

Frequentist



# A Comparison of Two Worlds

How does Bayes hold up to the  
status quo?

Alice Pressman (Kaiser, UC Berkeley)

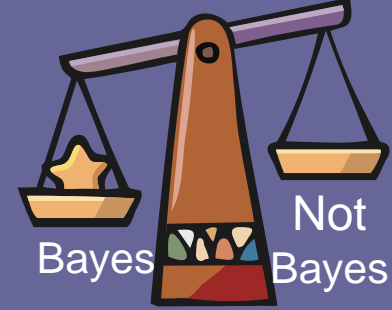
Alan Hubbard (UC Berkeley)

Andy Avins (Kaiser, UCSF)



Bayesian

# Introduction



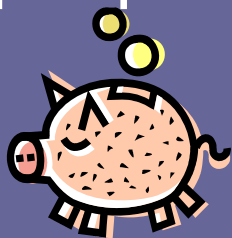
- Everyone **\*here\*** knows that Bayesian methods are superior

- The rest of the world isn't so sure

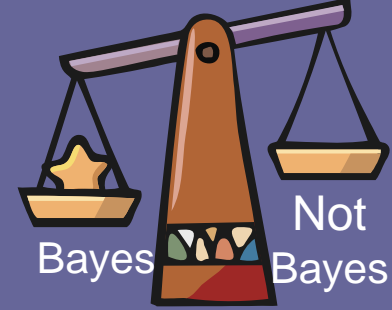
- I am not so sure



- The people who pay my salary aren't so sure



# Introduction




- Simulation suggests Bayesian methods are effective
- Practicing researchers may not be persuaded by “fake” data
- Use real data

# Questions

- Can we use these methods in non-device clinical trials?
- If we propose to use them, will we get funded??
- Can we offer evidence to the skeptics??



# Design

- Use existing REAL data from 2 completed non-device clinical trials
- Turn back the clock
- Perform the trials again comparing
  - Frequentist tried and true methods
  - Bayesian  methods

# Frequentist

- Group sequential methods
- 3 Interim analyses
  - O'Brien-Fleming
  - Lan-DeMets

Pete and Repeat sitting on a fence, Pete fell off, who was left?...Repeat

Pete and Repeat sitting on a fence, Pete fell off, who was left?...Repeat

Pete and Repeat sitting on a fence, Pete fell off, who was left?...Repeat

Pete and Repeat sitting on a fence, Pete fell off, who was left?...Repeat

Pete and Repeat sitting on a fence, Pete fell off, who was left?...Repeat

Pete and Repeat sitting on a fence, Pete fell off, who was left?...Repeat

Pete and Repeat sitting on a fence, Pete fell off, who was left?...Repeat

# Bayesian (simple)

- 3 interim analyses
- Conjugate prior distributions
- 85% probability stopping rule
- Two types of priors
  - Informative
  - Non-informative

Yesterday,  
All my troubles seemed so far away,  
Now it looks as though they're here to stay,  
Oh, I believe in yesterday.

Suddenly,  
I'm not half the man I used to be,  
There's a shadow hanging over me,  
Oh, yesterday came suddenly.

Why she  
Had to go I don't know, she wouldn't say.  
I said,  
Something wrong, now I long for yesterday.

Yesterday,  
Love was such an easy game to play,  
Now I need a place to hide away,  
Oh, I believe in yesterday.

Why she  
Had to go I don't know, she wouldn't say.  
I said,  
Something wrong, now I long for yesterday.

Yesterday,  
Love was such an easy game to play,  
Now I need a place to hide away,  
Oh, I believe in yesterday.

# The Trials



## I. SPINE

- The **S**timulation of **P**oints to **I**nvestigate **N**eedling **E**fficacy
- Cohort of ~300 men and women with low back pain
- Standard care versus acupuncture
- Outcome = change from baseline to 8 weeks in Roland Morris disability score (t-test)

# The Trials



## II. SOLVD-TT

- The **S**tudies **O**f **L**eft **V**entricular **D**ysfunction – **T**reatment **T**rial
- Cohort of ~2600 men and women with heart failure
- Enalapril versus placebo
- Outcome = all cause mortality (Cox PH)

# The Trials

## Published results



### I. SPINE

- Had no interim analyses in design
- Found significant result after all participants completed 8 weeks of study
- Mean difference (CI) = 2.47(1.40, 3.53)

(manuscript in press)

# The Trials

## Published results



## II. SOLVD-TT

- Lan-DeMets exponential spending function in place for trial. (Unknown number and spacing of interim analyses)
- Did not stop trial early
- At close, determined significantly lower mortality in treatment arm compared with placebo
- HR (95% CI) = 0.84 (0.73, 0.97)

# The Trials

## My results



### I. SPINE (published results: 2.47(1.40, 3.53) )

#### Frequentist O'Brien-Fleming

##### – Actual

- Significant at third interim look: 628 days into the 838-day study
- Difference at stopping time: 2.85 points (1.08, 4.63)

