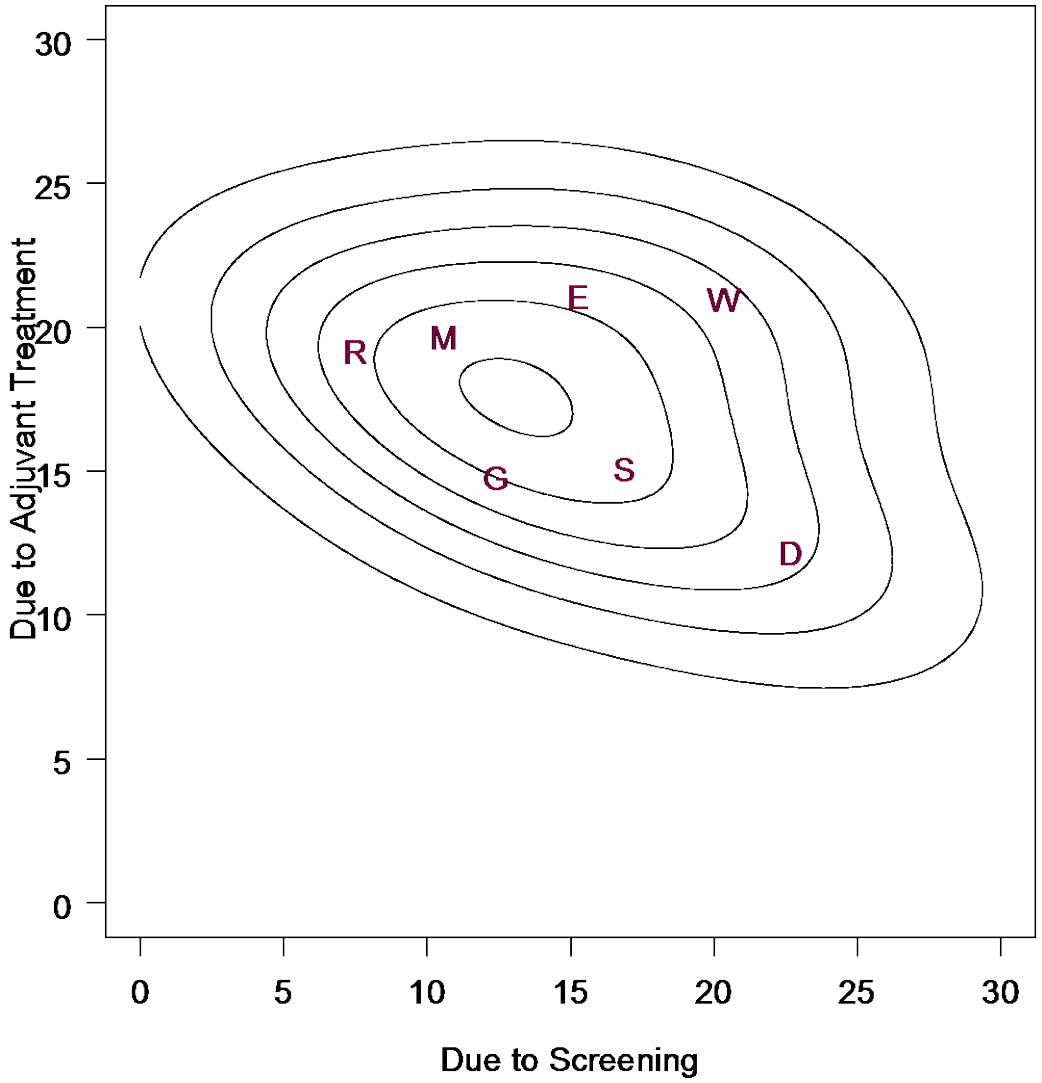
Adjuvant Therapy Dissemination Ages 50-69, node-positive



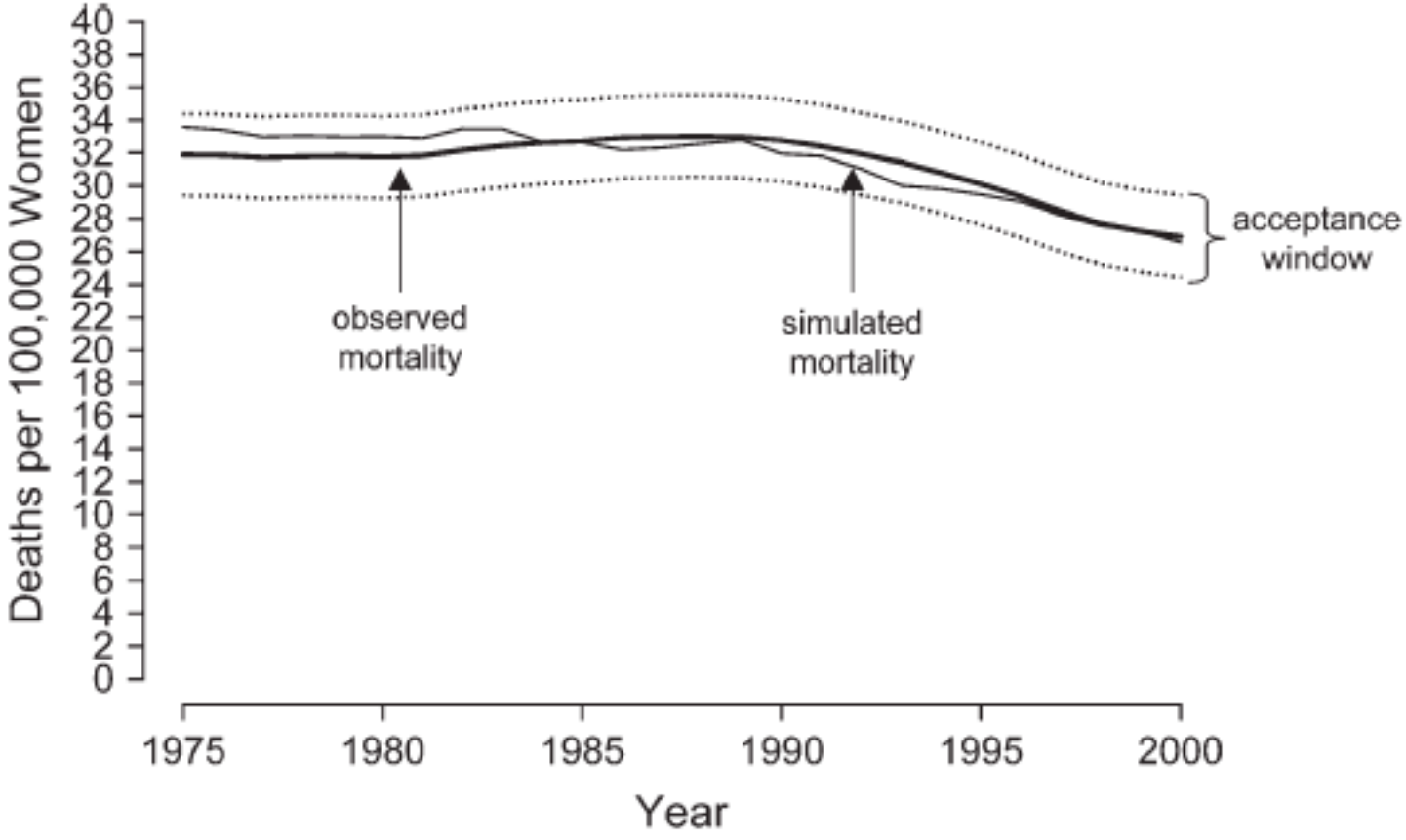
Model W's Mortality Rate per 100,000 Women 40-79 under Various Scenarios



