THE UNIVERSITY OF BRITISH COLUMBIA

DEPARTMENT OF STATISTICS

TECHNICAL REPORT #247

What are the limits of posterior distributions arising from nonidentified models, and why should we care?

Paul Gustafson

May 2009

What are the limits of posterior distributions arising from nonidentified models, and why should we care?

Paul Gustafson
Department of Statistics
University of British Columbia
gustaf@stat.ubc.ca

Version of: May 12, 2009

Abstract

In health research and other fields, the observational data available to researchers often fall short of the data that would ideally be available, due to inherent limitations of study design and data acquisition. Were they available, the ideal data might readily be analyzed via straightforward statistical models with desirable properties such as parameter identifiability. Conversely, realistic models for the available data, which incorporate uncertainty about the link between ideal and available data, may be nonidentified. While there is no conceptual difficulty in implementing Bayesian analysis with nonidentified models and proper prior distributions, it is important to know to what extent data can be informative about parameters of interest. Determining the large-sample limit of the posterior distribution is one way to characterize the informativeness of data. In some nonidentified models it is relatively straightforward to determine the limit, via a particular reparameterization of the model. In other nonidentified models, however, there is no such obvious approach. Thus an algorithm is developed to determine the limiting posterior distribution for at least some such harder models. The work is motivated by two specific nonidentified models which arise quite naturally, and the algorithm is applied to reveal how informative data are for these models.

Keywords: asymptotics; Bayesian inference; nonidentified models.

Paul Gustafson is Professor, Department of Statistics, University of British Columbia, Vancouver, BC V6T 1Z2 Canada (E-mail: *gustaf@stat.ubc.ca*). This research was funded via a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada and an Operating Grant from the Canadian Institutes of Health Research. The author would like to thank two anonymous referees for suggestions which improved the manuscript.

1 Introduction

In many scientific arenas where statistical methods are applied, resource limits and ethical concerns constrain study design and measurement quality. For instance, when observational epidemiological studies are conducted, seldom can concerns about measurement error, selection bias, and unobserved confounding be entirely mollified. Realistic statistical models for such studies then ought to involve two components: a model for the ideal but unavailable data, and a model linking this unattainable data to the available data. Typically one or more parameters in the former model are of scientific interest. However, particularly when faced with limited knowledge about the link between ideal and available data, concern about identification of the overall model may arise. Formally, a statistical model is nonidentified if multiple sets of parameter values correspond to the same distribution of observables. Identification is of course a key assumption underlying nice properties of model-based parameter estimators. Without it, for instance, one cannot expect root-n consistent estimation of parameters, even in highly parametric models. This is an inconvenient truth in light of the strong argument that realistic models for observational data will often be nonidentified. For general discussion of this point, see Greenland (2003, 2005).

From the mechanical perspective of transforming a proper prior distribution into a posterior distribution via the acquisition of data, lack of model identification presents no conceptual difficulties for a Bayesian analysis. Thus Bayesian analysis is often advocated in settings where study and data limitations do indeed lead to a nonidentified model (see, for example, Dendukuri and Joseph 2001; Georgiadis *et. al.* 2003; Gustafson and Greenland 2006; Hanson *et. al.* 2003; Menten, Boelaert and Lesaffre 2008; Scharfstein, Daniels and Robins 2003). It still seems very important, however, to gain theoretical understanding of how such inferential procedures will perform. And the situation is not very clear. For instance, the large-sample theory which guarantees similar performance for maximum likelihood and Bayesian procedures is clearly not applicable in a nonidentified model context. Thus a stand-alone theory is needed to elucidate the large-sample behaviour of Bayesian inferences arising from a nonidentified model.

To be more specific, given a nonidentified model, a proper prior distribution for its parameters, and a given set of true values for these parameters, we seek the limit of the posterior distribution as the sample size tends to infinity. Particularly, it would be helpful to know where the limiting posterior marginal distribution on a parameter of interest lies on the spectrum

between the extremes of (i), the corresponding prior distribution, and (ii), a point-mass at the true value. This would indicate the utility of collecting data for learning about the target parameter. Other proposals for measuring the informativeness of data in various nonidentified model contexts can be found in Neath and Samaniego (1997), Poirer (1998), and Xie and Carlin (2006).

1.1 The easier case

Toward understanding the ramifications of nonidentifiability for a Bayesian analysis, say that the nonidentified statistical model in question for observable data $\boldsymbol{D}$ is parameterized in scientific terms via a parameter vector $\boldsymbol{\lambda}$, and a proper prior distribution $\pi(\boldsymbol{\lambda})$ has been specified. In some cases the identification issue can be understood via a special reparameterization from $\boldsymbol{\lambda}$ to $(\boldsymbol{\phi}, \boldsymbol{\psi})$, such that the distribution of $\boldsymbol{D}$ is completely determined by $\boldsymbol{\phi}$ alone (but not by any lower-dimensional function of $\boldsymbol{\phi}$). This is termed a *transparent* reparameterization in Gustafson (2005a). Presuming the 'induced' model for $(\boldsymbol{D}|\boldsymbol{\phi})$ obeys regular asymptotics, the posterior marginal distribution for $\boldsymbol{\phi}$ will asymptotically tend to a point mass at the corresponding true value. In contrast, the posterior conditional distribution of $(\boldsymbol{\psi}|\boldsymbol{\phi}, \boldsymbol{D})$ will not depend on the data or sample size. Rather, it will simply be the same as the prior conditional distribution for $(\boldsymbol{\psi}|\boldsymbol{\phi})$. Thus it immediately follows that the large-sample limit of the posterior marginal distribution for $\boldsymbol{\psi}$ will equal the prior conditional given the true value of $\boldsymbol{\phi}$. Moreover, because $(\boldsymbol{\phi}, \boldsymbol{\psi})$ arises as a one-to-one reparameterization of $\boldsymbol{\lambda}$, determining the joint prior density for $(\boldsymbol{\phi}, \boldsymbol{\psi})$, and hence the prior conditional density for $(\boldsymbol{\psi}|\boldsymbol{\phi})$, is typically quite straightforward, as long as the mapping from $\boldsymbol{\lambda}$ to $(\boldsymbol{\phi}, \boldsymbol{\psi})$ can be inverted explicitly, or at least computationally.

This approach to examining how informative data are within a given nonidentified model has been applied in Gustafson (2005b, 2006, 2007), Gustafson and Greenland (2006), and Gustafson, Le and Saskin (2001). Interestingly, in many nonidentified models the transparent reparameterization is such that a sensible prior for $\boldsymbol{\lambda}$ induces substantial prior dependence between $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$. In fact, often the support of $\boldsymbol{\psi}$ depends on $\boldsymbol{\phi}$, inducing a structural dependence no matter what prior is specified in the original parameterization. Consequently the large-sample limit of the posterior marginal on $\boldsymbol{\psi}$ may be quite different from, and particularly more concentrated than, the prior marginal.

Overall then, an infinite data sample contains complete information about $\boldsymbol{\phi}$ and partial information about $\boldsymbol{\psi}$. Together this determines the limiting posterior distribution for a given

parameter of interest (often a function of both $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$). Without actually carrying out such determinations though, it is hard to intuit the extent to which inferences arising from a non-identified model might be usefully narrow versus uselessly wide. On balance, the nonidentified modelling contexts studied in previous work are characterized by narrower limiting posterior distributions than might be anticipated, particularly given the common view that identification is a basic requirement for useful statistical inference.

1.2 The Harder Case

Unfortunately, not all nonidentified models are amenable to the determination of limiting posterior distributions via reparameterization as described above. In some nonidentified models, a transparent reparameterization does not exist. In others there may be a mapping from $\boldsymbol{\lambda}$ to a 'candidate' $(\boldsymbol{\phi}, \boldsymbol{\psi})$ pair which *might* constitute a transparent reparameterization. However, it may not be possible to verify that the mapping is actually invertible. Moreover, a verification of invertibility is not enough. It must be possible to actually evaluate the inverse mapping in order to determine large-sample limits. Thus we seek an alternative approach to determining large-sample limits of posterior distributions arising from nonidentified models. In particular, an algorithm is developed which is applicable to at least some nonidentified models falling into the 'harder' category. The amount of analytic structure concerning parameters required to operationalize this algorithm is less than that of identifying a transparent reparameterization.

To motivate the work, Section 2 provides two examples of models which arise naturally, are nonidentified because of limitations on what variables can be observed, and do not admit a transparent reparameterization. Then Section 3 describes a Monte Carlo based algorithm for determining limiting posterior distributions when the structure that can be elucidated falls short of a transparent reparameterization. In Section 4 the algorithm is applied to both motivating examples. This addresses the question of what potential there is for useful inference notwithstanding the lack of identification. Some concluding remarks appear in Section 5. Three appendices provide technical details.

2 Motivating Examples

2.1 Two imperfect assessments for a binary trait applied to multiple populations

Consider a binary trait X, coded as absent ($X = 0$) or present ($X = 1$). Say that the prevalences of X in each of k populations, denoted by $r_i = Pr_i(X = 1)$, for $i = 1, \ldots, k$, are of interest. However, X can only be measured with error on individuals sampled from these populations. In particular, say that two imperfect diagnostic or screening tests are available, so that while X is unobservable, (X_1^*, X_2^*) is observable. The quality of each test can be described by its sensitivity and specificity, denoted $SN_j = Pr(X_j^* = 1|X = 1)$ and $SP_j = Pr(X_j^* = 0|X = 0)$, for $j = 1, 2$. We make the assumption of *nondifferential misclassification*, which posits that the distribution of $(X_1^*, X_2^*|X)$ is the same in each population.

There is a rich literature on this problem; see, for instance, Goetghebeur *et. al.* (2000), Hui and Zhou (1998), and Qu, Tan and Kutner (1996) for quite general discussions. If sensitivities and specificities are unknown, then identification is typically a concern. Consider the case of $k = 2$ populations, for instance. The general situation involves eight unknown parameters: $(r_1, r_2, SN_1, SN_2, SP_1, SP_2)$ plus two further parameters describing the conditional dependence between X_1^* and X_2^*, given $X = 0$ or $X = 1$. The data structure, comprised of a 2×2 table for (X_1^*, X_2^*) in each population, involves six degrees-of-freedom. Thus identification is ruled out. However, under the assumption that the two tests are conditionally independent given the trait, i.e., $(X_1^* \perp X_2^*|X)$, the number of parameters is reduced to six. Hui and Walter (1980) verify that this model is identified, by explicitly inverting the mapping from the parameters to the cell probabilities underlying the data tables. As a caveat though, this inversion breaks down if the trait prevalence is the same in both populations, and Gustafson (2005a,b) emphasizes that correspondingly poor estimation can result if the trait prevalences are similar in both populations.

Unfortunately, the assumption that X_1^* and X_2^* are conditionally independent given X can be problematic in many applications. For instance, consider the setting where the two tests are based on two different biologic assays. Any degree of commonality in the chemical bases of the two assays is likely to induce a positive conditional dependence. Or, in the setting of ascertaining X via both subject questionnaires and examination of subject medical records, dependence could arise if subjects who are more prone to give erroneous answers on a questionnaire are also more

prone to provide incomplete or wrong information to their physicians.

Given these sorts of concerns, Vacek (1985), Torrance-Rynard and Walter (1998), and Albert and Dodd (2004) look at the robustness of inferences when the conditional dependence structure of the tests is misspecified. Others have proposed models incorporating conditional dependence between tests, including the Bayesian models of Dendukuri and Joseph (2001), Hanson *et. al.* (2003), and Menten, Boelaert and Lesaffre (2008). With such modelling, however, identification issues are of paramount importance.

To focus on a specific question, say that data from $k = 3$ populations are available, and the conditional independence assumption is *not* invoked. Consequently there are nine unknown parameters: six to describe $(X_1^*, X_2^*|X)$ plus three trait prevalences. Commensurately, there are nine degrees-of-freedom represented in the three (X_1^*, X_2^*) data tables for the populations. Thus the parameter count is *compatible* with identification. However, Hanson and Johnson (2005) prove that this model is nonidentified nonetheless. Moreover, the form of this nonidentification cannot be addressed via a transparent reparameterization; the dimensions of $\boldsymbol{\lambda}$ and $\boldsymbol{\phi}$ would be the same, leaving no component $\boldsymbol{\psi}$. Thus a different tack must be followed to discover the extent to which the data are informative about parameters in this model.

2.2 Instrumental variable analysis with binary response

Say we wish to make inferences about an X-Y relationship adjusted for potential confounding variables $\boldsymbol{C} = (C_1, \ldots, C_p)$ and U, with the difficulty that U is not observable. We can attempt this task if we can observe an *instrumental variable* W that is associated with X and satisfies two conditional independencies. Specifically,

$$W \perp Y \quad | \quad (X, \boldsymbol{C}, U), \tag{1}$$

$$W \perp U \quad | \quad \boldsymbol{C}. \tag{2}$$

Conceptually, (1) says that the instrument itself does not 'drive' the outcome, while (2) says that within levels of the observed confounders, the instrument is independent of the unobserved confounder.

