Impact of Metabolite Abundance on Aging in Dogs

Bayesian Analysis

Stella Jia

May 2023

Contents

1 Abstract

Through Bayesian hierarchical linear models (HLM) we show the value of a Bayesian approach in understanding the metabolome of dogs. This report highlights the importance of considering interactions between different traits and utilizing shrinkage techniques to address uncertainty, particularly in cases with smaller sample sizes. The findings emphasized the need for careful interpretation and the selection of appropriate priors to enhance confidence in the estimated relationships between metabolite abundance and age.

2 Background

Metabolomics has emerged as a valuable tool in aging research, offering unique insights into the molecular changes that occur during the aging process (1). By studying the metabolite profiles of organisms, metabolomics provides a comprehensive snapshot of the biochemical pathways associated with aging which can be used to create age prediction models (2).

A metabolite is a small molecule that is involved in the chemical reactions and metabolic processes within a living organism. The metabolome encompasses the complete set of metabolites present within a biological sample or organism. The metabolome is highly influenced by various factors including genetics, environmental conditions, diet, lifestyle, and disease states (3,4,5). Aging is also dependent on genetic and physiological factors which makes metabolomics a promising approach to understanding aging (6).

To explore aging through metabolomics, we collaborated with the Dog Aging Project (DAP) to obtain two datasets related to the biometric profile and metabolite abundance across 138 metabolites of 726 dogs. DAP aims to understand how genes, lifestyle, and environment influence aging to help pets and people increase their healthspan.

Broadly, the goal of this project is to bring insight into the metabolome of dogs through a Bayesian approach. Taking a Bayesian approach to the data can be fruitful in combating small sample sizes when priors are intentionally used (7). In this dataset we group the data under certain characteristics using hierarchical linear models which results in smaller sample sizes. Domesticated dogs tend to exhibit a longer lifespan in dogs with smaller body mass than larger body mass (8). Thus, we were particularly interested in grouping the data by size and seeing how metabolite abundance varies among different size groups.

3 Visualization and Exploratory Data Analysis

By adjusting the grouping and predictor variables included in the HLM we can see how different traits interact and impact the relationship of metabolite abundance Y_{ij} vs age X_{ij}.

$$Y_{ij}|\beta_{0|ij}, \beta_{1|ij}, \sigma_y \sim N(\mu_{ij}, \sigma_y^2) \text{ where } \mu_{ij} = \beta_{0|ij} + \beta_{1|ij} X_{ij} \tag{1}$$

In this case, $i = \{Tiny, Small, Medium, Large, Giant\}$ which represents size groups and j is each dog in that size group. For this model, we set priors for the mean and variance of β_0 and β_1:

$$\begin{aligned} \mu_{\beta_0} &\sim N(0,1), \quad \mu_{\beta_1} \sim N(0,1) \\ \sigma_{\beta_0} &\sim Cauchy(0,1), \quad \sigma_{\beta_1} \sim Cauchy(0,1) \end{aligned} \tag{2}$$

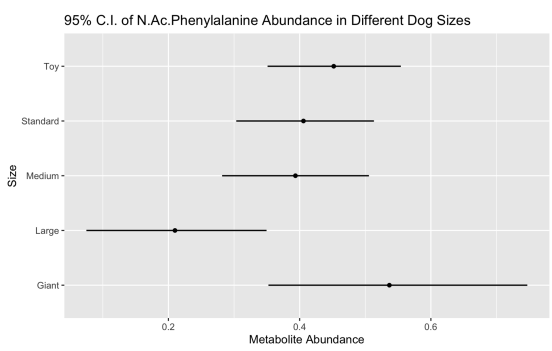
Since the data was standardized, theoretically the β_0 for the model should be 0 for all metabolites. Thus, the prior of $\mu_{\beta_0} \sim N(0,1)$ is not necessarily the most useful. $\mu_{\beta_1} \sim N(0,1)$ is saying that before seeing the data we are giving the value of β_1 for each size group an equal chance of being positive or negative. The Cauchy priors for both σ's are highly diffused which means there will not be a lot of shrinkage in the metabolite abundance of each size group to the overall mean. The sample size for each group was:

Giant	Large	Medium	Standard	Toy
60	185	153	174	154

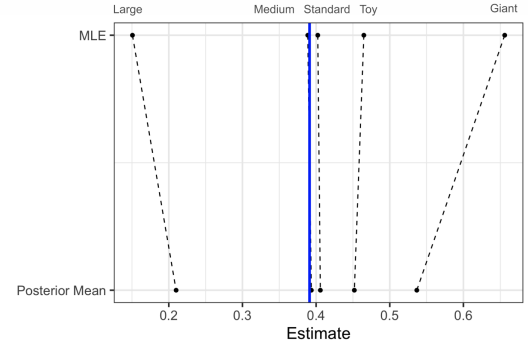
3.1 Stan Model Fit

Multiple HLMs were created to understand the Bayesian framework and how it fits into this data. Before running any models, I standardized the metabolite abundance and age data. To fit the model, I used the cmdstanr package.

After fitting it to the data, I used shinystan to check the stan diagnostics. Our stan fit had little divergence and the chains mixed well. Now we can take a look at the posterior confidence intervals and shrinkage plot to see what our stan model outputted:



(a) 95% Confidence Intervals



(b) Shrinkage Plot of β_{age}

Figure 1: Posterior Interval and Shrinkage Plot (1)

Instead of a consistent metabolite abundance pattern from toy to giant dog, the deviance in large dogs displayed in Figure 1a would be interesting to explore further. In Figure 1b, we see a shrinkage plot where the group with the most shrinkage are giant dogs due to its small sample size (n=60) relative to the rest. From this simple example, we can see how a Bayesian approach can help cushion uncertainty in a small sample.

Looking beyond age, we also created an HLM that grouped based on size and sterilization status. The same stan file was run expect we changed the grouping parameter so that each group was a unique pair of size and sterilization status which resulted in ten different groups.

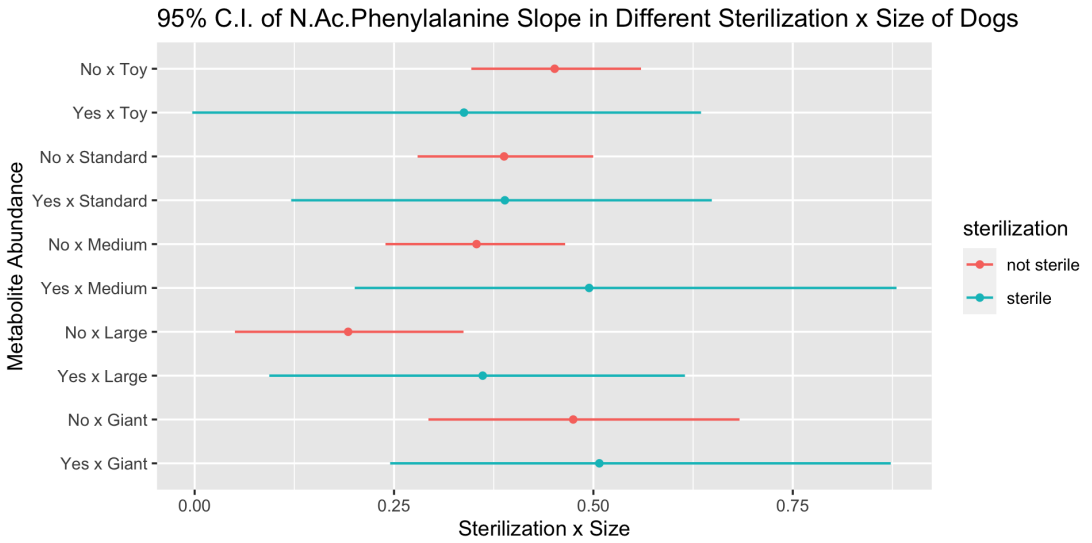


Figure 2: Grouping by Sterilization x Size on Model (1)