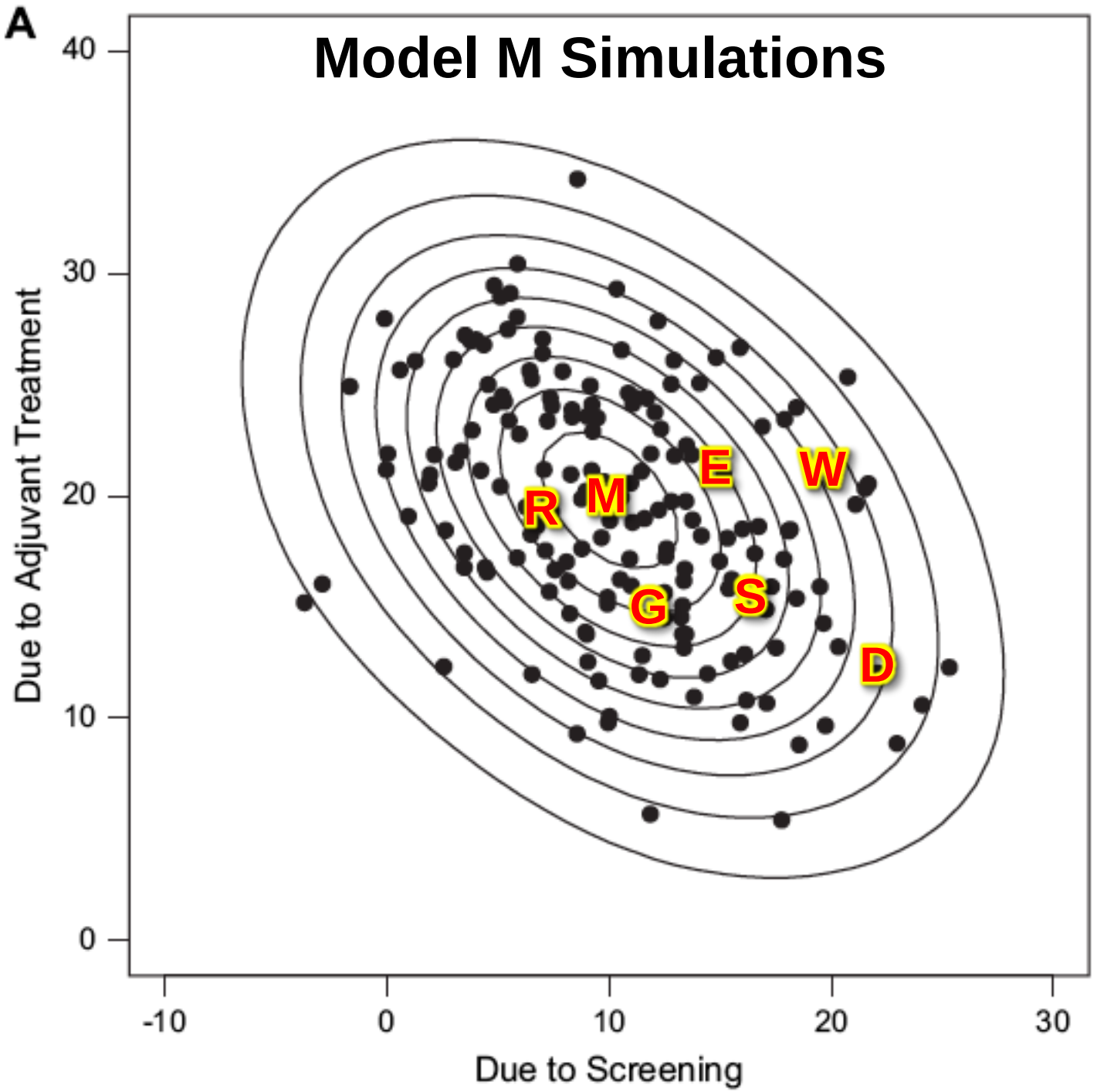
Percent reductions in BC mortality due to adjuvant Rx and screening



Bayesian Models (M)



37


A
Model M Simulations
40
30
20
10
0
Due to Adjuvant Treatment
-10
0
10
20
30
Due to Screening
R
M
E
W
G
S
D

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.

