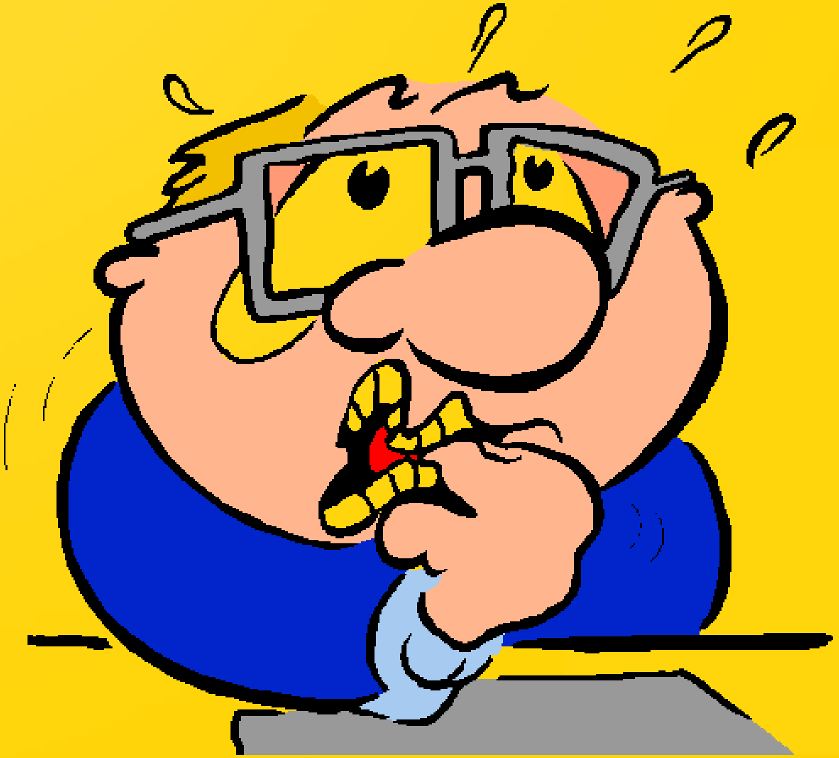
Universiteit Utrecht

# Bayes: When to worry?

Rens van de Schoot & Sarah Depaoli



Dear dr. ,

We would kindly invite you to review this paper about …

Because of the small sample size (n=20) we used Bayesian estimation. Hox et al. (2012) showed that a multilvel model with only 20 clusters could be estimated with Bayesian statistics whereas maximum likelihood estimation could not.

Hox, J., van de Schoot. R., & Matthijsse, S. (2012). How few countries will do? Comparative survey analysis from a Bayesian perspective. Survey Research Methods, 6, 87-93.

Since we are no experts in Bayesian estimation we relied on the default settings

The results are completely in line with our hypothesis: there is a significant difference between the two groups.  All is fine, please accept our paper for publication.

# ACCEPT
# OR
# REJECT??

# Making Decisions when Implementing Bayes

- **Naively applying Bayesian methods can be dangerous** for three main reasons:

    - First, the exact influence of the priors is often not well understood and priors might have a huge impact on the study results;

    - Second, akin to many elements of frequentist statistics, some Bayesian features can be easily misinterpreted;

    - Third, reporting on Bayesian statistics follows its own rules since there are elements included in the Bayesian framework that are fundamentally different from frequentist settings.

# Making Decisions when Implementing Bayes

- **Naively apply Bayesian metho s can be dangerous** for e main re
  - well t on the
  - wn rules since th are e s in th ian framework that are funda entally di ent from frequentist settings.

WAMBS-checklist
When to worry and how to Avoid the Misuse of Bayesian Statistics

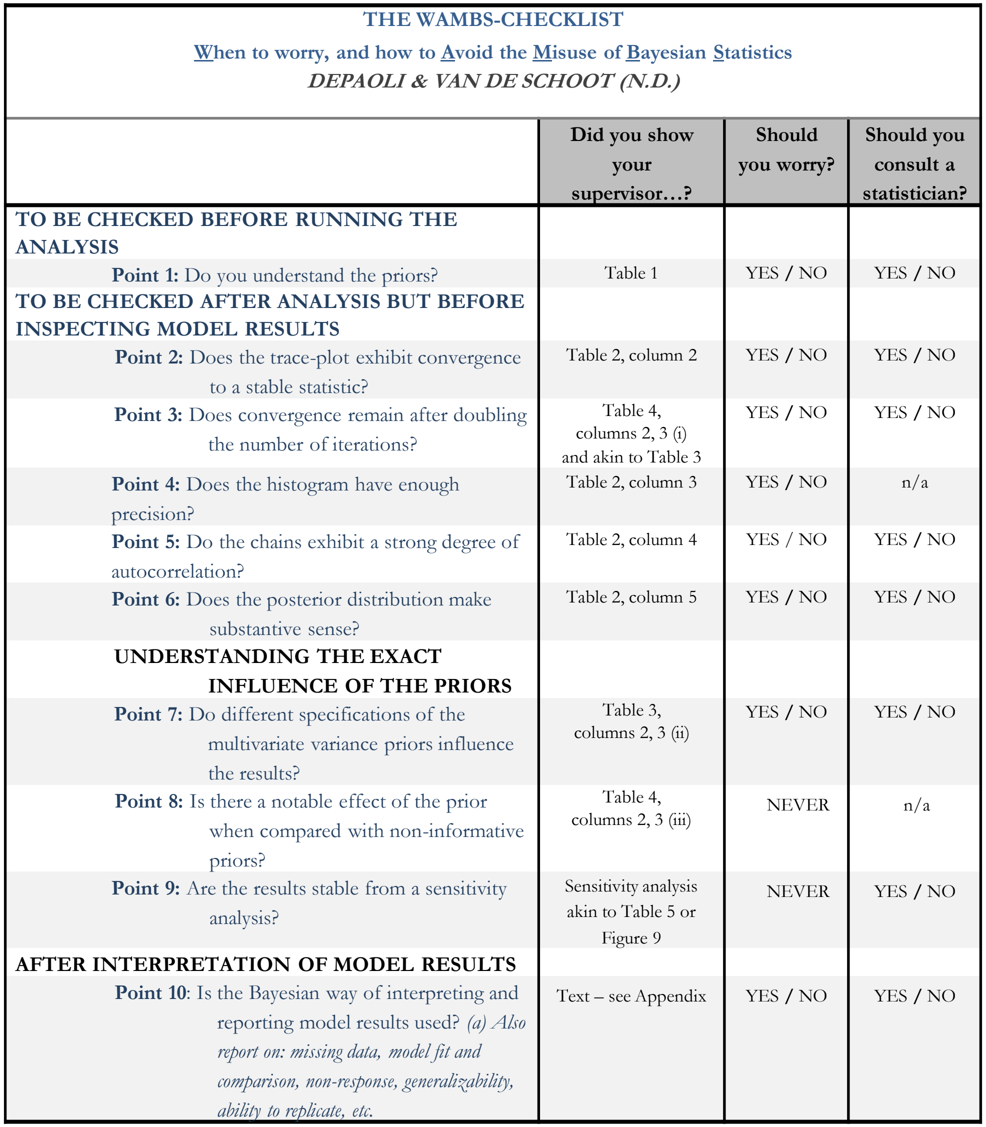
# WAMBS checklist

- 10 main points that should be thoroughly checked when applying Bayesian analysis:

(a) issues to check before running the analysis,
(b) issues to check after running the analysis but before interpreting results,
(c) understanding the influence of priors, and
(d) steps after interpreting results

Depaoli, Sarah, Van de Schoot, R. (2015). **The WAMBS-Checklist: When to Worry, and how to Avoid the Misuse of Bayesian Statistics**. *Psychological Methods.*

# WAMBS checklist

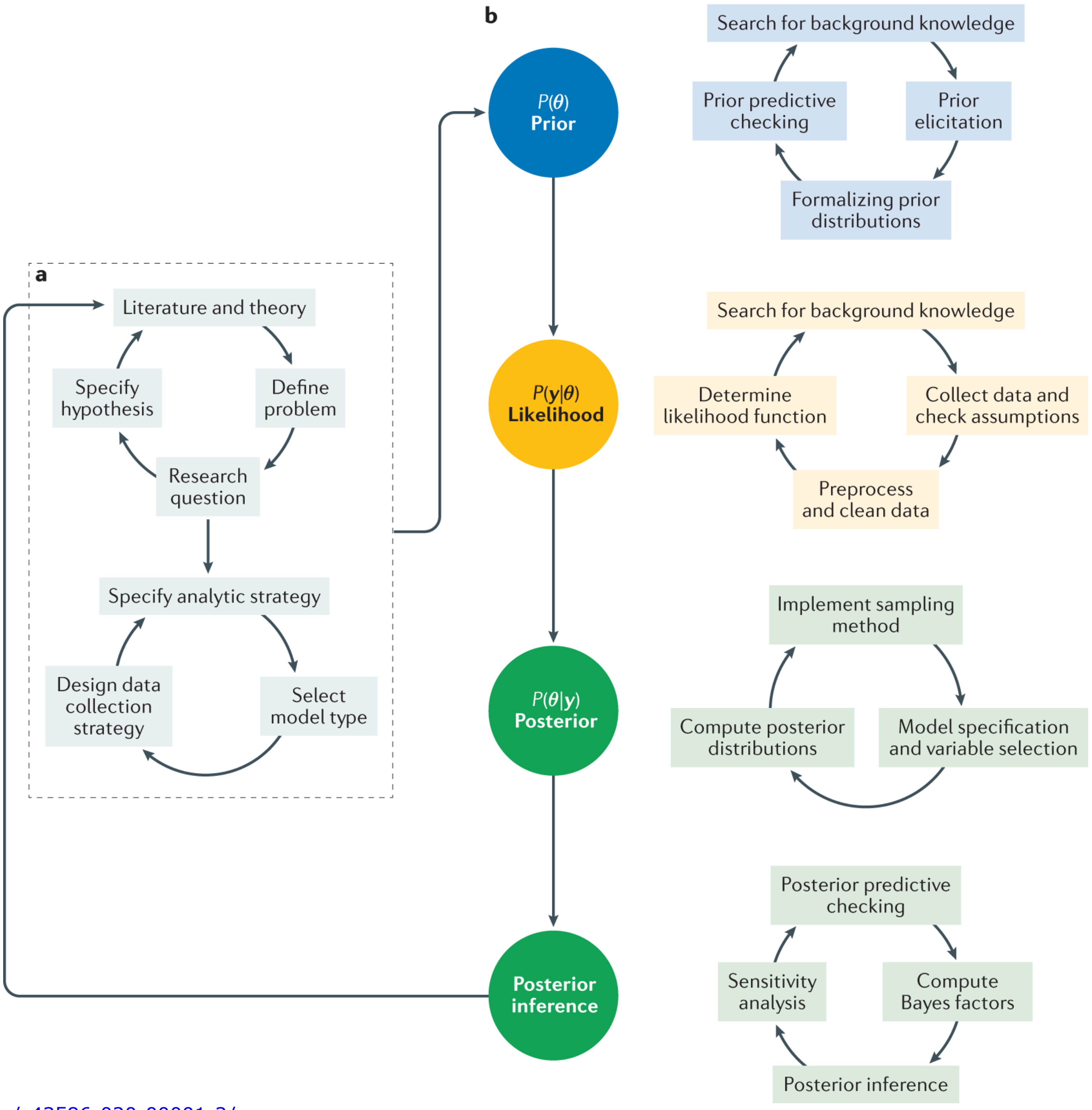


THE WAMBS-CHECKLIST

When to worry, and how to Avoid the Misuse of Bayesian Statistics

*DEPAOLI & VAN DE SCHOOT (N.D.)*

| | Did you show your supervisor…? | Should you worry? | Should you consult a statistician? |
|---|---|---|---|
| **TO BE CHECKED BEFORE RUNNING THE ANALYSIS** | | | |
| **Point 1:** Do you understand the priors? | Table 1 | YES / NO | YES / NO |
| **TO BE CHECKED AFTER ANALYSIS BUT BEFORE INSPECTING MODEL RESULTS** | | | |
| **Point 2:** Does the trace-plot exhibit convergence to a stable statistic? | Table 2, column 2 | YES / NO | YES / NO |
| **Point 3:** Does convergence remain after doubling the number of iterations? | Table 4, columns 2, 3 (i) and akin to Table 3 | YES / NO | YES / NO |
| **Point 4:** Does the histogram have enough precision? | Table 2, column 3 | YES / NO | n/a |
| **Point 5:** Do the chains exhibit a strong degree of autocorrelation? | Table 2, column 4 | YES / NO | YES / NO |
| **Point 6:** Does the posterior distribution make substantive sense? | Table 2, column 5 | YES / NO | YES / NO |
| **UNDERSTANDING THE EXACT INFLUENCE OF THE PRIORS** | | | |
| **Point 7:** Do different specifications of the multivariate variance priors influence the results? | Table 3, columns 2, 3 (ii) | YES / NO | YES / NO |
| **Point 8:** Is there a notable effect of the prior when compared with non-informative priors? | Table 4, columns 2, 3 (iii) | NEVER | n/a |
| **Point 9:** Are the results stable from a sensitivity analysis? | Sensitivity analysis akin to Table 5 or Figure 9 | NEVER | YES / NO |
| **AFTER INTERPRETATION OF MODEL RESULTS** | | | |
| **Point 10**: Is the Bayesian way of interpreting and reporting model results used? *(a) Also report on: missing data, model fit and comparison, non-response, generalizability, ability to replicate, etc.* | Text – see Appendix | YES / NO | YES / NO |


a
Literature and theory
Define problem
Research question
Specify hypothesis
Specify analytic strategy
Select model type
Design data collection strategy
b
$P(\theta)$ Prior
Search for background knowledge
Prior elicitation
Formalizing prior distributions
Prior predictive checking
$P(\mathbf{y}|\theta)$ Likelihood
Search for background knowledge
Collect data and check assumptions
Preprocess and clean data
Determine likelihood function
$P(\theta|\mathbf{y})$ Posterior
Implement sampling method
Model specification and variable selection
Compute posterior distributions
Posterior inference
Posterior predictive checking
Compute Bayes factors
Posterior inference
Sensitivity analysis

www.nature.com/articles/s43586-020-00001-2/

# Stage 1:

# To be Checked before Running the Analysis

# Where do your priors come from?

# Priors

- When specifying priors, it is important to recognize that prior distributions fall into three main classes related to the amount of (un)certainty they contribute to the model about a given parameter:

(1) *non-informative priors,*
(2) *weakly-informative priors* and
(3), *informative priors*

The term “non-informative prior” refers to the case where researchers supply vague information about the population parameter value; the prior is typically defined with a very wide variance (Gill, 2008). Although “non-informative” is one term commonly used in the Bayesian literature to describe this type of prior (see e.g., Gelman et al., 2004), other phrases such as “diffuse” (see e.g., Gill, 2008), or “flat” (Jeffreys, 1961) are also used to describe this type of prior. We use “non-informative” and “diffuse” interchangeably in the current paper.

