Comparing Priors: Sensitivity Analysis

Bayesian Mixed Effects Models with brms for Linguists

Job Schepens

2025-12-17

Table of contents

1 Comparing Priors: Influence on Coefficients and Effect Sizes

How sensitive are your results to prior choice? Validate robustness by fitting models with different plausible priors.

1.1 Why Compare Priors?

Prior sensitivity analysis shows whether your conclusions depend heavily on specific prior choices or whether they're robust across reasonable alternatives. This is especially important for:

- **Publication**: reviewers will ask "how robust is your result?"
- **Model criticism**: if results change dramatically with different priors, something's wrong
- **Theory building**: consistent results across priors = stronger evidence

1.2 Setup

1.3 Create Reaction Time Data

```
Data summary:

Sample size: 3000 observations

Mean log-RT: 6.07

SD log-RT: 0.331
```

1.4 Fit Models with Different Priors

We'll fit three models with different prior specifications:

1. **Domain-informed priors** (our best guess based on RT literature)
2. **Wide priors** (less informative, more uncertainty)
3. **Narrow priors** (more informative, regularizing)

1.4.1 Define Prior Specifications

Table 1: Prior Specifications for Sensitivity Analysis

Parameter	Domain	Wide	Narrow
Intercept	N(6, 1.5)	N(6, 3)	N(6, 0.8)
Slopes (b)	N(0, 0.5)	N(0, 1)	N(0, 0.3)
Residual SD	Exp(1)	Exp(0.5)	Exp(2)
Random Effect SD	Exp(1)	Exp(0.5)	Exp(2)
Correlation	LKJ(2)	LKJ(2)	LKJ(2)

1.4.2 Fit Models

```
Loading domain prior model from cache...

Loading wide prior model from cache...

Loading narrow prior model from cache...
```

```
All models fitted successfully
```

1.5 Compare Posterior Summaries

Table 2: Fixed Effects Posterior Summaries Across Prior Specifications

Prior	Parameter	Mean	Q5	Q95
Domain	b_Intercept	5.991	5.966	6.020
Domain	b_conditionB	0.161	0.139	0.182
Wide	b_Intercept	5.991	5.966	6.015
Wide	b_conditionB	0.161	0.139	0.182
Narrow	b_Intercept	5.990	5.963	6.017
Narrow	b_conditionB	0.161	0.139	0.181