It is straightforward to understand how IV inference proceeds in the presence of linear mean structures. Say that $E(Y|X, U, \boldsymbol{C}) = \beta_0 + \beta_1 X + \beta_2 U + \boldsymbol{\beta_3 C'}$, with β_1 being of primary inferential interest. Also, say $E(X|W, \boldsymbol{C}) = \alpha_0 + \alpha_1 W + \boldsymbol{\alpha_2 C'}$, and $E(U|C) = \gamma_0 + \boldsymbol{\gamma_1 C'}$. Then assuming

(1) and (2) leads to

$$\begin{aligned} E(Y|W,\boldsymbol{C}) &= E\{E(Y|X,W,U,\boldsymbol{C})|W,\boldsymbol{C}\} \\ &= E\{E(Y|X,U,\boldsymbol{C})|W,\boldsymbol{C}\} \\ &= E\{\beta_0+\beta_1X+\beta_2U+\boldsymbol{\beta_3C'}|W,\boldsymbol{C}\} \\ &= \beta_0+\beta_1E(X|W,\boldsymbol{C})+\beta_2E(U|\boldsymbol{C})+\boldsymbol{\beta_3C'} \\ &= \delta_0+\delta_1W+\boldsymbol{\delta_2C'}, \end{aligned}$$

where, in particular, $\delta_1 = \alpha_1\beta_1$. Since δ_1 and α_1 are clearly estimable from the regressions corresponding to $(Y|W,\boldsymbol{C})$ and $(X|W,\boldsymbol{C})$ respectively, β_1 can be consistently estimated as $\hat{\beta}_1 = \hat{\delta}_1/\hat{\alpha}_1$.

It is also instructive to consider the further specialized case where (i) W is binary, and (ii) there are no measured confounders, i.e., $\boldsymbol{C}$ is absent. Then the estimator of β_1 specializes to

$$\hat{\beta}_1 = \frac{\hat{E}(Y|W=1)-\hat{E}(Y|W=0)}{\hat{E}(X|W=1)-\hat{E}(X|W=0)}. \tag{3}$$

In this case it is very apparent that the IV estimator operates by measuring change in Y with X as the ratio of change in Y with W to change in X with W.

By way of contrast, consider the situation, common in health research settings, where Y, X, W and U are all binary. Then a natural analogue to the linear models discussed above would be logistic regression models with main effect terms. To contrast with (3) in particular, consider the case where $\boldsymbol{C}$ is absent. Logistic regression models of the form

$$\text{logit}Pr(Y=1|X,U,W) = \beta_0+\beta_1X+\beta_2U, \tag{4}$$

$$\text{logit}Pr(X=1|U,W) = \alpha_0+\alpha_1W+\alpha_2U, \tag{5}$$

$$Pr(U=1|W) = \gamma, \tag{6}$$

could be employed in lieu of linear models.

As a whole, the model arising from (4) to (6) is nonidentified. It contains 7 unknown parameters, whereas the observable data take the form of two 2×2 tables corresponding to the distribution of $(Y,X|W)$ for $W=0$ and $W=1$. Since 6 cell probabilities cannot determine 7 unknown parameters uniquely, the model is nonidentified. More particularly, we will see that β_1, the parameter of typical interest, is not uniquely determined by the distribution of $(Y,X|W)$. Hence there is a marked qualitative difference. This target quantity is consistently estimable in

the linear model case via the simple and intuitive estimator (3), but is not consistently estimable in the logistic regression case. Both Terza (2006) and Johnston *et. al.* (2008) offer related comments on the challenges in translating instrumental variable methods from continuous outcome settings to binary outcome settings. We also note that more complex and sophisticated models have been put forth for the binary IV setting (for instance, see Pearl 2000, Ch. 8). For present purposes though, the important point is that the model defined by (4) through (6) is nonidentified, and it does not seem possible to obtain a transparent reparameterization. Thus we resort to the algorithm described in the next section, in order to understand the extent to which data are informative about the target parameter.

3 Determining the Large-Sample Behaviour

In what follows we give a Monte Carlo based algorithm for determining the large-sample behaviour of the posterior distribution in a nonidentified model, when a transparent parameterization cannot be determined but some lesser amount of structure can be elucidated. We express the algorithm in quite general terms, as we suspect it, or variants, may be useful in a range of nonidentified model settings.

Let $\boldsymbol{\theta} \in \Theta$ represent the p-dimensional parameter vector, with the parameterization perhaps chosen specially for the purpose of applying the following algorithm (hence may differ from the initial scientific parameterization λ). Let $\boldsymbol{D}$ represent the observable data, assumed to be generated under a specific true value of the parameter vector, denoted by $\boldsymbol{\theta}^*$. In recognition of the lack of identification, say that the distribution of $\boldsymbol{D}$ depends on $\boldsymbol{\theta}$ only through the q-dimensional parameter vector $\boldsymbol{\phi} = t(\boldsymbol{\theta})$, with $q \leq p$. Moreover, this reduction is assumed to be minimal, i.e., any function of $\boldsymbol{\phi}$ which completely determines the distribution of the data would necessarily be bijective. In the parlance of Barankin (1961), $\boldsymbol{\phi}$ is a 'minimal sufficient parameter'; see also Florens, Mouchart and Rolin (1990) for discussion of sufficiency on the parameter space. The 'induced' model for $(\boldsymbol{D}|\boldsymbol{\phi})$ is assumed to obey regular asymptotics, so that the posterior distribution of $(\boldsymbol{\phi}|\boldsymbol{D})$ will converge to a point mass at the true value $\boldsymbol{\phi}^* = t(\boldsymbol{\theta}^*)$ as the sample size tends to infinity. And since the posterior conditional distribution of $(\boldsymbol{\theta}|\boldsymbol{\phi}, \boldsymbol{D})$ equals the prior conditional distribution of $(\boldsymbol{\theta}|\boldsymbol{\phi})$, the limiting posterior distribution is simply the prior conditional for $(\boldsymbol{\theta}|\boldsymbol{\phi} = \boldsymbol{\phi}^*)$.

To determine the limiting posterior, we wish to partition the parameter vector as $\boldsymbol{\theta} =$

$(\boldsymbol{\theta_a}, \boldsymbol{\theta_b})$, such that the value of $\boldsymbol{\theta}$ can be recovered from $\boldsymbol{\phi} = t(\boldsymbol{\theta})$ and $\boldsymbol{\theta_a}$. In particular, say that $dim(\boldsymbol{\theta_a}) = p - q + u$ (with $0 \leq u < q$), and that the system of q equations $t(\boldsymbol{\theta}) = \phi$ can be reexpressed in terms of functions $g(\cdot;\cdot)$ and $h(\cdot;\cdot)$ as

$$\begin{aligned} \boldsymbol{\theta_b} &= g(\boldsymbol{\theta_a}; \boldsymbol{\phi}) \\ \mathbf{0_v} &= h(\boldsymbol{\theta_a}; \boldsymbol{\phi}) \end{aligned} \tag{7}$$

for some $v \in \{0, \ldots, u\}$. Note that if such a representation can be achieved with $u = v = 0$, then $(\boldsymbol{\phi}, \boldsymbol{\theta_a})$ comprises a reparameterization of $\boldsymbol{\theta}$, and the straightforward analysis outlined in Section 1.1 can be applied. With $u > 0$, however, $(\boldsymbol{\phi}, \boldsymbol{\theta_a})$ has $p+u$ elements, and might be called an 'overparameterization' from which $\boldsymbol{\theta}$ can be readily determined. Note also that (7) comprises $q - (u - v)$ equations, allowing for the possibility that the original system of q equations may not be in a reduced form.

The structure of (7) is exploited as follows. For a given set of true parameter values $\boldsymbol{\theta^*}$ giving rise to $\boldsymbol{\phi^*} = t(\boldsymbol{\theta^*})$ and a given prior $\pi(\boldsymbol{\theta}) = \pi_a(\boldsymbol{\theta_a})\pi_{b|a}(\boldsymbol{\theta_b}|\boldsymbol{\theta_a})$, the large-sample limit of the posterior distribution on $\boldsymbol{\theta}$ is identically the prior conditional $\pi(\boldsymbol{\theta}|\boldsymbol{\phi} = \boldsymbol{\phi^*})$. An algorithm to produce an arbitrarily large Monte Carlo sample from this distribution is comprised of the following three steps. We term this the *Large-Sample Limit via Monte Carlo* (LSLMC) algorithm.

Step 1: Generate an *iid* sample of size m from the density:

$$f(\boldsymbol{\theta_a}, \boldsymbol{\theta_b}) \;\propto\; \pi_a(\boldsymbol{\theta_a}) I\{[\boldsymbol{\theta_a}, g(\boldsymbol{\theta_a}; \boldsymbol{\phi^*})] \in \Theta\} \delta_{g(\boldsymbol{\theta_a}; \boldsymbol{\phi^*})}(\boldsymbol{\theta_b}),$$

where $\delta_x()$ represents a point-mass at x. Thus the $\boldsymbol{\theta_a}$ marginal under $f()$ is the prior marginal $\pi_a(\boldsymbol{\theta_a})$ truncated to ensure that $[\boldsymbol{\theta_a}, g(\boldsymbol{\theta_a}; \boldsymbol{\phi^*})] \in \Theta$, while the $(\boldsymbol{\theta_b}|\boldsymbol{\theta_a})$ conditional is a point-mass at $g(\boldsymbol{\theta_a}; \boldsymbol{\phi^*})$.

Step 2. Weight the sample drawn in Step 1 to make it representative of $\pi(\boldsymbol{\theta}|g(\boldsymbol{\theta_a}; \boldsymbol{\phi^*}) = \boldsymbol{\theta_b})$. The requisite weights are readily seen to take the form

$$w(\boldsymbol{\theta}) \;\propto\; \pi_{b|a}(\boldsymbol{\theta_b}|\boldsymbol{\theta_a}).$$

Step 3 (only necessary if $v > 0$). Further weight the sample to make it (approximately) representative of $\pi(\boldsymbol{\theta}|g(\boldsymbol{\theta_a}; \boldsymbol{\phi^*}) = \boldsymbol{\theta_b}, h(\boldsymbol{\theta_a}; \boldsymbol{\phi^*}) = \mathbf{0_v})$, which in light of (7) is the desired distribution. The further modified weights take the form

$$\tilde{w}(\boldsymbol{\theta}) \;\propto w(\boldsymbol{\theta}) k\left[b^{-1} \| h(\boldsymbol{\theta_a}; \boldsymbol{\phi^*}) \|\right], \tag{8}$$

for some small bandwidth $b > 0$, where $k()$ is say the standard normal density function. Thus this kernel weighting favours points with $h(\boldsymbol{\theta_a}; \boldsymbol{\phi^*})$ close to zero. As b decreases to zero then, the weighting becomes representative of the desired distribution.

In practice there is a tradeoff between making b small for the sake of a good approximation to the desired conditioning, versus keeping b large enough so that the weights (8) are not too variable to represent the desired distribution well. Overall care is required in choosing m and b, to ensure that the approximation is sufficiently accurate. One useful notion in this regard is that of the *effective sample size* with which a weighted sample represents a distribution (see, for instance, Doucet *et. al.* 2001). Presuming the realized weights in (8) are normalized, the Monte Carlo sample of size m and choice of bandwidth b give an effective sample size of $ess(m, b) = \{\sum_{i=1}^{m} \tilde{w}_i^2\}^{-1}$. That is, the weighted sample is as good as an *iid* sample of this size in representing the limiting posterior distribution. In Appendix A we describe a scheme relying on the effective sample size for automatically choosing m and b to be sufficiently large and small respectively.

Note also that the $v = 0$ case, whereby that $t(\boldsymbol{\theta}) = \boldsymbol{\phi} \iff \boldsymbol{\theta_b} = g(\boldsymbol{\theta_a}; \boldsymbol{\phi})$, leads to a particularly simple form of the algorithm. Here the third step of the algorithm is not needed, obviating the issue of bandwidth selection.

4 Examples of Limiting Posterior Distributions

4.1 Two imperfect trait assessments, continued

The model for two possibly dependent trait assessments applied to three populations of varying trait prevalence can be parameterized in meaningful scientific terms as follows. The nine unknown parameters are taken as $\boldsymbol{\lambda} = (r_1, r_2, r_3, SN_1, SN_2, \gamma_N, SP_1, SP_2, \gamma_P)$, where, as described earlier, r_i is the prevalence of trait X in the i-th population, while (SN_j, SP_j) are the sensitivity and specificity of the j-th trait assessment X_j^*. The parameters γ_P and γ_N describe dependence between the two assessments. In particular, we take $\gamma_P = \log \mathrm{OR}(X_1^*, X_2^* | X = 0)$ and $\gamma_N = \log \mathrm{OR}(X_1^*, X_2^* | X = 1)$. Several restrictions on the parameters are introduced. First, both assessments are assumed to be 'better than chance,' as represented by the constraint

$$SN_j + SP_j \;>\; 1, \tag{9}$$

for $j = 1, 2$. This assumption is commonly made in diagnostic testing models, to rule out a trivial 'label-switched' inference, since replacing (r_i, SN_j, SP_j) with $(1 - r_i, 1 - SN_j, 1 - SP_j)$ leads to the same observed prevalence of X_j^* in the i-th population. Second, γ_P and γ_N are assumed to be nonnegative, given the intent of allowing for possible positive dependence between the two tests.

