Bayesian Adjustment for Multiplicity

Jim Berger
Duke University

with James Scott
University of Texas

2011 Rao Prize Conference
Department of Statistics, Penn State University
May 19, 2011
Outline

• Background on multiplicity

• Illustration of the Bayesian approach through simpler examples
  – Multiple testing under exclusivity
  – Multiple testing under non-exclusivity
  – Sequence multiple testing

• The general Bayesian approach to multiplicity adjustment

• Multiple models

• Variable selection (including comparison with empirical Bayes)

• Subgroup analysis
Some Multiplicity Problems in SAMSI Research Programs

- **Stochastic Computation / Data Mining and Machine Learning**
  - *Example:* Microarrays, with 100,000 mean gene expression differentials $\mu_i$, and testing $H_0 : \mu_i = 0$ versus $H_1 : \mu_i \neq 0$.
  - *Multiplicity problem:* Even if all $\mu_i = 0$, one would find that roughly 500 tests reject at, say, level $\alpha = 0.05$, so a correction for this effect is needed.

- **Astrostatistics and Phystat**
  - *Example:* 1.6 million tests of Cosmic Microwave Background radiation for non-Gaussianity in its spatial distribution.
  - *Example:* At the LHC, they are considering using up to $10^{12}$ tests for each particle event to try to detect particles such as the Higgs boson. And recently (pre LHC), there was an 8$\sigma$ event that didn’t replicate.

- **Multiplicity and Reproducibility in Scientific Studies**
  - In the USA, drug compounds entering Phase I development today have an 8% chance of reaching market, versus a 14% chance 15 years ago.
  - 70% phase III failure rates, versus 20% failure rate 10 years ago.
  - Reports that 30% of phase III successes fail to replicate.
Simple Examples of the Bayesian Approach to Multiplicity Adjustment

**Key Fact:** Bayesian analysis deals with multiplicity adjustment solely through the assignment of prior probabilities to models or hypotheses.

**Example: Multiple Testing under Exclusivity**
Suppose one is testing mutually exclusive hypotheses $H_i$, $i = 1, \ldots, m$, so each hypothesis is a separate model. If the hypotheses are viewed as exchangeable, choose $P(H_i) = 1/m$.

**Example:** 1000 energy channels are searched for a signal:
- if the signal is known to exist and occupy only one channel, but no channel is theoretically preferred, each channel can be assigned prior probability 0.001.
- if the signal is not known to exist (e.g., it is the prediction of a non-standard physics theory) prior probability 1/2 should be given to ‘no signal,’ and probability 0.0005 to each channel.

*This is the Bayesian solution regardless of the structure of the data.*
In contrast, frequentist solutions depend on the structure of the data.

**Example:** For each channel, test $H_{0i} : \mu_i = 0$ versus $H_{1i} : \mu_i > 0$.

**Data:** $X_i, i = 1, \ldots, m$, are normally distributed with mean $\mu_i$, variance 1, and correlation $\rho$.

If $\rho = 0$, one can just do individual tests at level $\alpha/m$ (Bonferroni) to obtain an overall error probability of $\alpha$.

If $\rho > 0$, harder work is needed:

- Choose an overall decision rule, e.g., “declare channel $i$ to have the signal if $X_i$ is the largest value and $X_i > K$.”
- Compute the corresponding error probability, which can be shown to be

$$\alpha = \Pr(\max_i X_i > K \mid \mu_1 = \ldots = \mu_m = 0) = E^Z \left[ 1 - \Phi \left( \frac{K - \sqrt{\rho}Z}{\sqrt{1-\rho}} \right)^m \right],$$

where $\Phi$ is the standard normal cdf and $Z$ is standard normal.

