

Sphincter analysis

Teddy Groves

Table of contents

Introduction	1
Overall strategy	2
Data	2
Data processing	3
Modelling approach	3
Software	4
Validation of statistical computation	5
Reports of findings	5
Reproducibility	6
Main findings	6
Whisker stimulation	6
Pulsatility	7
Red blood cell flow	10
Hypertensive challenge	11
Capillary density	11
Vessel tortuosity	12
Pressure	12
Baseline Diameters	14
Collaterals	14
Branchpoints	15
Details of the whisker stimulation analysis	15
Dependent variable	17
Statistical Models	17
Results	19

Details of the pulsatility analysis	21
Dependent variable	21
Questions	22
Description of the dataset	22
Statistical models	22
Results	27
Answers to specific questions	28
Details of the red blood cell flow analysis	32
Data processing	32
Dependent variable	33
Statistical models	34
Results	35
Shared parameters	35
Interactions	35
Details of the hypertensive challenge analysis	39
Dependent variable	39
Statistical model	40
Results	41
Details of the density analysis	44
Dependent variable	44
Statistical model	45
Results	46
Details of the tortuosity analysis	46
Dependent variable	48
Statistical model	48
Results	49
Details of the pressure analysis	50
Measurements	50
Statistical model	51
Results	52
Details of the diameter analysis	52
Data	52
Statistical model	53
Results	54

Details of the collaterals analysis	55
Software	57
Data processing	57
Statistical model	57
Results	58
Details of the branchpoint analysis	59
Software	59
Data processing	60
Statistical model	60
Results	60

Introduction

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

Measurement types 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)
- Capillary density
- Vessel tortuosity
- Counts and measurements of collateral vessels
- Counts and measurements of vessel branchpoints

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 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. The capillary density dataset also benefitted from a Bayesian approach, as this made it straightforward to implement smoothing of adjacent capillary orders.

After verification of computational results, our analyses aimed to evaluate whether each of our models adequately captured the information contained in the available data, and then to interpret and present the results of the best models.

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.

Data

This analysis involved three raw datasets, which we call the “ablation” dataset, the “hypertension” dataset and the “angio-architecture” dataset.

The ablation and hypertension datasets consist of 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)

The ablation and hypertension datasets contain different types of measurement. The ablation dataset includes measurements of whisker stimulation response, pulsatility and blood flow at various points of the treatment protocol. The hypertension dataset includes measurements of pressure to diameter correlation before and after two rounds of induced hypertension.

The angio-architecture dataset consists of real-valued measurements of vessel density and tortuosity.

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 abrupt 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.

In order to implement graphical predictive checking, we calculated quantiles of our models' prior and posterior predictive distributions and plotted these alongside the measurements. We then inspected the graphs to ascertain whether the measurements were generally consistent with the predictions.

We calculated estimated leave-one-observation-out log predictive densities using the Python package `arviz` (Kumar et al. 2019).

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.

Statistical computation was carried out using the probabilistic programming framework Stan (Carpenter et al. 2017) via the interface cmdstanpy (Stan Development Team 2022). For the analyses of collaterals and branchpoints, we used pymc (Abril-Pla et al. 2023) via the interface bambi [capretto_bambi_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.

Reports of findings

In general the findings we report were cases where our best performing model’s posterior probability distribution gave a high probability to an interesting and interpretable quantity having an interesting value. Since most of the interpretable parameters in our models are normalised so that zero indicates no

effect, in practice our main findings are mostly that a parameter, or the difference between some parameters, is likely to be substantially different from zero according to our best model's posterior distribution. We present results with this form using histograms of the relevant marginal posterior distributions.

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.

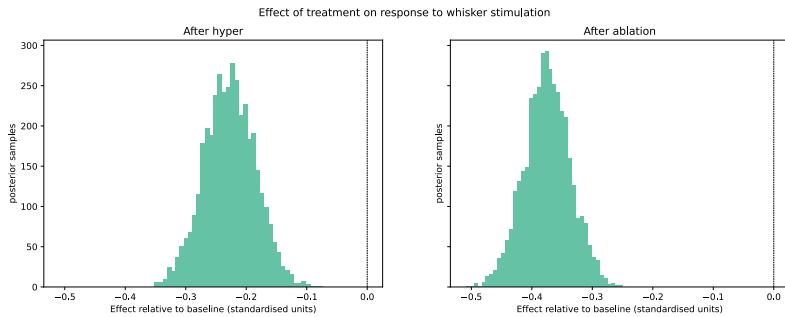


Figure 1: Marginal posterior histograms for treatment effects, relative to the baseline treatment.

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.

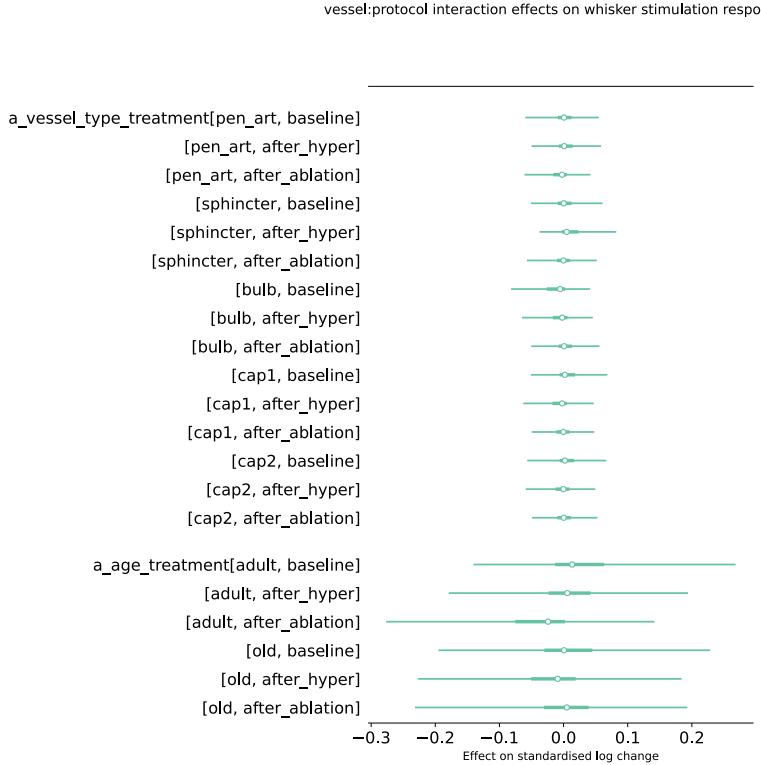


Figure 2: Marginal 2.5%-97.5% posterior intervals for protocol effects

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.

Pulsatility

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

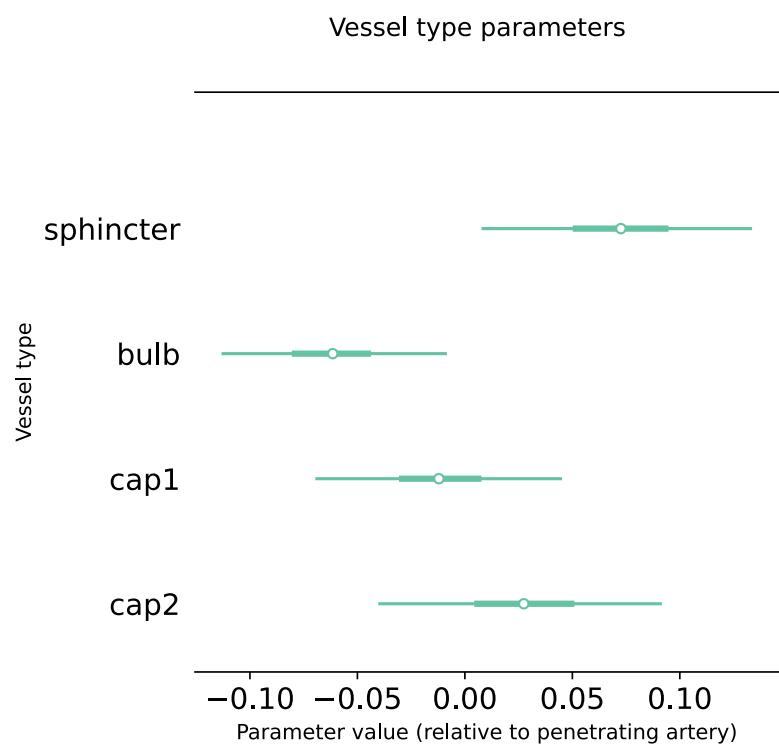


Figure 3: Marginal 2.5%-97.5% posterior intervals for vessel type effects

- 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, according to our model older mice 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.

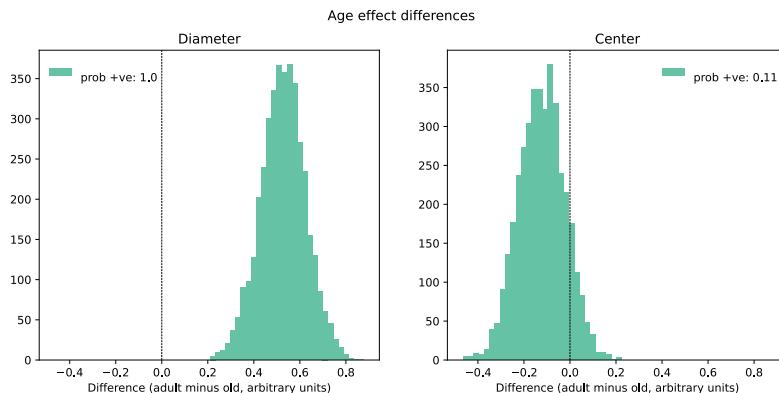


Figure 4: Posterior distribution of age effect differences for each measurement type.

Figure 27 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.

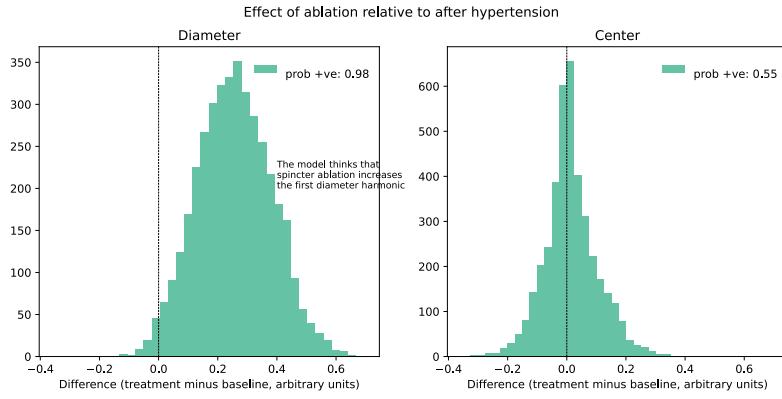


Figure 5: Posterior distribution of treatment effect differences for each measurement type.

Red blood cell flow

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

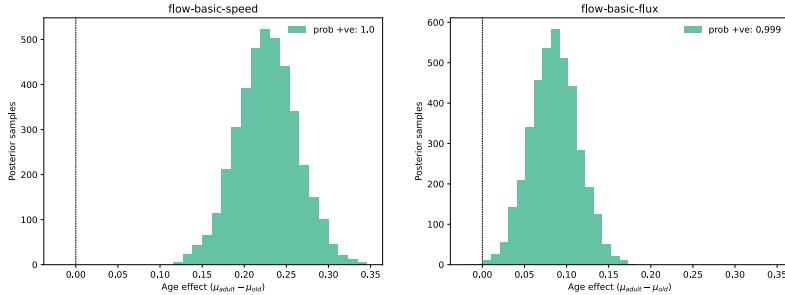


Figure 6: Posterior distributions of age effects on red blood cell speed and flux.

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

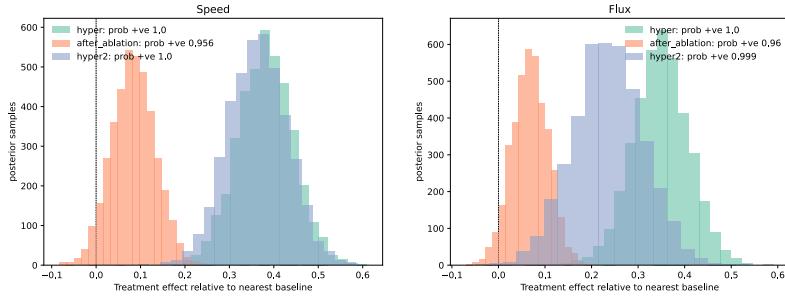


Figure 7: Posterior distributions of treatment effects on red blood cell speed and flux. The nearest baseline for treatment `hyper` is `baseline`, and for treatments `after_ablation` and `hyper2` it is `after_hyper`.

