PORT SECURITY, ANTHRAX, AND DRUG SAFETY: A DIMACS MEDLEY

David Madigan
Columbia University

Back in 2001…

Hi David, lets brainstorm on the new special focus on computational epidemiology

Sure Fred. Somebody should make the drug safety people talk to the disease surveillance people but I’m too busy to organize it

Interesting…

time passes…

arms twisted…

DIMACS

Center for Discrete Mathematics & Theoretical Computer Science
Founded as a National Science Foundation Science and Technology Center

---

DIMACS Working Group on Adverse Event/Disease Reporting, Surveillance, and Analysis

October 16 - 18, 2002
DIMACS Center, CoRE Building, Rutgers University

Organizers:

Donald Hoover, Rutgers, Statistics, drhoover@stat.rutgers.edu
David Madigan, Rutgers, Statistics, madigan@stat.rutgers.edu
Henry Rolka, (CDC), hrr2@cdc.gov

University of Pittsburgh

sas

Carnegie Mellon

FDA U.S. Food and Drug Administration
Protecting and Promoting *Your* Health

DARPA

WALTER REED
Army Institute of Research
Soldier Health • World Health

IBM

CDC

NYC Health
NEW YORK CITY DEPARTMENT of HEALTH and MENTAL HYGIENE

Los Alamos
NATIONAL LABORATORY
EST. 1943

RAND
CORPORATION

STANFORD
UNIVERSITY

at&t

UNIVERSITY OF
BATH

PROSANOS®
CORPORATION

APL
The Johns Hopkins University
Applied Physics Laboratory

GEORGE MASON
UNIVERSITY

Penn
UNIVERSITY of PENNSYLVANIA

HARVARD UNIVERSITY

Drug Safety + Disease Surveillance

PhRMA
New Medicines. New Hope.

Signal detection
methods project

Food and Drug Administration Amendments Act (FDAAA) of 2007

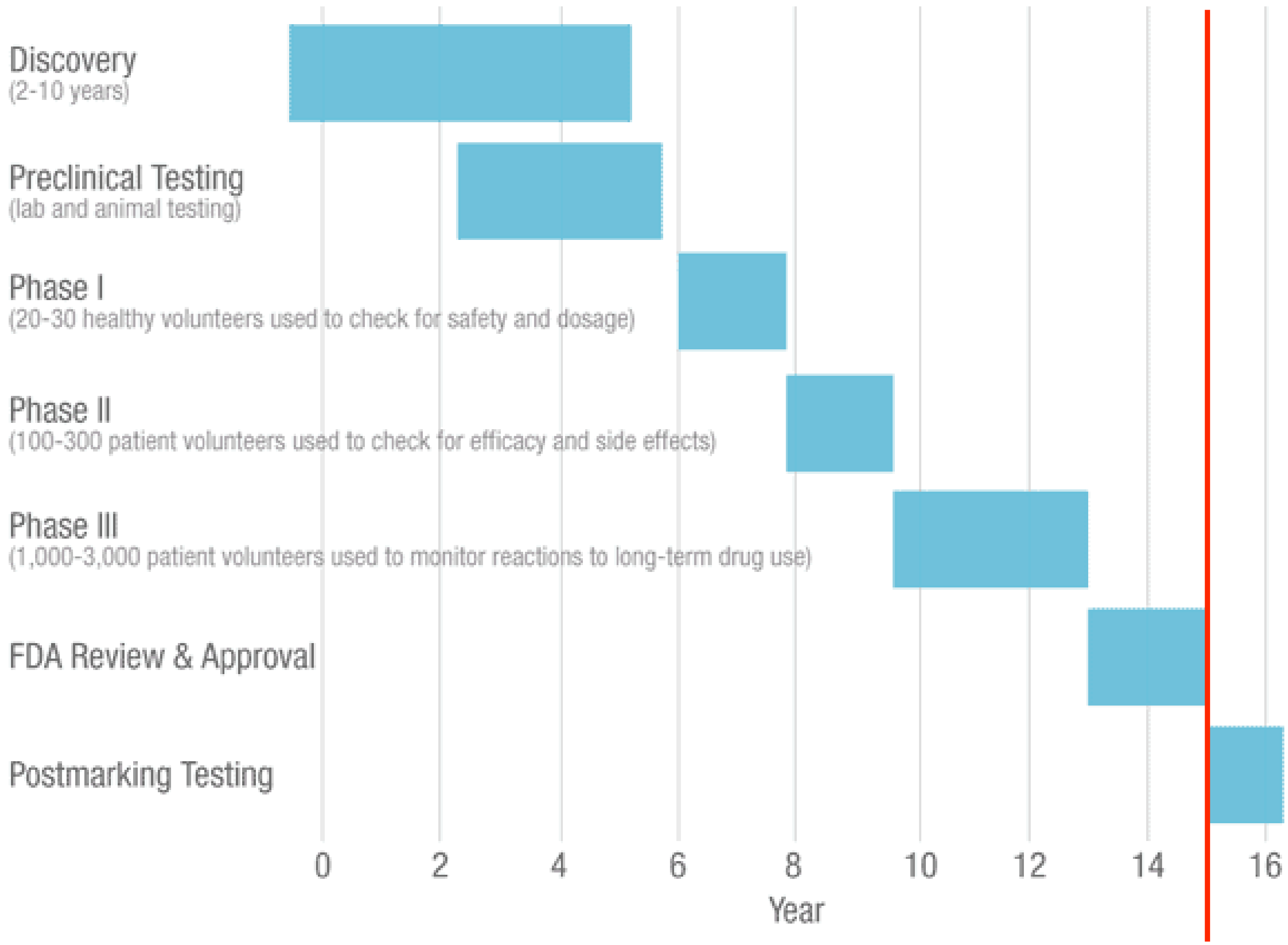
Sentinel Initiative

FOUNDATION FOR THE National Institutes of Health

Observational Medical Outcomes Partnership

Welcome to the 2009/2010 OMOP Cup!

Safety in Lifecycle of a Drug/Biologic product



Drug Safety Post-Approval

- Low quality data
- Extensive use of "data mining"

U.S. Department of Health and Human Services

Form Approved: OMB No. 0910-0291, Expires: 10/31/08
See OMB statement on reverse.

