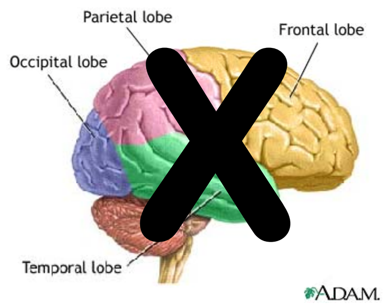


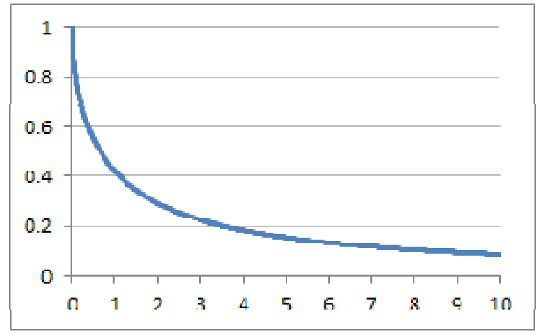


# *Bayes to the Rescue*

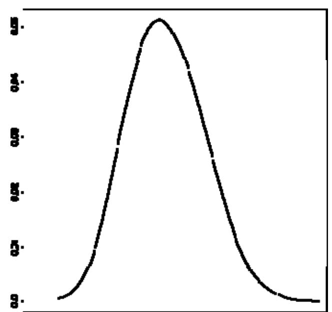
**David Madigan**
**Columbia University**


# Bayesian Learning Paradigm




prior



posterior

Bayes' Rule

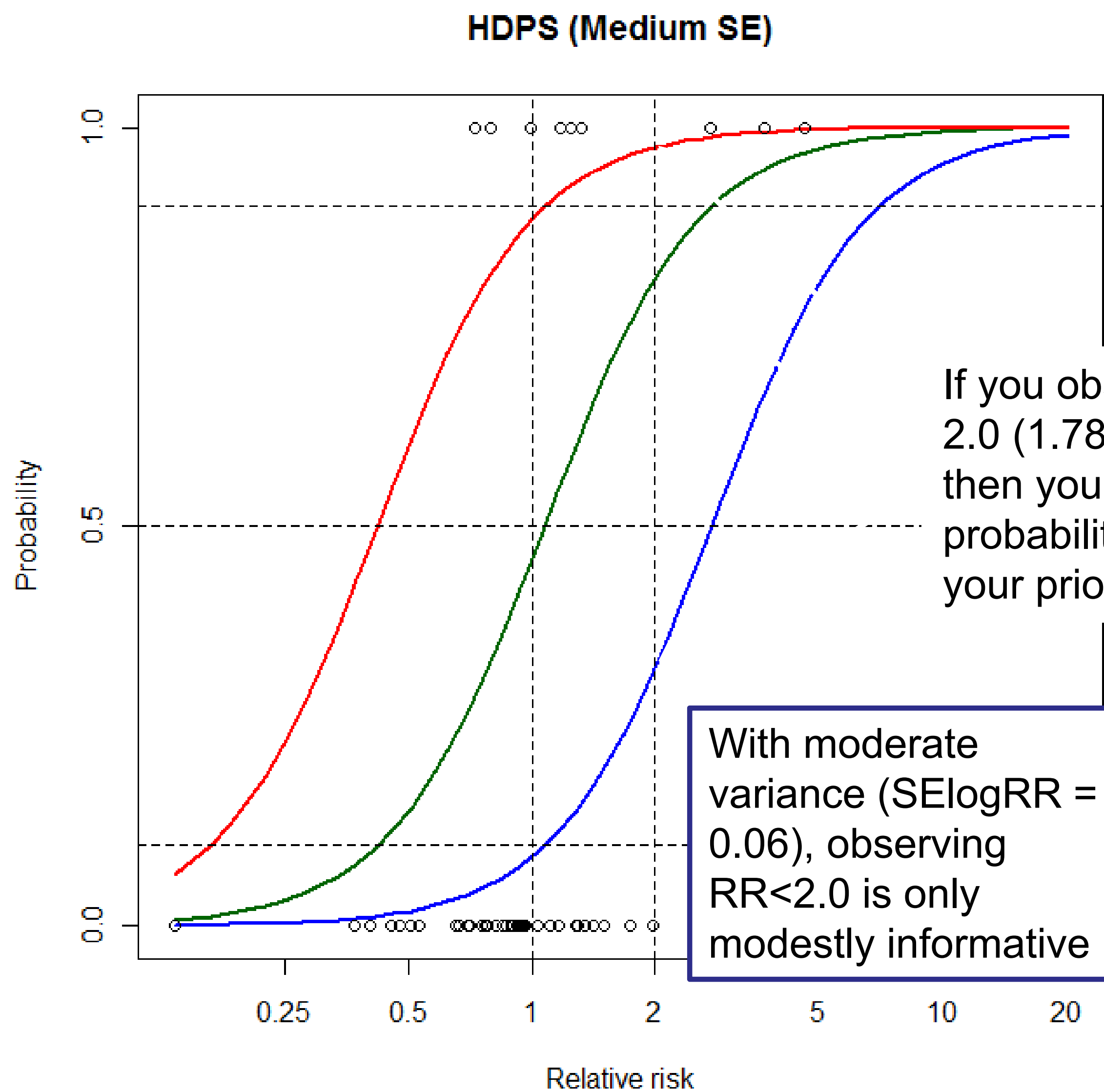
| Name | Thread pitch (mm) | Minor diameter tolerance | Nominal diameter (mm) | Head shape | Price for 50 screws | Available at factory outlet? | Number in stock | Flat or Phillips head? |
|---|---|---|---|---|---|---|---|---|
| M4 | 0.7 | 4g | 4 | Pan | $10.08 | Yes | 276 | Flat |
| M5 | 0.8 | 4g | 5 | Round | $13.89 | Yes | 183 | Both |
| M6 | 1 | 5g | 6 | Button | $10.42 | Yes | 1043 | Flat |
| M8 | 1.25 | 5g | 8 | Pan | $11.98 | No | 298 | Phillips |
| M10 | 1.5 | 6g | 10 | Round | $16.74 | Yes | 488 | Phillips |
| M12 | 1.75 | 7g | 12 | Pan | $18.26 | No | 998 | Flat |
| M14 | 2 | 7g | 14 | Round | $21.19 | No | 235 | Phillips |
| M16 | 2 | 8g | 16 | Button | $23.57 | Yes | 292 | Both |
| M18 | 2.1 | 8g | 18 | Button | $25.87 | No | 664 | Both |
| M20 | 2.4 | 8g | 20 | Pan | $29.09 | Yes | 486 | Both |
| M24 | 2.55 | 9g | 24 | Round | $33.01 | Yes | 982 | Phillips |
| M28 | 2.7 | 10g | 28 | Button | $35.66 | No | 1067 | Phillips |
| M36 | 3.2 | 12g | 36 | Pan | $41.32 | No | 434 | Both |
| M50 | 4.5 | 15g | 50 | Pan | $44.72 | No | 740 | Flat |

data

# “Data”

| Effect estimates of HDPS against CCAE (RR, SE) | Angioedema #1 | Aplastic Anemia #1 | Acute Liver Failure #1 | Bleeding #1 | Acute myocardial Infarction #1 | Hip Fracture #1 | Mortality after Myocardial Infarction #1 | Acute Renal Failure #1 | Upper GI Ulcer Hospitalization #1 |
|---|---|---|---|---|---|---|---|---|---|
| OMOP ACE Inhibitor | 1.80 (0.15) | 0.40 (0.05) | | | | 0.91 (0.12) | | | 0.87 (0.03) |
| OMOP Amphotericin B | | 3.30 (0.99) | 1.05 (0.24) | | | | | 4.01 (0.99) | |
| OMOP Antibiotics | | 1.22 (0.08) | 1.00 (0.01) | 1.14 (0.01) | 1.06 (0.03) | 1.05 (0.09) | | 1.44 (0.06) | |
| OMOP Antiepileptics | 1.74 (0.38) | 4.60 (0.80) | | | | | | 1.63 (0.21) | 0.54 (0.05) |
| OMOP Benzodiazepines | 0.13 (0.01) | 1.10 (0.06) | 0.98 (0.01) | 1.11 (0.01) | 1.18 (0.03) | 1.41 (0.12) | | 1.06 (0.05) | |
| OMOP Beta blockers | 0.81 (0.07) | 0.63 (0.06) | 0.95 (0.02) | | | 1.69 (0.19) | | 0.78 (0.04) | 0.88 (0.03) |
| OMOP Bisphosphonates | | 0.27 (0.05) | 0.85 (0.03) | | 0.82 (0.07) | | | 0.40 (0.04) | 0.90 (0.06) |
| OMOP Tricyclic antidepressa | | 0.63 (0.07) | 1.02 (0.02) | 0.96 (0.01) | 0.80 (0.04) | | | 0.82 (0.06) | |
| OMOP Typical antipsychotic | | | | | 0.96 (0.08) | | | 1.97 (0.16) | 3.46 (0.21) |
| OMOP Warfarin | 0.53 (0.11) | 0.47 (0.04) | | 2.13 (0.04) | | 1.2 (0.09) | 0.49 (0.07) | 0.76 (0.05) | |