# Prior source

The information embedded in the informative prior can come from a variety of places, for example:

- an expert, or a panel of experts,
- results of a previous publication as prior specification
- meta-analysis
- a pilot study
- data-based priors can be derived based on a variety of methods including:
  - maximum likelihood
  - or sample statistics
  - Training data
  - Data splitting priors

Note that there are some arguments against using such “double-dipping” procedures where the sample data are used to derive priors and then used in estimation

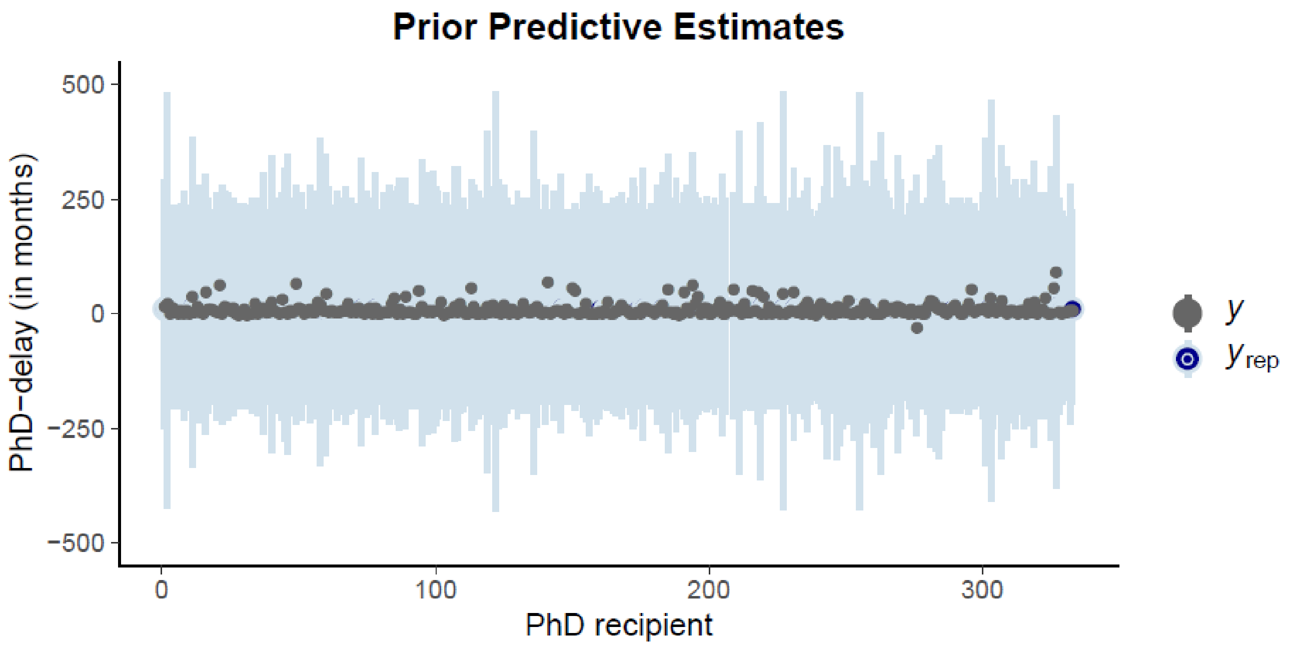
# Guidelines

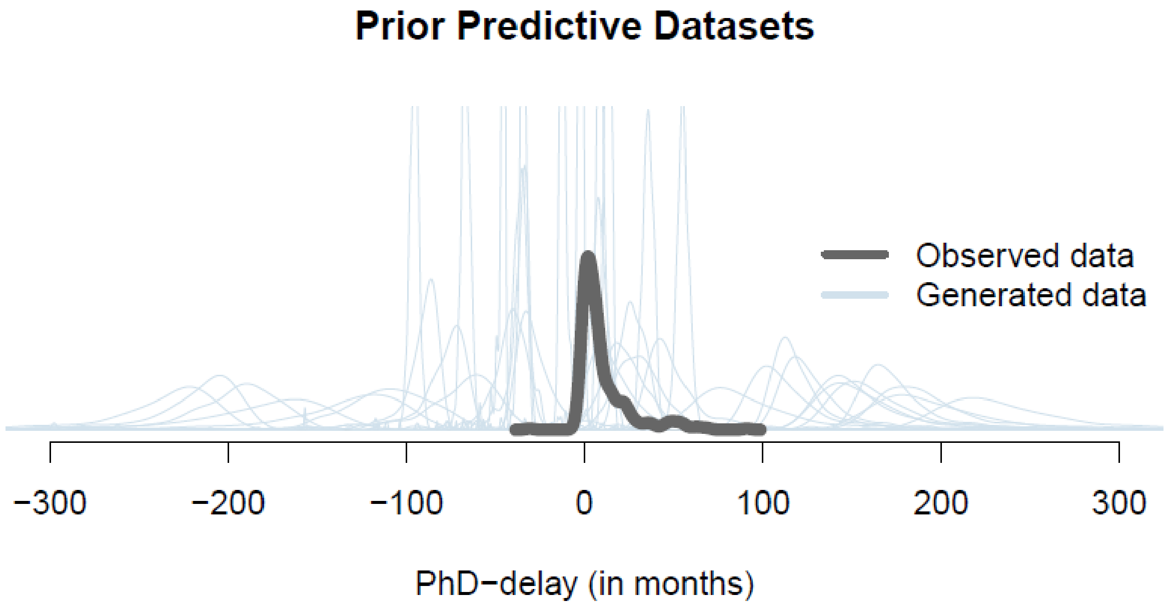
- Determine what **strategy** suits the project of interest best with questions like:
  - Could prior information likely be found in the ***literature*** (e.g., meta-analyses, reviews, empirical studies)? Note that the quantification of prior information is more straightforward when the literature covers the same variables obtained with the same measures as the data of interest.
  - Are there ***experts*** on the subject matter, and who are they? How can experts contribute? Would experts be able to specify priors for the parameters in the model at hand, or can they contribute in a different manner?
  - What ***general knowledge*** is available about the model parameters?

- Determine how to gather the information **systematically**. Keep a **log** of every decision

# Guidelines

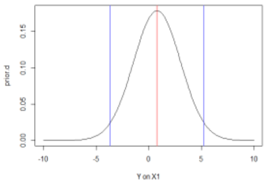
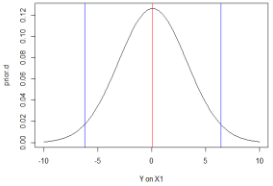
- When you intend to construct informative priors, **visualize** them. A visualization (e.g., with R, or www.wolframalpha.com) quickly shows whether the prior specifications that you consider are reasonable.
- When conducting a Bayesian analysis, always **provide** the following: (1) the **origin** of and **reason** behind the priors, and (2) the exact **specifications** of the priors. See Depaoli and Van de Schoot (2015) for further instructions on reporting Bayesian analyses.
- Conduct a **sensitivity** analysis and show the impact of various priors on the posterior estimates (Van de Schoot et al., 2016). Consider at least the derived informative priors and default priors, but conservative or skeptical priors may be interesting to examine as well.
- Try to **understand** and interpret differences between analyses with different priors.


Prior Predictive Estimates
PhD−delay (in months)
500
250
0
−250
−500
0
100
200
300
PhD recipient
y
$y_{rep}$


Prior Predictive Datasets
Observed data
Generated data
−300
−200
−100
0
100
200
300
PhD−delay (in months)

# 1. Do you understand your priors?

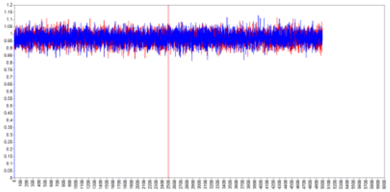
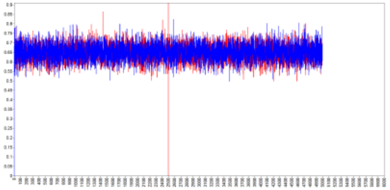
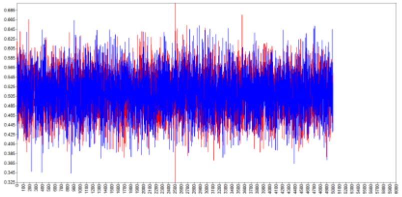
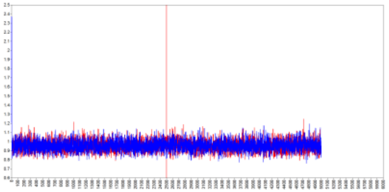
- Ensure the prior distributions and the model or likelihood are well understood and described in detail in the text. Prior-predictive checking can help identify any prior–data conflict.

| Parameters | Distributional form of the priors (e.g., normal, inverse gamma, etc) | Type of prior (non-, weakly, highly informative) | Source of background information | Picture of Plot | Hyperparameters |
|---|---|---|---|---|---|
| Y on $X_1$ | Normal | Highly Informative | Table x on page xx of the meta-analysis of Author et al. (2000) |  | N(.8,5); |
| Y on $X_2$ | Normal | Highly Informative | Obtained from expert knowledge, see Appendix X for more information. |  | N(.1,10); |
| Y: Mean | Normal | Non-Informative (software default) | n/a | n/a | $N(0,10^{10})$; |
| Y: Residual variance | Inverse Gamma | Non-Informative (software default) | n/a | n/a | IG(-1,0); |

# Stage 2:

## To be Checked after Analysis but Before Inspecting Model Results

# 2. Does the trace-plot exhibit convergence to a stable statistic?

| **Parameters** | **Trace plot**<br>**(Point 2)** |
|---|---|
| Y on $X_1$ |  |
| Y on $X_2$ |  |
| Y: Mean |  |
| Y: Residual<br>variance |  |

# 3. Does convergence remain after doubling the number of iterations?

- Assess each parameter for convergence, using multiple convergence diagnostics if possible. This may involve examining trace plots or ensuring diagnostics (Rˆ statistic or effective sample size) are being met for each parameter.
- Sometimes, convergence diagnostics such as the Rˆ statistic can fail at detecting non-stationarity within a chain. Use a subsequent measure, such as the split-Rˆ , to detect trends that are missed if parts of a chain are non-stationary but, on average, appear to have reached diagnostic thresholds.

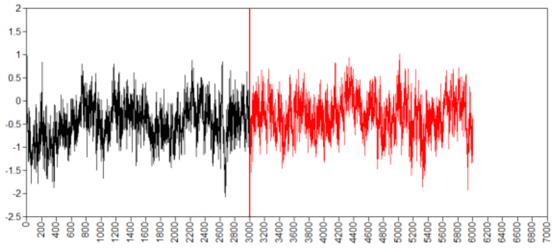
(1) another visual check after doubling the number of iterations;

(2) a convergence diagnostic,

(3) computation of relative bias.

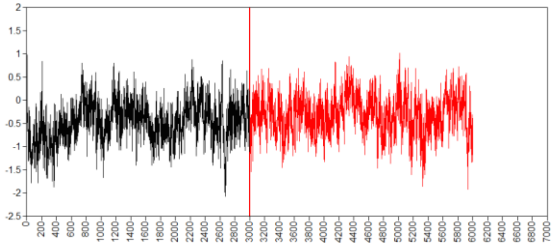
# 3. Does convergence remain after doubling the number of iterations?

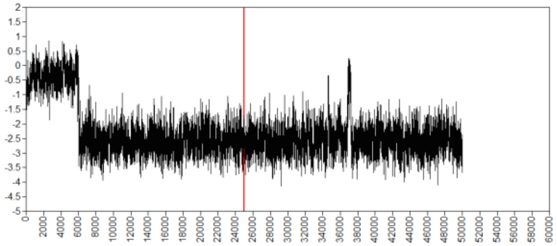
This second check is specifically to avoid obtaining what we call *local convergence*.

| Length of Chain | Parameter<br>Estimate (SD) | Trace Plot |
|---|---|---|
| Shorter chain:<br>6,000 iterations | -0.309(0.417) |  |

# 3. Does convergence remain after doubling the number of iterations?

This second check is specifically to avoid obtaining what we call *local convergence*.

