No Large Differences Among Centers in a Multi-Center Neurosurgical Clinical Trial

Emine O Bayman[1,2], K Chaloner[2,3],
BJ Hindman[1] and MM Todd [1]

1:Anesthesia, 2:Biostatistics, 3: Stat and Actuarial Sc,
The Univ of Iowa.
e-mail: emine-bayman@uiowa.edu

Outline

- Outcome differences among centers in multi-center clinical trials.
- Why an experimental therapy or other intervention was effective in one center but not in another.
- Healthcare management: identify medical centers with superior outcomes.

- Use Intraoperative Hypothermia for Aneurysm Surgery Trial (IHAST) to demonstrate the application of Bayesian methods to determine if outcome differences exist among centers, and if differences in center-specific practices affected outcomes.

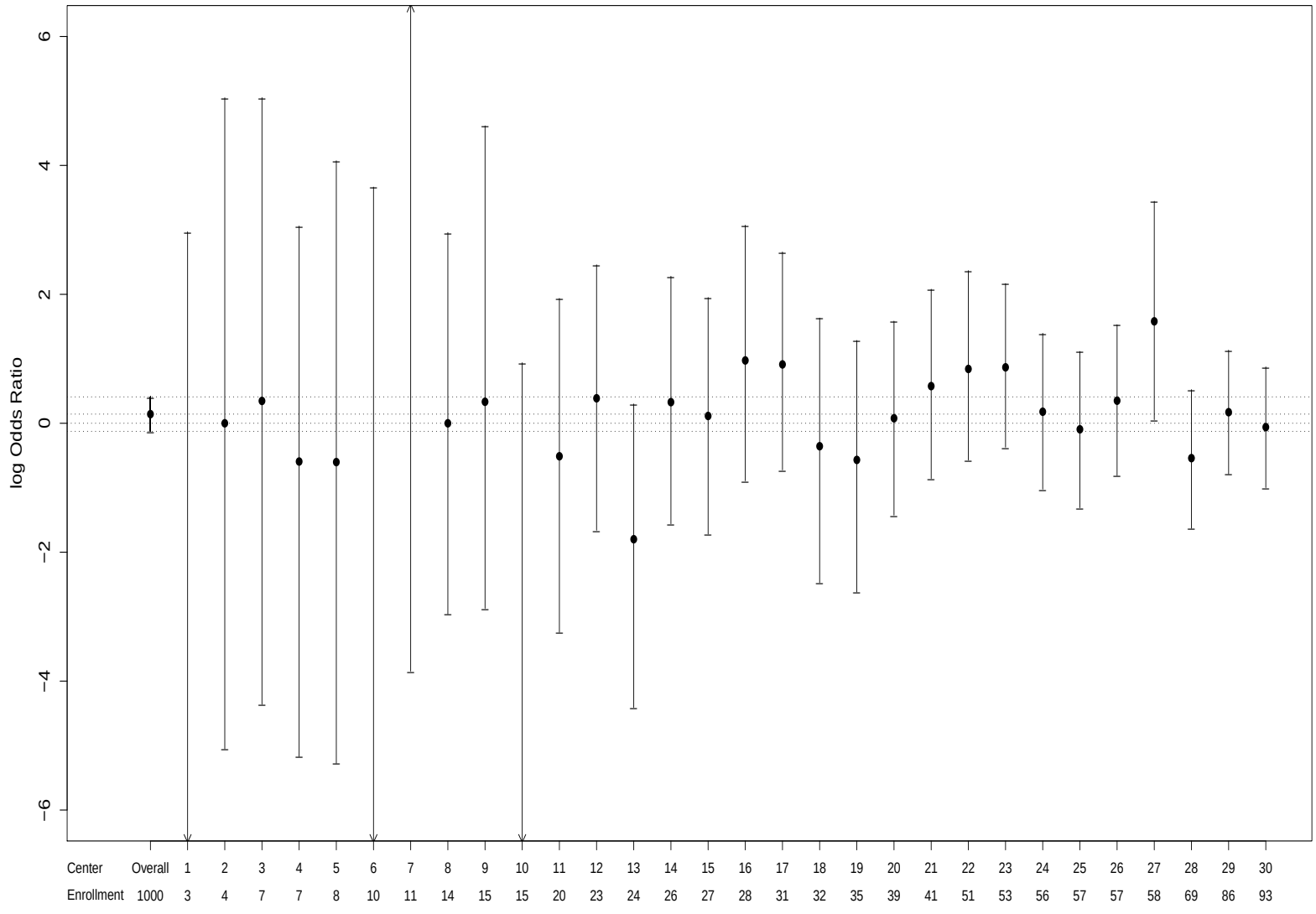
Todd, et al., NEJM, 352(2) : 135 – 145, 2005