# Revising prior expectations in light of new evidence from an active surveillance system

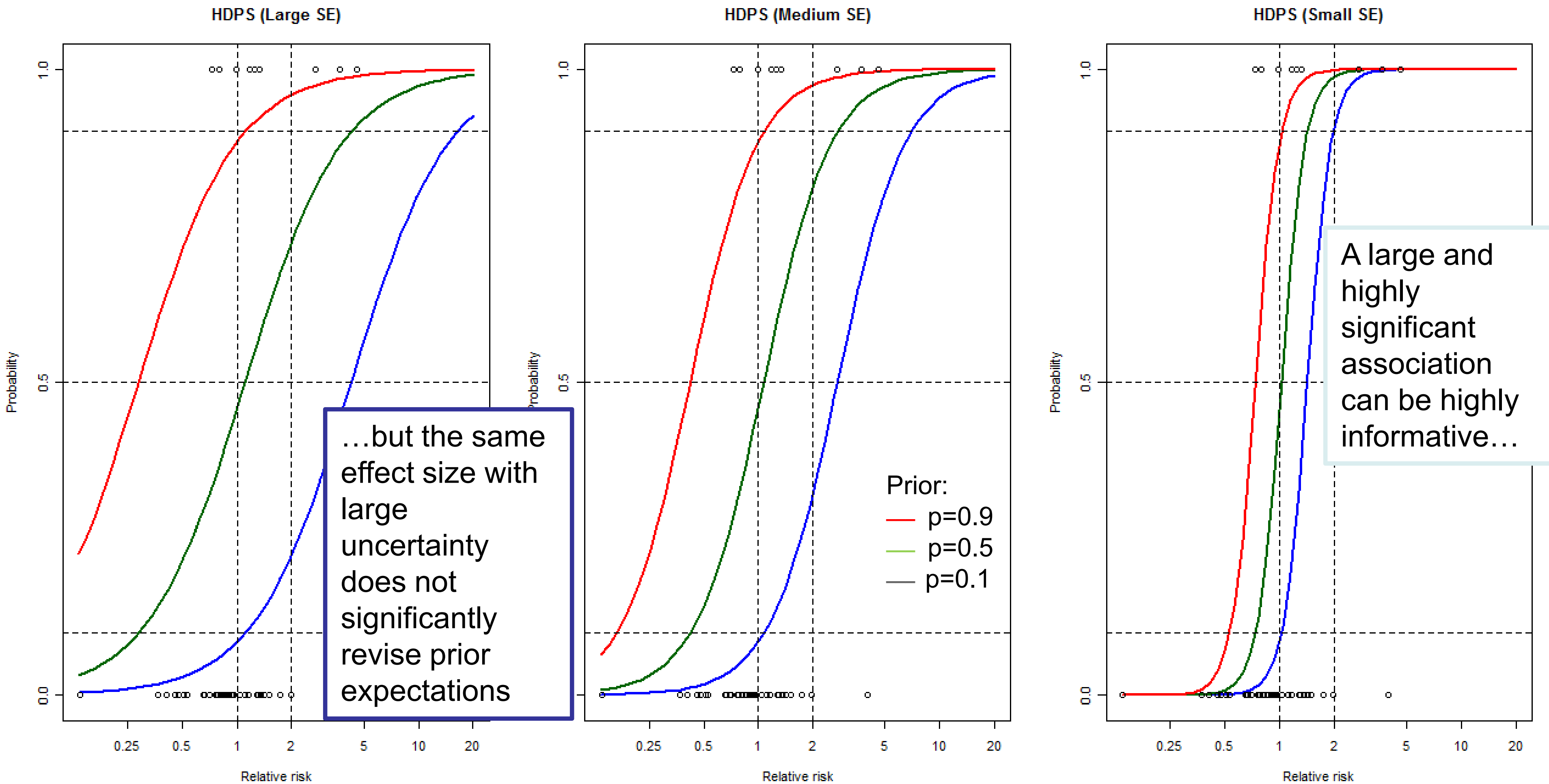



If you observe a RR = 2.0 (1.78 – 2.25), then your posterior probability depends on your prior expectations

With moderate variance (SElogRR = 0.06), observing RR<2.0 is only modestly informative

Prior:
- p=0.9
- p=0.5
- p=0.1

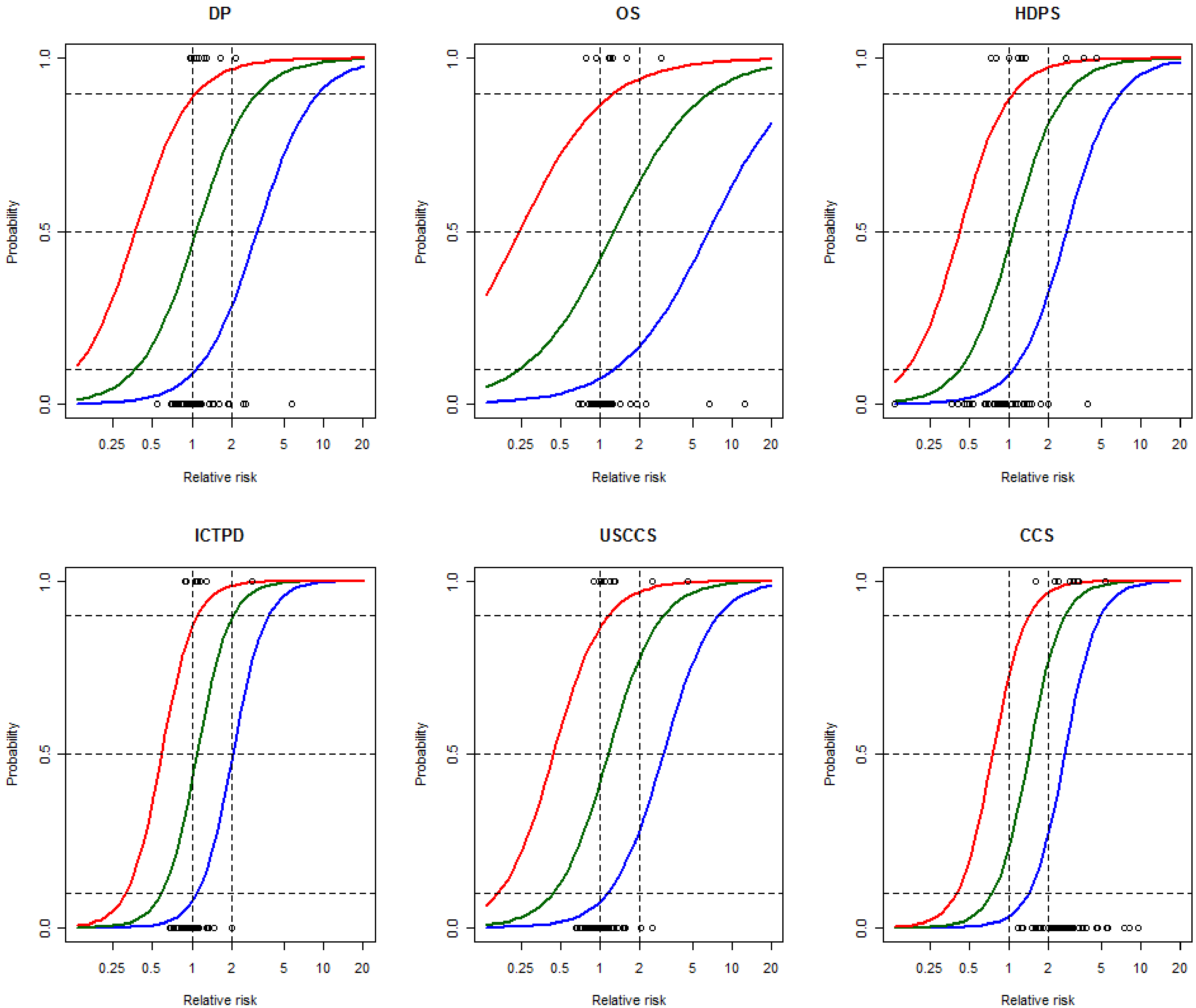
# Impact of precision of observed estimates