ABSTRACT

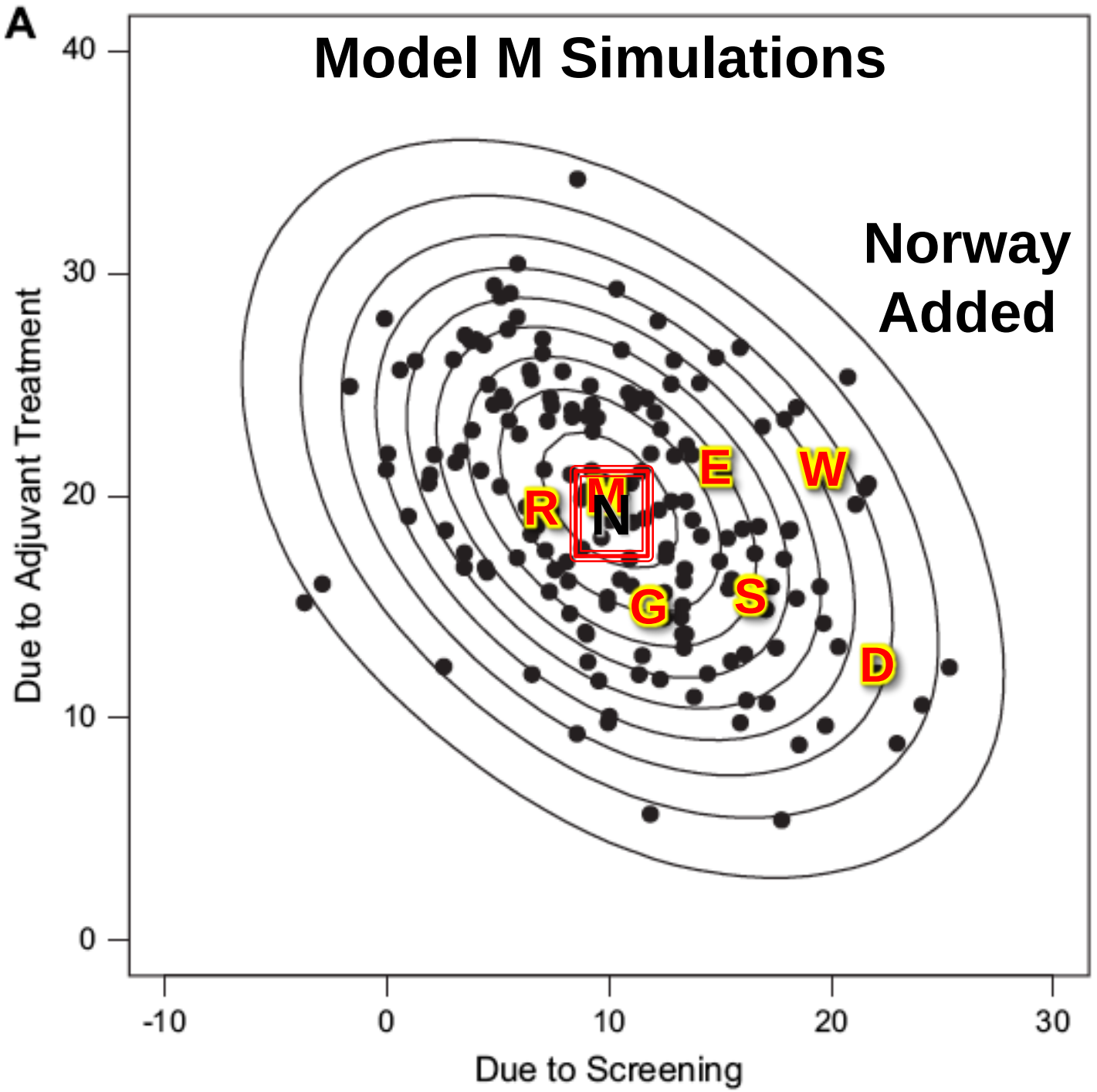
BACKGROUND

A challenge in quantifying the effect of screening mammography on breast-cancer mortality is to provide valid comparison groups. The use of historical control subjects does not take into account chronologic trends associated with advances in breast-cancer awareness and treatment.

METHODS

The Norwegian breast-cancer 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. We compared the inci-

From the Cancer Registry of Norway, Oslo (M.K., F.L., H.-O.A.); the Departments of Epidemiology (M.K., H.-O.A.) and Biostatistics (M.Z.), Harvard School of Public Health; and the Dana–Farber Cancer Institute and Harvard Medical School (M.Z., H.-O.A.) — all in Boston; and the Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm (H.-O.A.). Address reprint requests to Dr. Kalager at Oslo University Hospital, Department of Surgery, Monte-


A
Model M Simulations
Norway
Added
Due to Adjuvant Treatment
Due to Screening
40
30
20
10
0
-10
0
10
20
30
R
M
N
E
W
G
S
D

Conclusions from Model M

- Bayesian approach ideally suited for "comparative effectiveness research"
- Bayesian approach encompasses other six models
- Probability distributions of parameters are necessary for assessing predictive uncertainty

NYT Editorial

“What seems most important is that each team found at least some benefit from mammograms. The likelihood that they are beneficial seems a lot more solid today than it did four years ago, although the size of the benefit remains in dispute”

CISNET

2009 USPSTF employed modeling + RCTs

Recommendation: Repeat of 1997

Annals of Internal Medicine | Clinical Guidelines

Effects of Mammography Screening Under Different Screening Schedules

Model Estimates of Potential Benefits and Harms

Jeanne S. Mandelblatt, MD, MPH; Kathleen A. Cronin, PhD; Stephanie Bailey, PhD; Donald A. Berry, PhD; Harry J. de Koning, MD, PhD; Gerrit Draisma, PhD; Hui Huang, MS; Sandra J. Lee, DSc; Mark Munsell, MS; Sylvia K. Plevritis, PhD; Peter Ravdin, MD, PhD; Clyde B. Schechter, MD, MA; Bronislava Sigal, PhD; Michael A. Stoto, PhD; Natasha K. Stout, PhD; Nicolien T. van Ravesteyn, MSc; John Venier, MS; Marvin Zelen, PhD; and Eric J. Feuer, PhD; for the Breast Cancer Working Group of the Cancer Intervention and Surveillance Modeling Network (CISNET)*

Background: Despite trials of mammography and widespread use, optimal screening policy is controversial.

Objective: To evaluate U.S. breast cancer screening strategies.

Design: 6 models using common data elements.

Data Sources: National data on age-specific incidence, competing mortality, mammography characteristics, and treatment effects.

Target Population: A contemporary population cohort.

Time Horizon: Lifetime.

Perspective: Societal.

Interventions: 20 screening strategies with varying initiation and cessation ages applied annually or biennially.

