

Bayesian and Hierarchical **Models for Health Policy** **Research**

Cindy L. Christiansen, Ph.D., Boston
University SPH, Health Services Dept. and
CHQOER VA HSR&D, Bedford, MA

Outline

- Vocabulary
- Non-informative prior – Profile of VA hospitals return to the ICU rate
- Subjective prior – Gusto example

Outline continued

- References and Software (handout)
- Copies of articles

Vocabulary

- Statistic
 - Calculated from data
 - Example: “x-bar” or the average
- Parameter
 - A characteristic of a defined population
 - Unknown quantity
 - Usually a Greek letter: ρ , μ , θ

Vocabulary continued

- Coefficient of variation (CV)
= (100)(standard deviation) / mean
 - A measure of spread
 - Standard deviation (SD) relative to the mean
 - Unit-free
 - CV = 20.0 means the SD is about 20% of the mean

Vocabulary continued

- Heterogeneity
 - Over-dispersion
 - Unexplained variation between groups
 - Extra variation

Vocabulary continued

- Frequentist methods
 - Traditional
 - Statistics can have distributions
 - Parameters are fixed, unknown quantities
 - Inference is about observed data given assumptions about parameter values
 - Includes P-values, confidence intervals, hypothesis testing

Vocabulary continued

- Frequentist methods
 - P-value
 - = Probability of the data given the null hypothesis
 - = $P(\text{observed data} \mid \text{the null hypothesis assumption about the parameter})$
 - Confidence interval
 - A range of values calculated from the sample that are thought to contain the true parameter value

Vocabulary continued

- Frequentist methods

- Confidence interval (CI)

A range of values calculated from the sample that are thought to contain the true parameter value

A 95% CI implies that, were the estimation process repeated again and again and again,

then 95% of the calculated intervals would be expected to contain the true parameter value

Vocabulary continued

- Bayesian methods
 - Typically not taught
 - Parameters have distributions
 - Parameters are unknown quantities; Data are known quantities
 - Inference is about unknown quantities given known quantities (probability statements)
 - Includes posterior distributions, Bayes factor

Vocabulary continued

- Bayesian methods
 - Probability statement
 - = Probability the unknown parameter is some value given the observed data
 - = $P(\mu \mid \text{data})$
 - Probability interval (PI)
 - A 95% PI implies that the calculated interval contains the true parameter value with probability .95

Vocabulary continued

- Bayes Theorem

$$P(\mu | \text{data}) = \frac{P(\text{data} | \mu) P(\mu)}{P(\text{data})}$$

= probability distribution for the parameter
of interest given the observed data

$$\text{Posterior prob.} = \frac{\text{(Likelihood function)}(\text{Prior})}{\text{(Marginal distribution)}}$$

Vocabulary continued

- Bayesian model

Inference involves probability statements about parameters of interest

1. Prior information:

yes = informative prior, no = non-informative prior

2. Estimation method:

Approximate methods, Monte Carlo Markov Chain methods, REML, others

Vocabulary continued

- Hierarchical models include
 - Bayesian models
 - Multilevel models
 - Mixture models
 - Models with clusters
- Hierarchical models have more than one estimate of variance

Non-Informative Prior

Example, Profiling Hospitals

- References
 - Christiansen CL, Morris CN. **Improving the Statistical Approach to Health Care Provider Profiling.** *Annals of Internal Medicine.* 127(8):764-768, October 15, 1997.
 - Burgess JF, et.al., Medical Profiling: Improving standards and risk adjustments using hierarchical models. *Journal of Health Economics* 19 (2000) 291-309

P-values *versus* (Bayesian) probability statements

- For medical profiling, what question would we like to answer?
 - What is the probability that the hospital's *observed rate* was $\geq 4.3\%$ if the *hospital's true rate* was the overall sample rate (2.0%)?
 - What is the probability that a *hospital's true rate* $> \underline{3\%}$ given the hospital's *observed rate* was 4.3% (and knowing the observed rates at other hospitals)?