Scenarios: You observe RR=2.0 with confidence intervals based on standard error (SE):

Large SE: (1.01 – 3.97) Medium SE: (1.78 – 2.25) Small SE: (1.96 – 2.04)

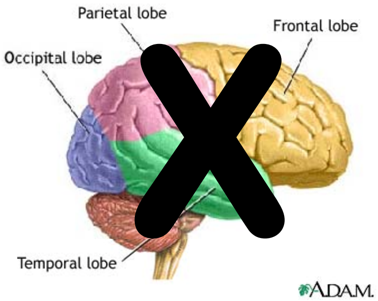
# Impact of using estimates from different methods

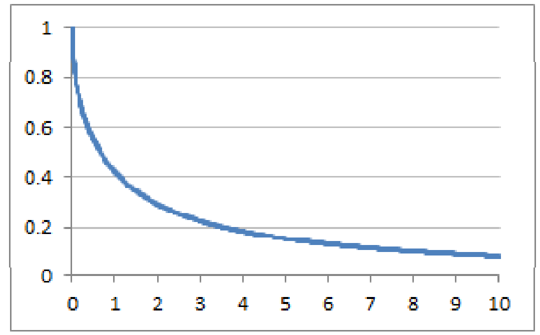



Across all methods, observing a RR=2.0 with moderate variance (SElogRR = 0.06) does not increase a skeptical prior beyond indifference

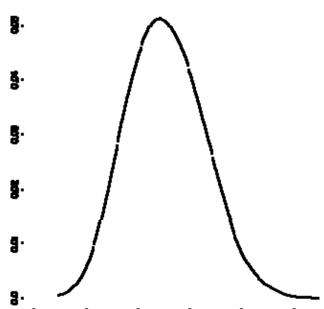
Prior:
- p=0.9
- p=0.5
- p=0.1

# Bayesian Learning Paradigm




prior

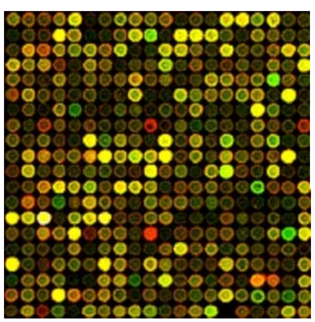


Bayes' Rule

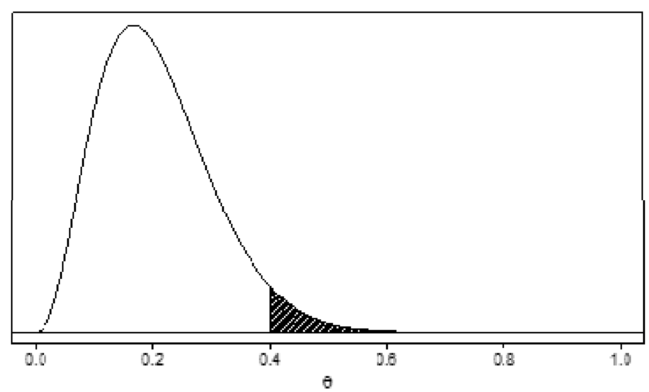
| Name | Thread pitch (mm) | Minor diameter tolerance | Nominal diameter (mm) | Head shape | Price for 50 screws | Available at factory outlet? | Number in stock | Flat or Phillips head? |
|---|---|---|---|---|---|---|---|---|
| M4 | 0.7 | 4g | 4 | Pan | $10.08 | Yes | 276 | Flat |
| M5 | 0.8 | 4g | 5 | Round | $13.89 | Yes | 183 | Both |
| M6 | 1 | 5g | 6 | Button | $10.42 | Yes | 1043 | Flat |
| M8 | 1.25 | 5g | 8 | Pan | $11.98 | No | 298 | Phillips |
| M10 | 1.5 | 6g | 10 | Round | $16.74 | Yes | 488 | Phillips |
| M12 | 1.75 | 7g | 12 | Pan | $18.26 | No | 998 | Flat |
| M14 | 2 | 7g | 14 | Round | $21.19 | No | 235 | Phillips |
| M16 | 2 | 8g | 16 | Button | $23.57 | Yes | 292 | Both |
| M18 | 2.1 | 8g | 18 | Button | $25.67 | No | 664 | Both |
| M20 | 2.4 | 8g | 20 | Pan | $29.00 | Yes | 486 | Both |
| M24 | 2.55 | 9g | 24 | Round | $33.01 | Yes | 982 | Phillips |
| M28 | 2.7 | 10g | 28 | Button | $35.66 | No | 1067 | Phillips |
| M36 | 3.2 | 12g | 36 | Pan | $41.32 | No | 434 | Both |
| M50 | 4.5 | 15g | 50 | Pan | $44.72 | No | 740 | Flat |

prior

data

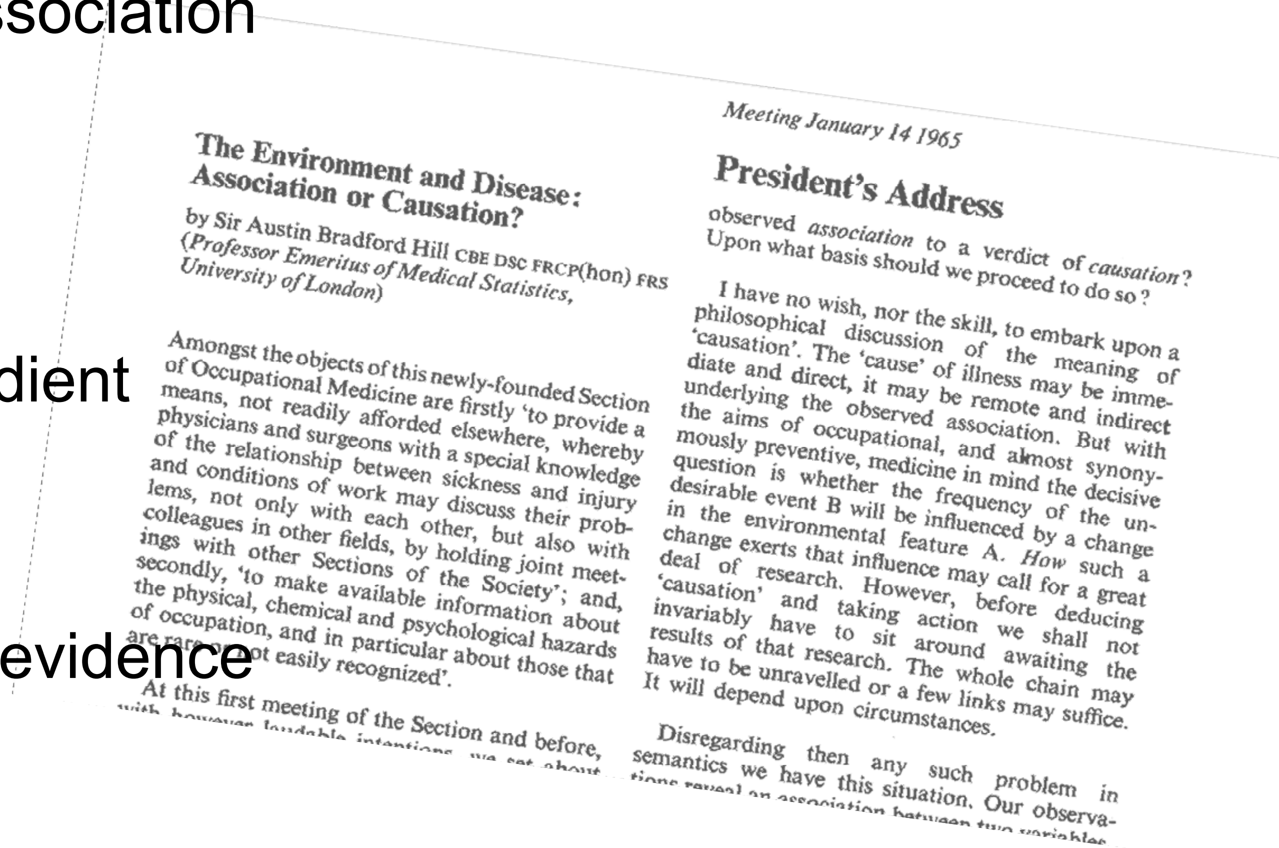


data



# Hill’s causality viewpoints

- Strength of association
- Consistency
- Specificity
- Temporality
- Biological gradient
- Plausibility
- Coherence
- Experimental evidence
- Analogy



