

# Sphincter analysis

Teddy Groves

## Table of contents

<b>Introduction</b>	<b>1</b>
<b>Overall strategy</b>	<b>2</b>
Features . . . . .	2
Data processing . . . . .	2
Modelling approach . . . . .	3
Software . . . . .	4
Validation of statistical computation . . . . .	4
Reproducibility . . . . .	5
<b>Main findings</b>	<b>5</b>
Whisker stimulation . . . . .	5
Pulsatility . . . . .	7
Red blood cell flow . . . . .	8
Hypertensive challenge . . . . .	10
<b>Details of the whisker stimulation analysis</b>	<b>10</b>
Dependent variable . . . . .	10
Statistical Models . . . . .	11
Results . . . . .	13
<b>Details of the pulsatility analysis</b>	<b>15</b>
Dependent variable . . . . .	15
Questions . . . . .	16
Description of the dataset . . . . .	16
Statistical models . . . . .	16
Results . . . . .	21
Answers to specific questions . . . . .	22

<b>Details of the red blood cell flow analysis</b>	<b>26</b>
Data processing . . . . .	26
Dependent variable . . . . .	27
Statistical models . . . . .	28
Results . . . . .	29
Shared parameters . . . . .	29
Interactions . . . . .	29
<b>Hypertensive challenge</b>	<b>33</b>
Dependent variable . . . . .	33
Statistical model . . . . .	34
Results . . . . .	35

## Introduction

This project aimed to model brain blood vessel measurements of mice during a treatment protocol designed to elicit stress responses.

Measurements included:

- Whisker stimulation response (vessel diameter before and after stimulation)
- Vessel centre and diameter pulsatility
- Red blood cell flow (i.e. speed and flux)
- Hypertensive challenge response (Correlation between blood pressure and diameter under different pressure conditions)

We believed that the mechanisms underlying each kind of measurement were independent, and moreover each measurement required different data filtering choices. We therefore conducted a separate analysis for each measurement type.

Our overall modelling approach broadly followed the recommendations in Gelman et al. (2020). For each analysis, we first constructed a simple mathematical model of the data generating process, then iteratively improved it, taking into account estimated predictive performance in and out of sample, simplicity, interpretability and computational feasibility.

We decided to employ a Bayesian modelling approach primarily because of the availability of non-experimental information, particularly structural information concerning the data making it important to take into partially pool information between categories given the relatively small number of measurements. In a Bayesian context partial pooling can be achieved using a prior distribution on the random effects.

## Overall strategy

Although our project involved several statistical analyses, we used a similar overall strategy in each case. This section describes the aspects of this strategy that were common to all of our analyses.

### Features

All of our analyses involved a common data structure, with real-valued measurements each with multiple categorical features, namely

- age of the measured mouse (adult or young)
- identity of the measured mouse
- stage of the treatment protocol when measured
- measured vessel type (penetrating artery, sphincter, bulb, first order capillary, etc)

### Data processing

We ignored data from one mouse (id 310321) that was determined to be an outlier. 310321 is a mouse where we did not see any whisker response, it reacted to angiotensin II, but the BP increase was abrupted for a short while and then re-established. Perhaps due to a clot or a bubble in the venous catheter. This resulted in a biphasic and slow BP increase.

In all of our analyses we assumed that any missing measurements were caused by factors that were unrelated to our main target process, or in other words that the absent measurements

were “missing at random”. We therefore did not attempt to model the measurement removal process explicitly.

## Modelling approach

All of our models had a common structure, with generalised linear models used to describe information from measurements and multi-level prior distributions used to describe non-experimental information. The modelling choices open to us concerned the following questions:

1. What generalised linear model to use to model measurements?
2. Which interaction effects to estimate?
3. What structure to use for the prior model, and in particular how and to what extent to pool information between categories?
4. What quantitative values to use for our prior model?
5. In cases where a phenomenon of interest was assessed using multiple, potentially related measurements, whether to model the possible relatedness?

In order to answer these questions for each analysis, we started with a simple but plausible model, then iteratively added and removed components as described in Gelman et al. (2020). Our general aim was to achieve a better quantitative and qualitative description of the data generating process while avoiding computational issues. In particular, we focused on the estimated out of sample predictive performance as measured by the estimated leave-one-observation-out log predictive density (Vehtari, Gelman, and Gabry 2017) and qualitative agreement between predictive and observed observations in graphical checks.

We address these questions below for each analysis in its corresponding section, providing model comparisons and illustrative results where appropriate.

## Software

The raw data are found in csv files which are available from our project’s github repository at the following urls:

- [https://github.com/teddygroves/sphincter/blob/main/data/raw/hyper\\_challenge.csv](https://github.com/teddygroves/sphincter/blob/main/data/raw/hyper_challenge.csv)
- [https://github.com/teddygroves/sphincter/blob/main/data/raw/data\\_sphincter\\_paper.csv](https://github.com/teddygroves/sphincter/blob/main/data/raw/data_sphincter_paper.csv)

For each analysis, we conducted filtering and reshaping operations using the standard scientific Python stack and validated the resulting prepared datasets against custom data models constructed using the Python libraries pydantic (Pydantic developers 2022) and pandera (Niels Bantilan 2020). These models can be inspected at this url: [https://github.com/teddygroves/sphincter/blob/main/sphincter/data\\_preparation.py](https://github.com/teddygroves/sphincter/blob/main/sphincter/data_preparation.py).

Statistical computation was carried out using the probabilistic programming framework Stan (Carpenter et al. 2017) via the interface cmdstanpy (Stan Development Team 2022).

Analysis and serialisation of posterior samples was carried out using the Bayesian inference library arviz (Kumar et al. 2019).

Our analysis was orchestrated by the Python package bibat (Groves 2023).

## Validation of statistical computation

We validated our statistical computation using standard Hamiltonian Monte Carlo diagnostics, including the improved  $\hat{R}$  statistic (Vehtari et al. 2021) as well as inspection for post-warmup divergent transitions (Betancourt 2017), problematic EBF1 statistics, tree depth or effective sample size to total sample size ratios. All reported models had improved  $\hat{R}$  close to 1 and no divergent transitions or other signs of algorithm failure.

## Reproducibility

See the repository readme for instructions on reproducing our analysis: <https://github.com/teddygroves/sphincter/blob/main/README.md>

## Main findings

### Whisker stimulation

Our analysis of whisker stimulation data indicates that the hypertension and sphincter ablation treatments are associated with lower whisker stimulation response, as measured by log diameter change, compared with the baseline treatment. The effect from whisker stimulation is greater than from hypertension.

Figure Figure 1 illustrates this finding by showing the distribution of posterior samples for treatment effects relative to baseline from our best whisker stimulation model.

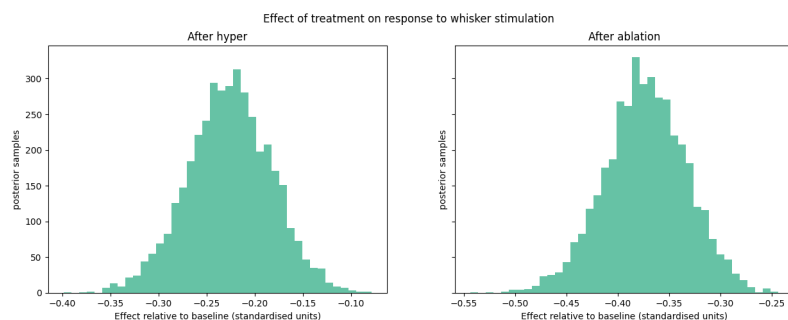


Figure 1: Marginal posterior histograms for treatment effects, relative to the baseline treatment.

Our analysis did not indicate any substantial difference between old and adult mice, or any noticeable vessel type:treatment interaction effects. This can be seen from figure Figure 2, which shows posterior quantiles for age and vessel type:treatment interaction effects in our model that included both of these.

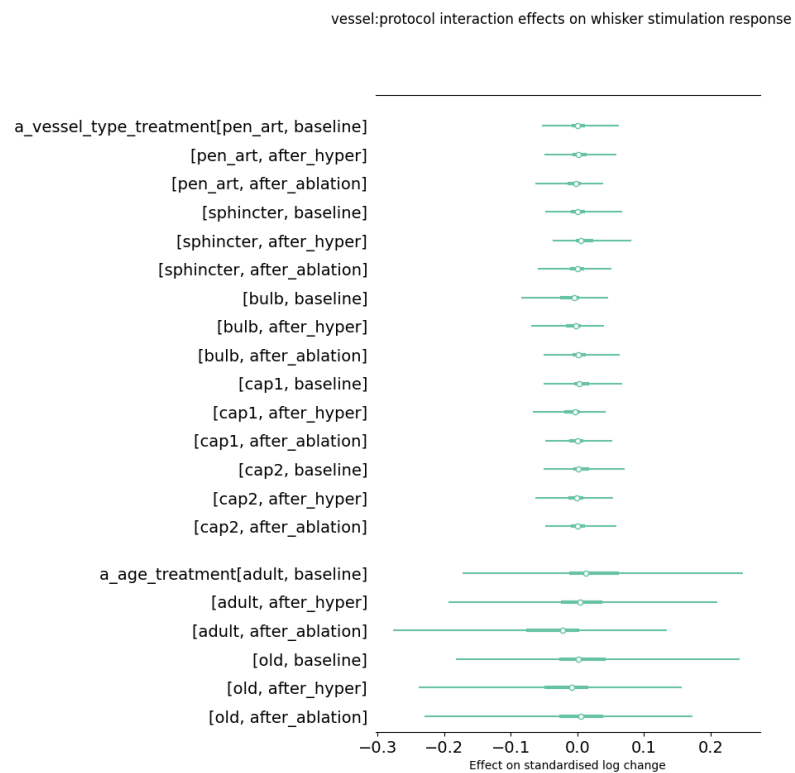


Figure 2: Marginal 2.5%-97.5% posterior intervals for protocol effects

We found some difference between vessel type effects: sphincters had the greatest relative diameter change in response to whisker stimulation, and bulbs the smallest. Figure Figure 3 shows these.