MEDWATCH

For VOLUNTARY reporting of adverse events, product problems and product use errors

The FDA Safety Information and Adverse Event Reporting Program

Page ____ of ____

FDA USE ONLY

Triage unit sequence #

PLEASE TYPE OR USE BLACK INK

A. PATIENT INFORMATION

1. **Patient Identifier** In confidence
2. **Age at Time of Event, or Date of Birth:**
3. **Sex** ☐ Female ☐ Male
4. **Weight** ______ lb or ______ kg

B. ADVERSE EVENT, PRODUCT PROBLEM OR ERROR

Check all that apply:

1. ☐ **Adverse Event** ☐ **Product Problem** *(e.g., defects/malfunctions)*
☐ **Product Use Error** ☐ **Problem with Different Manufacturer of Same Medicine**

2. **Outcomes Attributed to Adverse Event** *(Check all that apply)*
☐ Death: ____________ *(mm/dd/yyyy)*
☐ Disability or Permanent Damage
☐ Life-threatening
☐ Congenital Anomaly/Birth Defect
☐ Hospitalization - initial or prolonged
☐ Other Serious (Important Medical Events)
☐ Required Intervention to Prevent Permanent Impairment/Damage (Devices)

3. **Date of Event** *(mm/dd/yyyy)*
4. **Date of this Report** *(mm/dd/yyyy)*
5. **Describe Event, Problem or Product Use Error**
6. **Relevant Tests/Laboratory Data, Including Dates**
7. **Other Relevant History, Including Preexisting Medical Conditions** *(e.g., allergies, race, pregnancy, smoking and alcohol use, liver/kidney problems, etc.)*

C. PRODUCT AVAILABILITY

Product Available for Evaluation? *(Do not send product to FDA)*

☐ Yes ☐ No ☐ Returned to Manufacturer on: ____________ *(mm/dd/yyyy)*

D. SUSPECT PRODUCT(S)

1. **Name, Strength, Manufacturer** *(from product label)*
#1
#2

2. | | Dose or Amount | Frequency | Route |
|---|---|---|---|
| #1 | | | |
| #2 | | | |

3. **Dates of Use** *(If unknown, give duration) from/to (or best estimate)*
#1
#2

4. **Diagnosis or Reason for Use** *(Indication)*
#1
#2

5. **Event Abated After Use Stopped or Dose Reduced?**
#1 ☐ Yes ☐ No ☐ Doesn't Apply
#2 ☐ Yes ☐ No ☐ Doesn't Apply

6. **Lot #**
#1
#2

7. **Expiration Date**
#1
#2

8. **Event Reappeared After Reintroduction?**
#1 ☐ Yes ☐ No ☐ Doesn't Apply
#2 ☐ Yes ☐ No ☐ Doesn't Apply

9. **NDC # or Unique ID**

E. SUSPECT MEDICAL DEVICE

1. **Brand Name**
2. **Common Device Name**
3. **Manufacturer Name, City and State**
4. **Model #** **Lot #**
Catalog # **Expiration Date** *(mm/dd/yyyy)*
Serial # **Other #**
5. **Operator of Device**
☐ Health Professional
☐ Lay User/Patient
☐ Other:
6. **If Implanted, Give Date** *(mm/dd/yyyy)*
7. **If Explanted, Give Date** *(mm/dd/yyyy)*
8. **Is this a Single-use Device that was Reprocessed and Reused on a Patient?**
☐ Yes ☐ No
9. **If Yes to Item No. 8, Enter Name and Address of Reprocessor**

F. OTHER (CONCOMITANT) MEDICAL PRODUCTS

Product names and therapy dates *(exclude treatment of event)*

G. REPORTER *(See confidentiality section on back)*

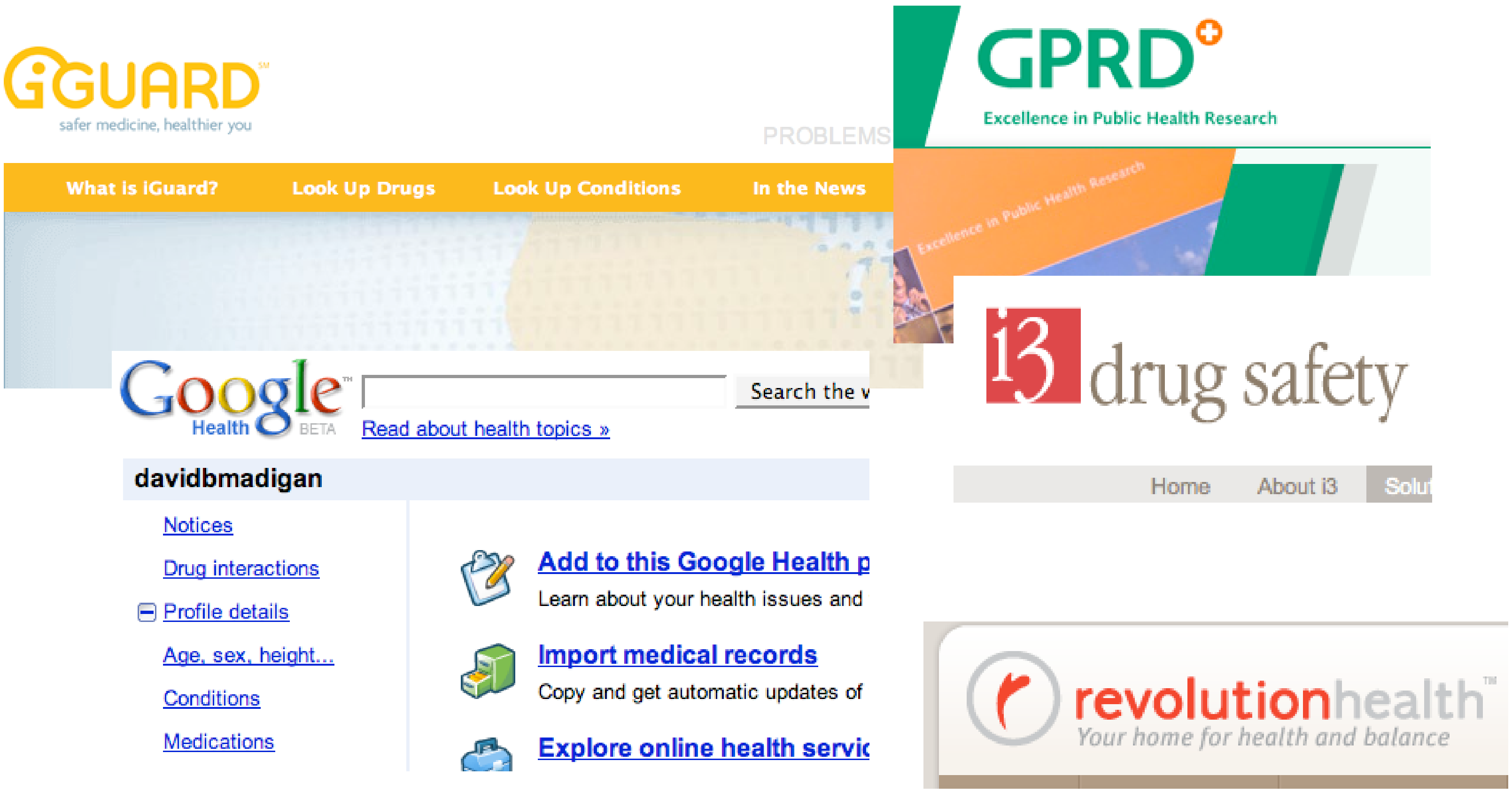
1. **Name and Address**
Phone # **E-mail**
2. **Health Professional?** ☐ Yes ☐ No
3. **Occupation**
4. **Also Reported to:**
☐ Manufacturer
☐ User Facility
☐ Distributor/Importer
5. **If you do NOT want your identity disclosed to the manufacturer, place an "X" in this box:** ☐

 Submission of a report does not constitute an admission that medical personnel or the product caused or contributed to the event.