P-values *versus* (Bayesian) probability statements

- Probability(observed data is some value or greater *given* we know the value of the parameter of interest)
= $P(\text{data} \mid \mu)$
- Probability (the parameter of interest is some value or greater *given* we observed the data)
= $P(\mu \mid \text{data})$

VA's 1995 Return to the ICU data

- 148 Hospitals
- Return to the ICU is a rare event (Poisson)
- Observed (unadjusted) rates vary from 0 to 4.3%; Median value is 1.9%
- # of patients varies by a factor of 30
- Decision to control for hospital type and for case-mix

Two Standard Methods

- Method 1.
 - Null model:
 - Observed/expected \sim Normal (1, 1/expected)
individually for every hospital
 - Z-score and P-values calculated and outliers are flagged (usually) if P-value $< .025$ or $> .975$

Two Standard Methods

- Problems with Method 1.
 - Normal approximation, could use exact distribution
 - Hospitals usually are compared with the population mean without thought given to this choice
 - P-values are indirect measurements of performance

Two Standard Methods

- Method 2.
 - Binomial or Poisson regression is fit to the data
 - Residuals are used to profile and compare hospitals

Two Standard Methods

- Problems with Method 2.
 - Residuals are not probabilities
 - Poisson and Binomial residuals cannot be interpreted the same as those from a Normal distribution
 - Model does not account for unexplained heterogeneity so precision is overstated

Improvements to Analysis Using Hierarchical Models (HM)

- HM accounts for regression-to-the-mean
- Provides probability statements about each hospital's true performance
- Permits case-mix adjustments (as does Method 2)
- Model pools information across hospitals – helps with analysis of small hospitals

Improvements to Analysis Using Hierarchical Models (HM)

- Compromise (usually determined by the data) between observed rate and regression or population mean
- HM estimate equals
$$= (1-\text{shrinkage})(\text{observed rate})$$
$$+ (\text{shrinkage})(\text{population rate})$$
$$= (1-B)(\text{obs rate}) + B\mu$$

Features of the Hierarchical Model (HM)

- Shrinkage = ratio of the within variance to the total variance
- Shrinkage is different for each hospital because of different number of patients and case-mix
- Information is pooled across sites while allowing individual observed rates to ‘speak’

Criteria matters graphs (see figure 1 from Burgess, et.al
paper)

Graph of posterior distributions of some hospitals (see figure 2 from Burgess et.al paper)

Comparing P-Values and Posterior Probabilities for 4 ICU Units

Data				Adjusted for Hospital Type		
# of Patients	obs	expect	obs/expect	P-Value	Posterior Probabilities CV = ∞ CV estimated (20.5)	
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Possible Substandard Hospitals:					P ($\rho < 1$)	P ($\rho < 1$)
459	14	10.0	1.4	.087	.087	.260
1294	42	32.0	1.3	.091	.091	.142
Possible Exemplary Hospitals:					P ($\rho > 1$)	P ($\rho > 1$)
418	4	9.4	0.4	.031	.031	.146
1275	26	34.9	0.7	.051	.051	.097

CV = coefficient of variation for true ratios

ρ = true performance ratio

The New England Journal of Medicine

Volume 329(10)

2 Sep 1993

pp 673-682

An International Randomized Trial Comparing Four Thrombolytic Strategies For Acute Myocardial Infarction.

[Original Articles]

The GUSTO Investigators.

OUTCOME	STREPTOKINASE AND SUBCUTANEOUS HEPARIN (N = 9796)	STREPTOKINASE AND INTRAVENOUS HEPARIN (N = 10,377)	ACCELERATED t-PA AND INTRAVENOUS HEPARIN (N = 10,344)	BOTH THROMBOLYTIC AGENTS AND INTRAVENOUS HEPARIN (N = 10,328)	P VALUE, ACCELERATED t-PA vs. BOTH STREPTOKINASE GROUPS
	<i>percent of patients</i>				
24-hr mortality	2.8	2.9	2.3	2.8	0.005
30-day mortality	7.2	7.4	6.3	7.0	0.001
Or nonfatal stroke	7.9	8.2	7.2	7.9	0.006
Or nonfatal hemorrhagic stroke	7.4	7.6	6.6	7.4	0.004
Or nonfatal disabling stroke	7.7	7.9	6.9	7.6	0.006