Note that this gives (essentially) the Bonferroni correction when $\rho = 0$, and converges to $1 - \Phi[K]$ as $\rho \to 1$ (the one-dimensional solution).
An example of non-mutually exclusive Bayesian multiple testing

(Scott and Berger, 2006 JSPI; other, more sophisticated full Bayesian analyses are in Gönen et. al. (03), Do, Müller, and Tang (02), Newton et all. (01), Newton and Kendziorski (03), Müller et al. (03), Guindani, M., Zhang, S. and Mueller, P.M. (2007), ...; many empirical Bayes such as Storey, J.D., Dai, J.Y and Leek, J.T. (2007))

- Suppose $x_i \sim N(\mu_i, \sigma^2)$, $i = 1, \ldots, m$, are observed, $\sigma^2$ known, and test $H_{0i} : \mu_i = 0$ versus $H_{1i} : \mu_i \neq 0$.

- Most of the $\mu_i$ are thought to be zero; let $p$ denote the unknown common prior probability that $\mu_i$ is zero.

- Assume that the nonzero $\mu_i$ follow a $N(0, V)$ distribution, with $V$ unknown. q

- Assign $p$ the uniform prior on $(0, 1)$ and $V$ the prior density $\pi(V) = \sigma^2/(\sigma^2 + V)^2$. 
Then the posterior probability that $\mu_i \neq 0$ is

$$p_i = 1 - \frac{\int_0^1 \int_0^1 p \prod_{j \neq i} \left( p + (1 - p)\sqrt{1 - w \cdot e^{w x_j^2/(2\sigma^2)}} \right) dp dw}{\int_0^1 \int_0^1 \prod_{j=1}^m \left( p + (1 - p)\sqrt{1 - w \cdot e^{w x_j^2/(2\sigma^2)}} \right) dp dw}.$$ 

$(p_1, p_2, \ldots, p_m)$ can be computed numerically; for large $m$, it is most efficient to use importance sampling, with a common importance sample for all $p_i$.

**Example:** Consider the following ten ‘signal’ observations:

-8.48, -5.43, -4.81, -2.64, -2.40, 3.32, 4.07, 4.81, 5.81, 6.24

- Generate $n = 10, 50, 500,$ and $5000$ $N(0, 1)$ noise observations.

- Mix them together and try to identify the signals.
The ten ‘signal’ observations

<table>
<thead>
<tr>
<th>$n$</th>
<th>-8.5</th>
<th>-5.4</th>
<th>-4.8</th>
<th>-2.6</th>
<th>-2.4</th>
<th>3.3</th>
<th>4.1</th>
<th>4.8</th>
<th>5.8</th>
<th>6.2</th>
<th>#noise $p_i &gt; .6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>.94</td>
<td>.89</td>
<td>.99</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>50</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>.71</td>
<td>.59</td>
<td>.94</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>500</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>.26</td>
<td>.17</td>
<td>.67</td>
<td>.96</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>5000</td>
<td>1</td>
<td>1.0</td>
<td>.98</td>
<td>.03</td>
<td>.02</td>
<td>.16</td>
<td>.67</td>
<td>.98</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 1: The posterior probabilities of being nonzero for the ten ‘signal’ means.

**Note 1:** The penalty for multiple comparisons is automatic.

**Note 2:** **Theorem:** $E[\#i : p_i > .6 \mid \text{all } \mu_j = 0] = O(1)$ as $m \to \infty$, so the Bayesian procedure exerts medium-strong control over false positives. (In comparison, $E[\#i : \text{Bonferroni rejects } \mid \text{all } \mu_j = 0] = \alpha$.)
Figure 1: For four of the observations, $1 - p_i = \Pr(\mu_i = 0 | y)$ (the vertical bar), and the posterior densities for $\mu_i \neq 0$. 
Sequence Multiple Testing

San Jose Mercury News

Friday, September 25, 2009

AIDS MILESTONE

New path for HIV vaccine

Some in study protected from infection, but trial raises more questions

By Karen Kaplan and Thomas H. Maugh II
Los Angeles Times

Hours after HIV researchers announced the achievement of a milestone that had eluded them for a quarter of a century, reality began to set in: Tangible progress could take another decade.