##### – Bootstrap

- Average stopping time: after collection of 72% of the data (corresponding to 608 days)
- Average difference: 2.69 points (1.34 – 4.01)
- 2% of simulated trials never reached significance

# The Trials

## My results



### I. SPINE (published results: 2.47(1.40, 3.53) )

## Frequentist Lan-DeMets

### – Actual

- Significant at third interim look: 628 days into the 838-day study
- Difference at stopping time: 2.85 (1.22, 4.58)

### – Bootstrap

- Average stopping time: after collection of 66% of the data (corresponding to 563 days)
- Average difference: 2.68 (1.35, 4.33)
- 2% of simulated trials never reached significance

# The Trials

## My results



### I. SPINE (published results: 2.47(1.40, 3.53) )

#### Bayesian Non-informative Prior

- Actual
  - Significant at third interim look: 628 days into the 838-day study
  - Difference at stopping time: 2.97(0.74) mean (SD)
- Type I error
  - 0.2%
- Bootstrap
  - Average stopping time: after collection of 84% of the data (corresponding to 704 days)
  - Average difference: 3.32 (0.14, 6.72)
  - 43% of simulated trials never reached significance

# The Trials

## My results



### I. SPINE (published results: 2.47(1.40, 3.53) )

#### Bayesian Informative Prior

- Actual
  - Significant at third interim look: 628 days into the 838-day study
  - Difference at stopping time: 2.85(0.68)
- Type I error
  - 0.3%
- Bootstrap
  - Average stopping time: after collection of 82% of the data (corresponding to 691 days)
  - Average difference: 3.37 (0.14, 6.68) mean (SD)
  - 43% of simulated trials never reached significance

# The Trials

## Comparison of Results



### I. SPINE (published results: 2.47(1.40, 3.53) )

Analysis	Avg Time	Avg Difference	Type I Error
Frequentist			
OBF	608 days	2.68	5.1%
LD	565 days	2.68	5.3%
Bayesian			
Non-Inform	704 days	3.32	0.2%
Informative	691 days	3.39	0.3%

# The Trials

## Published results



II. SOLVD-TT published results: (0.84 (0.73, 0.97) )

### Frequentist O'Brien-Fleming

#### – Actual

- Significant at final look: 1688 days into the 1688-day study
- Hazard ratio at stopping time: 0.84 (0.74 – 0.96)

#### – Bootstrap

- Average stopping time: after collection of 74% of the data (corresponding to 1254 days)
- Average hazard ratio: 0.79 (0.57 – 0.95)
- 22% of simulated trials never reached significance

# The Trials

## Published results



II. SOLVD-TT published results: (0.84 (0.73, 0.97) )

### Frequentist Lan - DeMets

#### – Actual

- Significant at final look: 771 days into the 1688-day study
- Hazard ratio at stopping time: 0.76 (0.56 – 0.99)

#### – Bootstrap

- Average stopping time: after collection of 63% of the data (corresponding to 1062 days)
- Average hazard ratio: 0.74 (0.51 – 0.95)
- 20% of simulated trials never reached significance

# The Trials

## Published results



II. SOLVD-TT published results: (0.84 (0.73, 0.97) )

### Bayesian Non-Informative Prior

#### – Actual

- Significant at final look: 1688 days into the 1688-day study
- Hazard ratio at stopping time: 0.83 (0.73, 0.94)

#### – Type I error

- 1%

#### – Bootstrap

- Average stopping time: after collection of 69% of the data (corresponding to 1165 days)
- Average hazard ratio: 0.78 (0.61, 0.95)
- 42% of simulated trials never reached significance

# The Trials

## Published results



II. SOLVD-TT published results: (0.84 (0.73, 0.97) )

### Bayesian Informative Prior

#### – Actual

- Significant at first look: 519 days into the 1688-day study
- Hazard ratio at stopping time: 0.71 (0.56, 0.90)

#### – Type I error

- 1.5%

#### – Bootstrap

- Average stopping time: after collection of 56% of the data (corresponding to 944 days)
- Average hazard ratio: 0.72 (0.51, 0.95)
- 34% of simulated trials never reached significance

# The Trials

## Comparison of Results



II. SOLVD-TT published results: (0.84 (0.73, 0.97) )

Analysis	Avg Time	Avg HR	Type I Error
Frequentist			
OBF	1254 days	0.79	5.4%
LD	1062 days	0.74	5.4%
Bayesian			
Non-Inform	1165 days	0.78	1.0%
Informative	944 days	0.72	1.5%

# What does it all mean?

1. For both trials, on average, frequentist methods dictated earlier stopping times than original trial.
2. Bayesian methods gave similar results as frequentist methods.
  - Cox models - type I error is higher
  - t-tests - effect size is larger

# What does it all mean?

3. Methods do not disagree appreciably for these two trials.
4. Bottom line: Easy to understand in simple situation. Quickly gets more complex as the models get more intricate, and no longer have closed form.