EVENT RATE (%)
STREPTO-
KINASE t-PA

Odds Ratio and 95% CI

30-day mortality

Streptokinase and SC heparin

7.2 6.3



Streptokinase and IV heparin

7.4 6.3



Both streptokinase groups

7.3 6.3



30-day mortality or disabling stroke

Streptokinase and SC heparin

7.7 6.9



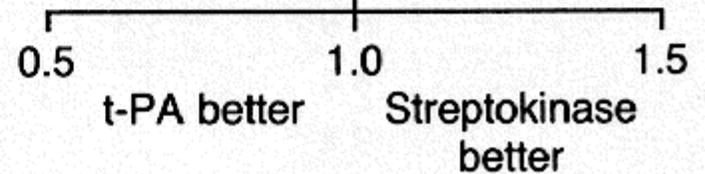
Streptokinase and IV heparin

7.9 6.9



Both streptokinase groups

7.8 6.9



GUSTO trial – example of subjective analysis

- References:

- Hively W, The Mathematics of Making Up Your Mind, Discover, May 1996, 90-97.

- <http://208.245.156.153/archive/output.cfm?ID=753>

- Brophy JM, Joseph, L. **Placing Trials in Context Using Bayesian Analysis: GUSTO Revisited by Reverend Bayes.** *JAMA*. 273(11):871-875, Mar 15, 1995

JAMA, The Journal of the American Medical Association

Volume 273(11)

15 Mar 1995

pp 871-875

Placing Trials in Context Using Bayesian Analysis: GUSTO Revisited by Reverend Bayes

[Special Communications]

Brophy, James M.; Joseph, Lawrence

Table 1.—Data From GUSTO, GISSI-2, and ISIS-3*

Trial	Agent	No. of Patients	No. (%) of Deaths	No. (%) of Nonfatal Strokes	Combined Deaths or Strokes
GUSTO†	SK	20 173	1473 (7.3)	101 (0.5)	1574 (7.8)
	t-PA	10 343	652 (6.3)	62 (0.6)	714 (6.9)
GISSI-2	SK	10 396	929 (8.9)	56 (0.5)	985 (9.5)
	t-PA	10 372	993 (9.6)	74 (0.7)	1067 (10.3)
ISIS-3	SK	13 780	1455 (10.6)	75 (0.5)	1596 (11.6)
	t-PA	13 746	1418 (10.3)	95 (0.7)	1513 (11.0)

*SK indicates streptokinase; and t-PA, tissue-type plasminogen activator.

†The 10 374 patients who received both SK and t-PA are not included here.