As alluded to earlier, Hanson and Johnson (2005) prove that this model is not identified. In fact, Jones *et. al.* (2009) make more detailed statements by examining the Jacobian of the mapping from scientific parameters to cell probabilities for observables. Their scientific parameterization is as above, but with covariances rather than log odds-ratios describing the between-test dependence. They show that the 9×9 Jacobian has two zero eigenvalues and are able to give explicit forms for the two corresponding eigenvectors. Appealing to classical results on local identifiability (e.g., Rothenberg 1971, Goodman 1974), they note that none of the scientific parameters have corresponding zero entries in *both* eigenvectors. Thus, none of the scientific parameters are uniquely determined by the distribution of the observable data.

Toward illustrating the available information about the scientific parameters, we start with an initial prior distribution under which all nine parameters are independent, with $U(0, 1)$ marginals for all parameters, except for (γ_P, γ_N). Particularly then, we are not attempting to infuse any prior information about prevalences, sensitivities, and specificities. The dependence parameters γ_P and γ_N are assigned exponential priors with means μ_P and μ_N respectively. This induces some downweighting of stronger dependence, which seems appropriate in envisioned applications. In what follows we set the hyperparameters as $\mu_P = \mu_C = \log 2$, downweighting an odds-ratio of 2 by a factor of $e \approx 2.72$ compared to an odds-ratio of 1, in terms of density ratio on the log-OR scale. The final prior is obtained by truncating the initial prior to obey (9). This induces some mild dependence between SN_j and SP_j, and results in Beta$(2, 1)$ marginals for these parameters.

In order to apply the LSLMC algorithm of Section 3, it is necessary to move from the scientific parameterization to an algorithm-amenable parameterization denoted $\boldsymbol{\theta}$. In the present instance this parameterization is based on trait prevalences and cell probabilities for the $(X_1^*, X_2^*|X)$ distribution. Full details of the parameterization and the LSLMC implementation for this model are given in Appendix B.

To illustrate results in a detailed manner for one underlying set of parameter values, say that the true values are $\boldsymbol{r} = (0.1, 0.25, 0.5)$, $(SN_1, SP_1) = (0.83, 0.85)$, $(SN_2, SP_2) = (0.87, 0.75)$,

$\gamma_P = 0.486$, and $\gamma_N = 1.36$. Figure 1 displays the large-sample limiting posterior marginal distribution of each parameter (in the form of a histogram given that the limiting posterior is represented via a weighted Monte Carlo sample). The 95% central interval for the limiting posterior, the true parameter value, and the corresponding prior marginal density are all indicated as well. The extent of prior-to-posterior updating ranges from moderate to strong for these parameters, and clearly the lack of model identification does *not* equate with a complete lack of information in the data. Note that the extent of updating is particularly strong for the two specificities and the first of the three trait prevalences, and particularly weak for the two dependence parameters. Importantly, all the limiting credible intervals contain the true parameter values. Thus a large-sample does have considerable and reliable information content under this model, though the information does not become perfect as the sample size tends to infinity.

We also examine the bivariate behaviour of the limiting posterior distribution. There is substantial positive dependence amongst the three trait prevalences (Figure 2). Also, there is deterministic bivariate dependence amongst elements of (SN_1, SN_2, γ), and amongst elements of (SP_1, SP_2, γ_P), as evidenced in Figure 3. This is not surprising given that knowledge of the dependence parameters (γ_N, γ_P) should result in an identified model. That is, the limiting posterior distribution on $\boldsymbol{\theta}$ should reduce to a point mass when conditioned on specific values of the dependence parameters.

Of course studying inferential performance for a single set of true parameter values is not necessarily indicative of performance for other values. In fact it seems that extra caution is required in this regard. In other nonidentified model contexts the extent to which the posterior concentrates as the sample size goes to infinity has been seen to vary substantially across the parameter space (see, for instance, Gustafson 2005ab, 2007). Thus we must consider the limiting posterior marginal distribution arising from a given prior distribution for a large ensemble of true parameter values. It is convenient to simulate this ensemble of values from a chosen distribution, which we refer to as a *parameter generating distribution* (PGD). One possibility is to equate the PGD and the prior distribution. The common distribution would then both (i) dictate where in the parameter space inferential performance is to be assessed, and (ii) provides this information as an input to guide the analysis that is carried out for each (infinite-sized) dataset. On the other hand, as stressed by Wang and Gelfand (2002), one may be more comfortable in ascribing a relatively narrow swath of the parameter space as the region where performance is of interest than one is in providing the same swath as an input to guide the analysis of a given dataset.

In their (our) terminology, one may wish to specify a more concentrated *sampling prior* (PGD) than *fitting prior* (prior).

We proceed with a PGD based on simulating prevalences from the $\mathrm{U}(0.05, 0.95)$ distribution, sensitivities and specificities from the $\mathrm{U}(0.6, 0.95)$ distribution, and dependence parameters γ_N and γ_P from the $\mathrm{U}(0, \log 2)$ distribution. In contrast, the prior distribution specification is the same as given earlier. To reiterate the point made above, the PGD is chosen to study inferential performance under somewhat typical conditions. In contrast, to mimic practice and for reasons of conservatism, the prior specification is arguably 'wider' than typical-use conditions. Also, the PGD specification avoids extreme points in the parameter space, involving prevalences, sensitivities, or specificities which are near zero or one, which can be problematic for the LSLMC algorithm. Particularly, the Step 2 weights can be too variable to be relied upon in such cases. In fact, even with the chosen PGD we discard 20 of the 500 simulated parameter vectors, since they give Step 2 weightings with effective sample size less than 500 when the Step 1 sample is of size $m = 20000$.

Table 1 describes the aggregate inferential performance (with respect to the PGD) of the limiting posterior marginal distributions. For prevalences, sensitivities, and specificities we see quite small average absolute discrepancy between the true value and the limiting posterior mean. That is, point estimators of these parameters tend to have moderate asymptotic biases as a result of the nonidentification. Biases for the dependence parameters are much larger. Average widths of limiting 95% credible intervals speak to data being quite informative for sensitivities and specificities, but less so for prevalences. While the average widths for the dependence parameters are quite high, it is worth pointing out that they are about half the width of the corresponding prior intervals, so even these parameters are somewhat informed by data. Importantly, coverage (in the sense of proportion of parameter values generated by the PGD for which the limiting interval contains the true value) is near nominal. Thus the asymptotic bias of the Bayesian point estimator is acknowledged in the Bayesian interval estimator, obviating concern about reporting falsely precise results.

To be clear, Table 1 summarizes how well parameters could be inferred upon collecting an infinite amount of data, and thus represent 'best possible' bounds on what a finite dataset can achieve. Since the asymptotic bias of the point estimator is fundamental consequence of the lack of identification, the issue at hand is not whether a different procedure might have more attractive properties. Rather, the issue is whether the best-possible results are sufficiently good

	AAD	ALEN	COV
r_i	0.076	0.369	0.997
SN_j	0.061	0.165	0.935
SP_j	0.060	0.161	0.929
γ_N	0.415	1.213	0.935
γ_P	0.396	1.190	0.935

Table 1: Aggregate properties of the limiting posterior marginal distributions, with respect to the PGD described in the text. The columns give the average absolute discrepancy (AAD) between the limiting posterior mean and the true value, the average length (ALEN) of the limiting 95% central credible interval, and the coverage (COV) in terms of proportion of parameter values for which the limiting 95% central credible interval contains the true value.

to merit the allocation of resources to implementing a study under the specified conditions. While results such as those in Table 1 could inform such a decision, they cannot be used in isolation. The decision would necessarily involve subject-area considerations, such as the utility of a given magnitude of uncertainty about the target parameter and the per-unit cost of collecting data. In general, a formal scheme for linking study design (which includes 'no study' as one option) and inferential performance under nonidentified models remains to be developed. See Gustafson (2006), however, for some work in this direction.

In a related vein, it should be mentioned that some investigators might not view good average (across the parameter space) performance as sufficient justification to launch a study, if this average involves poor performance in a minority region under the PGD (offset by very good performance elsewhere). Thus Figure 4 reinforces Table 1 by showing the relationships between the true parameter values and their limiting posterior means, across the PGD. Figure 5 then sheds some light on regions in the parameter space where the limiting posterior distribution is a better or worse inferential quantity. When inferring trait prevalences, for instance, the discrepancy between the limiting posterior mean and the true value and the width of the limiting posterior credible interval both tend to be smaller when the the smallest gap between the three true prevalences is larger. That is, two populations having similar prevalences leads to less posterior information about these prevalences. This is in keeping with the findings of Gustafson (2005a) in the context of two conditionally independent tests. Figure 5 also shows that inferences about specificities tend to improve when the true prevalences are smaller (and symmetrically, inferences about sensitivities tend to improve when the true prevalences are larger). This makes sense, as lower prevalences correspond to more trait-absent subjects whose data can inform the specificities.

	male $<$ 90kg		female $<$ 90kg		male $>$ 90kg	
	$X_2^*=0$	$X_2^*=1$	$X_2^*=0$	$X_2^*=1$	$X_2^*=0$	$X_2^*=1$
$X_1^*=0$	92	49	49	13	40	26
$X_1^*=1$	10	8	1	5	5	9

Table 2: Data on imperfect assessment of haemodynamically obstructive disease, from Kosinski and Flanders (1999). The two surrogates X_1^* and X_2^* for true status X are as described in the text.

A referee raised the question of how much information is contributed by data from the population having the middle prevalence amongst the three, since the geometric arguments of Hanson and Johnson 2005 suggest that the inference will be driven by the populations with lowest and highest trait prevalence. Assume without loss of generality that $r_1 < r_2 < r_3$. In Appendix B we show that indeed the limiting posterior distribution on $(r_1, r_3, SN_1, SN_2, \gamma_N, SP_1, SP_2, \gamma_P)$ arising from observable data from all three populations coincides with the limiting posterior from observable data on the first and third populations only. This illuminates some structure in the problem, and might provide an alternate route to determining the limiting posterior distribution in the original problem with three populations.

To underscore the connection between real study data and large-sample limiting posteriors, consider the data in Table 2 from Kosinski and Flanders (1999). A study sample of patients with multi-vessel coronary artery disease are stratified according to weight (compared to a threshold of 90 kg) and gender. For purposes of connecting with the theory, the smallest stratum (8 female, $>$ 90kg patients) is ignored, leaving 307 subjects across three strata. The trait of interest X is whether the subject's coronary artery disease is haemodynamically obstructive or not. (In fact all subjects had haemodynamically obstructive disease at study entry, but the trait assessment is carried out after treatment, which renders some patients free of this condition.) For measurement of this trait, the first imperfect surrogate X_1^* arises from an exercise stress test. The second surrogate X_2^* arises from a single-photon-emission computed tomography (SPECT) thallium test. It is clear from the data that the two surrogates are discordant for a substantial proportion of study subjects.

The model is fit to these data using a Markov Chain Monte Carlo algorithm. Note that Bayesian inference problems involving unobserved variables are often tackled by applying MCMC to the joint posterior distribution of unobserved variables and parameters, given observed variables. However, for both the present model and other nonidentified models we have found this

strategy to be ineffective due to very poor MCMC mixing. Thus we instead work directly with the posterior distribution of parameters given observed variables. Specifically, we apply random walk Metropolis-Hastings updates to three blocks of parameters: $\boldsymbol{r}$, $(\boldsymbol{SN}, \gamma_N)$, and $(\boldsymbol{SP}, \gamma_P)$. For further discussion of MCMC applied to nonidentified models, see, Gelfand and Sahu (1999).

Posterior marginal densities for all parameters appear in Figure 6, indicating that these data are somewhat informative. By way of contrast, posterior marginals arising from dividing all observed data cell counts by four (i.e., pretending one has only a quarter of the data) are also displayed. These are actually not much less peaked than the full-data posteriors. This ramification of nonidentifiability stands in stark contrast with the situation for identified models with large-sample asymptotics having 'kicked in' (whereby a fourfold change in sample size corresponds roughly to a twofold change in posterior width).