Four out of five of the size groups have a larger median (although sterile standard dogs are only slightly larger than not sterile standard dogs) when sterilized than when not sterile. Thus, on average sterile dogs have more extreme slopes in the N.Ac.Phenylalanine metabolite which may indicate shorter lifespan. However, since we have a smaller sample size for sterile dogs (n=70) compared to non-sterile (n=656), there is low confidence around the actual relationship of metabolite vs age in sterile dogs. It would be helpful to see how sterility impacts other metabolites and use existing domain literature on sterility to further inform our prior.

3.2 Multiple Metabolites

Looking at one metabolite is limiting. Thus, the next model we ran looked at all metabolites with a significant positive linear relationship with age. To determine whether a metabolite is significant, we looked at the correlation of age versus each metabolite (across all groups) and selected the ones that had p-values less than 0.05. When plotting all metabolites, we interacted with the shrinkage component to show how much 'cushion' we can control.

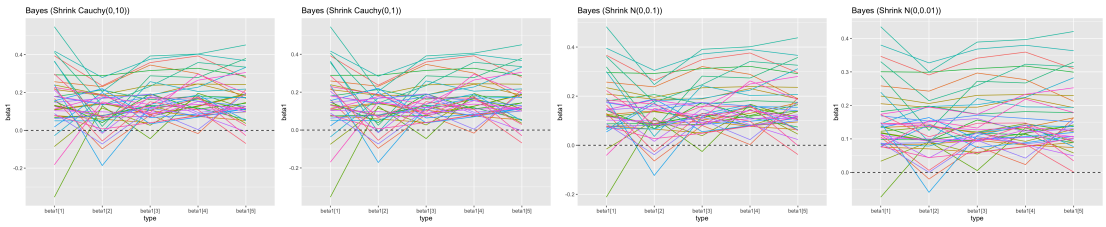


Figure 3: Comparing All Positive Metabolites on Model (1)

The first chart to the left adjusts the variance of the shrinkage (σ_{age} in $\beta_{age} \sim N(\mu_{age}, \sigma_{age})$ such that $\sigma_{age} \sim \text{Cauchy}(0, 10)$. Since Cauchy(0, 10) is very dispersed, the plot resembles what a frequentist approach to the calculation of β_{age} for each size group might look like. Theoretically, the values of β_{age} for a frequentist would vary greatly among each size because there

is no shrinkage towards the global mean. As you move towards the charts on the right, the less variance there is in each metabolite across different sizes (line is more horizontal because all values shrinking towards singular global mean value). This allows us to see the outcome of β_1 when taking a no shrinkage to gradually stronger shrinkage approach. With proper domain knowledge, we could make an informative call on which approach is preferred.

After looking at significant metabolites across all groups, we then explored the significant metabolites within each group. Instead of calculating correlation across all dogs, we calculated it across the dogs in each size group. The group with the most significant metabolites is Toy and the least is Giant and Medium. **The six metabolites that are significant across all groups are: N.Ac.Tryptophan, N.Ac.Phenylalanine, Glutaric Acid, Acetylcarnitine, Glycine, and Isocalerylcarnitine.** It would be interesting to take a closer look at the behavior of these metabolites and what the existing domain literature says.