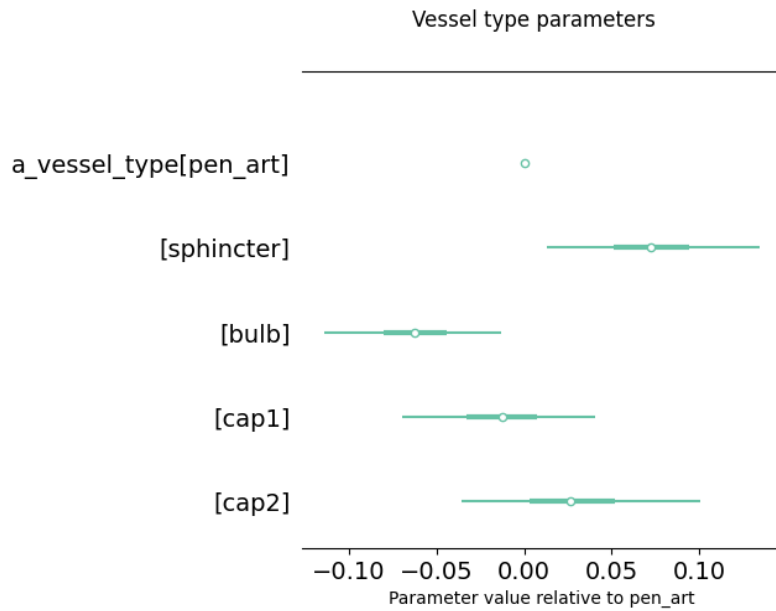


Figure 3: Marginal 2.5%-97.5% posterior intervals for vessel type effects

## Pulsatility

Our analysis of vessel centre and diameter pulsatility yielded the following conclusions:

- Adult mice have higher vessel diameter pulsatility than old mice, whereas old mice have slightly higher centre pulsatility.
- Sphincter ablation correlates with increased diameter pulsatility, with no strong interaction effects. On the other hand there is no clear effect of sphincter ablation on centre pulsatility.



Figure 4 plots the distribution of age effect differences (adult minus old) for each measurement type in our final model. This graph shows that, in this model, the age effect for adult mice was higher than for old mice in every single posterior sample: in other words there is a clear trend for older mice to have lower diameter pulsatility. There is a smaller opposite trend for centre pulsatility measurements, but it is not clearly separated from zero, indicating that the direction of the effect is not fully settled.

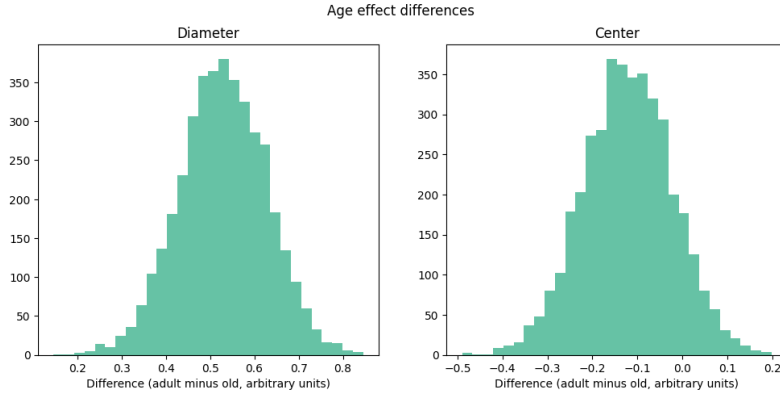


Figure 4: Posterior distribution of age effect differences for each measurement type.

Figure 20 shows the distribution of posterior draws for sphincter ablation effects compared with the immediately prior protocol stage (“after hypertension”). The ablation/diameter parameter is greater than the after hypertension/diameter parameter in 98% of posterior samples, whereas there is no clear effect on centre pulsatility.

## Red blood cell flow

Our main result regarding red blood cell flow is that both RBC speed and flux tend to be higher in adult mice compared with old mice. Figure 6 illustrates this finding by plotting the relevant marginal posterior histograms.

We also quantified treatment effects on red blood cell flow, as shown in Figure 7:

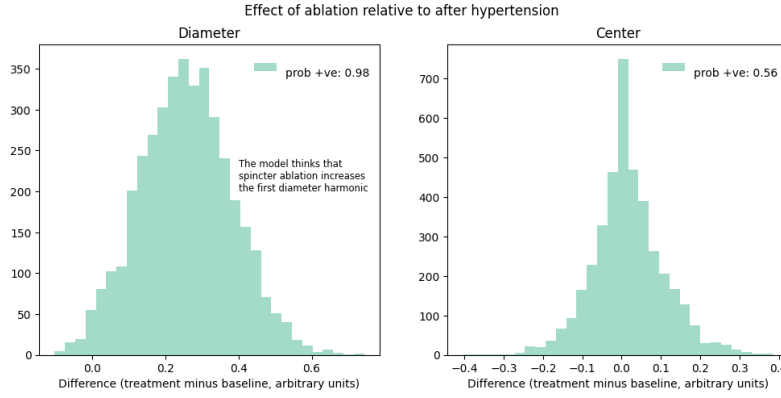


Figure 5: Posterior distribution of treatment effect differences for each measurement type.

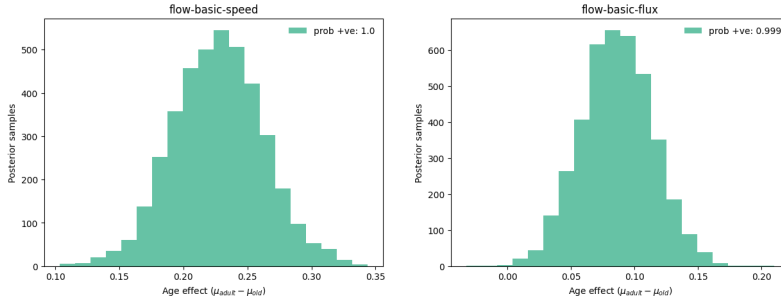


Figure 6: Posterior distributions of age effects on red blood cell speed and flux.

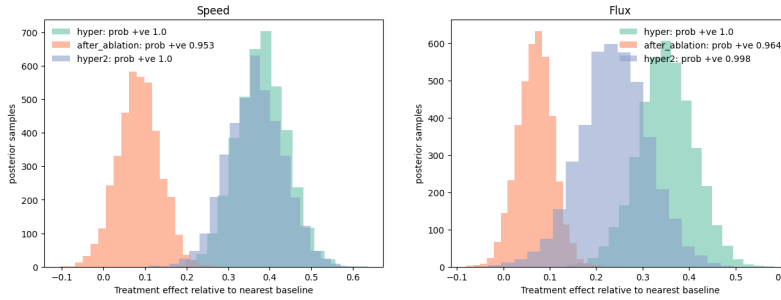


Figure 7: Posterior distributions of treatment effects on red blood cell speed and flux. The nearest baseline for treatment **hyper** is **baseline**, and for treatments **after\_ablation** and **hyper2** it is **after\_hyper**.

## Hypertensive challenge

Our hypertensive challenge data also showed pronounced age and treatment differences, as shown in figure Figure 8. Specifically, we found that, for adult mice, blood pressure and vessel diameter tended to be less correlated, and that the **hyper2** treatment tended to increase pressure-diameter correlation compared with the **hyper1** treatment.

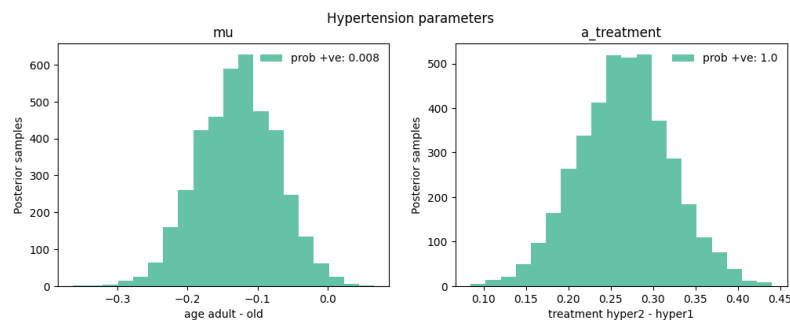


Figure 8: Posterior distributions of relative age and treatment effects for hypertensive challenge data.

## Details of the whisker stimulation analysis

In order to measure how vascular responsiveness changed during our experimental protocol, the diameters of different vessel types were recorded before and during whisker stimulation, at baseline, post-hypertension and post-ablation stages. We aimed to assess statistical relationships between the known factors and the relative change in vessel diameter before and after stimulation.

In particular, we were interested in differences between old and adult mice and in the effect of sphincter ablation.

### Dependent variable

The ratio of the peak compared with the pre-stimulation level for each mouse at each stage, on natural logarithmic scale, also

known as the ‘log change’, was standardised by subtracting the overall mean and dividing by the standard deviation, then treated as a single measurement. This way of the measurements was chosen to facilitate modelling, as log change is a symmetric and additive measure of relative change (see Tornqvist, Vartia, and Vartia (1985)).