<table>
<tr><th>Length of Chain</th><th>Parameter<br>Estimate (SD)</th><th>Trace Plot</th></tr>
<tr><td>Shorter chain:<br>6,000 iterations</td><td>-0.309(0.417)</td><td>


</td></tr>
<tr><td>Longer chain:<br>50,000 iterations</td><td>-2.574(0.535)</td><td>


</td></tr>
</table>

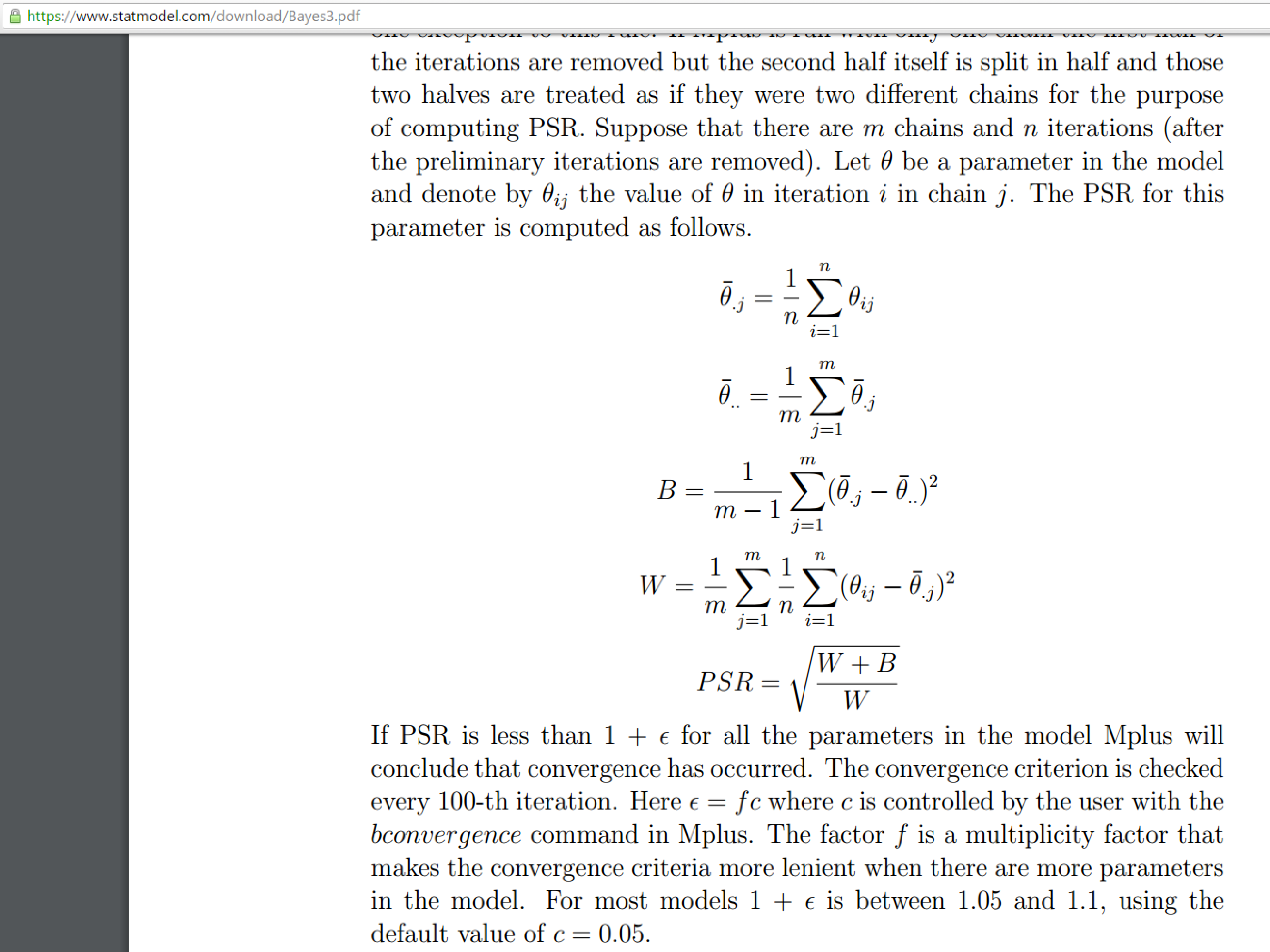
# Rˆ statistic convergence diagnostic

- Gelman and Rubin’s convergence diagnostic (1992)

- based a comparison of within-chain and between-chain variances, and is similar to a classical analysis of variance
  => computed per variable

- The multivariate a version of Gelman and Rubin’s diagnostic was proposed by Brooks and Gelman (1998).

- Values substantially above 1 indicate lack of convergence



https://www.statmodel.com/download/Bayes3.pdf

one exception to this rule. If Mplus is run with only one chain the first half of the iterations are removed but the second half itself is split in half and those two halves are treated as if they were two different chains for the purpose of computing PSR. Suppose that there are $m$ chains and $n$ iterations (after the preliminary iterations are removed). Let $\theta$ be a parameter in the model and denote by $\theta_{ij}$ the value of $\theta$ in iteration $i$ in chain $j$. The PSR for this parameter is computed as follows.

$$\bar{\theta}_{.j} = \frac{1}{n}\sum_{i=1}^{n}\theta_{ij}$$

$$\bar{\theta}_{..} = \frac{1}{m}\sum_{j=1}^{m}\bar{\theta}_{.j}$$

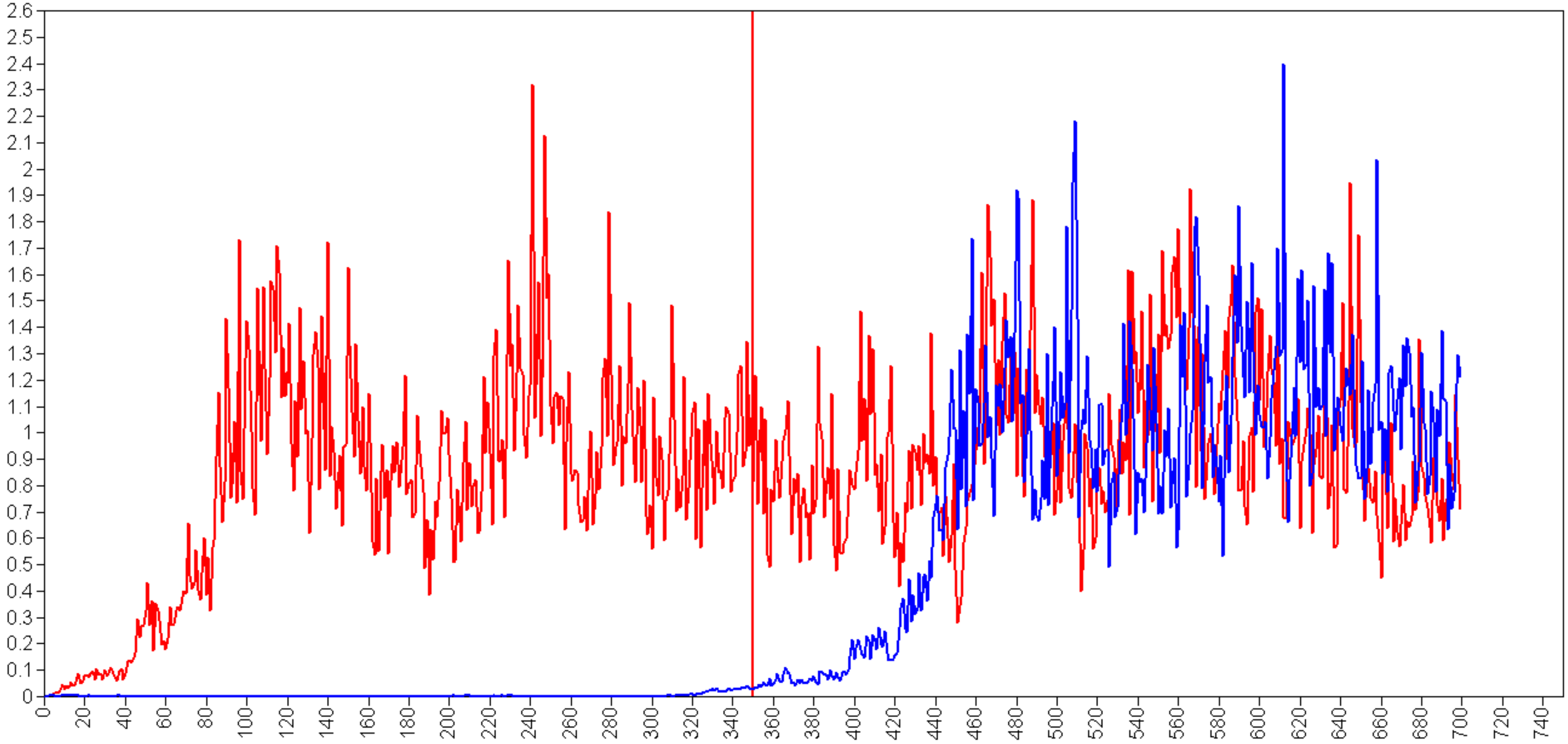
$$B = \frac{1}{m-1}\sum_{j=1}^{m}(\bar{\theta}_{.j} - \bar{\theta}_{..})^2$$

$$W = \frac{1}{m}\sum_{j=1}^{m}\frac{1}{n}\sum_{i=1}^{n}(\theta_{ij} - \bar{\theta}_{.j})^2$$

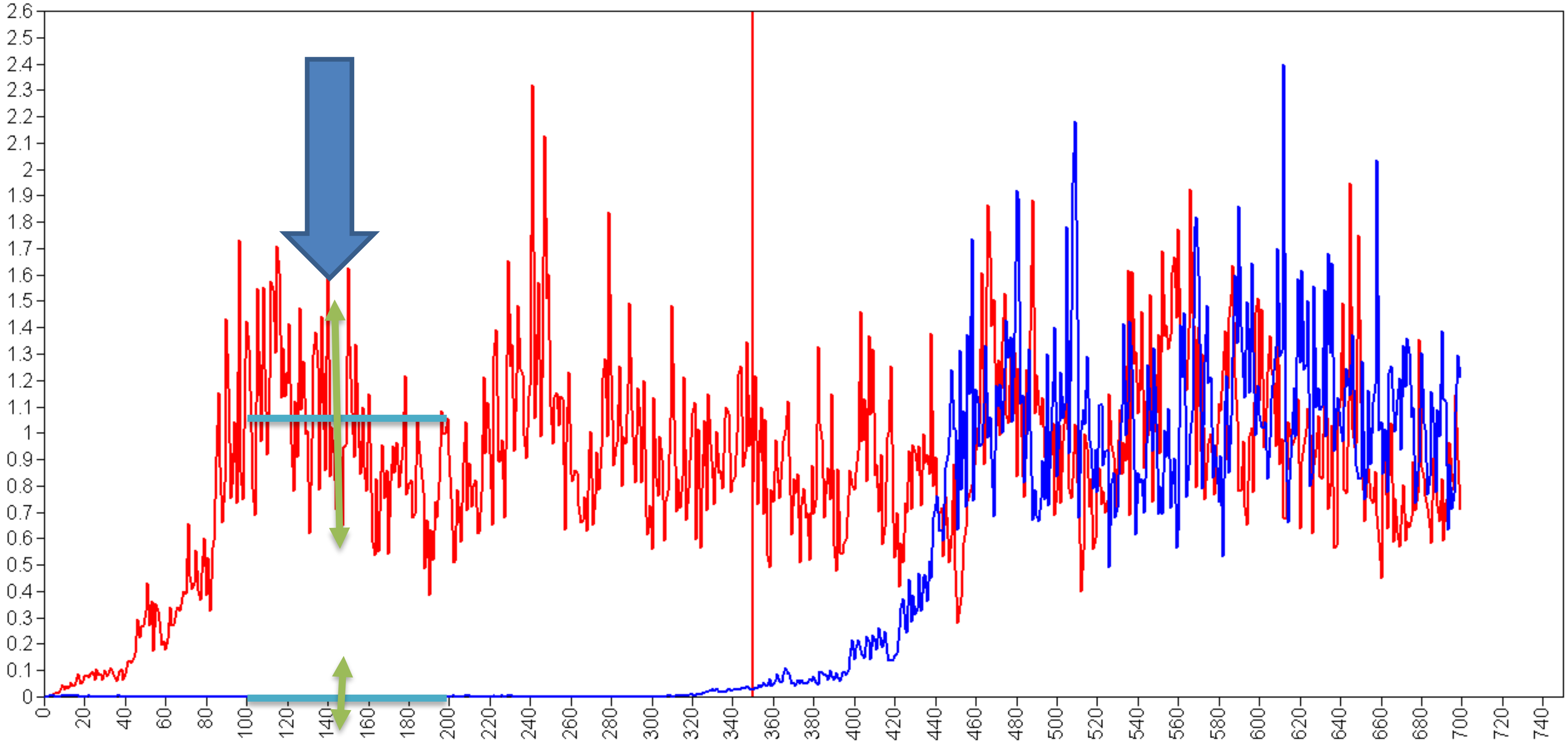
$$PSR = \sqrt{\frac{W+B}{W}}$$

If PSR is less than $1+\epsilon$ for all the parameters in the model Mplus will conclude that convergence has occurred. The convergence criterion is checked every 100-th iteration. Here $\epsilon = fc$ where $c$ is controlled by the user with the *bconvergence* command in Mplus. The factor $f$ is a multiplicity factor that makes the convergence criteria more lenient when there are more parameters in the model. For most models $1+\epsilon$ is between 1.05 and 1.1, using the default value of $c = 0.05$.