Hypertensive challenge

Our hypertensive challenge data also showed pronounced age and treatment differences, as shown in figure Figure 8. Specifically, we found that, for the adult mice, blood pressure and vessel diameter were less correlated, and that the `hyper2` treatment was associated with increased pressure-diameter correlation compared with the `hyper1` treatment.

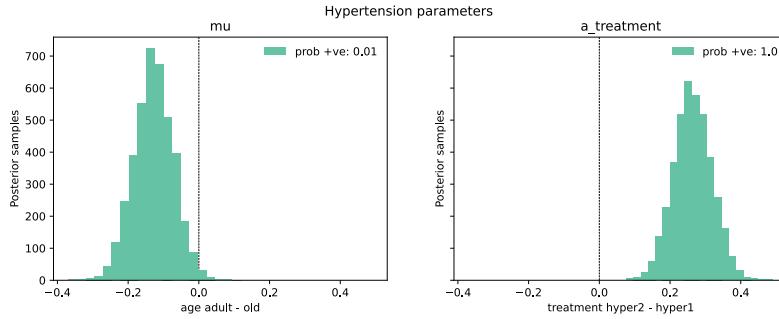


Figure 8: Posterior distributions of relative age and treatment effects for hypertensive challenge data.

Capillary density

The capillary density dataset had a somewhat different structure to the other data, with no treatments, more vessel types,

with clear correlation between measurements corresponding to adjacent vessel types. We therefore used a different statistical approach, with smoothing components for parameters of adjacent vessel types. See Section for more about this model.

Our analysis indicated that the old mice tended to have lower density than the adult mice for capillaries of order 1 to 5 and higher density for capillaries of order 10 to 12, and for ascending venules, as shown in figure Figure 9.

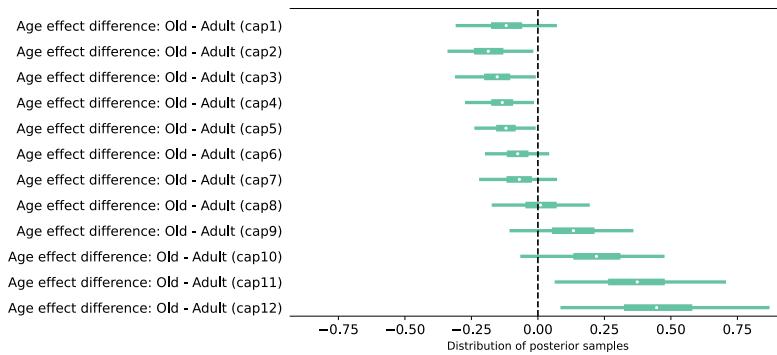


Figure 9: Posterior distributions of differences in age-dependent parameters by vessel type.

Vessel tortuosity

We fit a smoothed regression model to measurements of vessel tortuosity from the angio-architecture dataset. See Section for details about this model.

As shown in Figure 10, our analysis showed a clear difference between the adult and old mice in the upper vasculature, with the old mice tending to have substantially more tortuous pial arteries and penetrating arterioles. Other vessels were similarly tortuous for both age categories.

Pressure

We modelled measurements of three variables related to blood pressure—mean arterial pressure (MAP), pulse pressure (PP)

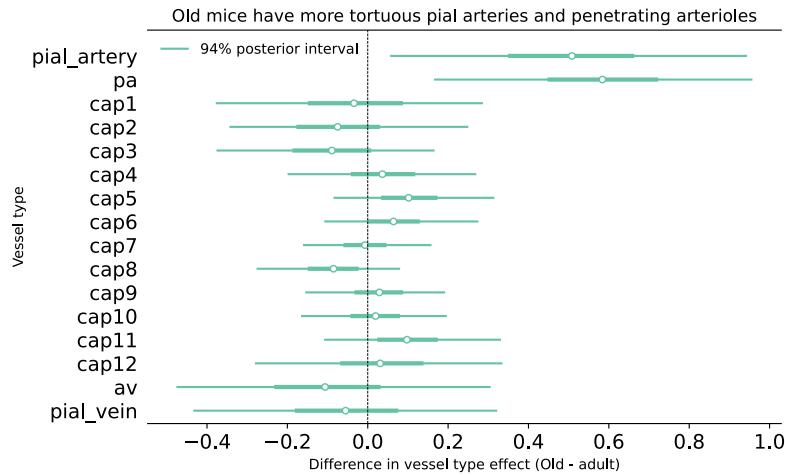


Figure 10: Posterior distributions of differences in age-dependent tortuosity effects by vessel type

and heart rate (HR)—aiming to quantify any treatment and age effects.

The main results of our model are shown in Figure 11.

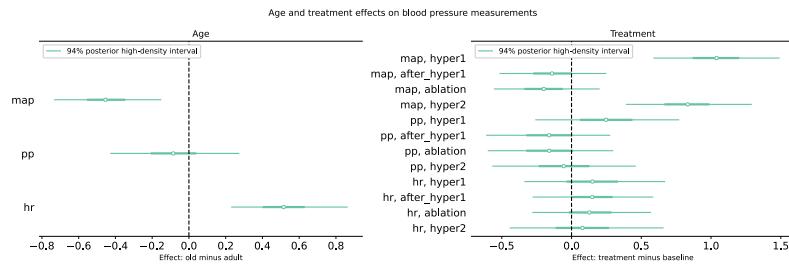


Figure 11: Posterior distributions of age and treatment effects in our models of blood pressure measurements.

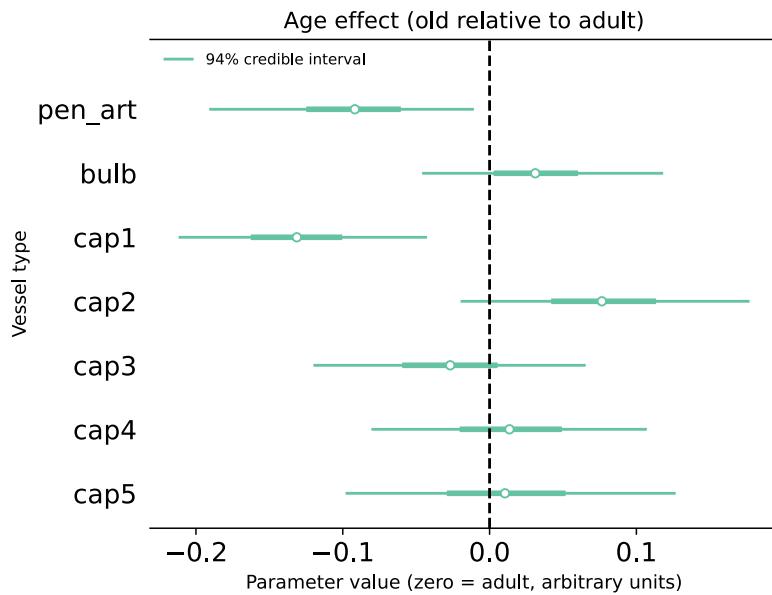
Compared with young mice, old mice tended to have lower mean arterial pressures, similar pulse pressures and higher heart rates.

Our MAP model allocated clear positive effects to the treatments “hyper1” and “hyper2” relative to the baseline treatment whereas other treatment/measurement pairs were not clearly differentiated from the baseline.

Baseline Diameters

We modelled measurements of baseline vessel diameters, aiming to quantify general and vessel-type specific age effects.

Our results are shown in Figure 12: old mice showed smaller diameters of penetrating arterioles and first order capillaries, while other vessel types tended to have similar diameters in adult and old mice.



Marginal posterior distributions of aggregate age effects on modelled vessel diameters for the measured vessel types.
Penetrating arterioles and first order capillaries tended to have smaller diameters in old mice.

Figure 12: The aggregate effect was found by adding the general age-specific effect to the vessel-type specific age effect, then subtracting the adult value from the old value. See notebook “Diameter.ipynb” for details.

Collaterals

We modelled the density of collateral vessels in adult and old mice, finding that old mice tended to have fewer collaterals, as

shown in Figure 13.

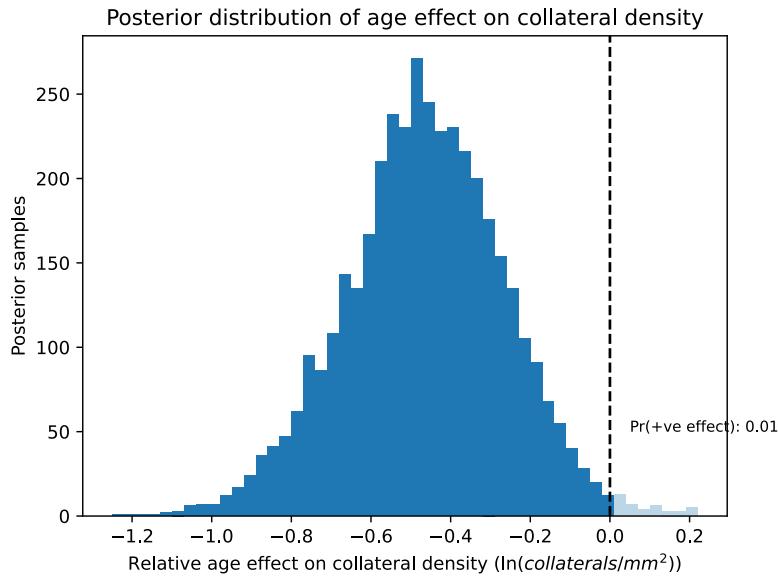


Figure 13: Posterior distribution of differences in age effects on collateral density.

Applying a similar analysis to measurements of collateral diameters, lengths and tortuosities, we found negligible age effects: see Figure 14

Branchpoints

Our models of whether a branchpoint contained a bulb or sphincter indicated no particular age effect for sphincters, and a clear tendency for branchpoints in old mice to have more bulbs. These results are shown in Figure 15.

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

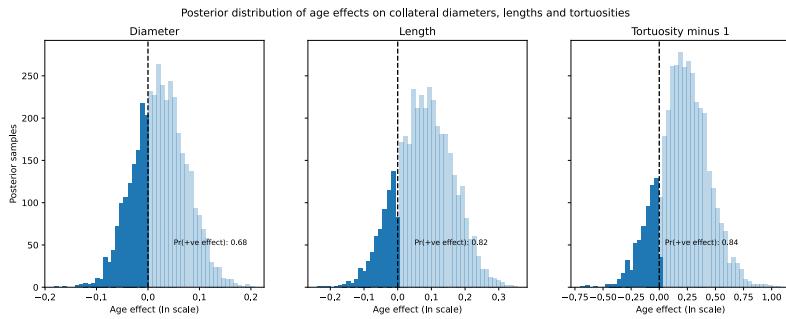


Figure 14: Posterior distribution of differences in age effects on collateral density.

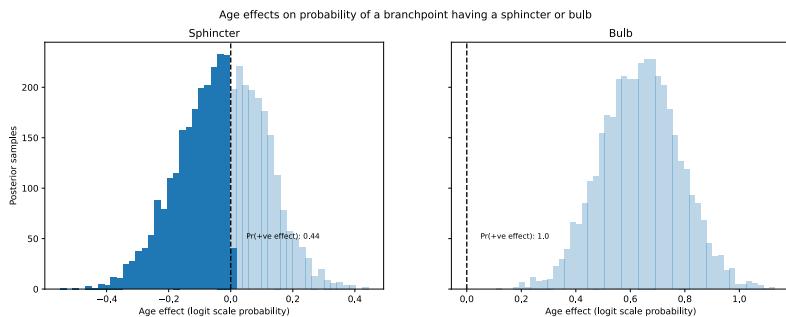


Figure 15: Posterior distribution of differences in age effects on branchpoint sphincter and bulb probability.

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 v_1 and v_2 is far less than 1, the log change $\ln \frac{v_2}{v_1}$ is approximately the same as the more widely used relative difference measure $\frac{v_2-v_1}{v_1}$.

Statistical Models

The final model is shown below:

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

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 (α^{vessel}) and age (μ). 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 heterogenous—in this case the corresponding τ parameter will be large—and on the other hand high similarity, leading to small τ parameters. The student t distribution was chosen to reflect the heavy tails we observed in the data, with the parameter ν 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 ν 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):

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

Results

Figure 16 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 17 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.