Meeting January 14 1965

**The Environment and Disease: Association or Causation?**

by Sir Austin Bradford Hill CBE DSC FRCP(hon) FRS
(*Professor Emeritus of Medical Statistics, University of London*)

Amongst the objects of this newly-founded Section of Occupational Medicine are firstly ‘to provide a means, not readily afforded elsewhere, whereby physicians and surgeons with a special knowledge of the relationship between sickness and injury and conditions of work may discuss their problems, not only with each other, but also with colleagues in other fields, by holding joint meetings with other Sections of the Society’; and, secondly, ‘to make available information about the physical, chemical and psychological hazards of occupation, and in particular about those that are rare or not easily recognized’.

At this first meeting of the Section and before,

**President’s Address**

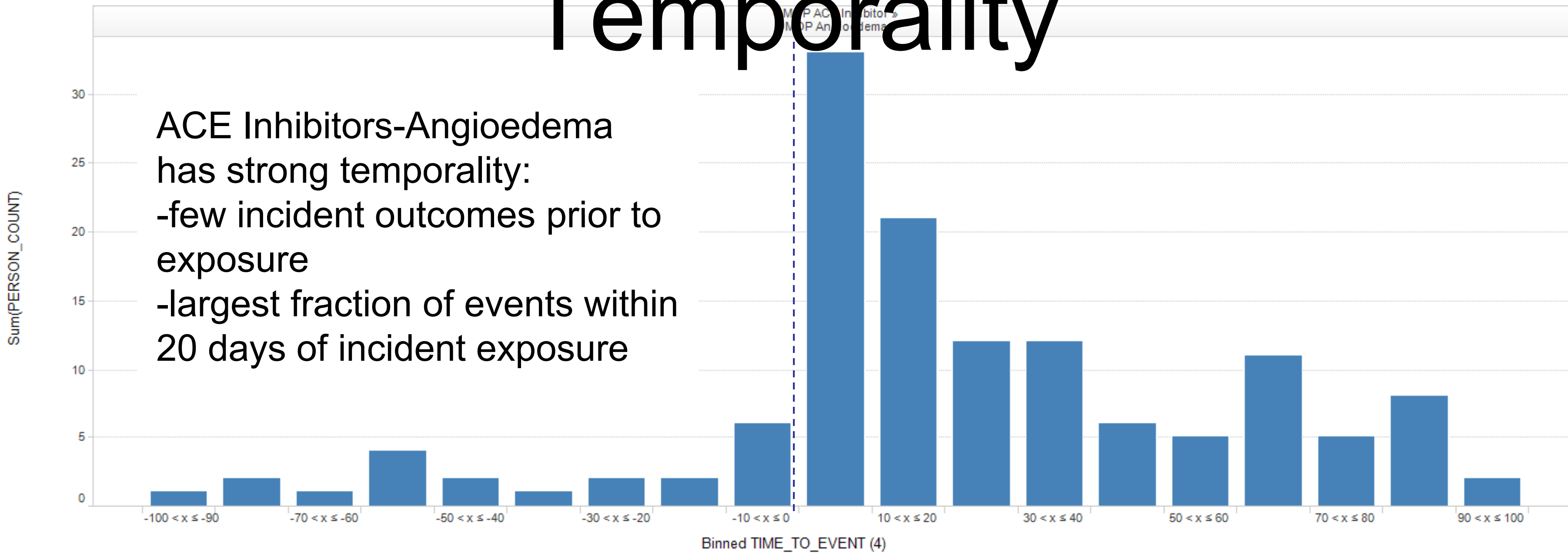
observed *association* to a verdict of *causation*? Upon what basis should we proceed to do so?

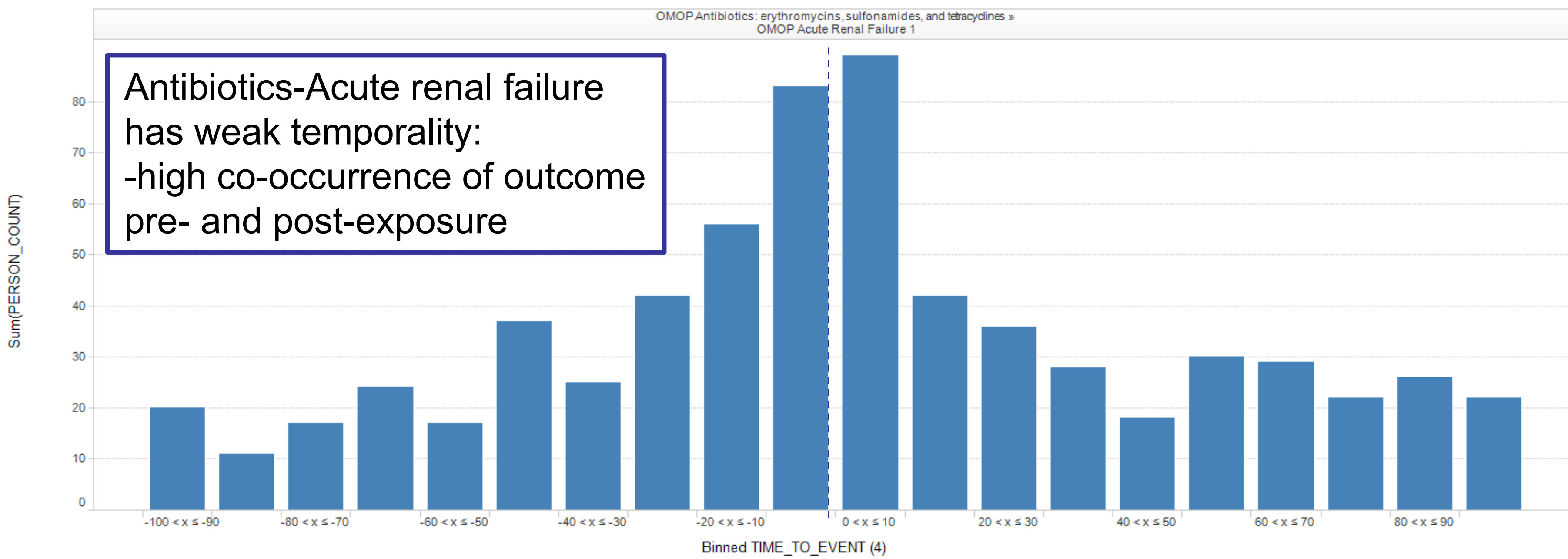
I have no wish, nor the skill, to embark upon a philosophical discussion of the meaning of ‘causation’. The ‘cause’ of illness may be immediate and direct, it may be remote and indirect underlying the observed association. But with the aims of occupational, and almost synonymously preventive, medicine in mind the decisive question is whether the frequency of the undesirable event B will be influenced by a change in the environmental feature A. *How* such a change exerts that influence may call for a great deal of research. However, before deducing ‘causation’ and taking action we shall not invariably have to sit around awaiting the results of that research. The whole chain may have to be unravelled or a few links may suffice. It will depend upon circumstances.

Disregarding then any such problem in semantics we have this situation. Our observa-

Austin Bradford Hill, “The Environment and Disease: Association or Causation?,” *Proceedings of the Royal Society of Medicine*, 58 (1965), 295-300.