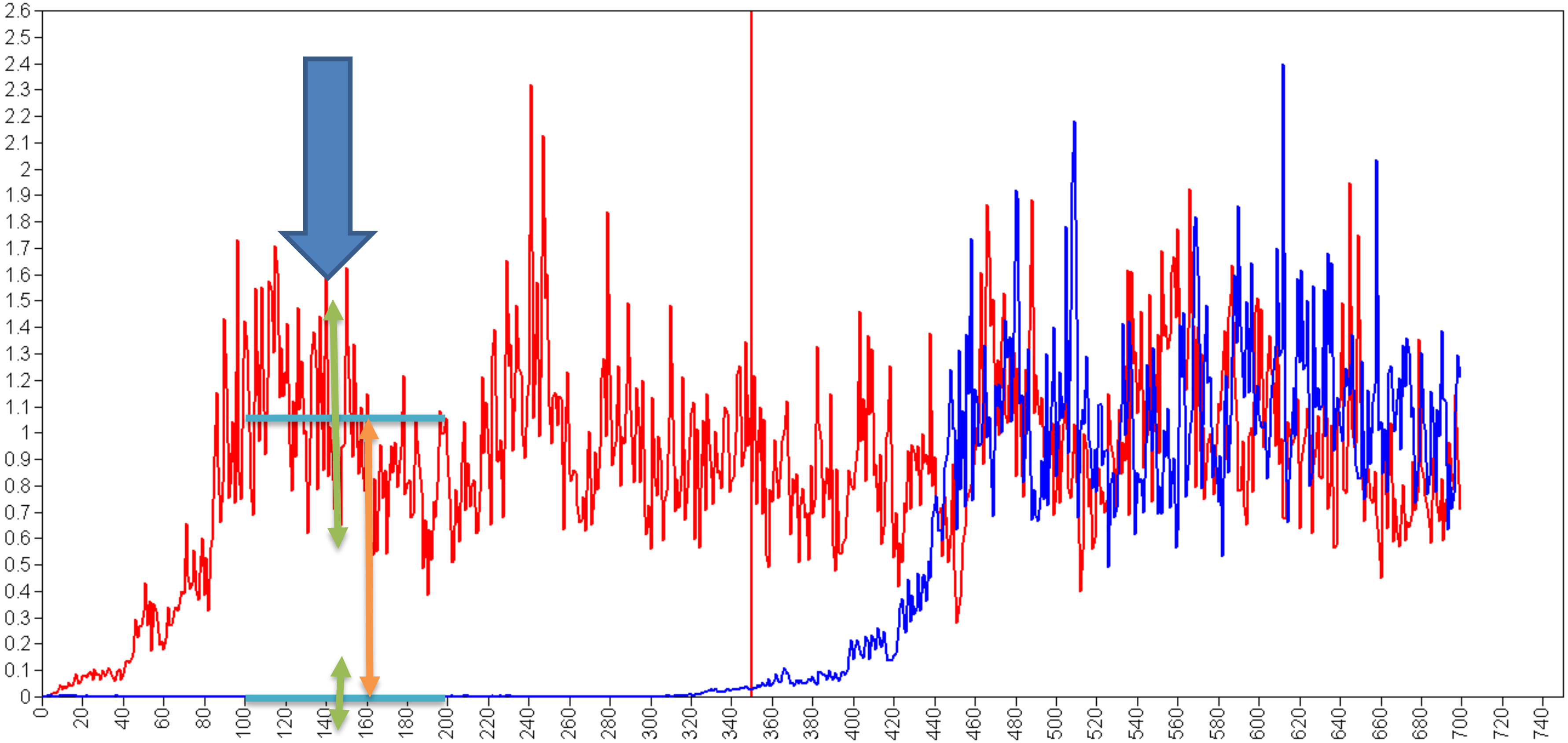
# Gelman and Rubin's convergence diagnostic



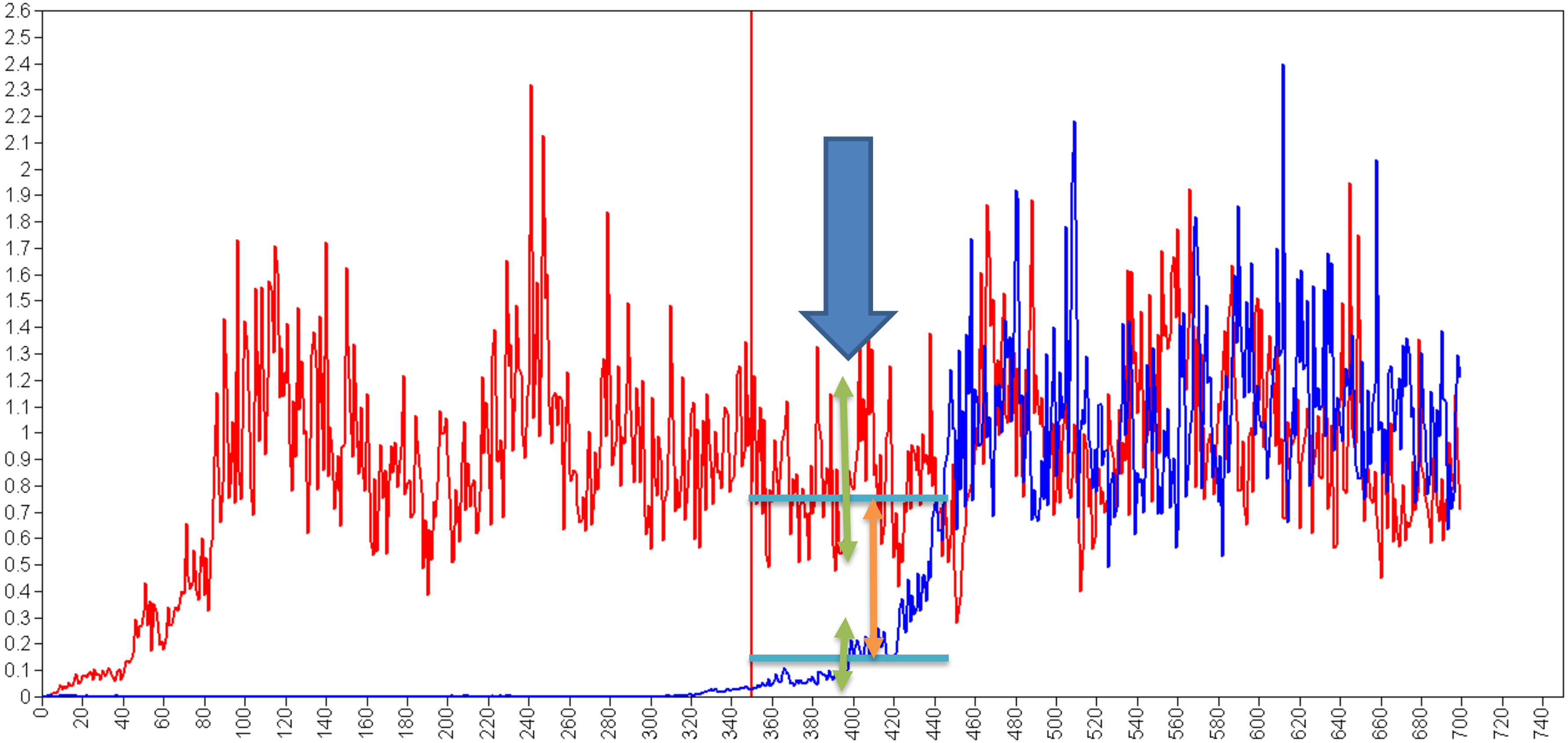
# Gelman and Rubin's convergence diagnostic



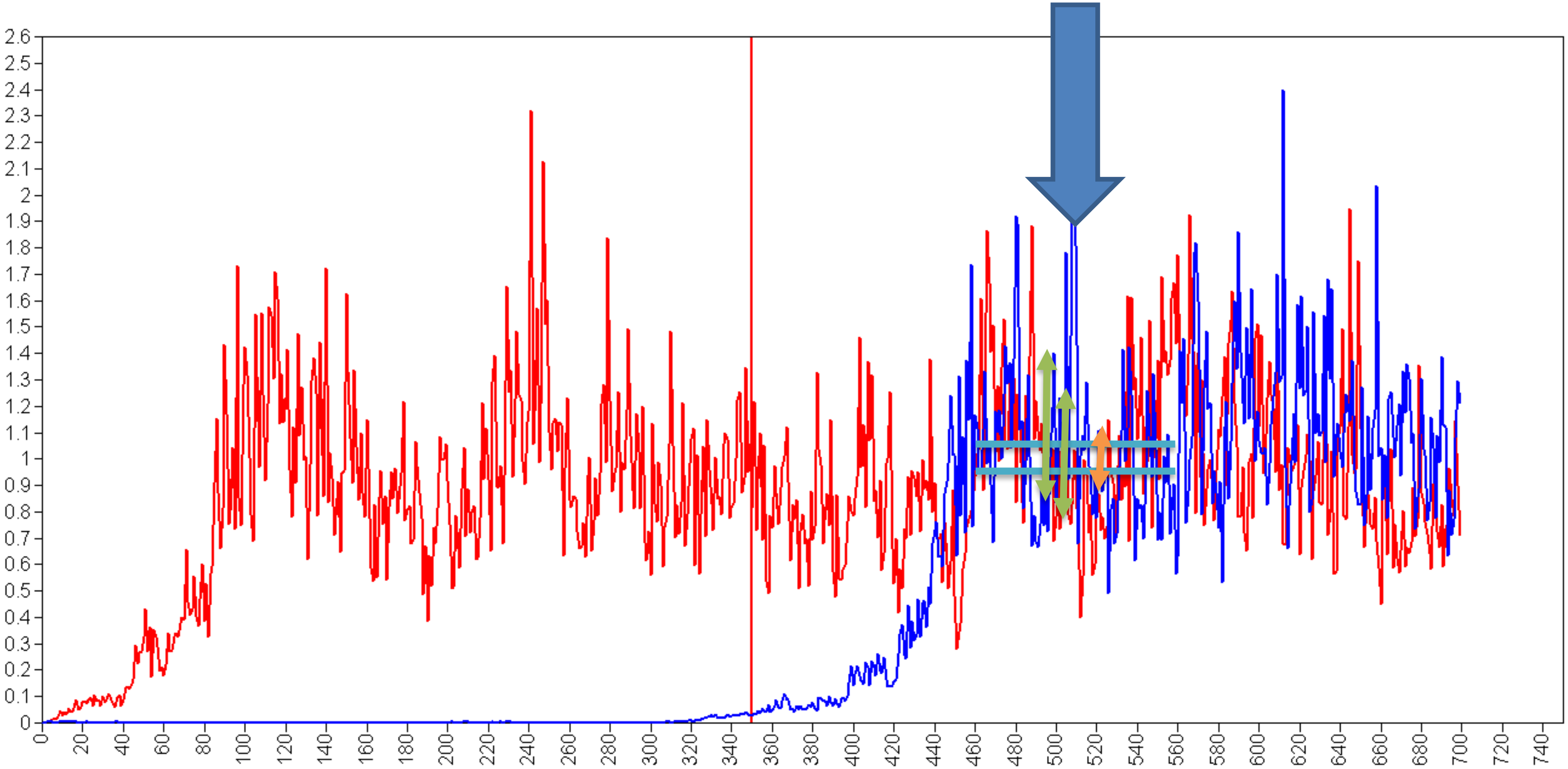
# Gelman and Rubin's convergence diagnostic