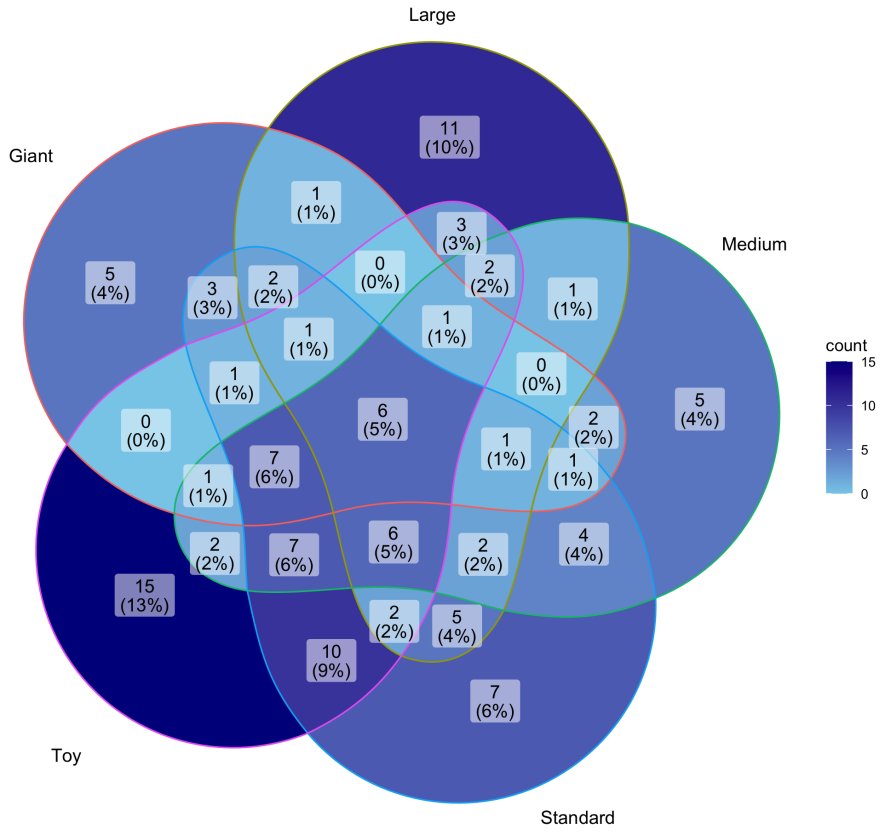


Figure 4: Number of Significant Metabolites Per Size Group

For even more granularity, we looked at the posterior interval of β_{age} for all metabolites (includes non-significant metabolites) within a group. In the figure below, we specifically looked at giant and toy dogs. Both charts are scaled the same on the x-axis. Giant dogs have larger intervals and intersects zero more often. This is likely a result of the smaller sample size of giant dogs which makes us less confident in the values. To increase our confidence, we would need to set a

more "confident" prior to make up for less data.