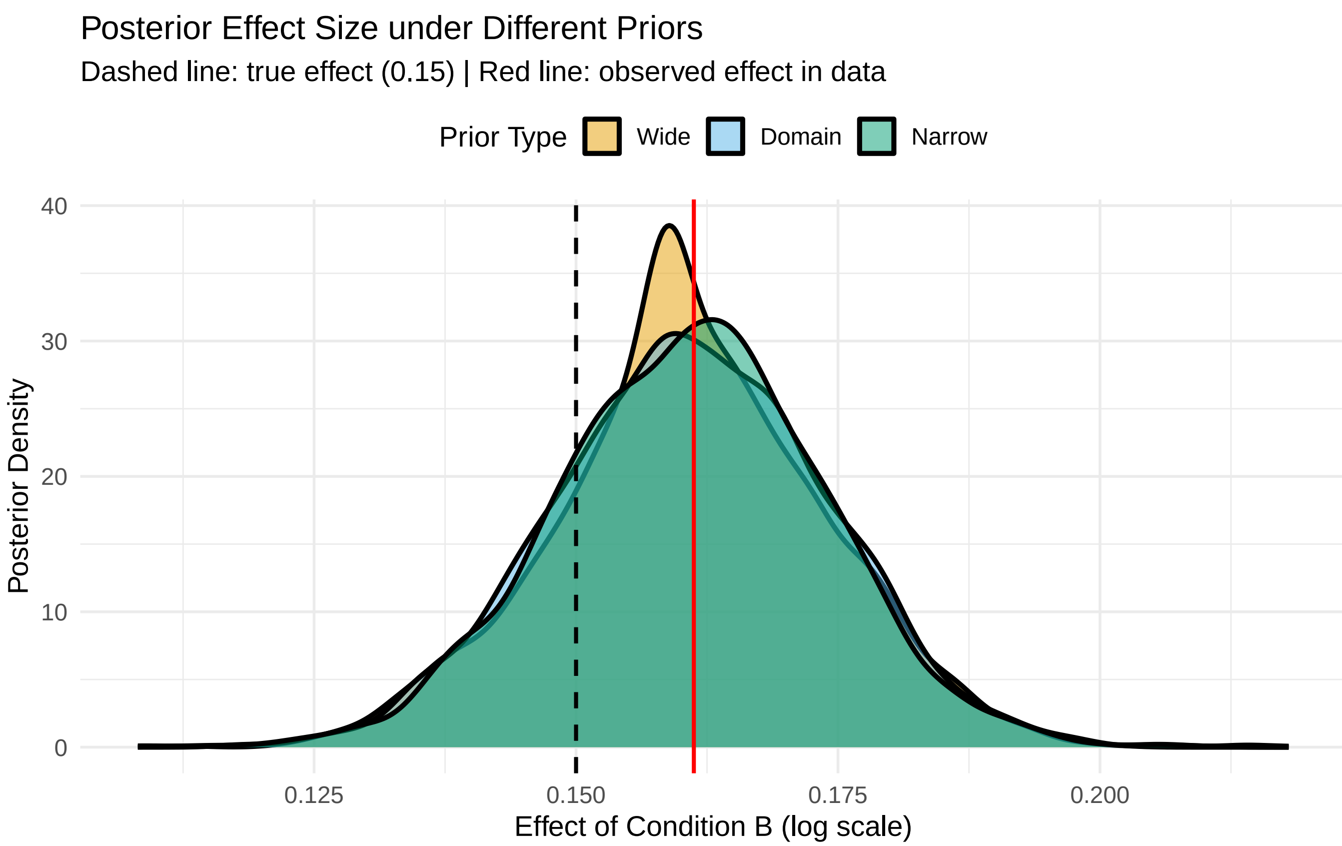
1.5.1 Compact Comparison Table

Table 3: Posterior Summaries: 95% Credible Intervals Across Prior Specifications

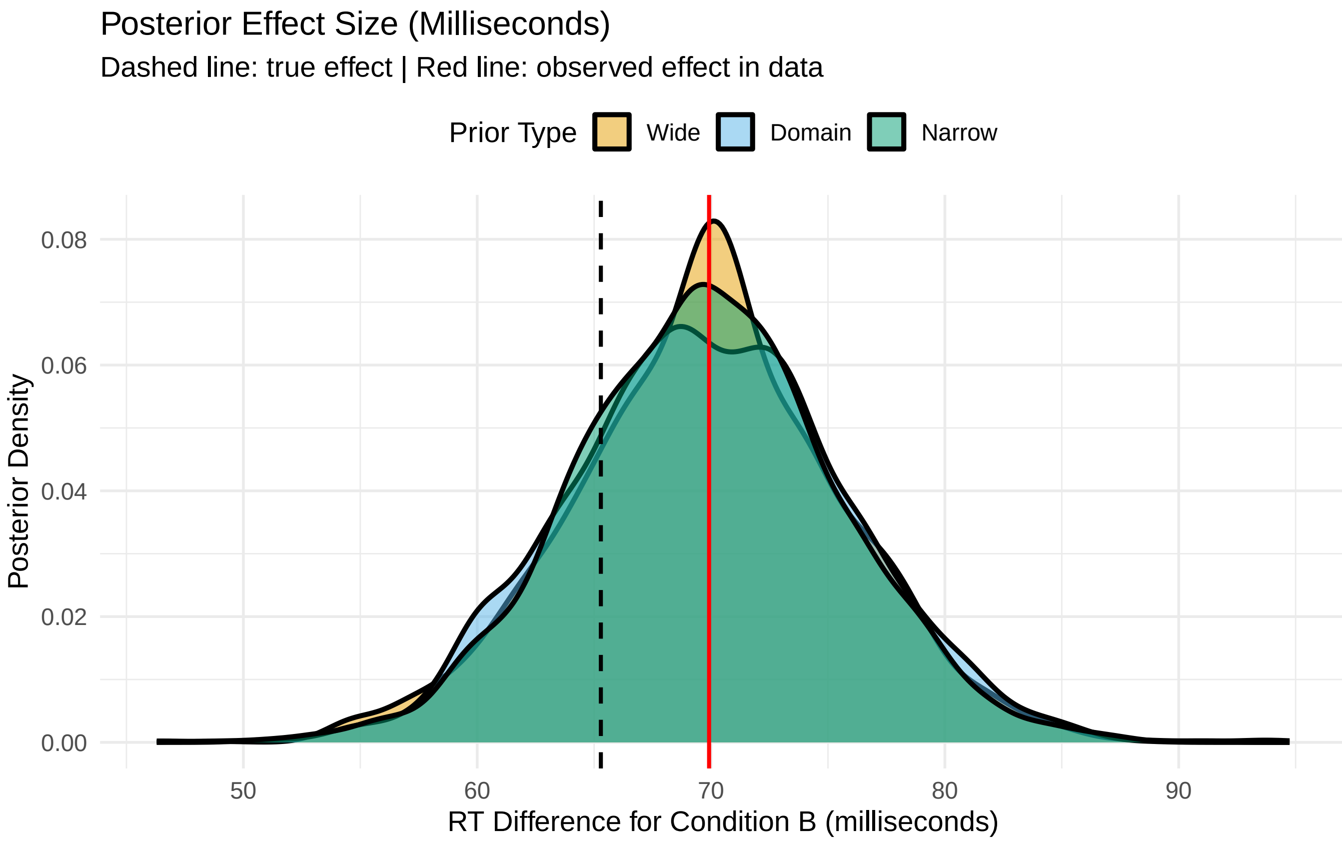
Parameter	Prior	2.5%	Median	97.5%
Intercept	Domain	5.959	5.991	6.031
Intercept	Wide	5.957	5.991	6.021
Intercept	Narrow	5.951	5.991	6.028
Condition B Effect	Domain	0.135	0.161	0.186
Condition B Effect	Wide	0.135	0.160	0.186
Condition B Effect	Narrow	0.136	0.161	0.186
Residual SD	Domain	0.312	0.320	0.329
Residual SD	Wide	0.313	0.320	0.328
Residual SD	Narrow	0.312	0.320	0.329