Note that when the difference between the two values  $v1$  and  $v2$  is far less than 1, the log change  $\ln \frac{v2}{v1}$  is approximately the same as the more widely used relative difference measure  $\frac{v2-v1}{v1}$ .

## Statistical Models

The final model is shown below:

$$\begin{aligned}
 y_{vtm} &\sim ST(\nu, \hat{y}_{vtm}, \sigma_t) \\
 \hat{y}_{vtm} &= \mu_a \\
 &\quad + \alpha_t^{treatment} \\
 &\quad + \alpha_v^{vesseltype} \\
 \alpha_t^{treatment} &\sim N(0, \tau^{treatment}) \\
 \alpha_v^{vesseltype} &\sim N(0, \tau^{vesseltype}) \\
 \nu &\sim Gamma(2, 0.1) \\
 \sigma_t &\sim HN(0, 0.5) \\
 \mu &\sim N(0, 0.7) \\
 \tau^{treatment} &\sim HN(0, 0.7) \\
 \tau^{vesseltype} &\sim HN(0, 0.7)
 \end{aligned} \tag{1}$$

In equation (1), the term  $ST$  indicates the student t distribution,  $N$  indicates the normal distribution,  $Gamma$  the gamma distribution and  $HN$  the ‘half-normal’ distribution, i.e. the normal distribution with support only for non-negative numbers. Subscripts indicate indexes and superscripts indicate parameter labels.

This model has independent effects for treatments ( $\alpha^{treatment}$ ), vessel types ( $\alpha^{vessel}$ ) and age ( $\mu$ ). The treatment and vessel

type effects have hierarchical priors, reflecting the need to partially pool information between different treatment and vessel type categories. This structure allows our model to accommodate the full spectrum between different categories being highly heterogeneous—in this case the corresponding  $\tau$  parameter will be large—and on the other hand high similarity, leading to small  $\tau$  parameters. The student t distribution was chosen to reflect the heavy tails we observed in the data, with the parameter  $\nu$  controlling the level of heaviness.

The prior standard deviation 0.7 was chosen because it led to what we judged to be a reasonable allocation of prior probability mass over possible data realisations. The prior for the student t degrees of freedom parameter  $\nu$  was set following the recommendation in Juárez and Steel (2010).

As well as this model, we also present results from a more complex model that adds vessel type:treatment and age:treatment interaction effects. The full specification of this “big” model is as follows, using the same notation as equation (1):

$$\begin{aligned}
y_{vtm} &\sim ST(\nu, \hat{y}_{vtm}, \sigma_t) \\
\hat{y}_{vtm} &= \mu_a \\
&+ \alpha_t^{treatment} \\
&+ \alpha_v^{vesseltype} \\
&+ \alpha_{vt}^{vesseltype:treatment} \\
&+ \alpha_{vt}^{age:treatment} \\
\alpha_t^{treatment} &\sim N(0, \tau^{treatment}) \\
\alpha_v^{vessel} &\sim N(0, \tau^{vessel}) \\
\alpha_{vt}^{vesseltype:treatment} &\sim N(0, \tau^{vesseltype:treatment}) \\
\alpha_{at}^{age:treatment} &\sim N(0, \tau^{age:treatment}) \\
\nu &\sim Gamma(2, 0.1) \\
\sigma &\sim HN(0, 0.5) \\
\mu &\sim N(0, 0.7) \\
\tau^{treatment} &\sim HN(0, 0.7) \\
\tau^{vessel} &\sim HN(0, 0.7) \\
\tau^{age:treatment} &\sim HN(0, 0.7) \\
\tau^{vesseltype:treatment} &\sim HN(0, 0.7)
\end{aligned} \tag{2}$$

## Results

Figure 9 shows the observed log change measurements with colours illustrating the various categorical information. Note that the measurements have different dispersion depending on the treatment, indicating a the need for a model with distributional effects.

Figure 10 compares the measurements with our model's posterior predictive distribution. This shows that our model achieved a reasonable fit to the observed data. There is a pattern in the model's bad predictions, in that these tend to be for higher baseline measurements. However, we judged that this pattern was small enough that for our purposes we could disregard it.

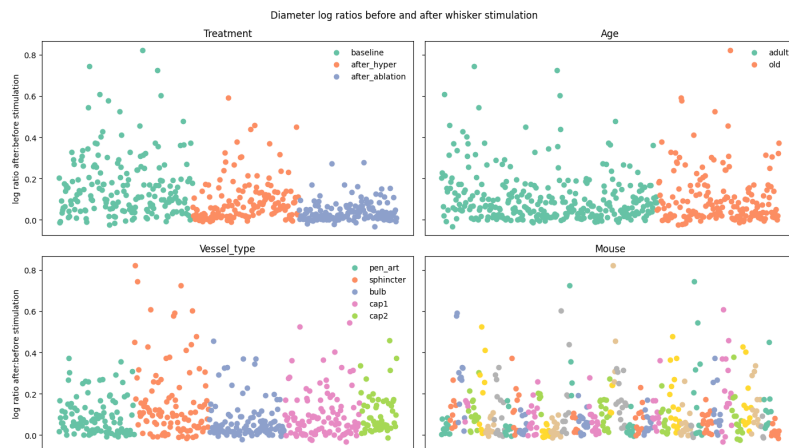


Figure 9: Raw measurements

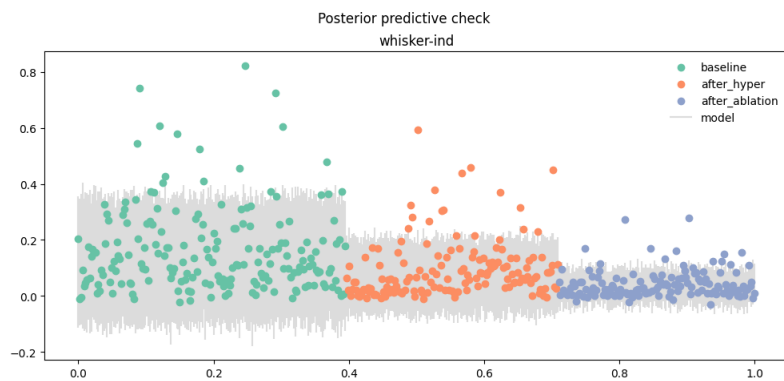


Figure 10: Graphical posterior predictive check

To back up our finding that there were no important interaction effects, we compared the predictive performance of our final model **whisker-ind** with **whisker-big**, the best-performing model with more interactions. The results are shown below in figure Figure 11:

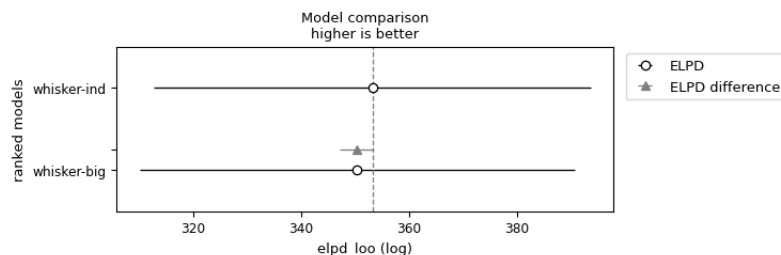


Figure 11: Comparison of estimated leave-one-out log predictive density between the final model **whisker-ind** and the best performing interaction model **whisker-big**.

The two models have similar estimated predictive performance, indicating that there is no gain from considering interaction effects in this case.

## Details of the pulsatility analysis

The pulsatility data consisted of fast measurements of diameter and center point for the same mice. These measurements were Fourier-transformed, and the harmonics of the transformed data were interpreted as representing the pulsatility of the measured quantities.

### Dependent variable

We used the first harmonic of each transformed time series as a dependent variable. It might have been preferable to aggregate all the available power harmonics, but this would have complicated our measurement model, and in any case power at the subsequent harmonics was typically negligible compared with the first.



## Questions

As well as the results reported in the main findings section, we were also interested in these additional questions:

**Question 0.1.** *How does blood pressure affect diameter and centre pulsatility?*

**Question 0.2.** *Do hypertension and sphincter ablation influence diameter and centre pulsatility differently?*

## Description of the dataset

As well as the categorical data described above, our pulsatility analysis also took into account measurements of each mouse's blood pressure at the femoral artery.

The final dataset included 514 joint measurements of diameter and centre pulsatility, calculated as described above. These measurements are shown in Figure 12.

Figure 13 shows the relationship between pressure and the measurements in our dataset for both age categories. The light dots show raw measurements and the darker dots show averages within evenly sized bins.

Figure 14 shows the relationship between diameter and the measurements in our dataset for all vessel type categories. The light dots show raw measurements and the darker dots show averages within evenly sized bins. There is a clear positive relationship between measured absolute diameter and diameter pulsatility, and it is approximately the same for all vessel types.