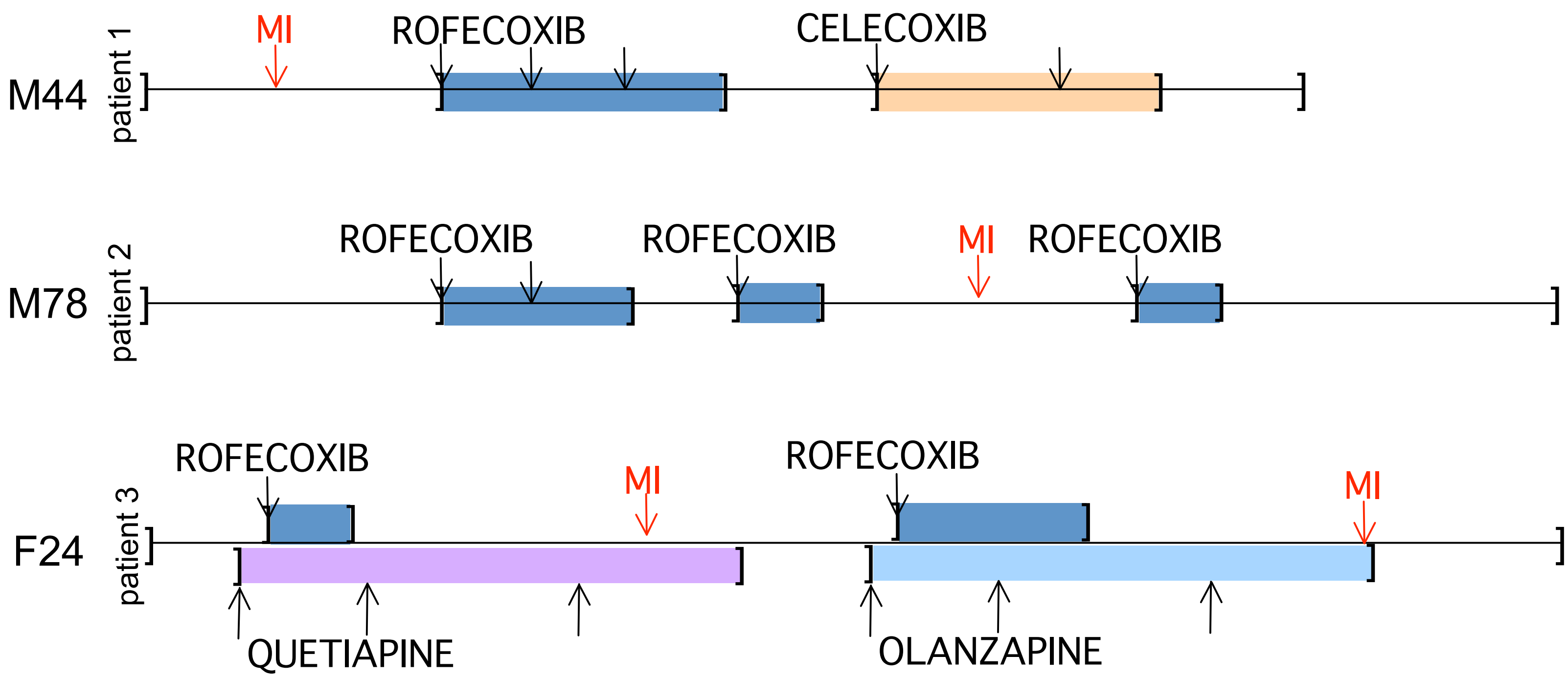
Problems with Spontaneous Reports

- Under-reporting
- Duplicate reports
- No temporal information
- No denominator

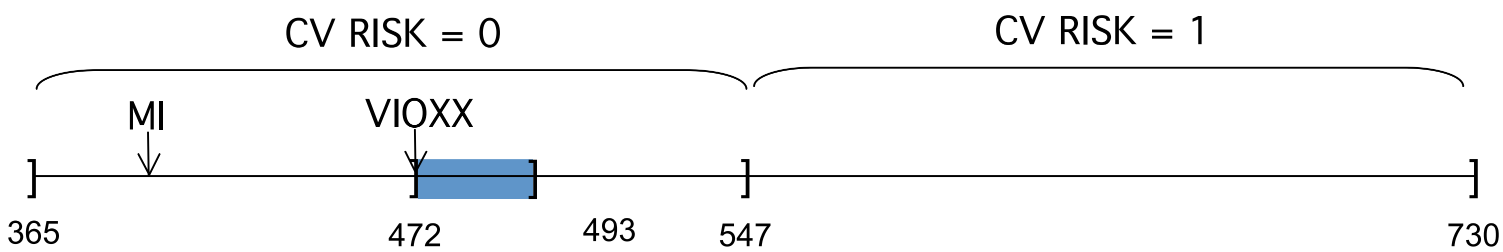
Newer Data Sources for PV



Longitudinal Claims Data



Self Controlled Case Series



- assume diagnoses arise according to a non-homogeneous Poisson process

$e^{!_i}$ baseline incidence for subject *i*

$e^{!_1}$ relative incidence associated with CV risk group 1

$e^{!_1}$ relative incidence associated with Vioxx risk level 1

$!_1 = 107 e^{\;1}$ Poisson rate for subject 1, period 1 overall Poisson rate for subject 1:

$$\Lambda = 107e^{\phi_1} + 21e^{\phi_1}e^{\beta_1} + 54e^{\phi_1} + 183e^{\phi_1}e^{\alpha_1}$$

cohort study contribution to the likelihood:

$$\left(\lambda_1 e^{-\lambda_1}\right) \times e^{-\lambda_2} \times e^{-\lambda_3} \times e^{-\lambda_4} = \lambda_1 e^{-\Lambda}$$

conditional likelihood:

$$\begin{aligned}
\frac{\lambda_1 e^{-\Lambda}}{\Lambda e^{-\Lambda}} &= \frac{\lambda_1}{\Lambda} \\
&= \frac{107e^{\phi_1}}{107e^{\phi_1} + 21e^{\phi_1}e^{\beta_1} + 54e^{\phi_1} + 183e^{\phi_1}e^{\alpha_1}} \\
&= \frac{107}{107 + 21e^{\beta_1} + 54 + 183e^{\alpha_1}}.
\end{aligned}$$

Self-Controlled Case Series Method

Farrington et al.

equivalent multinomial likelihood:

$$l(\alpha_1, \beta_1) = \left(\frac{107}{107 + 21e^{\beta_1} + 54 + 183e^{\alpha_1}}\right)^1 \times \left(\frac{21e^{\beta_1}}{107 + 21e^{\beta_1} + 54 + 183e^{\alpha_1}}\right)^0 \times \left(\frac{54}{107 + 21e^{\beta_1} + 54 + 183e^{\alpha_1}}\right)^0 \times \left(\frac{183e^{\alpha_1}}{107 + 21e^{\beta_1} + 54 + 183e^{\alpha_1}}\right)^0$$