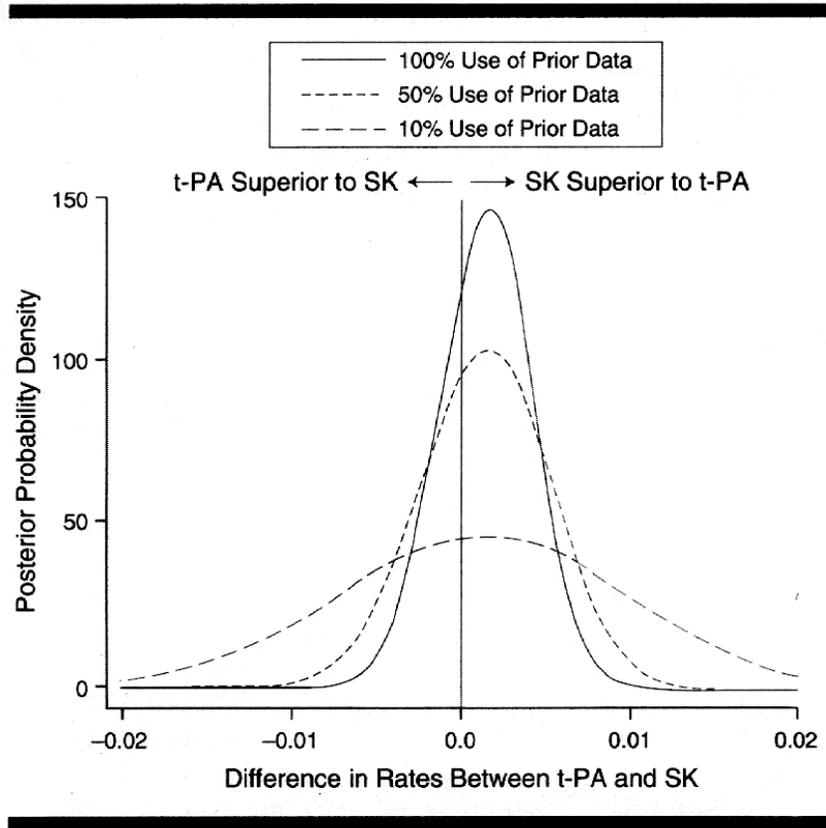


Figure 1. Plot of the prior distributions for the difference in mortality rates between tissue-type plasminogen activator (t-PA) and streptokinase (SK) using weights of 100%, 50%, and 10% of the GISSI-2 and ISIS-3 data, representing a range in prior beliefs in the relevance of these trials to the GUSTO trial. The area under the curve between any two points on the x-axis is the posterior probability that the difference in mortality rates lies between those limits. Numbers to the right of zero indicate the superiority of SK, while those to the left of zero indicate the superiority of t-PA

From: Brophy: JAMA, Volume 273(11).Mar 15, 1995.871-875

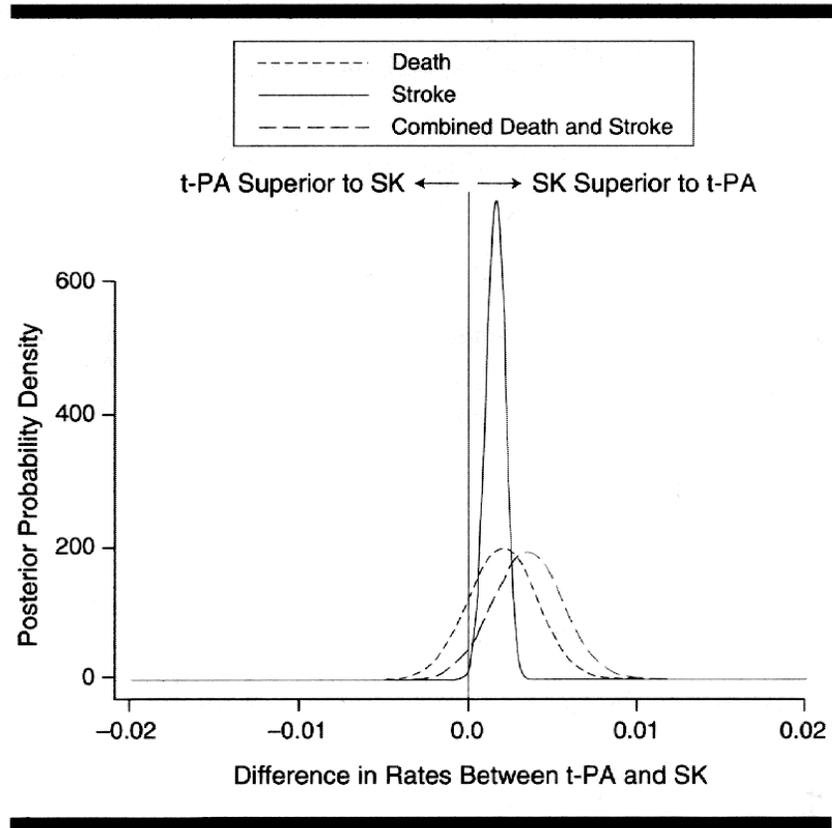


Figure 2. Plot of the posterior distribution for the difference in mortality, nonfatal stroke, and combined stroke and mortality rates between tissue-type plasminogen activator (t-PA) and streptokinase (SK), using data from the GUSTO trial, with full prior use of data from the GISSI-2 and ISIS-3 trials. The area under the curve between any two points on the x-axis is the posterior probability that the difference in rates lies between those limits. Numbers to the right of zero indicate the superiority of SK, while those to the left of zero indicate the superiority of t-PA