# Gelman and Rubin's convergence diagnostic

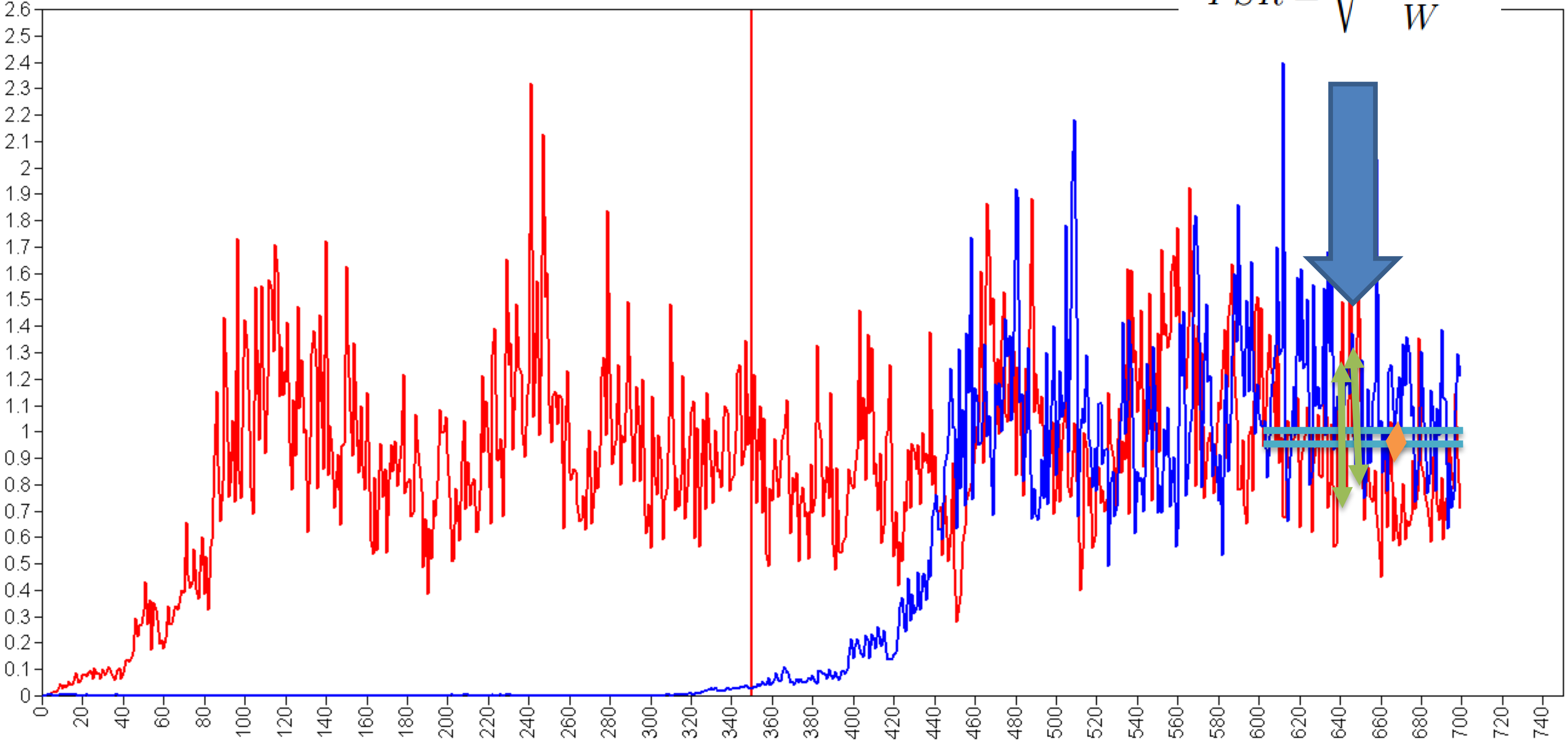


# Gelman and Rubin’s convergence diagnostic

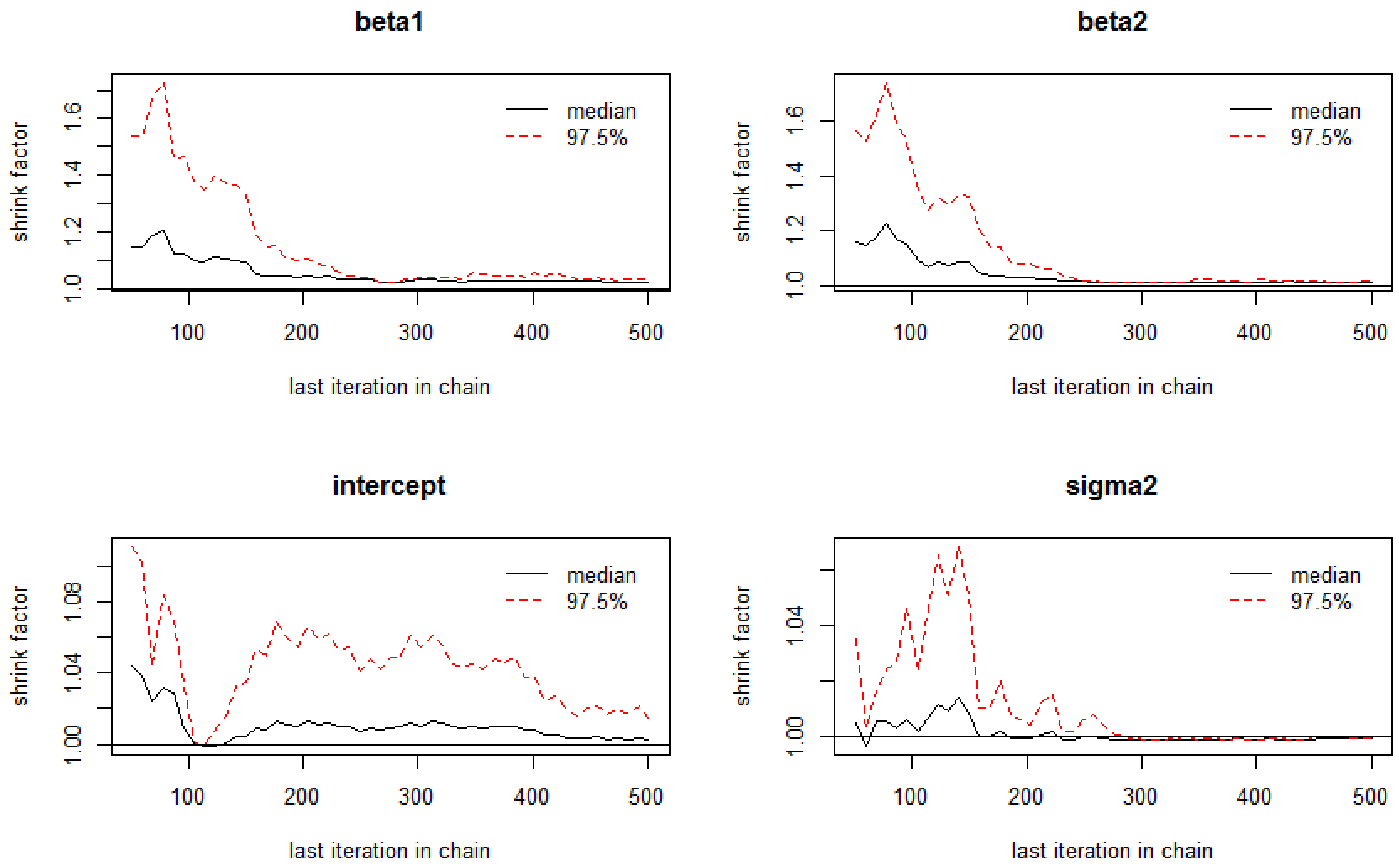


# Gelman and Rubin's convergence diagnostic

$$PSR = \sqrt{\frac{W+B}{W}}$$



# Gelman and Rubin’s convergence diagnostic



# Geweke diagnostic

- Geweke diagnostic (Geweke, 1992).

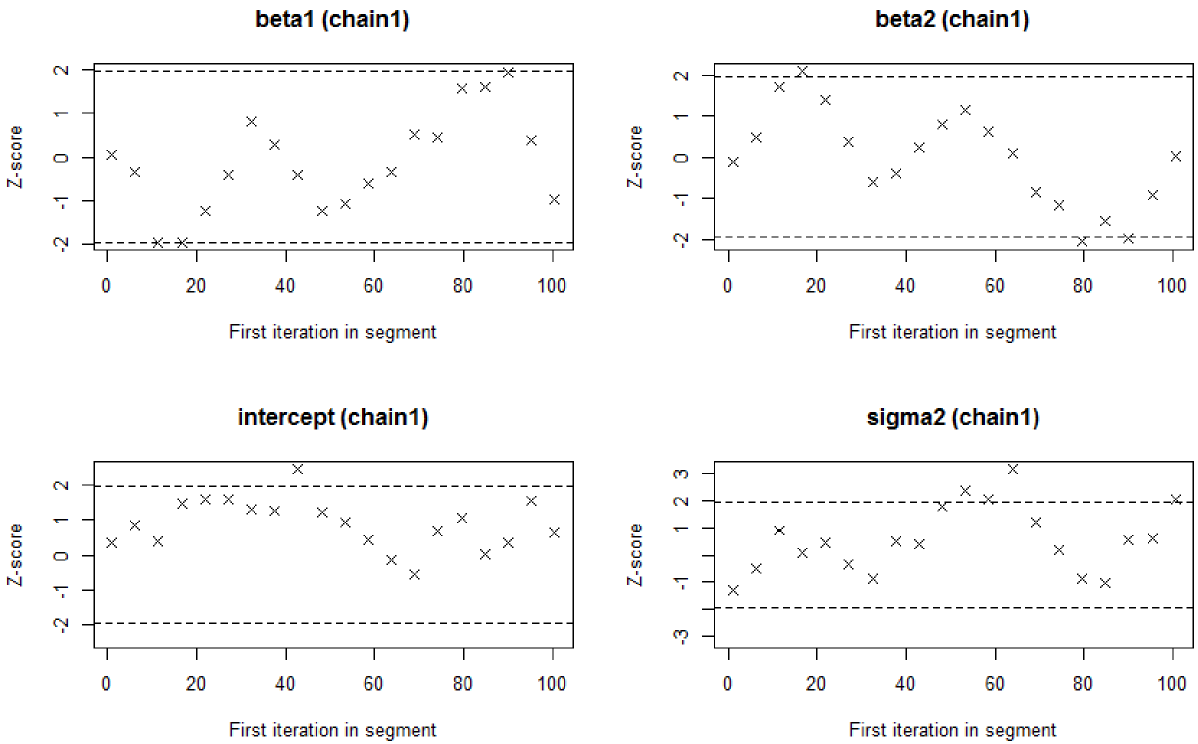
- a test for equality of the means of the first (10%) and last (50%) portions of a chain

- The test statistic is a standard Z-score: the difference between the two sample means divided by its estimated standard error.

- If the *z*-test yields a significant test statistic, then the two portions of the chain significantly differ and full chain convergence was not obtained.

# Geweke diagnostic

If Geweke indicates that the first and last part of a sample from a Markov chain are not drawn from the same distribution, it may be useful to discard the first few iterations to see if the rest of the chain has "converged". This plot shows what happens to Geweke’s Z-score when successively larger numbers of iterations are discarded from the beginning of the chain.



# Relative bias

- bias can also be computed between the converged result obtained for the initial model (Model 1) and the model where the number of iterations was doubled (Model 2);

- Percent Bias = [(initial converged model – model with double iterations)/model with double iterations]*100.

- Bias should be small (note that the relative bias, 10% might be much or nothing depending on the estimate itself)

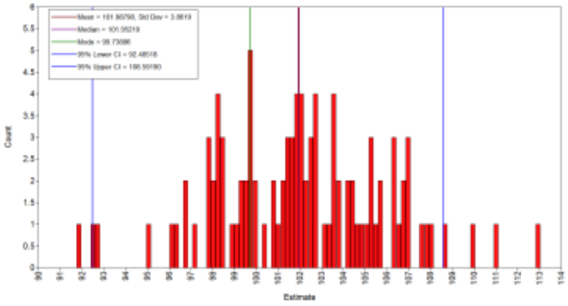
# Relative bias

| **Parameters** | **Bias or Size of Effect** | **Convergence Diagnostic** |
|---|---|---|
| **(i)** | **Bias for Point 3**[a]<br>[(initial converged model – model with double iterations)/model with double iterations]*100 | Geweke *z*-statistic<br>(Significant or not): |
| Y on $X_1$ | [(0.969-0.970)/ 0.970]*100= -0.10 | Non-significant |
| Y on $X_2$ | [(0.650-0.650)/ 0.650]*100= 0.00 | Non-significant |
| Y: Mean | [(0.510-0.511)/ 0.511]*100= -0.19 | Non-significant |
| Y: Residual variance | [(0.953-0.951)/ 0.951]*100= 0.21 | Non-significant |

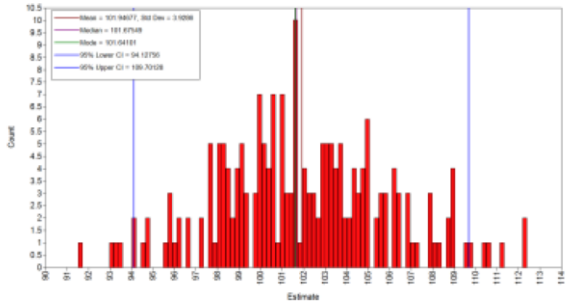
# 4. Does the histogram have enough precision?

- The precision, or smoothness, of the histogram should be checked visually for each model parameter.
- Notice that the plots for our simple example show histograms with no gaps or other abnormalities,

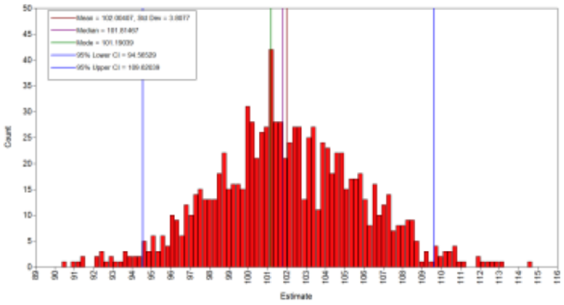
# 4. Does the histogram have enough precision?



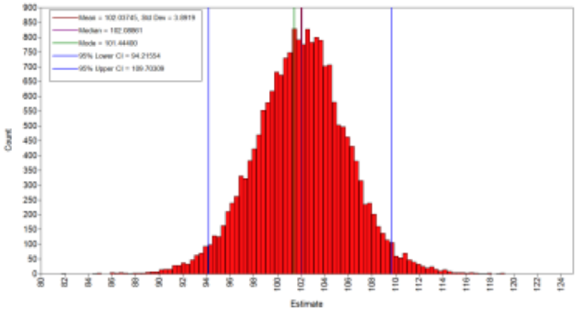
(A)



(B)



(C)



(D)

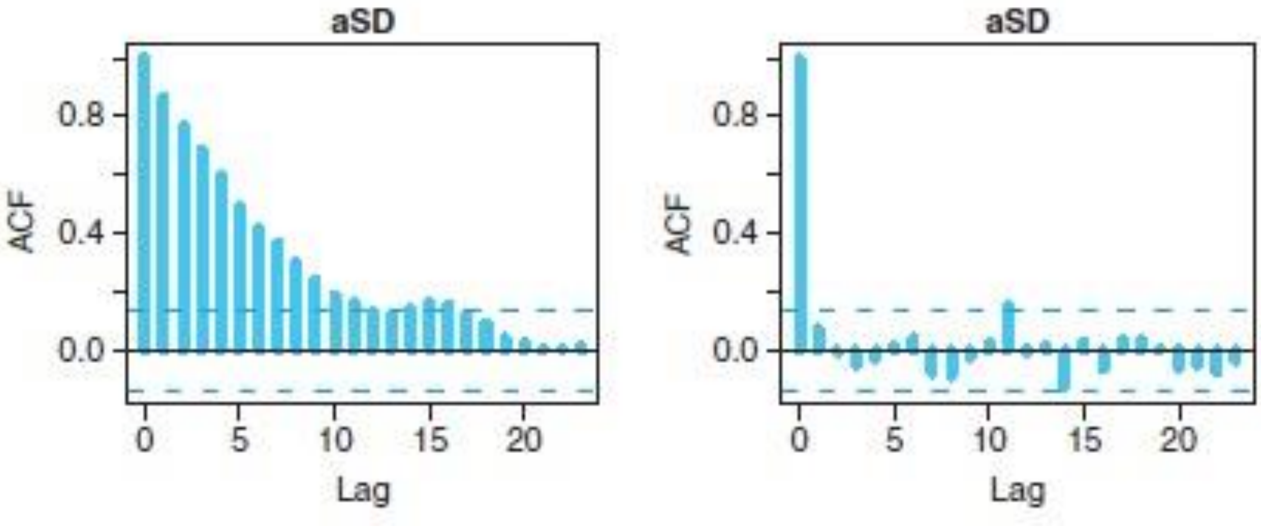
# 5. Do the chains exhibit a strong degree of autocorrelation?