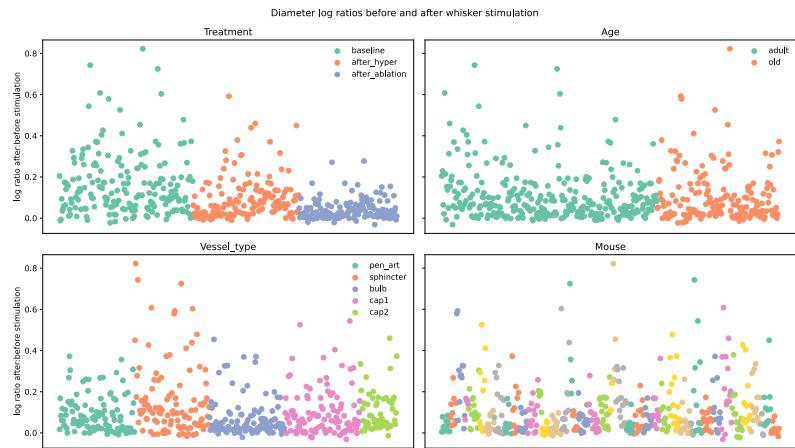


Figure 16: Raw measurements

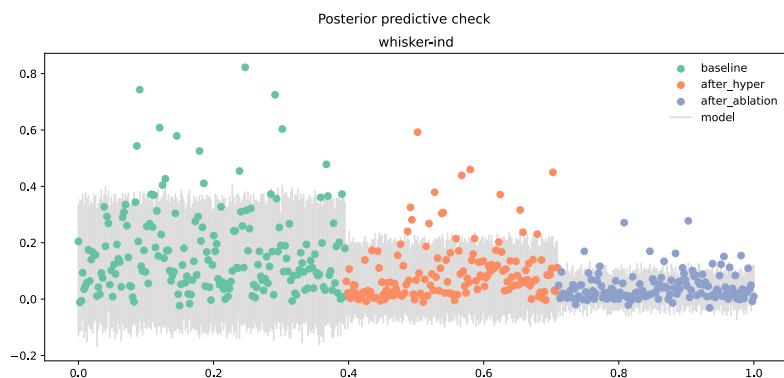


Figure 17: Graphical posterior predictive check

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 18:

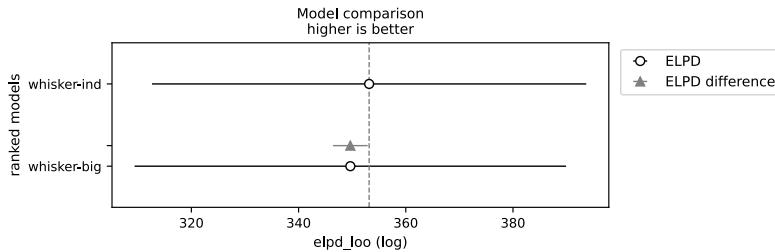


Figure 18: Comparison of estimated leave-one-out log predictive density between the final model `whisker-ind` and the best performing interaction model `whisker-big`.

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 19.

Figure 20 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 21 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

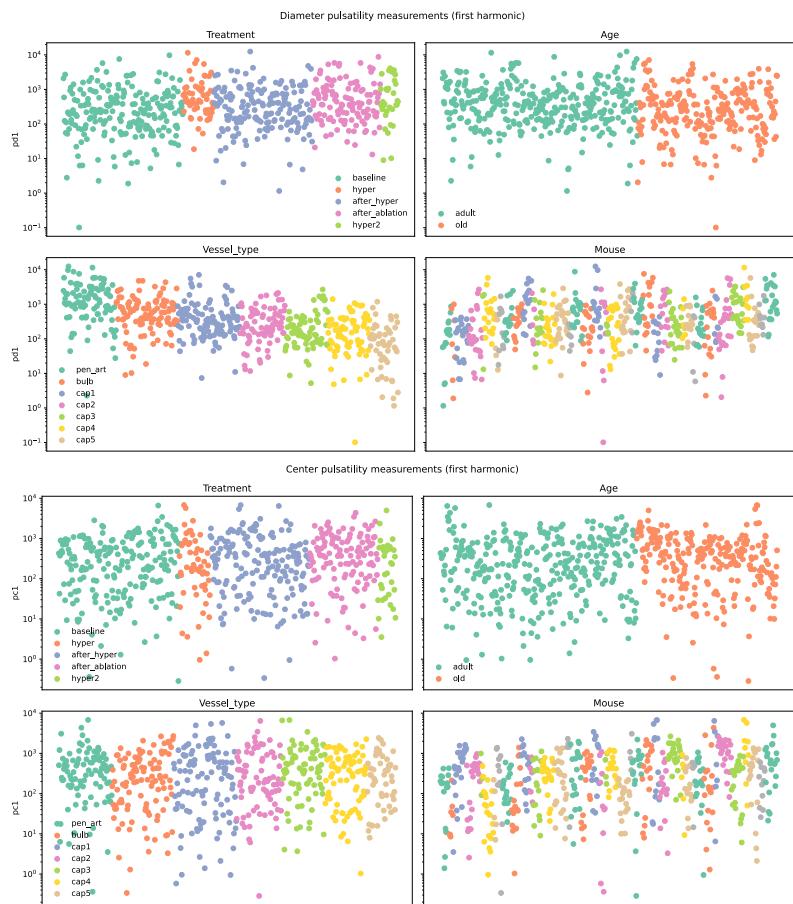


Figure 19: The modelled measurements, shown in order of the coloured categories.

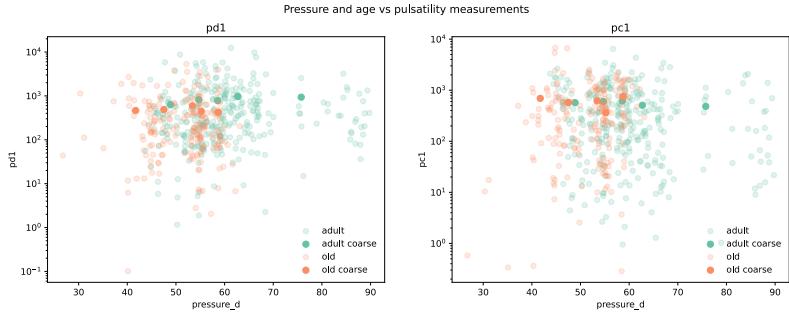


Figure 20: Pulsatility measurements plotted against the corresponding pressure measurements and coloured according to age. Darker dots indicate averages within evenly sized pressure bins.

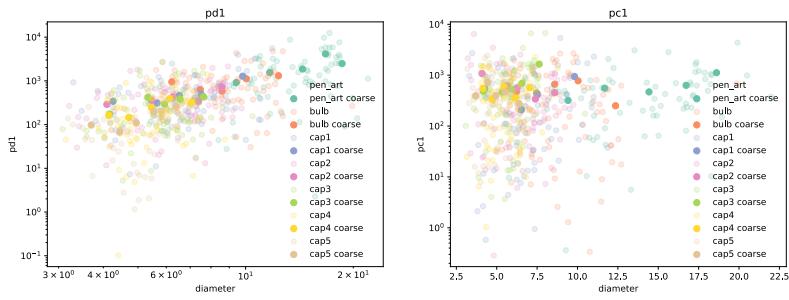


Figure 21: Pulsatility measurements plotted against the corresponding diameter measurements and coloured according to vessel type. Darker dots indicate averages within evenly sized pressure bins.

and linear predictor η the measurement probability density is given by this equation:

$$p(y | \eta) = \text{Exponential}(y, \lambda) \quad (3)$$

$$= \lambda e^{-\lambda y}$$

$$\ln \frac{1}{\lambda} = \eta \quad (4)$$

The log link function (4) was chosen so that linear changes in the term η 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 η 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 η_{vtad}^{basic} for an observation with vessel type v , treatment t , age a and diameter d as follows:

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

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:

$$\eta_{vatad}^{interaction} = \mu_a + \alpha_t^{treatment} \quad (6)$$

$$+ \alpha_{d,vesseltype(n)}^{vesseltype} + \alpha_{tv}^{treatment:vesseltype} + \beta_d^{diameter} \cdot \ln d \quad (7)$$

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 $\eta_{vatdp}^{pressure no age}$ are calculated as shown in equation (9). Note that, unlike in equation (8), the μ 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 α parameters were given independent, semi-informative, hierarchical prior distributions to allow for appropriate information sharing. The β and μ 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 22.

We evaluated our models' fit to data using prior and posterior predictive checking, with the results for the pressure model shown in Figure 46.

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 24.

From this result, together with the worse estimated out of sample predictive performance as shown in Figure 22, we concluded that there were no important interaction effects, so that we could essentially discard the interaction model.

Figure 25 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.

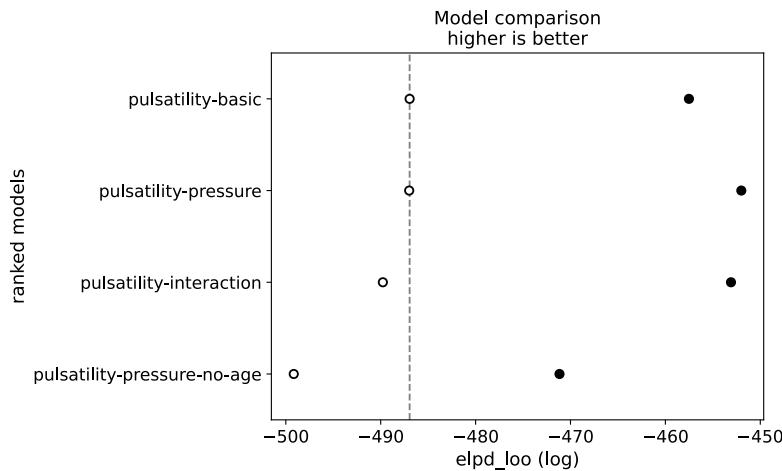


Figure 22: Comparison of estimated leave-one-out log predictive density (ELPD) for our pulsatility models. The main result is that the pressure-no-age and interaction models are clearly worse than the pressure model, as shown by the separation of the relevant grey and dotted lines.

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 22, 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 26 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

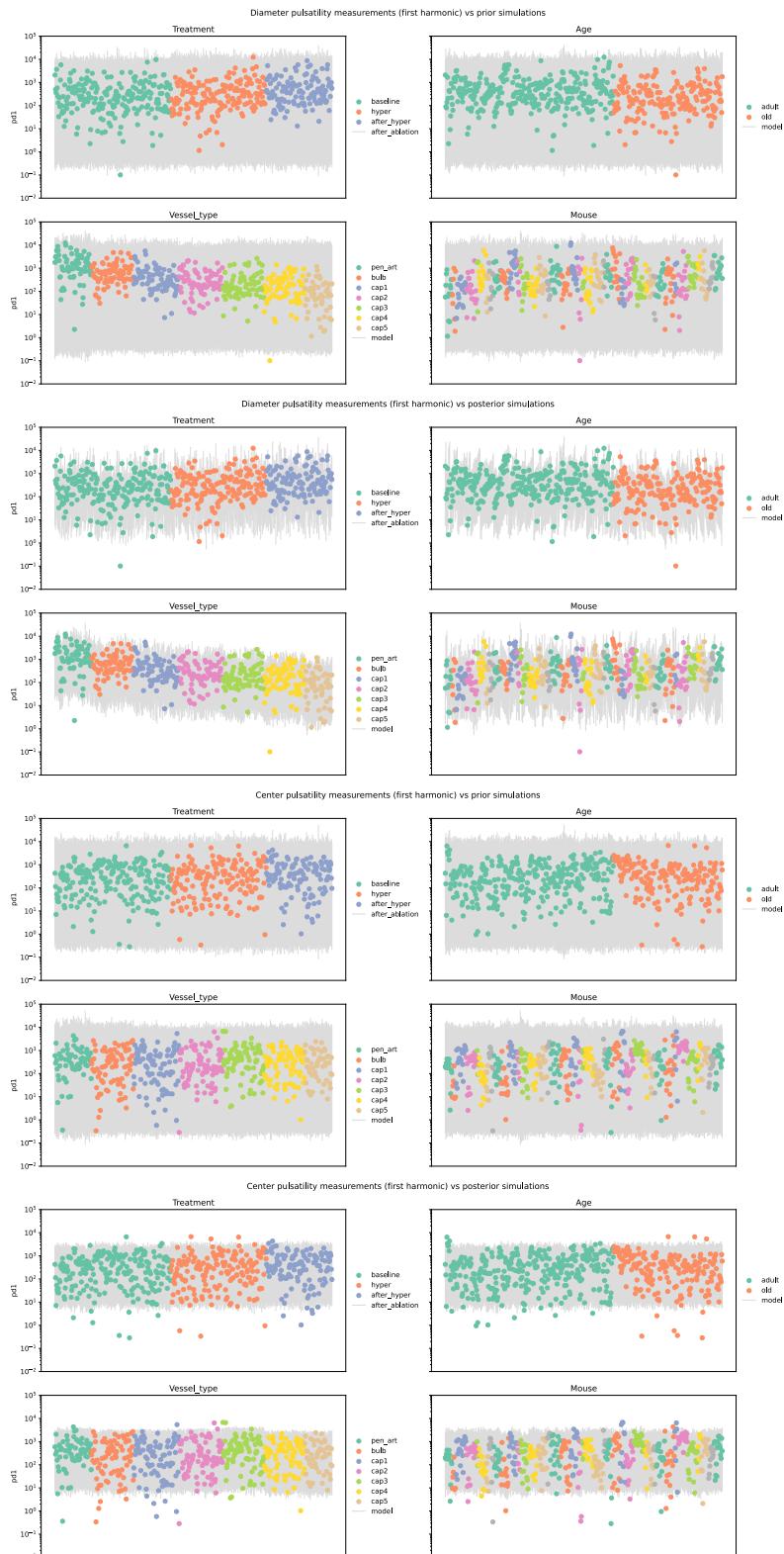


Figure 23: Prior and posterior predictive checks for the pressure model.