# Temporality

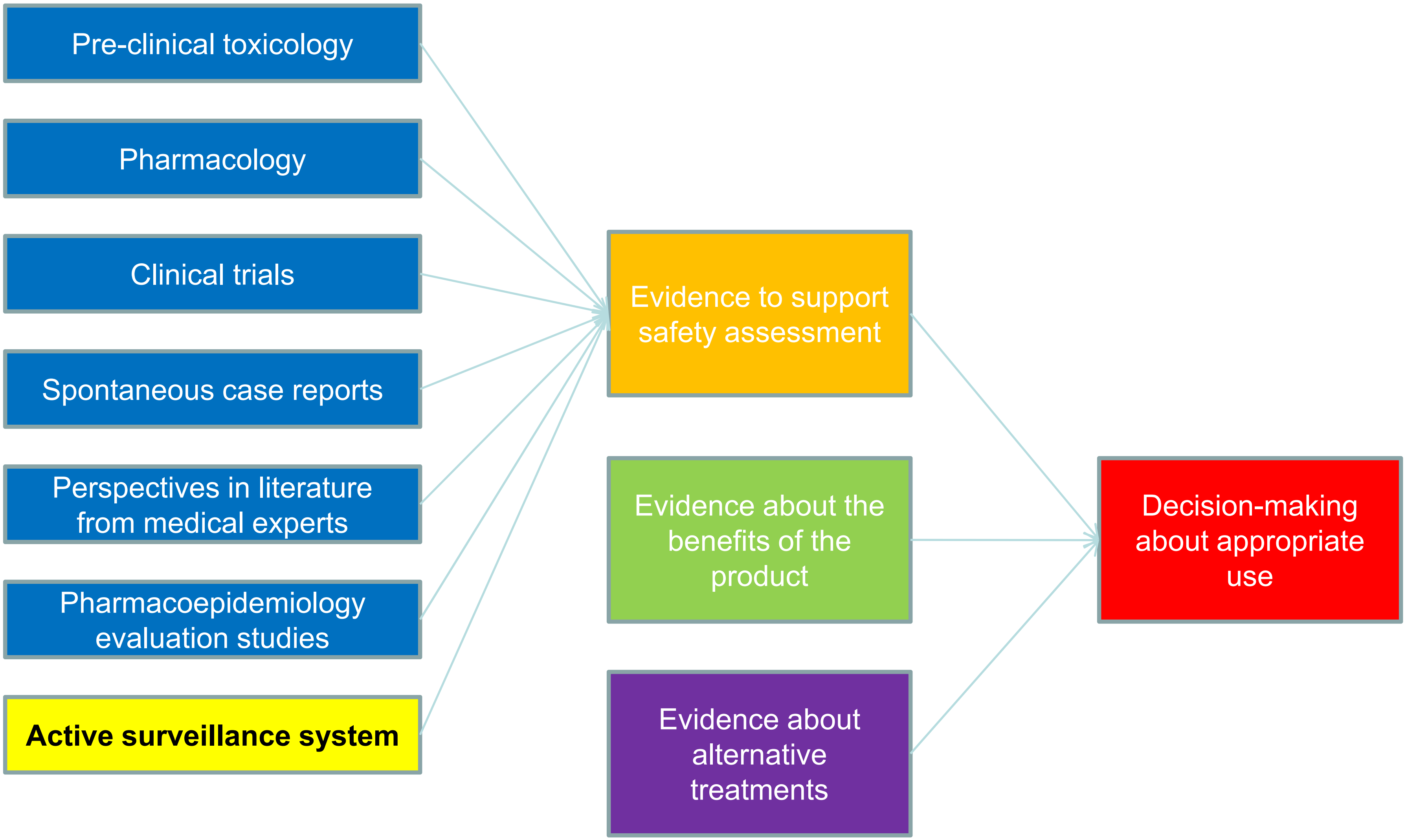


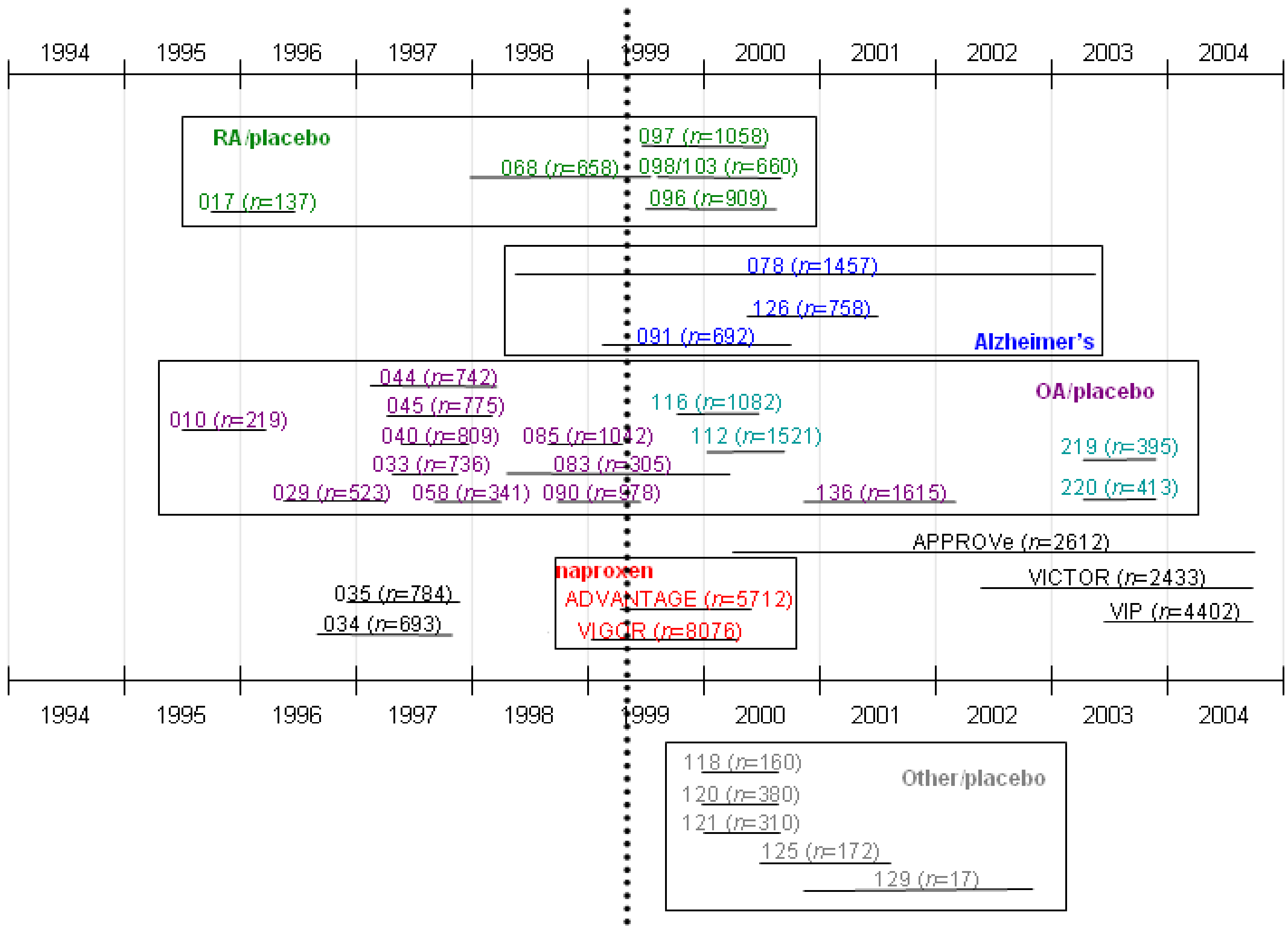
# Harnessing Hill

- Previously p(true | RR, SE)
  - Logistic regression with 2 predictors
- Using Hill: p(true | RR, SE, temporality, coherence, consistency, etc.)
  - Logistic regression with many predictors
- Thus we have a framework to formally integrate diverse evidence into the causal judgment

