Adjusting for multiplicities in variable selection using neutral-data comparisons

Dan J. Spitzner

April 7, 2014

Abstract

A multiplicity adjustment is proposed for variable selection problems that is specified through prior model probabilities in a model-choice context. This is distinct from many traditional approaches to multiple testing, whose focus is the choice of loss function, rather than the choice of prior. The proposed multiplicity adjustment is described through a simple rule that is expressed within a recently developed framework for calibrating Bayes factors, known as neutral-data comparisons. Theoretical analysis suggests that the approach is effective in high-dimensional, sparse-signal contexts. It is illustrated on a data set of adverse-event frequencies in a vaccine trial, and compared with an existing, related approach, where it is shown to provide a beneficial clarifying effect.

KEY WORDS: Variable selection; Bayes factors; neutral-data comparisons; multiple testing; adverse-event data.

ABBREVIATED TITLE: Adjusting for multiplicities in variable selection.

1 Introduction

In testing problems, the "multiplicity" concept (a.k.a., "multiple testing" or "multiple comparisons") extends the goal of choosing between two models, \mathcal{M}_0 vs \mathcal{M}_1 , say, to that of choosing from among a class of multiple models, \mathcal{M}_s for $s \in S$. It emphasizes the difficult challenge of making sensible inferences when there are many, possibly *very* many, models to consider. In recent years, the development of multiplicity-adjustment techniques has become, and remains, one of the most active areas of statistical research. Berry and Hochberg (1999) express its importance in stating, "Multiplicities are present in virtually every application of statistics. Multiple comparisons... are among the most difficult of problems faced by statisticians and other researchers."

Multiplicity is often treated using decision theory in various ways, focusing primarily on the choice of loss function; see, *e.g.*, Duncan (1965), Berry and Hochberg (1999), Scott and Berger (2006), Müller, Parmigiani, and Rice (2007), and Polson and Scott (2011). This article focuses instead on the choice of prior model probabilities. The essence of the approach is to formulate each prior probability of a "signal" conditionally on the number of existing signals, assigning a low or high value according to whether there a low or high number of existing signals. The exact probability value to be assigned depends on the multipletesting problem under consideration. This article develops a rule for variable selection with independent components, in which the likelihood function factors into component-specific likelihood functions. Theoretical evaluation suggests that the resulting procedure is remarkably effective. For example, it is able to detect sparse signals in ultra-high dimensions, according to stringent asymptotic criteria described in Fan and Lv (2008).

Dan Spitzner is Associate Professor, Department of Statistics, University of Virginia, P. O. Box 400135, Charlottesville, VA 22904-4135, USA. The author is grateful for invaluable support in preparing this technical report from the National Science Foundation (grant number SES-1260803).

An example variable selection problem, an adverse-event analysis of data from a clinical trial, motivates and guides the development of the proposed methodology. The data appear in Berry and Berry (2004), whose analysis is especially relevant for demonstrating an existing methodology that treats multiplicity through the assignment of prior model probabilities. That analysis applies a hierarchical formulation of prior model probabilities, and is found to produce a very strong shrinkage effect, while the proposed approach shrinks more modestly and better clarifies the evidence for adverse events (and nonevents) in the data.

As a means to make sense of the proposed approach, its ideas are organized within the framework of "neutral-data comparisons," a recent technique for calibrating Bayes factors introduced in Spitzner (2011). Neutral-data comparisons situate among Bayesian testing procedures that have been developed for use under vague prior information. Notable examples are the criterion of Schwarz (1978), the calibrated improper Bayes factors of Spiegelhalter and Smith (1982), the fractional Bayes factors of O'Hagan (1995), the unit-information priors of Kass and Wasserman (1995), and the intrinsic Bayes factors of Berger and Pericchi (1996). See also Bayarri *et al.* (2012) and references therein. In what follows, ideas discussed in Spiegelhalter and Smith (1982) are especially relevant for guiding the formulation of "imaginary data," which anchors the calibration underlying the neutral-data comparisons technique.

The article is organized as follows. Section 2 describes the variable selection problem and proposed methodology. Section 3 demonstrates the methodology on Berry and Berry's adverse-events data set, and compares results. Concluding discussion appears in Section 4, and all technical derivations appear in the appendix.

2 Variable selection by neutral-data comparisons

In the variable selection problem treated here, the data consist of p independent sets of sample measurements, $\mathbf{Y}_1, \ldots, \mathbf{Y}_p$ such that the *i*'th set is $\mathbf{Y}_i = (\mathbf{Y}_{i1}, \ldots, \mathbf{Y}_{in_i})$ for independent \mathbf{Y}_{ij} . A model \mathcal{M}_s is defined from a subset $A_s \subset \{1, \ldots, p\}$, according to which the parameter "of interest" is $\boldsymbol{\theta}_s = (\boldsymbol{\theta}_i : i \notin A_s)$, the "nuisance" parameter is $\boldsymbol{\phi} = (\boldsymbol{\phi}_i : i = 1, \ldots, p)$, and the likelihood function is

$$L_s(\boldsymbol{\theta}_s, \boldsymbol{\phi}; \boldsymbol{Y}) = \prod_{i \in A_s} L(\boldsymbol{\theta}_i^0, \boldsymbol{\phi}_i; \boldsymbol{Y}_i) \times \prod_{i \notin A_s} L(\boldsymbol{\theta}_i, \boldsymbol{\phi}_i; \boldsymbol{Y}_i)$$
(1)

in which the components of $\theta_s^0 = (\theta_i^0 : i \in A_s)$ are fixed "null" values of the parameter. In other words, the model \mathcal{M}_s reflects the hypothesis of "no signal," $\theta_i = \theta_i^0$, in components $i \in A_s$. In some contexts (such as the Gaussian case described below), the nuisance parameters ϕ_i are not distinct across *i*, but are identified with a common parameter.

The prior is understood through the specification

$$\boldsymbol{\theta}_i | \boldsymbol{\phi}_i, \boldsymbol{\tau} \sim G(\boldsymbol{\theta}_i^0, \tau_i^2 \boldsymbol{\Delta}_i(\boldsymbol{\phi}_i, \boldsymbol{\tau})),$$
(2)

where $\Delta_i(\phi_i, \tau)$ is a positive-definite covariance matrix, and τ is a hierarchical parameter that includes the scale parameter, τ_i^2 . Though (2) is specifically Gaussian, its dependence on τ admits treatment of possibly complex hierarchical formulations that extend beyond the Gaussian family.

Several specific cases of variable selection problems will be examined in detail to develop the proposed methodology and understand its properties.