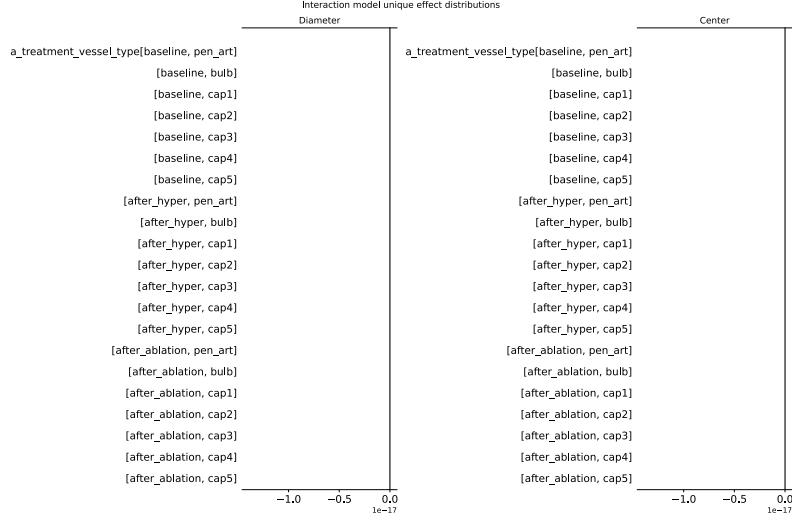


Figure 24: Marginal posterior quantiles for the unique effects in the interaction model.

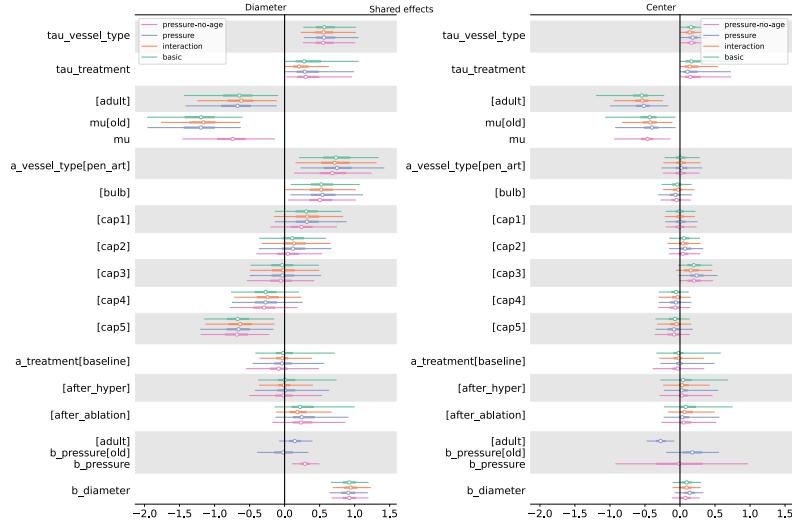


Figure 25: Marginal posterior quantiles for shared model effects.

for adult mice but not for old mice.

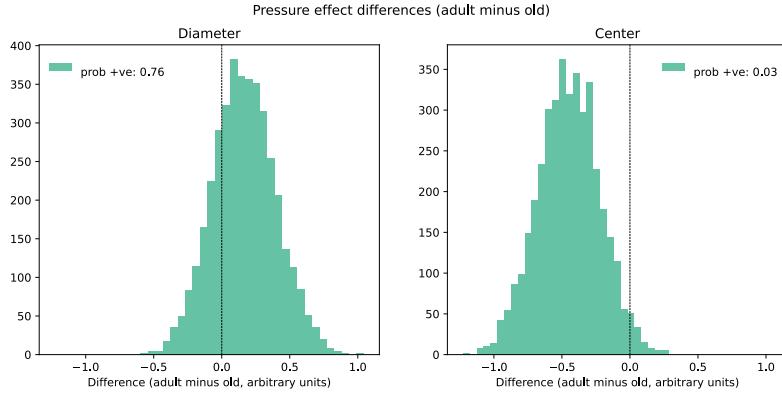


Figure 26: Posterior distribution of pressure effect differences for each measurement type.

To illustrate the effect of treatments, and specifically sphincter ablation relative to hypertension (i.e. to answer Question 0.2) Figure 27 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.

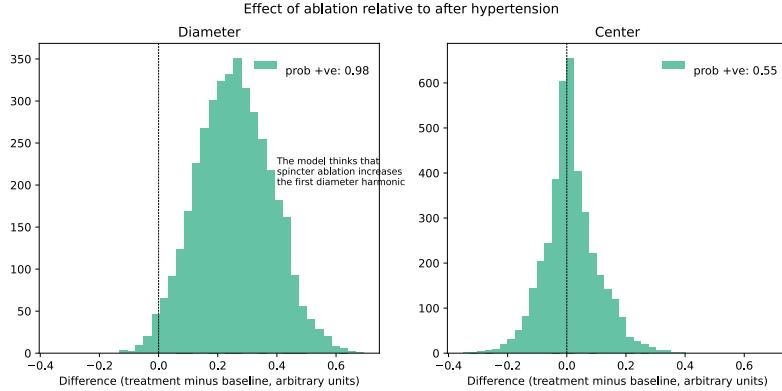


Figure 27: Posterior distribution of treatment effect differences for each measurement type.

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 28 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.

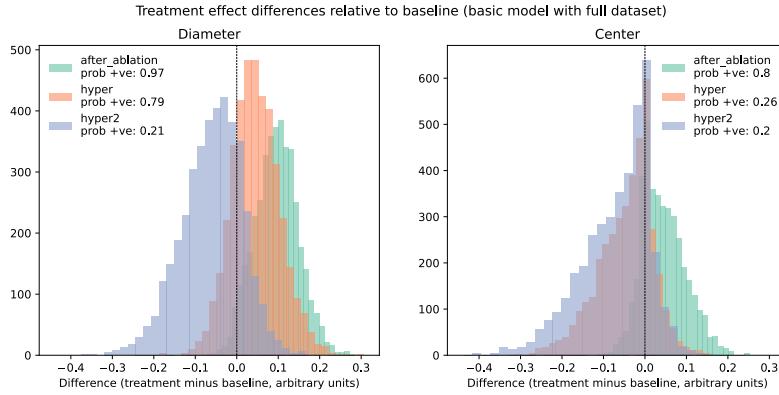


Figure 28: Treatment effect distributions relative to baseline in the basic model when fit to the full dataset including all treatments.

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 29.

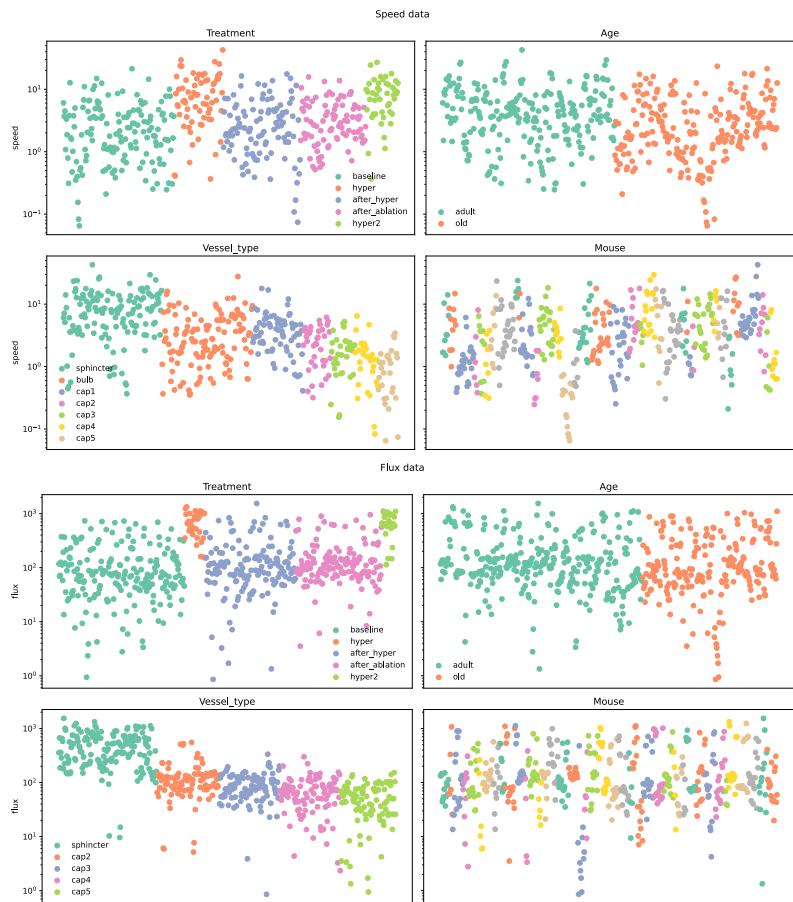


Figure 29: Red blood cell speed and flux data

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:

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

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 30

Shared parameters

Figure 31 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 to the vessel type sphincter. According to our models, the sphincter has 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 32.

For completeness the interaction effects are shown in figure Figure 33. No effects are clearly separated from zero. The sphincter/ablation effect on RBC speed is notably different from the