1.6 Visualize Posterior Comparisons

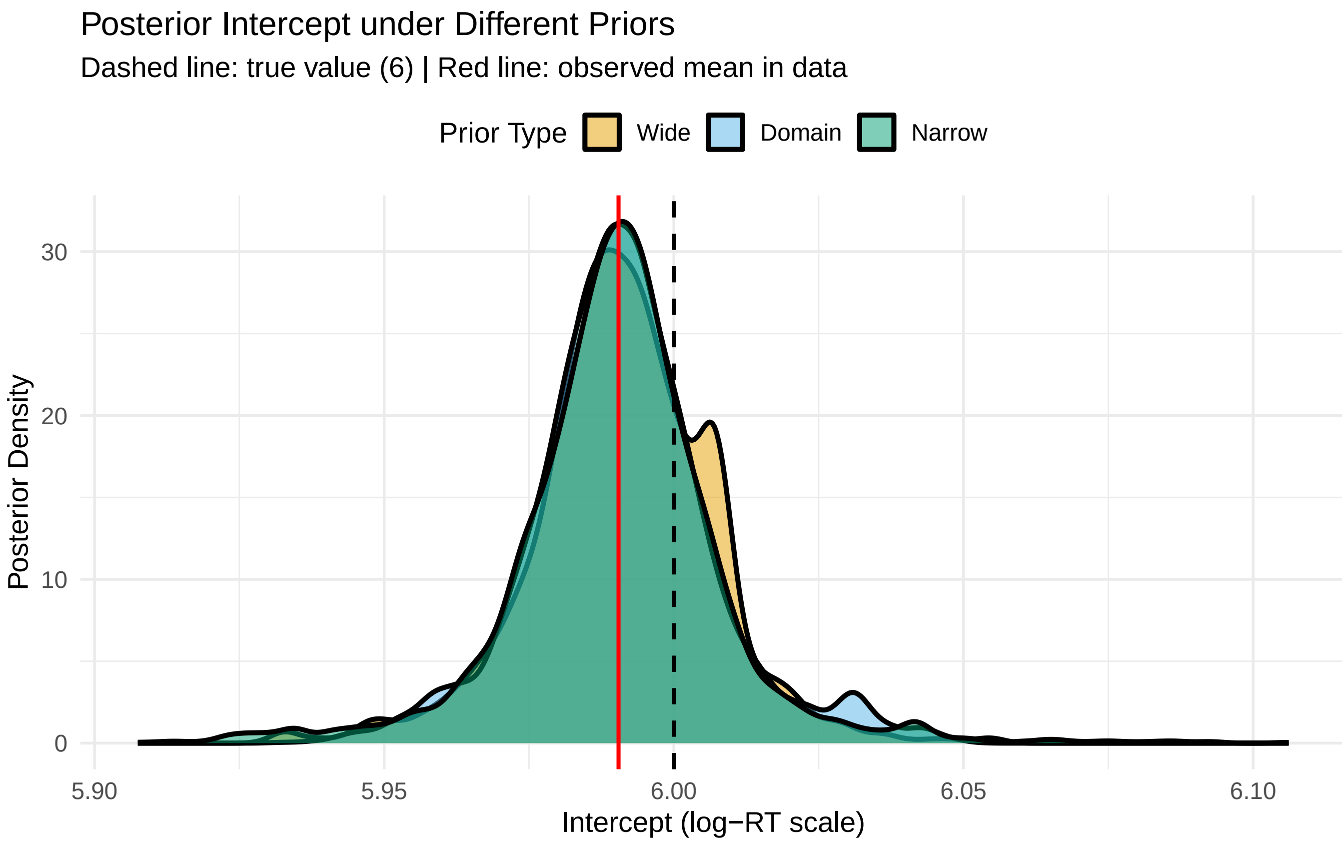
1.6.1 Effect Size Distributions (Log Scale)



