

Supplementary Table S2. Model validity and accuracy parameters definition and interpretation.

Model criteria	Parameter	Definition	Interpretation	Reference
Model Validity	Acceptance rate (AR)	The acceptance rate is the proportion of proposed values of θ that were included in our final Markov chain Monte Carlo (MCMC) sample.	The asymptotically optimal acceptance rate is 0.234 under quite general conditions	Roberts, Gelman, and Gilks (1997).
	Efficiency (Ef)	Efficiency describes mixing properties of the Markov chain.	High efficiency means good mixing (low autocorrelation) in the MCMC sample, and low efficiency means bad mixing (high autocorrelation) in the MCMC sample. An efficient MH sampler has an AR between 15% and 50% and low autocorrelation and thus relatively large effective sample size (ESS) for all model parameters.	Roberts and Rosenthal (2001)
	Monte Carlo standard error (MCSE)	The Monte Carlo standard error (MCSE) is an approximation of the error in estimating the true posterior mean.	The lowest the MCSE the most accurate is the prediction of true posterior mean	Flegal (2008)
Model Accuracy	The posterior predictive P values provides (PPP values)	Computed as a goodness of fit measure for the model being tested. PPP values is the proportion of time during an MCMC run that a chosen test statistic, generated from a distribution predicted by the model, is higher than the test statistic generated from the distribution of the actual input data.	A value around 0.5 indicating a plausible good-fitting model and values toward the extremes of 0 or 1 indicating that the model is not plausible.	Defined in Appendix C of Lee and Song (2003)
	The marginal likelihood, also known as the evidence (Ev)	Denominator of the Bayes equation. The Log marginal likelihood is the combination of a data fit term and complexity penalty.	The model reporting the highest log marginal likelihood is precisely the model that is the best sequential predictor of the data tested according to the log scoring rule. A difference of 0.01 between two log-likelihood values is considered to be the same model. A difference of more than 3 log likelihood units (considered as “strong evidence against competing model” can be used as threshold for accepting a more parameter-rich model. A positive log likelihood means that the likelihood is larger than 1. This is possible because the log likelihood is not itself the probability of observing the data, but just proportional to it.	Chickering and Heckerman (1996), Aydin, Marcussen, Ertekin, and Oxelman (2014) and Edwards (1992).