We could similarly multiply the cell counts by a factor larger than one to investigate the utility of collecting further data beyond the $n = 307$ study subjects. In fact, we might contemplate determining the large-sample limit of the posterior distribution when the cell probabilities match the sample proportions in Table 2. This does not work, however, since sample proportions will not generally lie in the image of Θ under $t()$. Instead we use the full data to estimate $\boldsymbol{\theta}$ by its posterior mean vector $\hat{\boldsymbol{\theta}}$, and then determine the large-sample limit of the posterior as if this were the true value of $\boldsymbol{\theta}$. The limiting posterior marginals are also given in Figure 6. They suggest that further increases in sample size beyond the current level would at best have modest impact on inferring $\boldsymbol{r}$, but could have more impact in learning about the other parameters governing the trait assessment.

4.2 Instrumental variable analysis with binary response, continued

The binary instrumental variable model defined earlier by (4) through (6) is initially parameterized by $\boldsymbol{\lambda} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma)$. As detailed in Appendix C, application of the LSLMC algorithm requires reparameterization, from these scientific parameters to algorithm-amenable parameters. This is given by $\boldsymbol{\theta} = (\gamma, q_{00}, q_{01}, q_{10}, p_{00}, p_{01}, p_{10})$, where $\boldsymbol{q}$ and $\boldsymbol{p}$ comprise cell probabilities for $(X|W,U)$ and $(Y|X,U)$ respectively, i.e., $q_{wu} = Pr(X = 1|W = w, U = u)$, $p_{xu} = Pr(Y = 1|X = x, U = u)$. The components of $t(\boldsymbol{\theta})$ are six cell probabilities which completely characterize the distribution of $(Y, X|W)$, while the required partition of $\boldsymbol{\theta} = (\boldsymbol{\theta_a}, \boldsymbol{\theta_b})$ is based on $\boldsymbol{\theta_a} = (r, q_{00})$. Details of the functions g and h which are required for the LSLMC algorithm appear in Appendix C.

Before proceeding further with the LSLMC algorithm, we consider what light classical identifiability results shed on this model. The model as a whole must be nonidentified, since seven parameters give rise to six cell probabilities for observables. This alone, however, does not preclude a parameter of interest, such as β_1, being consistently estimable. Again motivated by Jones *et. al.* (2009), we consider the Jacobian of the mapping from parameters to cell probabilities. The Jacobian is conveniently calculated upon taking the parameters to be $(\beta_1, \gamma, p_{00}, p_{01}, q_{00}, q_{01}, q_{10})$. While we are not able to obtain closed-form expressions associated with a singular value decomposition of the Jacobian, numerical evaluation at trial values of the parameters indicates that the Jacobian is of full rank (i.e., rank six), implying that only one singular value is zero. Moreover, the corresponding right singular vector is seen to not have any zero entries, implying that none of these parameters are uniquely determined by the distribution of observables, even locally. In particular then, β_1, which describes the exposure-disease association, is not consistently estimable.

To illustrate the application of the LSLMC algorithm to this model, say the specified prior distribution involves $r \sim Beta(c, c)$, $\alpha_0, \beta_0 \sim N(0, \nu_1^2)$, $\alpha_1, \beta_1, \beta_2 \sim N(0, \nu_2^2)$. To reflect the 'label-switching' constraint, we take $\alpha_2 \sim N^+(0, \nu_2^2)$, i.e., a half-normal distribution. More specifically we set $c = 2.5$ on the grounds that extremely low or high prevalence for U is implausible. We set $2\nu_1 = -\text{logit } 0.02 = \text{logit } 0.98$ to probabilistically constrain away from the cases of zero/one prevalence for X and Y, and set $2\nu_2 = \log 6$ to reflect the notion that extremely large associations are uncommon in observational epidemiological studies.

To describe one specific execution of the algorithm, consider true parameter values $\gamma = 0.2$, $\boldsymbol{\alpha} = (\text{logit } 0.15, 1, 0.5)$, $\boldsymbol{\beta} = (\text{logit } 0.1, 0.5, 0.25)$. To gain some geometric insight into how the LSLMC algorithm works, the upper-left panel of Figure 7 depicts the set of $\boldsymbol{\theta_a}$ values for which $[\boldsymbol{\theta_a}, g(\boldsymbol{\theta_a}; \boldsymbol{\phi^*})] \in \Theta$. This indicates immediately that the posterior support of $\boldsymbol{\theta_a}$ is considerably smaller than the prior support, speaking to substantial prior-to-posterior updating which is not driven by the particular choice of prior distribution. This plot also indicates values of $\boldsymbol{\theta_a}$ satisfying $h(\boldsymbol{\theta_a}; \boldsymbol{\phi^*}) = 0$. Interestingly, we see instances of two solutions to this equation having the same value of $\theta_{a,1} = \gamma$. This implies that the initially plausible candidate $(\boldsymbol{\phi}, \gamma)$ would not in fact constitute a transparent reparameterization, and speaks to the difficulty of finding such a reparameterization or verifying that one exists.

The second panel in Figure 7 plots sampled values of the target parameter β_1 against $h(\boldsymbol{\theta_a}; \boldsymbol{\phi^*})$, with the suggestion that a deterministic relationship is present (in which case con-

ditioning on $h(\boldsymbol{\theta_a}; \boldsymbol{\phi^*}) = 0$ would produce a point-mass limiting posterior for β_1). However, plotting the relationship on a much finer scale in a neighbourhood of $h = 0$ (third panel) reveals that the relationship is in fact stochastic. One cannot directly intuit the effect of conditioning on $h = 0$ from the plot, since the plotted points have corresponding weights $w(\boldsymbol{\theta})$. These weights are seen to vary only modestly near $h = 0$ though (fourth panel), so that the $h = 0$ slice of the β_1 versus h plot should roughly correspond to the limiting posterior distribution.

More formally, the automated scheme described in Appendix A selects a Monte Carlo sample size of $m = 70000$, and a bandwidth of $b = 6.27 \times 10^{-6}$. The corresponding effective sample size associated with the $\tilde{w}()$ weighting in Step 3 of the algorithm is $ess(m, b) = 1333$. The resulting limiting marginal distribution for β_1 is depicted in the last panel of Figure 7. This distribution is sufficiently narrow (on the log odds-ratio scale) to be effectively a point mass for inferential purposes. Moreover, its location is consistent with the true value of $\beta_1 = 0.5$. Thus, at least for the particular set of underlying parameter values considered, the lack of formal identification has essentially no deleterious impact. As the sample size goes to infinity the posterior marginal distribution of β_1 converges to a near point-mass effectively at the true value.

As was emphasized in the previous example of Section 4.1, performance of posterior inferences in nonidentified models may vary considerably across the parameter space. Thus we consider other points in the parameter space as follows. We fix $\alpha_0 = \text{logit } 0.1$, $\beta_0 = \text{logit } 0.1$, and consider all $2^5 = 32$ combinations arising from $\gamma \in \{0.1, 0.4\}$, $\alpha_1 \in \{0.5, 1\}$, $\alpha_2 \in \{0.1, 0.3\}$, $\beta_1 \in \{0.2, 0.6\}$, $\beta_2 \in \{0.1, 0.3\}$. These values are chosen to be compatible with 'typical use' conditions, and can be regarded as forming a discrete PGD. Using the same prior distribution described above, the LSLMC algorithm is applied to determine the limiting posterior for every combination of parameter values.

The limiting posterior marginal distributions for β_1 arising under these parameter combinations are depicted in Figure 8, via means and 95% central intervals. Across combinations, the width of the limiting posterior marginal ranges from effectively zero in some cases (as was seen above), to as much as 0.04 on the log odds-ratio scale. Even this widest case, however, is extremely narrow in practical terms of knowledge about an exposure-disease relationship. Moreover, in all cases the 95% central interval covers the true value. Thus we conclude that this model shouldn't be discarded *because* it is nonidentified. The price to be paid in terms of lack of posterior concentration ranges from effectively none to extremely modest.

In fact, the results presented above correspond to one-eighth of a 2^8 factorial experiment.

The additional three factors are the signs of β_1, β_2, and α_1 (all set at '+' above). In order to avoid settings corresponding to extreme cell probabilities, in the general experiment the intercept terms are set as $\beta_0 = \text{logit } 0.1 - \min\{0, \beta_1\} - \ min\{0, \beta_2\}$ and $\alpha_0 = \text{logit } 0.1 - \min\{0, \alpha_1\} - min\{0, \alpha_2\}$, so that the smallest values of $Pr(Y = 1|X, U)$ and $Pr(X = 1|U, W)$ are both 0.1. Plots describing results for the other seven settings of the additional factors are available as supplemental material. The qualitative results are very consistent with those seen in Figure 8. That is, the limiting posterior distributions for β_1 are consistent with the true value, and the widths range from 'effectively' zero in many cases, to at most 0.05 on the log odds-ratio scale. Interestingly, in cases where the limiting posterior has appreciable width, the distribution is always favoring larger values of $|\beta_1|$. In none of the $2^8 = 256$ cases does the limiting posterior marginal give appreciable weight to values which are closer to the null than the true value.

5 Discussion

To return to the question posed in the title, why should we care about determining the large-sample limits of posterior distributions arising from nonidentified models? In short, we should care because realistic modelling of observational data will often lead to a nonidentified model, and consequently we need to understand the extent to which data contain information about parameters in such models.

In the example involving two imperfect measurements of a binary trait, the limiting posterior distributions for various underlying parameter values show that data are somewhat informative for parameters, but that very concentrated posteriors are unattainable at any sample size. In the instrumental variable model for binary variables, the limiting posterior distributions show that for all practical purposes this particular nonidentified model is more or less as good as an identified model. If this model seems appropriate on scientific grounds then, it would be wasteful to discard it because it is nonidentified. Without having evaluated the limiting posterior though, and knowing only that the model is nonidentified, one would likely have far less confidence in using the model for inference. The contrast between the two examples underscores that the important issue with identification is not whether it is lacking, but rather the extent to which a lack of identification impacts inference.

Of course simulation studies would be an alternative to evaluation of large-sample limits in determining the informativeness of data for a given nonidentified model. However, simulation

studies of Bayesian estimators can be computationally burdensome when MCMC is required to fit the model to every simulated dataset. Moreover, as alluded to in Section 4.1, the MCMC burden can be particularly great in the nonidentified model context. In fact our experience is that algorithmic performance of standard MCMC algorithms on posterior distributions arising from nonidentified models tends to worsen as the sample size increases. Thus simulation may be particularly problematic for indicating what happens as the sample size grows.

Admittedly, working with nonidentified models requires a shift in statistical mindset. Resulting point estimators (such as posterior means of target parameters) usually have biases which does not vanish asymptotically. Typical statistical thinking would then rule these to be 'bad' rather than 'good' estimators. Several points must be considered, however, before drawing such conclusions. First, in general there is no hope of constructing consistent estimators for nonidentified model settings. Second, since the posterior marginal distribution does not shrink to a single point, the estimator bias is reflected by the corresponding interval estimator. In fact, a detailed study of interval estimator performance in nonidentified models is given by Gustafson and Greenland (2009). Most important for present purposes, though, is the fact that we are *not* in the bad asymptotic regime of being more and more confident in a wrong answer as the sample size grows. Third, the lack of identification typically arises via real limitations on study design and data acquisition. One strategy would be to pretend these limitations are less than they really are, in order to obtain an identified model. This gets rid of an estimator bias due to nonidentification, but introduces an estimation bias due to model misspecification. This bias tradeoff is emphasized in Gustafson (2005a, 2007), and is generally seen to be unfavorable in the contexts considered there. Moreover, the identified but misspecified model paradigm does indeed involve becoming more and more confident about a wrong answer as the sample size grows. In all, it seems the utility of inferences driven by nonidentified models must be studied on a case-by-case basis. A blanket policy that nonidentified models are never usable, regardless of how realistic they may be, is not supported by the facts.

Appendix A: Choice of Monte Carlo Sample Size and Bandwidth

As mentioned in Section 3, one must take the Monte Carlo sample size m to be sufficiently large, and the bandwidth b in Step 3 to be sufficiently small, to ensure that the sampled points

under the $\tilde{w}()$ weighting comprise a good empirical representation of the limiting posterior distribution. The effective sample size $ess(m, b)$ associated with the weighted sample is a key quantity to monitor in this regard.

Our ad-hoc automated scheme for determining m and b proceeds as follows. Start with an initial Monte Carlo sample of size m_0 (we use $m_0 = 20000$), and choose the bandwidth b_0 as small as possible subject to $ess(m_0, b_0) \approx \tilde{m}_0$ (we use $\tilde{m}_0 = 500$). Then repeatedly:

- Increment the sample size, via $m_{j+1} \leftarrow m_j + \Delta_m$ (we use $\Delta_m = 5000$).
- If $ess(m_{j+1}, b_j) > ess(m_j, b_j)$ (as will usually be the case), then choose a new smaller bandwidth b_{j+1} such that $ess(m_{j+1}, b_{j+1}) = \{ess(m_j, b_j) + ess(m_{j+1}, b_j)\}/2$. Otherwise, set $b_{j+1} \leftarrow b_j$.

The rationale for this scheme for moving from (m_j, b_j) to (m_{j+1}, b_{j+1}) is that half of the *potential* improvement in ess when enlarging the sample is actually realized, while the other half is sacrificed toward the purpose of reducing the bandwidth. Hence at each iteration typically there is a meaningful improvement in ess and a meaningful reduction in bandwidth.