- Case 1: In the Gaussian case, each $Y_{ij}|\mathcal{M}_s, \Sigma \sim G(\mathbf{0}, \Sigma)$ if $i \in A_s$ and $Y_{ij}|\mathcal{M}_s, \theta_i, \Sigma \sim G(\theta_i, \Sigma)$ if $i \notin A_s$, where $\phi_i = \Sigma$ is a positive definite covariance matrix. The prior is specified as $\theta_i | \tau^2 \sim G(\mathbf{0}, \tau^2 \Delta)$. Each Y_i is assumed to have the same sample size, $n_i = n$, and each θ_i to have the same dimension, ν .
- Case 2: The component likelihoods are instead defined from an exponential family, or some other parametric family that is suitable regular. For instance, the case of binary outcomes is the basis of many disease-incidence models, such as those found in spatial epidemiology, in which the sum T_i = ∑_j Y_{ij} has T_i|M_s ~ binomial(n_i, θ_i⁰) if i ∈ A_s and T_i|M_s ~ binomial(n_i, θ_i) if i ∉ A_s. In general, justification of the concepts discussed below will follow from the existence of maximum-likelihood values, θ̂_i, calculated at fixed φ_i, that satisfy and asymptotic conditional Gaussian approximation, θ̂_i|φ_i~G(θ_i, n_i⁻¹I(θ_i, φ_i)⁻¹), as n_i → ∞, where n_iI(θ_i, φ_i) is the component-specific Fisher information matrix. See, *e.g.*, Shao (2003) for conditions under which this property is satisfied.

The article's objective is to develop ideas sufficiently well to apply them in a reanalysis of Berry and Berry's (2004) adverse-event data set. That context is described as follows, using slightly revised subscripting.

• *Case 3:* The data are an array of incidence-count totals from a vaccine trial that involved control and treatment groups of $n_1 = 132$ and $n_2 = 148$ subjects. The counts are of forty pre-defined "adverse event" (AE) occurrences (*e.g.*, a rash or nausea), which are uniquely grouped into eight body systems. Corresponding notation identifies pairs of triple-subscripted data, $Y_{jk} = (Y_{1jk}, Y_{2jk})$, where k indexes AE-type $k \in K_j$ within body system $j \in J$, and the order of pairing reflects "control" *versus* "treatment" conditions. The data-analysis objective is to "flag" any AE-types whose occurrence-rates are greater under the vaccine treatment. Each $Y_{ijk} \sim$ binomial (n_i, p_{ijk}) , independently across i = 1, 2 and $(j, k) \in \Omega = \{(j, k) : j \in J, k \in K_j\}$. Each model \mathcal{M}_s is characterized by three subsets: $A_{s,0}$, which collects index-pairs (j, k) such that $p_{1jk} = p_{2jk}$; $A_{s,1}$, which collects the (j, k) such that $p_{1jk} > p_{2jk}$; and, $A_{s,2}$, which collects the (j, k) such that $p_{1jk} < p_{2jk}$. Define $\phi_{jk} = \frac{1}{2}(\eta_{1jk} + \eta_{2jk})$ and $\theta_{jk} = \frac{1}{2}(\eta_{1jk} - \eta_{2jk})$, having set $\eta_{ijk} = \log\{p_{ijk}/(1 - p_{ijk})\}$, and collect them into the parameters $\phi = (\phi_{jk} : (j, k) \in \Omega)$ and $\theta_s = (\theta_{jk} : (j, k) \notin A_{s,0})$, which respectively record ν_0 and $\nu_{s,1}$ "free" parameters of model \mathcal{M}_s . The likelihood function factors into components, as in (1) but indexed by (j, k), such that $\theta_{jk}^0 = 0$ replaces θ_{jk} for components with $(j, k) \in A_{s,0}$. Each component is from an exponential family,

$$L(\theta_{jk}, \phi_{jk}; \boldsymbol{Y}_{jk}) = C_{\zeta}(\boldsymbol{Y}_{jk}) \exp\{(Y_{1jk} - Y_{2jk})\theta_{jk} + (Y_{1jk} + Y_{2jk})\phi_{jk} - \zeta(\theta_{jk}, \phi_{jk})\},\$$

where $C_{\zeta}(\boldsymbol{Y}_{jk}) = \binom{n_1}{Y_{1jk}}\binom{n_2}{Y_{2jk}}$ and

$$\zeta(\theta_{jk}, \phi_{jk}) = n_1 \log \left(1 + e^{\phi_{jk} + \theta_{jk}} \right) + n_2 \log \left(1 + e^{\phi_{jk} - \theta_{jk}} \right)$$

The prior for this case is understood through a hierarchical formulation that makes use of the arrangement of AE-types within body systems; its detailed description appears in Section S:AEs.

2.1 Bayes factors and neutral-data comparisons

Conditional on ϕ and τ , it is assumed the prior on \mathcal{M}_s specifies mutual independence among the components of θ_s . It follows that the conditional Bayes factor for choosing between two models \mathcal{M}_s vs \mathcal{M}_t factors according to

$$BF_{st}(\boldsymbol{Y}|\boldsymbol{\phi},\boldsymbol{\tau}) = \prod_{i \in A_s \cap A_t^c} BF_i(\boldsymbol{Y}_i|\boldsymbol{\phi},\boldsymbol{\tau}) \times \prod_{i \in A_s^c \cap A_t} \frac{1}{BF_i(\boldsymbol{Y}_i|\boldsymbol{\phi},\boldsymbol{\tau})},$$
(3)

where

$$BF_i(\boldsymbol{Y}_i|\boldsymbol{\phi},\boldsymbol{\tau}) = \frac{L(\boldsymbol{\theta}_i^0,\boldsymbol{\phi}_i;\boldsymbol{Y}_i)}{\int L(\boldsymbol{\theta}_i,\boldsymbol{\phi}_i;\boldsymbol{Y}_i)\pi(\boldsymbol{\theta}_i|\boldsymbol{\phi}_i,\boldsymbol{\tau})d\boldsymbol{\theta}_i}$$
(4)

is the conditional Bayes factor associated with the hypothesis that component i has "no signal."

The proposed variable selection scheme is formulated by calibrating the $BF_{st}(\mathbf{Y}|\boldsymbol{\phi}, \tau)$ in a manner that reflects the complexity of \mathcal{M}_s and \mathcal{M}_t . It is sufficient to only work directly with elementary tests defined by \mathcal{M}_s vs \mathcal{M}_t such that $A_s = A_t \cup \{i\}$, for some *i*, since the formula (3) induces a relationship among Bayes factors that will extend any calibration made in these tests to the Bayes factor of any two models. Assuming an elementary test, the proposed calibration acts to shrink the Bayes factor if $|A_s|$ is small and magnify the Bayes factor if $|A_s|$ is large. The specific calibration is not formulated in an *ad hoc* manner, but within the neutral-data comparisons framework of Spitzner (2011), whose basic elements are now described.