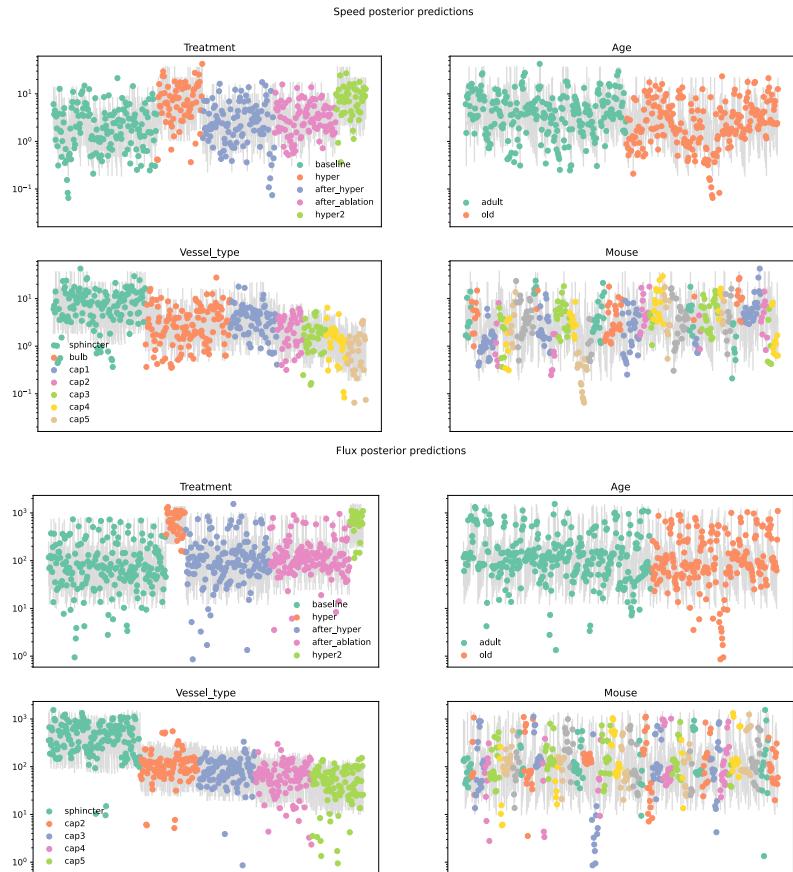


Figure 30: Posterior predictive checks for flow models

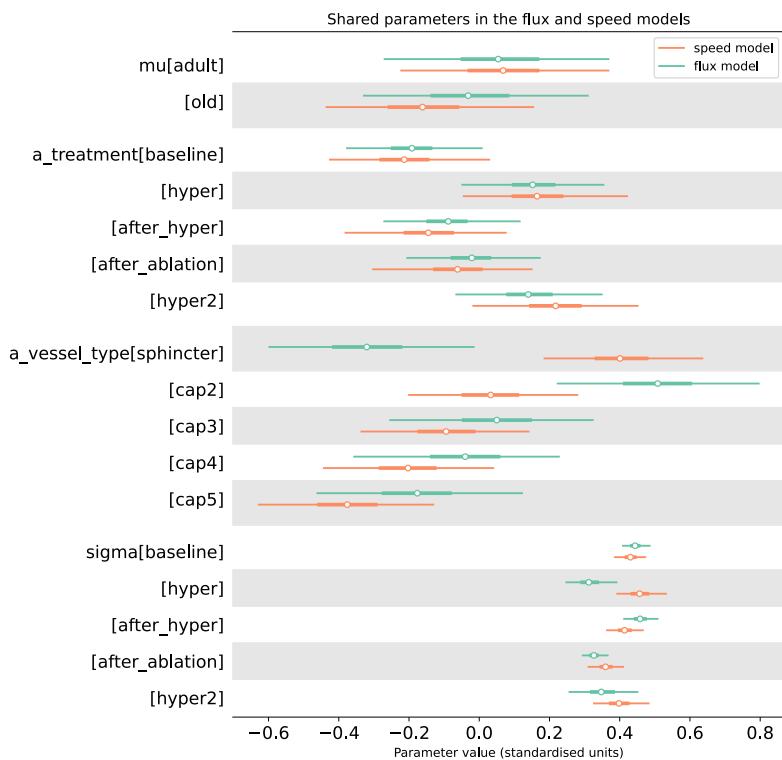


Figure 31: Posterior distributions of parameters that appear in both our flow and speed models.

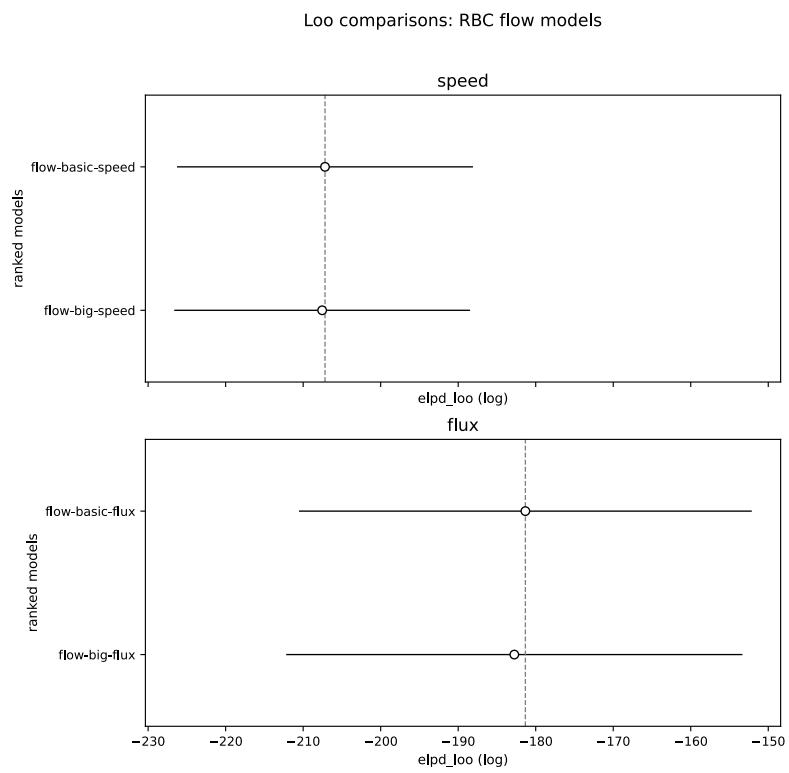


Figure 32: Out of sample predictive performance comparison for red blood cell flow models.

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.

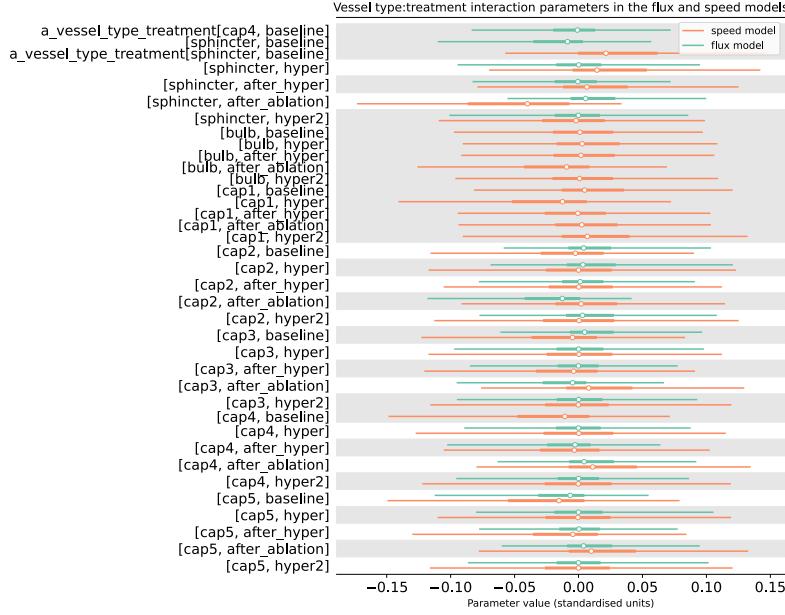


Figure 33: Interaction effects on red blood cell flow.

Details of the hypertensive challenge analysis

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 34. The transformed correlation coefficients do not appear extremely dispersed, in-

dicating that standard modelling techniques ought to be able to describe them.

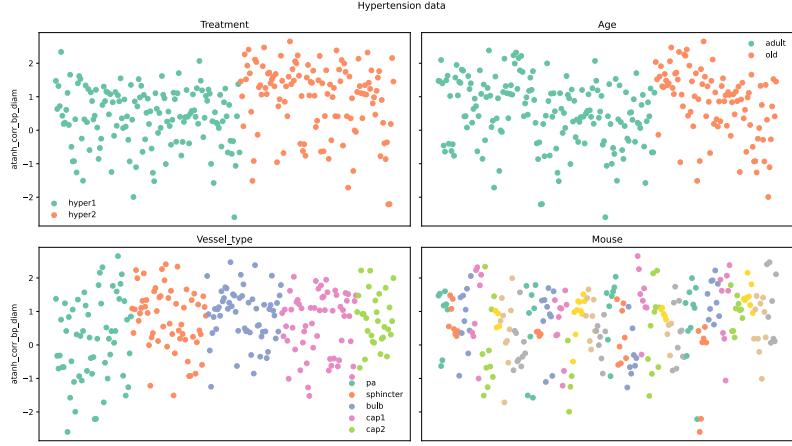


Figure 34

Statistical model

Our final statistical model had the following form.

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

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 σ 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 35 shows that, as in the other cases, including interaction effects did not improve estimated predictive performance. The expected leave-one-observation-out predictive density is estimated to be higher for the model **hypertension-basic**, which did not have any interaction effects, than for the interaction model **hypertension-big**. Furthermore, the standard error for the difference in these two quantities is much smaller than the difference itself, indicating that the ranking is robust.

Figure 36 shows the marginal distributions for the non-hierarchical parameters in our final model. As shown in Figure 8, there are clear differences between the different **mu** and **a_treatment** parameters, indicating that there are important age and treatment effects. There is also a difference between parameters corresponding to penetrating arterioles and the other vessel type parameters. This reflects the fact that the penetrating arterioles had lower pressure vs diameter correlation coefficients than other vessels, and that the penetrating arteriole measurements were more dispersed.

Figure 37 and Figure 38 show 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 more observations lie outside the plotted intervals than might be expected.

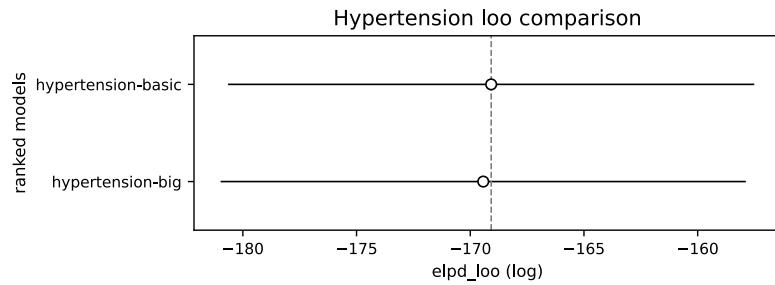


Figure 35: Comparison of out-of-sample predictive performance of our hypertension models, as measured by estimated leave-one-out expected log predictive density. The two models have similar estimated performance, but the **hypertension-big** model is clearly worse.

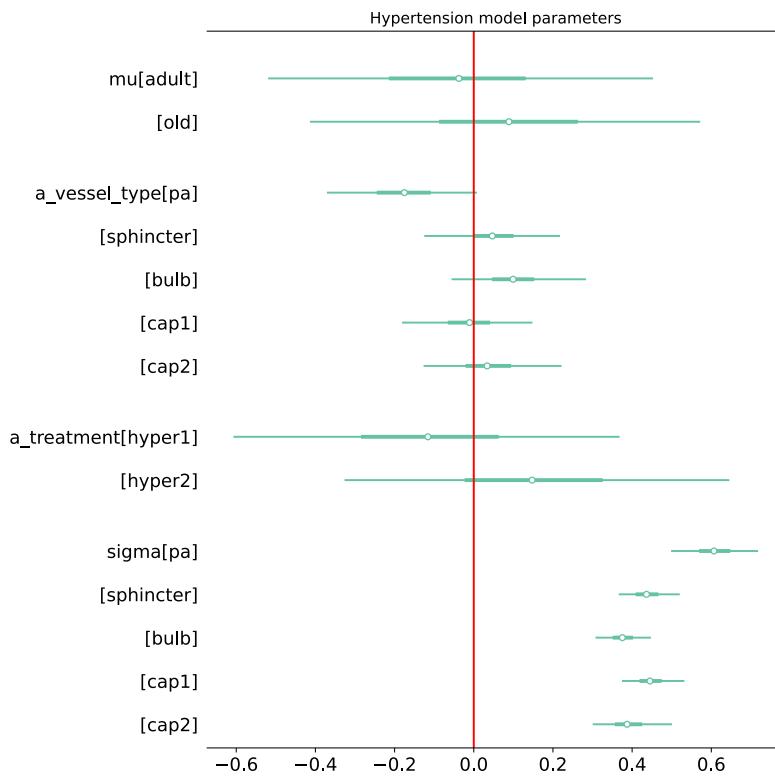


Figure 36: 1%-99% posterior intervals for parameters in our final hypertension model.

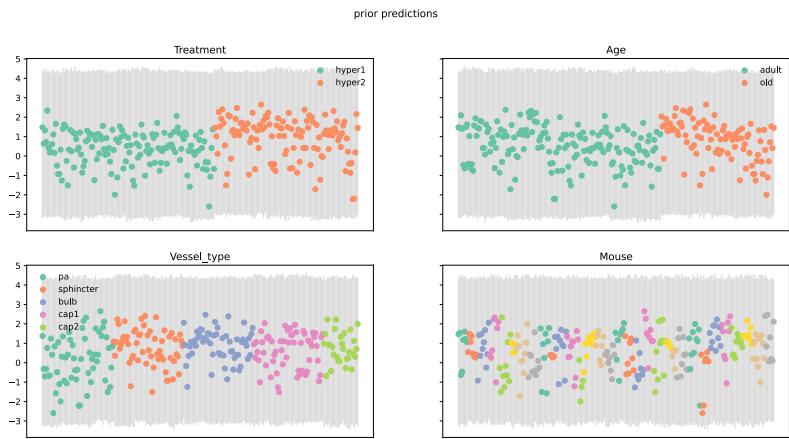


Figure 37

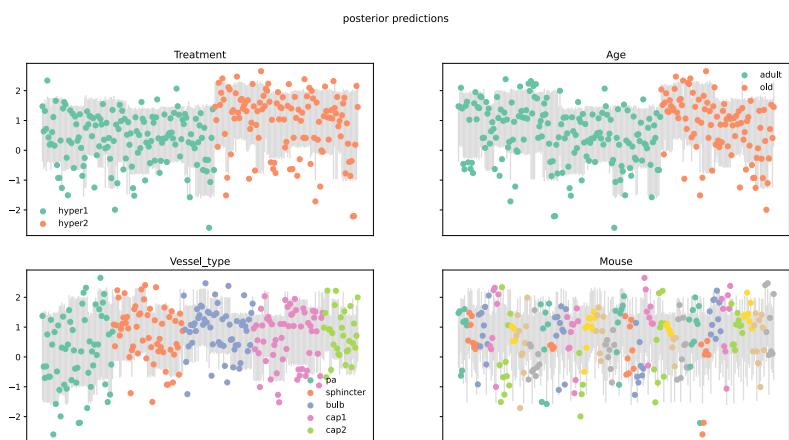


Figure 38

Details of the density analysis

The density dataset consisted of 144 measurements from five adult and four old mice.

