SUPPLEMENTARY MATERIALS: On the importance of the Jacobian determinant in parameter inference for random parameter and random measurement error models *

David Swigon†, Shelby R. Stanhope‡, Sven Zenker§, and Jonathan E. Rubin¶

SM1. Bayesian formulation for random parameter model. A perceptive reader may wonder whether the density of the random parameter model could be estimated using Bayesian inference methodology. Bayesian nonparametric population models have been studied for example by Wakefield and Walker [SM4]. Since it is generally difficult to define probability on the space of distributions, results have been obtained for maximum likelihood estimation of the most probable distribution that corresponds to the data [SM2, SM3].

In some random parameter systems, one may have prior information about the variance of the parameters but not about their mean, i.e., one may assume that the random variable A is a sum of a fixed unknown parameter a and random effect B with zero mean and known density $\beta(b)$:

$$Y = F(a + B). \tag{SM1}$$

One can then formulate the Bayesian inference problem for the parameter a given a knowledge of the distribution of B and the data D. The parameter a in this context plays the role of the mean of the parameter distribution $\rho(a)$.

For the statistical extension (SM1) the likelihood of a given a single data vector y is given by[1]

$$L(a|y) = \beta(F^{-1}(y) - a)J(F^{-1}(y))^{-1}. \tag{SM2}$$

By BayesTheorem, the posterior distribution is

$$\begin{aligned}\sigma(a|y) &\propto \beta(F^{-1}(y) - a)J(F^{-1}(y))^{-1}\pi(a) \\ &\propto \beta(F^{-1}(y) - a)\pi(a)\end{aligned} \tag{SM3}$$

(Note that here and in equation (SM4) below the normalization over a removes any multiplicative factor that does not depend on a.)

*Submitted to the editors DATE.

†Department of Mathematics, University of Pittsburgh, Pittsburgh, PA (swigon@pitt.edu).

‡Department of Mathematical Sciences, United States Air Force Academy, USAF Academy, CO (shelby.stanhope@usafa.edu).

§Department of Anesthesiology and Intensive Care Medicine, University of Bonn Medical Center, Bonn, Germany (zenker@uni-bonn.de).

¶Department of Mathematics, University of Pittsburgh, Pittsburgh, PA (jonrubin@pitt.edu).

[1] The Jacobian factor is required in order for $L(a|y)$ to integrate to 1 over y at fixed a.

For statistical extension (SM1), the likelihood of a given the data set $D = \{y^1, y^2, \dots y^N\}$ is the product likelihood $L(a|D) = \prod_{i=1}^N L(a|y^i)$, and, by Bayes Theorem, the posterior distribution is given by

$$\sigma(a|D) \propto \prod_{i=1}^{N} \beta(F^{-1}(y^i) - a)\pi(a). \tag{SM4}$$

In the limit of a large data set with density $\eta(y)$, the Bayesian posterior reduces to a singular distribution localized at the value a_{KL} given by (see [SM1])

$$\begin{aligned} a_{KL} = \arg\min_a D_{KL}(\eta||L) &= \arg\min_a \int -\ln[\beta(F^{-1}(y) - a)J(F^{-1}(y))^{-1}]\eta(y)dy \\ &= \arg\min_a \int -\ln[\beta(x-a)J(x)^{-1}]\eta(F(x))J(x)dx \\ &= \arg\min_a \int -\ln[\beta(x-a)]\rho(x)dx \end{aligned} \tag{SM5}$$

where $\rho(x) = \eta(F(x))J(x)$ is the pullback of the density η under the map F.