IHAST

- IHAST: prospective randomized partially blinded multi-center trial designed to determine whether mild intraoperative hypothermia (33° C), compared to normothermia (36.5° C), resulted in improved neurologic outcome in patients with an acute subarachnoid hemorrhage (SAH).
- 1000 patients in 30 centers were randomized to treatment (hypothermia vs. normothermia) and were followed postoperatively for 3 months.

- The primary outcome measure: the three-month modified Glasgow Outcome Score (GOS).
- GOS: five-point functional outcome scale between 1 (good outcome) and 5 (death).
- The primary result of IHAST: intraoperative hypothermia did not affect neurological outcome. 66% good outcome (GOS $= 1$) with hypothermia (329/499) vs. 63% with normothermia (314/501), odds ratio (OR) $= 1.15$, 95% confidence interval (CI): 0.89 to 1.49).

Frequentist Center-specific logOR and 95% CI



- A positive log OR: there were more good outcomes in patients randomized to hypothermia versus those randomized to normothermia.
- Centers are sorted in ascending order of patients randomized.
- Overall study failed to find any difference between the two treatments, but there may be important center-to-center differences.

Limitations of Frequentist Approach

- Multi-center clinical trials: many centers, variable and sometimes small sample size in each center
- Problems when all success or all failures are observed in all subjects in a center for subjects assigned to a treatment
- The usual frequentist estimate (maximum likelihood estimate, MLE) of an OR is zero or infinite
- The 95% CI of log OR is unbounded
- Bayesian methods

Aim of the Study

Using Bayesian methods determine whether outcome variability among IHAST centers is consistent with a normal distribution and/or whether outcome differences can be explained by characteristics and practices of the centers and /or characteristics of the patients.

Bayesian Approach

- A Bayesian hierarchical generalized linear model for the log OR of a good outcome (3-month GOS score of 1), θ_j.
- The log odds of a good outcome are exchangeable: different in each center, but similar.

- Potential predictors: World Federation of Neurological Surgeons (WFNS) score, age, gender and treatment.
- Compare models by deviance information criteria (DIC).

Bayesian Model

$y_j \mid p_j \sim Bin(n_j, p_j)$ for $j = 1, \ldots 30$ centers,

$\theta_j = logit(p_j) = log(\frac{p_j}{1-p_j})$,

$\theta_j = \mu + \beta_1 Hypo + \beta_2 WFNS1 + \beta_3 males + \beta_4 age + \delta_j$.

Non-informative prior distributions for μ, β_1 to β_4.

$\delta_j \sim Normal(0, \sigma^2)$,

$\sigma^2 \sim Inv - Gamma(\alpha, \beta)$.

Bayesian analysis

- Posterior distribution of the random center effect and between-center standard deviation
- Diagnostic probabilities for centers corresponding to "outliers", and graphical diagnostic tools
- Sensitivity to the prior distribution

Apply Bayesian methods to determine if

- the treatment (hypothermia vs. normothermia) had a different effect on outcome ($GOS = 1$ vs. $GOS > 1$) among any of the 30 IHAST centers,
- center-specific variables (number of subjects, geographic location, various medical practices) had an effect on outcome.