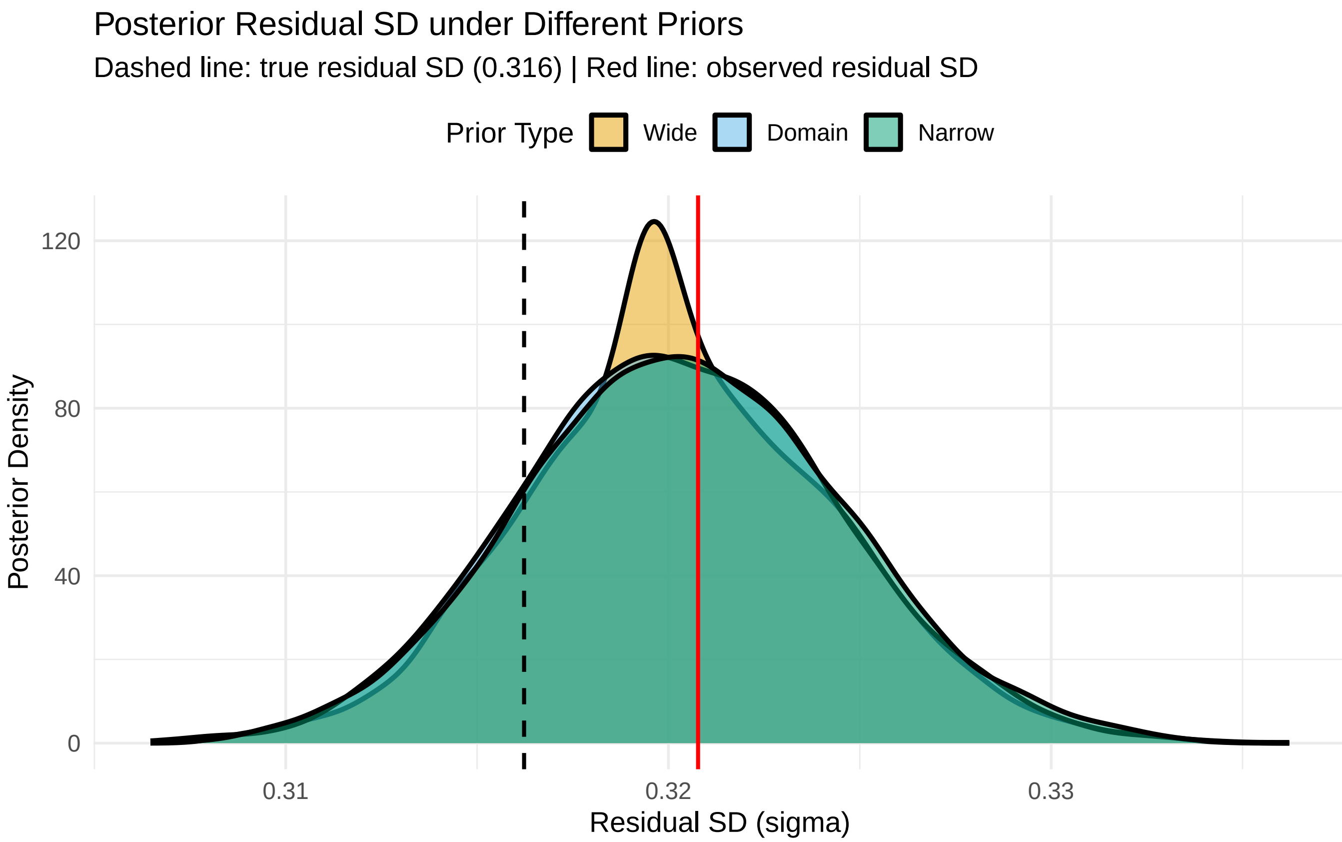
1.6.2 Effect Size in Milliseconds



1.6.3 Intercept Comparisons



1.6.4 Residual SD Comparisons



2 Prior Sensitivity with Limited Data (n = 10, 40, 100)

To see the impact of priors more clearly, let's compare results from three dataset sizes: extremely small (n=10), very small (n=40), and moderate (n=100).