Outcome Measures: Number of mammograms, reduction in deaths from breast cancer or life-years gained (vs. no screening), false-positive results, unnecessary biopsies, and overdiagnosis.

false-positive results. Screening biennially from ages 50 to 69 years achieved a median 16.5% (range, 15% to 23%) reduction in breast cancer deaths versus no screening. Initiating biennial screening at age 40 years (vs. 50 years) reduced mortality by an additional 3% (range, 1% to 6%), consumed more resources, and yielded more false-positive results. Biennial screening after age 69 years yielded some additional mortality reduction in all models, but overdiagnosis increased most substantially at older ages.

Results of Sensitivity Analysis: Varying test sensitivity or treatment patterns did not change conclusions.

Limitation: Results do not include morbidity from false-positive results, patient knowledge of earlier diagnosis, or unnecessary treatment.

Conclusion: Biennial screening achieves most of the benefit of annual screening with less harm. Decisions about the best strategy depend on program and individual objectives and the weight placed on benefits, harms, and resource considerations.

Primary Funding Source: National Cancer Institute.

Reactions to 2009 USPSTF

- *New York Times*: New Guidelines on Breast Cancer Draw Opposition
- *LA Times*: Mammogram guidelines spark heated debate. A government panel's recommendation that women under 50 do not need regular mammograms is attacked by oncologists, gynecologists and cancer groups
- *Boston Globe*: Breast screening advice upended
- Obama critics charge "rationing healthcare" and "death squads"

CISNET Contributions to 2009 USPSTF Recommendations

- Mortality for
  - Biennial vs annual screening
  - Benefit for women in 40s
  - Context of modern chemotherapy & hormonal therapy
- False positives/biopsies

Update to today

Cooler heads are prevailing

Look for 2014 USPSTF update: Focus of overdiagnosis

MSA Partners LLC

2006年6月30日発行 第138号

MSA Partners LLC
516 Fifth Avenue, 9th Fl.
10036
760
pr.com
http://www.msapr.com

The Coming Bayesian Tsunami of Clinical Development

目次

ベイズ統計学パート2

臨床開発にベイズ統計学の「津波」が到来

テキサス大学アンダーソン癌センター生物統計学科主任、ドナルド・A・ベリー博士

ドナルド・A・ベリー博士

30年以上もの間、ドナルド・ベリー博士は臨床試験のデザインと分析におけるベイズ統計学の利用を提唱してきた。この発想は当初、行政機関や製薬業界、および臨床研究における教育を受けておらず認識の乏しい統計学者から無視された。しかし、同氏が統計学の薬剤開発への適用に情熱を傾け続けた結果、数十年後、FDAと製薬会社が同氏に耳を傾け、ベイズ的アプローチの利点を理解するようになった。ベリー博士にこの変化について聞いた。

――薬剤開発のデザインと分析において、統計は乱用されているとお考えですか。

ベリー　統計学者を含め、臨床試験のデザ

Janet Woodcock
Dir CDER FDA (2006)

“Improved utilization of adaptive and Bayesian methods” could help resolve low success rate of and expense of phase 3 clinical trials

Current use of Bayesian designs

- MDACC (> 300 trials)
- Device companies
  (> 25 PMAs, many IDEs)
- Drug companies
  (Most of top 40;
  many biotechs)
- NIH & NCI?

Bayesian adaptive trials

- Stopping early (or late)
  - Efficacy
  - Futility
- Dose finding (& dose dropping)
- Seamless phases
- Population finding
- Adaptive randomization
- Ramping up accrual

Why?

- Smaller trials (usually!)
- More accurate conclusions
- Addresses more questions
- Better treatment of patients in trials

June 13, 2012, Vol 307, No. 22 >

Viewpoint | June 13, 2012

Adaptive Clinical Trials

A Partial Remedy for the Therapeutic Misconception?

William J. Meurer, MD, MS; Roger J. Lewis, MD, PhD; Donald A. Berry, PhD

JAMA. 2012;307(22):2377-2378. doi:10.1001/jama.2012.4174

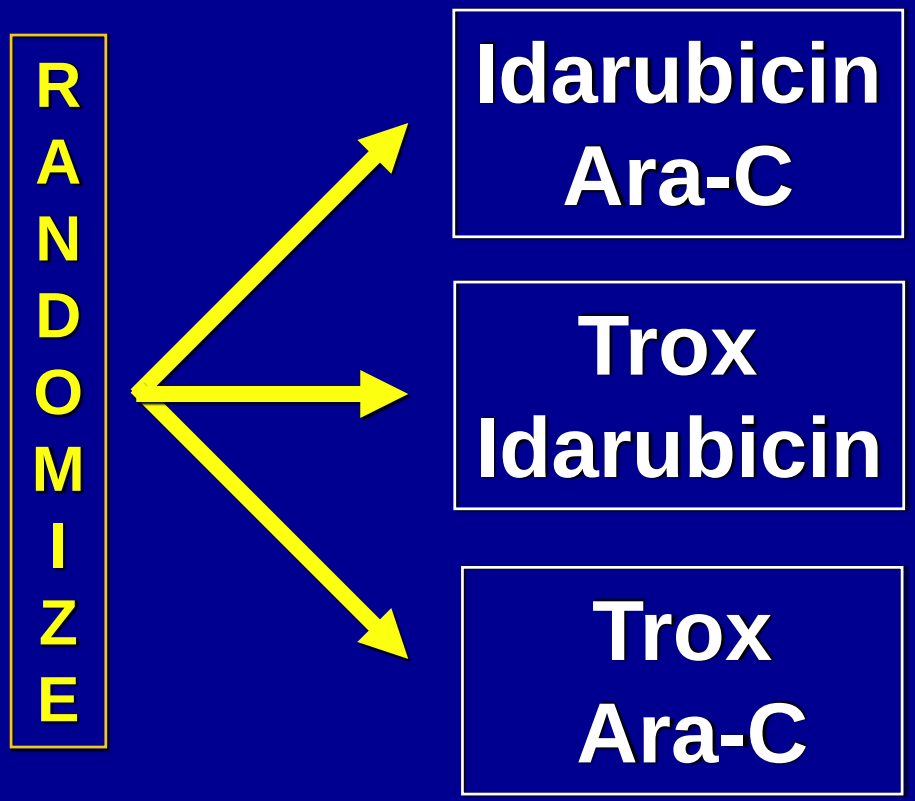
There is a common "therapeutic misconception" among patients considering participation in clinical trials.[1] Some trial participants and family members believe that the goal of a clinical trial is to improve their outcomes—a misperception often reinforced by media advertising of clinical research.[2] Clinical trials have primarily scientific aims and rarely attempt to collectively improve the outcomes of their participants. The overarching goal of most clinical trials is to evaluate the effect of a treatment on disease outcomes.[3] Comparisons are usually made with placebo for conditions having no established treatments and with standard care for conditions having effective treatments. Any benefit to an individual trial participant is a chance effect of randomization and the true, but unknown, relative effects of the treatments. Available evidence is conflicting regarding whether patients receive some benefit from simply participating in a clinical trial.[3] Thus, even though serving as a research participant is essentially an altruistic activity, many clinical trial volunteers do not participate in research out of altruism.[4] An adaptive clinical trial design can be used to increase the likelihood that study participants will benefit by being in a clinical trial.