From: Brophy: JAMA, Volume 273(11).Mar 15, 1995.871-875

Table 2.—Probability of t-PA Superiority as a Function of Prior Belief in GISSI-2 and ISIS-3 Data After Consideration of the GUSTO Data*

Prior Belief in GISSI-2 and ISIS-3, %	Probability of t-PA Mortality Higher Than SK Mortality	Probability of t-PA Net Clinical Benefit Greater Than SK Benefit	Probability of t-PA Net Clinical Benefit Greater Than SK Benefit by at Least 1%
100	.17	.05	<.001
50	.44	.24	<.001
10	.98	.94	.03
0	.999	.998	.36

*See footnote to Table 1 for expansions of abbreviations. Net clinical benefit is the combined death and stroke rate.

Presenter Contact Information

Cindy L. Christiansen, Ph.D.

CHQOER & Boston University

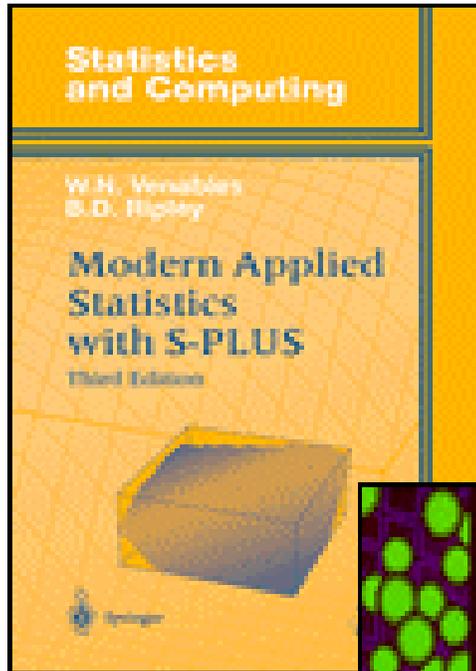
Edith Nourse Rogers Memorial Veterans Hospital

200 Springs Road (152)

Bedford, MA 01730

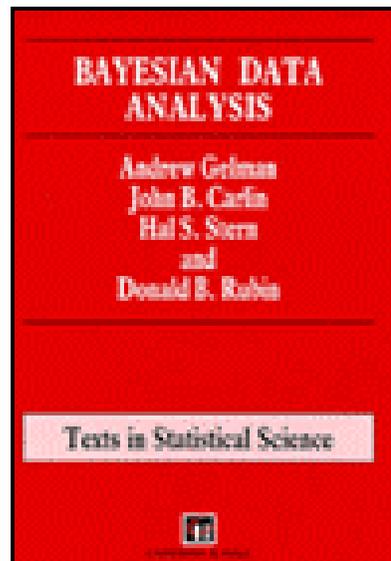
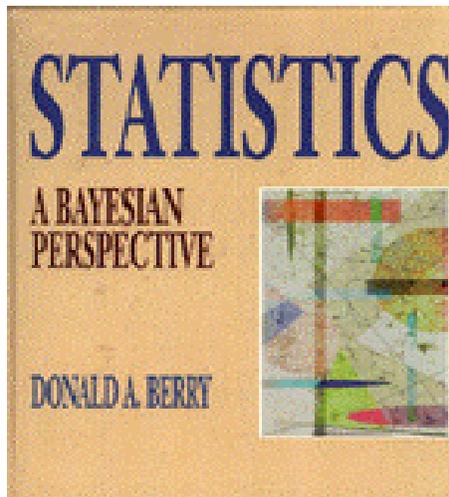
(781) 687-2915