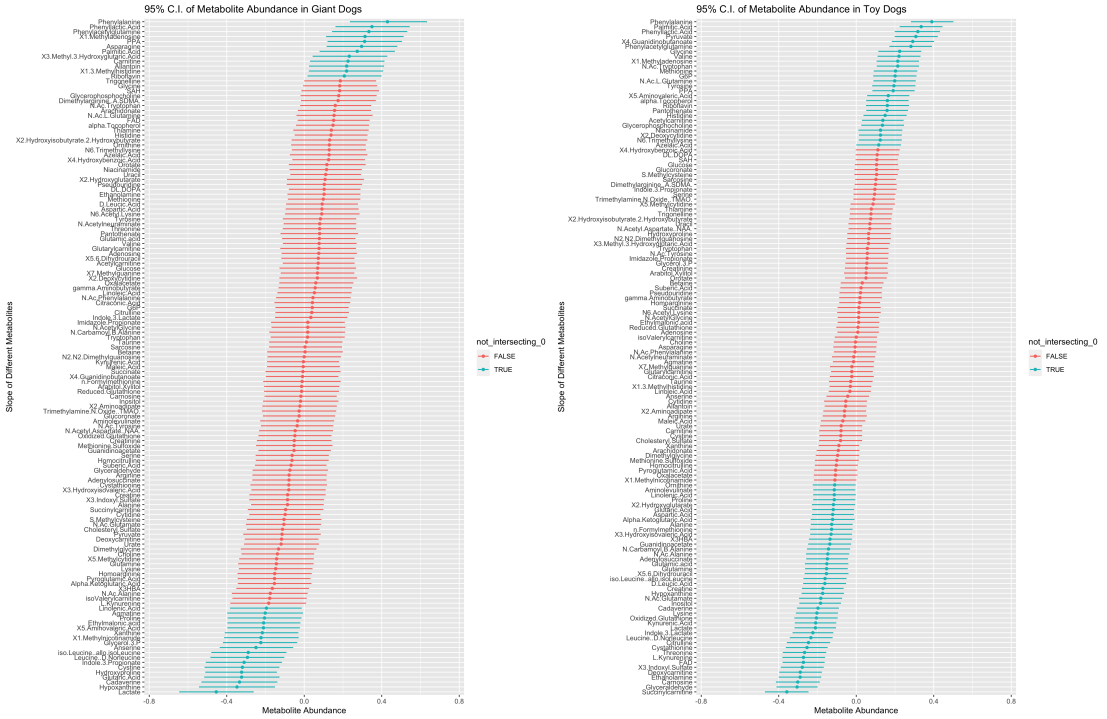
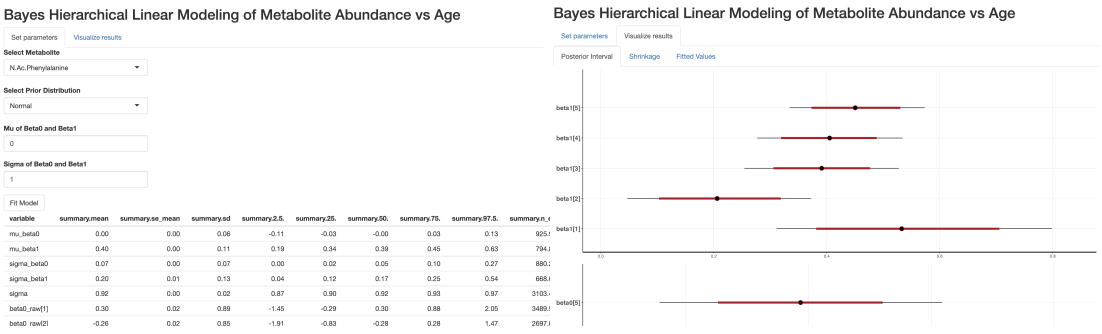


Figure 5: Comparing All Metabolites on Model (1)

3.3 Shiny App

When working with 137 metabolites, it was mildly tedious running the stan model multiple times in a file with a lot of existing code. To address this, I decided to leverage the capabilities of R Shiny to create an interactive dashboard. This dashboard serves as a user-friendly interface that simplifies the analysis process and enhances accessibility. While it is more user friendly, there is a trade off on the flexibility of analysis that can be executed.



(a) Main Page

(b) Posterior Analysis Page

Figure 6: Basic Interface of R Shiny App

Within the dashboard, I incorporated options for selecting prior distributions and setting priors for the global mean and shrinkage variance. By utilizing the dashboard, I could easily run the code for each metabolite, reducing the tediousness of navigating and executing complex code. The R Shiny dashboard also offers the potential for further sophistication, like the ability to include other predictors (size, sterilization, sex, etc), and pre-fitting models for more efficient user experience.

4 Conclusion

4.1 Reflection

From a research perspective, this project broadened my understanding of the challenges and intricacies involved in Bayesian analysis. Through visualization tools like shinystan and creating my own visuals for posterior analysis, I learned to critically analyze and interpret the results obtained from Bayesian models while considering uncertainties from the assumptions being made. Strategically setting priors and understanding how it will impact the data is important and something I hope to be more critical about if supplementing it with more domain knowledge.