The formula (SM5) does not require that the distributions β and η be consistent, in the sense that they are related by a transformation induced by the model (SM1). One can obtain β consistent with the data by transformation of the probability density as $\rho(a) = \eta(F(a))J(a)$ and taking $\beta(b) = \rho(b + \bar{a})$, where $\bar{a}$ is the mean of the density ρ. In that case, in view of the non-negativity of the Kullback-Leibler divergence,

$$a_{KL} = \bar{a} = \int x\ \rho(x)dx = \int F^{-1}(y)\eta(y)dy = \mathbb{E}[F^{-1}(Y)]. \tag{SM6}$$

Note that the singular Bayesian posterior differs from that (equation (8) of the main paper) for the random measurement error model: whereas for (8) the distribution is centered at the inverse of the mean observation, for the random parameter model it is centered at the mean of the inverse observations. The two estimates coincide when F is a linear map.

Application of Jeffreys principle to the statistical extension (SM1) can be made again as follows: In view of (SM2), the Fisher information matrix $I(a)$ is here given by

$$\begin{aligned} I(a)_{ij} &= \int \left(\frac{\partial}{\partial a_i} \ln\beta(F^{-1}(y) - a)\right)\left(\frac{\partial}{\partial a_j} \ln\beta(F^{-1}(y) - a)\right)\beta(F^{-1}(y) - a)J(F^{-1}(y))^{-1}\,dy \\ &= \int \frac{\partial\beta(w)}{w_i}\frac{\partial\beta(w)}{w_j}\beta(w)^{-1}J(w+a)^{-1}\,|\det D_aF(w+a)|\,dw \end{aligned}$$

and hence $I(a) = W$, a positive-definite symmetric matrix with $W_{ij} = \int \frac{\partial\ln\beta(w)}{w_i}\frac{\partial\ln\beta(w)}{w_j}\beta(w)\,dw$. The corresponding invariant noninformative prior for model (SM1) is therefore the uniform prior $\pi(a) = (\det I(a))^{1/2} = (\det W)^{1/2}$.

Using the results above, one can extend the Table 9 of the main paper as shown in Table SM1, with the appropriate Jeffreys prior used in each Bayesian inference, and with densities γ and β consistent with the data density η.

Data\Model	$Y = F(a) + G$ (Bayesian inference)	$Y = F(A)$ (measure transformation)	$Y = F(a + B)$ (Bayesian inference)
$D = \{y\}$	$\gamma(y - F(a))J(a)$	$\delta(F^{-1}(y) - a)$	$\beta(F^{-1}(y) - a)$
$D = \{y_i\}_{i=1}^N$	$\prod_i \gamma(y_i - F(a))J(a)$	$\sum_i \delta(F^{-1}(y_i) - a)$	$\prod_{i=1}^N \beta(F^{-1}(y^i) - a)$
$D = \eta(y)$	$\delta(F^{-1}(\mathbb{E}[Y]) - a)$	$\eta(F(a))J(a)$	$\delta(\mathbb{E}[F^{-1}(Y)] - a)$

Table SM1
Inferred parameter densities.

Furthermore, a result similar to the Theorem in the Appendix of the main paper can be obtained for the posterior distribution for model $Y = F(a + B)$ and its relation to the parameter density of $Y = F(A)$. Since both results refer to the random parameter model, this result is less surprising and less practical:

Theorem SM1. *Given model $y = F(a)$ and data density $\eta(y)$, the parameter density $\rho(a)$ for the statistical extension $Y = F(A)$ derived from η is identical to the Bayesian posterior density $\sigma(a|\bar{y})$ for extension $Y = F(a + B)$ with error density β derived from ρ and with $\bar{y} = F(\mathbb{E}[F^{-1}(Y)])$, provided $\rho(a)$ is symmetric and the prior distribution is uniform.*

Proof. For the statistical extension $Y = F(a + B)$, the error density β derived from the data density obeys $\beta(b) = \rho(b + \bar{a})$, where $\bar{a} = \mathbb{E}[F^{-1}(Y)]$. Given the symmetry of β, the likelihood of a given data $\bar{y} = F(\bar{a})$ is given by $L(a|\bar{y}) = \beta(F^{-1}(\bar{y}) - a)J(\bar{a})^{-1} = \beta(a - \bar{a})J(\bar{a})^{-1}$ (see (SM2)). Thus, the posterior distribution with prior π is given by

$$\sigma(a|\bar{y}) \propto L(a|\bar{y})\pi(a) = \beta(a - \bar{a})\pi(a) = \rho(a)\pi(a).$$