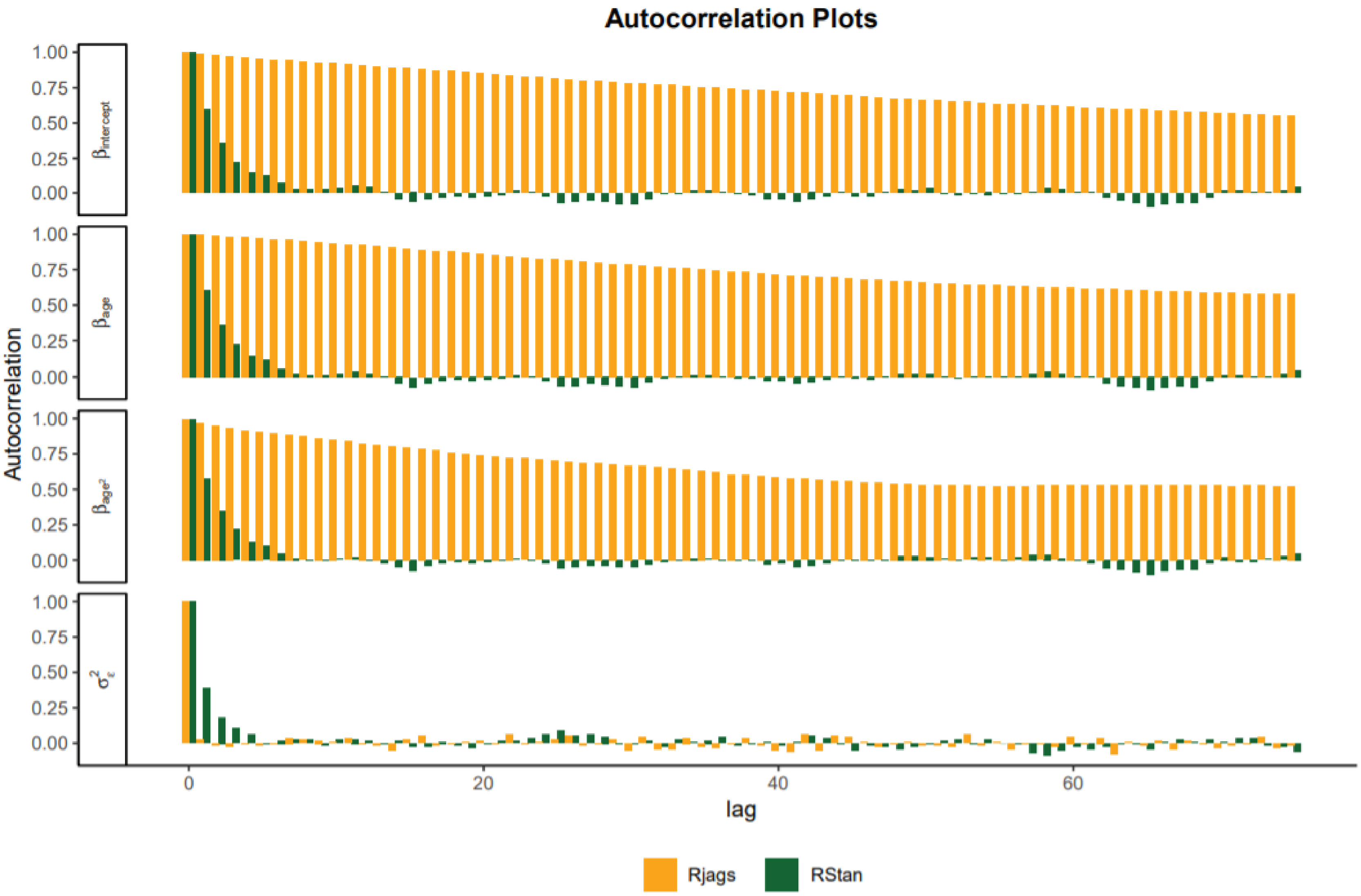
- The very nature of a Bayesian Markov chain is that the iterations in the chain are dependent on one another.

- For example, if iteration t of a Markov chain produces an estimate of .34 for a regression coefficient, then iteration t+1 will produce an estimate correlated with the previous one.

- This dependency is captured by the amount of autocorrelation present in a chain.

# 5. Do the chains exhibit a strong degree of autocorrelation?




Autocorrelation Plots
Autocorrelation
$\beta_{intercept}$
$\beta_{age}$
$\beta_{age^2}$
$\sigma^2_\varepsilon$
1.00
0.75
0.50
0.25
0.00
0
20
40
60
lag
Rjags
RStan

# Thinning might help… or not

Take the estimate of every n$^{th}$ iteration where n>1

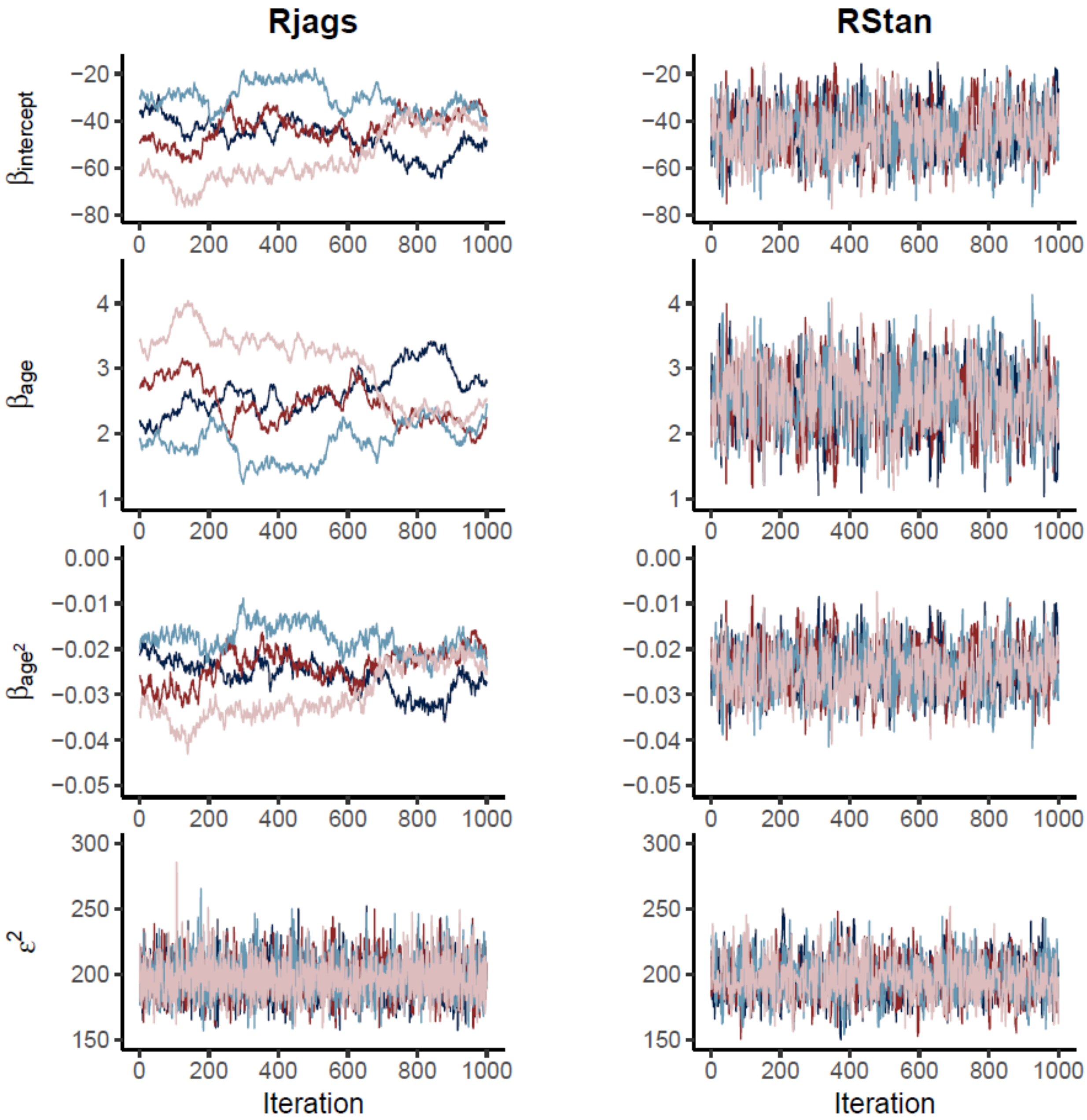
“Thinning merely produces correct results less efficiently (on average) than using the full chain from which the thinned chain was extracted.”
*Link, W. A. & Eaton, M. J. (2011) On thinning of chains in MCMC. Methods in Ecology and Evolution. doi: 10.1111/j.2041-210X.2011.00131.x*

“Perhaps if you're tempted to thin by *n* to reduce autocorrelation, just use a chain *n* times as long without thinning.”
*http://doingbayesiandataanalysis.blogspot.nl/2011/11/thinning-to-reduce-autocorrelation.html*

# Effective sample size

- indication of the efficiency of the algorithm.
- roughly expresses how many independent sampled parameter values contain the same information as the autocorrelated MCMC samples;
- it is the effective length of the MCMC chain
- a small effective sample size could point towards potential problems in the model estimation
  - Example + debugging: Veen, D., & Egberts, M. (2020). The Importance of Collaboration in Bayesian Analyses with Small Samples. In *Small Sample Size Solutions* (pp. 50-70). Routledge

- Effective sample size is also useful for diagnosing the sampling efficiency for a large number of variables.

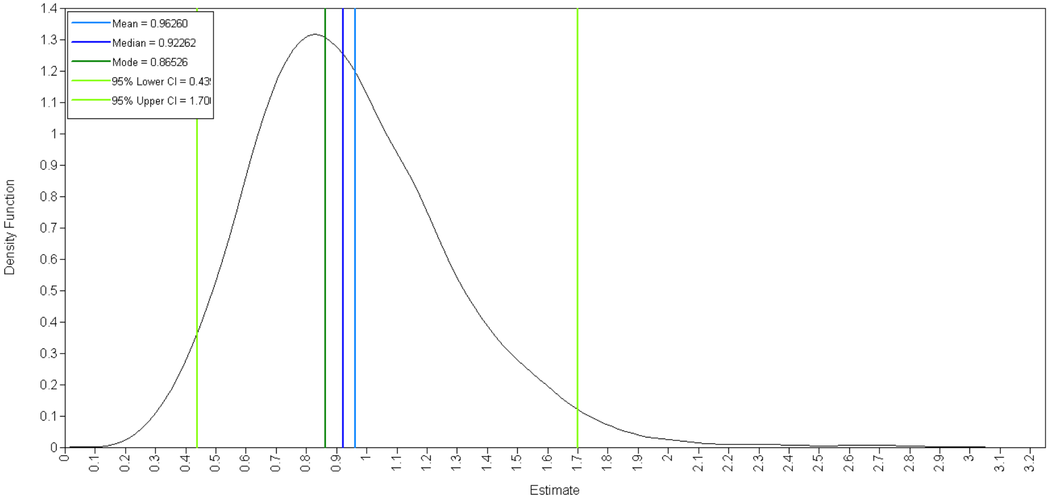

Rjags
RStan
$\beta_{intercept}$
−20
−40
−60
−80
$\beta_{age}$
4
3
2
1
$\beta_{age^2}$
0.00
−0.01
−0.02
−0.03
−0.04
−0.05
$\varepsilon^2$
300
250
200
150
0
200
400
600
800
1000
Iteration
Iteration

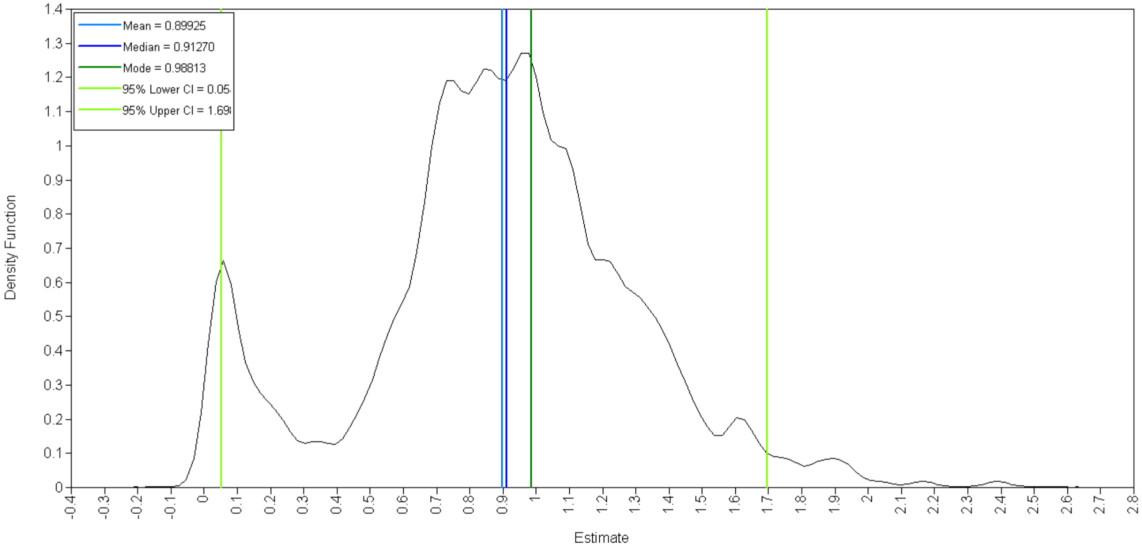
# 6. Does the posterior distribution make substantive sense?

- Substantive abnormalities in the posterior distribution should be examined (e.g., through Kernel density plots).