## Statistical models

We knew from prior studies that the power harmonics should individually follow exponential distributions [REFERENCE FOR THIS]. This consideration motivated the use of exponential generalised linear models for both the centre and diameter pulsatility measurements. In this model, given measurement  $y$

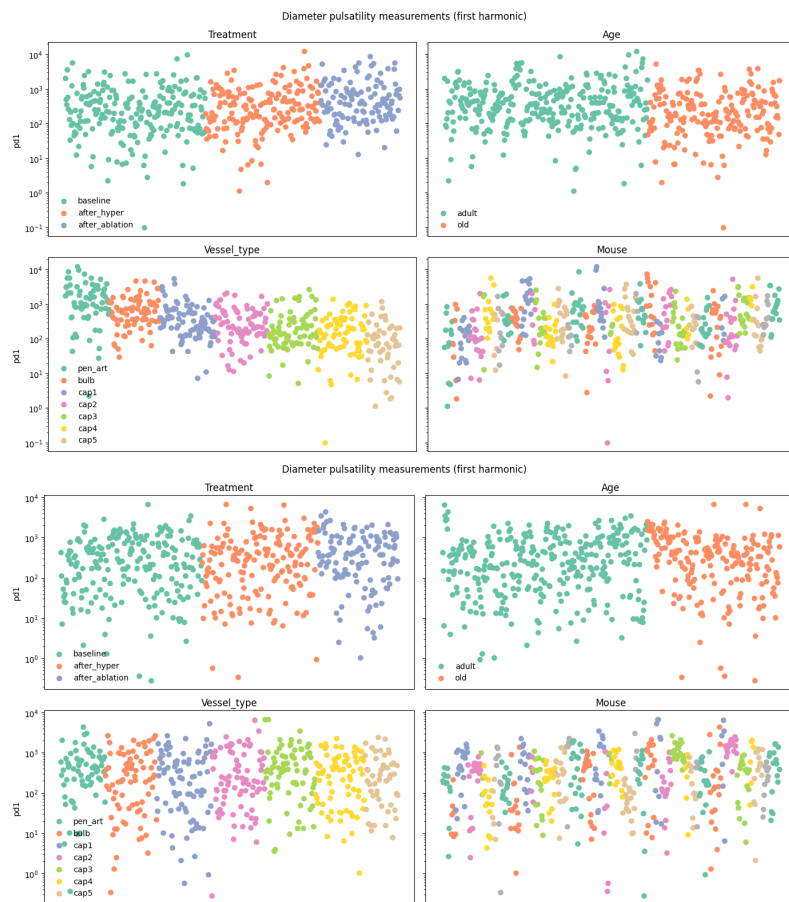


Figure 12: The modelled measurements, shown in order of the coloured categories.

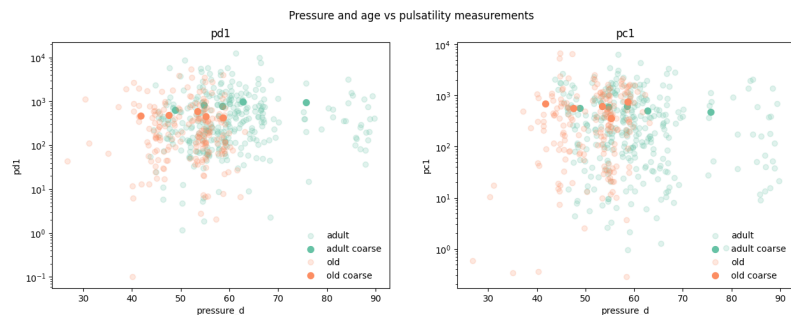


Figure 13: Pulsatility measurements plotted against the corresponding pressure measurements and coloured according to age. Darker dots indicate averages within evenly sized pressure bins.

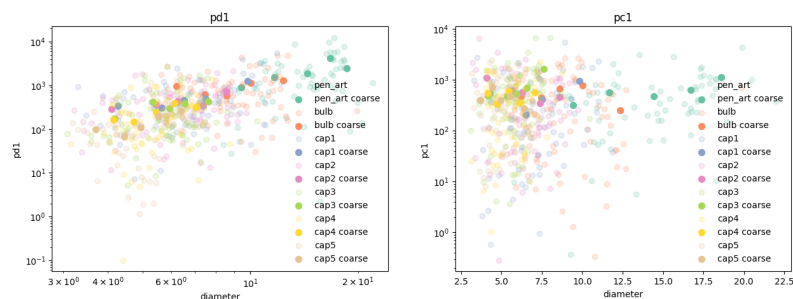


Figure 14: Pulsatility measurements plotted against the corresponding diameter measurements and coloured according to vessel type. Darker dots indicate averages within evenly sized pressure bins.

and linear predictor  $\eta$  the measurement probability density is given by this equation:

$$\begin{aligned} p(y | \eta) &= \textit{Exponential}(y, \lambda) \\ &= \lambda e^{-\lambda y} \end{aligned} \tag{3}$$

$$\ln \frac{1}{\lambda} = \eta \tag{4}$$

The log link function (4) was chosen so that linear changes in the term  $\eta$  induce multiplicative changes in the mean  $\frac{1}{\lambda}$  of the measurement distribution, as we believed the effects we wanted to model would be multiplicative.

We compared four different ways of parameterising  $\eta$  based on the information available about a given measurement, corresponding to three hypotheses about the way the data were generated.

The simplest model, which we labelled “basic”, calculates the linear predictor  $\hat{\eta}_{vtad}^{\text{basic}}$  for an observation with vessel type  $v$ , treatment  $t$ , age  $a$  and diameter  $d$  as follows:

$$\begin{aligned} \eta_{vtad}^{\text{basic}} &= \mu_a \\ &+ \alpha_t^{\text{treatment}} \\ &+ \alpha_v^{\text{vesseltype}} \\ &+ \beta^{\text{diameter}} \cdot \ln d \end{aligned} \tag{5}$$

The basic model provided a plausible baseline against which to compare the other models.

Next we constructed a more complex model by extending the basic model with interaction effects, resulting in the following linear predictor:

$$\eta_{vtad}^{interaction} = \mu_a \quad (6)$$

$$\begin{aligned} &+ \alpha_t^{treatment} \\ &+ \alpha_{d,vesseltype(n)}^{vesseltype} \\ &+ \alpha_{at}^{age:treatment} \\ &+ \alpha_{tv}^{age:treatment:vesseltype} \\ &+ \beta_d^{diameter} \cdot \ln d \end{aligned} \quad (7)$$

Next we constructed a model that adds to the basic model parameters that aim to capture possible effects corresponding to the blood pressure measurements. To compensate for collinearity between age and pressure, our “pressure” model does not use the observed pressure as a predictor, but rather the age-normalised pressure, calculated by subtracting the mean for each age category from the observed pressure measurement. The model for the linear predictors  $\eta_{vatdp}^{pressure}$  with age-normalised pressure measurement  $p$  is then

$$\begin{aligned} \eta_{vatdp}^{pressure} &= \mu_a \\ &+ \alpha_t^{treatment} \\ &+ \alpha_v^{vesseltype} \\ &+ \beta_d^{diameter} \cdot \ln d \\ &+ \beta_a^{pressure} \cdot p \end{aligned} \quad (8)$$

Finally, we made a model that includes a pressure effect but no age-specific parameters from the pressure model. This was to test whether any age effects are due to the collinearity between age and pressure. The pressure-no-age model’s linear predictors  $\eta_{vatdp}^{pressure \text{ no age}}$  are calculated as shown in equation (9). Note that, unlike in equation (8), the  $\mu$  and  $\beta^{pressure}$  parameters in equation (9) have no age indexes.

$$\begin{aligned}
\eta_{vatdp}^{pressure \ no \ age} = & \mu \\
& + \alpha_t^{treatment} \\
& + \alpha_v^{vesseltype} \\
& + \beta_d^{diameter} \cdot \ln d \\
& + \beta^{pressure} \cdot p
\end{aligned} \tag{9}$$

In all of our models the  $\alpha$  parameters were given independent, semi-informative, hierarchical prior distributions to allow for appropriate information sharing. The  $\beta$  and  $\mu$  parameters were given independent, semi-informative, non-hierarchical prior distributions.

## Results

We estimated the leave-one-out log predictive density for each model using the method described in Vehtari, Gelman, and Gabry (2017) and implemented in Kumar et al. (2019). The results of the comparison are shown below in Figure 15.

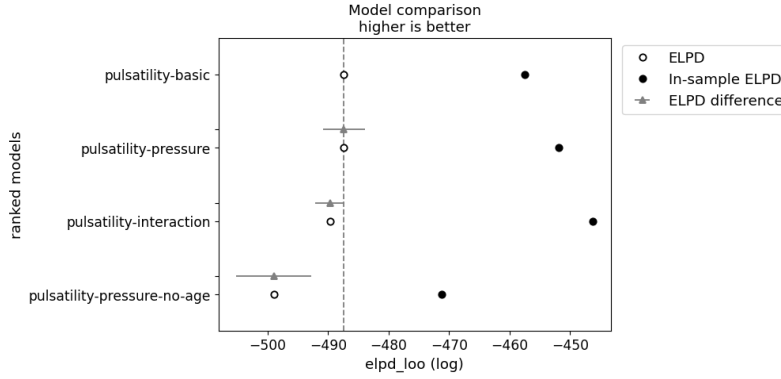


Figure 15: Comparison of estimated leave-one-out log predictive density (ELPD) for our pulsatility models. The main result is that the pressure-no-age and interaction models are clearly worse than the pressure model, as shown by the separation of the relevant grey and dotted lines.

We evaluated our models' fit to data using prior and posterior predictive checking, with the results for the pressure model shown in Figure 16.

Inspecting of the interaction model output showed that none of the interaction effect parameters that differed substantially from zero, as can be seen in Figure 17.

From this result, together with the worse estimated out of sample predictive performance as shown in Figure 15, we concluded that there were no important interaction effects, so that we could essentially discard the interaction model.

Figure 18 shows the marginal posterior distributions for other effect parameters in all three models. Note that the parameters `b_diameter` are strongly positive for diameter pulsatility in all models and also mostly positive for centre pulsatility. There is also a strong trend for diameter pulsatility to decrease with the order of the vessel and no particular vessel type trend for centre pulsatility.