regularization => Bayesian approach

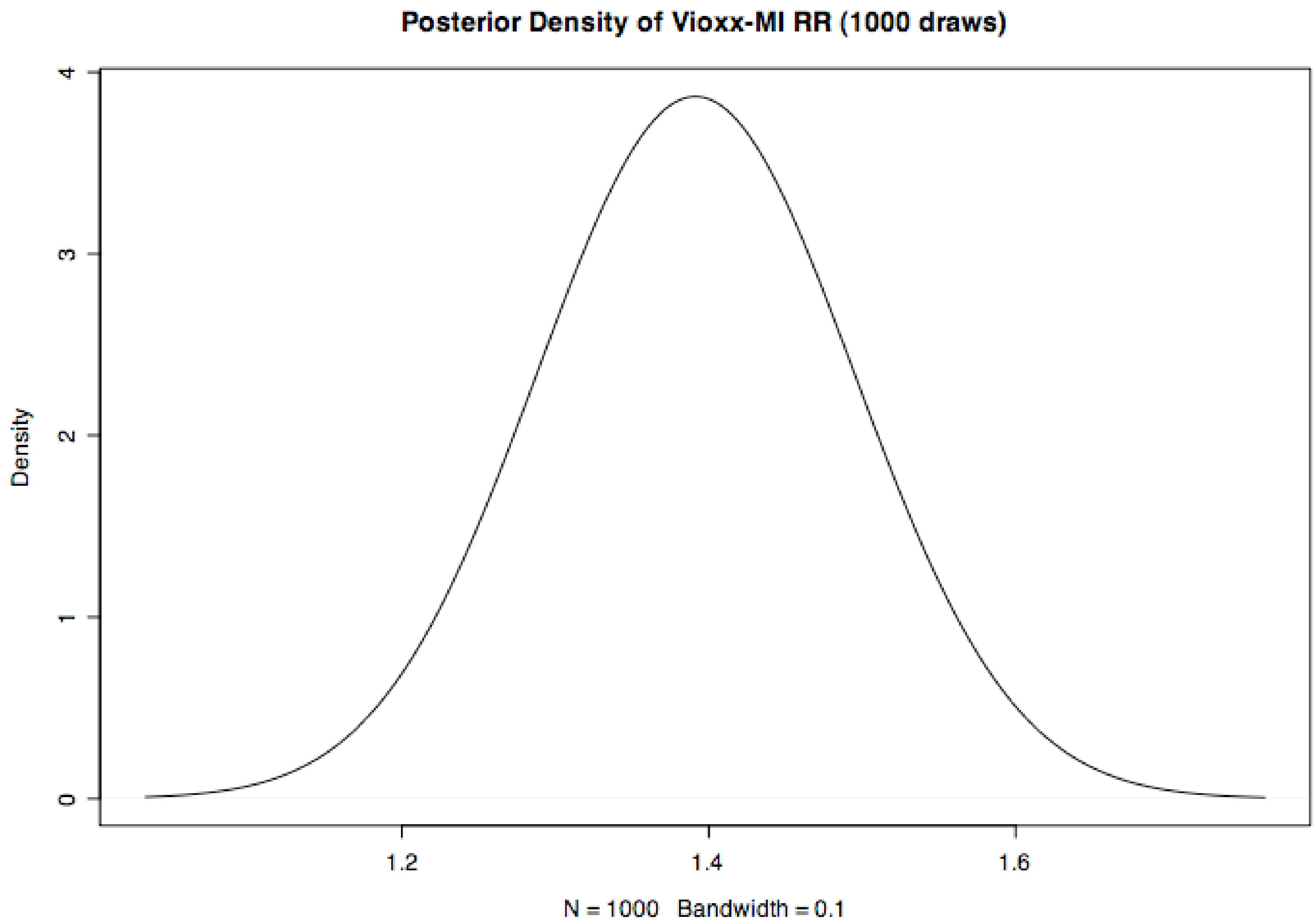
scale to full database?

Vioxx & MI: SCCS RRs
i3 claims database

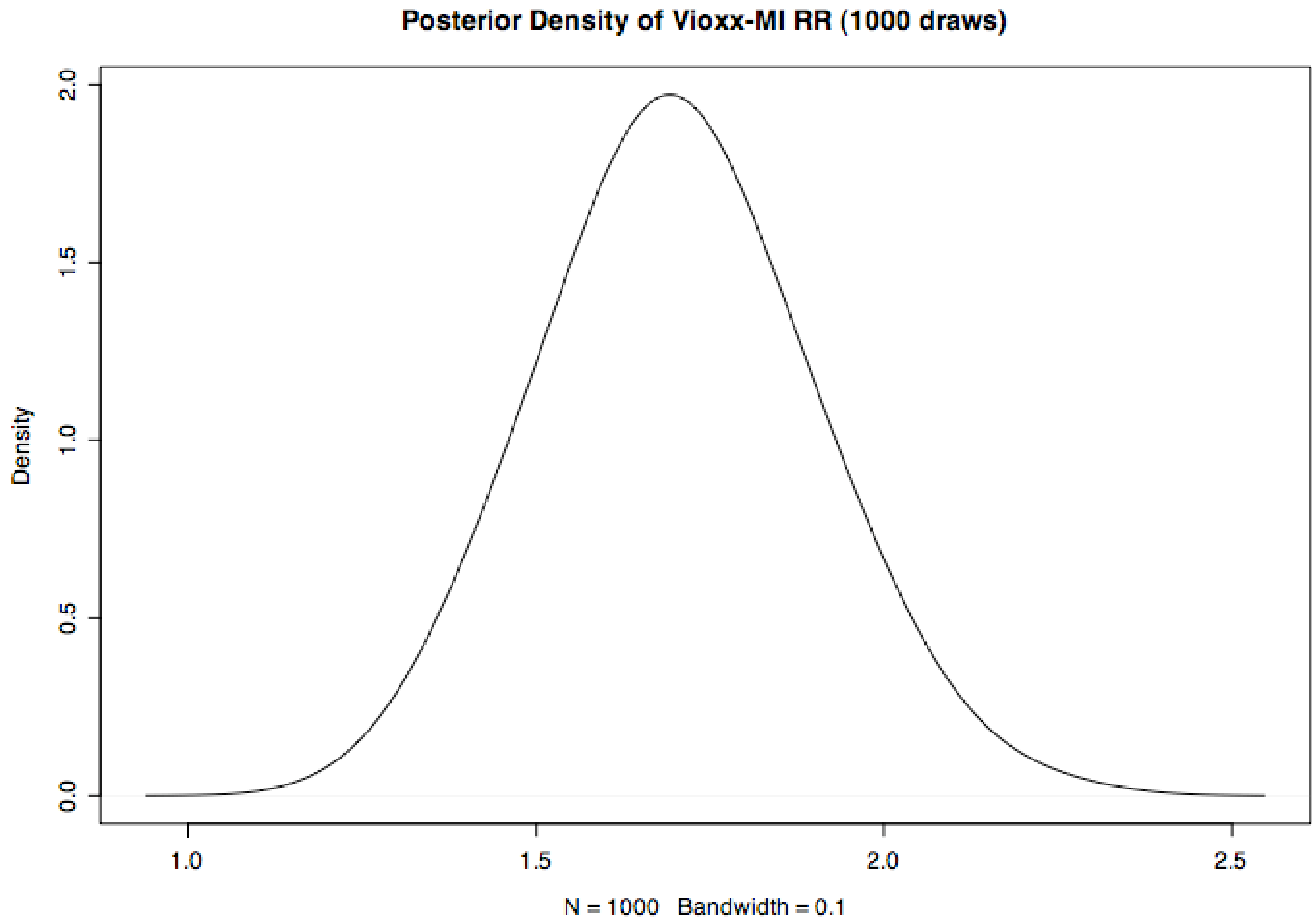
- Bayesian analysis N(0,10) prior + MCMC