Of course such an iterative scheme requires a stopping criterion. We have found it effective to keep iterating until the total variation distance between the empirical marginal distribution of the target parameter at successive iterations falls below a threshold. In particular, we evaluate the empirical cdf of β_1 based on both the smaller (size m_j) and larger (size m_{j+1}) weighted samples, at the β_1 values realized in the smaller sample. We stop if the maximum absolute difference in cdfs across these points does not exceed 0.005.

Appendix B: Details for the two imperfect assessments model

The LSLMC algorithm in Section 3 is applicable to this model via reparameterizing from $(\boldsymbol{r}, \boldsymbol{SN}, \gamma_N, \boldsymbol{SP}, \gamma_P)$ to $\boldsymbol{\theta} = (\boldsymbol{r}, \boldsymbol{p}, \boldsymbol{q})$, where $\boldsymbol{p}$ and $\boldsymbol{q}$ denote cell probabilities which describe the joint distribution of (X_1^*, X_2^*) conditioned on $X = 0$ and $X = 1$ respectively. That is, $p_{ij} = Pr(X_1^* = i, X_2^* = j|X = 0)$ and $q_{ij} = Pr(X_1^* = i, X_2^* = j|X = 1)$. Thus both $\boldsymbol{p}$ and $\boldsymbol{q}$ must have nonnegative entries which sum to one. Using a 'dot' notation to indicate summation, we have $SP_1 = p_{0\cdot}$, $SP_2 = p_{\cdot 0}$, $SN_1 = q_{1\cdot}$, $SN_2 = q_{\cdot 1}$, while $\gamma_P = \log(p_{00}p_{11}) - \log(p_{01}p_{10})$ and

$\gamma_N = \log(q_{00}q_{11}) - \log(q_{01}q_{10})$. We write $(\boldsymbol{p}, \boldsymbol{q}) \in R$ to denote the constraints introduced in the original parameterization, as outlined in Section 4.1. Thus the parameter space Θ is $(0,1)^3 \times R$.

Now let $\boldsymbol{\phi} = t(\boldsymbol{\theta})$ denote the cell probabilities underlying the observed data tables, with $\phi_{ij}^{(a)} = Pr_a(X_1^* = i, X_2^* = j)$ in the a-th population. Clearly

$$\boldsymbol{\phi^{(a)}} \;=\; (1 - r_a)\boldsymbol{p} + r_a\boldsymbol{q}, \tag{10}$$

for $a = 1, 2, 3$. Now, we can partition $\boldsymbol{\theta}$ as $\boldsymbol{\theta_a} = (r_1, r_2)$ and $\boldsymbol{\theta_b} = (r_3, \boldsymbol{p}, \boldsymbol{q})$. Given $\boldsymbol{\phi}$ and $\boldsymbol{\theta_a}$, from (10) we see that the mapping $\boldsymbol{\theta_b} = g(\boldsymbol{\theta_a}; \boldsymbol{\phi})$ is given by

$$\begin{aligned}
\boldsymbol{p} &= \frac{r_1\boldsymbol{\phi^{(2)}} - r_2\boldsymbol{\phi^{(1)}}}{r_1 - r_2} \\
\boldsymbol{q} &= \frac{(1 - r_1)\boldsymbol{\phi^{(2)}} - (1 - r_2)\boldsymbol{\phi^{(1)}}}{r_2 - r_1} \\
r_3 &= \frac{\phi_{00}^{(3)} - p_{00}}{q_{00} - p_{00}}.
\end{aligned}$$

Note that this mapping can lead to values of $(\boldsymbol{p}, \boldsymbol{q})$ falling outside of R, or a value of r_3 outside of $(0, 1)$. The Step 1 sampling in the LSLMC algorithm will reject such values of $\boldsymbol{\theta_a}$. Note also that $t(\boldsymbol{\theta}) = \boldsymbol{\phi} \iff g(\boldsymbol{\theta_a}; \boldsymbol{\phi}) = \boldsymbol{\theta_b}$, so we are using the $u = 2$, $v = 0$ version of the algorithm, for which Step 3 is not needed.

As further implementation details, the prior specified in the original parameterization induces independence between $\boldsymbol{\theta_a}$ and $\boldsymbol{\theta_b}$, which simplifies the determination of weights $w()$ in Step 2 of the algorithm. In particular, the prior marginal density of $\boldsymbol{\theta_b}$ is determined to be

$$\pi_b(r_3, \boldsymbol{p}, \boldsymbol{q}) \;\propto\; \exp\left\{-\mu_P^{-1}\gamma_P(\boldsymbol{p}) - \mu_N^{-1}\gamma_N(\boldsymbol{q})\right\} \left(\sum_{i,j} p_{ij}^{-1}\right) \left(\sum_{i,j} q_{ij}^{-1}\right) I_R(\boldsymbol{p}, \boldsymbol{q}) I_{(0,1)}(r_3).$$

Thus the $w()$ weights are readily computed.

As a further point concerning this model, we take up the issue of the information provided by the population having intermediate prevalence amongst the three populations. Consider the parameterization $(r_1, w, r_3, \boldsymbol{\phi^{(1)}}, \boldsymbol{\phi^{(3)}})$, where $w = (r_2 - r_1)/(r_3 - r_1)$. Note here that $\boldsymbol{\phi^{(2)}} = (1 - w)\boldsymbol{\phi^{(1)}} + w\boldsymbol{\phi^{(3)}}$, hence observing infinite samples from all three populations equates with conditioning on true values of $(w, \boldsymbol{\phi^{(1)}}, \boldsymbol{\phi^{(3)}})$. In the case that the joint prior density takes the form $\pi(\boldsymbol{p}, \boldsymbol{q}) I_{(0,1)}(r_1) I_{(0,1)}(r_2) I_{(0,1)}(r_3)$, via change of variables the reparameterization is seen to induce the prior

$$\begin{aligned}
\pi(\boldsymbol{\phi^{(1)}}, \boldsymbol{\phi^{(3)}}, r_1, r_3, w) \;=\; & \pi\left\{\boldsymbol{p}(\boldsymbol{\phi^{(1)}}, \boldsymbol{\phi^{(3)}}, r_1, r_3), \boldsymbol{q}(\boldsymbol{\phi^{(1)}}, \boldsymbol{\phi^{(3)}}, r_1, r_3)\right\} \times \\
& I_{(0,1)}(r_1) I_{(0,1)}\{(1 - w)r_1 + wr_3\} I_{(0,1)}(r_3) |r_3 - r_1|.
\end{aligned}$$

Clearly over the range $w \in (0,1)$ this expression is constant in w. Thus we deduce that *a priori* (r_1, r_3) and w are conditionally independent given that $w \in (0,1)$ and given the values of $(\boldsymbol{\phi^{(1)}}, \boldsymbol{\phi^{(3)}})$. Consequently, we deduce the same limiting distribution for (r_1, r_3) whether we condition on just the cell probabilities for populations 1 and 3, or the cell probabilities for all three populations (via the additional observation of w). Moreover, since $\boldsymbol{p}$ and $\boldsymbol{q}$ are both functions of $(\boldsymbol{\phi^{(1)}}, \boldsymbol{\phi^{(3)}}, r_1, r_3)$, it follows that the limiting distributions on $(r_1, r3, SN_1, SN_2, \gamma_N, SP_1, SP_2, \gamma_P)$ coincide (again, with the presumption that the prevalence for the second population is intermediate between the first and third).

Appendix C: Details for the IV model

Let $\boldsymbol{p} = (p_{00}, p_{01}, p_{10})$ parameterize $(Y|X,U)$ with $p_{xu} = Pr(Y=1|X=x, U=u)$. Note that (4) implies $p_{11} = s(\boldsymbol{p})$, where $s(a,b,c) = \text{expit}(\text{logit } b + \text{logit } c - \text{logit } a)$. Note also that

$$\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} \text{logit } p_{00} \\ \text{logit } p_{10} - \text{logit } p_{00} \\ \text{logit } p_{01} - \text{logit } p_{00} \end{pmatrix}. \tag{11}$$

Similarly, let $\boldsymbol{q} = (q_{00}, q_{01}, q_{10})$, with $Pr(X=1|W=w, U=u) = q_{wu}$, and $q_{11} = s(q)$ from (5). The relationship between $\boldsymbol{q}$ and $\boldsymbol{\alpha}$ follows the same pattern as (11).

Our expedient parameterization is now $\boldsymbol{\theta} = (\boldsymbol{p}, \boldsymbol{q}, \gamma)$. The distribution of $(Y|X,W)$ is characterized by $\boldsymbol{\phi} = t(\boldsymbol{\theta})$, where $\phi_{yxw} = Pr(Y=y, X=x|W=w)$ is given by

$$\begin{pmatrix} \phi_{000} \\ \phi_{010} \\ \phi_{100} \\ \phi_{001} \\ \phi_{011} \\ \phi_{101} \end{pmatrix} = (1-\gamma) \begin{pmatrix} (1-p_{00})(1-q_{00}) \\ (1-p_{10})q_{00} \\ p_{00}(1-q_{00}) \\ (1-p_{00})(1-q_{10}) \\ (1-p_{10})q_{10} \\ p_{00}(1-q_{10}) \end{pmatrix} + \gamma \begin{pmatrix} (1-p_{01})(1-q_{01}) \\ \{1-s(\boldsymbol{p})\}q_{01} \\ p_{01}(1-q_{01}) \\ (1-p_{01})\{1-s(\boldsymbol{q})\} \\ \{1-s(\boldsymbol{p})\}s(\boldsymbol{q}) \\ p_{01}\{1-s(\boldsymbol{q})\} \end{pmatrix}. \tag{12}$$

Using the partition $\boldsymbol{\theta} = (\boldsymbol{\theta_a}, \boldsymbol{\theta_b})$ with $\boldsymbol{\theta_a} = (\gamma, q_{00})$, we want to construct families of functions functions $g(\cdot;\boldsymbol{\phi})$ and $h(\cdot;\boldsymbol{\phi})$ satisfying (7). Say that given values of $\boldsymbol{\theta_a} = (r, q_{00})$ are consistent with $\boldsymbol{\phi}$. It then follows from (12) that

$$q_{01} = 1 - \gamma^{-1}\{\phi_{000} + \phi_{100} - (1-\gamma)q_{00}\}. \tag{13}$$

Subsequently q_{10} is determined as the solution to $u(z) = 0$, where

$$u(z) = (1-\gamma)(1-z) + \gamma\{1 - s(q_{00}, q_{01}, z))\} - (\phi_{001} + \phi_{101}). \tag{14}$$

Note that $u()$ decreases monotonically, from $1-\phi_{001}-\phi_{101}$ when $z=0$ to $-(\phi_{001}+\phi_{101})$ when $z=1$. Now armed with values of both γ and $\boldsymbol{q}$, we can determine

$$p_{00} \;=\; 1-\frac{\{1-s(\boldsymbol{q})\}\phi_{000}-(1-q_{01})\phi_{001}}{(1-\gamma)[(1-q_{00})\{1-s(\boldsymbol{q})\}-(1-q_{01})(1-q_{10})]} \tag{15}$$

and similarly

$$p_{01} \;=\; 1-\frac{(1-q_{00})\phi_{001}-(1-q_{10})\phi_{000}}{\gamma[(1-q_{00})\{1-s(\boldsymbol{q})\}-(1-q_{01})(1-q_{10})]}. \tag{16}$$

Finally, from this point we obtain p_{10} as the solution to $v(z)=0$, where

$$v(z) \;=\; (1-\gamma)q_{00}(1-z)+\gamma q_{01}\{1-s(p_{00},p_{01},z)\}-\phi_{010}, \tag{17}$$

which is monotonically decreasing. Taken together then, (13) through (17) determine the function $g(\boldsymbol{\theta_a};\boldsymbol{\phi})$ in (7). In practical terms, any out-of-range solution in (13) through (17) corresponds to a situation where $[\boldsymbol{\theta_a}, g(\boldsymbol{\theta_a};\boldsymbol{\phi})] \notin \Theta$, and such values are excluded in Step 1 of the LSLMC algorithm.

Note that (13) through (17) guarantee that all but one of the six equations in (12) must hold, with the fifth equation for ϕ_{011} being the exception. Thus $g(\boldsymbol{\theta_a};\boldsymbol{\phi})=\boldsymbol{\theta_b}$ is necessary, but not sufficient, for $t(\boldsymbol{\theta})=\boldsymbol{\phi}$. The decomposition (7) is completed by re-expressing the fifth equation in the form $h(\boldsymbol{\theta_a};\boldsymbol{\phi})=0$. That is, we take $h(\boldsymbol{\theta_a};\boldsymbol{\phi})=\tilde{h}((\boldsymbol{\theta_a}, g(\boldsymbol{\theta_a};\boldsymbol{\phi}));\boldsymbol{\phi})$, where

$$\tilde{h}(\boldsymbol{\theta};\boldsymbol{\phi}) \;=\; (1-\gamma)(1-p_{10})q_{10}+\gamma\{1-s(\boldsymbol{p})\}s(\boldsymbol{q})-\phi_{011}.$$

Armed with $g()$ and $h()$, the LSLMC algorithm is immediately applicable, with $u=1$, $v=1$.