CONTEMPORARY CLINICAL TRIALS: ADAPTIVE VS FIXED RANDOMIZATION RATIOS

Adaptive Randomization

Relation to “bandit problems”?

Example: Troxacitabine in AML



Our design

Adaptive randomization to learn, while effectively treating patients in trial

n = 25

n = 25

Adaptive Randomization

- Assign with higher probability to better performing therapies
- TI dropped after 24th patient
- Trial stopped after 34 patients

Summary of AML trial results

CR by 50 days:

IA	10/18 = 56%
TA	3/11 = 27%
TI	0/5 = 0%

Cure Magazine (2006)

“I see no rationale to further delay moving to these designs,” says Dr. Giles, who is currently involved in eight Bayesian-based leukemia studies. “They are more ethical, more patient-friendly, more conserving of resources, more statistically desirable.”

Simulations Usually Required

- To find operating characteristics:
  - Type I error rate
  - Power
  - Sample size distribution
  - Trial duration
  - Amount of drug required
- Prospective design essential
- Longitudinal modeling
- Many scenarios
- Accrual rate matters

Two Recent Bayesian Clinical Trials … with Smaller Sample Size

- Bayesian predictive probabilities
- Longitudinal modeling

ORIGINAL CONTRIBUTION

JAMA. 2010;303(4):333-340

Comparison of Antiarrhythmic Drug Therapy and Radiofrequency Catheter Ablation in Patients With Paroxysmal Atrial Fibrillation

A Randomized Controlled Trial

David J. Wilber, MD
Carlo Pappone, MD, PhD
Petr Neuzil, MD
Angelo De Paola, MD
Frank Marchlinski, MD
Andrea Natale, MD
Laurent Macle, MD
Emile G. Daoud, MD
Hugh Calkins, MD
Burr Hall, MD
Vivek Reddy, MD
Christine Y. Liu, MPH
Scott M. Berry, PhD
Donald A. Berry, PhD
for the ThermoCool AF Trial Investigators

Context Antiarrhythmic drugs are commonly used for prevention of recurrent atrial fibrillation (AF) despite inconsistent efficacy and frequent adverse effects. Catheter ablation has been proposed as an alternative treatment for paroxysmal AF.

Objective To determine the efficacy of catheter ablation compared with antiarrhythmic drug therapy (ADT) in treating symptomatic paroxysmal AF.

Design, Setting, and Participants A prospective, multicenter, randomized (2:1), unblinded, Bayesian-designed study conducted at 19 hospitals of 167 patients who did not respond to at least 1 antiarrhythmic drug and who experienced at least 3 AF episodes within 6 months before randomization. Enrollment occurred between October 25, 2004, and October 11, 2007, with the last follow-up on January 19, 2009.

Intervention Catheter ablation (n=106) or ADT (n=61), with assessment for effectiveness in a comparable 9-month follow-up period.

Main Outcome Measures Time to protocol-defined treatment failure. The pro-

Design, Setting, and Participants A prospective, multicenter, randomized (2:1), unblinded, Bayesian-designed study conducted at 19 hospitals of 167 patients who

compared with 16% of patients treated with ADT. The hazard ratio of catheter ablation to ADT was 0.30 (95% confidence interval, 0.19-0.47; P• .001). Major 30-day treatment-related adverse events occurred in 5 of 57 patients (8.8%) treated with ADT and 5 of 103 patients (4.9%) treated with catheter ablation. Mean quality of life scores improved significantly in patients treated by catheter ablation compared with ADT at 3 months; improvement was maintained during the course of the study.

Conclusion Among patients with paroxysmal AF who had not responded to at least

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 MAY 14, 2009 VOL. 360 NO. 20

Adjuvant Chemotherapy in Older Women with Early-Stage Breast Cancer

Hyman B. Muss, M.D., Donald A. Berry, Ph.D., Constance T. Cirrincione, M.S., Maria Theodoulou, M.D., Ann M. Mauer, M.D., Alice B. Kornblith, Ph.D., Ann H. Partridge, M.D., M.P.H., Lynn G. Dressler, Ph.D., Harvey J. Cohen, M.D., Heather P. Becker, Patricia A. Kartcheske, B.S., Judith D. Wheeler, M.P.H., Edith A. Perez, M.D.,

A Bayesian statistical design was used with a range in sample size from 600 to 1800 patients.

BACKGROUND

Older women with breast cancer are underrepresented in clinical trials, and data on the effects of adjuvant chemotherapy in such patients are scant. We tested for the noninferiority of capecitabine as compared with standard chemotherapy in women with breast cancer who were 65 years of age or older.

METHODS

We randomly assigned patients with stage I, II, IIIA, or IIIB breast cancer to standard chemotherapy (either cyclophosphamide, methotrexate, and fluorouracil or cyclophosphamide plus doxorubicin) or capecitabine. Endocrine therapy was recommended after chemotherapy in patients with hormone-receptor–positive tumors. A Bayesian statistical design was used with a range in sample size from 600 to 1800 patients. The primary end point was relapse-free survival.