A Thai and American team announced early Thursday in Bangkok that they had found a combination of vaccines providing modest protection against infection with the virus that causes AIDS, unleashing excitement worldwide. The idea of a vaccine to prevent infection with the human immunodeficiency virus, HIV, had long been frustrating and fruitless.

But by Thursday afternoon, initial euphoria gave way to a more sober assessment. There is still a very long way to go before reaching the goal of producing a vaccine that reliably shields people from HIV.

Some researchers questioned whether the apparent 31 percent reduction in infections was a sta-

See VACCINE, Page 14

A researcher during the Thai phase III HIV Vaccine Trial, also known as RV 144, tests the "prime-boost" combination of two vaccines.

ASSOCIATED PRESS
Hypotheses and Data:

• Alvac had shown no effect
• Aidsvax had shown no effect

Question: Would Alvac as a primer and Aidsvax as a booster work?

The Study: Conducted in Thailand with 16,395 individuals from the general (not high-risk) population:

• 74 HIV cases reported in the 8198 individuals receiving placebos
• 51 HIV cases reported in the 8197 individuals receiving the treatment
The test that was performed:

- Let $p_1$ and $p_2$ denote the probability of HIV in the placebo and treatment populations, respectively.

- Test $H_0 : p_1 = p_2$ versus $H_1 : p_1 \neq p_2$

- Normal approximation okay, so

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{\sigma}\{\hat{p}_1 - \hat{p}_2\}}} = \frac{.009027 - .006222}{.001359} = 2.06$$

is approximately $N(\theta, 1)$, where $\theta = (p_1 - p_2)/(0.001359)$.

We thus test $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$, based on $z$.

- Observed $z = 2.06$, so the $p$-value is 0.04.

Questions:

- Is the $p$-value useable as a direct measure of vaccine efficacy?

- Should the fact that there were two previous similar trials be taken into account (the multiple testing part of the story)?
Bayesian Analysis of the Single Trial:

Prior distribution:

- $Pr(H_i) = \text{prior probability that } H_i \text{ is true, } i = 0, 1,$
- On $H_1 : \theta > 0,$ let $\pi(\theta)$ be the prior density for $\theta.$

Note: $H_0$ must be believable (at least approximately) for this to be reasonable (i.e., no fake nulls).

Subjective Bayes: choose these based on personal beliefs

Objective (or default) Bayes: choose

- $Pr(H_0) = Pr(H_1) = \frac{1}{2},$
- $\pi(\theta) = \text{Uniform}(0, 6.46), \text{ which arises from assigning}$
  - uniform for $p_2$ on $0 < p_2 < p_1,$
  - plug in for $p_1.$
Posterior probability of hypotheses:

\[ Pr(H_0|z) = \text{probability that } H_0 \text{ true, given data } z \]

\[ = \frac{f(z|\theta = 0) Pr(H_0)}{Pr(H_0) f(x|\theta = 0) + Pr(H_1) \int_0^\infty f(z|\theta)\pi(\theta)d\theta} \]

For the objective prior, \( Pr(H_0 | z = 2.06) \approx 0.33 \) (recall, p-value \( \approx .04 \))

Posterior density on \( H_1 : \theta > 0 \) is

\[ \pi(\theta|z = 2.06, H_1) \propto \pi(\theta)f(2.06 | \theta) = (0.413)e^{-\frac{1}{2}(2.06-\theta)^2} \]

for \( 0 < \theta < 6.46 \).
**Robust Bayes:** Report the *Bayes factor* (the odds of $H_0$ to $H_1$) as a function of $\pi_C(\theta) \equiv \text{Uniform}(0, C)$:

\[
B_{01}(C) = \frac{\text{likelihood of } H_0 \text{ for observed data}}{\text{average likelihood of } H_1} = \frac{\frac{1}{\sqrt{2\pi}} e^{-(2.06-\theta)^2/2}}{\int_0^C \frac{1}{\sqrt{2\pi}} e^{-(2.06-\theta)^2/2} C^{-1} d\theta}
\]

*Note:* $\min_C B_{01}(C) = 0.265$ (while $B_{01}(6.46) = 0.51$).