It follows that $\sigma(a|\bar{y}) = \rho(a)$ if and only if $\pi(a)$ is constant for all a in the support of $\rho(a)$. ■

SM2. Additional Examples. In this section we report several examples that complement the results presented in the main paper. We here focus on the effect of the choice of the time points at which data are sampled on the accuracy of inferred distributions for the random parameter model.

Example SM1. In this example, with prescribed $\rho(a)$, we choose the same baseline parameters and coefficients of variation as in Examples 7 and 8 of the main paper, and observation time points $\{t_1, t_2\} = \{2, 3\}$ as in Example 6. The new observation times better capture data in the decreasing portion of the V component of the solution, and there is an improvement in the accuracy of estimation of the marginal in c (Figure SM1(b)), but a degradation in accuracy of the estimate of β and r. In this example, all of the posteriors perform similarly but they show significant deviation from A in all parameters but H_0, in contrast to the case when $\eta(y)$ was prescribed as shown in Example 6, highlighting the effect of data aggregation.

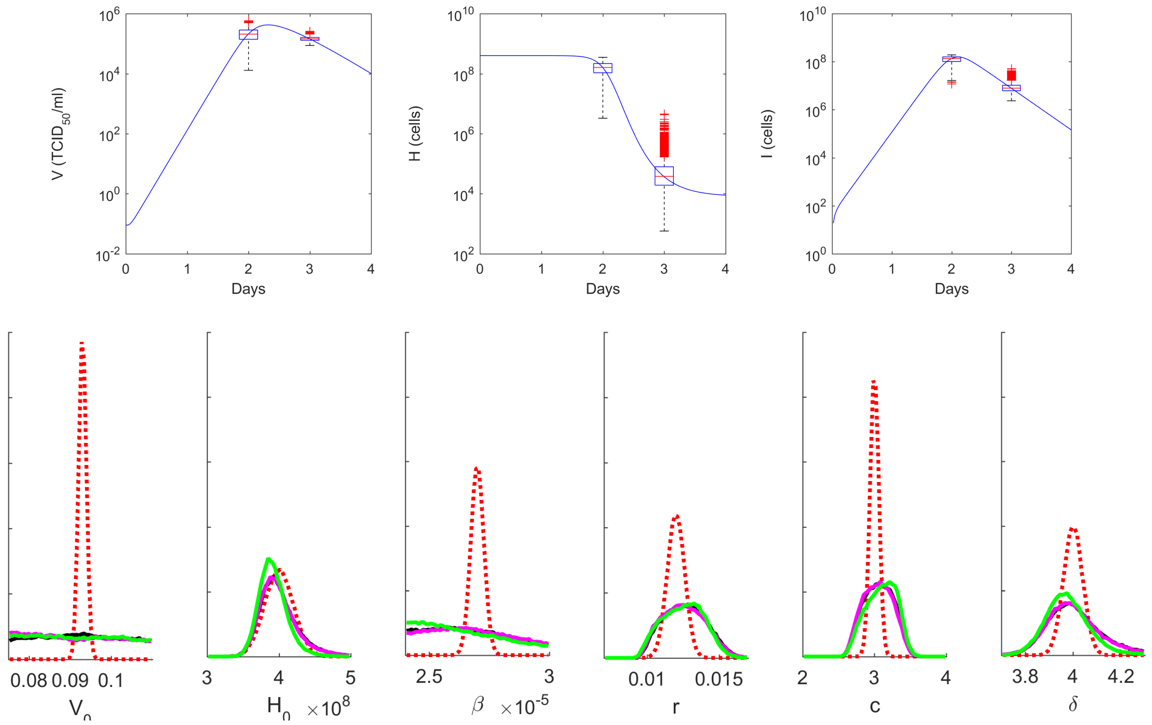

Figure SM1. *Example SM1: (Top row) Solution curves for the mean parameter values and box plot representations of Y. (Bottom row) Marginalized histograms for A (dotted red), M_{Rec} (magenta), M_{Unif} (black), and M_{Jac} (green).*

In the next series of examples for influenza model (19), all with prescribed $\rho(a)$, we choose the same baseline parameters as in Examples 7 and 8 of the main paper, but with the coefficient of variation set to $s = 0.05$ for all parameters.