From the University of Vermont, Burlington (H.B.M.); the M.D. Anderson Cancer Center, Houston (D.A.B.); the Cancer and Leukemia Group B (CALGB) Statistical Center, Duke University Medical Center (C.T.C., P.A.K.) and Duke University Medical Center (H.J.C., J.D.W., A.A.M.) — both in Durham, NC; Memorial Sloan-Kettering Cancer Center, New York (M.T., L.N., C.A.H.); CALGB, Chicago (A.M.M., H.P.B.); the Dana–Farber Cancer Institute, Boston (A.B.K., A.H.P., H.J.B., E.P.W.); the University of North Carolina, Chapel Hill (L.G.D.); the North Central Cancer Treat-

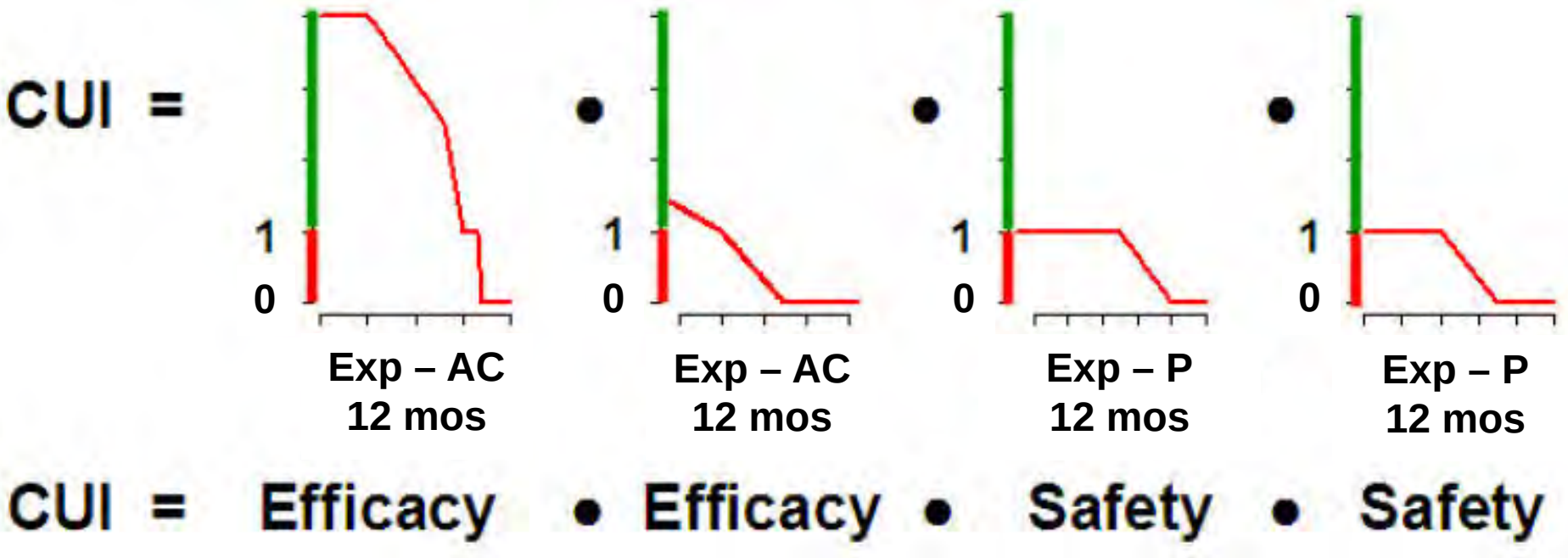
Example from Critical Path Initiative

- Type II diabetes
- Seamless Phase II/III: Dose finding via plus confirmation
- Active comparator & placebo
- Primary endpoint: Clinical Utility Index (12 months)

Some Details

- Phase II: 7 doses experimental drug, adaptively randomized
- Phase III
  - 1 or 2 doses experimental drug
  - Sample size via predictive power considering available Phase II data
  - Adaptive transition: Bayesian predictive probs
- Both phases driven by CUI
- Role of longitudinal modeling

Clinical Utility Index



- Dose-response modeling
- Longitudinal modeling

October 22, 2012

Lilly Diabetes Announces Positive Results of Phase III Trials of Dulaglutide in Type 2 Diabetes

Company Shares Top-line Results on 3 Completed AWARD Trials

INDIANAPOLIS, Oct. 22, 2012 /PRNewswire/ -- Eli Lilly and Company (NYSE: LLY) today announced positive top-line results of three completed Phase III AWARD trials for dulaglutide, an investigational, long-acting glucagon-like peptide 1 (GLP-1) analog being studied as a once-weekly treatment for type 2 diabetes. Primary efficacy endpoints, as measured by reduction in hemoglobin A1c (HbA1c) at the 1.5 mg dose, were met in three studies (AWARD-1, AWARD-3 and AWARD-5). Having met the primary endpoints, superiority for HbA1c lowering was examined, and both doses of dulaglutide (0.75mg and 1.5mg) demonstrated statistically superior reduction in HbA1c from baseline compared to: exenatide twice-daily injection at 26 weeks (AWARD-1); metformin at 26 weeks (AWARD-3); and sitagliptin at 52 weeks (AWARD-5).

When the I-SPY 2 trial launched in 2010, oncologists heralded it as the future of cancer research. Five pharmaceutical companies put aside their differences to participate in the landmark phase 2 breast cancer trial, which adaptively and efficiently randomized patients to one of seven experimental therapies. Now, even as I-SPY 2 propels its first two drugs into phase 3 trials, researchers in other areas of medicine are catching on to the benefits of this collaborative approach. On 11 December, Europe's Innovative Medicines Initiative (IMI) announced a €53 million call for proposals for a similarly designed trial in Alzheimer's disease. Already, at least 12 drug companies are keen to participate.

Fanatic Studio / Alamy

Across the divide: Drug companies are teaming up to run trials in many disease areas.

Two of BC's federal grants:

- PCORI: Adaptive Design Guidance for Clinical Trials
- FDA/NIH Regulatory Science

ScienceInsider

Breaking news and analysis from the world of science policy

FDA's $25 Million Pitch for Improving Drug Regulation

by Jennifer Couzin-Frankel on 7 October 2010, 3:17 PM | Permanent Link | 0 Comments

"Our current approach [to trials] is horribly inefficient, and we need to do something better," says Roger Lewis, an emergency medicine physician at Harbor-University of California, Los Angeles, Medical Center. Lewis helps advise a company called Berry Consultants …

FDA is trying to move forward nevertheless, in part by linking up with more flush agencies. Last week, in conjunction with the National Institutes of Health (NIH), it announced four sizable grants, totaling $9.4 million, in regulatory science. (FDA contributed just under $1 million and NIH gave the rest.) They include support for a heart-lung system that can test potential drugs and an effort to dramatically streamline clinical trials.