On a personal level, this project allowed me to develop a deeper appreciation for the power and flexibility of Bayesian statistics. I also enjoyed taking a utility approach to the data by creating the R Shiny app. As I continue exploring, I hope to be more intentional with thinking about how each visual or tools plays into the overarching story and other existing stories (i.e. existing literature, domain knowledge, etc) of the research.

4.2 Future Steps

In the future, there are a few areas I'd like to explore deeper:

1. Since we are working with 137 metabolites, we have attempted to filter the metabolites based on ones that show statistical significance of correlation between metabolite abundance and age variable. This approach is limiting because we are only looking at the relationship between age and metabolite abundance. Next time, it would be interesting to target a few metabolites that have **biological** significance based on past research and domain expertise.

2. Based on our brief exploration, it seems dogs that are sterile may have more extreme slopes. Since we have only observed this relationship among a few metabolites, it would be interesting to see whether this pattern applies to all metabolites. Taking a Bayesian approach could be particularly useful because of the limited number of non-sterile dogs (dataset contains 70 non-sterile dogs and 656 sterile dogs). Using Bayes and domain knowledge, create informed priors for a model that **groups by size group and sterilization status.**

3. Using more complex models to analyze how metabolite abundance varies with age for **different breeds** instead of just grouping by size groups. Also, including more biological predictors in the model.

4. Run the model on all metabolites and compare the pattern of metabolite abundance across groups. From Figure 1a, we know that N.Ac.Phenylalanine has an obvious difference in the metabolite abundance of large dogs compared to the rest. Does this pattern occur in other metabolites?

4.3 Credit

This work was supported by the guidance of Professor Alex Franks (UC Santa Barbara), Professor Daniel Promislow (University of Washington), and Maria Partida-Aguilar (University of Washington).

5 References

1. Adav, Sunil S, and Yulan Wang. "Metabolomics Signatures of Aging: Recent Advances." Aging and Disease, vol. 12, no. 2, 2021, p. 646, https://doi.org/10.14336/ad.2020.0909.

2. Hwangbo, Nathan, et al. "A Metabolomic Aging Clock Using Human Cerebrospinal Fluid." The Journals of Gerontology: Series A, vol. 77, no. 4, 2021, pp. 744–754, https://doi.org/10.1093/gerona/glab212.

3. Jin, Kelly, Kenneth A. Wilson, Jennifer N. Beck, Christopher S. Nelson, George W. Brownridge, et al. "Genetic and Metabolomic Architecture of Variation in Diet Restriction-Mediated Lifespan Extension in Drosophila." PLOS Genetics, vol. 16, no. 7, 2020, https://doi.org/10.1371/journal.pgen.1008835.

4. Tounta, Vivian, et al. "Metabolomics in Infectious Diseases and Drug Discovery." Molecular Omics, vol. 17, no. 3, 2021, pp. 376–393, https://doi.org/10.1039/d1mo00017a.

5. Almanza-Aguilera, Enrique, et al. "Impact in Plasma Metabolome as Effect of Lifestyle Intervention for Weight-Loss Reveals Metabolic Benefits in Metabolically Healthy Obese Women." Journal of Proteome Research, vol. 17, no. 8, 2018, pp. 2600–2610, https://doi.org/10.1021/acs.jproteor

6. Sgarbieri, Valdemiro Carlos, and Maria Teresa Pacheco. "Healthy Human Aging: Intrinsic and Environmental Factors." Brazilian Journal of Food Technology, vol. 20, no. 0, 2017, https://doi.org/10.1590/1981-6723.00717.

7. McNeish, Daniel. "On Using Bayesian Methods to Address Small Sample Problems." Structural Equation Modeling: A Multidisciplinary Journal, vol. 23, no. 5, 2016, pp. 750–773, https://doi.org/10.1080/10705511.2016.1186549.

8. Brookes, Paul S., and Ana G. Jimenez. Metabolomics of Aging in Primary Fibroblasts from Small and Large Breed Dogs, 2021, https://doi.org/10.1101/2021.02.25.432888.