*Note:* The robustness analysis applies to all non-increasing priors.
Incorporation information from multiple tests: To adjust for the two previous similar failed trials, the (exchangeable) Bayesian solution

- assigns each trial common unknown probability $p$ of success, with $p$ having a uniform distribution;

- computes the resulting posterior probability that the current trial exhibits no efficacy

$$Pr(H_0 \mid x_1, x_2, x_3) = \left(1 + \frac{B_{01}(x_1)B_{01}(x_2) + B_{01}(x_1) + B_{01}(x_2) + 3}{3B_{01}(x_1)B_{01}(x_2) + B_{01}(x_1) + B_{01}(x_2) + 1} \times \frac{1}{B_{01}(x_3)}\right)^{-1}$$

where $B_{01}(x_i)$ is the Bayes factor of “no effect” to “effect” for trial $i$.

The result is $Pr(H_0 \mid x_1, x_2, x_3) = 0.54$. 
General Approach to Bayesian Multiplicity Adjustment

1. Represent the problem as a *model uncertainty* problem: Models $M_i$, with densities $f_i(x \mid \theta_i)$ for data $x$, given unknown parameters $\theta_i$; prior distributions $\pi_i(\theta_i)$; and marginal likelihoods

   $$m_i(x) = \int f_i(x \mid \theta_i) \pi_i(\theta_i) \, d\theta_i.$$ 

2. Specify prior probabilities, $P(M_i)$, of models to reflect the multiplicity issues; **Bayesian analysis controls multiplicity through $P(M_i)$**

   - *Subjective Bayesian Analysis:* If the $P(M_i)$ are real subjective probabilities, that’s it: multiplicity correction has been done.
   - *Objective Bayesian Analysis:* One has to be careful to make choices of the $P(M_i)$ that ensure multiplicity correction (e.g., specifying equal prior probabilities does not generally control multiplicity)!

3. Implement Bayesian model averaging (model selection?), based on

   $$P(M_i \mid x) = \frac{P(M_i) \, m_i(x)}{\sum_{j=1}^k P(M_j) \, m_j(x)}.$$

---

*see, e.g., Jeffreys 1961; Waller and Duncan 1969; Meng and Dempster 1987; Berry 1988; Westfall, Johnson and Utts 1997; Carlin and Louis 2000.*
Choice of transformation/model

Bayesian solution: model averaging.

• Assign each model/transformation a prior probability.
• Compute model/transformation posterior probabilities.
• Perform inference with weighted averages over the models/transformations. (An overwhelmingly supported model/transformation will receive weight near one.)
**Example:** From i.i.d. vehicle emission data $\mathbf{X} = (X_1, \ldots, X_n)$, one desires to determine the probability that the vehicle type will meet regulatory standards.

Traditional models for this type of data are Weibull and lognormal distributions given, respectively, by

$$
\mathcal{M}_1 : f_W(x; \beta, \gamma) = \frac{\gamma}{\beta} \left(\frac{x}{\beta}\right)^{\gamma-1} \exp\left[-\left(\frac{x}{\beta}\right)^\gamma\right]
$$

$$
\mathcal{M}_2 : f_L(x; \mu, \sigma^2) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\log x - \mu)^2}{2\sigma^2}\right].
$$

Note that both distributions are in the location-scale family (the Weibull being so after a log transformation).
Model Averaging Analysis:

- Assign each model prior probability 1/2.
- Because of the common location-scale invariance structures, assign the right-Haar prior densities $\pi_W(\beta, \gamma) = 1/(\beta\gamma)$ and $\pi_L(\mu, \sigma) = 1/(\sigma)$, respectively (Berger, Pericchi and Varshavsky, 1998 Sankhyā).
- The posterior probabilities (and conditional frequentist error probabilities) of the two models are then

$$P(M_1 \mid \mathbf{x}) = 1 - P(M_2 \mid \mathbf{x}) = \frac{B(\mathbf{x})}{1 + B(\mathbf{x})},$$

where $z_i = \log x_i, \bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i, s_{\bar{z}}^2 = \frac{1}{n}\sum_{i=1}^{n} (z_i - \bar{z})^2$, and

$$B(\mathbf{x}) = \frac{\Gamma(n) n^{n} \pi^{(n-1)/2}}{\Gamma(n - 1/2)} \int_{0}^{\infty} \left[ \frac{y}{n} \sum_{i=1}^{n} \text{exp} \left( \frac{z_i - \bar{z}}{s_{\bar{z}} y} \right) \right]^{-n} dy.$$

- For the studied data set, $P(M_1 \mid \mathbf{x}) = .712$. Hence,

$$P(\text{meeting standard}) = .712 \ P(\text{meeting standard} \mid M_1) + .288 \ P(\text{meeting standard} \mid M_2).$$
Variable Selection

Problem: Data $X$ arises from a normal linear regression model, with $m$ possible regressors having associated unknown regression coefficients $\beta_i, i = 1, \ldots, m$, and unknown variance $\sigma^2$.

Models: Consider selection from among the submodels $M_i, i = 1, \ldots, 2^m$, having only $k_i$ regressors with coefficients $\beta_i$ (a subset of $(\beta_1, \ldots, \beta_m)$) and resulting density $f_i(x | \beta_i, \sigma^2)$.

Prior density under $M_i$: Zellner-Siow priors $\pi_i(\beta_i, \sigma^2)$.

Marginal likelihood of $M_i$: $m_i(x) = \int f_i(x | \beta_i, \sigma^2) \pi_i(\beta_i, \sigma^2) d\beta_i d\sigma^2$

Prior probability of $M_i$: $P(M_i)$

Posterior probability of $M_i$:

$$P(M_i | x) = \frac{P(M_i)m_i(x)}{\sum_j P(M_j)m_j(x)}.$$
Common Choices of the $P(M_i)$

Equal prior probabilities: $P(M_i) = 2^{-m}$

Bayes exchangeable variable inclusion:

- Each variable, $\beta_i$, is independently in the model with unknown probability $p$ (called the prior inclusion probability).
- $p$ has a $\text{Beta}(p \mid a, b)$ distribution. (We use $a = b = 1$, the uniform distribution, as did Jeffreys 1961, who also suggested alternative choices of the $P(M_i)$. Probably $a = b = 1/2$ is better.)
- Then, since $k_i$ is the number of variables in model $M_i$,
  $$P(M_i) = \int_0^1 p^{k_i} (1 - p)^{m-k_i} \text{Beta}(p \mid a, b) dp = \frac{\text{Beta}(a + k_i, b + m - k_i)}{\text{Beta}(a, b)}.$$ 

Empirical Bayes exchangeable variable inclusion: Find the MLE $\hat{p}$ by maximizing the marginal likelihood of $p$, $\sum_j p^{k_j} (1 - p)^{m-k_j} m_j(x)$, and use $P(M_i) = \hat{p}^{k_i} (1 - \hat{p})^{m-k_i}$ as the prior model probabilities.
Controlling for multiplicity in variable selection

Equal prior probabilities: $P(M_i) = 2^{-m}$ does not control for multiplicity here (as it did in the simpler examples); it corresponds to fixed prior inclusion probability $p = 1/2$ for each variable.

Empirical Bayes exchangeable variable inclusion does control for multiplicity, in that $\hat{p}$ will be small if there are many $\beta_i$ that are zero.

Bayes exchangeable variable inclusion also controls for multiplicity (see Scott and Berger, 2008), although the $P(M_i)$ are fixed.

Note: The control of multiplicity by Bayes and EB variable inclusion usually reduces model complexity, but is different than the usual Bayesian Ockham’s razor effect that reduces model complexity.