"Our current approach [to trials] is horribly inefficient, and we need to do something better," says Roger Lewis, an emergency medicine physician at Harbor-University of California, Los Angeles, Medical Center. Lewis helps advise a company called Berry Consultants founded by Donald Berry, a biostatistician at M.D. Anderson Cancer Center in Houston, Texas. He and Berry, along with emergency medicine physician William Barsan at the

ScienceInsider
Breaking news and analysis from the world of science policy

FDA's $25 Million Pitch for Improving Drug Regulation

by Jennifer Couzin-Frankel on 7 October 2010, 3:17 PM | Permanent Link | 0 Comments

Lewis and Berry, along with emergency medicine physician William Barsan at the University of Michigan, will be studying whether "adaptive" trial designs that incorporate new information in midcourse can answer medical questions. They also want to learn what concerns researchers might have about this approach.

conjunction with the National Institutes of Health (NIH), it announced four sizable grants, totaling $9.4 million, in regulatory science. (FDA contributed just under $1 million and NIH gave the rest.) They include support for a heart-lung system that can test potential drugs and an effort to dramatically streamline clinical trials.

"Our current approach [to trials] is horribly inefficient, and we need to do something better," says Roger Lewis, an emergency medicine physician at Harbor-University of California, Los Angeles, Medical Center. Lewis helps advise a company called Berry Consultants founded by Donald Berry, a biostatistician at M.D. Anderson Cancer Center in Houston, Texas. He and Berry, along with emergency medicine physician William Barsan at the

I-SPY2

http://clinicaltrials.gov/ct2/show/NCT01042379?term=I-SPY2&rank=1

The NEW ENGLAND JOURNAL of MEDICINE

Perspective

Development of Novel Combination Therapies

Janet Woodcock, M.D., Joseph P. Griffin, J.D., and Rachel E. Behrman, M.D., M.P.H.

For example, in 2010, the Biomarkers Consortium–a public-private partnership that includes the NIH, the FDA, patient groups, and pharmaceutical and biotech–initiated a groundbreaking trial in breast cancer to predict drug responsiveness based on the presence or absence of genetic and biological markers, … I-SPY 2 (ClinicalTrials.gov NCT01042379).

This article (10.1056/NEJMp1101548) was published on February 16, 2011, at NEJM.org.

THE WALL STREET JOURNAL.
WSJ.com

THE SATURDAY ESSAY | OCT

A New Rx for Med

Fed up with slow drug trials, ca… back to personalized treatments.

By RON WINSLOW

New trial design

Uses genetic profiles to highlight 'biomarker' differences among patients and to match drugs to patients with biomarkers that predict a benefit.

PERSONALIZED MEDICINE | How … rug development

1 cube = 10 patients

Traditional clinical trial

Takes essentially all patients with a disease being studied and is typically intended to eliminate differences in patient characteristics that could bias measures of drug effectiveness.

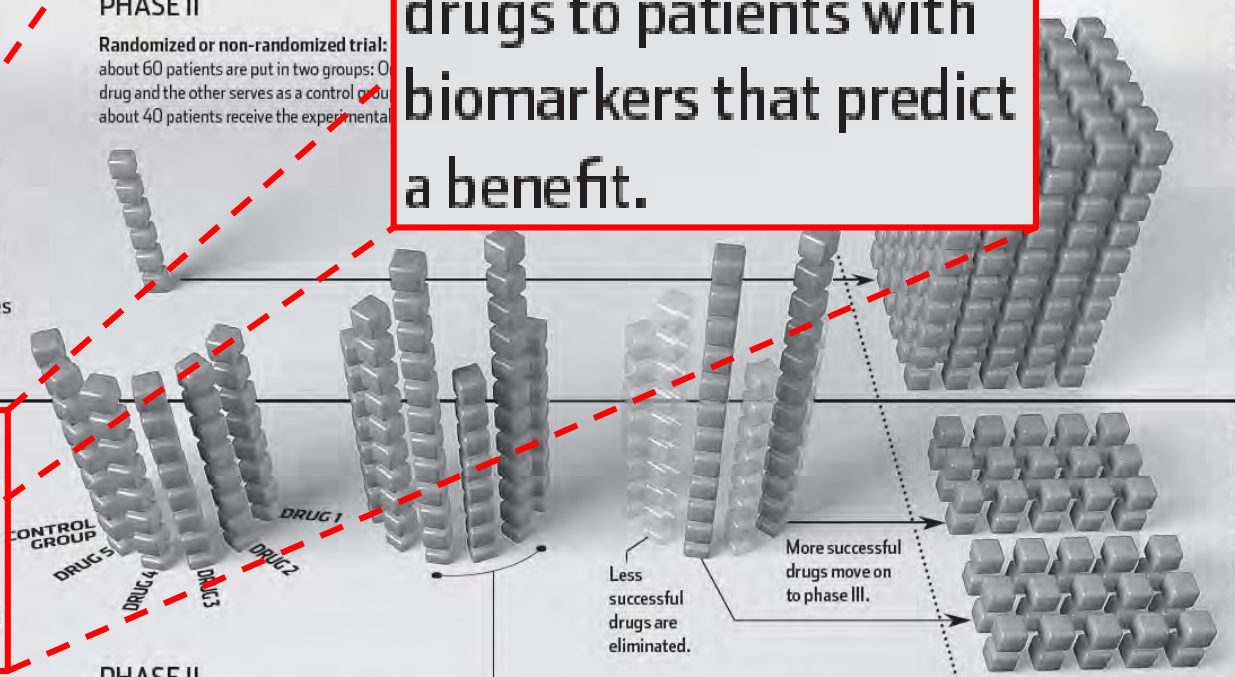
PHASE II

Randomized or non-randomized trial: about 60 patients are put in two groups: O… drug and the other serves as a control g… about 40 patients receive the experimental…

PHASE III

If a drug graduates to phase III, it typically takes **3,000 patients** and about three years to determine if it is safe and effective enough for approval.

HISTORIC SUCCESS RATE
30 TO 40%



New trial design

Uses genetic profiles to highlight 'biomarker' differences among patients and to match drugs to patients with biomarkers that predict a benefit.