2.1 Create Small Datasets

```
Extremely small dataset (n=10) summary:

Sample size: 10 observations

Mean log-RT: 6.02

SD log-RT: 0.399


Very small dataset (n=40) summary:

Sample size: 40 observations

Mean log-RT: 6.07

SD log-RT: 0.369
```

2.2 Fit Models on Small Datasets

```
 n=10 models fitted successfully
Loading small domain prior model from cache...
Loading small wide prior model from cache...
Loading small narrow prior model from cache...

 n=40 models fitted successfully

 All small dataset models fitted successfully
```

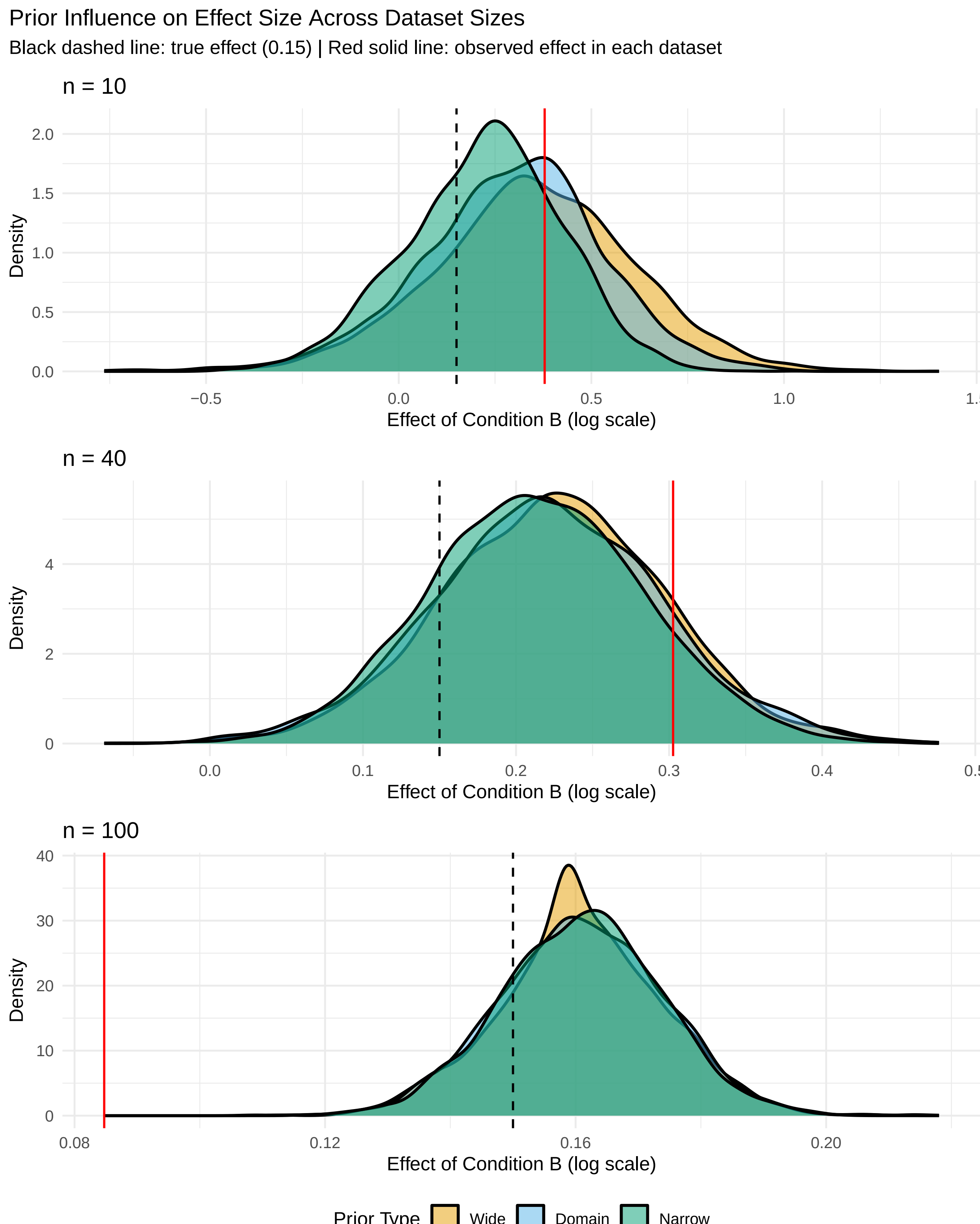
2.3 Compare Across Dataset Sizes

Table 4: Condition B Effect: Comparing Prior Influence with Different Sample Sizes

