Slowly but Surely, Bayesian Ideas Revolutionize Medical Research

Donald A. Berry
dberry@mdanderson.org
Rough History of Biostatistics

1800: Medicine Dominated by Case Study & Anecdote

W.R. Thompson & Bayes

1900: Bradford Hill

1933: Frequentist
Half Cent—No Bayes

1940s: Bradford Hill & RCT
Rough History of Biostatistics

1800: Medicine Dominated by Case Study & Anecdote

W.R. Thompson & Bayes

1900

Bradford Hill & RCT

1940s: Frequentist

Half Cent—No Bayes

1997: FDAMA & Bayes Credibility
My exposures

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

- 1970: Bandit Problems
- 1980: 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
- Ancestry of corn, soybeans, cattle, cabernet sauvignon, etc.
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


**Purpose:** To compare genetic test results for deleterious mutations of BRCA1 and BRCA2 with estimated probabilities of carrying such mutations; to assess sensitivity of genetic testing; and to assess the relevance of other susceptibility genes in familial breast and ovarian cancer.

**Patients and Methods:** Data analyzed were from six high-risk genetic counseling clinics and concern individuals from families for which at least one member was tested for mutations at BRCA1 and BRCA2. Predictions of genetic predisposition to breast and ovarian cancer 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 and 29 for BRCA2 mutations. BRCAPRO performed well: for the 150 probands with the smallest BRCAPRO carrier probabilities (average, 29.0%), the proportion testing positive was 32.7%; for the 151 probands with the largest carrier probabilities (average, 95.2%), 78.8% tested positive. Genetic testing sensitivity was estimated to be at least 85%, with false-negatives including mutations of susceptibility genes heretofore unknown.

**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.


**INDIVIDUALS WITH a family history of breast and/or ovarian cancer are at increased risk of carrying deleterious and second-degree relatives. (We use “carrier probability” to mean the probability of carrying a deleterious mutation of...**
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.
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,
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 faced...
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-
The problem with multiples

Landis seemed to have an unusual test result. Because he was among the leaders he provided 8 pairs of urine samples (of the total of approximately 126 sample-pairs in the 2006 Tour de France). So there were 8 opportunities for a true positive — and 8 opportunities for a false positive. If he never doped and assuming a specificity of 95%, the probability of all 8 samples being labelled ‘negative’ is the eighth power of 0.95, or 0.66. Therefore, Landis’s false-positive rate for the race as a whole would be about 34%. Even a very high specificity of 99% would mean a false-positive rate of about 8%. The single-test specificity would have to be increased to much greater than 99% to have an acceptable false-positive rate. But we don’t know the single-test specificity because the appropriate studies have not been performed or published.
Assessing Probability of Ancestry Using Simple Sequence Repeat Profiles: Applications to Maize Hybrids and Inbreds

Donald A. Berry,*† 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.


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 is subtracted from that of the offspring and the deduced...
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.
"CONCLUSIONS: 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
Mortality reduction = 18%

Fig. 1, Berry JNCI 1998
Fig. 3, Berry JNCI 1998

Shaded area is 1.4 days per woman
Fig. 3, Berry JNCI 1998

Shaded area is 1.4 days per woman

Total about 5 days
“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 NCI’s budget until NCI agreed to recommend screening for women in their 40s
Berry JNCI 1998 ... with 2009 updates

Mortality reduction = 18%

Bayes estimates
Berry JNCI 1998 ... with 2009 updates

**A**

Screening rate = control rate

Mortality reduction = 18%

**B**

Percent reduction in breast cancer mortality due to screening

Updates
2009 USPSTF employed modeling + RCTs

Recommendation: Repeat of 1997
2009 USPSTF employed modeling + RCTs

Recommendation:
Repeat of 1997
Reactions to 2009 USPSTF

- **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”
CANCER INTERVENTION AND SURVEILLANCE MODELING NETWORK
Trends in Female Breast Cancer Death Rates* by Race and Ethnicity, US,
Age-standardised (European) mortality rates, breast cancer, females, UK, 1971-2008
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*