# Active surveillance: One additional piece of evidence to inform medical decision-making



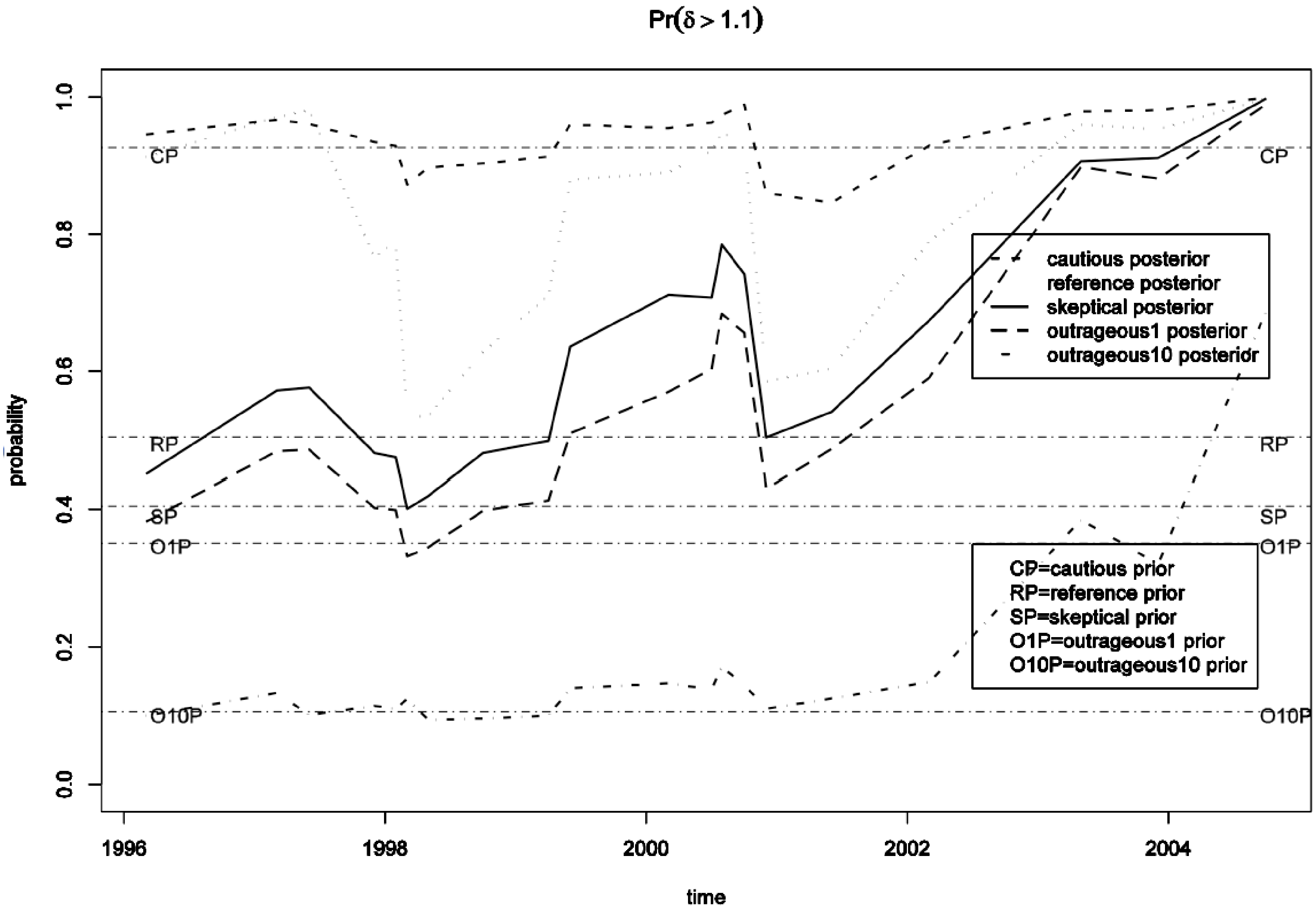
# A Case Study: Trials for Vioxx - Timeline