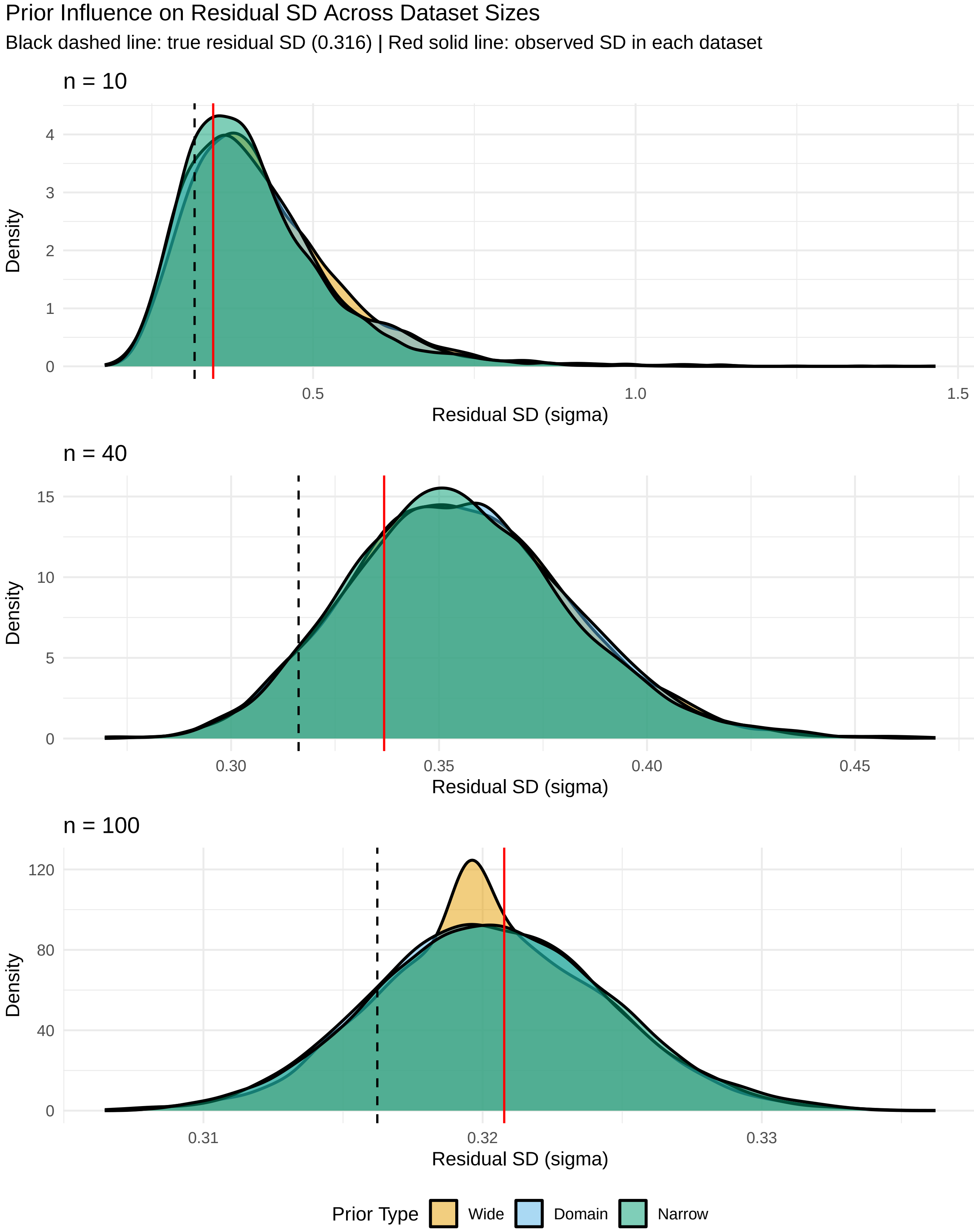
Dataset	Prior	2.5%	Median	97.5%
n=10	Domain	-0.198	0.308	0.751
n=10	Wide	-0.180	0.351	0.850
n=10	Narrow	-0.194	0.234	0.597
n=40	Domain	0.069	0.220	0.372
n=40	Wide	0.080	0.226	0.369
n=40	Narrow	0.079	0.212	0.351
n=100	Domain	0.135	0.161	0.186
n=100	Wide	0.135	0.160	0.186
n=100	Narrow	0.136	0.161	0.186