Dependent variable

The dependent variable in this case was capillary density, measured as in length of vessel per unit of volume. These measurements are shown in Figure 39 on both natural and logarithmic scales.

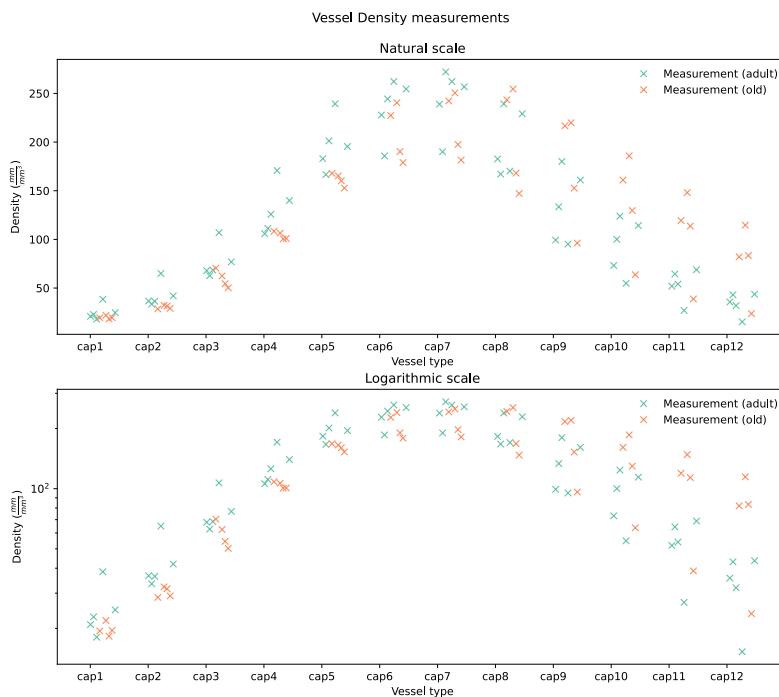


Figure 39: Vessel density measurements

We noticed several interesting patterns in this data.

First, there is a clear trend for capillary density to increase with vessel order until the cap6 vessel type and then to decrease, with adjacent vessel types tending to have similar densities.

Second, we noticed that the measurements are somewhat more dispersed for higher-order capillaries, with the level of dispersion increasing somewhat smoothly.

Statistical model

We modelled the measurement process using a linear model on logarithmic scale. In this model, given measurement y , linear predictor \hat{y} and a vessel type specific measurement error parameter σ , the measurement probability density is given by this equation:

$$p(y | \alpha, \sigma) = \text{Normal}(\ln y | \alpha, \sigma) \quad (12)$$

In order to capture the observed smoothness between adjacent vessel types, we used Gaussian random walk priors for the $\alpha^{age:vesseltype}$ parameters:

$$\alpha_{a,v} \sim N(\alpha_{a,v-1}, \lambda_a^\alpha) \quad (13)$$

$$(14)$$

We also used a Gaussian random walk prior on log standardised scale for the measurement error parameter σ , where $\sigma^{std} = \frac{\sigma}{sd(\ln y)}$:

$$\ln \sigma_v^{std} \sim N(\ln \sigma_{v-1}^{std}, \lambda^s) \quad (15)$$

$$(16)$$

This approach to smoothing parameters corresponding to ordered categories is essentially the same as that used in Gao et al. (2019) to model age effects on voting behaviour. As explained in that paper, the random walk priors allow for information sharing between categories, without the need for detailed assumptions about the functional form of the overall relationship.

The other priors in our model were as follows (units are on standardised logarithmic scale):

$$\begin{aligned} \alpha_1 &\sim N(0, 0.1) \\ \ln \sigma_1^{std} &\sim N(-1.5, 1) \\ \lambda^\alpha &\sim N(0, 0.3) \\ \lambda^\sigma &\sim N(0, 0.3) \end{aligned} \quad (17)$$

(18)

Results

Figure 40 shows the overall fit of our model to the observed data. We judged that the overall fit to the observed measurements was adequate and did not attempt model evaluation using estimated leave-one-out density as given the highly correlated data it would be difficult to perform the necessary estimates with acceptable accuracy. In particular, our model captured the increase and decrease in density with increasing capillary order for both age categories, and the higher dispersion at higher capillary orders.

Figure 41 shows posterior 94% credible intervals for the quantity $\alpha_2 - \alpha_1$ for each vessel type; in other words, the overall difference, on standardised log scale, in densities between old and adult mice for each vessel type.

There is a clear trend, with four low order capillary types separated from zero on the left and two high order types separated on the right.

Details of the tortuosity analysis

Like the density analysis, the tortuosity analysis modelled the architecture dataset.

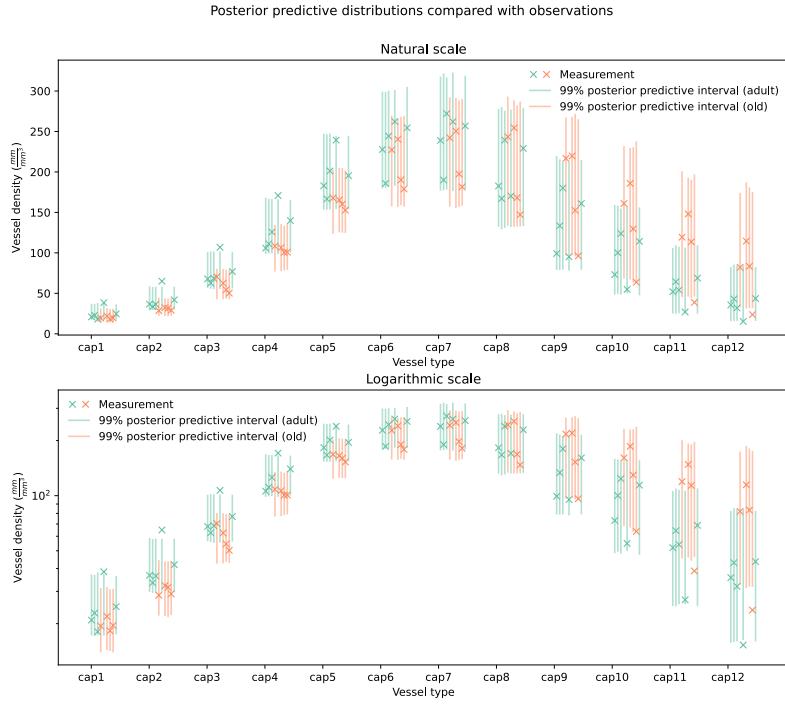


Figure 40: Posterior predictive check for our final vessel density model, shown on natural and logarithmic scale.

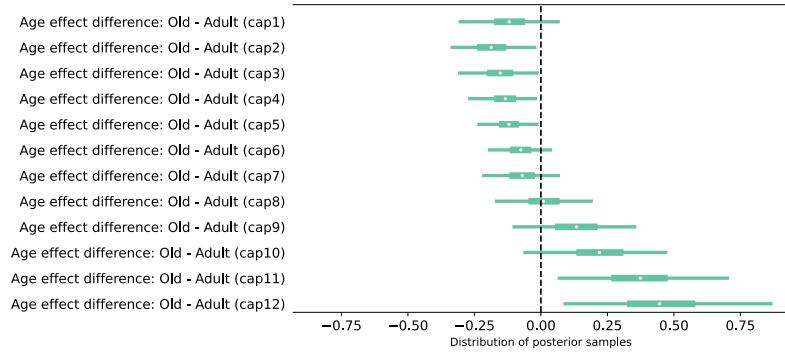


Figure 41: Main result of our density analysis: old and adult mice have different vessel density patterns.

Dependent variable

The dependent variable in this case was vessel tortuosity, measured as total vessel length divided by straight-line length.

The raw measurements are shown in Figure 42.

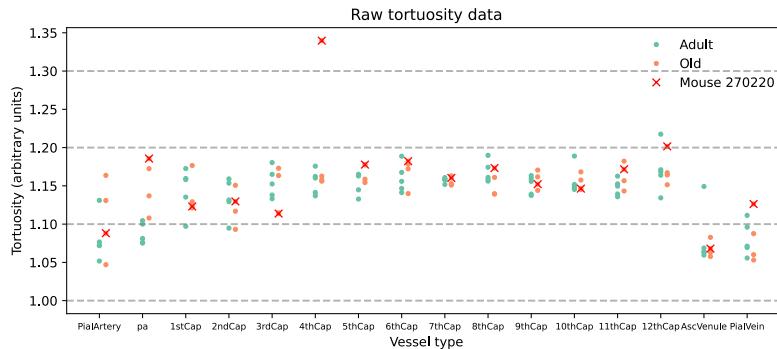


Figure 42: Vessel tortuosity measurements

There is a clear outlier for mouse 270220 at the 4th capillary: we excluded this measurement but kept other measurements from this mouse as they did not seem particularly anomalous.

Otherwise, we observed a similar pattern of smoothness between adjacent vessel as with the density measurements, motivating the use of a model with smoothed latent parameters. Unlike with the density measurements, however, there is some variety in the size of the jumps from vessel type to vessel type. In particular, the adult mice show a sharp upward jump from vessel pa to the first capillary, and both the adult and old mice show a downward jump in tortuosity at the ascending venule.

Statistical model

As with the density dataset, we used a regression model with random-walk priors to smooth adjacent vessel type effects within age categories and measurement errors for adjacent vessel types. In order to appropriately capture the heterogeneity in jump sizes, we used a student-t distribution rather than a normal distribution for the random walk in vessel

type effects, with the degrees of freedom parameter modelled hierarchically.

The model specification is as follows, with $\ln y^{std}$ representing the log-transformed and then standardised tortuosity measurements:

$$\begin{aligned}
 p(y | \alpha, \sigma) &= Normal(\ln y^{std} | \alpha, \sigma) \\
 \alpha_{a,v} &\sim ST(\nu, \alpha_{a,v-1}, \lambda_a^\alpha) \\
 \ln \sigma_v^{std} &\sim N(\ln \sigma_{v-1}^{std}, \lambda^s) \\
 \nu &\sim \Gamma(2, 0.1) \\
 \alpha_{,0} &\sim N(0, 1) \\
 \ln \sigma_0^{std} &\sim N(-1.5, 0.8) \\
 \lambda &\sim N(0, 0.2) \\
 \lambda^s &\sim N(0, 0.2)
 \end{aligned} \tag{19}$$

Results

Figure 43 shows this model's posterior predictive distribution alongside the retained measurements, indicating a reasonably good fit.

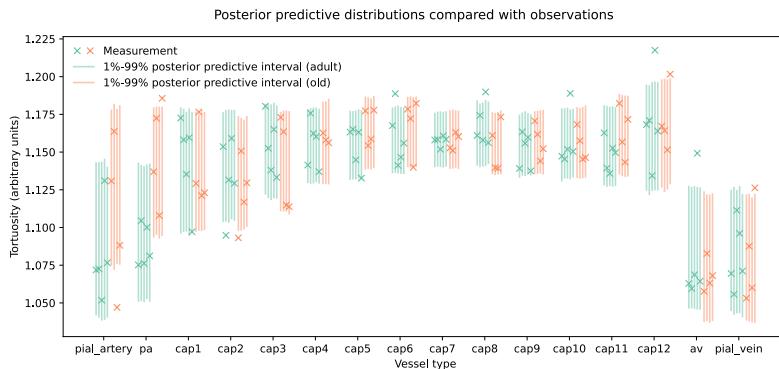


Figure 43: Marginal posterior predictive distributions alongside tortuosity measurements

The posterior predictive distributions for upper vessels “pial artery” and “pa” appear different for old and adult mice. To

ascertain the extent and certainty of this contrast we plotted the differences between vessel type effects for old and adult mice in Figure 44. From this plot it is clear that, according to our model, there is a pronounced and robust difference, with over 97% probability of each effect being greater for old mice.