- The main things that should be checked in a posterior distribution are that it:
  - is smooth,
  - makes substantive sense,
  - does not have a posterior standard deviation that is greater than the scale of the original parameter,
  - does not have a range of the posterior credibility interval greater than the underlying scale of the original parameter,
  - and does not show great fluctuations in the variance of the posterior.

# 6. Does the posterior distribution make substantive sense?





# Stage 3:

# Understanding the Exact Influence of the Priors

# Stage 3:

# Understanding the Exact

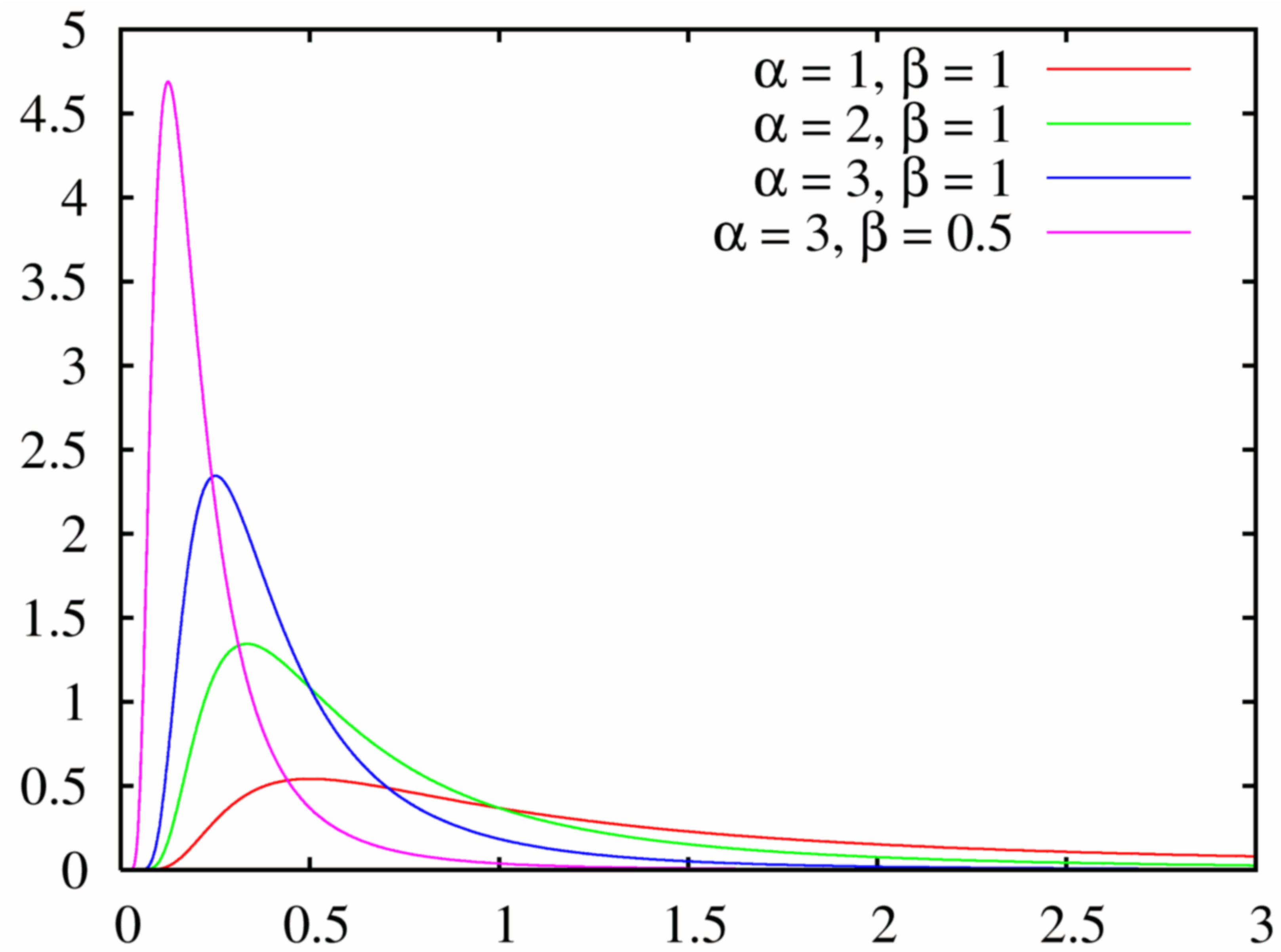
# In

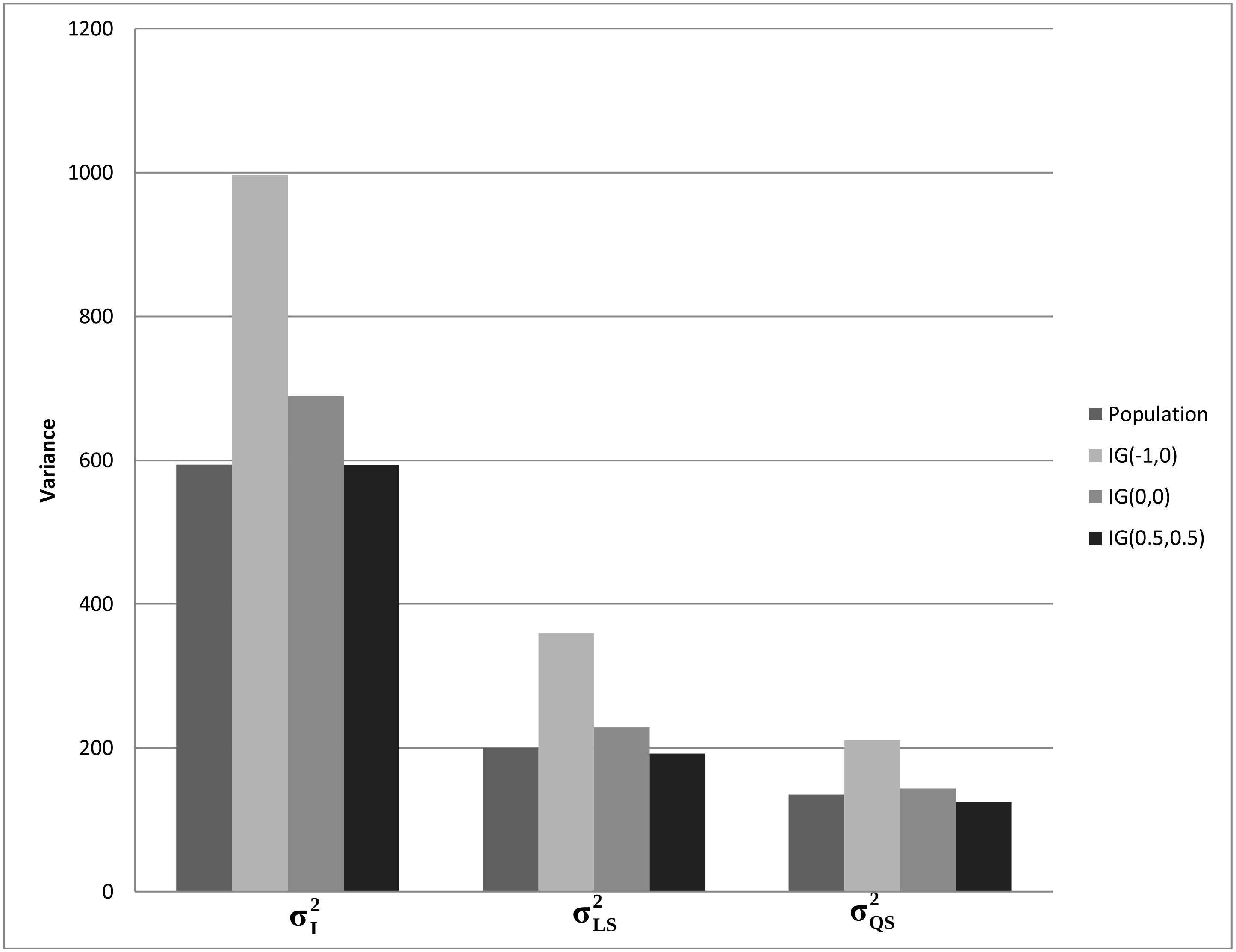
**Warning:** Do not change the priors determined in Point 1 above since these priors represent the current state of affairs

# 7. Do different specifications of the multivariate variance priors influence the results?

- Not so easy, but can have a huge impact on the results

Prior is Inverse Gamma
$\alpha$ (shape), $\beta$ (scale)





*Van de Schoot, Broere, Perryck,Zondervan-Zwijnenburg, & Van Loey, (2015). Analyzing Small Data Sets using Bayesian Estimation: The case of posttraumatic stress symptoms following mechanical ventilation in burn survivors. European Journal of Psychotraumatology, 6: 25216*

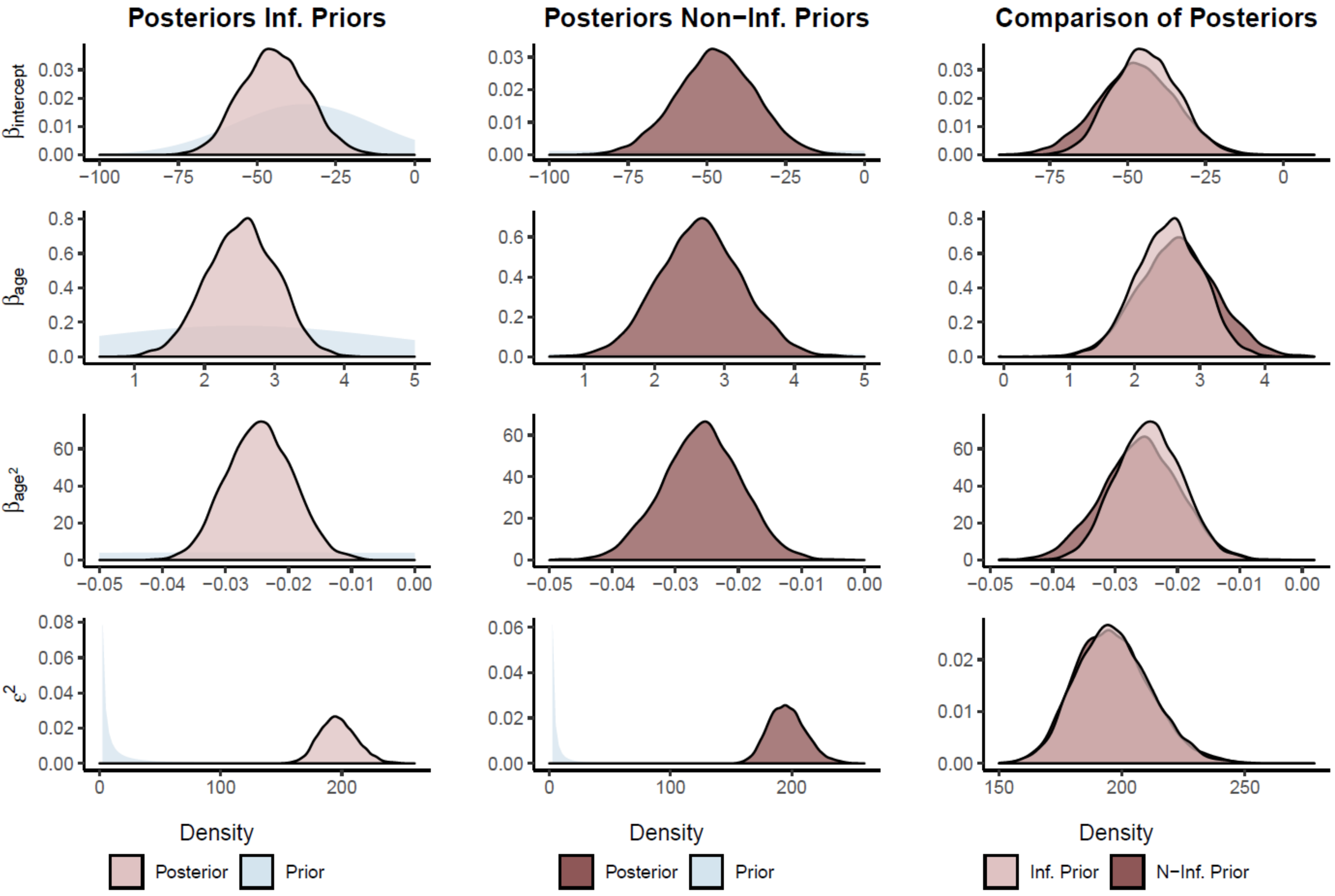
# 7. Do different specifications of the multivariate variance priors influence the results?

- Effect of the prior = [(initial prior specification – subsequent prior specification)/subsequent prior specification]*100.

| | **Size of the effect for Point 7**<br>[(initial priors – default/non-informative priors)/ default/non-informative priors]*100 |
|---|---|
| Y on $X_1$ | [(0.969-0.969)/ 0.969]*100= 0.00 |
| Y on $X_2$ | [(0.650-0.650)/ 0.650]*100= 0.00 |
| Y: Mean | [(0.510-0.510)/ 0.510]*100= 0.00 |
| Y: Residual variance | [(0.953-0.949)/ 0.949]*100= 0.42 |

# 8. Is there a notable effect of the prior when compared with non-informative priors?

- Compare your priors against non-informative priors



Posteriors Inf. Priors
Posteriors Non−Inf. Priors
Comparison of Posteriors
$\beta_{intercept}$
$\beta_{age}$
$\beta_{age^2}$
$\epsilon^2$
0.00
0.01
0.02
0.03
−100
−75
−50
−25
0
0.0
0.2
0.4
0.6
0.8
1
2
3
4
5
0
20
40
60
−0.05
−0.04
−0.03
−0.02
−0.01
0.00
0.02
0.04
0.06
0.08
100
200
150
250
Density
Posterior
Prior
Inf. Prior
N−Inf. Prior

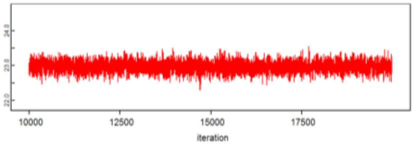
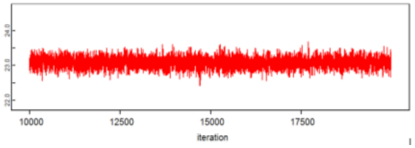
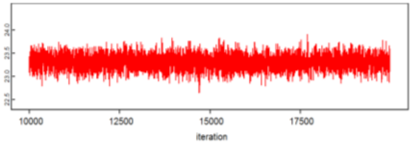
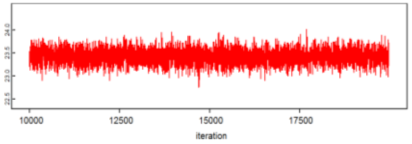
# 9. Are the results stable from a sensitivity analysis?