- The Bayesian Ockham’s razor operates through the effect of model priors $\pi_i(\beta, \sigma^2)$ on $m_i(x)$, penalizing models with more parameters.
- Multiplicity correction occurs through the choice of the $P(M_i)$. 
Table 2: Posterior inclusion probabilities for 10 real variables in a simulated data set.
Comparison of Bayes and Empirical Bayes Approaches

**Theorem 1** In the variable-selection problem, if the null model (or full model) has the largest marginal likelihood, \( m(x) \), among all models, then the MLE of \( p \) is \( \hat{p} = 0 \) (or \( \hat{p} = 1 \)). (The naive EB approach, which assigns \( P(M_i) = \hat{p}^{k_i} (1 - \hat{p})^{m-k_i} \), concludes that the null (full) model has probability 1.)

A simulation with 10,000 repetitions to gauge the severity of the problem:

- \( m = 14 \) covariates, orthogonal design matrix
- \( p \) drawn from \( U(0,1) \); regression coefficients are 0 with probability \( p \) and drawn from a Zellner-Siow prior with probability \( (1-p) \).
- \( n = 16, 60, \) and 120 observations drawn from the given regression model.

<table>
<thead>
<tr>
<th>Case</th>
<th>( \hat{p} = 0 )</th>
<th>( \hat{p} = 1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n = 16 )</td>
<td>820</td>
<td>781</td>
</tr>
<tr>
<td>( n = 60 )</td>
<td>783</td>
<td>766</td>
</tr>
<tr>
<td>( n = 120 )</td>
<td>723</td>
<td>747</td>
</tr>
</tbody>
</table>
Is empirical Bayes at least accurate asymptotically as $m \to \infty$?

Posterior model probabilities, given $p$:

$$ P(\mathcal{M}_i \mid x, p) = \frac{p^{k_i}(1 - p)^{m-k_i} m_i(x)}{\sum_j p^{k_j}(1 - p)^{m-k_j} m_j(x)} $$

Posterior distribution of $p$: $\pi(p \mid x) = K \sum_j p^{k_j}(1 - p)^{m-k_j} m_j(x)$

This does concentrate about the true $p$ as $m \to \infty$, so one might expect that

$$ P(\mathcal{M}_i \mid x) = \int_0^1 P(\mathcal{M}_i \mid x, p)\pi(p \mid x)dp \approx P(\mathcal{M}_i \mid x, \hat{p}) \propto m_i(x) \hat{p}^{k_i}(1 - \hat{p})^{m-k_i}.$$

This is not necessarily true; indeed

$$ \int_0^1 P(\mathcal{M}_i \mid x, p)\pi(p \mid x)dp = \int_0^1 \frac{p^{k_i}(1 - p)^{m-k_i} m_i(x)}{\pi(p \mid x)/K} \times \pi(p \mid x) \, dp $$

$$ \propto m_i(x) \int_0^1 p^{k_i}(1 - p)^{m-k_i} dp \propto m_i(x) P(\mathcal{M}_i). $$

Caveat: Some EB techniques have been justified; see Efron and Tibshirani (2001), Johnstone and Silverman (2004), Cui and George (2006), and Bogdan et. al. (2008).
Theorem 2 Suppose the true model size $k_T$ satisfies $k_T/m \to p_T$ as $m \to \infty$, where $0 < p_T < 1$. Consider all models $M_i$ such that $k_T - k_i = O(\sqrt{m})$, and consider the optimal situation for EB in which

$$\hat{p} = p_T + O\left(\frac{1}{\sqrt{m}}\right) \quad \text{as} \quad m \to \infty.$$ 