Center-specific Variations

- **Size of centers:** very large ($n_j > 69$ patients; 3 centers (c), 248 patients (p)), large ($56 \leq n_j < 68$; $c : 4$, $p : 228$), medium ($31 \leq n_j < 55$, $c : 7$, $p : 282$) and small ($n_j < 31$, $c : 16$, $p : 242$).
- **Geographic location:** North American (US and Canada, $c : 22$, $p : 637$) or non-North American (Europe, Australia, New Zealand, $c : 8$, $p : 363$).
- **Learning:** outcomes of the first 50% of patients vs the second 50%.
- **Learning within each center:** the outcomes of first 50% patients within each center were compared to the second 50%.

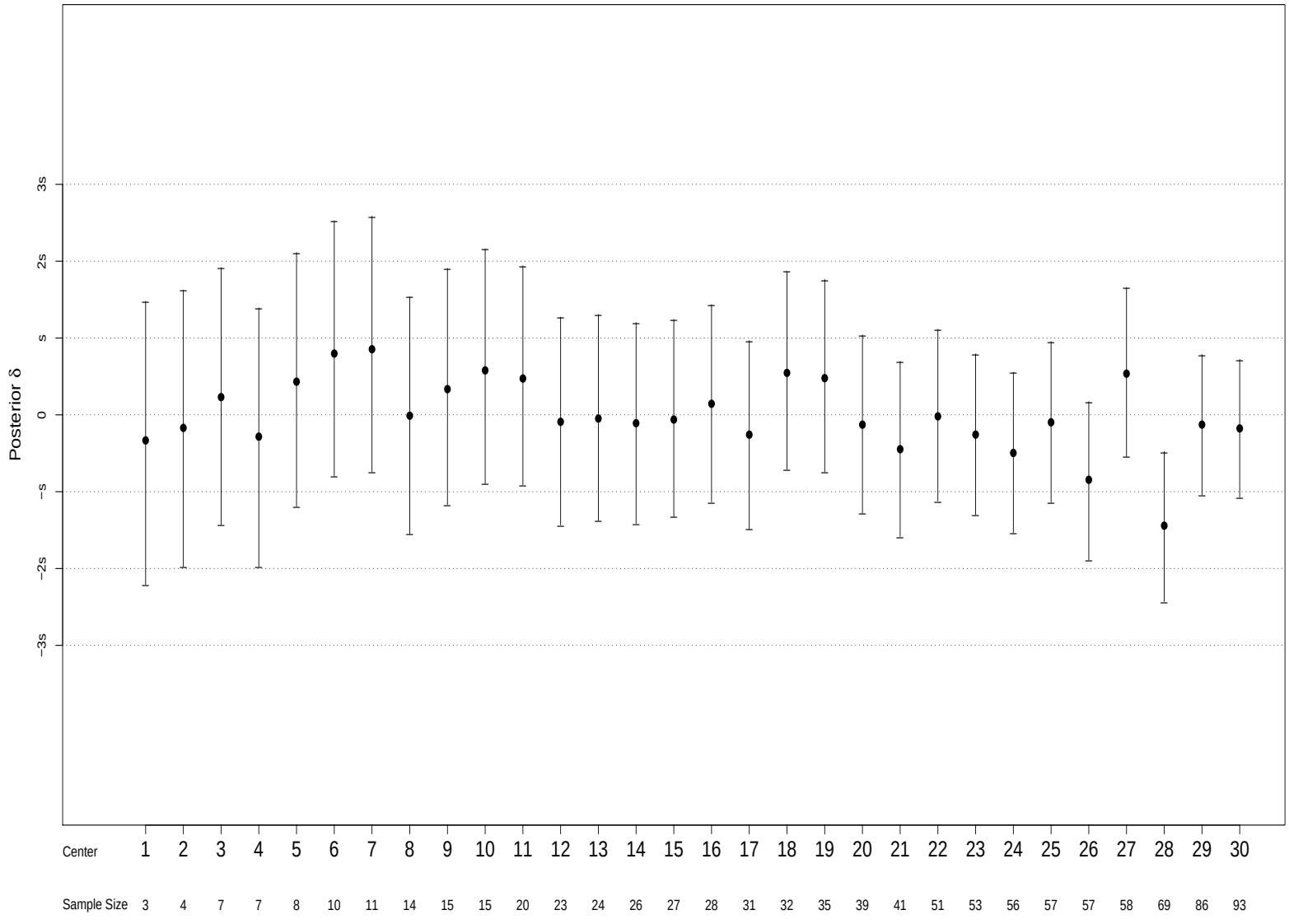
Clinical Practice Differences

- **Nitrous oxide user centers:** low ($\leq 25\%$ of the cases; 13 centers), medium (26% to 74% of cases, 8 centers) or high ($\geq 75\%$ of cases; 9 centers).
- **Nitrous oxide use, patient level:** (yes, $n = 627$; no, $n = 372$).
- **Temporary clipping during aneurysm surgery, center:** low: ($\leq 30\%$ of cases; 6 centers), medium: (30% to 69% of cases; 21 centers) and high: ($\geq 70\%$ of the cases; $n = 3$ centers).
- **Temporary clipping, patient level:** yes ($n = 441$); no ($n = 553$).

Model Comparing 30 Centers

- The best model adjusts for the main effects of treatment (hypothermia vs. normothermia), WFNS score, gender, and age.

Posterior δ_j Comparing 30 Centers



Bayesian Outlier Detection Methods

- The prior probability of at least one center being an outlier: 0.05.
- 30 centers; a center is a potential outlier if the magnitude of the posterior random center effect, δ_j, is greater than 3.137σ in absolute value.
- Treat any center with a posterior probability of being an outlier larger than the prior probability (0.0017) as a potential outlier.
- A center is outlier if Bayes factor (BF) indicates a strong evidence or more.

Chaloner K, Brant RA, Biometrika, 75(4) : 651 – 659, 1988.