2.4 Side-by-Side Visualizations

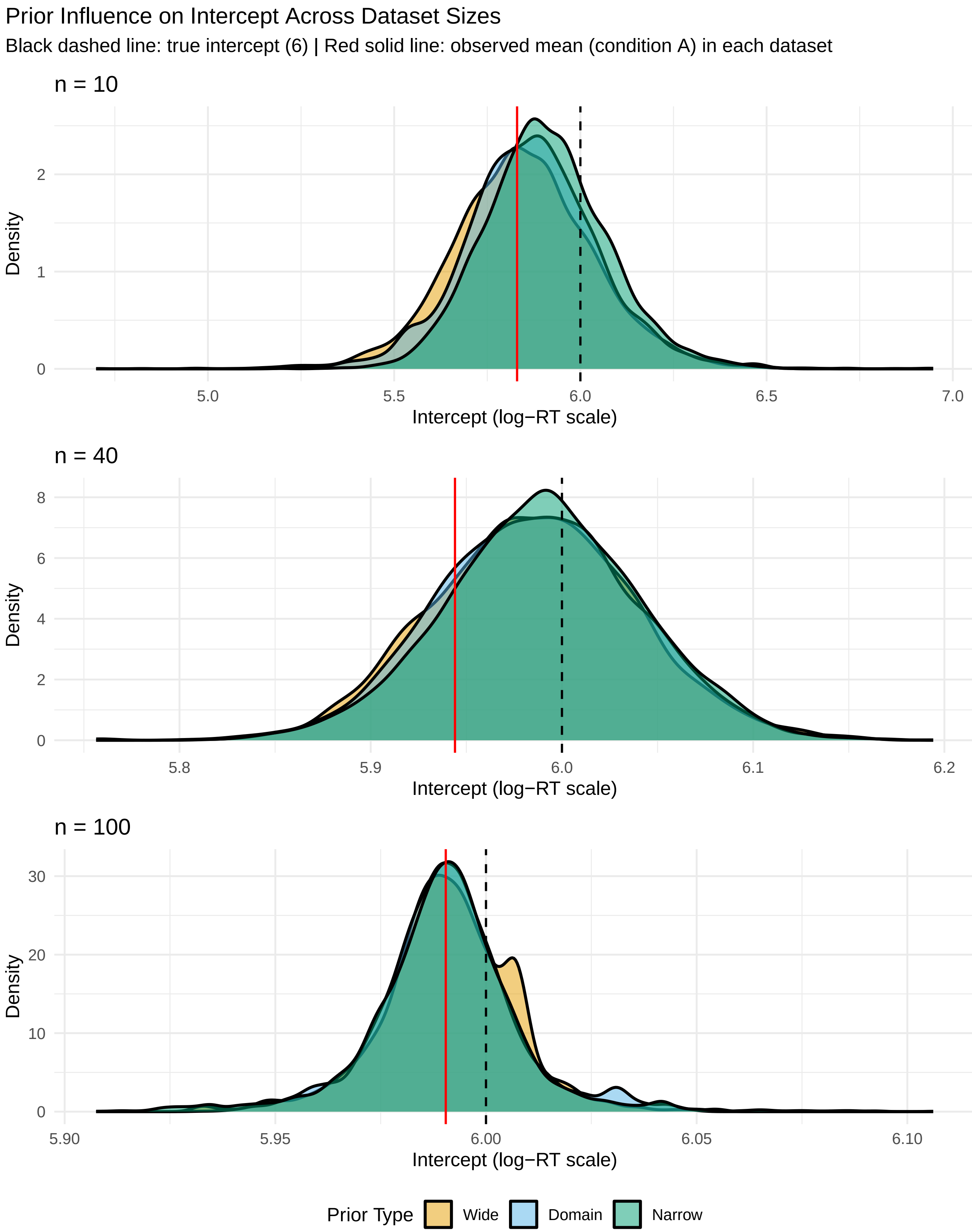
2.4.1 Effect Size Comparison



2.4.2 Residual SD Comparison



2.4.3 Intercept Comparison



2.4.4 Key Observations

With n = 10:

- Posteriors show **dramatic separation** between prior specifications
- **Extremely wide credible intervals** - massive uncertainty
- Prior **dominates** the posterior - data barely influences results
- Narrow prior strongly regularizes toward zero
- Conclusions are **highly sensitive** to prior choice
- Wide prior may produce unrealistic estimates due to limited data

With n = 40:

- Posteriors show **substantial separation** between prior specifications
- **Very wide credible intervals** - high uncertainty
- Priors have **strong influence** but data is starting to matter
- Narrow prior pulls estimates toward zero (regularization visible)
- Risk of conclusions depending heavily on prior choice

With n = 100:

- Posteriors show **moderate overlap** - priors still matter but less
- **Narrower credible intervals** - reduced uncertainty
- Data begins to dominate, but prior influence still visible
- Estimates converging toward similar values

Conclusion: With n < 20, priors dominate and results are highly sensitive. With n = 20-100, prior choice still matters significantly. For robust conclusions regardless of reasonable prior choice, aim for n > 200-300 per group.

2.5 Interpretation Guide

2.5.1 What to Look For

Robust results:

- Posteriors roughly overlap across prior specifications
- Conclusions (e.g., "effect exists" vs. "effect absent") consistent
- Differences are small relative to uncertainty

Fragile results:

- Posteriors diverge substantially
- Conclusions flip depending on prior
- Suggests your data isn't informative enough or model is misspecified

2.5.2 Assess Your Results

```
**Condition B Effect - Overlap Analysis:**
```

Table 5: Posterior Overlap: Percentage of draws from first prior within 95% CI of second

Comparison	Overlap (%)
Domain vs Wide	95.0
Domain vs Narrow	94.6
Wide vs Narrow	94.4

2.6 Common Questions & Answers

2.6.1 "Isn't using domain priors just imposing my beliefs?"