Example SM2. Here the observation times are taken as in Example 7: $\{t_1, t_2\} = \{1, 2\}$. Uniformity in the coefficient of variation leads to an improvement in the accuracy of estimates for V_0, β, and δ compared to Example 7, as shown in Figure SM2. M_{Jac} performs the best for V_0, β, and δ but all marginals still show significant deviation from A for V_0, r, and especially c, in contrast to the case when $\eta(y)$ was prescribed as shown in Example 6, again highlighting the effect of data aggregation.

Example SM3. Here the observation times are taken as in Example SM1: $\{t_1, t_2\} = \{2, 3\}$. Uniformity in coefficient of variation leads to an improvement in the accuracy of estimates for β and δ, compared to Example SM1, as shown in Figure SM3. M_{Jac} performs the best for δ but is uncharacteristically skewed for H_0 and β compared to other cases. All other marginals show significant deviation from A, again highlighting the effect of data aggregation.

Example SM4. Here the observation times are taken as in Example 8: $\{t_1, t_2\} = \{1, 3\}$. Uniformity in coefficient of variation leads to an improvement in the accuracy of estimates

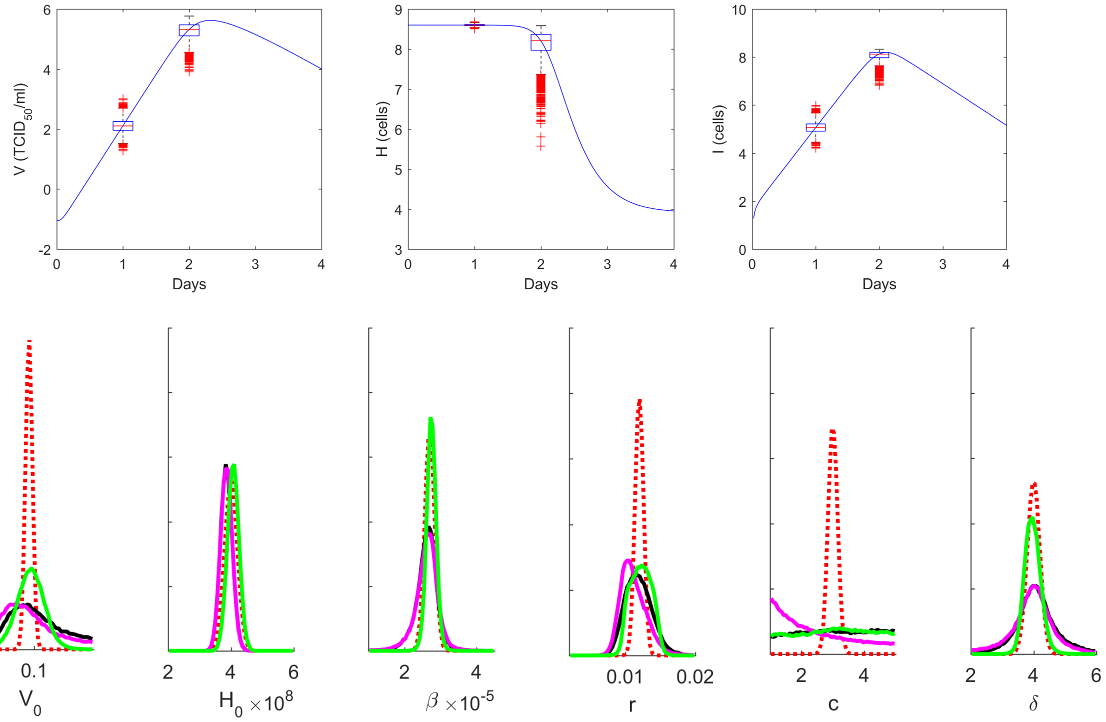

Figure SM2. *Example SM2: (Top row) Solution curves for the mean parameter values and box plot representations of Y. (Bottom row) Marginalized histograms for A (dotted red), $M_{\rm Rec}$ (magenta), $M_{\rm Unif}$ (black), and $M_{\rm Jac}$ (green).*

for β and δ, and degradation in accuracy for r, compared to Example 8, as shown in Figure SM4. There is little difference in the performance of different priors, similarly as in Example 8. The marginals show slight deviation from A, except for V_0 which has significant deviation.

REFERENCES

[1] B. Kleijn, A. Van der Vaart, et al., *The Bernstein–von–Mises theorem under misspecification*, Electronic Journal of Statistics, 6 (2012), pp. 354–381.

[2] A. Mallet, *A maximum likelihood estimation method for random coefficient regression models*, Biometrika, 73 (1986), pp. 645–656.

[3] A. Mallet, F. Mentré, J.-L. Steimer, and F. Lokiec, *Nonparametric maximum likelihood estimation for population pharmacokinetics, with application to cyclosporine*, Journal of Pharmacokinetics and Pharmacodynamics, 16 (1988), pp. 311–327.

[4] J. Wakefield and S. Walker, *Bayesian nonparametric population models: formulation and comparison with likelihood approaches*, Journal of Pharmacokinetics and Pharmacodynamics, 25 (1997), pp. 235–253.