## Answers to specific questions

The poorer estimated out of sample predictive performance of the pressure-no-age model compared with the other models, as shown in Figure 15, indicates that our pressure measurements did not fully explain the observed difference between adult and old mice. It is nonetheless possible that different pressure explains the difference between old and adult mice, but that the pressure measurements did not reflect the true pressure at the measured vessels. This is plausible since the pressure measurements were taken at a different location.

Figure 19 shows the difference in  $\beta^{pressure}$  parameters for old and adult mice in the pressure model in order to answer Question 0.1. This shows a weak tendency of the pressure effect on diameter pulsatility to be more positive for adult mice than for old mice, and a strong opposite tendency for centre pulsatility. Taking the absolute values into account, the analysis suggests that greater measured pressure is not strongly related to diameter pulsatility and correlates with reduced centre pulsatility for adult mice but not for old mice.

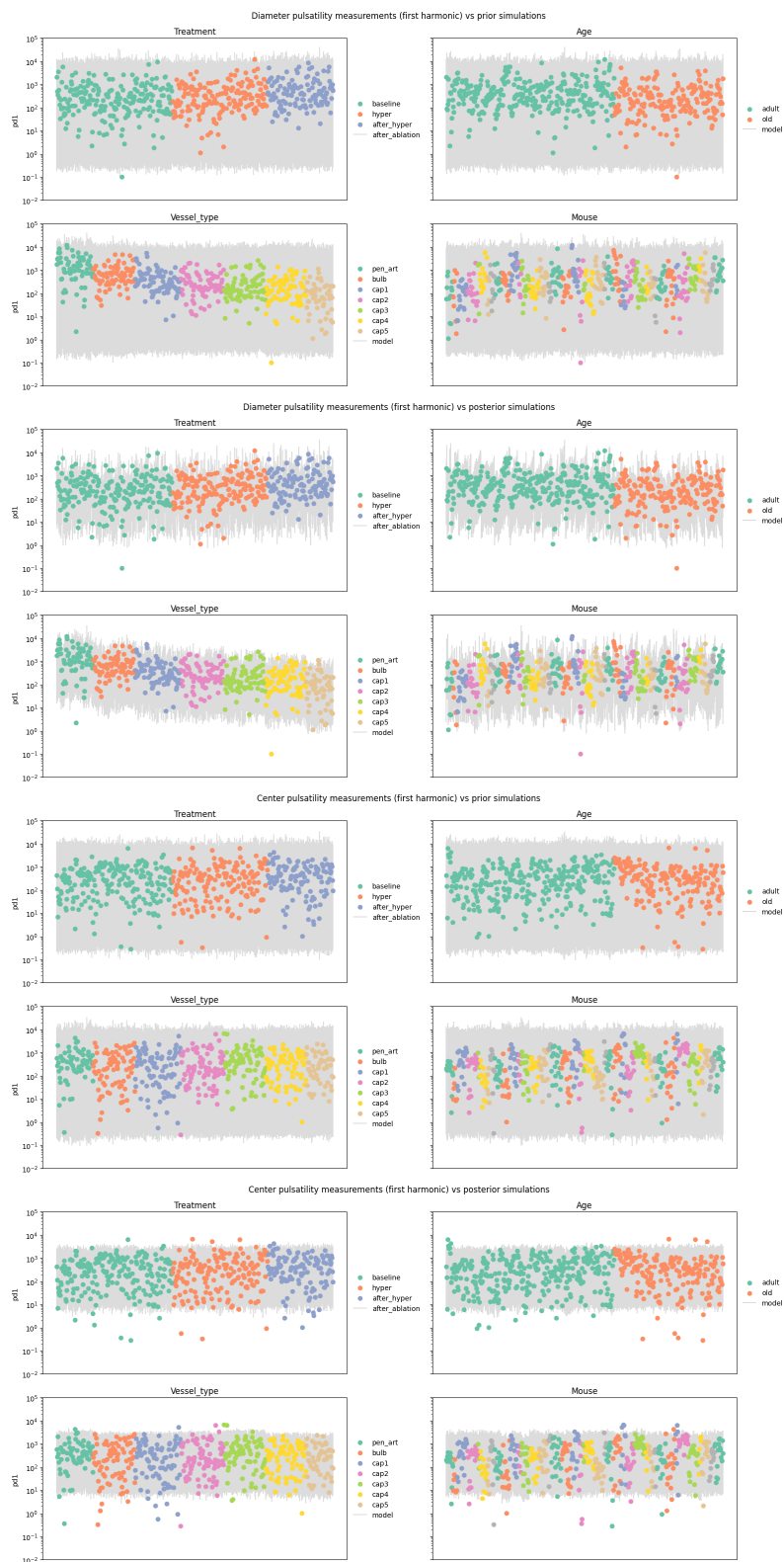


Figure 16: Prior and posterior predictive checks for the pressure model.



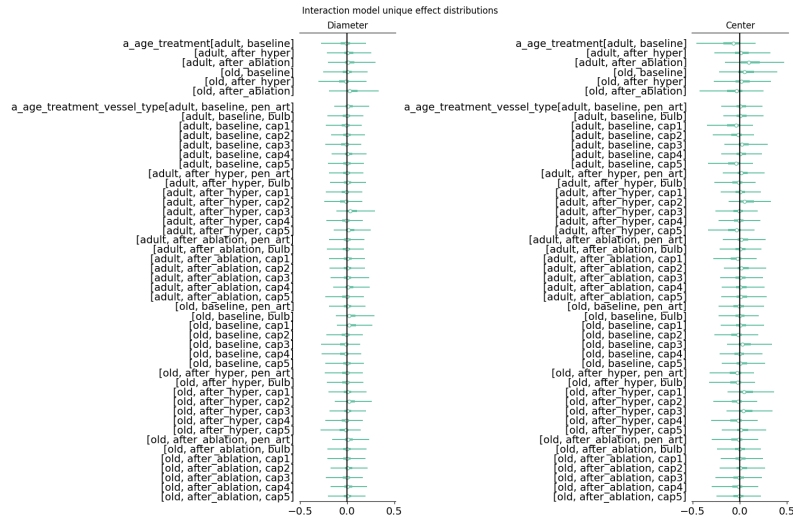


Figure 17: Marginal posterior quantiles for the unique effects in the interaction model.

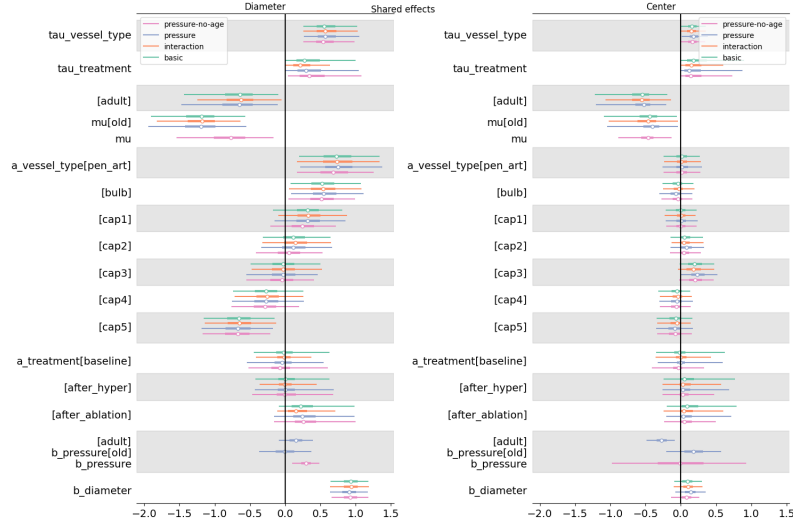


Figure 18: Marginal posterior quantiles for shared model effects.

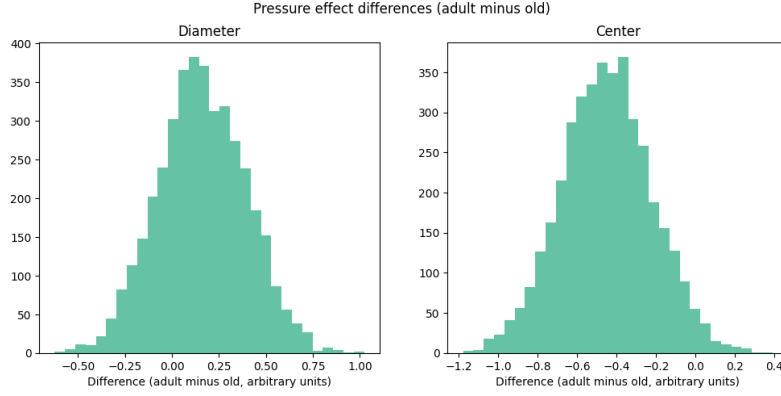


Figure 19: Posterior distribution of pressure effect differences for each measurement type.

To illustrate the effect of treatments, and specifically sphincter ablation relative to hypertension (i.e. to answer Question 0.2) Figure 20 shows the difference between the effect for each treatment and the baseline treatment effect. There is a clear effect of ablation to increase diameter pulsatility and no clear effects of hypertension on diameter pulsatility or of either treatment on centre pulsatility.