From M.D. Anderson Cancer Center, Houston (D.A.B.); the National Cancer Institute, Bethesda, Md. (K.A.C., E.J.F.); Stanford University, Stanford, Calif. (S.K.P.); the University of Wisconsin–Madison, Madison (D.G.F.); Cornerstone Systems, Lynden, Wash. (L.C.); Dana–Farber Cancer Institute, Boston (M.Z.); Georgetown University, Washington, D.C. (J.S.M.); the University of Rochester, Rochester, N.Y. (A.Y.Y.); and Erasmus University Medical Center, 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.

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 mortality in the United States from 1975 to 2000.

METHODS
A consortium of investigators developed seven independent statistical models of breast-cancer incidence and mortality. All seven groups used the same sources to obtain data on the use of screening mammography, adjuvant treatment, and benefits 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).
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. Habbenema, 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 mortality and 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).
“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 from NEJM

Women 40-79

Node-positive BC
Percent reductions in BC mortality due to adjuvant Rx and screening
Bayesian Models (M)

The graph shows the trend of deaths per 100,000 women from 1975 to 2000. It includes observed and simulated mortality rates. The acceptance window is indicated on the right side of the graph.
% Reduction in Breast Cancer Mortality

Due to Screening vs. Due to Treatment

Models M
% Reduction in Breast Cancer Mortality

Due to Screening

Due to Treatment

Models M

2010 Study from Norway
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
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

30年以上の間、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案してきた。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、数少な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、数年後、FDAが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、臨床試験のデザイン...
The Coming Bayesian Tsunami of Clinical Development
“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)
Some areas of application of Bayesian adaptive drug trials

- Oncology
- Migraine
- Rheum Arthritis
- Lupus
- Sepsis
- Diabetes
- Obesity
- Stroke
- Acute heart failure

- Spinal Cord Injury
- HIV
- Hepatitis C
- Pre-term labor
- Constipation
- Micturition
- Alzheimer’s
- Parkinson’s
- Pandemic flu (H1N1)
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
Adaptive Clinical Trials
A Partial Remedy for the Therapeutic Misconception?

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


There is a common “therapeutic misconception” among patients considering participation in clinical trials. 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. 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. 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. 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. 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

Standard design

- Idarubicin
  - Ara-C
  - n = 25
- Trox
  - Idarubicin
  - n = 25
- Trox
  - Ara-C
  - n = 25
Example: Troxacitabine in AML

Our design

Adaptive randomization to learn, while effectively treating patients in trial
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:**

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>IA</td>
<td>10/18 = 56%</td>
<td></td>
</tr>
<tr>
<td>TA</td>
<td>3/11 = 27%</td>
<td></td>
</tr>
<tr>
<td>TI</td>
<td>0/5 = 0%</td>
<td></td>
</tr>
</tbody>
</table>
“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
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
Giuseppe Augello, MD
Matthew R. Reynolds, MD, MSc
Chandan Vinekar, MS
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 proportion of patients who experienced major treatment-related adverse events within 30 days of catheter ablation or ADT was also reported.

Results At the end of the 9-month effectiveness evaluation period, 66% of patients in the catheter ablation group remained free from protocol-defined treatment failure 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 1 antiarrhythmic drug, catheter ablation was more effective than ADT at reducing the cumulative risk of treatment failure during 9 months of follow-up.
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

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 protocol-defined endpoint in the catheter ablation group remained free from protocol-defined treatment failure 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 1 antiarrhythmic drug, catheter ablation was superior to ADT at reducing the time to protocol-defined treatment failure.
Adjuvant Chemotherapy in Older Women with Early-Stage Breast Cancer


ABSTRACT

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 Treatment Group, Chicago (M.N.E., E.A. P.).
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.
Two of several BC federal grants:

- PCORI: Adaptive Design Guidance for Clinical Trials
- FDA/NIH Regulatory Science
FDA's $25 Million Pitch for Improving Drug Regulation

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

The U.S. Food and Drug Administration (FDA) is pressing for a big funding boost for "regulatory science"—research that can help it evaluate new treatments better and faster. Yesterday, FDA chief Margaret Hamburg laid out her case for regulatory research at the National Press Club while FDA released a report on the subject. The agency wants to devote $25 million next year to regulatory science, a small slice of the $4 billion President Barack Obama's Administration has requested for the agency in 2011. Congress has not yet approved that request.

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 University of Colorado, are working on a computer program that could help speed up the testing of new drugs.
"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 University of Pennsylvania, learned the hard way when they tried to recruit volunteers for a clinical trial of a promising brain IPC drug. They found it impossible to get the patients they needed, ultimately forcing them to abandon the effort.
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.

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 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.
I-SPY2

http://clinicaltrials.gov/ct2/show/NCT01042379?term=I-SPY2&rank=1
Development of Novel Combination Therapies
Janet Woodcock, M.D., Joseph P. Griffin, J.D., and Rachel E. Behrman, M.D., M.P.H.

Innovative drug development requires science and regulation to advance in concert. Nowhere is this need more apparent or urgent than in the development of combination therapies. Advances in genomics and cell biology have increased the opportunity for rational design of targeted drugs to inhibit the function of specific molecules, including those contributing to the proliferation of cancer cells and microorganisms. Although targeted drugs added to a standard regimen is compared with the standard regimen alone.

Successful development of future targeted therapies will require modernizing this paradigm to provide the flexibility needed to rapidly evaluate combination regimens involving new targeted agents in a single development program. Increasingly, tumors will be screened for pertinent pathway dependencies, as is currently done for breast cancer, and patients

This article (10.1056/NEJMp1101548) was published on February 16, 2011, at NEJM.org.
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).
Driving Biomedical Innovation:
Initiatives to Improve Products for Patients
To respond to these challenges, FDA will hold a series of scientific meetings with academic investigators, patient groups, drug developers, statistical and methodological experts, and ethicists to achieve a common understanding of steps that can be taken when an investigational drug being studied for a serious disease with no acceptable treatment option shows exceptional promise. CDER will then publish a draft guidance on an expedited development pathway based on the outcome of these meetings.

FDA is also working on two more immediate and related steps toward expedited drug development. First, the Agency is developing a draft guidance on enrichment strategies in clinical drug development. This is a major step forward for speeding progress for targeted therapies and will lay out many strategies for selecting the patients most likely to benefit from a particular drug. These enrichment strategies are expected to improve the efficiency of clinical trials and serve as a source of expedited drug development.

Second, as a working example of an expedited pathway, CDER will publish a draft guidance on the use of pathologic complete response (pCR)—when no clinical evidence of a disease remains—as a surrogate endpoint for accelerated approval in primary high-risk breast cancer. This guidance will outline a relatively seamless pathway that could be followed from a multi-drug screening trial such as I-SPY 2 to an accelerated approval. This would speed the availability of targeted therapies for breast cancer.

U.S. Food and Drug Administration / Driving Innovation
I-SPY 2

In March 2010, the I-SPY 2 Trial was launched. This is a groundbreaking clinical trial model that will help quickly and efficiently test promising drugs in development for women with high-risk, rapidly growing breast cancers. During the trial, drugs are individually targeted to the biology of each woman’s tumor. By applying an innovative trial design, researchers then use data from one set of patients’ treatments to decide treatment for future women who join the trial. This will help the researchers learn more quickly which investigational drugs will be most beneficial for women with certain biomarkers.
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 WINSWLOW

PERSONALIZED MEDICINE | How redesigning a clinical trial can speed drug development

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.