Neutral-data comparisons derive from an exercise involving *imaginary* data, in which the analyst, when considering \mathcal{M}_s vs \mathcal{M}_t , subjectively chooses a set of "neutral data," \tilde{Y}_{st} , which is to represent a configuration of "no evidence for either model." (Guidelines for specifying \tilde{Y}_{st} are discussed in Section 2.2 below.) The component-specific conditional neutral-data comparison paralleling (4) is

$$NDC_{st,i}(\boldsymbol{Y}_i|\boldsymbol{\phi},\boldsymbol{\tau}) = \frac{BF_i(\boldsymbol{Y}_i|\boldsymbol{\phi},\boldsymbol{\tau})}{BF_i(\boldsymbol{\tilde{Y}}_{st,i}|\boldsymbol{\phi},\boldsymbol{\tau})}.$$
(5)

This quantity may be interpreted in various ways discussed in Spitzner (2011), and is to serve in place of (4) when assessing the weight of evidence for a "signal." The complete conditional neutral-data comparison, to substitute for (3), is

$$NDC_{st}(\boldsymbol{Y}|\boldsymbol{\phi},\boldsymbol{\tau}) = \frac{BF_{st}(\boldsymbol{Y}|\boldsymbol{\phi},\boldsymbol{\tau})}{BF_{st}(\tilde{\boldsymbol{Y}}_{st}|\boldsymbol{\phi},\boldsymbol{\tau})}.$$
(6)

One practical advantage of neutral-data comparisons over Bayes factors is that they are drastically less sensitive to the dispersion of the prior, and thus produce an informative assessments even when the prior is vague.

Although both Bayes factors and neutral-data comparisons are invariant to prior model probabilities, it is straightforward to deduce that the choice of neutral data has a direct impact on those probabilities, which, in turn, impacts the posterior distribution as a whole. Such connections are understood through an alternative characterization of a Bayes factor, by which formula (3) is identical to the ratio of posterior to prior odds; by rearranging terms, this characterization implies

$$\rho_{st}(\boldsymbol{\phi}, \boldsymbol{\tau}) = \tilde{\rho}_{st}(\boldsymbol{\phi}, \boldsymbol{\tau}) / BF_{st}(\boldsymbol{Y}|\boldsymbol{\phi}, \boldsymbol{\tau}), \tag{7}$$

having written $\rho_{st}(\phi, \tau) = P[\mathcal{M}_s|\phi, \tau]/P[\mathcal{M}_t|\phi, \tau]$ and $\tilde{\rho}_{st}(\phi, \tau) = P[\mathcal{M}_s|\tilde{\mathbf{Y}}_{st}, \phi, \tau]/P[\mathcal{M}_t|\tilde{\mathbf{Y}}_{st}, \phi, \tau]$. In formulating an analysis, one of $\rho_{st}(\phi, \tau)$ or $\tilde{\rho}_{st}(\phi, \tau)$ is to be specified by the analyst; typical settings are $\rho_{st}(\phi, \tau) = 1$ when working with Bayes factors, or $\tilde{\rho}_{st}(\phi, \tau) = 1$ when working with neutral-data comparisons, which are applied in the analysis of Section 3. Instead of solving (7), the analyst may instead work with the formulas

$$P[\mathcal{M}_s|\mathbf{Y}, \boldsymbol{\phi}, \boldsymbol{\tau}] / P[\mathcal{M}_t|\mathbf{Y}, \boldsymbol{\phi}, \boldsymbol{\tau}] = \rho_{st}(\boldsymbol{\phi}, \boldsymbol{\tau}) BF_{st}(\mathbf{Y}|\boldsymbol{\phi}, \boldsymbol{\tau}) = \tilde{\rho}_{st}(\boldsymbol{\phi}, \boldsymbol{\tau}) NDC_{st}(\mathbf{Y}|\boldsymbol{\phi}, \boldsymbol{\tau}), \tag{8}$$

which are implied from (3), (6), and (7).

2.2 Selecting neutral data

The proposed scheme for specifying $\tilde{\mathbf{Y}}_{st}$ builds on insights deduced in Spitzner (2011), in which asymptotic settings of neutral data are identified that induce the neutral-data comparison to reflect the standard asymptotic behavior of a Bayes factor. In particular, a proposed default setting for *Case 1*, the Gaussian case, has $\|\tilde{\mathbf{Z}}_{st,i}\|^2 \approx \nu \log n$, as $n \to \infty$, where $\tilde{\mathbf{Z}}_{st,i}$ is the neutral-data analogue to $\mathbf{Z}_i = n^{1/2} \Sigma^{-1/2} \bar{\mathbf{Y}}_i$, writing $\bar{\mathbf{Y}}_i = n^{-1} \sum_{j=1}^n \mathbf{Y}_{ij}$. The present investigation contributes to that line of inquiry by translating Spitzner's (2011) asymptotic setting to a meaningful *exact* setting that may be manipulated for good performance within the variable selection context. The approach concentrates on identifying a suitable target value for $BF_i(\tilde{\mathbf{Y}}_{st} | \boldsymbol{\phi}, \tau)$, the denominator in formula (5).

To this end, let us continue to work within *Case 1*, and consider the formula for the relevant Bayes factor implied by the prior $\theta_i | \tau^2 \sim G(\mathbf{0}, \tau^2 \mathbf{\Delta})$,

$$BF_{i}(\tilde{\boldsymbol{Y}}_{st}|\boldsymbol{\phi}_{i},\boldsymbol{\tau}) = |\boldsymbol{I} + \tau^{2}n\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Delta}\boldsymbol{\Sigma}^{-1/2}|^{1/2}\exp\left\{-\frac{1}{2}\tilde{\boldsymbol{Z}}_{st,i}^{T}\boldsymbol{W}\tilde{\boldsymbol{Z}}_{st,i}\right\},$$
(9)

where $W = \{I + \Sigma^{1/2} \Delta^{-1} \Sigma^{1/2} / (\tau^2 n)\}^{-1}$. Identification of a target value for (9) is made using ideas borrowed from Spiegelhalter and Smith (1982), by which the desired value is to reflect two criteria:

- a. The sample size, n, is the smallest possible that permits comparison of \mathcal{M}_s vs \mathcal{M}_t
- b. The target value indicates maximum support for \mathcal{M}_s .