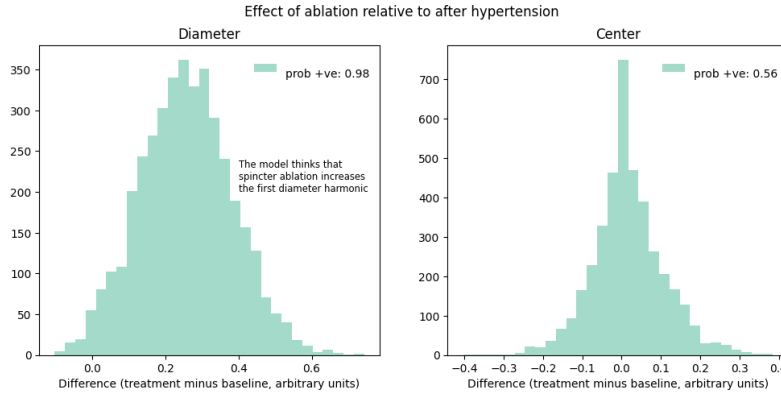


Figure 20: Posterior distribution of treatment effect differences for each measurement type.

To get an idea about how the effect of sphincter ablation on diameter pulsatility compares quantitatively with the effect of hypertension, we fit the basic model to the full dataset, without

excluding measurements from either hypertension treatment. Figure 21 shows the main result from fitting this model: ablation and hypertension had similarly positive effects on diameter pulsatility. Interestingly there is no clear effect from the second hypertension treatment.

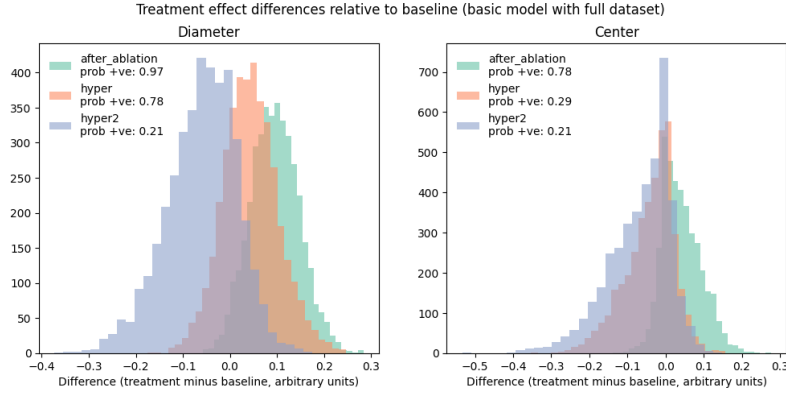


Figure 21: Treatment effect distributions relative to baseline in the basic model when fit to the full dataset including all treatments.

## Details of the red blood cell flow analysis

Our measurements included flow data recording the measured speed and flux of red blood cells through some vessels. This data is interesting because it allows for inference of the local blood pressure, which determines the speed.

We were interested in whether the speed and flux tended to be different between old and adult mice for a given vessel type and treatment, as this would indicate that the pressure would likely be similar as well.

## Data processing

There is a significant missing data issue in this case: both speed and flux measurements were available for only 271 out of 1525 raw measurements. We therefore conducted separate analyses

for speed and flux even though we suspected that these two quantities are related.

## Dependent variable

We modelled speed and flux measurements on natural logarithmic scale as we expected multiplicative effects and this transformation ensures support on the whole real number line, simplifying modelling.

The resulting measurements are shown in figure Figure 22.

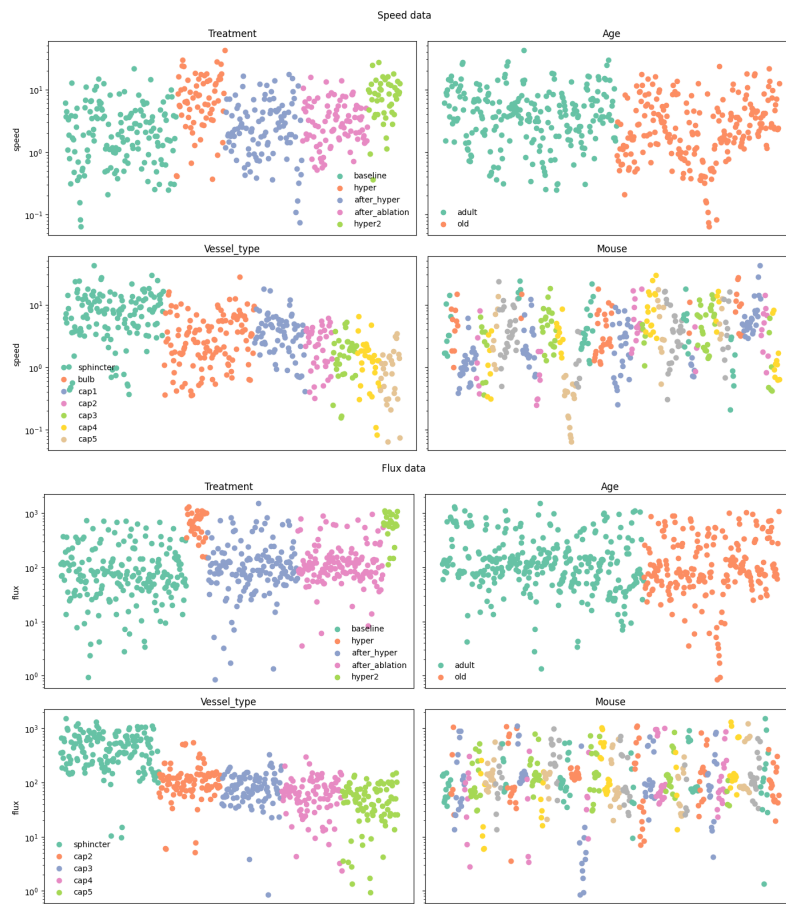


Figure 22: Red blood cell speed and flux data

From a glance at these graphs it is clear that there were treatment effects on both measurement types, particularly from the hypertension treatments, that there are treatment-related distributional effects, and that both speed and flux reduce as vessel order increases.

## Statistical models

As in the whisker case we investigated the results of fitting models with and without interaction effects. Again we found no large or fully resolved interactions and therefore used the smaller model for further analysis.

Our final model had this specification:

$$\begin{aligned}
\ln(y_{vtm}) &\sim N(\hat{y}_{vtm}, \sigma_t) \\
\hat{y}_{vtm} &= \mu_a \\
&\quad + \alpha_t^{treatment} \\
&\quad + \alpha_v^{vesseltype} \\
\alpha_t^{treatment} &\sim N(0, \tau^{treatment}) \\
\alpha_v^{vesseltype} &\sim N(0, \tau^{vesseltype}) \\
\sigma_t &\sim HN(0, 0.5) \\
\mu &\sim N(0, 0.5) \\
\tau^{treatment} &\sim HN(0, 0.5) \\
\tau^{vesseltype} &\sim HN(0, 0.5)
\end{aligned} \tag{10}$$

We chose the prior standard deviation 0.5 after prior predictive checking to ensure reasonably tight coverage of the observed data.

For investigation of interaction effects we fit another model that extended our final model with a vessel type:treatment interaction effect as follows:

$$\begin{aligned}
\hat{y}_{vtm} &= \mu_a \\
&+ \alpha_t^{treatment} \\
&+ \alpha_v^{vesseltype} \\
&+ \alpha_{vt}^{vesseltype:treatment} \\
\alpha_v^{vesseltype:treatment} &\sim N(0, \tau^{vesseltype:treatment}) \\
\tau^{vesseltype:treatment} &\sim HN(0, 0.5)
\end{aligned}$$

## Results

Our statistical model successfully captured all of these trends, as can be seen from figure [Figure 23](#)

### Shared parameters

[Figure 24](#) summarises the marginal posterior distributions for parameters that appear in both our final flux and speed models. There is quite a lot of agreement, unsurprisingly given that red blood cell speed and flux are closely related. We conclude from this plot that, given sufficient data, a joint model including both measurement types would be a good topic for future analysis.

It is also interesting to note the main area where our flux and speed models disagree, namely the effect corresponding the vessel type sphincter. According to our model, the sphincter tends to have the lowest RBC speed of all vessels but the highest flux.

### Interactions

To test for important interaction effects we fit a model with vessel type:treatment interaction parameters. This model achieved marginally worse loo elpd scores as shown in figure [Figure 25](#).

For completeness the interaction effects are shown in figure [Figure 26](#). No effects are clearly separated from zero. The sphincter/ablation effect on RBC speed is notably different from the