# Back to the original questions

- Can we use these methods in non-device clinical trials? **YES... ??in all cases??**
- If we propose to use them, will we get funded?? **???**
- Can we offer evidence to the skeptics??  
**YES!**

# What's next??

- Non-conjugate priors
  - Robust priors
  - MCMC methods
- Studies with different types of outcomes (non-normally distributed)
- Stopping early for futility

# References

## **1. Interim analyses in clinical trials: classical vs. Bayesian approaches**

Berry, DA

*Stat Med* 1985 Oct-Dec; 4(4):521-526

## **2. Bayesian Methods for Data Analysis, third edition**

Carlin BP, Louis TA

*Chapman and Hall/CRC Press* New York, New York, 2009

## **3. Group Sequential Method with Applications to Clinical Trials**

Jennison C, Turnbull BW

*Chapman and Hall/CRC Press* New York, New York, 2000

## **4. Group sequential methods in the design and analysis of clinical trials**

Pocock SJ

*Biometrika* 1977 64;2:191-199

## **5. A multiple testing procedure for clinical trials**

O'Brien & Fleming,

*Biometrics* 1979; 35, 549-556

## **6. Interim analysis: the alpha spending function approach.**

DeMets DL, Lan KK

*Stat Med* 1994 Jul15-30; 13(13-14):1341-1352

## **7. Effect of enalapril on survival in patients with reduced left ventricular ejection fractions and congestive heart failure. The SOLVD Investigators.**

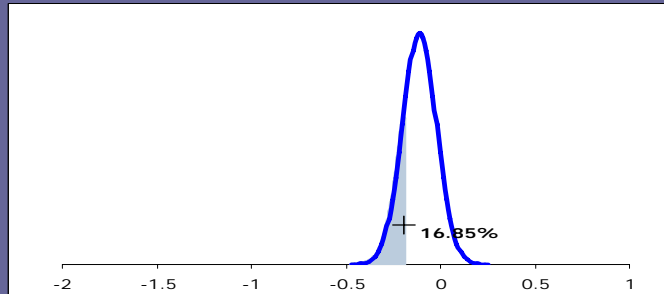
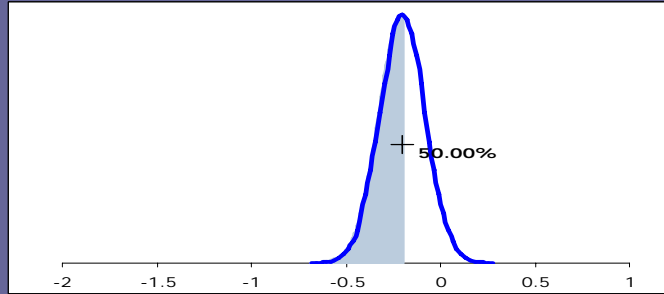
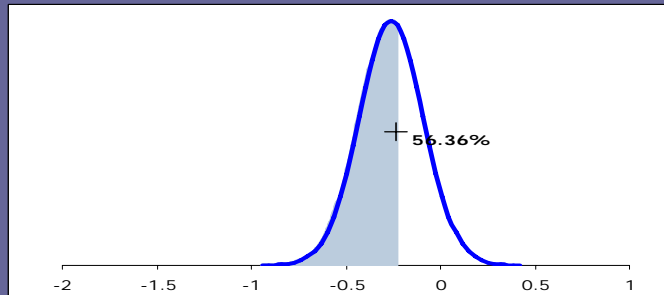
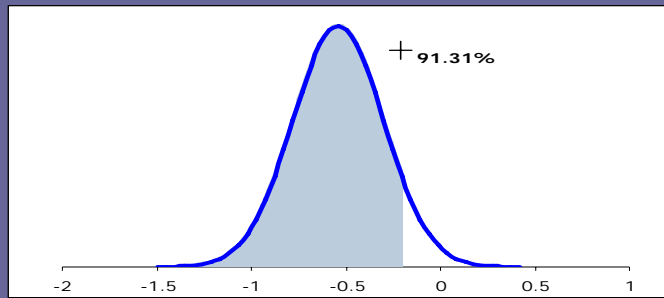
*N Engl J Med.* 1991;325:293-302.

## **8. Efficacy of acupuncture for chronic low back pain: protocol for a randomized controlled trial**

Daniel C Cherkin,<sup>1</sup> Karen J Sherman,<sup>1</sup> Charissa J Hogeboom, Janet H Erro,<sup>1</sup> William E Barlow,<sup>1,2</sup> Richard A Deyo,<sup>3</sup> and Andrew L Avins<sup>4</sup>

*Trials.* 2008; 9: 10. Published online 2008 February 28. doi: 10.1186/1745-6215-9-10.

# Posterior Probability Distributions



Actual Trial data  
SOLVD-TT

TRIAL  
START



Bayes  
inf  
prior



Bayes  
non-inf  
prior



Frq  
OBF, LD



END