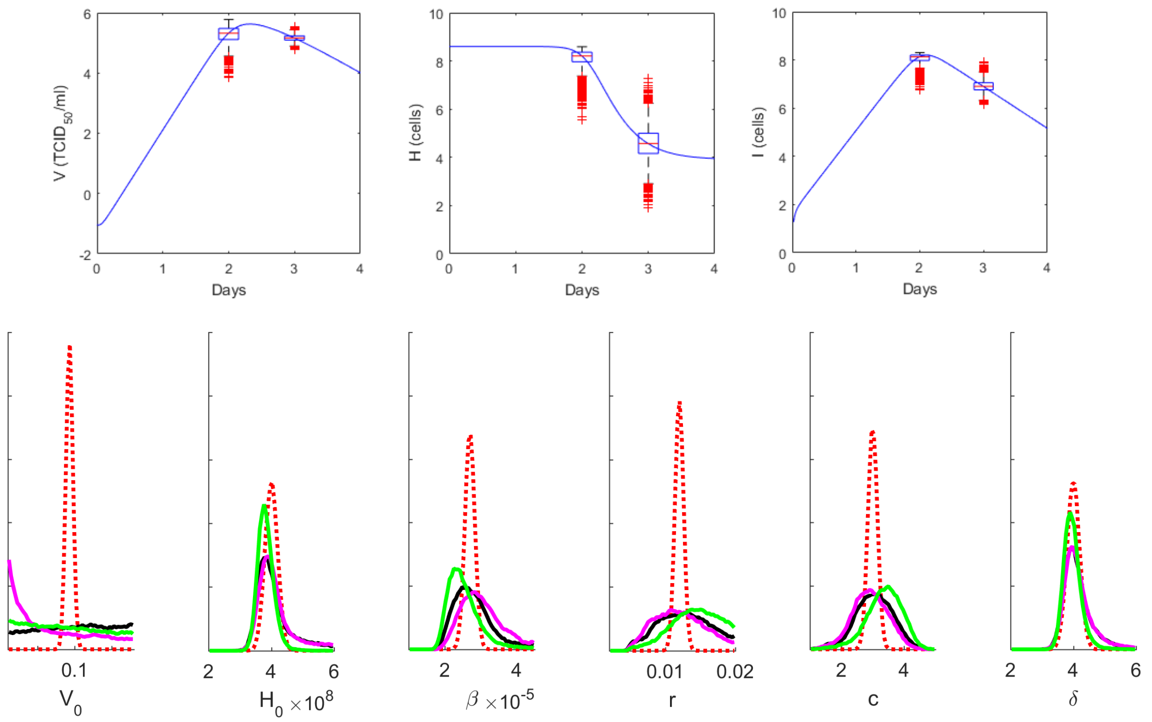

Figure SM3. *Example SM3: (Top row) Solution curves for the mean parameter values and box plot representations of Y. (Bottom row) Marginalized histograms for A (dotted red), M_{Rec} (magenta), M_{Unif} (black), and M_{Jac} (green).*