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
Randomized or non-randomized trial: In a randomized trial, about 60 patients are put in two groups: One receives the experimental drug and the other serves as a control group. In a non-randomized trial, about 40 patients receive the experimental drug.

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%

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.

Probability of success: 85%

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

Graphic by Maryanne Manay/WSJ

Source: Donald Berry, M.D., Anderson Cancer Center
New trial design

Uses genetic profiles to highlight ‘biomarker’ differences among patients and to match drugs to patients with biomarkers that predict a benefit.
Fed up with slow drug trials, cancer patients and doctors are testing a fast track to personalized treatments.

By RON WINSLOW

PERSONALIZED MEDICINE | How redesigning a clinical trial can speed drug development

PHASE II
Randomized or non-randomized trial: In a randomized trial, about 60 patients are put in two groups. One receives the experimental drug and the other serves as a control group. In a non-randomized trial, about 40 patients receive the experimental drug.

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
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%

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 I. It is hoped that this will shorten the time to approval.

PROBABILITY OF SUCCESS
85%

Source: Donald Berry, M.D. [Andersson Cancer Center]
I-SPY2: The Cartoon

*<http://ispy2.org>
Standard Phase 2 Cancer Drug Trials

Population of patients

Outcome: Tumor shrinkage?

Experimental arm

Population of patients

Outcome: Longer time disease free

Experimental arm

Randomize

Standard therapy
Consequence: 34% Success of Phase 3 Trials
I-SPY2 TRIAL

Population of patients

Randomly

Experimental arm 1
Experimental arm 2
Experimental arm 3
Experimental arm 4
Experimental arm 5
Standard therapy

Outcome: Complete response at surgery
I-SPY2 TRIAL

Arm 2 graduates to small focused Phase 3 trial

Outcome: Complete response at surgery
I-SPY2 TRIAL

Population of patients

Adaptively Randomize

Experimental arm 1
Experimental arm 3
Experimental arm 4
Experimental arm 5
Standard therapy

Arm 3 drops for futility

Outcome: Complete response at surgery
I-SPY2 TRIAL

Population of patients

Randomly

Arm 3 drops for futility

Experimental arm 1
Experimental arm 4
Experimental arm 5
Standard therapy

Outcome: Complete response at surgery
I-SPY2 TRIAL

Arm 5 graduates to small focused Phase 3 trial

Outcome: Complete response at surgery

Experimental arm 5
I-SPY2 TRIAL

Population of patients

- Adaptively randomized

- Experimental arm 1
- Experimental arm 4
- Experimental arm 6
- Standard therapy

Arm 6 is added to the mix

Outcome: Complete response at surgery
I-SPY-like TRIAL for Combinations

Population of patients

Randomly

A
B
C
D
C + D
SOC
A + SOC
B + SOC
C + SOC
D + SOC

Outcome: pathCR or PFS or OS
I-SPY-like TRIAL for Combinations

Population of patients

Randomly

A + SOC
B + SOC
C + SOC
D + SOC
C + D + SOC
SOC

Substudy: Adaptively randomized factorial

Outcome: pathCR or PFS or OS
I-SPY-like TRIAL for Combinations

Goal: Greater than 85% success rate in Phase 3, with focus on patients who benefit

Substudy: Adaptively randomized factorial
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: colorectal cancer, melanoma, lymphoma, Alzheimer’s, HIV, acute heart failure, SARI & H1N1, …
Rough History of Biostatistics

1800: Medicine Dominated by Case Study & Anecdote

1900: W.R. Thompson & Bayes

1940s: Bradford Hill & RCT

1933: Frequentist Half Cent—No Bayes

1997: FDAMA & Bayes Credibility
Rough History of Biostatistics

1800

Medicine Dominated by Case Study & Anecdote

W.R. Thompson & Bayes

1900

1933

Bradford Hill & RCT

1940s

2000

1997: FDAMA & Bayes Credibility

Frequentist Half Cent—No Bayes