Results

- Posterior probabilities of being outlier for centers 6 and 7 are: 0.00295 and 0.0034.
- BF's for these two centers: 0.499 and 0.578 ("negligible" evidence).
- The prior probability that any one of the 30 centers is an outlier: 0.05.
- The joint posterior probability that any one of the 30 centers is an outlier: 0.013.
- The posterior probability for any center to be an outlier is very small and none are identified as possible outlier.

Other Comparisons

Separate similar analyses for outcome ($GOS = 1$ vs. $GOS > 1$)

- center size,
- geography,
- nitrous oxide use,
- temporary clip use and
- learning

no outliers are detected.

Conclusions

- Use of Bayesian analysis methods demonstrates that there are differences between centers in the IHAST, but these differences are consistent with the random variability of a normal distribution.
- Outcome was largely determined by patient characteristics such as WFNS, age, and gender.
- Even if centers differ in their size, location, and clinical practices, the disease (subarachnoid hemorrhage) appears to have the same result everywhere and it is the disease, and not center variation or practice, that determines patient outcome.
- Different prior distributions give similar results.

References

- Todd MM, Hindman BJ, Clarke WR, Torner JC. Mild intraoperative hypothermia during surgery for intracranial aneurysm. New England Journal of Medicine, $352(2) : 135 - 145, 2005$.
- Chaloner K, Brant R. A Bayesian approach to outlier detection and residual analysis. Biometrika, $75(4) : 651 - 659, 1988$.
- Spiegelhalter DJ, Abrams KR, Myles JP. Bayesian Approaches to Clinical Trials and Health-Care Evaluation. England: John Wiley & Sons.; 2004.
- Bayman EO, Chaloner K, Cowles MK. Detecting qualitative interaction: A bayesian approach. Statistics in Medicine, $29 : 455 - 463, 2010$.