Application of criterion (b), especially, requires a certain level of judgment by the analyst. Spiegelhalter and Smith (1982) argue for the target value $BF_i(\tilde{Y}_{st}|\phi_i, \tau) = 1$, having set n = 1, which is sensible for their particular context. In the present context, the target value will ultimately be determined from a performance evaluation described in Section 2.3, below. However, a step in that direction is possible by adding a third criterion, which customizes Spiegelhalter and Smith's guidelines to the motivating concerns of the neutral-data comparisons framework:

c. The target value responds to changes in prior dispersion in the manner of a Bayes factor.

Germane to criterion (c) is the property that if the prior on a ν -dimensional parameter is from a scale family, $\pi(\theta) = \tau^{-\nu} \pi^*(\theta/\tau)$, then, in regular problems, the corresponding Bayes factors are related according to $BF \approx \tau^{\nu} BF^*$ as $\tau \to \infty$. For example, this property is evident in formula (9). Bringing together criteria (a), (b), and (c) points to a specific parameterization of the target value given by

$$BF_i(\tilde{\boldsymbol{Y}}_{st}|\boldsymbol{\phi},\boldsymbol{\tau}) = \tau^{\nu} |\boldsymbol{\Sigma}^{-1/2} \boldsymbol{\Delta} \boldsymbol{\Sigma}^{-1/2}|^{1/2} / \gamma_{st}, \qquad (10)$$

where the quantity γ_{st} is to be determined in Section 2.3.

Once identifying a target value, it is straightforward to solve (9) for a suitable specification of neutral data, which produces

$$\tilde{\boldsymbol{Z}}_{st,i}^{T} \boldsymbol{W} \tilde{\boldsymbol{Z}}_{st,i} = \nu \log n - \log |\boldsymbol{W}| + 2 \log \gamma_{st}.$$
(11)

Of course, it is not strictly necessary to solve for neutral data, since only the target value (10) is needed to deduce, by (4), that the component-specific conditional neutral-data comparison is

$$NDC_{st,i}(\boldsymbol{Y}|\boldsymbol{\phi},\boldsymbol{\tau}) = \gamma_{st} n^{\nu/2} |\boldsymbol{W}|^{1/2} \exp\left\{-\frac{1}{2}\boldsymbol{Z}_i^T \boldsymbol{W} \boldsymbol{Z}_i\right\}.$$
 (12)

These ideas readily extend to *Case 2*, the regular non-Gaussian case, in which asymptotic Gaussianity of the maximum-likelihood value, $\hat{\theta}_i$, sets up a parallel argument for specifying neutral data. In this case, it is proposed that one work as if the $\hat{\theta}_i$ serve in place of the data, strictly for the purpose of selecting neutral data, and subsequently recalculate the neutral-data conditional Bayes factor (9) using the inverted "null-value" Fisher information matrix $I(\theta_i^0, \phi_i)^{-1}$ in place of Σ . Taking the prior to have the form (2), the target value is

$$BF_{n_{i},i}(\tilde{\boldsymbol{Y}}_{st}|\boldsymbol{\phi}_{i},\boldsymbol{\tau}) = \tau_{i}^{\nu_{i}}|\boldsymbol{I}(\boldsymbol{\theta}_{i}^{0},\boldsymbol{\phi}_{i})^{1/2}\boldsymbol{\Delta}_{i}(\boldsymbol{\phi}_{i},\boldsymbol{\tau})\boldsymbol{I}(\boldsymbol{\theta}_{i}^{0},\boldsymbol{\phi}_{i})^{1/2}|^{1/2}/\gamma_{st},$$
(13)

where ν_i is the dimension of θ_i . The desired conditional neutral-data comparison, the analogue to (12), is calculated as the ratio of (4) to (13). This proposed solution for *Case 2* cannot justly be called an exact solution, as (10) has been described for *Case 1*, but it nevertheless refines the type of asymptotic results derived in Spitzner (2011).

2.3 Asymptotic consistency in variable selection

Translating the calibration scheme outlined descriptively in Section 2.1, the constants γ_{st} are to be selected in such a way that, in an elementary test for which \mathcal{M}_s vs \mathcal{M}_t is such that $A_s = A_t \cup \{i\}$, the constant γ_{st} is small when $|A_s|$ is small and large when $|A_s|$ is large. The specific proposed setting of this quantity is

$$\gamma_{st} = |A_s|. \tag{14}$$

Justification for this setting derives from an examination of asymptotic consistency in a specialized version of the current setup. However, before considering asymptotic consistency it is interesting to note that (14) reduces to $\gamma_{st} = 1$ where there are only two models. The setting $\gamma_{st} = 1$ regarded here, for illustrative purposes, to define an "unadjusted" setting for neutral data, when multiplicity is to be ignored. Spitzner (2014) argues that $\gamma_{st} = 1$ would also serve well as a general "default" setting to use when defining neutral data, but such an interpretation is not pursued here.

Regarding asymptotic consistency, the specialized setup alluded to above constrains the choice of prior model odds, the $\rho_{st}(\phi, \tau)$ in (7), to reflect that case where the events { $\theta_i = \theta_i^0$ } (*i.e.*, "non-signals") are conditionally independent. It follows that any conditional posterior model probability factors according to

$$P[\mathcal{M}_s|\mathbf{Y}, \boldsymbol{\phi}, \boldsymbol{\tau}] = \left\{ \prod_{i \in A_s} P[\{\boldsymbol{\theta}_i = \boldsymbol{\theta}_i^0\} | \mathbf{Y}_i, \boldsymbol{\phi}_i, \boldsymbol{\tau}] \right\} \left\{ \prod_{i \notin A_s} P[\{\boldsymbol{\theta}_i = \boldsymbol{\theta}_i^0\}^c | \mathbf{Y}_i, \boldsymbol{\phi}_i, \boldsymbol{\tau}] \right\}.$$
(15)

In this context, neutral data need no longer be test-specific (*i.e.*, the subscript of \tilde{Y}_{st} can go away), but the same \tilde{Y} , specified component by component, can be applied to every \mathcal{M}_s vs \mathcal{M}_t . Thus, for example, in *Case 1*, the component posterior probabilities in (15) are

$$P[\{\boldsymbol{\theta}_i = \boldsymbol{\theta}_i^0\} | \boldsymbol{Y}_i, \boldsymbol{\phi}_i, \boldsymbol{\tau}] = \left\{ 1 + e^{\frac{1}{2}(\boldsymbol{Z}_i^T \boldsymbol{W} \boldsymbol{Z}_i - \tilde{\boldsymbol{Z}}_i^T \boldsymbol{W} \tilde{\boldsymbol{Z}}_i)} \right\}^{-1},$$
(16)

assuming $\tilde{\rho}_{st}(\phi, \tau) = 1$, where the same \tilde{Z}_i always serves as neutral-data in component *i*.