As remaining implementation details, the required Monte Carlo sampling in Step 1 of the algorithm is based on sampling from the prior marginal on $\boldsymbol{\theta_a}=(\gamma,q_{00})$ and simply rejecting draws for which $[\boldsymbol{\theta_a}, g(\boldsymbol{\theta_a};\boldsymbol{\phi^*})] \notin \Theta$. The conditional prior density required to form weights in Step 2 of the algorithm is readily determined upon transforming the joint prior density in the original scientific parameterization (under which all components are independent) to the joint prior density in the $\boldsymbol{\theta}$ parameterization. Note in particular that the requisite Jacobian term for the mapping from $\boldsymbol{\theta}$ to the original parameter vector is $q_{00}^{-1}(1-q_{00})^{-1}q_{01}^{-1}(1-q_{01})^{-1}q_{10}^{-1}(1-q_{10})^{-1}p_{00}^{-1}(1-p_{00})^{-1}p_{01}^{-1}(1-p_{01})^{-1}p_{10}^{-1}(1-p_{10})^{-1}$. Finally, the choice of bandwidth b for the $\tilde{w}()$ weighing in Step 3 of the algorithm is chosen in an automated manner, as described in Appendix A.

Supplemental Materials

Additional plots: Additional plots of results from Section 4.2 (pdf file)

References

Albert, P.S. and Dodd, L.E. (2004). A Cautionary Note on the Robustness of Latent Class Models for Estimating Diagnostic Error without a gold standard. *Biometrics* **60**, 427–435.

Barankin, E. (1961). Sufficient parameters: solution to the minimal dimensionality problem. *Annals of the Institute of Statistical Mathematics* **12**, 91-118.

Doucet, A., de Freitas, N., and Gordon, Neil (2001). *Sequential Monte Carlo Methods in Practice*. Springer-Verlag: New York.

Dendukuri, N. and Joseph, L. (2001). Bayesian Approaches to Modeling the Conditional Dependence Between Multiple Diagnostic Tests. *Biometrics* **57**, 158–167.

Florens, J.P., Mouchart, M., and Rolin, J.M. (1990). *Elements of Bayesian Statistics*. Marcel Dekker: New York.

Gelfand, A.E. and Sahu, S.K. (1999). Identifiability, improper priors, and Gibbs sampling for generalized linear models. *Journal of the American Statistical Association* **94**, 247–253.

Georgiadis, M.P., Johnson, W.O., Gardner, I.A., Singh, R. (2003). Correlation-adjusted estimation of sensitivity and specificity of two diagnostic tests. *Journal of the Royal Statistical Society C* **52**, 63–76.

Goetghebeur, E., Liinev, J., Boelaert, M., Van der Stuyft, P. (2000). Diagnostic test analyses in search of their gold standard: latent class analyses with random effects. *Statistical Methods in Medical Research* **9**, 231–248.

Goodman, L.A. (1974). Exploratory latent structure using both identifiable and unidentifiable models. *Biometrika* **61**, 215-231.

Greenland, S. (2003). The impact of prior distributions for uncontrolled confounding and response bias: a case study of the relation of wire codes and magnetic fields to childhood leukemia. *Journal of the American Statistical Association* **98**, 47–54.

Greenland, S. (2005). Multiple bias modeling for analysis of epidemiologic data (with dis-

cussion). *Journal of the Royal Statistical Society A* **168**, 267–306.

Gustafson, P. (2005a). On model expansion, model contraction, identifiability, and prior information: two illustrative scenarios involving mismeasured variables (with discussion). *Statistical Science* **20**, 111–140.

Gustafson, P. (2005b). The utility of prior information and stratification for parameter estimation with two screening tests but no gold standard. *Statistics in Medicine* **24**, 1203–1217.

Gustafson, P. (2006). Sample size implications when biases are modeled rather than ignored. *Journal of the Royal Statistical Society A* **165**, 865–881.

Gustafson, P. (2007). Measurement error modelling with an approximate instrumental variable. *Journal of the Royal Statistical Society B* **69**, 797–815.

Gustafson, P. and Greenland, S. (2006). The performance of random coefficient regression in accounting for residual confounding. *Biometrics* **62**, 760–768.

Gustafson, P. and Greenland, S. (2009). Interval estimation for messy observational data. Unpublished report (`ftp.stat.ubc.ca/pub/gustaf/GuGr_N.pdf`).

Gustafson, P., Le, N.D., and Saskin, R. (2001). Case-control analysis with partial knowledge of exposure misclassification probabilities. *Biometrics* **57**, 598–609.

Hanson, T.E. and Johnson, W.O. (2005). Discussion of 'On model expansion, model contraction, identifiability, and prior information: two illustrative scenarios involving mismeasured variables.' *Statistical Science* **20**, 131–134.

Hanson T.E., Johnson W.O., Gardner I.A., Georgiadis, M.P. (2003). Determining the infection status of a herd. *Journal of Journal of Agricultural, Biological & Environmental Statistics* **8**, 469–485.

Hui, S.L. and Zhou, X.H. (1998). Evaluation of diagnostic tests without gold standards. *Statistical Methods in Medical Research* **7**, 354–370.

Hui, S.L. and Walter, S.D. (1980). Estimating the error rates of diagnostic tests. *Biometrics* **36**, 167–71.

Johnston, K., Gustafson, P., Levy, A.R., and Grootendorst, P. (2008). Use of instrumental variables in the analysis of generalized linear models in the presence of unmeasured confounding with applications to epidemiological research. *Statistics in Medicine* **27**, 1539-1556.

Jones, G., Johnson, W.O., Hanson, T.E., and Christensen, R. (2009). Identifiability of

models for multiple diagnostic testing in the absence of a gold standard. Unpublished report (`www-ist.massey.ac.nz/GJones/IdDiagTest_91.pdf`).

Kosinski, A. S. and Flanders, W. D. (1999). Evaluating the exposure and disease relationship with adjustment for different types of exposure misclassification: A regression approach. *Statistics in Medicine* **18**, 2795-2808.

Menten, J., Boelaert, M., Lesaffre, E. (2008). Bayesian latent class models with conditionally dependent diagnostic tests: A case study. *Statistics in Medicine* **27**, 4469–4488.

Neath, A.E., and Samaniego, F.J. (1997). On the efficacy of Bayesian inference for nonidentifiable models. *American Statistician* **51**, 225–232.

Pearl, J. (2000). *Causality: models, reasoning, and inference.* Cambridge University Press: New York.

Poirier, D.J. (1998). Revising beliefs in nonidentified models. *Econometric Theory* **14**, 483–509.

Qu, Y., Tan, M., Kutner, M.H. (1996). Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. *Biometrics* **52**, 797-810.

Rothenberg, T.J. (1971). Identification in parametric models. *Econometrica* **39**, 577-591.

Scharfstein, D.O., Daniels, M.J., Robins, J.M. (2003). Incorporating prior beliefs about selection bias into the analysis of randomized trials with missing outcomes. *Biostatistics* **4**, 495–512.

Terza, J.V. (2006). Estimation of policy effects using parametric nonlinear models: a contextual critique of the generalized method of moments. *Health Services Outcomes and Research Methodology* **6**, 177-198.

Torrance-Rynard, V.L. and Walter, S.D. (1998). Effects of dependent errors in the assessment of diagnostic test performance. *Statistics in Medicine* **16**, 2157–2175.

Vacek, P.M. (1985). The effect of conditional dependence on the evaluation of diagnostic tests. *Biometrics* **41**, 959–968.

Wang, F. and Gelfand, A.E. (2002). A simulation-based approach to Bayesian sample size determination for performance under a given model and for separating models. *Statistical Science* **17**, 193-208.

Xie, Y. and Carlin, B.P. (2006). Measures of Bayesian learning and identifiability in hierar-

chical models. *Journal of Statistical Planning and Inference* **136**, 3458–3477.

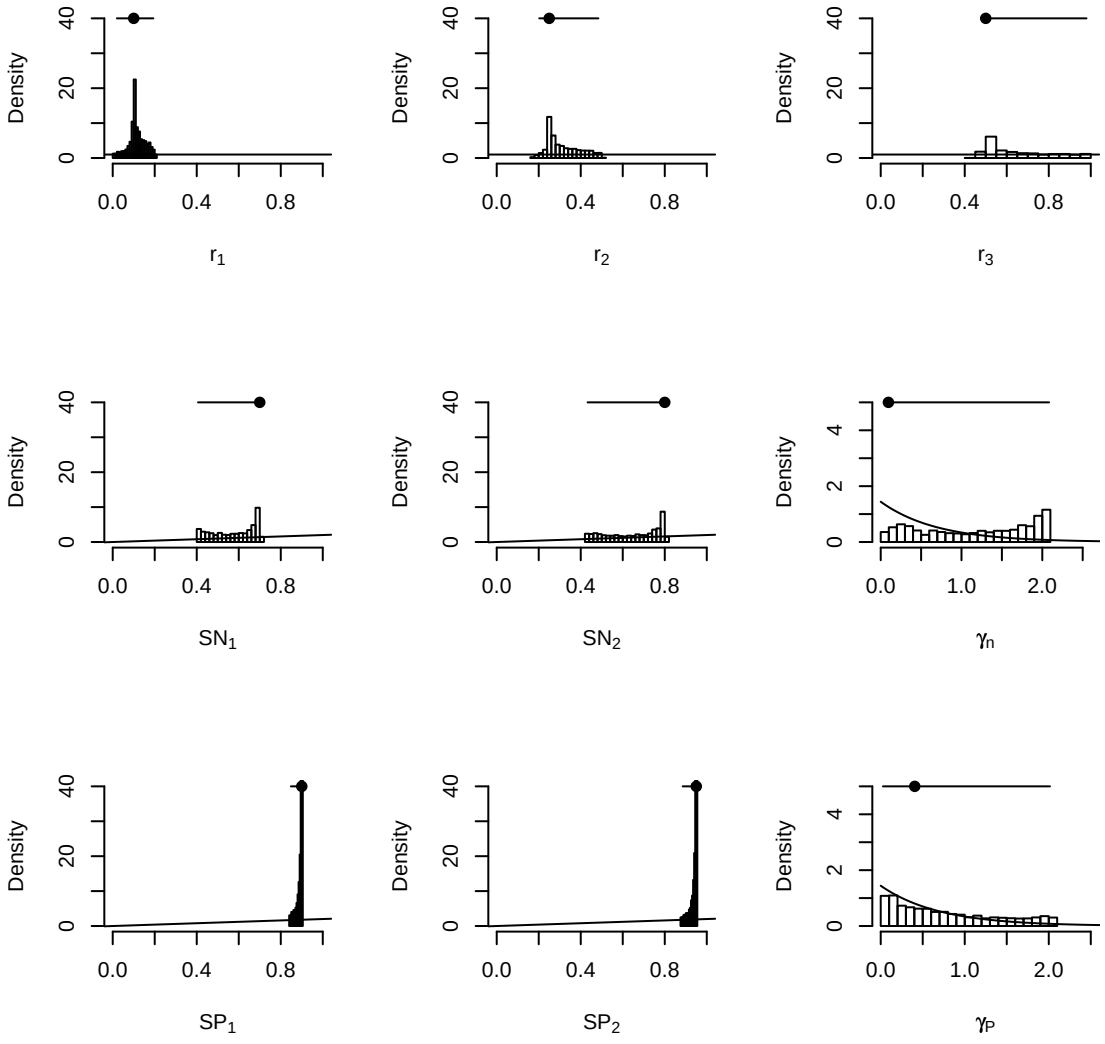

Figure 1: Limiting posterior marginal distributions, for a particular set of true parameter values. The true parameter value and the 95% central interval of the limiting posterior distribution are indicated at the top of each panel.