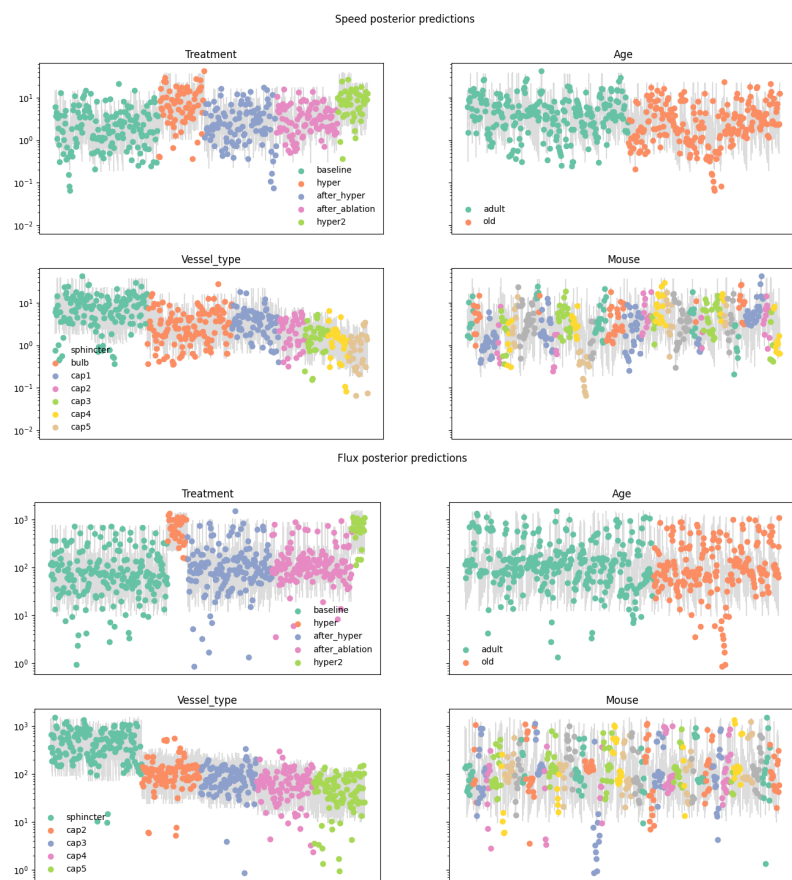


Figure 23: Posterior predictive checks for flow models

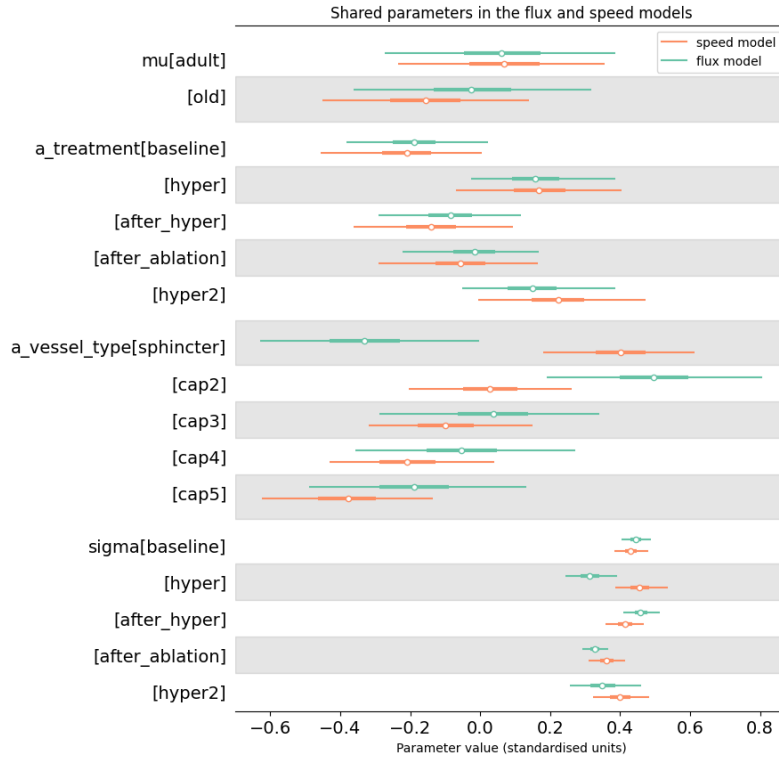


Figure 24: Posterior distributions of parameters that appear in both our flow and speed models.



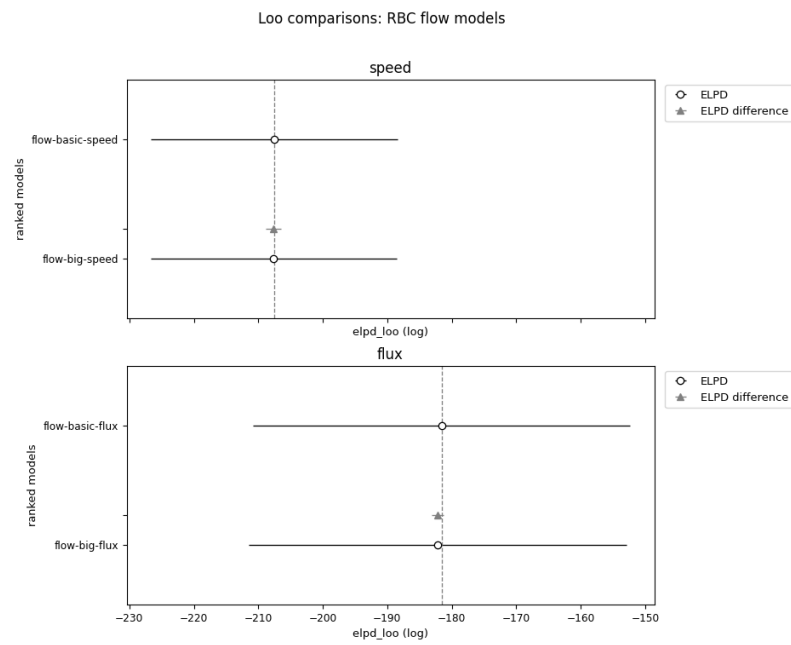


Figure 25: Out of sample predictive performance comparison for red blood cell flow models.

others. We fit several models with sparsity-inducing priors including the regularised horseshoe Piironen and Vehtari (2017) to see if it was possible to resolve this effect, but were unsuccessful. From this we conclude that any real effect is too small to be easily detected in our dataset.

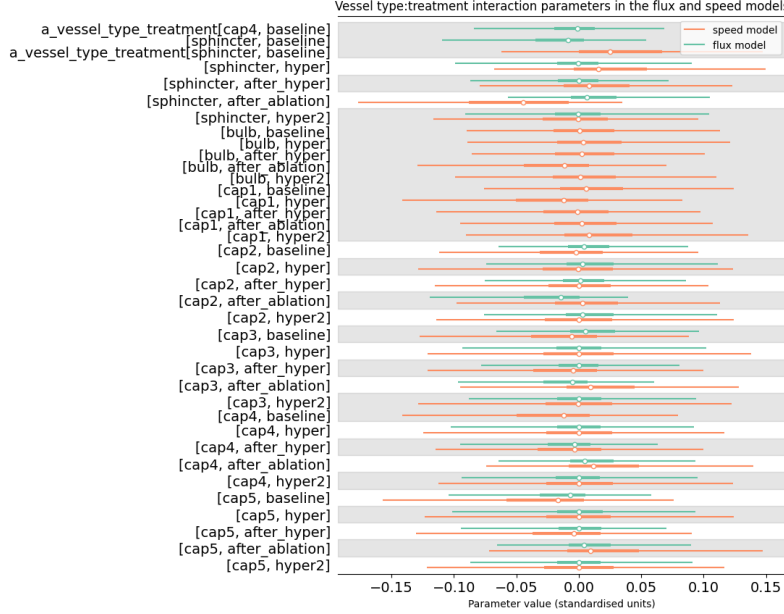


Figure 26: Interaction effects on red blood cell flow.

## Hypertensive challenge

### Dependent variable

The raw data on hypertension challenge were correlation coefficients relating blood pressure and vessel diameter, which are constrained to lie on the  $[-1, 1]$  interval. For easier modelling we transformed these by applying an inverse hyperbolic tangent function for use in modelling. The dependent variables then had support on the entire real number line.

The resulting dataset is shown in figure Figure 27. The transformed correlation coefficients do not appear extremely dis-

persed, indicating that standard modelling techniques ought to be able to describe them.

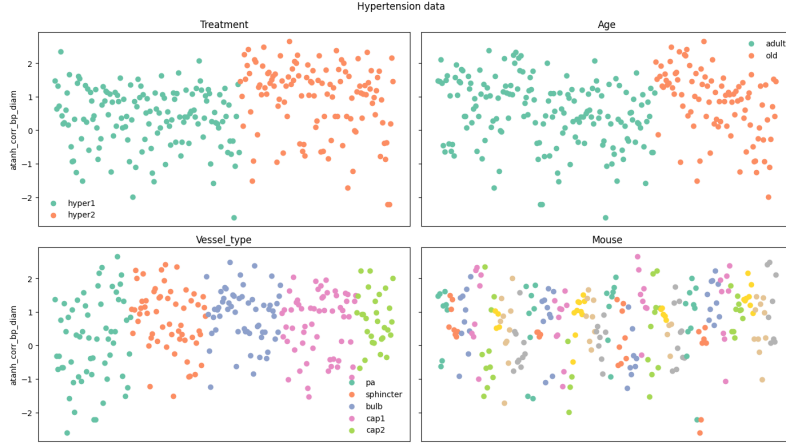


Figure 27

## Statistical model

Our final statistical model had the following form.

$$\begin{aligned}
 \ln(y_{vtm}) &\sim N(\hat{y}_{vtm}, \sigma_v) \\
 \hat{y}_{vtm} &= \mu_a \\
 &\quad + \alpha_t^{treatment} \\
 &\quad + \alpha_v^{vesseltype} \\
 \alpha_t^{treatment} &\sim N(0, 0.5) \\
 \alpha_v^{vesseltype} &\sim N(0, \tau^{vesseltype}) \\
 \sigma_v &\sim HN(0, 0.5) \\
 \mu &\sim N(0, 0.5) \\
 \tau^{vesseltype} &\sim HN(0, 0.5)
 \end{aligned} \tag{11}$$

This model is different from the others in that we did not partially pool the treatment effects, since there were only two of these in this case. We also allowed the measurement error parameters  $\sigma$  to vary according to vessel type, since this improved model fit and predictive performance.

As in the other analyses, for investigation of interaction effects we fit another model that extended our final model with a vessel type:treatment interaction effect as follows:

$$\begin{aligned}
\hat{y}_{vtm} &= \mu_a \\
&+ \alpha_t^{treatment} \\
&+ \alpha_v^{vesseltype} \\
&+ \alpha_{vt}^{vesseltype:treatment} \\
\alpha_v^{vesseltype:treatment} &\sim N(0, \tau^{vesseltype:treatment}) \\
\tau^{vesseltype:treatment} &\sim HN(0, 0.5)
\end{aligned}$$

## Results

Figure 28 shows that, as in the other cases, including interaction effects did not improve estimated predictive performance.