One way to implement a formal asymptotic analysis is to suppose there is a hypothetical connection between sample size, n, and the number of components, $p = p_n$, such that $p_n \to \infty$ as $n \to \infty$. Let us adopt that supposition, and furthermore suppose that one particular model, \mathcal{M}_n^* , with associated subset $A_n^* \subset \{1, \ldots, p_n\}$ and parameter $\theta^* = (\theta_i^* : i \notin A_n^*)$, is singled out as the "true" model. The goal of the analysis is to identify settings of \tilde{Y} that induce asymptotic consistency, $P[\mathcal{M}_n^*|Y] \to 1$ as $n \to \infty$, for data generated under \mathcal{M}_n^* . (See Polson and Scott, 2011, for an alternative to this sort of analysis.) The following result identifies a class of solutions to this problem.

Theorem 1. Suppose, in Case 1, that $\tilde{\rho}_{st}(\phi, \tau) = 1$, and data are generated under the model \mathcal{M}_n^* . One has $P[\mathcal{M}_n^*|\mathbf{Y}] \to 1$ if, for some sequence $C_n \to \infty$, neutral-data $\tilde{\mathbf{Z}}_i$ are such that

(i) $\tilde{\boldsymbol{Z}}_{i}^{T} \boldsymbol{W} \tilde{\boldsymbol{Z}}_{i} \geq \nu \log n - \log |\boldsymbol{W}| + 2 \log |A_{n}^{*}| + C_{n}$

(ii) $\tilde{\boldsymbol{Z}}_i^T \boldsymbol{W} \tilde{\boldsymbol{Z}}_i \leq n \boldsymbol{\theta}_i^{*T} \boldsymbol{\Sigma}^{-1/2} \boldsymbol{W} (\boldsymbol{I} + \boldsymbol{W})^{-1} \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\theta}_i^* - 2 \log |A_n^{*c}| - C_n.$

Similarly, one has $\liminf_n P[\mathcal{M}_n^*|\mathbf{Y}] > 0$ if the criteria (i) and (ii) are instead satisfied for a convergent sequence C_n .

For perspective on Theorem 1, it is worthwhile to consider the asymptotic context laid out by Fan and Lv (2008) that articulates the notion of "sparse signals" in "ultra-high dimensional" space. In that context, the number of components is allowed to increase at up to an exponential rate, constrained only by $\log p_n = O(n^a)$ for some a > 0; concurrently, signals may arise at a very slow rate at continually weaker strength: for some b > 0 with a < 1-b, it is only necessary that some c > 0 exist such that $\min_{i \in A_n^*} \theta_i^{*T} \Sigma^{-1} \theta_i^* \ge cn^{-b}$. Under these constraints, the conditions of Theorem 1 are satisfied by the settings $\tilde{Z}_i^T W \tilde{Z}_i = n^d$ and $C_n = n^{(d-a)/2}$, whenever a < d < 1-b. (To see this, note that $\log |A_n^*| \le \log p_n \le n^a$, eventually, in criterion *i*, and that $W \to I$ in criterion *i*). This means that the neutral-data formulation to the variable selection problem is among the very few statistical procedures (*cf.* Fan and Lv, 2010, for others) that are able to identify a sparse "true" model in ultra-high dimensions.

Nevertheless, despite its desirable properties, the setting $\tilde{Z}_i^T W \tilde{Z}_i = n^d$ may be difficult to work with in practice, because the exponent, d, is a vaguely defined parameter of the asymptotic context. The setting $\gamma_{st} = |A_s|$, as in (14), is preferred for its definitiveness and precision, and is regarded to define an *adaptive* solution that is intended to induce one of the asymptotically consistent settings identified in Theorem 1. To see how (14) forms an adaptive solution, suppose $\mathcal{M}_s vs \mathcal{M}_t$ arises within a reversible-jump MCMC algorithm that is evolving in a region near \mathcal{M}_n^* . (It is assumed the reader has a basic familiarity with the reversible-jump algorithm; otherwise, see Robert and Casella, 1999.) In that scenario, the count, $|A_s|$, of non-signals in \mathcal{M}_s is near to the same count in the true model, $|A_n^*|$, hence the $\tilde{Z}_{st,i}^T W \tilde{Z}_{st,i}$ defined by the setting (14) within formula (11) likely resembles an asymptotically consistent setting of $\tilde{Z}_i^T W \tilde{Z}_i$ identified in Theorem 1.i. Such arguments do not constitute a proof of asymptotic consistency, but they are compelling, nonetheless, in suggesting that the rule (14) gives rise to an effective multiplicity adjustment.

The adaptive solution (14) may be regarded as incorporating a mechanism for "estimating" $|A_n^*|$, the count of non-signals in the true model. Other possible mechanisms intended for the same purpose might also be considered. For example, Berry and Berry, formulate a hierarchical solution in which a hyperprior is specified on $|A_n^*|$. It is not clear, however, which hyper-priors, if any, will yield an asymptotically consistent solution. Through the argument made above, Theorem 1 is seen to offer support for the efficacy of the proposed adaptive solution. As will be seen in Section 3.3, below, the two solutions produces widely differing results.

3 Illustration: Adverse events in a vaccine trial

The proposed multiplicity adjustment is now explored in the analysis of the adverse-event data examined in Berry and Berry (2004). The context and likelihood function are described as *Case 3* in Section 2.1. Raw relative frequencies, $\hat{p}_{ijk} = Y_{ijk}/n_i$, and AE-type groupings into body systems are listed below in Table 1. Recall that, under model \mathcal{M}_s , the subset $A_{s,2}$ indicates the most worrisome AE-types, for which there is an increased probability of an adverse event under the vaccine treatment.