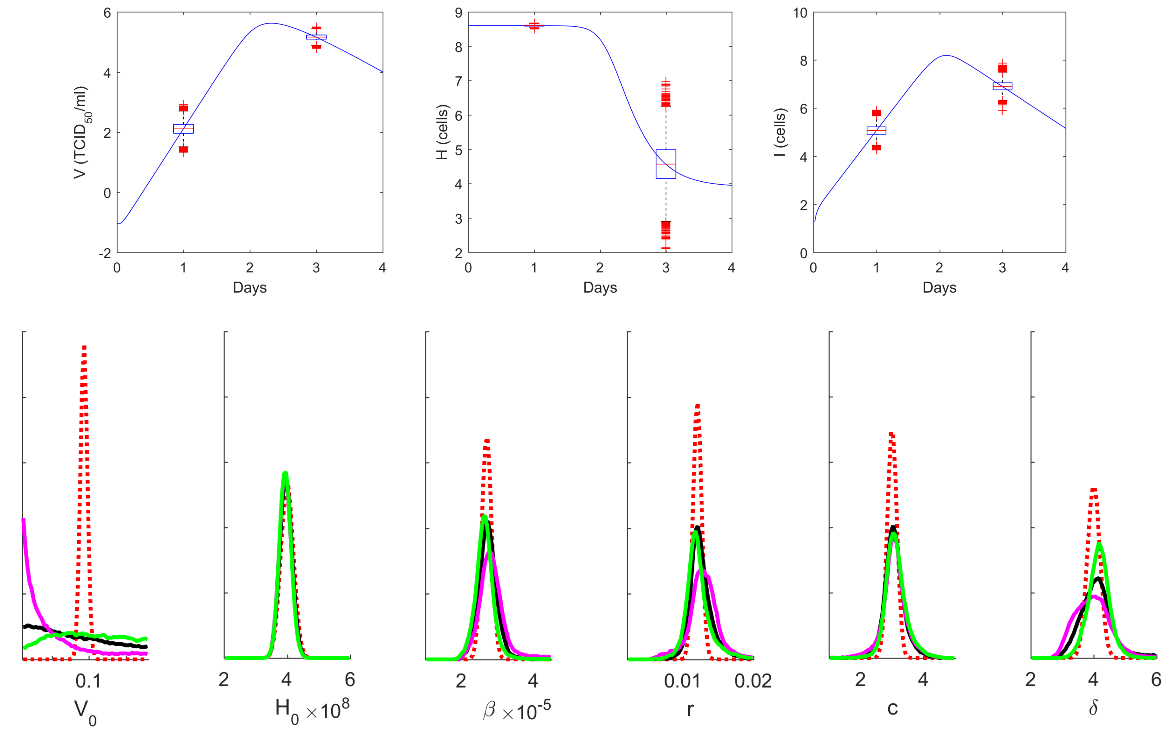

Figure SM4. *Example SM4: (Top row) Solution curves for the mean parameter values and box plot representations of Y. (Bottom row) Marginalized histograms for A (dotted red), M_{Rec} (magenta), M_{Unif} (black), and M_{Jac} (green).*