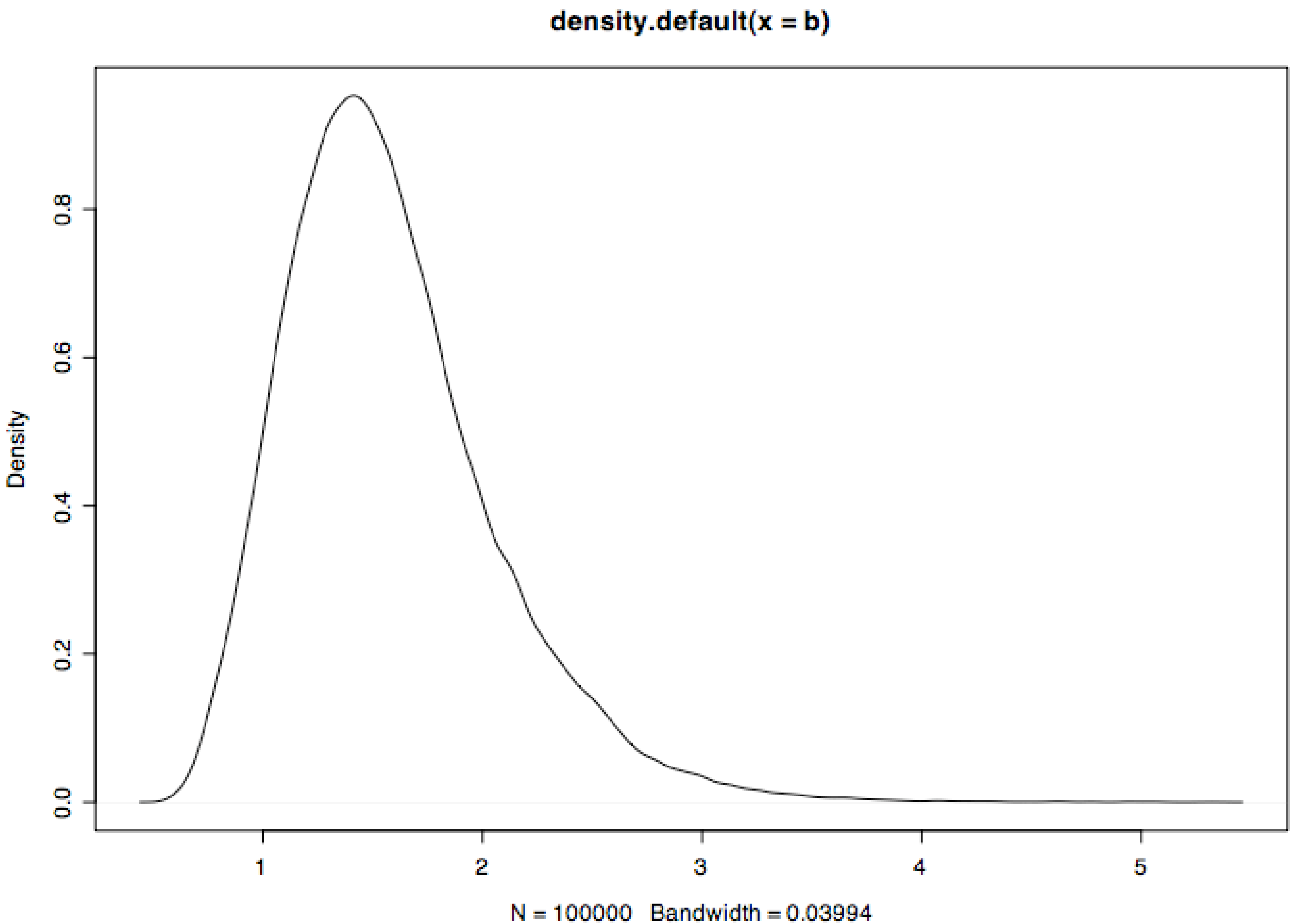
- Overall: 1.38 (n=11,581)
- Male: 1.41   Female: 1.36
- Age >= 80: 1.48
- Male + Age >= 80: 1.68

overall (n=11,581)

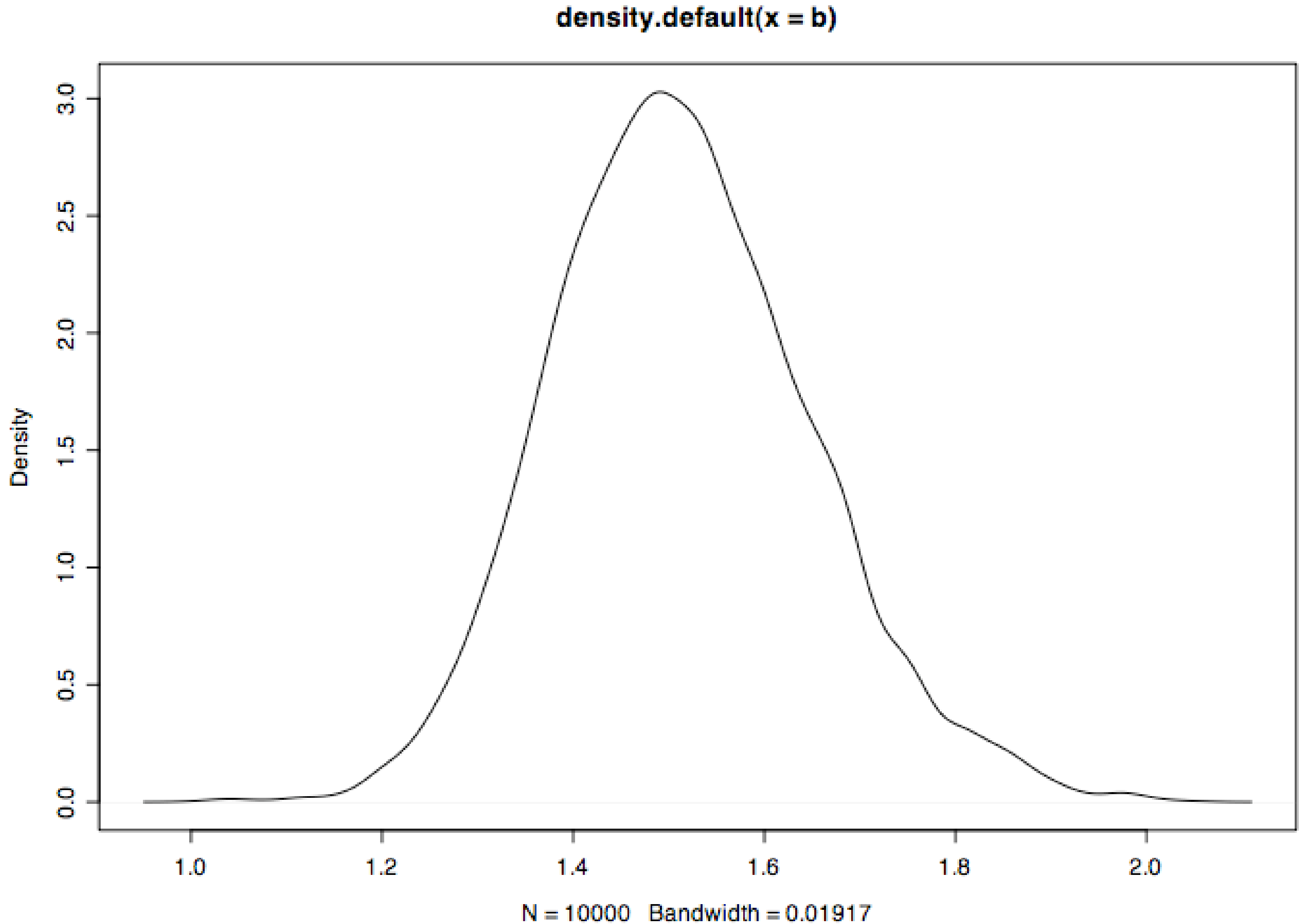


males 80 and over (n=440)