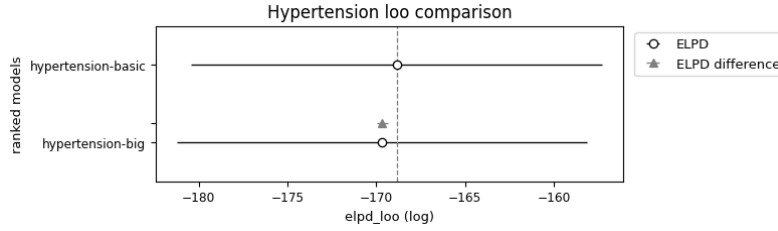


Figure 28: Comparison of out-of-sample predictive performance of our hypertension models, as measured by estimated leave-one-out expected log predictive density. The two models have similar estimated performance, but the **hypertension-big** model is clearly worse.

Figure 29 shows the marginal distributions for the non-hierarchical parameters in our final model.

Figure 30 shows graphical prior and posterior predictive checks for our final hypertension model. The fit is fairly good, with no obvious systematic pattern in the errors, though slightly

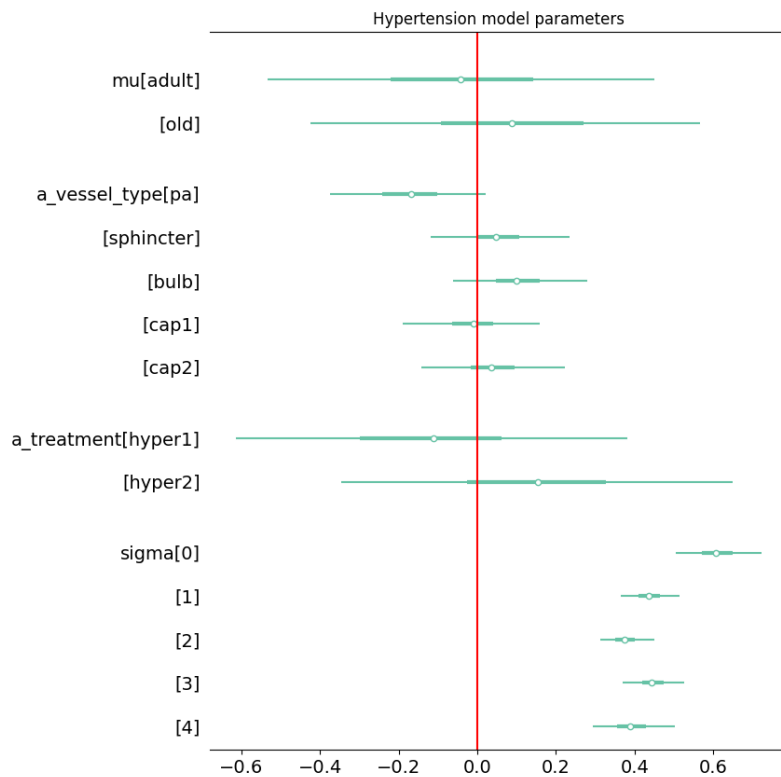


Figure 29: 1%-99% posterior intervals for parameters in our final hypertension model.

more observations lie outside the plotted intervals than might be expected.

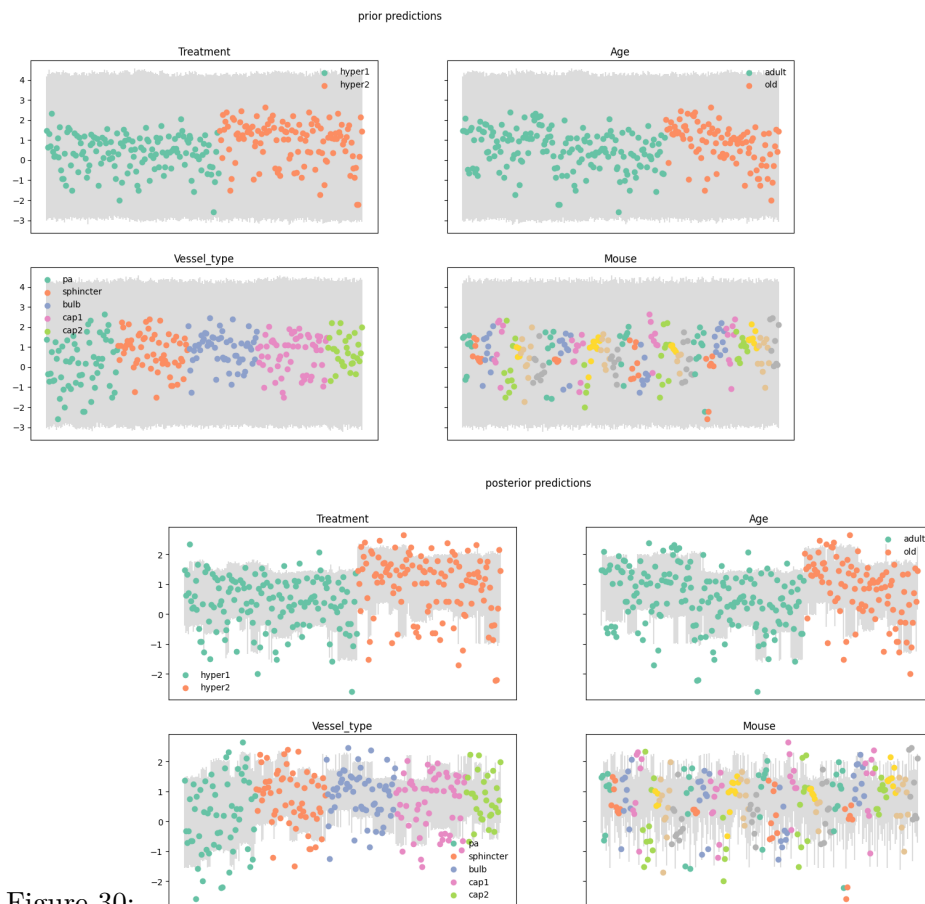


Figure 30:

- Betancourt, Michael. 2017. “Diagnosing Biased Inference with Divergences.” *Betanalpha.github.io*. [https://github.com/betanalpha/knitr\\_case\\_studies/tree/master/divergences\\_and\\_bias](https://github.com/betanalpha/knitr_case_studies/tree/master/divergences_and_bias).
- Carpenter, Bob, Andrew Gelman, Matthew D. Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2017. “Stan: A Probabilistic Programming Language.” *Journal of Statistical Software* 76 (1): 1–32. <https://doi.org/10.18637/jss.v076.i01>.
- Gelman, Andrew, Aki Vehtari, Daniel Simpson, Charles C. Margossian, Bob Carpenter, Yuling Yao, Lauren Kennedy,

- Jonah Gabry, Paul-Christian Bürkner, and Martin Modrák. 2020. “Bayesian Workflow.” *arXiv:2011.01808 [Stat]*, November. <https://arxiv.org/abs/2011.01808>.
- Groves, Teddy. 2023. “Bibat: Batteries-Included Bayesian Analysis Template.” <https://doi.org/10.5281/zenodo.7775328>.
- Juárez, Miguel A., and Mark F. J. Steel. 2010. “Model-Based Clustering of Non-Gaussian Panel Data Based on Skew-t Distributions.” *Journal of Business & Economic Statistics* 28 (1): 52–66. <https://doi.org/10.1198/jbes.2009.07145>.
- Kumar, Ravin, Colin Carroll, Ari Hartikainen, and Osvaldo Martin. 2019. “ArviZ a Unified Library for Exploratory Analysis of Bayesian Models in Python.” *Journal of Open Source Software* 4 (33): 1143. <https://doi.org/10.21105/joss.01143>.
- Niels Bantilan. 2020. “Pandera: Statistical Data Validation of Pandas Dataframes.” In *Proceedings of the 19th Python in Science Conference*, edited by Meghann Agarwal, Chris Calloway, Dillon Niederhut, and David Shupe, 116–24. <https://doi.org/10.25080/Majora-342d178e-010>.
- Piironen, Juho, and Aki Vehtari. 2017. “Sparsity Information and Regularization in the Horseshoe and Other Shrinkage Priors.” *Electronic Journal of Statistics* 11 (2). <https://doi.org/10.1214/17-EJS1337SI>.
- Pydantic developers. 2022. “Pydantic.” <https://pypi.org/project/pydantic/>.
- Stan Development Team. 2022. “CmdStanPy.” <https://github.com/stan-dev/cmdstanpy>.
- Tornqvist, Leo, Pentti Vartia, and Yrjö O. Vartia. 1985. “How Should Relative Changes Be Measured?” *The American Statistician* 39 (1): 43–46. <https://doi.org/10.2307/2683905>.
- Vehtari, Aki, Andrew Gelman, and Jonah Gabry. 2017. “Practical Bayesian Model Evaluation Using Leave-One-Out Cross-Validation and WAIC.” *Statistics and Computing* 27 (5): 1413–32. <https://doi.org/10.1007/s11222-016-9696-4>.
- Vehtari, Aki, Andrew Gelman, Daniel Simpson, Bob Carpenter, and Paul-Christian Bürkner. 2021. “Rank-Normalization, Folding, and Localization: An Improved  $\hat{R}$  for Assessing Convergence of MCMC (with Discussion).” *Bayesian Analysis* 16 (2): 667–718. <https://doi.org/10.1214/20-BA1221>.