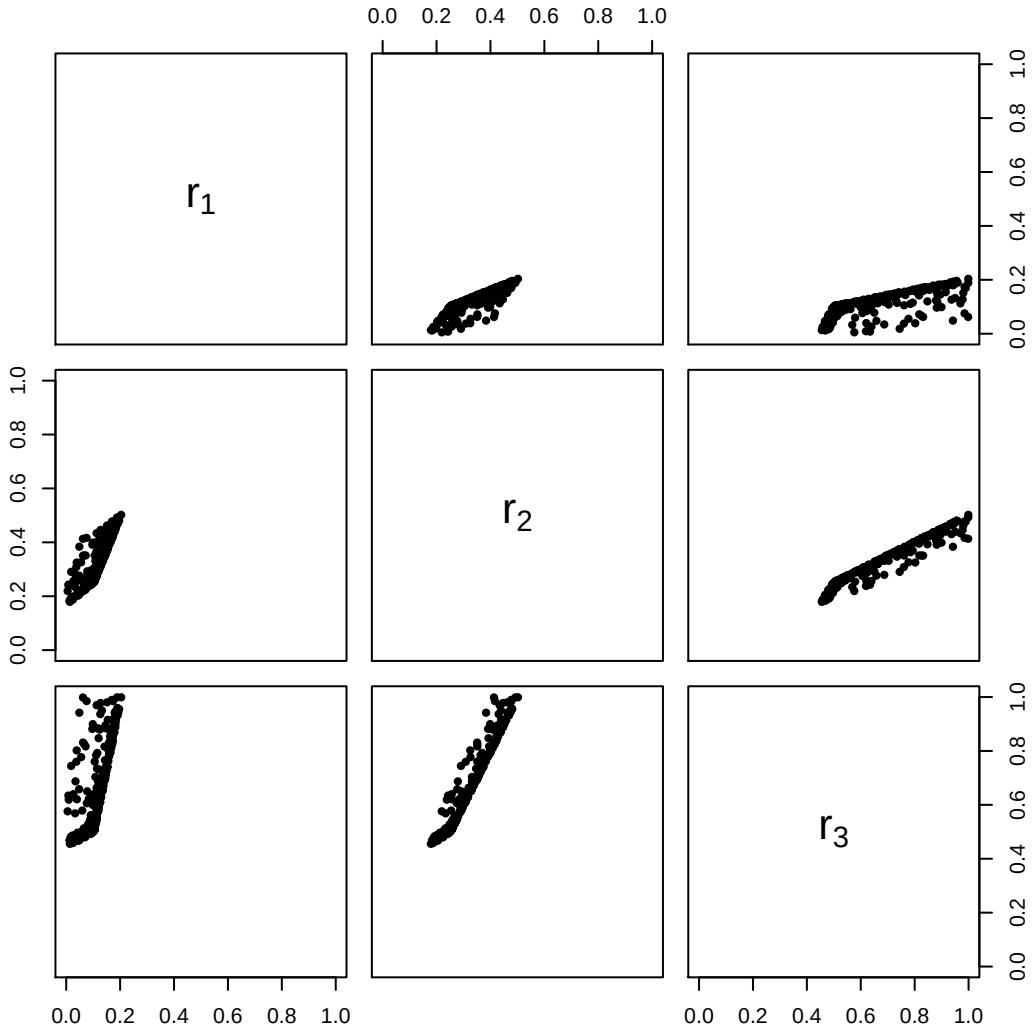

Figure 2: Limiting bivariate posterior distributions for trait prevalences.

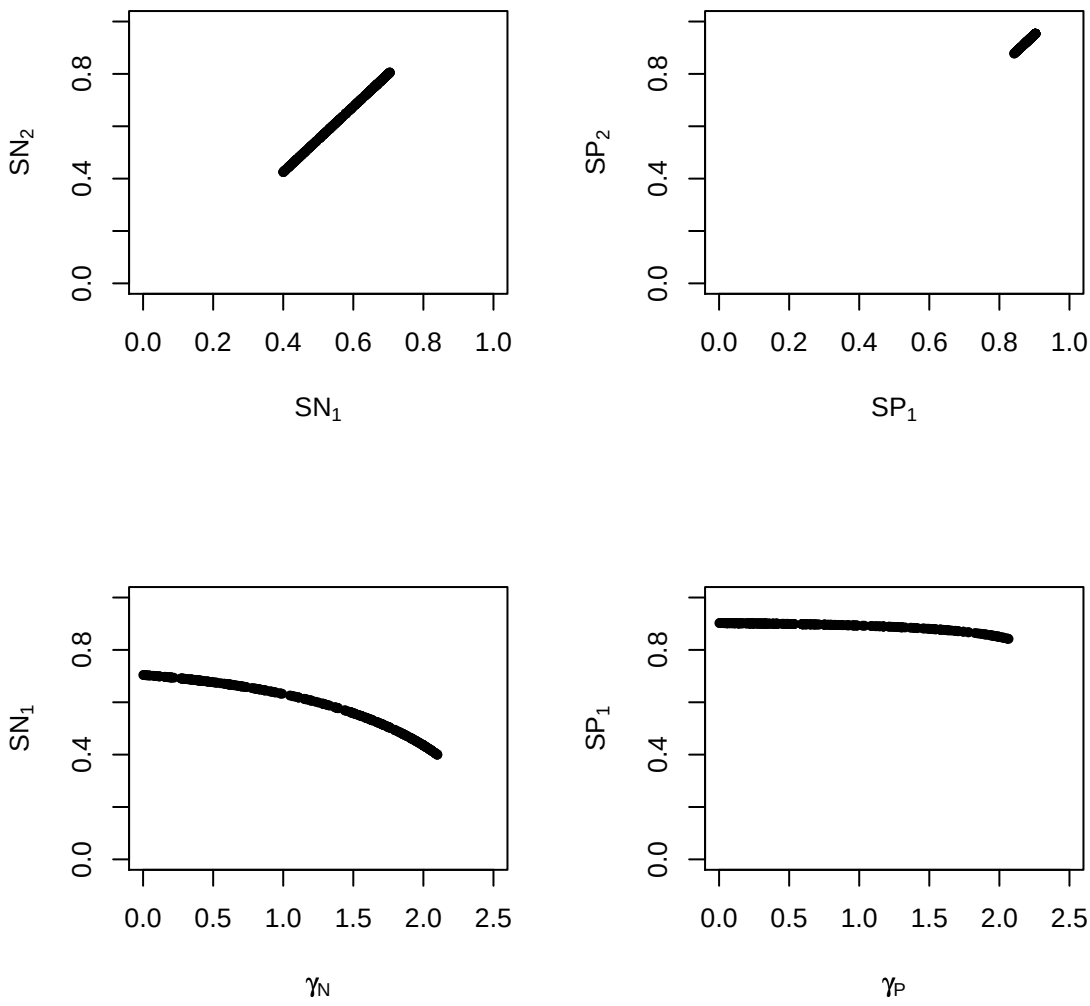

Figure 3: Limiting bivariate posterior distributions for trait assessment parameters.

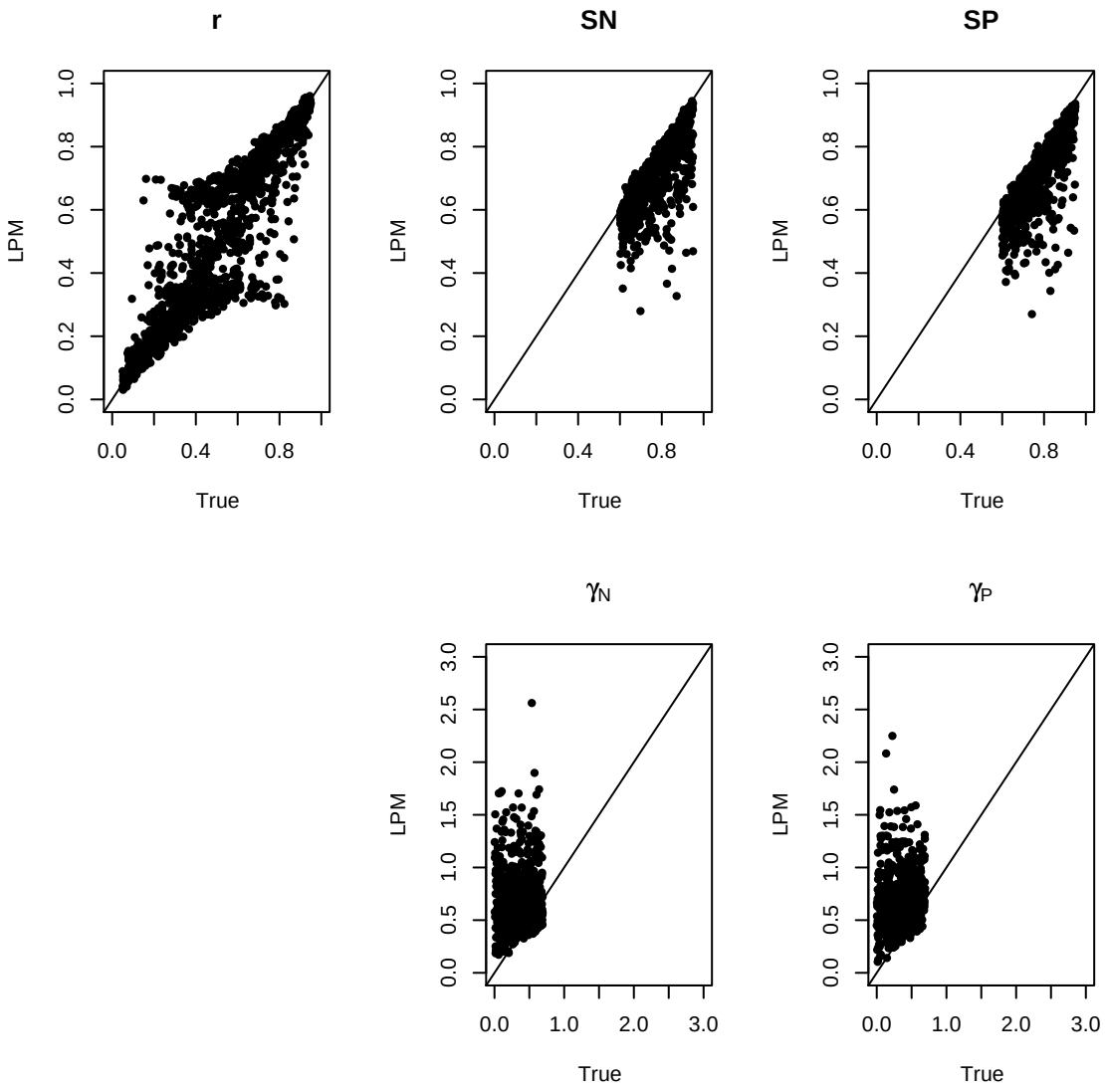

Figure 4: Limiting posterior mean (LPM) versus true value, for true values generated from the PGD described in the text.

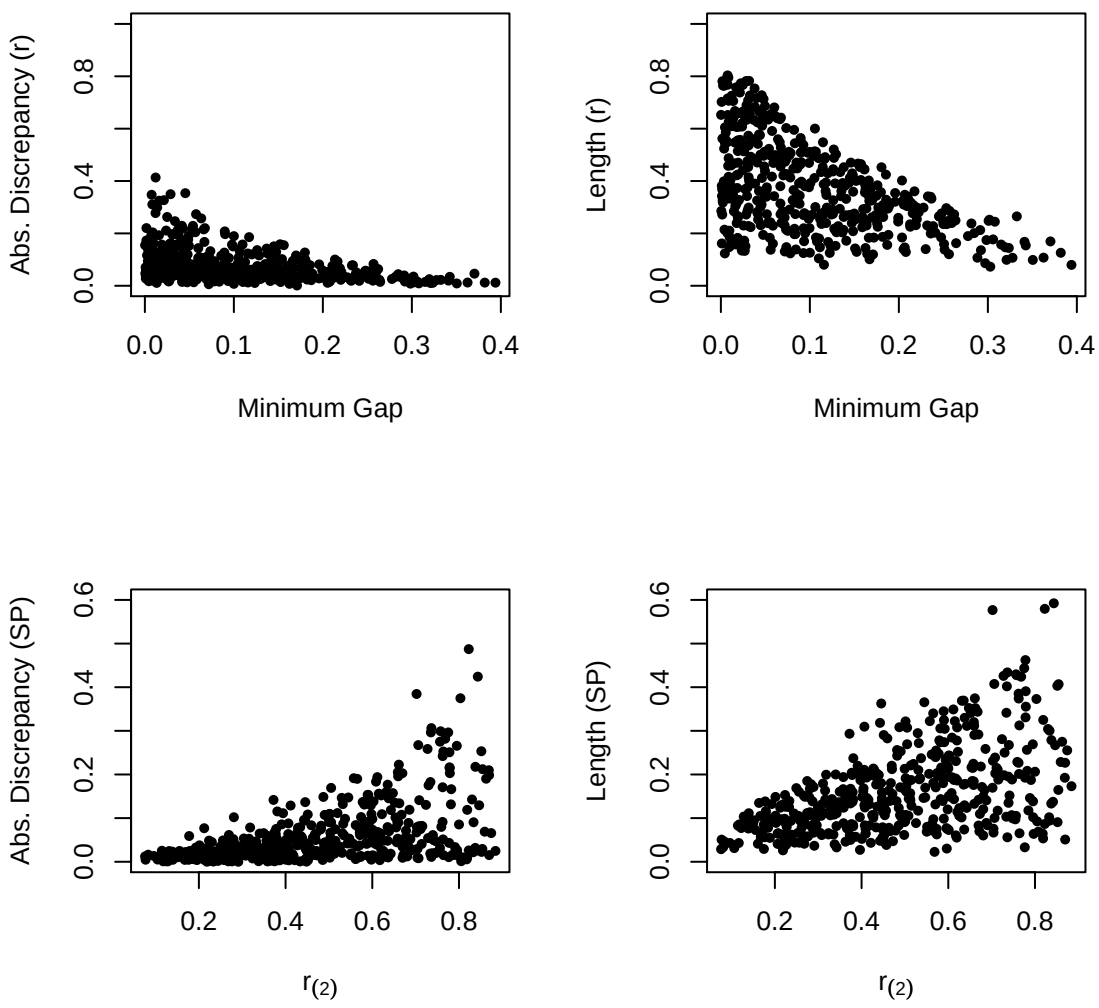

Figure 5: Variation in limiting posterior marginals with underlying prevalence values. For inference about the prevalences $\boldsymbol{r} = (r_1, r_2, r_3)$, the top panels plot absolute discrepancy between limiting posterior mean and true value, and length of limiting 95% credible interval, both as a function of the smallest absolute difference between elements of $\boldsymbol{r}$. For inferences about specificities, the bottom panels plot absolute discrepancy between limiting posterior mean and true value, and length of limiting 95% credible interval, both as a function of the median of $\boldsymbol{r}$, denoted $r_{(2)}$.