CindyLC@bu.edu

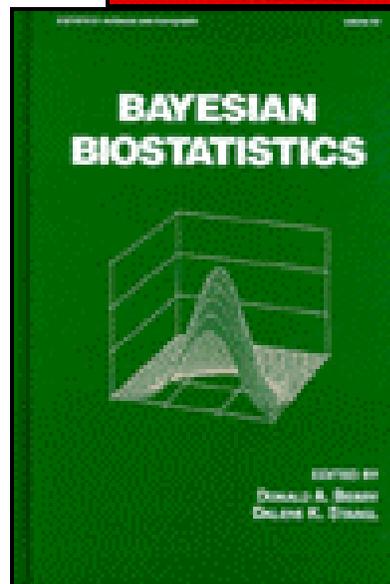
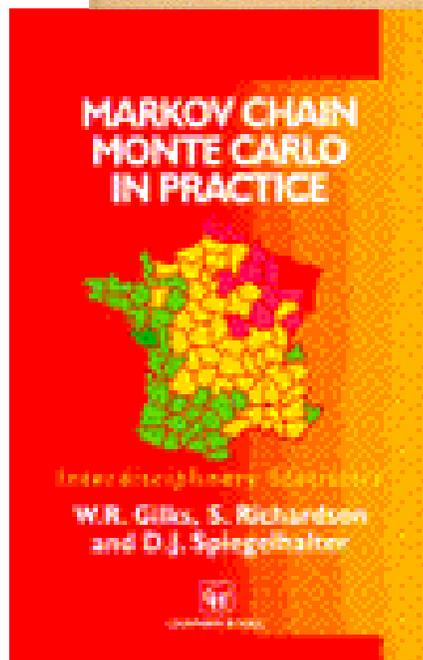


Software

- Proc MIXED
- GLIMMIX
- HLM
- ML3 and MLn
- VARCL
- S-PLUS
- BUGS & CODA

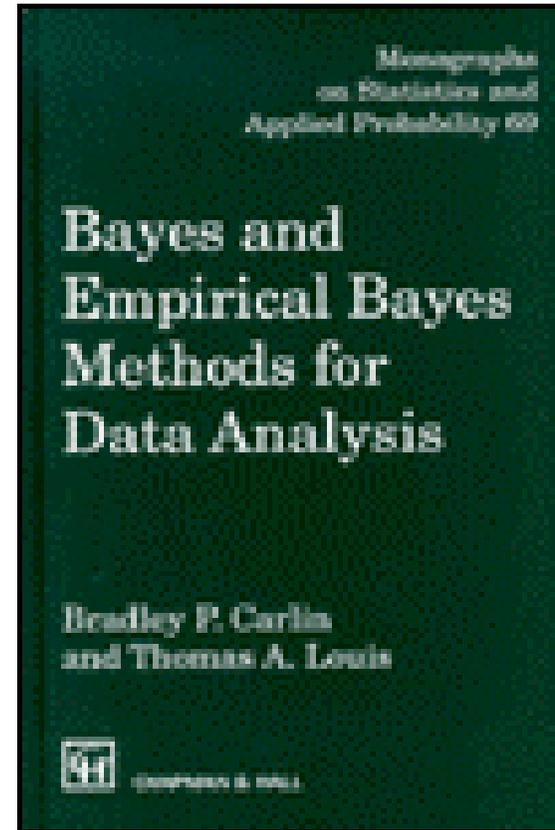


Some Books



Other Software and References

- **Listed in Carlin & Louis book**
- **StatLib website**
- **HLM website**
- **Multilevel website**
- **Zhou, et.al, 1999 Amer Stat (handout)**



<http://lib.stat.cmu.edu>

<http://www.ssicentral.com/hlm/hlmref.htm>

<http://www.ioe.ac.uk/multilevel/>

Websites of Some Bayesian and Hierarchical Modeling Researchers

- <http://www.bath.ac.uk/~masdd/>
– David Draper, Ph.D.
- <http://www.calstatela.edu/faculty/ikreft/>
– Ita Kreft, Ph.D.
- <http://gseweb.harvard.edu/~faculty/singer/>
– Judy Singer, Ph.D.
- <http://www.stat.duke.edu/-ds6e/>
– Dalene Stangl, Ph.D.

Additional references

- Lilford and Braunholtz, BMJ, 313(7057), 1996, 603-607
- Spiegelhalter e.al, 319 (7208), 1999, 508-512