PHASE II

Patients are placed in groups based on genetic profiles and are **randomly assigned to either standard therapy or one of five different drugs** plus standard care.

Early results increase chances that **patients entering the trial later will be assigned to a drug showing benefit** against tumors with their genetic profile.

It will take up to 120 patients for each drug to determine which ones graduate to phase III studies.

PHASE III

Researchers expect that drugs graduating from I-Spy 2 to phase III can be tested with **300 patients** selected according to genetic profiles found to respond to the drug in phase II. It is hoped that this will shorten the time to approval.

PROBABILTY OF SUCCESS
85%

Note: In all clinical trials, phase I consists of testing on human subjects to determine toxicity levels.

Graphic by Maryanne Murray/WSJ

Source: Donald Berry, M.D. Anderson Cancer Center

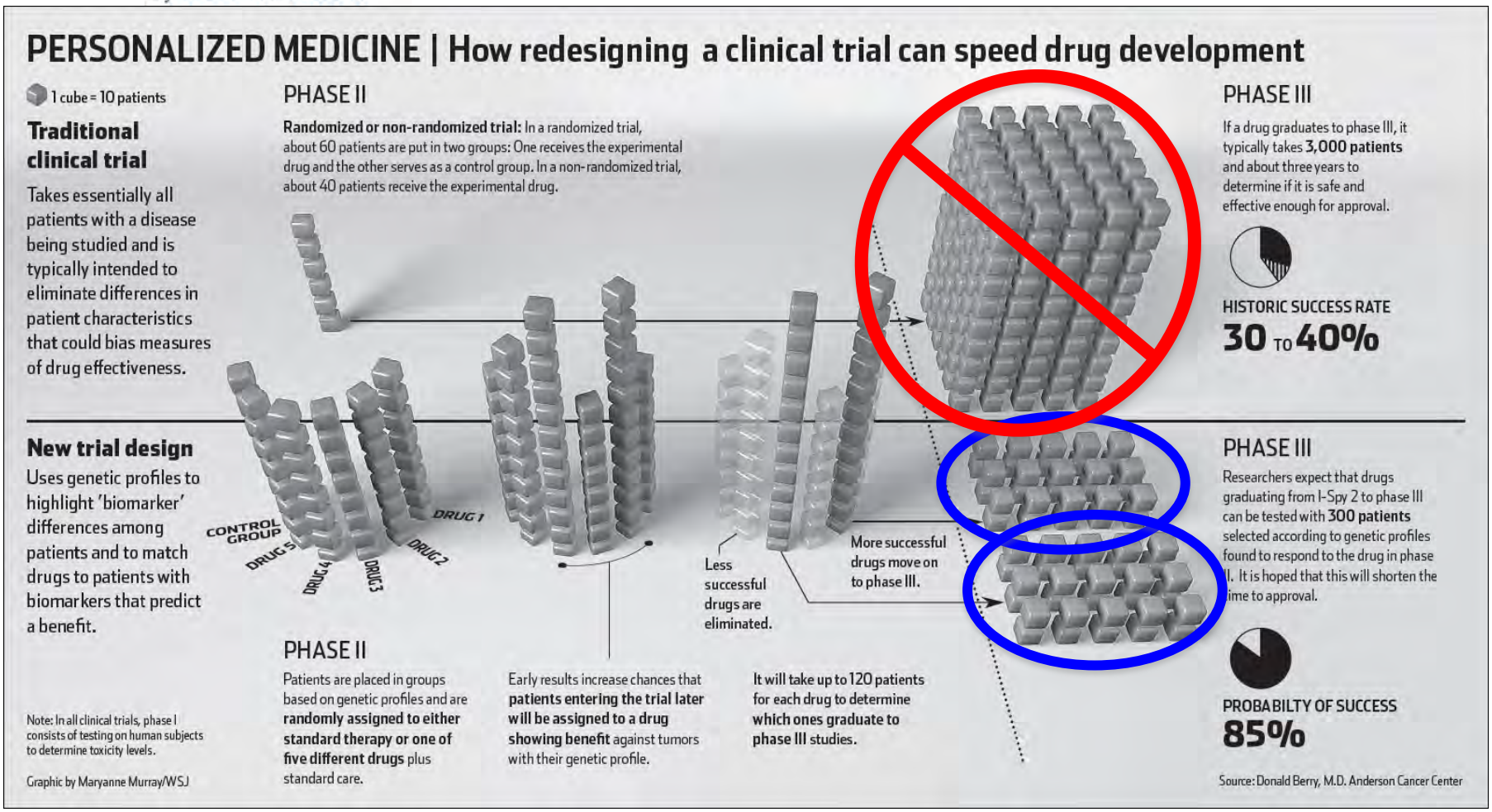
THE WALL STREET JOURNAL.
WSJ.com

THE SATURDAY ESSAY | OCTOBER 2, 2010

A New Rx for Medicine

Fed up with slow drug trials, cancer patients and doctors are testing a fast track to personalized treatments.

By RON WINSLOW



Effects of I-SPY Approaches

- Match drugs with biomarker signatures
- Savings from common control
- Better therapies move thru faster
- Successful drug/biomarker pairs graduate to small, focused, more successful Phase 3 based on Bayesian predictive probabilities
- Offspring of I-SPY 2: melanoma, GBM, Alzheimer's, HIV, acute heart failure, SARI/H1N1, …

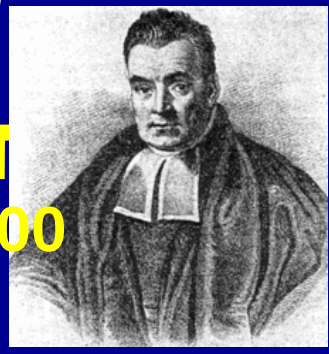
I-SPY 2 is a prototypic "platform trial"

- 5 pharmaceutical companies, going on 6
- 7 experimental agents (2 graduated), going on 8


Merck
TRONCS
AZ
gsk
GlaxoSmithKline

Biased History of Biostatistics

1997: FDAMA & Bayesian Credibility



1800

1900

1940s

2000

1933

Medicine Dominated by Case Study & Anecdote

W.R. Thompson & Bayes

Bradford Hill & RCT

Frequentist Half Cent— No Bayes