Then the ratio of the prior probabilities assigned to such models by the Bayes approach and the empirical Bayes approach satisfies

$$\frac{P_B(M_i)}{P_{EB}(M_i)} = \frac{\int_0^1 p^{k_i}(1 - p)^{m-k_i} \pi(p) dp}{(\hat{p})^{k_i}(1 - \hat{p})^{m-k_i}} = O\left(\frac{1}{\sqrt{m}}\right),$$

providing $\pi(\cdot)$ is continuous and nonzero.
Subgroup Analysis

**The New England Journal of Medicine**

**ORIGINAL ARTICLE**

**Clopidogrel and Aspirin versus Aspirin Alone for the Prevention of Atherothrombotic Events**

Deepak L. Bhatt, M.D., Keith A.A. Fox, M.B., Ch.B., Werner Hacke, M.D.,
Peter B. Berger, M.D., Henry R. Black, M.D., William E. Boden, M.D.,
Patrice Cacoub, M.D., Eric A. Cohen, M.D., Mark A. Creager, M.D.,
J. Donald Easton, M.D., Marcus D. Flather, M.D., Steven M. Haffner, M.D.,
Christian W. Hamm, M.D., Graeme J. Hankey, M.D., S. Claiborne Johnston, M.D.,
Koon-Hou Mak, M.D., Jean-Louis Mas, M.D., Gilles Montalescot, M.D., Ph.D.,
Thomas A. Pearson, M.D., P. Gabriel Steg, M.D., Steven R. Steinbuchl, M.D.,
Michael A. Weber, M.D., Danielle M. Brennan, M.S., Liz Fabry-Ribaudo, M.S.N., R.N.,
Joan Booth, R.N., and Eric J. Topol, M.D., for the CHARISMA Investigators*

**CONCLUSIONS**

In this trial, there was a suggestion of benefit with clopidogrel treatment in patients with symptomatic atherothrombosis and a suggestion of harm in patients with multiple risk factors. Overall, clopidogrel plus aspirin was not significantly more effective than aspirin alone in reducing the rate of myocardial infarction, stroke, or death from cardiovascular causes. (ClinicalTrials.gov number, NCT00050817.)
<table>
<thead>
<tr>
<th>Group</th>
<th>Hazard Ratio for MI, Stroke, or Death from Cardiovascular Causes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td></td>
</tr>
<tr>
<td>&lt;75 yr</td>
<td></td>
</tr>
<tr>
<td>≥75 yr</td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td></td>
</tr>
<tr>
<td>Diabetes</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Smoking</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Body mass index</td>
<td></td>
</tr>
<tr>
<td>Normal (≤25)</td>
<td></td>
</tr>
<tr>
<td>Overweight (25 to &lt;30)</td>
<td></td>
</tr>
<tr>
<td>Obese (≥30)</td>
<td></td>
</tr>
<tr>
<td>Hypertension</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Hypercholesterolemia</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td></td>
</tr>
<tr>
<td>History of CABG</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td></td>
</tr>
<tr>
<td>History of PCI</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td></td>
</tr>
<tr>
<td>History of MI</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td></td>
</tr>
<tr>
<td>History of stroke</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Inclusion group</td>
<td></td>
</tr>
<tr>
<td>Asymptomatic</td>
<td></td>
</tr>
<tr>
<td>Symptomatic</td>
<td></td>
</tr>
<tr>
<td>All patients—overall cohort:</td>
<td>0.93 (0.83–1.05)</td>
</tr>
</tbody>
</table>
Frequentist adjustment for performing 26 hypothesis tests

- Split the data into one part to suggest a subgroup and another part to confirm (or confirm with a new experiment).

- Bonferroni correction
  - To achieve an overall error probability level of 0.05 when conducting 26 tests, one would need to use a per-test rejection level of \( \alpha = 0.05/26 = 0.002 \).
  - This is likely much too conservative because of the dependence in the 26 tests.

- Various bootstrap types of correction to try to account for dependence.
Bayesian adjustment

Let $\mathbf{v}$ be the vector of 25 zeroes and ones indicating subgroup characteristics.

For each possible such vector, let $\mu_{\mathbf{v}}$ denote the mean of the intersected subgroup (e.g., young, male, diabetic, non-smoker,...).

Data: $\mathbf{x} \sim f(\mathbf{x} \mid \{\mu_{\mathbf{v}}, \text{all possible } \mathbf{v}\})$.