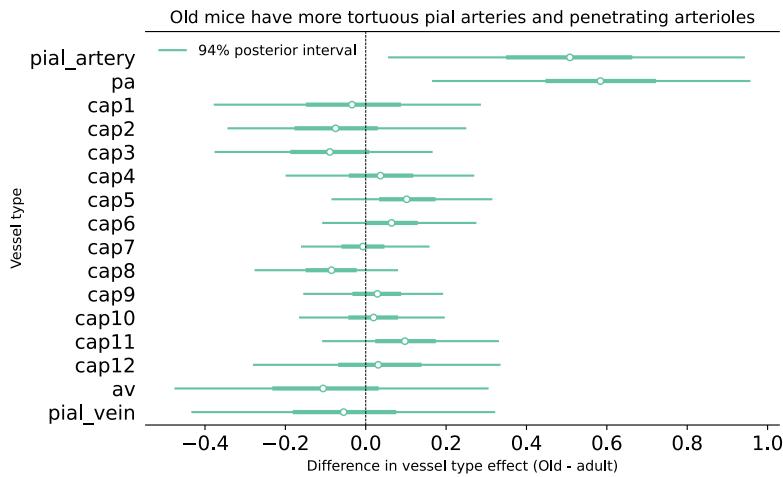


Figure 44: Main result of our tortuosity analysis: the vessel type effect for tortuosity is greater for old mice than for adult mice.

Details of the pressure analysis

Measurements

To model blood pressure we used a new dataset consisting of 140 measurements of 24 adult and 15 old mice. The mice underwent the following treatment regime:

1. baseline
2. hyper1
3. after_hyper1
4. ablation
5. hyper2

Each measurement records the mean arterial pressure (acronym MAP, units mmHg), pulse pressure (acronym PP, units mmHg)

and heart rate (HR, units Hz) at a single point during this regime. The distribution of measurements across ages and treatments was as follows:

treatment	adult	old
baseline	14	12
hyper1	10	11
after_hyper1	22	13
ablation	21	13
hyper2	13	11

These measurements are shown in Figure 45

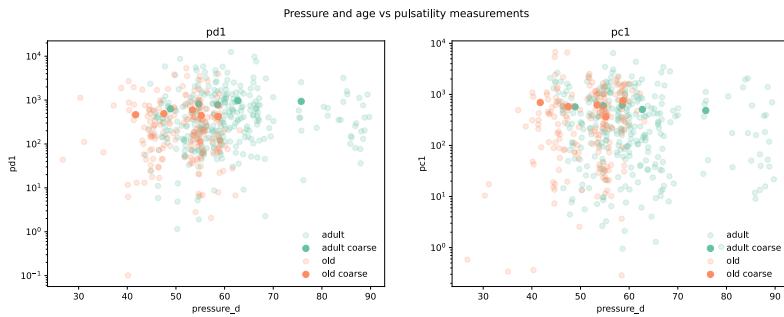


Figure 45

The most notable feature of these data is that the data are clearly more dispersed for some treatments than for others. For example, for all three measurement types, the measurements for treatment “hyper1” are more dispersed than for treatments “baseline” and “after_hyper1”.

Statistical model

For modelling we log-transformed and standardised the measurements, then fit Bayesian linear models with fixed effects on the dependent variable for age, treatment and age-treatment interaction, as well as fitting treatment-specific measurement errors to capture the observed heteroskedasticity.

We set informative priors for all of our model parameters based on our knowledge of the likely size of any effects.

In tilde notation, the model for a measurement i of a mouse with age a at treatment stage t was as follows:

$$\ln y_i \sim N(\hat{y}_{at}, \sigma_t) \quad (20)$$

$$\hat{y}_{at} = \alpha_a^{age} + \alpha_t^{treatment} + \alpha_{at}^{age:treatment} \quad (21)$$

$$\alpha_a^{age} \sim N(0, 0.3) \quad (22)$$

$$\alpha_t^{treatment} \sim N(0, 0.3) \quad (23)$$

$$\alpha_{at}^{age:treatment} \sim N(0, 0.2) \quad (24)$$

$$\sigma_t \sim HN(0, 0.5) \quad (25)$$

Results

Figure 46 shows the measurements with our model's marginal posterior predictive distributions. The fit is overall adequate apart from a few outlier measurements under the hyper1 and hyper2 treatments. In particular, our model successfully captured the treatment-based heteroskedasticity in the data.

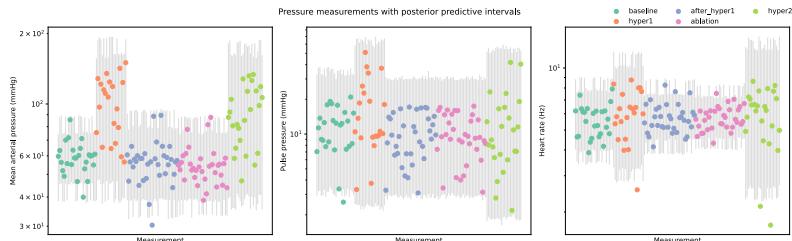


Figure 46: Grey bars represent our model's 1% to 99% posterior predictive interval for each measurement.

Details of the diameter analysis

Data

For the diameter analysis we used the same dataset and processing as for the pulsatility analysis.

Statistical model

We used a lognormal generalised linear model implemented in Stan: see the file `sphincter/stan/diameter.stan`. Our model's linear predictor η_{avt} for a vessel with age a , vessel type v and treatment t is determined by the following sum:

$$\eta_{avt} = \mu_a \tag{26}$$

$$+ \alpha_t^{treatment} \tag{27}$$

$$+ \alpha_v^{vessel\ type} \tag{28}$$

$$+ \alpha_{vt}^{vessel\ type:treatment} \tag{29}$$

$$+ \alpha_{av}^{age:vessel\ type} \tag{30}$$

Our model's likelihood for a measurement y_{avt} is given by

$$y_{avt} \sim LN(\eta_{avt}, \sigma) \tag{31}$$

where σ is a scalar parameter.

The prior model is as follows:

$$\sigma \sim N(0, 0.5) \tag{32}$$

$$\mu \sim N(2, 0.5) \tag{33}$$

$$\alpha^{treatment} \sim N(0, \tau^{treatment}) \tag{34}$$

$$\alpha_v^{vessel\ type} \sim N(0, \tau^{vessel\ type}) \tag{35}$$

$$\alpha_{vt}^{vessel\ type:treatment} \sim N(0, \tau^{vessel\ type:treatment}) \tag{36}$$

$$\alpha_{av}^{age:vessel\ type} \sim N(0, \tau^{age:vessel\ type}) \tag{37}$$

$$\tau^{vessel\ type} \sim N(0, 0.5) \tag{38}$$

$$\tau^{treatment} \sim N(0, 0.5) \tag{39}$$

$$\tau^{age:vessel\ type} \sim N(0, 0.5) \tag{40}$$

$$\tau^{vessel\ type:treatment} \sim N(0, 0.5) \tag{41}$$

Results

Figure 47 shows the measurements with our model's marginal posterior predictive distributions. The fit to the observed data is overall adequate.

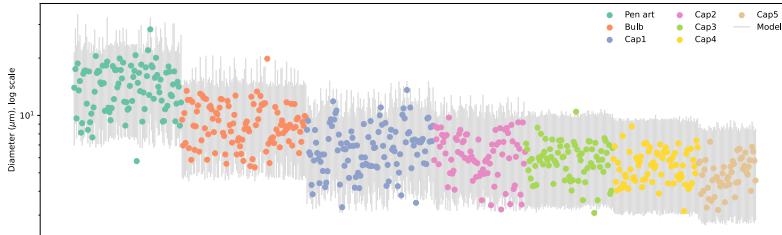


Figure 47: Grey bars represent our model's 1% to 99% posterior predictive interval for each measurement.

Figure 48 shows the marginal distributions of vessel type effects in our model, i.e. the parameters $\alpha^{vessel\ type}$. As expected, these show declining diameter moving further down the tree of vessels.

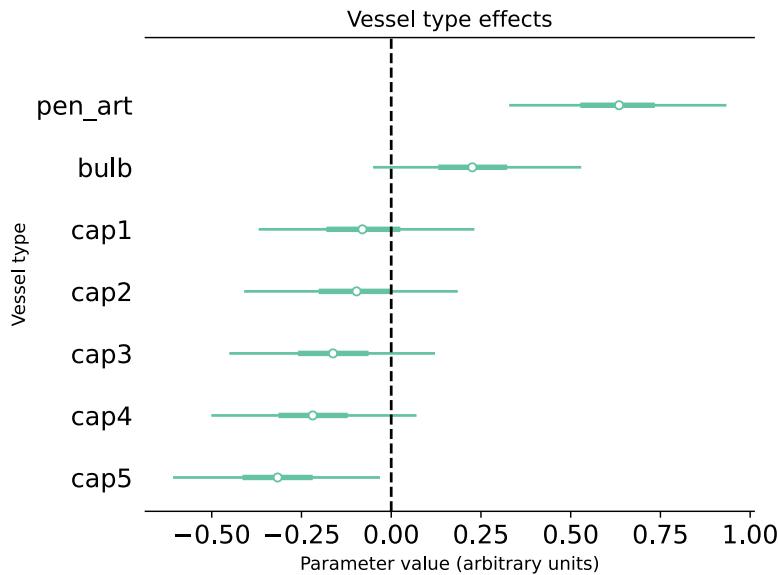


Figure 48: 94% posterior intervals for vessel type effects in our diameter model.

Figure 49 shows the marginal distributions of our model’s global age effect, i.e. the quantity $\mu_{old} - \mu_{adult}$. This plot shows that our model found no global difference between the diameters of adult and old mice.

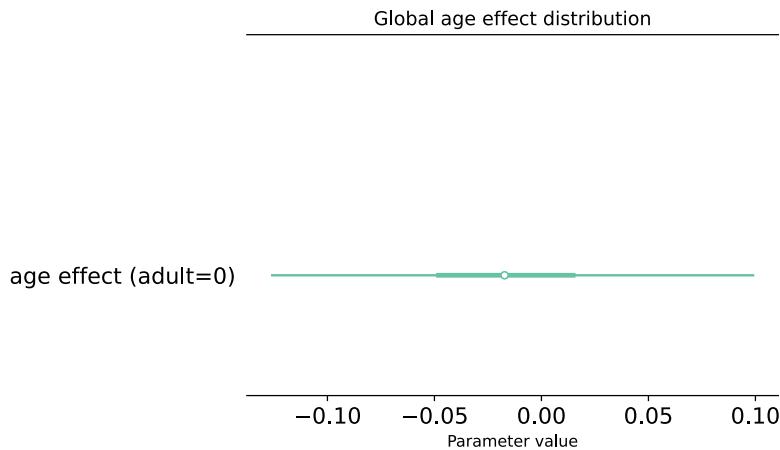


Figure 49: 94% posterior interval for the global age effect in our diameter model.

Figure 50 shows the marginal distributions of our model’s treatment effects, i.e. the quantity $\alpha_t^{treatment} - \alpha_{baseline}^{treatment} + \alpha_{vessel\ type:treatment} - \alpha_{v,baseline}^{vessel\ type:treatment}$ for each vessel type v and non-baseline treatment t . Our results show that the ablation and hyper2 treatments were associated with larger diameters in the upper blood vessels.

Details of the collaterals analysis

The collaterals data consisted of measurements of collateral diameter, length and tortuosity, as well as counts recording the number of collaterals per brain surface area. All of these measurements came with mouse ids, and all mice were labelled as adult or old.