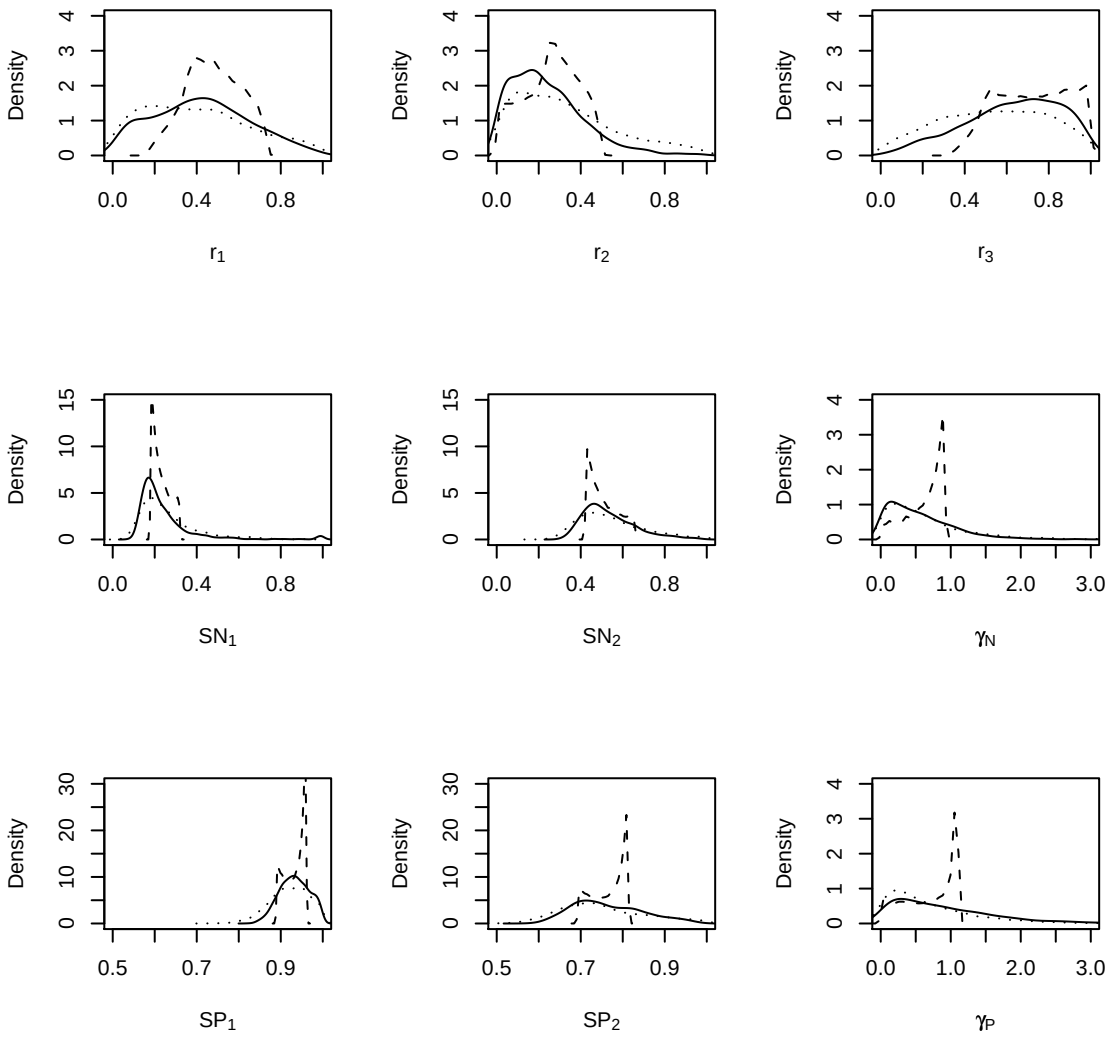

Figure 6: Posterior marginal distributions in the haemodynamic coronary artery disease example. The solid curves are posterior densities based on all the data. The dotted curves are posterior densities based on one-quarter of the data (i.e., all cell counts are divided by four). The dashed curves are large-sample limits of posterior densities, when the parameter values equal estimated values from the full-data fit.

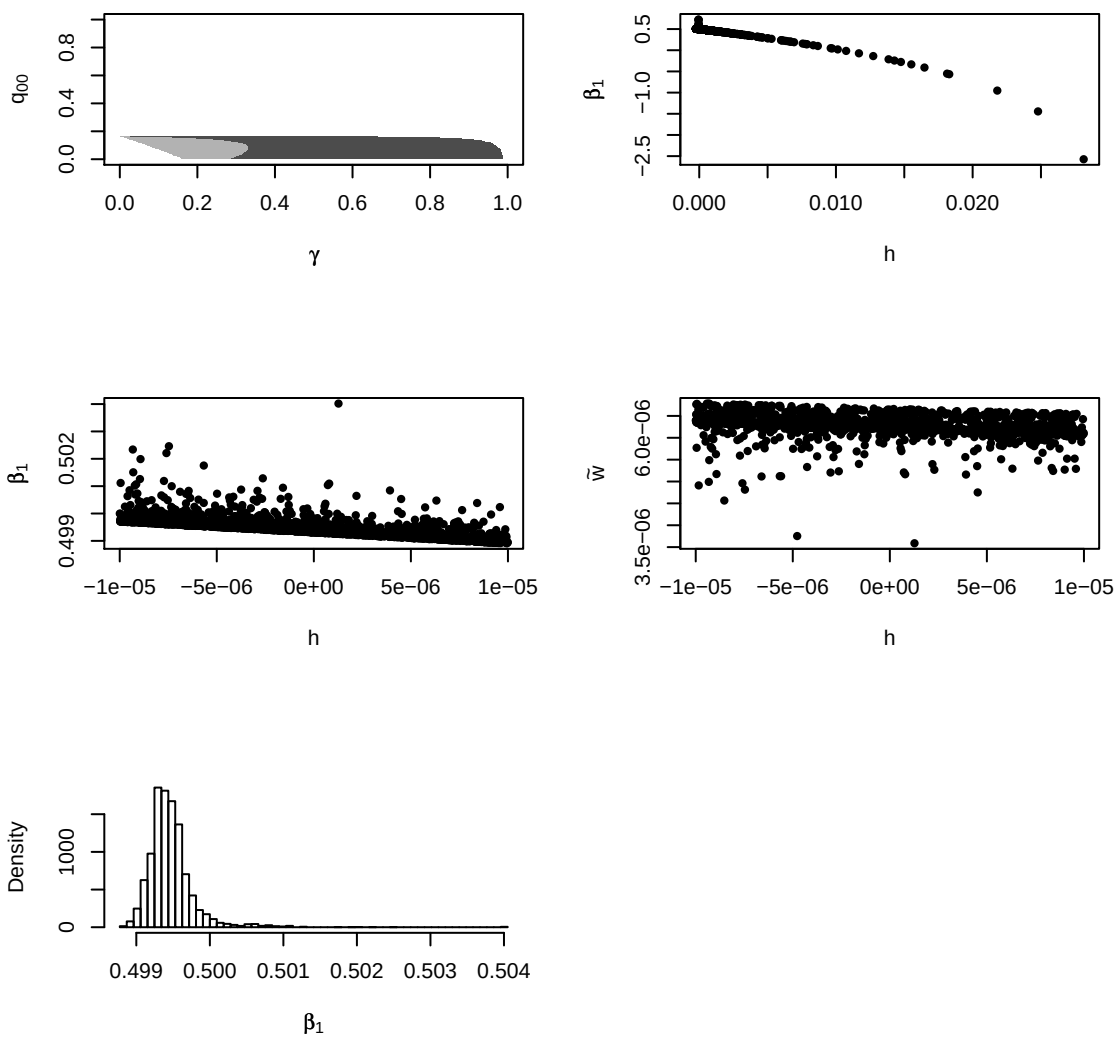

Figure 7: Aspects of the LSLMC algorithm implementation in the binary instrumental variable model. The upper-left panel depicts in grey values of $\boldsymbol{\theta_a} = (\gamma, q_{00})$ for which $[\boldsymbol{\theta_a}, g(\boldsymbol{\theta_a}; \boldsymbol{\phi^*})] \in \Theta$. The light/dark shading corresponding to negative/positive values of $h(\boldsymbol{\theta_a}; \boldsymbol{\phi^*})$. The next two panels give plots of β_1 versus h for the Monte Carlo sample, both for the whole range of h and in a neighbourhood of $h = 0$. For the same neighbourhood, the fourth panel plots $\tilde{w}(\boldsymbol{\theta})$ versus h. The last panel gives the large-sample limiting posterior marginal for β_1.

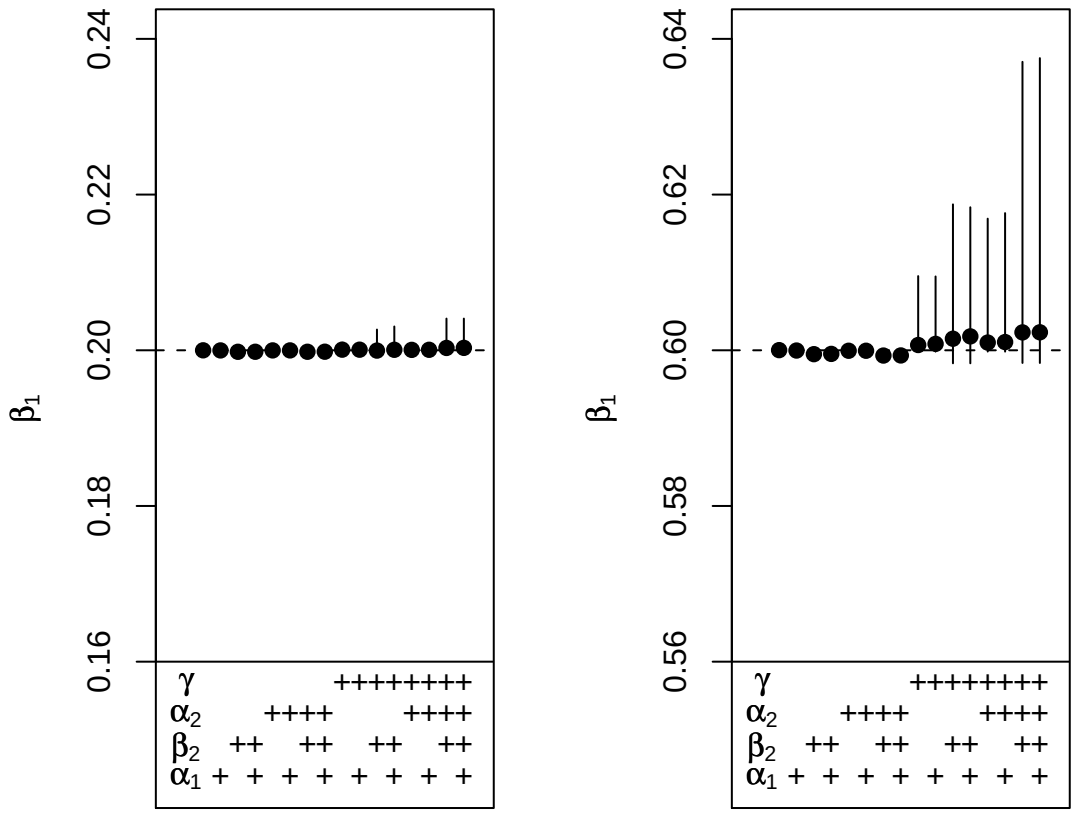

Figure 8: Limiting posterior marginal distribution of β_1 for $2^5 = 32$ different parameter combinations. The left/right panels correspond to true values $\beta_1 = 0.2$ and $\beta_1 = 0.6$ respectively. The legend indicates which parameters are set to the larger of the two possible values.