Two classes of approaches

• Factor-based approaches
• Aggregation-based approaches
An example factor-based approach

Model the intersected subgroup means additively as

$$\mu v = \mu + v\beta, \quad \beta = (\beta_1, \ldots, \beta_{25})',$$

where $\mu$ is an overall mean and $\beta_i$ is the effect corresponding to the $i^{th}$ subgroup factor.

Conversion to model selection:

- Let $\gamma = (\gamma_0, \gamma^*) = (\gamma_0, \gamma_1, \ldots, \gamma_{25})$ be the vector of zeroes and ones, indicating whether $\mu$ (corresponding to $\gamma_0$) and each factor $\beta_i$ is zero or not.
- This defines the model $\mathcal{M}_\gamma$. 
A reasonable objective choice of prior model probabilities:

- $P(\gamma_0 = 0) = P(\mu = 0) = 3/4$.
- Independently, $P(\gamma^* = 0) = 2/3$ and $\gamma^* \neq 0$ have probability

$$P(\gamma^*) = \frac{26}{75} \cdot \frac{Beta(1 + r, 1 + 25 - r)}{Beta(1, 1)}$$

where $r = \#$ zeroes in $\gamma^*$.

- Note that then
  - $P(\text{no effect}) = P(\mu = 0, \gamma^* = 0) = 1/2$
  - $P(\mu \neq 0, \gamma^* = 0) = 1/6$
  - $P(\mu = 0, \gamma^* \neq 0) = 1/4$
  - $P(\mu \neq 0, \gamma^* \neq 0) = 1/12$
  - $P(\gamma_i \neq 0) = 13/75$

The experimenter could (pre-experimentally) make different choices here, as long as $P(\text{no effect})$ is kept at 1/2. Post-experimentally, one would need to utilize an objective choice such as the above.
Possible Bayesian outputs of interest:

- $P(\text{effect of factor } i \neq 0 \mid x) = \sum_{\{\gamma_i = 1\}} P(M_{\gamma} \mid x)$.
- $P(\text{effect in subgroup } i \neq 0 \mid x) = \sum_{\{\gamma_0 = 1 \text{ or } \gamma_i = 1\}} P(M_{\gamma} \mid x)$.
- $P(\text{a constant effect } \neq 0 \mid x) = P(M_{(1,0)} \mid x)$.

Of course, posterior densities for all effects, conditional on their being nonzero, are also available.
Aggregation-based approaches

*Basic idea:* Recall that for every intersected subgroup (e.g., young, male, diabetic, non-smoker,...) there is an unknown mean $\mu_v$. Plausible models involve aggregation of these means into common effects, e.g. $\mu_{v_1} = \mu_{v_2}$. There are a number of ways to aggregate means, including

- Product partition models (Hartigan and Berry)
- Dirichlet process models (Gopalan and Berry use for multiplicity control)
- Generalized partition models
- Species sampling models
- Tree-based models (our current favorite)

*Surmountable problem:* Any of these aggregate means could be zero; with some work, this can typically be handled by adding “zero” to the list.

*Harder problem:* Not all (not even most) aggregations are sensible (e.g., $\mu_{F_1G_1} = \mu_{F_2G_2} \neq \mu_{F_1G_2} = \mu_{F_2G_1}$ versus $\mu_{F_1G_1} = \mu_{F_2G_1} \neq \mu_{F_1G_2} = \mu_{F_2G_2}$).
Summary

• Developing methods for controlling for multiplicity is a dramatically increasing need in science.

• Approaching multiplicity control from the Bayesian perspective has the attractions that
  – there is a single approach that can be applied in any situation;
  – since multiplicity is controlled solely through prior probabilities of models, it does not depend on the error structure of the model;
  – there is flexibility in the assignment of prior probabilities to hypotheses, from pure objective assignments to (pre-experimental) subjective assignments favoring scientifically preferred hypotheses;
  – objective Bayesian control can even be implemented retroactively.

• Associated empirical Bayes analysis exhibits multiplicity control, but cannot be assumed to be an approximation to the Bayesian analysis.

• Bayesian implementation of subgroup analysis is promising.
Thanks!