3.1 The multiplicity adjustment

The context of this problem is, at its core, a specialization of *Case 2*, for which the issue of specifying neutral data is discussed at the end of Section 2.2. The required asymptotic variance formula follows from standard asymptotic theory, by which the maximum-likelihood value, $\hat{\theta}_{jk}$, for θ_{jk} at fixed ϕ_{jk} is asymptotic cally Gaussian, with asymptotic variance

$$\sigma_{jk}^2 = \frac{1}{n_1 + n_2} \left\{ \frac{e^{\phi_{jk}}}{1 + e^{\phi_{jk}}} \left(1 - \frac{e^{\phi_{jk}}}{1 + e^{\phi_{jk}}} \right) \right\}^{-1}$$
(17)

when θ_{jk} is set to its "null" value, $\theta_{jk}^0 = 0$. Also needed, to apply criterion (a), is a notion of sample size "as small as possible," which in the present scenario is taken to mean $n_1 = n_2 = 1$, as is consistent with Spiegelhalter and Smith's (1982) handling multi-sample designs. It follows that, for a prior of the form $\theta_{jk}|\phi_{jk}, \tau \sim G(0, \tau^2 \delta_{jk}^2(\tau))$, which matches that in Berry and Berry (2004), the component-specific target



Figure 1: Evidence assessments of an adverse event on Berry and Berry's (2004) data for τ between 1 and 100, plotted on a standard scale of evidence. The left panel plots transformed Bayes factors, the middle panel plots transformed, unadjusted neutral-data comparisons, and the right panel plots transformed neutral-data comparisons that are adjusted for multiple models. Assessments are reported only of the four AE-types indexed by (j,k) = (3,4), (8,3), (10,4), and (10,6).

value (10) is

$$BF_{jk}(\tilde{\boldsymbol{Y}}_{st}|\boldsymbol{\phi},\boldsymbol{\tau}) = \frac{\tau \delta_{jk}(\boldsymbol{\tau})}{\gamma_{st}} \sqrt{\frac{1}{2} \left\{ \frac{e^{\phi_{jk}}}{1 + e^{\phi_{jk}}} \left(1 - \frac{e^{\phi_{jk}}}{1 + e^{\phi_{jk}}} \right) \right\}^{-1}}.$$
 (18)

The hierarchical parameters, τ , and other remaining aspects of the prior are defined in Section 3.2, below.

Recall that (18) is only needed when M_s vs M_t forms an elementary test such that $A_{s,0} = A_{t,0} \cup \{(j,k)\}$ for some $(j,k) \in \Omega$, which implies that either $A_{s,1} = A_{t,1} - \{(j,k)\}$ or $A_{s,2} = A_{t,2} - \{(j,k)\}$. Only in those cases is the multiplicity adjustment explicitly defined, through γ_{st} . In the present context, the correct setting, translating from (14), is $\gamma_{st} = |A_{s,0}|$.

3.2 The hierarchical prior

When formulating a prior for flagging AE-types, Berry and Berry suggest consideration of at least the following three issues: the scientific relationships among all of the AE-types; the total number of AE-types; and, the relationship between AE-types that are flagged and not flagged. Berry and Berry incorporate the latter two considerations into a hierarchical specification of prior model probabilities, which serves, in effect, as a hyper-prior on $|A_{s,0}|$. In the present analysis, the hyper-prior is replaced by the proposed adaptive scheme, which incorporates the same two considerations by alternative means.

Berry and Berry's first consideration, the scientific relationships among AE-types, is incorporated into the continuous portion of the prior. That portion is duplicated exactly in the present analysis: Each $\phi_{jk}|\tau_{0A}^2, \tau_{0B}^2 \sim G\left(0, \tau_{0A}^2 + \tau_{0B}^2 + \tau^2\right)$ and, independently, each $\theta_{jk}|\tau_{1A,j}^2, \tau_{1B}^2 \sim G\left(0, \tau_{1A,j}^2 + \tau_{1B}^2 + \tau^2\right)$, for hierarchical parameters $\tau^2, \tau_{H}^2, \tau_{0A}^2, \tau_{0B}^2, \tau_{1A,j}^2$ for $j \in J, \tau_{1B}^2$, which are collected into τ . (Berry and Berry construct their prior slightly differently through a "three-stage" hierarchy, but it readily collapses to the two-stage form just described.) Among the hierarchical parameters, τ^2 and τ_{H}^2 are fixed constants to be set explicitly, while $\tau_{0A}^2, \tau_{0B}^2, \tau_{1A,j}^2$, and τ_{1B}^2 are independent parameters such that $\tau_{H}^2/\tau_{0A}^2 \sim \chi_{\kappa}$, $\tau_{H}^2/\tau_{0B}^2 \sim \chi_{\kappa}, \tau_{H}^2/\tau_{1A,j}^2 \sim \chi_{\kappa}$, and $\tau_{H}^2/\tau_{1B}^2 \sim \chi_{\kappa}$ for an additional prior parameter κ . Berry and Berry specify $\tau^2 = 10, \tau_{H}^2 = 2$, and $\kappa = 6$; these and other settings are examined in the analysis below.

Calculations are made using the reversible-jump MCMC algorithm alluded to in Section 2.3. The algorithm is implemented by making reversible jumps on individual θ_{jk} at fixed values of the remaining parameters. Proposed jumps from $(\theta_{s,jk}^0, \phi_{s,jk})$ in model \mathcal{M}_s to $(\theta_{t,jk}, \phi_{t,jk})$ in model \mathcal{M}_t are defined through the invertible transformation $\phi_{t,jk} = \frac{1}{4}\phi_{s,jk}\{4 + a(\boldsymbol{u})\}$ and $\theta_{t,jk} = \frac{1}{4}\phi_{s,jk}\{4 - a(\boldsymbol{u})\}$, where $a(\boldsymbol{u}) = 1/\boldsymbol{u} - 1/(1 - \boldsymbol{u})$ and $\boldsymbol{u} \sim \text{Beta}(2, 2)$.