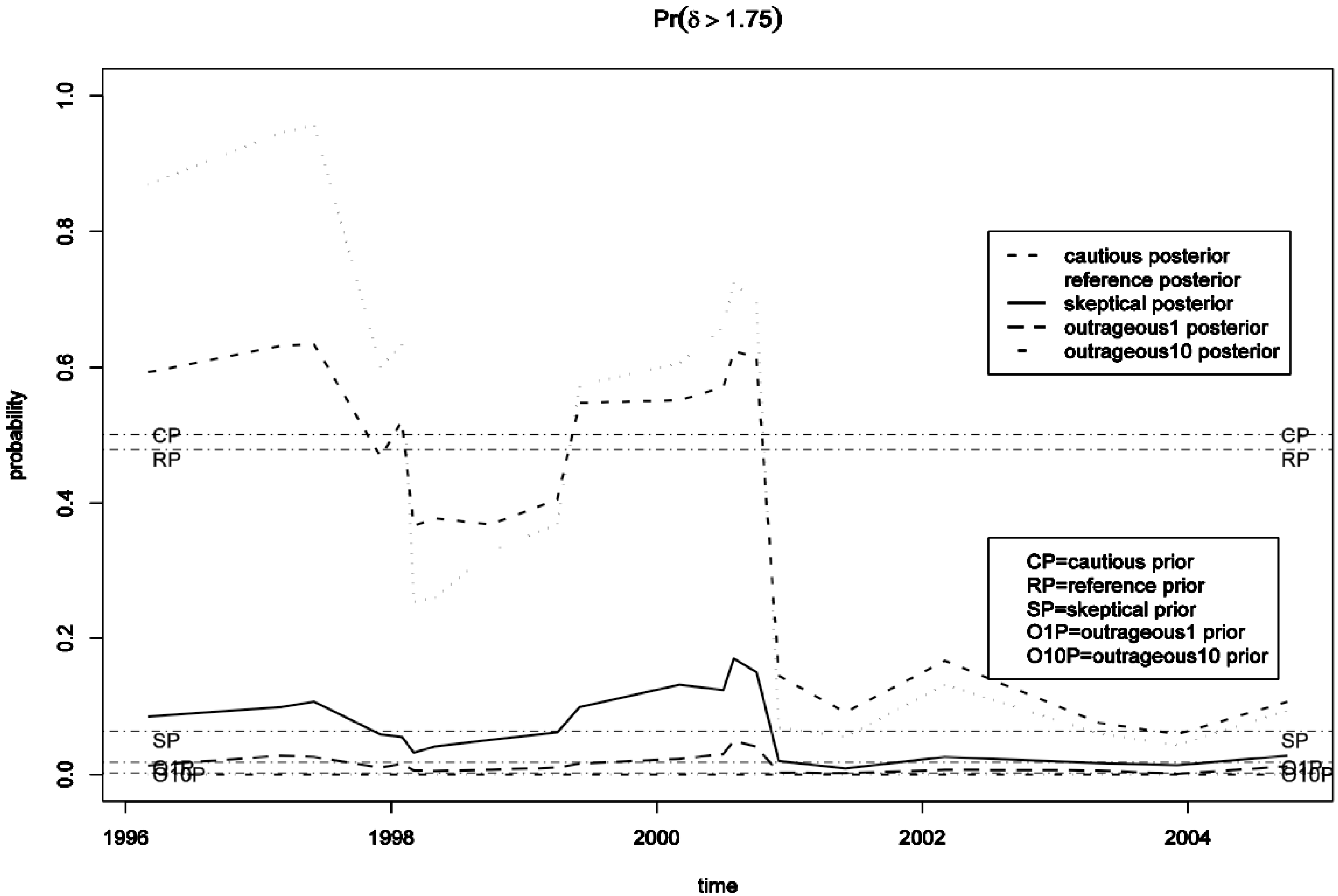
# Case Study: CVT Events

| | | | V | | Placebo | |
|---|---|---|---|---|---|---|
| Study Block | Study No. | LPO | Events | PYR | Events | PYR |
| RA | 017 | 5/21/1997 | 1 | 8 | 0 | 7 |
| | 068 | 9/10/1998 | 1 | 49 | 0 | 24 |
| | 096 | 7/21/2000 | 4 | 97 | 0 | 58 |
| | 097 | 6/6/2000 | 0 | 137 | 0 | 62 |
| | 098 | 7/6/2000 | 0 | 11 | 1 | 12 |
| | 103 | 7/6/2000 | 0 | 44 | 0 | 45 |
| | All RA | | 6 | 345 | 1 | 208 |
| OA | 010 | 2/8/1996 | 2 | 16 | 0 | 7 |
| | 029 | 2/5/1997 | 3 | 46 | 0 | 16 |
| | 033 | 11/18/1997 | 1 | 66 | 1 | 9 |
| | 040 | 1/1/1998 | 2 | 72 | 0 | 11 |
| | 044 | 2/18/1998 | 3 | 154 | 0 | 52 |
| | 045 | 2/18/1998 | 2 | 157 | 3 | 61 |
| | 058 | 4/1/1998 | 1 | 21 | 0 | 6 |
| | 083 | 2/9/2000 | 0 | 21 | 0 | 21 |
| | 085 | 3/3/1999 | 1 | 61 | 0 | 28 |
| | 090 | 5/17/1999 | 5 | 56 | 0 | 27 |
| | 112 | 9/8/2000 | 0 | 104 | 0 | 15 |
| | 116 | 6/22/2000 | 1 | 54 | 0 | 15 |
| | 136 | 2/5/2002 | 1 | 95 | 1 | 201 |
| | 219 | 11/28/2003 | 0 | 18 | 1 | 8 |
| | 220 | 11/24/2003 | 0 | 18 | 0 | 8 |
| | All OA | | 22 | 959 | 6 | 485 |
| OA/RA | All OA/RA | | 28 | 1304 | 7 | 692 |
| ALZ | 078 | 4/23/2003 | 75 | 1623 | 52 | 1762 |
| | 091 | 11/30/2000 | 13 | 369 | 14 | 381 |
| | 126 | 5/30/2001 | 11 | 193 | 7 | 197 |
| | All ALZ | | 99 | 2185 | 73 | 2341 |
| APP | 122 | 9/8/2004 | 75 | 5700 | 49 | 5828 |

# Posterior Probability (1) – block effect model



# Posterior Probability (2) – block effect model



# Sequential Uber-Analysis

- Sequential learning
- Silos: RCTs, observational data, spontaneous reports, pharmacology
- Little knowledge transfer
- Bayesian approach provides a snapshot at any point in time that reflects *all* available knowledge and data

# Think about this…

| | *Hospital* | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **A** | **B** | **C** | **D** | **E** | **F** | **G** | **H** | **I** | **J** | **K** | **L** |
| **No. of ops.** $n$ | 27 | 148 | 119 | 810 | 211 | 196 | 148 | 215 | 207 | 97 | 256 | 360 |
| **No. of deaths** $r$ | 0 | 18 | 8 | 46 | 8 | 13 | 9 | 31 | 14 | 8 | 29 | 24 |

**Denote by $\theta$ the probability that the next operation in Hospital A results in a death**

**Use the data to estimate (i.e., guess the value of) $\theta$**

# Some Theory (Bayes works!)

**Consider observations from a normal distribution:**

$$y_1,\ldots,y_n \sim N(\mu,1)$$

**Could use the sample mean to estimate *μ*:**

$$\bar{y} = \frac{y_1 + \ldots + y_n}{n}$$

**"unbiased"**

**Suppose you want to estimate *μ* with *δ(y)* to minimizes the expected squared loss:**

$$\underset{y|\mu}{E}\left[(\delta(y)-\mu)^2\right]$$

Bayesian has to specify a prior distribution for $\mu$, e.g.:

$$N(\theta, 10)$$

Bayesian would typically set $\delta(y)$ to be the posterior mean

$$c\theta + (1-c)\bar{y}$$

Can easily prove that, e.g., for $n$=30, the Bayesian posterior mean has smaller expected squared loss so long as:

$$|\theta - \mu| < 134.3$$

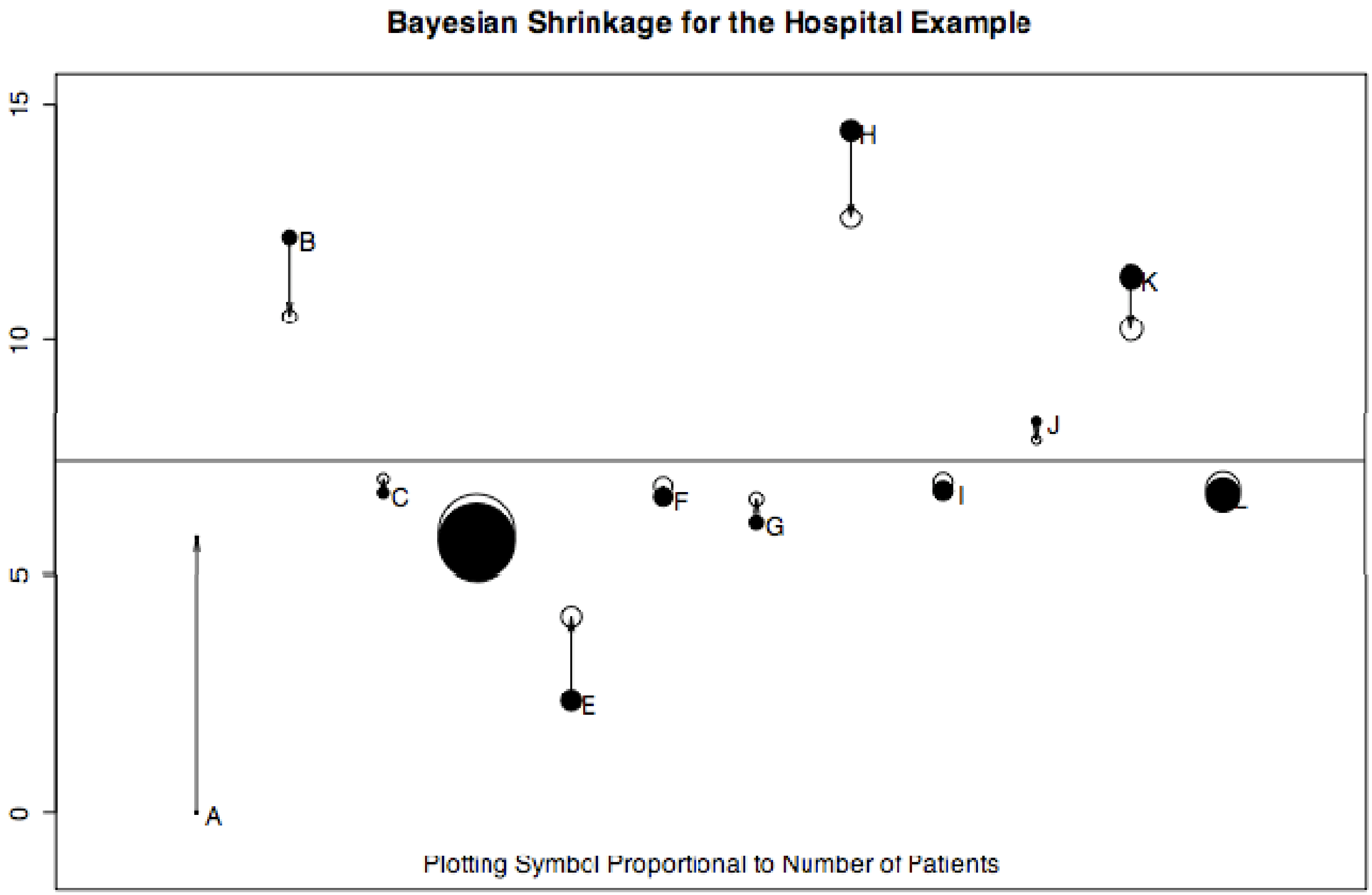
# Back to the hospital data

| | *Hospital* | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | A | B | C | D | E | F | G | H | I | J | K | L |
| No. of ops. $n$ | 27 | 148 | 119 | 810 | 211 | 196 | 148 | 215 | 207 | 97 | 256 | 360 |
| No. of deaths $r$ | 0 | 18 | 8 | 46 | 8 | 13 | 9 | 31 | 14 | 8 | 29 | 24 |