- Perform a robustness check to understand the impact of specifying different levels of the subjective priors.

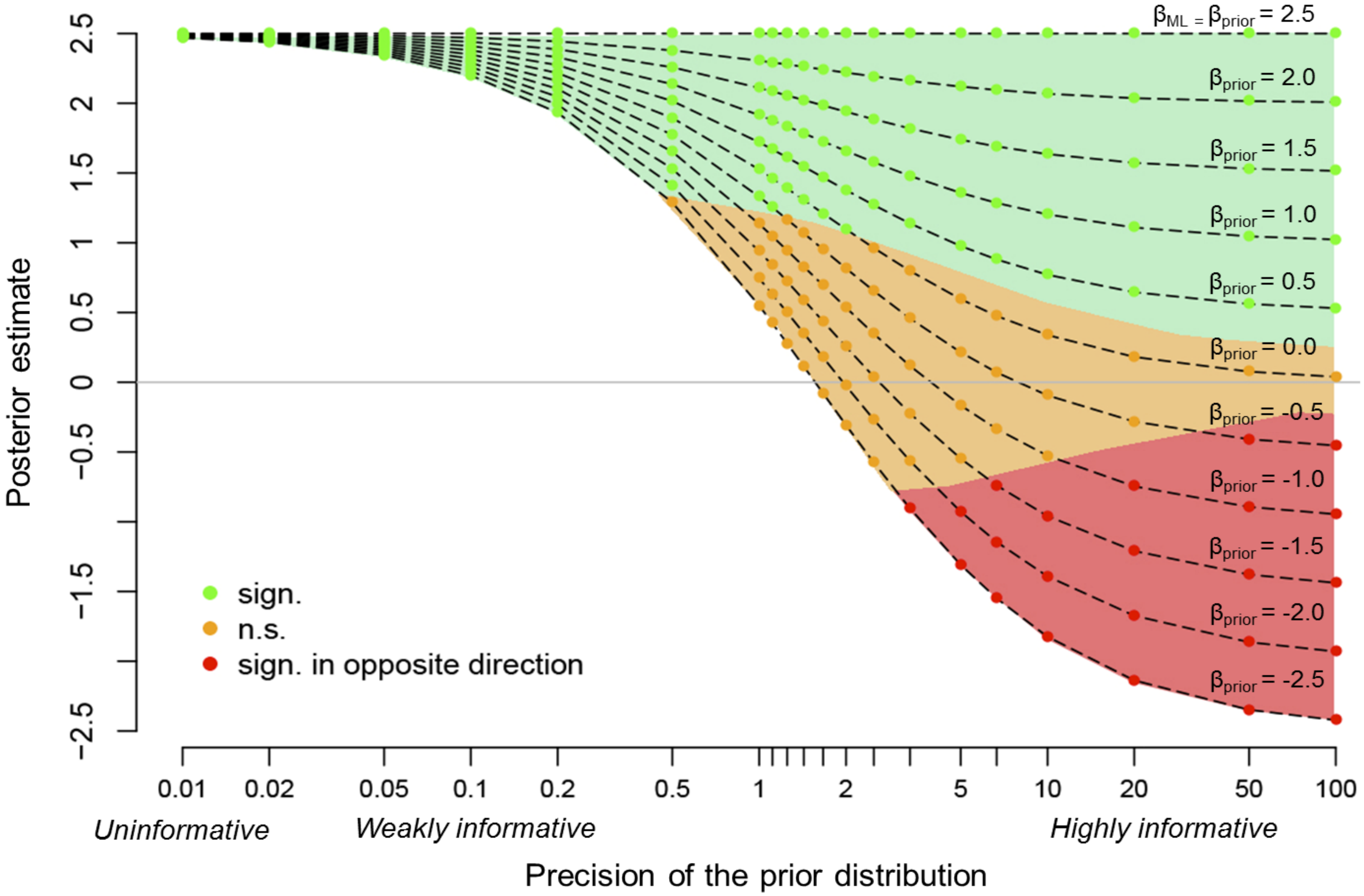
- A sensitivity analysis for priors would entail adjusting hyperparameters upward and downward and re-estimating the model with these varied priors.

- Several different hyperparameter specifications can be made in a sensitivity analysis, and results obtained will point toward the impact of small fluctuations in hyperparameter values.

# 9. Are the results stable from a sensitivity analysis?

| Chain Comparison | Intercept Estimate (SD) | Trace Plot | PSRF | Size of Effect (Percent Bias)[a] |
|---|---|---|---|---|
| Point 9: Sensitivity Analysis for Subjective Prior—Altering the Mean Hyperparameter (alter hyperparameters upward and downward) | | | | |
| Compared to: N(21.37, 1) | 22.97(0.149) |  | 1.645 | 0.948% |
| Compared to: N(26.37, 1) | 23.08(0.149) |  | 1.194 | 0.474% |
| Compared to: N(36.37, 1) | 23.31(0.150) |  | 1.194 | -0.517% |
| Compared to: N(41.37, 1) | 23.42(0.150) |  | 1.646 | -0.992% |

# 9. Are the results stable from a sensitivity analysis?



# Stage 4:

# After the Interpretation of the Model Results

# 10. Is the Bayesian way of interpreting and reporting model results used?

- the Bayesian framework no longer deals in terms of point estimates compared to frequentist approaches.

  => each parameter is estimated with a density capturing uncertainty in the true value.

  => summarize the posterior density with the mean, median, or mode

- Bayesian credibility intervals instead of Cis.
  - For example, a 95% frequentist confidence interval of [0.05, 1.12] for a regression coefficient would indicate that over long-run frequencies, 95% of the confidence intervals constructed in this manner (e.g., with the same sample size, etc.) would contain the true population value.
  - In contrast, the 95% Bayesian credibility interval of [0.05, 1.12] would be interpreted such that there is a .95 probability of the population regression coefficient falling between 0.05 and 1.12.

# 10. Is the Bayesian way of interpreting and reporting model results used?

- The statistical program used for analysis is an important detail to include since different methods (called *sampling methods*) are implemented + version

- A discussion of the priors needs to be in place. The researcher should thoroughly detail and justify all prior distributions that were implemented in the model

- A discussion of chain convergence must be included. Each model parameter estimated should be monitored to ensure that convergence was established for the posterior.

- Results of sensitivity analysis using different forms and levels of informativeness for the priors implemented.

- Basically everything needed to replicate the results (seed values, number of chains, number of iterations, etc)

- Model fit (DIC, ppp-values)

**WAMBS-v2, an updated version of the WAMBS-checklist (https://www.nature.com/articles/s43586-020-00001-2 ).**

1. Ensure the prior distributions and the model or likelihood are well understood and described in detail in the text. Prior-predictive checking can help identify any prior–data conflict.

2. Assess each parameter for convergence, using multiple convergence diagnostics if possible. This may involve examining trace plots or ensuring diagnostics (R^ statistic or effective sample size) are being met for each parameter.

3. Sometimes convergence diagnostics such as the R^ statistic can fail at detecting non-stationarity within a chain. Use a subsequent measure, such as the split-R^, to detect trends that are missed if parts of a chain are non-stationary but, on average, appear to have reached diagnostic thresholds.

4. Ensure that there were sufficient chain iterations to construct a meaningful posterior distribution. The posterior distribution should consist of enough samples to visually examine the shape, scale and central tendency of the distribution.

5. Examine the effective sample size for all parameters, checking for strong degrees of autocorrelation, which may be a sign of model or prior mis-specification.

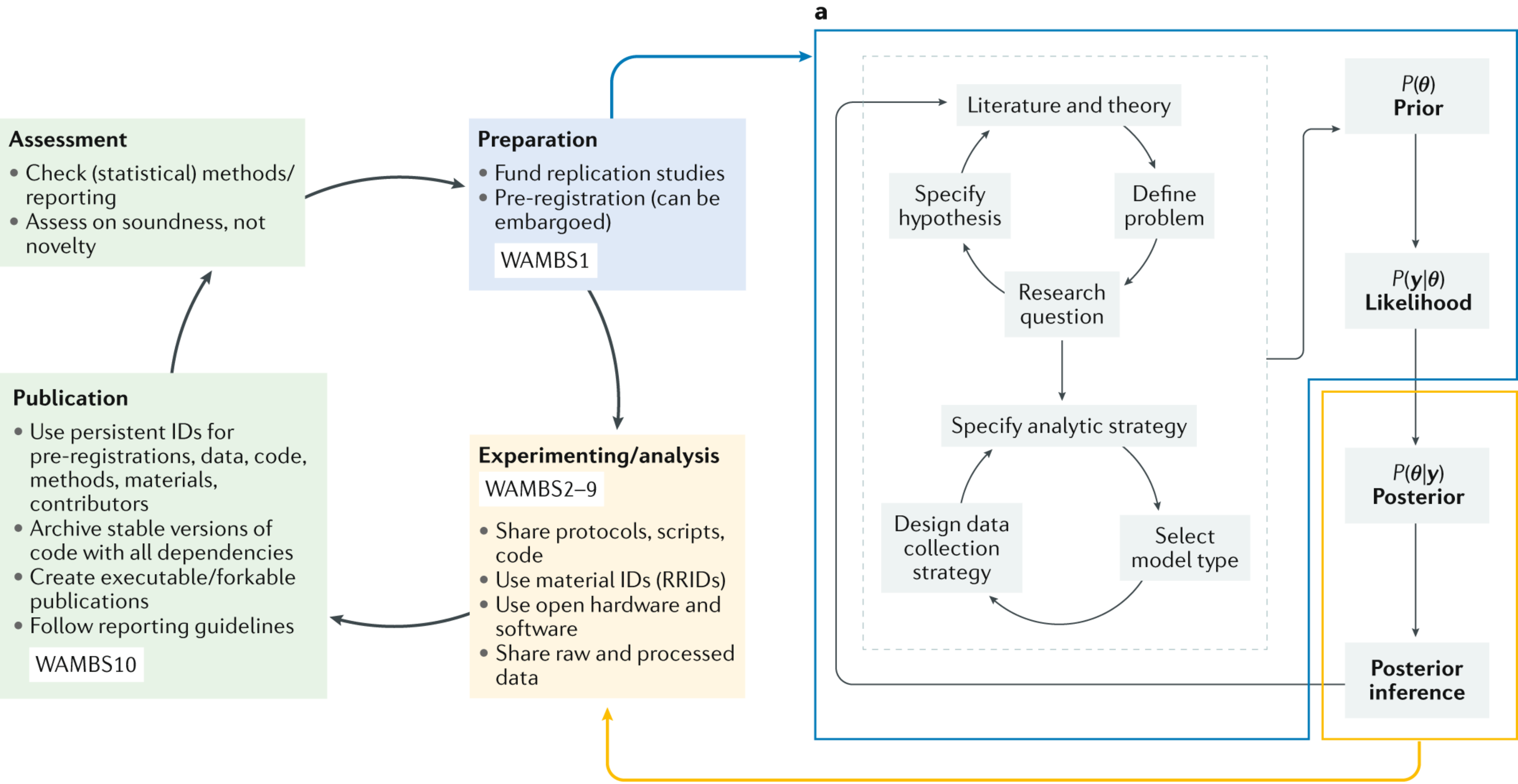
6. Visually examine the marginal posterior distribution for each model parameter to ensure that they do not have irregularities that could have resulted from misfit or non-convergence. Posterior predictive distributions can be used to aid in examining the posteriors.

7. Fully examine multivariate priors through a sensitivity analysis. These priors can be particularly influential on the posterior, even with slight modifications to the hyperparameters.

8. To fully understand the impact of subjective priors, compare the posterior results with an analysis using diffuse priors. This comparison can facilitate a deeper understanding of the impact the subjective priors have on findings. Next, conduct a full sensitivity analysis of all priors to gain a clearer understanding of the robustness of the results to different prior settings.

9. Given the subjectivity of the model, it is also important to conduct a sensitivity analysis of the model (or likelihood) to help uncover how robust results are to deviations in the model.

10. Report findings, including Bayesian interpretations. Take advantage of explaining and capturing the entire posterior rather than simply a point estimate. It may be helpful to examine the density at different quantiles to fully capture and understand the posterior distribution.


a
Assessment
• Check (statistical) methods/ reporting
• Assess on soundness, not novelty
Preparation
• Fund replication studies
• Pre-registration (can be embargoed)
WAMBS1
Publication
• Use persistent IDs for pre-registrations, data, code, methods, materials, contributors
• Archive stable versions of code with all dependencies
• Create executable/forkable publications
• Follow reporting guidelines
WAMBS10
Experimenting/analysis
WAMBS2–9
• Share protocols, scripts, code
• Use material IDs (RRIDs)
• Use open hardware and software
• Share raw and processed data
Literature and theory
Specify hypothesis
Define problem
Research question
Specify analytic strategy
Design data collection strategy
Select model type
$P(\theta)$
Prior
$P(\mathbf{y}|\theta)$
Likelihood
$P(\theta|\mathbf{y})$
Posterior
Posterior inference