				B&B		NDC, unadj.		NDC, adj.	
j	k	\hat{p}_{1jk}	\hat{p}_{2jk}	eq.	AE	eq.	AE	eq.	AE
1	1	0.303	0.385	0.762	0.211	0.392	0.601	0.956	0.043
1	2	0.197	0.230	0.827	0.122	0.798	0.180	0.992	0.007
1	3	0.000	0.014	0.796	0.101	0.054	0.944	0.947	0.053
1	4	0.008	0.020	0.813	0.100	0.765	0.216	0.992	0.008
1	5	0.152	0.182	0.826	0.116	0.798	0.179	0.993	0.007
3	1	0.015	0.047	0.821	0.117	0.328	0.666	0.949	0.050
3	2	0.000	0.014	0.835	0.083	0.067	0.932	0.738	0.261
3	3	0.000	0.014	0.812	0.101	0.068	0.930	0.859	0.141
3	4	0.076	0.162	0.743	0.231	0.030	0.969	0.517	0.483
3	5	0.008	0.020	0.823	0.093	0.767	0.214	0.992	0.008
3	6	0.053	0.014	0.805	0.050	0.211	0.003	0.910	0.000
3	7	0.144	0.128	0.849	0.076	0.890	0.044	0.996	0.002
5	1	0.015	0.020	0.717	0.136	0.882	0.083	0.996	0.003
6	1	0.015	0.000	0.666	0.087	0.039	0.001	0.431	0.000
8	1	0.000	0.014	0.655	0.185	0.023	0.977	0.749	0.251
8	2	0.015	0.014	0.661	0.153	0.898	0.048	0.997	0.001
8	3	0.326	0.507	0.214	0.780	0.001	0.999	0.033	0.967
9	1	0.008	0.027	0.900	0.059	0.560	0.428	0.981	0.019
9	2	0.015	0.027	0.901	0.058	0.828	0.147	0.994	0.005
9	3	0.015	0.007	0.896	0.040	0.858	0.026	0.995	0.001
9	4	0.061	0.088	0.906	0.062	0.758	0.223	0.991	0.008
9	5	0.152	0.189	0.897	0.083	0.754	0.228	0.990	0.009
9	6	0.008	0.014	0.898	0.047	0.870	0.101	0.996	0.003
9	7	0.061	0.088	0.906	0.061	0.758	0.223	0.991	0.008
9	8	0.106	0.101	0.904	0.051	0.894	0.055	0.997	0.002
9	9	0.008	0.020	0.903	0.051	0.766	0.215	0.992	0.008
9	10	0.008	0.014	0.905	0.042	0.870	0.101	0.996	0.003
9	11	0.008	0.020	0.907	0.050	0.769	0.212	0.992	0.008
10	1	0.000	0.027	0.859	0.087	0.001	0.999	0.065	0.935
10	2	0.000	0.014	0.860	0.070	0.001	0.999	0.945	0.054
10	3	0.008	0.014	0.868	0.062	0.872	0.099	0.996	0.003
10	4	0.023	0.088	0.784	0.190	0.011	0.989	0.287	0.713
10	5	0.015	0.041	0.852	0.099	0.540	0.450	0.978	0.022
10	6	0.008	0.054	0.836	0.126	0.014	0.986	0.364	0.636
10	7	0.015	0.027	0.862	0.076	0.828	0.148	0.994	0.005
10	8	0.015	0.000	0.852	0.048	0.009	0.000	0.798	0.000
10	9	0.015	0.007	0.855	0.055	0.857	0.026	0.995	0.001
11	1	0.015	0.000	0.721	0.079	0.008	0.000	0.789	0.000
11	2	0.106	0.122	0.757	0.102	0.858	0.111	0.995	0.004
11	3	0.008	0.014	0.749	0.121	0.872	0.099	0.996	0.003

Table 1: Posterior probabilities on Berry and Berry's (2004) adverse-event data. Index values are listed for body system (j) and AE-type (k) in the first pair of columns; the remaining columns list raw adverseevent relative frequencies, followed by the posterior probabilities from Berry and Berry's (2004) analysis (headed "B & B"), then those derived from unadjusted neutral-data comparisons, and finally those derived from neutral-data comparisons that are adjusted for multiple models. The columns labeled "eq." list posterior probabilities that the rates of adverse events between treatment and control conditions are equal, and those labeled "AE" list posterior probabilities of an adverse event.

3.3 Analysis results

Results of the analysis, in a variety of configurations, are indicated in Figure 1 and Table 1. Data analysis is carried out repeatedly across twenty values of the scale parameter τ^2 in the range $1 \le \tau \le 100$, which forms the horizontal axis in each panel of Figure 1, all while holding constant the ratio $\tau^2/\tau_H^2 = 5$, and the parameter $\kappa = 6$. Each panel plots the relevant assessments after having been transformed according to $2\log\{P[\mathcal{E}_{(j,k)}|\mathbf{Y}]/(1-P[\mathcal{E}_{(j,k)}|\mathbf{Y}])\}$, where $\mathcal{E}_{(j,k)}$ collects all models \mathcal{M}_s such that $(j,k) \in A_{s,2}$. That is, the values plotted in Figure 1 indicate support for an adverse event of type j in body-system k, transformed for interpretation on Kass and Raftery's (1995) scale of evidence. Only four AE-types are represented in Figure 1, those with index values (j,k) = (3,4), (8,3), (10,4), and (10,6), which were selected by Berry and Berry for having been flagged in a previous frequentist analysis. Selected results for all AE-types are listed in Table 1, but only at $\tau = 100$.

The three panels of Figure 1 show results of the proposed procedure in three configurations of the discrete prior: the configuration represented in the left panel has $\rho_{st}(\phi, \tau) = 1$ in formula (7), hence the results shown are from Bayes factors; that of the middle panel has $\tilde{\rho}_{st}(\phi, \tau) = 1$ and $\gamma_{st} = 1$, hence the results are from "unadjusted" neutral-data comparisons (*cf.* the discussion below formula 14); and, the right panel is like the middle one, but with $\gamma_{st} = |A_{s,0}|$, thereby specifying the proposed multiplicity adjustment. The latter two configurations are also represented in Table 1, as posterior probabilities, alongside Berry and Berry's results for comparison.

As expected, and illustrated in Figure 1, support for an adverse event drastically weakens as τ grows large when it is reported by a Bayes factor, but it eventually stabilizes when it is reported by a neutral-data comparison. "Strong" evidence of an AE (a reported value above 6) of every selected type is indicated in the middle panel, with each neutral-data comparison stabilizing (by coincidence) near the maximum of the corresponding Bayes factor in the left panel. Comparison with the right panel illustrates how the multiple-model adjustment weakens evidence across the board, so much that "strong" support of an AE remains only for the AE-type at (j, k) = (8, 3). The adjustment ultimately induces a beneficial clarifying effect of reducing the collection of several suspicious AE-types to just one that is to be flagged.

In Table 1, it is seen that Berry and Berry's prior weakens the reported evidence even more, to the point where no strong evidence of an AE is exhibited among any of the forty AE-types. Consider that on Berry and Berry's results the transformation $2\log\{P[\mathcal{E}_{(j,k)}|\mathbf{Y}]/(1 - P[\mathcal{E}_{(j,k)}|\mathbf{Y}])\}$ yields the values -2.41, 2.53, -2.90, and -3.87 for (j,k) = (3,4), (8,3), (10,4), and (10,6). These transformed assessments are much different than those of the adjusted neutral-data comparisons, and it is interesting that Berry and Berry's results are also hard to place among the patterns exhibited in Figure 1: even at $\tau = 100$, the Bayes factors in the left panel report much stronger evidence than those of Berry and Berry, and yet the neutral-data comparisons of the other two panels are well past the point of having stabilized with respect to τ . From this perspective, the prior dependencies introduced in Berry and Berry's hierarchical discrete prior are seen to have an astoundingly strong effect.