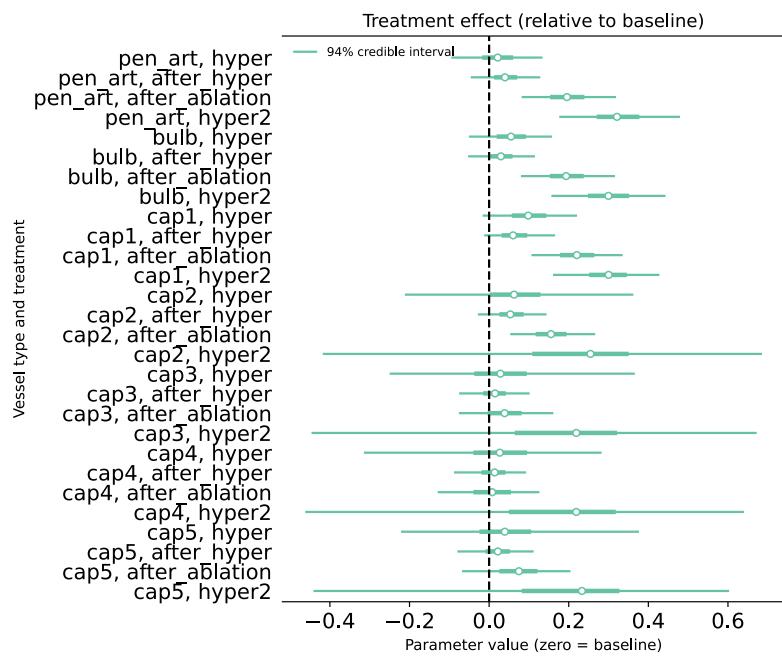


Figure 50: 94% posterior intervals for treatment effects in our diameter model.

Software

Since this data did not involve a nested structure of categorical features, we used the simpler formula-based model specification framework bambi (Capretto et al. 2022), avoiding the need to hand-write Stan programmes.

All code relating to this analysis can be found at <https://github.com/teddygroves/sphincter/blob/main/sphincter/collaterals.py>.

Data processing

We created two tables: one recording the diameter, length and tortuosity of each measured collateral and one with the measured collateral density per mouse.

We discarded measurements for which the measured tortuosity was not greater than one, or where any of the diameter, length or tortuosity were not measured.

The resulting tables can be inspected at <https://github.com/teddygroves/sphincter/blob/main/data/prepared/collaterals.csv> and <https://github.com/teddygroves/sphincter/blob/main/data/prepared/collaterals-mice.csv>.

Statistical model

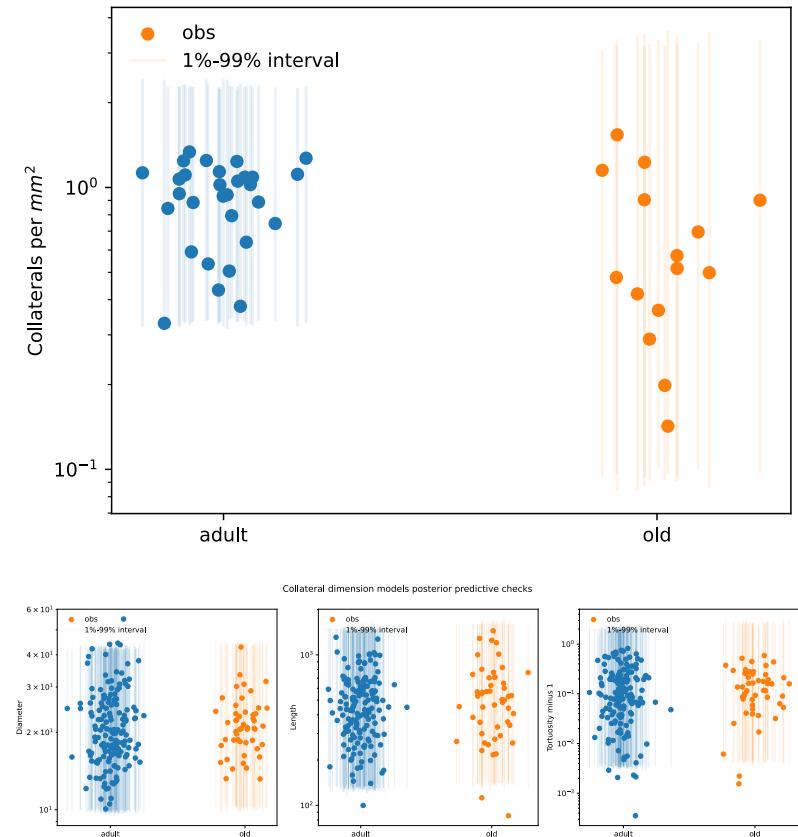
We used a linear regression model on natural logarithmic scale to model all collateral measurements, with effects for age. In bambi's formula language our model was written as " $\{y\} \sim \text{age}$ ", where " $\{y\}$ " can be substituted with the name of a dependent variable representing log-scale collateral density, diameter, length or tortuosity minus the theoretical minimum value 1.

The priors were the bambi defaults, which aim to be weakly informative while automatically scaling to match the predictors.

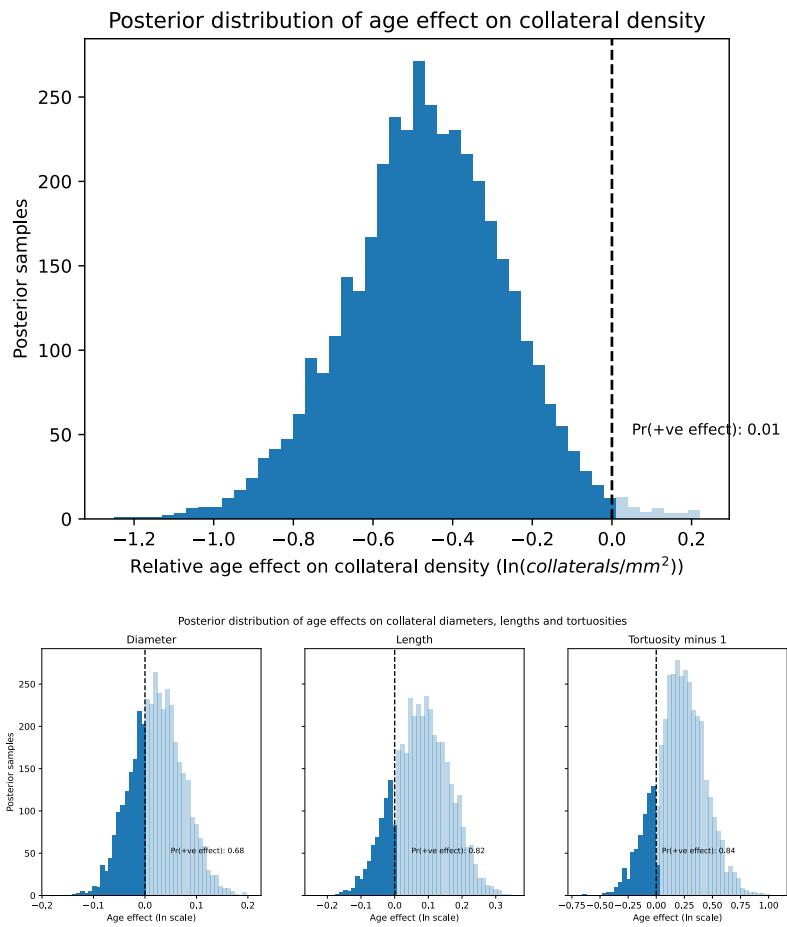
Results

The following figure show the fits of our models' posterior predictive distributions to the observed data: the fit is acceptable in all cases.

Collateral density model posterior predictive check



These figures show the distribution of estimated effects. There is a clear tendency for old mice to have lower collateral density, and no clear age effect on any of the collateral dimensions.



Details of the branchpoint analysis

The branchpoint data consisted classifications branchpoints as having or not having a sphincter and/or a bulb, as well the following predictive features: branch number, depth and ratio of first order capillaries to penetrating arterioles.

Software

Similarly to the collaterals analysis, we defined statistical models using pymc (Abril-Pla et al. 2023) via bambi (Capretto et al. 2022).

All code relating to this analysis can be found at <https://github.com/teddygroves/sphincter/blob/main/sphincter/branchpoints.py>.

Data processing

We discarded measurements for which the measured depth was zero, or where any dependent or explanatory variables were unmeasured.

The resulting table can be inspected at <https://github.com/teddygroves/sphincter/blob/main/data/prepared/branchpoints.csv>.

Statistical model

We used a logistic regression model to predict sphincter or bulb occurrence for each branchpoint in our dataset, with effects for age, branch number, log-scale depth and logit-scale first-order-to-penetrating-arteriole ratio.

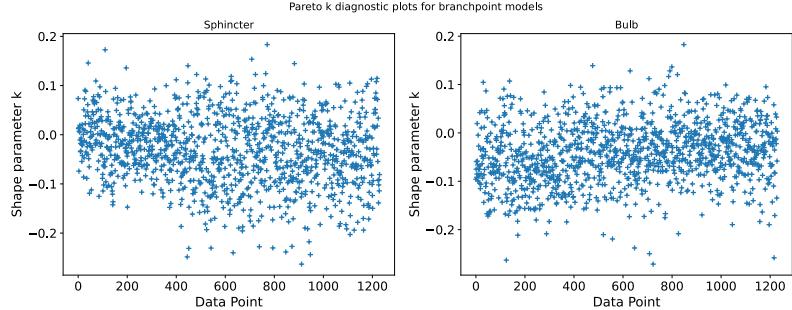
In bambi's formula language our model was written as
" $\{y\}['True'] \sim \text{age} + \text{branch_number} + \ln_{\text{depth}} + \text{logit_firstorder_per_pa}$ ", where " $\{y\}$ " can be substituted with the name of a boolean variable representing whether or not a branchpoint has a bulb or sphincter.

The priors were the bambi defaults, which aim to be weakly informative while automatically scaling to match the predictors.

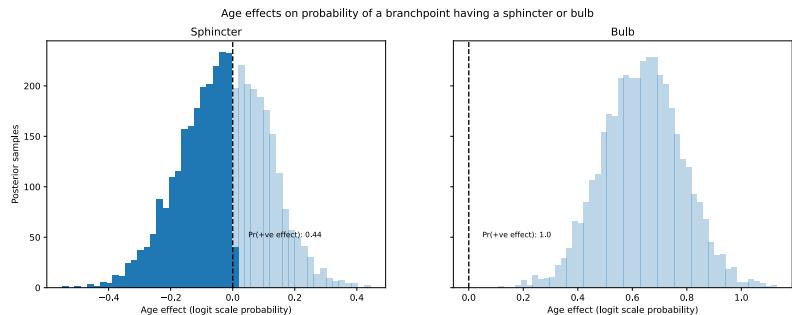
Results

We assessed model specification by estimating leave-one-out predictive probabilities and examining the distribution of pointwise pareto-k parameters. The results were acceptable in both cases, indicating clearly better-than-random prediction in both cases: on average our models probabilities 0.552 and 0.573 to the correct classifications. The figure below shows the distribution of pareto k parameters for the estimation method we

used, indicating that our estimated leave-one-out probabilities were accurate. See (Vehtari et al. 2024) for more about this method.



The figure below shows the distribution of estimated age effects. There is a clear tendency for branchpoints in old mice to have more frequent bulbs, and no effect of age on whether a branchpoint has a sphincter.



Abril-Pla, Oriol, Virgile Andreani, Colin Carroll, Larry Dong, Christopher J. Fonnesbeck, Maxim Kochurov, Ravin Kumar, et al. 2023. “PyMC: A Modern, and Comprehensive Probabilistic Programming Framework in Python.” *PeerJ Computer Science* 9 (September): e1516. <https://doi.org/10.7717/peerj-cs.1516>.

Betancourt, Michael. 2017. “Diagnosing Biased Inference with Divergences.” *Betanalpha.github.io*. https://github.com/betanalpha/knitr_case_studies/tree/master/divergence

Capretto, Tomás, Camen Piho, Ravin Kumar, Jacob Westfall, Tal Yarkoni, and Osvaldo A. Martin. 2022. “**Bambi** : A Simple Interface for Fitting Bayesian Linear Models in

- Python.” *Journal of Statistical Software* 103 (15). <https://doi.org/10.18637/jss.v103.i15>.*
- 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>.
- Gao, Yuxiang, Lauren Kennedy, Daniel Simpson, and Andrew Gelman. 2019. “Improving Multilevel Regression and Post-stratification with Structured Priors.” *arXiv:1908.06716 [Stat]*, September. <http://arxiv.org/abs/1908.06716>.
- 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>.

[com/stan-dev/cmdstanpy](https://github.com/stan-dev/cmdstanpy).

- Tornqvist, Leo, Pentti Vartia, and Yrjo 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 $R^{\hat{}}$ for Assessing Convergence of MCMC (with Discussion).” *Bayesian Analysis* 16 (2): 667–718. <https://doi.org/10.1214/20-BA1221>.
- Vehtari, Aki, Daniel Simpson, Andrew Gelman, Yuling Yao, and Jonah Gabry. 2024. “Pareto Smoothed Importance Sampling.” *Journal of Machine Learning Research* 25 (72): 1–58. <http://jmlr.org/papers/v25/19-556.html>.