June 30, 2000 RR=1.53 Pr(RR>1)=0.92



Dec 31, 2000 RR=1.51 Pr(RR>1)=1.0

The New York Times

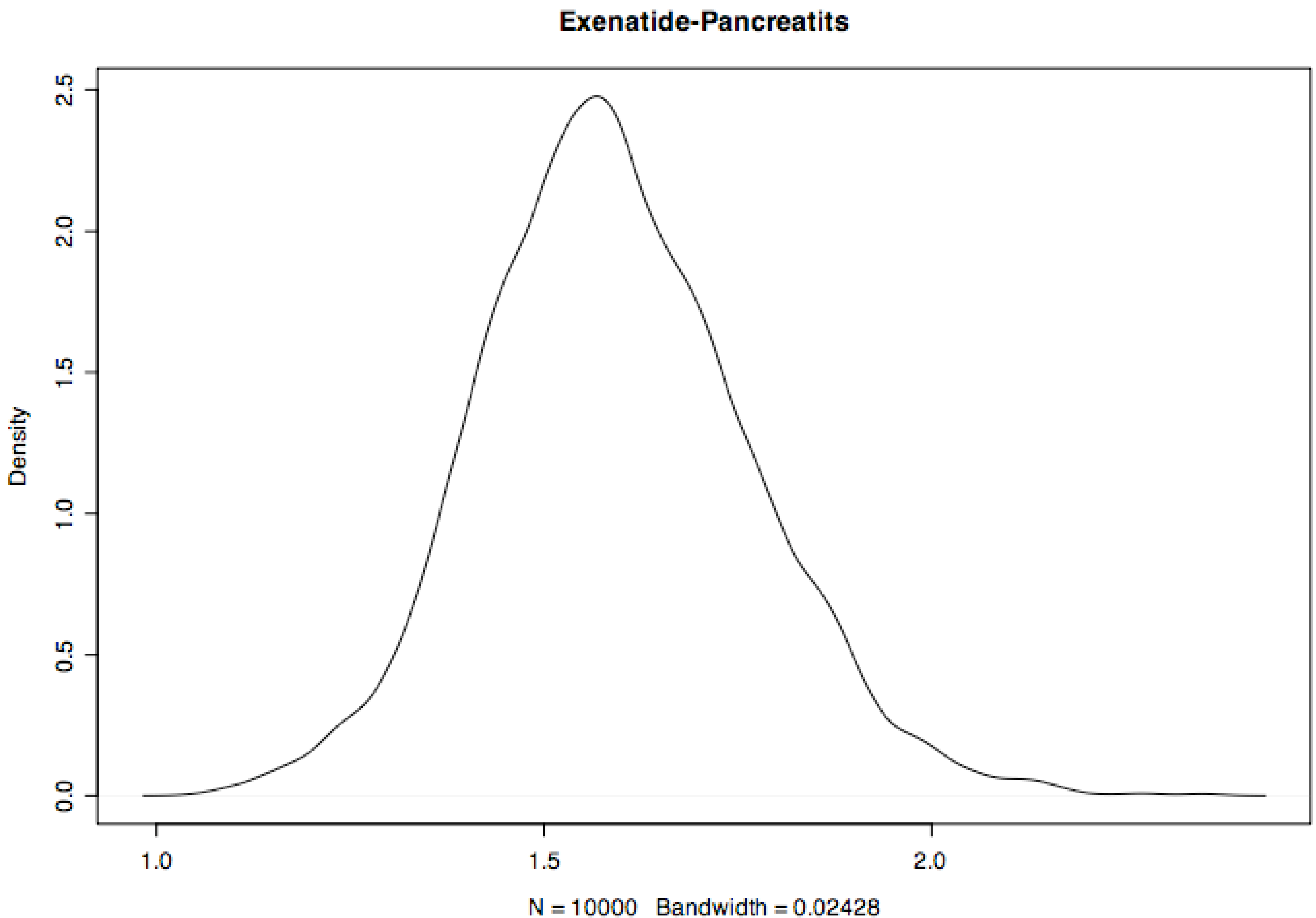
Diabetes Drug Tied to New Deaths

By BLOOMBERG NEWS
Published: August 26, 2008

The diabetes drug Byetta, marketed by Eli Lilly & Company and Amylin Pharmaceuticals, was linked to four more deaths in patients with pancreatitis, adding to two deaths announced by federal regulators last week.

No definite relationship between Byetta and the deaths has been proved, and the Food and Drug Administration was aware of the additional deaths when it made its announcement last week, Amylin's chief executive, Dan Bradbury, said on Tuesday. The company is talking with the F.D.A. about adding warnings on the drug's prescribing information.

Byetta is Amylin's leading product.


Exenatide-Pancreatits
Density
0.0
0.5
1.0
1.5
2.0
2.5
1.0
1.5
2.0
N = 10000   Bandwidth = 0.02428

Back in 2004...

Hi David, you might be interested in some of the port security work we are doing

Sounds interesting Fred but I'm too busy with the drug safety stuff

Let me tell you more...

time passes...

arms twisted...

Port of Entry Inspection Algorithms

Aim: Develop decision support algorithms that will help us to "optimally" intercept illicit materials and weapons subject to limits on delays, manpower, and equipment

Find inspection schemes that minimize total cost including cost of false positives and false negatives

Mobile VACIS: truck-mounted gamma ray imaging system

Sequential Decision Making Problem

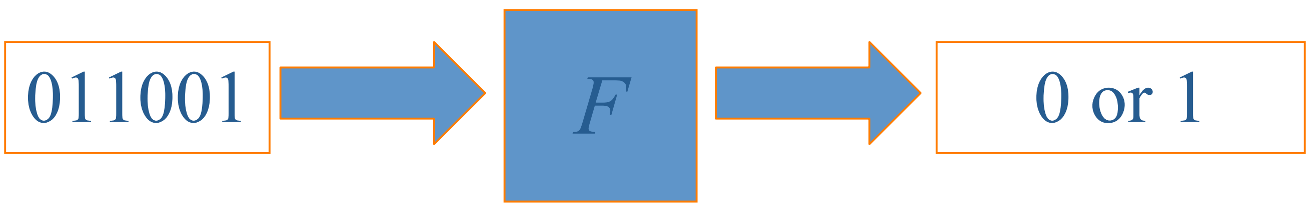
- Containers arriving are classified into categories
- Simple case: 0 = “ok”, 1 = “suspicious”
- Containers have attributes, either in state 0 or 1
- **Sample attributes**:
  - Does the ship’s manifest set off an alarm?
  - Is the neutron or Gamma emission count above certain threshold?
  - Does a radiograph image return a positive result?
  - Does an induced fission test return a positive result?
- **Inspection scheme:**
  - *specifies which inspections are to be made based on previous observations*
- Different “sensors” detect presence or absence of various attributes

Sequential Decision Making Problem

- Simplest Case: Attributes are in state 0 or 1

- Then: Container is a *binary string* like 011001

- So: Classification is a *decision function* F that assigns each binary string to a category.