4 Conclusions

The main conclusion is that the proposed multiplicity adjustment for variable selection is capable of incorporating relevant scientific goals, is straightforward and not difficult to implement using MCMC simulation, and is effective at clarifying evidence of signals and non-signals. The methodology is distinct from widely used decision-theoretic formulations of a multiplicity adjustment, and distinct from Berry and Berry's approach based on hierarchical modeling, although the latter is similar in formulating the adjustment through the discrete portion of the prior. A reanalysis of Berry and Berry's adverse-event data set suggests that the original analysis applies an adjustment that is perhaps too drastic, while the proposed solution, on the basis of its support from asymptotic theory, produces a more judicious assessment of evidence.

The article also contributes to the developing theory of neutral-data comparisons, by demonstrating a role for Spiegelhalter and Smith's (1982) criteria for eliciting imaginary data. Further refinement of the criteria laid out in Section 2, and of our general understanding of neutral data, is possible and desired as the concept is developed for more complex applications. For instance, the definition of "minimal sample size" can be unclear in some scenarios, although various ideas have been proposed to make that concept precise; *e.g.*, see Berger and Pericchi (1996). Such issues, and the impulse to resolve them, are no doubt important to future development of the concepts used here.

The proposed multiplicity adjustment is intriguing for its simplicity and apparent effectiveness, but it remains to explore the technique in more general scenarios. The rule (14) has been shown to yield outstanding performance when the Y_i are independent; as for the dependent case, some preliminary results

have been obtained pointing to the same rule proposed here, but that investigation is ongoing and its results will be reported elsewhere. Beyond variable selection, the general scheme explored here is so straightforward that it would seem to adapt well to complex settings such as partition analysis or graphical models. In these more complex scenarios, tracking the connections between models presents a considerable challenge to theoretical analysis; nevertheless, it is hoped that the ideas presented here offer a productive means of exploring these problems and developing effective analysis approaches to them.

A Appendix

Proof. (THEOREM 1) The inequality $\log(1+x) \leq x$ provides that $-\log P[\mathcal{M}_n^*|Y]$ is bounded above by

$$B_n = \sum_{i \in A_n^*} \exp\{\frac{1}{2} (\boldsymbol{Z}_i^T \boldsymbol{W} \boldsymbol{Z}_i - \boldsymbol{\tilde{Z}}_i^T \boldsymbol{W} \boldsymbol{\tilde{Z}}_i)\} + \sum_{i \notin A_n^*} \exp\{\frac{1}{2} (\boldsymbol{\tilde{Z}}_i^T \boldsymbol{W} \boldsymbol{\tilde{Z}}_i - \boldsymbol{Z}_i^T \boldsymbol{W} \boldsymbol{Z}_i)\}.$$

Because the terms in B_n are independent and each is non-negative, an extension of the Borel-Cantelli lemmas (*cf.* Billingsley, 1995, prob. 22.3, p. 294) provides that B_n converges almost surely, and therefore $\liminf_n P[\mathcal{M}_n^*|\mathbf{Y}_n] > 0$, whenever its expectation, $E[B_n]$, converges; also $P[\mathcal{M}_n^*|\mathbf{Y}] \to 1$ when $E[B_n] \to -\infty$. Each $\mathbf{Z}_i^T \mathbf{W} \mathbf{Z}_i$ is a quadratic form of Gaussian random variables, whose properties are well known. The corresponding moment generating functions, evaluated under \mathcal{M}_n^* , imply $E[B_n] = \sum_{i \in \mathcal{A}_n^*} |\mathbf{I} - \mathbf{I}| \mathbf{U}_n^*$.

 $W|^{-1/2} \exp\{-\frac{1}{2}\tilde{Z}_i^T W \tilde{Z}_i\} + \sum_{i \notin A_n^*} |I + W|^{-1/2} \exp[\frac{1}{2}\{\tilde{Z}_i^T W \tilde{Z}_i - n\theta_i^{*T} \Sigma^{-1/2} W (I + W)^{-1} \Sigma^{-1/2} \theta_i^*\}],$ which links to criteria (*i*) and (*ii*) in the manner they describe conditions under which $E[B_n]$ convergences.

REFERENCES

- Bayarri, M. J., Berger, J. O., Forte, A., and García-Donato, G. (2012). Criteria for model choice with application to variable selection. *Annals of Statistics*, 40:1550-1577.
 Berger, J. O., and Pericchi, L. (1996). The intrinsic Bayes factor for model selection and prediction.
- Berger, J. O., and Pericchi, L. (1996). The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association*, 91:109- 122.
- Berry, D. A., and Hochberg, Y. (1999), Bayesian perspectives on multiple comparisons, *Journal of Statistical Planning and Inference*, 82:215-227.
- Berry, S. M., and Berry, D. A. (2004), Accounting for multiplicities in assessing drug safety: a three-level hierarchical mixture model, *Biometrics*, 60:418-426.
- Billingsley, P. (1995). Probability and measure. 3rd ed. New York: Wiley.
- Duncan, D. B. (1965), A Bayesian approach to multiple comparisons, Technometrics, 7:171-222.
- Fan, J., and Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *Journal* of the Royal Statistical Society B, 70:849-911.
- Fan, J., and Lv, J. (2010). A selective overview of variable selection in high dimensional feature space. *Statistica Sinica*, 20:101-148.
- Kass, R. E., and Raftery, A. E., (1995) Bayes factors, *Journal of the American Statistical Association*, 90:773-795.
- Kass, R. E., and Wasserman, L., (1995) A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion, *Journal of the American Statistical Association*, 90:928-934.
- Müller, P., Parmigiani, G., and Rice, K. (2007), FDR and Bayesian multiple comparison rules, *Bayesian Statistics 8*, (eds. J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West), Oxford: Oxford University Press, 349-370.
- O'Hagan, A. (1995), Fractional Bayes factors for model comparisons, *Journal of the Royal Statistical* Society B, 57:99-138.
- Polson, N., and Scott, J. G. (2011), Shrink globally, act locally: sparse Bayesian regularization and prediction, *Bayesian Statistics 9*, (eds. J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West), Oxford: Oxford University Press, 501-538.
- Robert, C. P., and Casella, G. (1999), Monte Carlo Statistical Methods, New York: Springer.
- Schwarz, G. (1978). Estimating the dimension of a model. Annals of Statistics, 6:461-464.
- Scott, J. G., and Berger, J. O. (2006). An exploration of aspects of Bayesian multiple testing. *Journal of Statistical Planning and Inference*, 136:2144-2162.
- Spiegelhalter, D. J., and Smith, A. F. M. (1982), Bayes factors for linear and log-linear models with vague prior information, *Journal of the Royal Statistical Society B*, 44:377-387.
- Spitzner, D. J. (2011) Neutral-data comparisons for Bayesian testing, Bayesian Analysis, 6:603-638.
- Spitzner, D. J. (2014) Neutral-data comparisons defining a spectrum between Bayes factors and the Schwarz criterion. Under review.