Answer: Yes, exactly. The question is whether your beliefs are *reasonable*. Prior specification is:

- **Data**: "Everyone agrees this is fact"
- **Reasonable prior**: "Domain experts expect this range"
- **Unreasonable prior**: "I want results to look like this"

If experts in linguistics expect RTs of 200-1000ms, that's reasonable. If your prior forces results to match your hypothesis, that's not.

2.6.2 "How different should my alternative priors be?"

Answer: Use the range of *reasonable* specifications:

- **Narrow**: Informed by strong prior knowledge
- **Domain**: Your best guess (typically used for main analysis)
- **Wide**: Vague but still plausible (not completely flat)

Don't use:

- Priors that violate domain knowledge (e.g., negative RTs)
- Priors that are technically possible but implausible

2.6.3 "What if results change with different priors?"

Options:

1. **Collect more data** - let data dominate the prior
2. **Refine your prior** - discuss with domain experts
3. **Simplify the model** - maybe you're overfitting
4. **Report the sensitivity** - honest science: "Results depend on prior choice"

2.6.4 "Should I always compare priors?"

Recommended:

- Always: For main effects you're claiming are "real"
- Always: For publication
- Always: If anyone questions your priors

Optional:

- For exploratory analyses

- For well-established effects with strong data
- For parameters you're not making inferences about

2.7 Session Info

```
R version 4.4.1 (2024-06-14)
Platform: x86_64-pc-linux-gnu
Running under: Ubuntu 22.04.5 LTS

Matrix products: default
BLAS:   /usr/lib/x86_64-linux-gnu/openblas-pthread/libblas.so.3
LAPACK: /usr/lib/x86_64-linux-gnu/openblas-pthread/libopenblasp-r0.3.20.so;  LAPACK version 3.10.0

locale:
 [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C
 [3] LC_TIME=en_US.UTF-8        LC_COLLATE=en_US.UTF-8
 [5] LC_MONETARY=en_US.UTF-8    LC_MESSAGES=en_US.UTF-8
 [7] LC_PAPER=en_US.UTF-8       LC_NAME=C
 [9] LC_ADDRESS=C               LC_TELEPHONE=C
[11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C

time zone: Etc/UTC
tzcode source: system (glibc)

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
 [1] rstan_2.32.7         StanHeaders_2.32.10  patchwork_1.3.2
 [4] posterior_1.6.1.9000 bayesplot_1.14.0     lubridate_1.9.3
 [7] forcats_1.0.0        stringr_1.5.1        dplyr_1.1.4
[10] purrr_1.0.2          readr_2.1.5          tidyr_1.3.1
[13] tibble_3.2.1         ggplot2_4.0.0        tidyverse_2.0.0
[16] brms_2.23.0          Rcpp_1.0.13

loaded via a namespace (and not attached):
 [1] gtable_0.3.6          tensorA_0.36.2.1      QuickJSR_1.8.1
 [4] xfun_0.54             processx_3.8.4        inline_0.3.21
 [7] lattice_0.22-6        tzdb_0.4.0            vctrs_0.6.5
[10] tools_4.4.1           ps_1.8.1              generics_0.1.3
[13] stats4_4.4.1          parallel_4.4.1        fansi_1.0.6
[16] cmdstanr_0.9.0        pkgconfig_2.0.3       Matrix_1.7-0
[19] checkmate_2.3.3       RColorBrewer_1.1-3    S7_0.2.0
[22] distributional_0.5.0  RcppParallel_5.1.11-1 lifecycle_1.0.4
[25] compiler_4.4.1        farver_2.1.2          Brobdingnag_1.2-9
[28] tinytex_0.53          codetools_0.2-20      htmltools_0.5.8.1
[31] yaml_2.3.10           pillar_1.9.0          bridgesampling_1.1-2
[34] abind_1.4-8           nlme_3.1-164          tidyselect_1.2.1
```

```
[37] digest_0.6.37        mvtnorm_1.3-3        stringi_1.8.4
[40] labeling_0.4.3       fastmap_1.2.0        grid_4.4.1
[43] cli_3.6.5            magrittr_2.0.3       loo_2.8.0
[46] pkgbuild_1.4.8       utf8_1.2.4           withr_3.0.2
[49] scales_1.4.0         backports_1.5.0      timechange_0.3.0
[52] estimability_1.5.1   rmarkdown_2.30       matrixStats_1.5.0
[55] emmeans_2.0.0        gridExtra_2.3        hms_1.1.3
[58] coda_0.19-4.1        evaluate_1.0.1       knitr_1.50
[61] rstantools_2.5.0     rlang_1.1.6          xtable_1.8-4
[64] glue_1.8.0           jsonlite_1.8.9       R6_2.5.1
```