If attributes 2, 3, and 6 are present, assign container to category $F(011001)$.

Sequential Decision Making Problem

- If there are two categories, 0 and 1, decision function F is a *Boolean function*.

- Example:

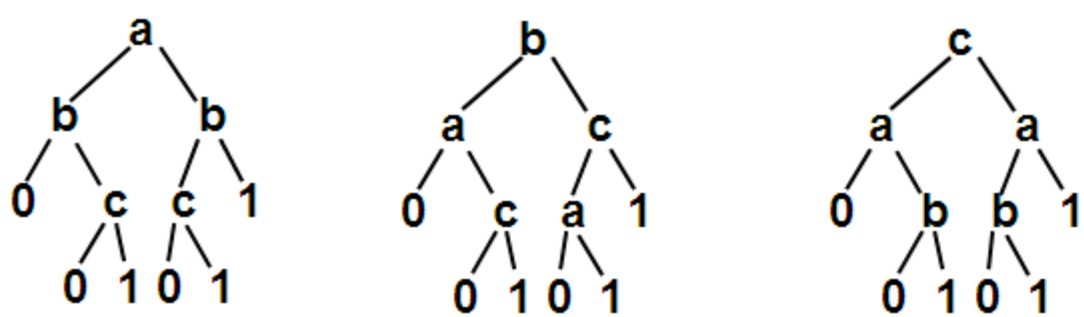
a	b	c	F(abc)
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	1

- This function classifies a container as positive iff it has at least two of the attributes.

Binary Decision Tree Approach

- Binary Decision Tree:
  - –Nodes are sensors or categories (0 or 1)
  - –Two arcs exit from each sensor node, labeled left and right.
  - –Take the right arc when sensor says the attribute is present, left arc otherwise

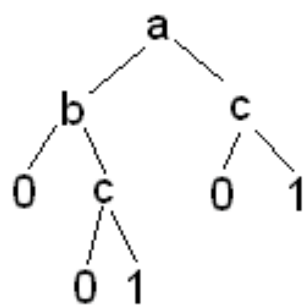
a	b	c	F(abc)
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	1



Cost of a BDT

- Cost of a BDT comprises of:
  - Cost of utilization of the tree and
  - Cost of misclassification

A BDT, τ
with n = 3



$$
\begin{aligned}
f(!) = {} & P_0(C_a + P_{a0|0}C_b + P_{a0|0}P_{b1|0}C_c + P_{a1|0}C_c) \\
& + P_1(C_a + P_{a0|1}C_b + P_{a0|1}P_{b1|1}C_c + P_{a1|1}C_c) \\
& + P_0(P_{a0|0}P_{b1|0}P_{c1|0} + P_{a1|0}P_{c1|0})C_{FP} \\
& + P_1(P_{a0|1}P_{b0|1} + P_{a0|1}P_{b1|1}P_{c0|1} + P_{a1|1}P_{c0|1})C_{FN}
\end{aligned}
$$

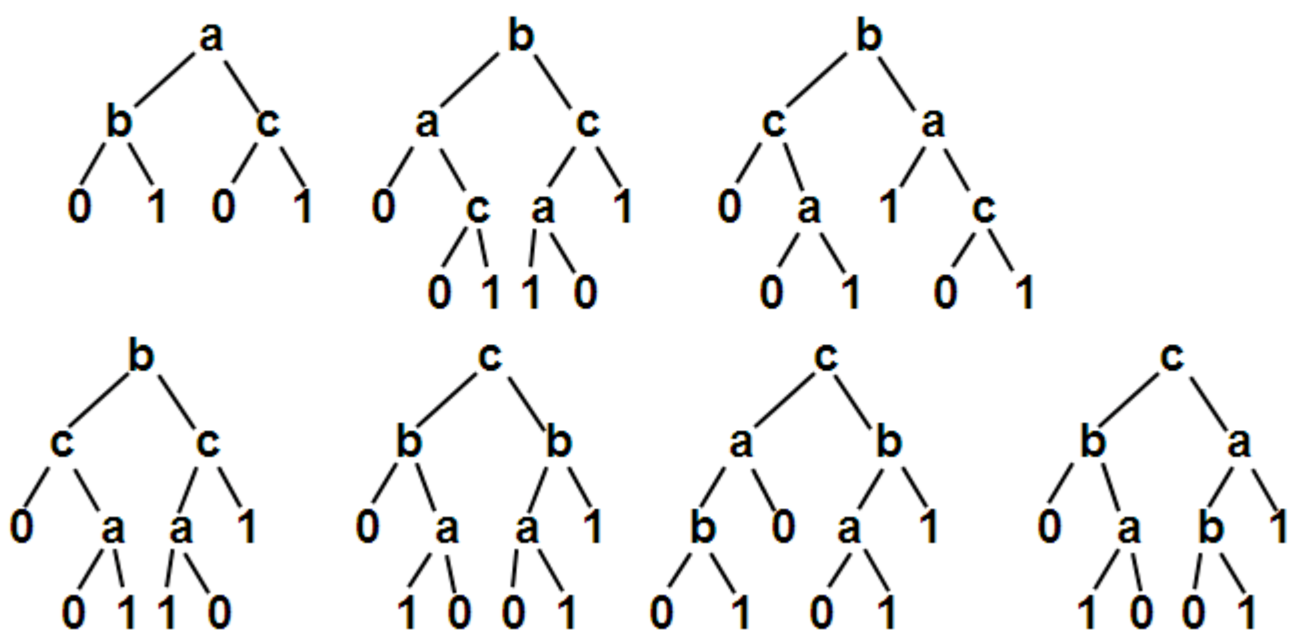
P_1 is prior probability of occurrence of a bad container

$P_{i|j}$ is the conditional probability that given the container was in state j, it was classified as i

Revisiting Monotonicity

- **Monotonic Decision Trees**
  - A binary decision tree will be called monotonic if all the left leafs are class "0" and all the right leafs are class "1".
- Example:

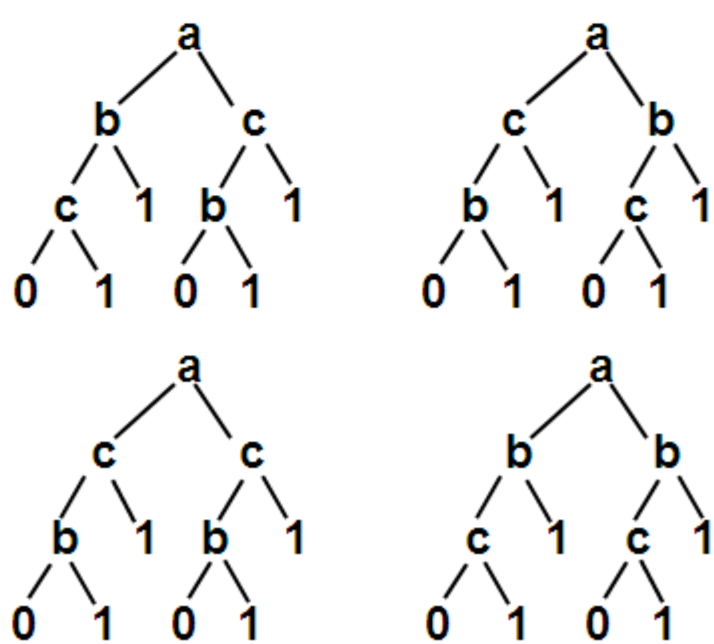
a	b	c	F(abc)
0	0	0	0
0	0	1	0
0	1	0	1
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	0
1	1	1	1



Revisiting Completeness

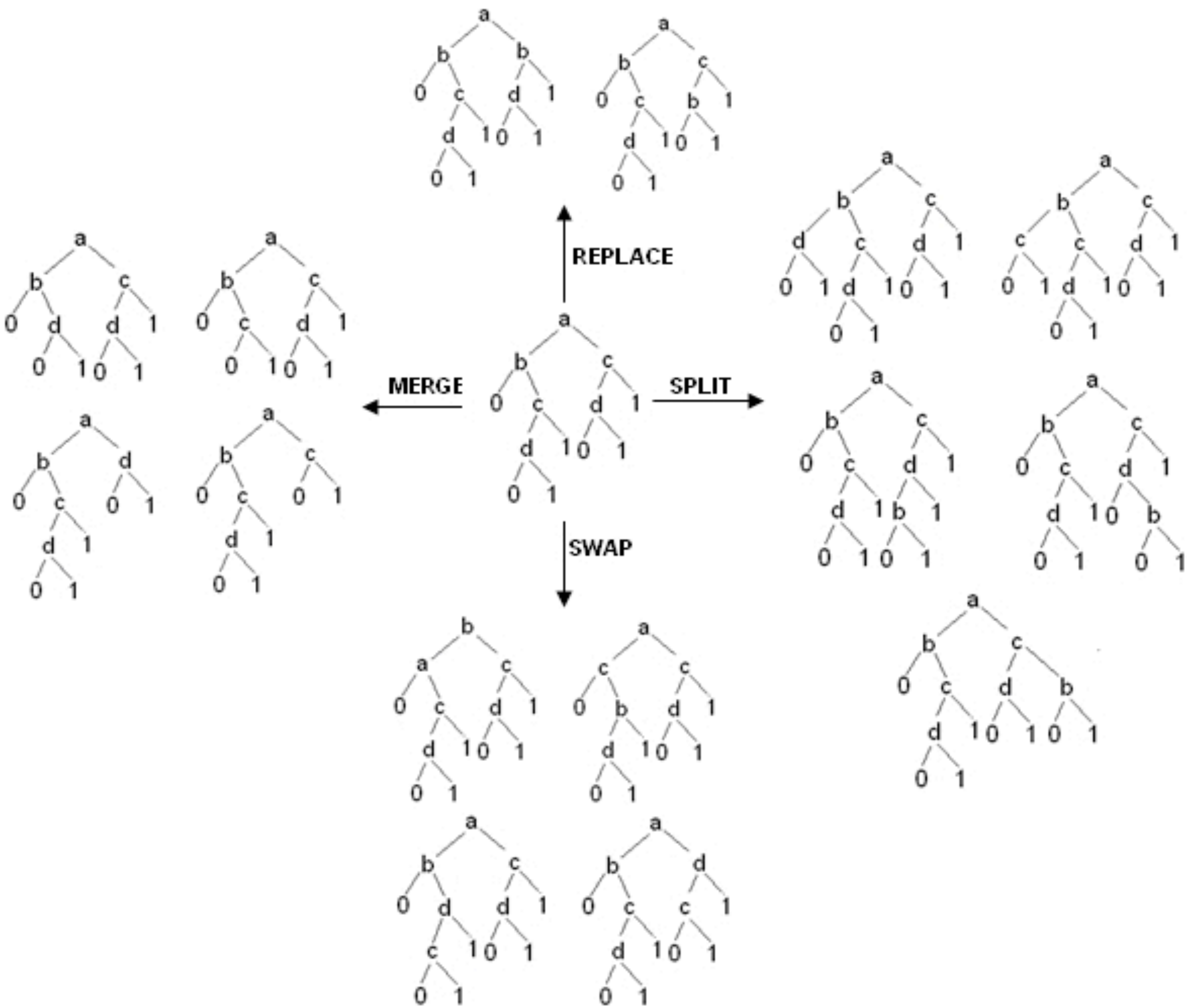
- **Complete Decision Trees**
  - A binary decision tree will be called complete if every sensor occurs at least once in the tree and at any non-leaf node in the tree, its left and right sub-trees are not identical.
- Example:

a	b	c	F(abc)
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	1



The CM Tree Space

No. of attributes	*Distinct BDTs*	*Trees From CM Boolean Functions*	*Complete and Monotonic BDTs*
2	*74*	*4*	*4*
3	*16,430*	*60*	*114*
4	*1,079,779,602*	*11,808*	*66,000*



REPLACE
MERGE
SPLIT
SWAP

Tree Space Traversal

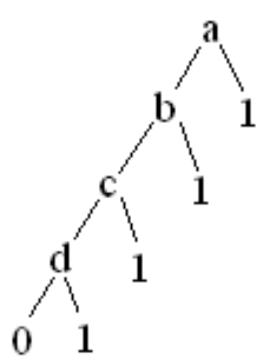
- Greedy Search
  1. Randomly start at any tree in the CM tree space
  2. Find its neighboring trees using neighborhood operations
  3. Move to the neighbor with the lowest cost
  4. Iterate till the solution converges

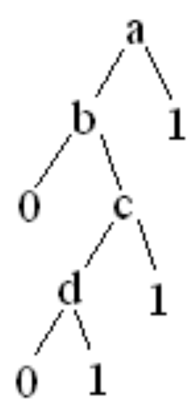
  – The CM Tree space has a lot of local minima. For example: 9 in the space of 114 trees for 3 sensors and 193 in the space of 66,000 trees for 4 sensors.

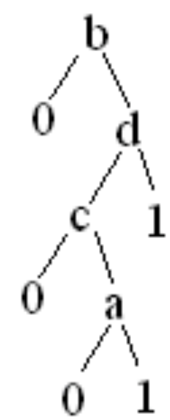
- Proposed Solutions
  - Stochastic Search Method with Simulated Annealing
  - Genetic Algorithms based Search Method

Tree Space Irreducibility

- We have proved that the CM tree space is irreducible under the neighborhood operations
- Simple Tree:
  - A simple tree is defined as a CM tree in which every sensor occurs exactly once in such a way that there is exactly one path in the tree with all sensors in it.







Results

- Significant computational savings over previous methods
- Have run experiments with up to 10 sensors
- Genetic algorithms especially useful for larger scale problems

Current Work

- Tree equivalence
- Tree reduction and irreducible trees
- Canonical form representation of the equivalence class of trees
- Revisiting completeness and monotonicity

Thank You!