**Denote by $\theta_i$ the probability that the next operation in Hospital *i* results in a death**

**Assume $\theta_i \sim$ beta(*a*,*b*)**

**Compute joint posterior distribution for all the $\theta_i$ simultaneously**

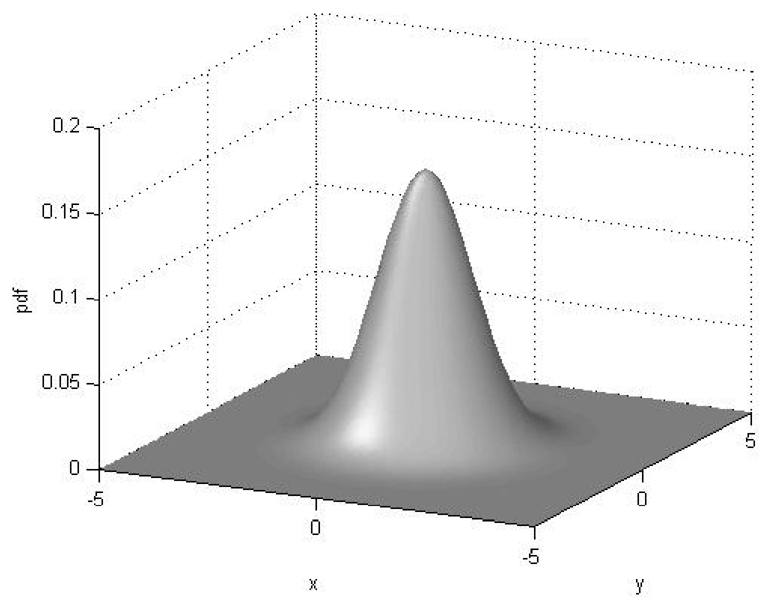

Bayesian Shrinkage for the Hospital Example
15
10
5
0
A
B
C
E
F
G
H
I
J
K
L
Plotting Symbol Proportional to Number of Patients

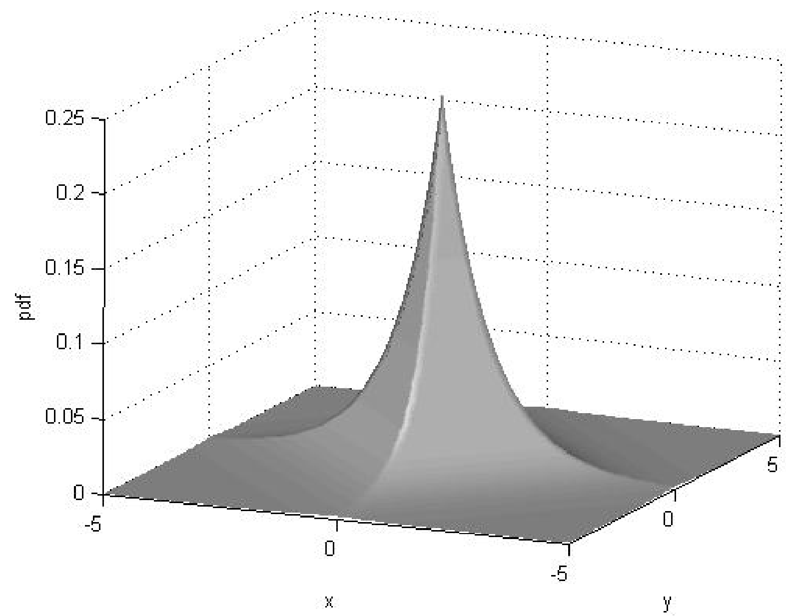
# Logistic Regression

**Linear model for log odds of category membership:**

$$\log \frac{p(y=1|\boldsymbol{x}_i)}{p(y=-1|\boldsymbol{x}_i)} = \sum b_j x_{ij} = \boldsymbol{b}\boldsymbol{x}_i$$

# Bayesian Perspective

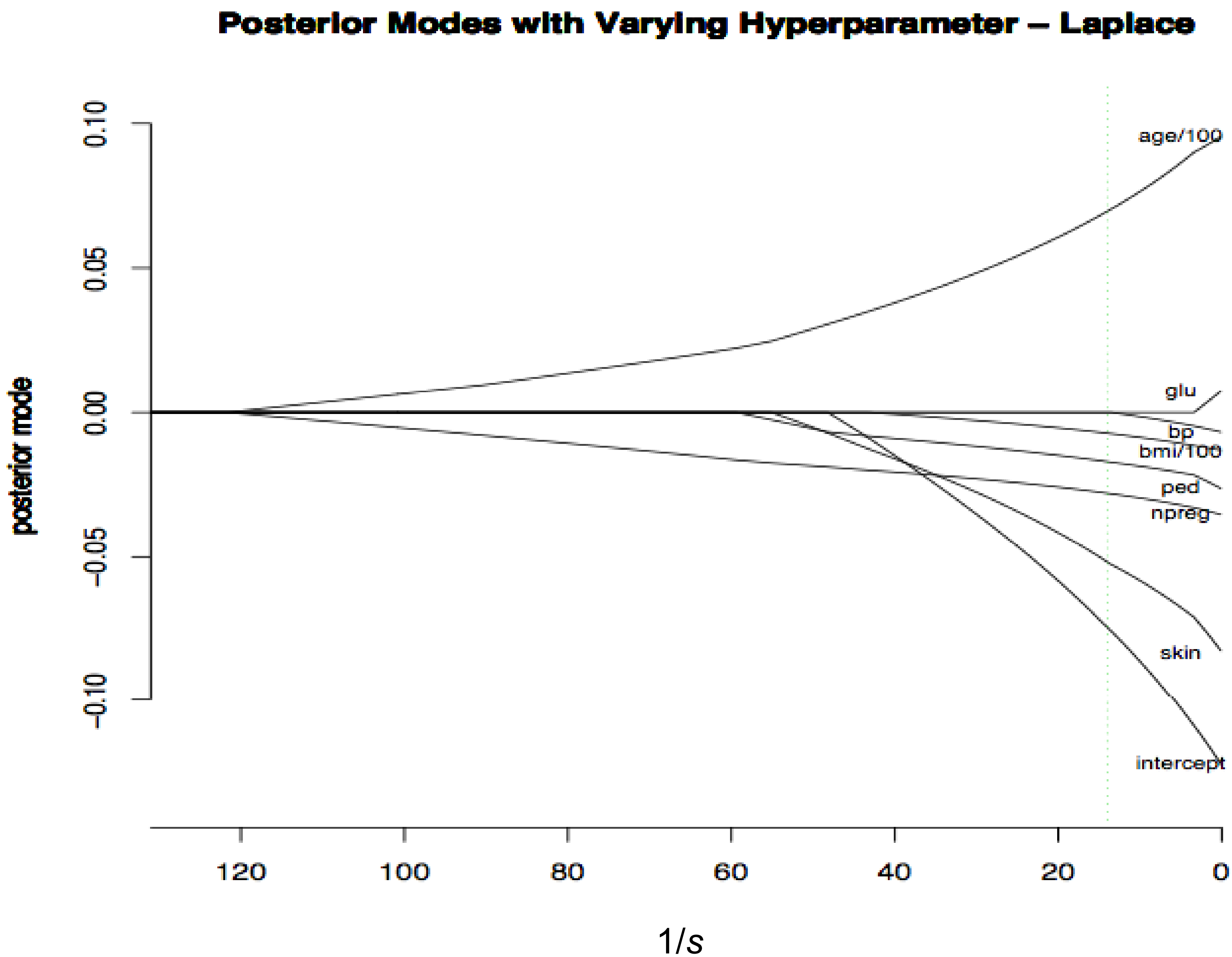



$$\beta_j \sim N(0, \tau^2)$$

$$\beta_j \sim N(0, \tau_j^2)$$

$$\tau_j^2 \sim \exp(\gamma)$$


Posterior Modes with Varying Hyperparameter – Laplace
posterior mode
0.10
0.05
0.00
−0.05
−0.10
age/100
glu
bp
bmi/100
ped
npreg
skin
intercept
120
100
80
60
40
20
0
1/s

# Family of Priors

- **Two thresholds for rate ratio of an adverse event**
  - $\delta_u$: the drug should not be on market if ratio $>\delta_u$
    - set $\delta_u = 1.75$ expert opinion,
  - $\delta_L$: the drug should be on market if ratio $<\delta_L$
    - set $\delta_L = 1.1$ weigh= the risk against the benefit.

✓ **Skeptical Prior**: Gaussian with $\mu = 0$ and $P(\delta > \delta_U) = 5\%$
- Strong prior belief that the drug is not dangerous

✓ **Cautious Prior**: Gaussian with $\mu = \delta_u$ and $P(\delta < \delta_L) = 5\%$
- Strong prior belief that the drug is not safe

✓ **Reference Prior**: Gaussian with $\mu = 1$, $\sigma = 10^3$

✓ **Outrageous10 Prior**: Gaussian with $\mu = 1$ and $P(\delta > \delta_L) = 10\%$

✓ **Outrageous1 Prior**: Gaussian with $\mu = 1$ and $P(\delta > \delta_U) = 1\%$