

Modeling Longitudinal Behavioural Data in Preclinical Home Cage Monitoring Studies

Simone Modolo
simone.modolo@evotec.com



Digital Home Cage Monitoring



Animal Behavior

Active (%)

Climbing (%)

Locomotion (%)

Inactive (%)

Drinking (%)

Feeding (%)

Sleeping (%)

Percentage-based
parameters

Distance travelled
(cm)

Activity (cm/s)

Social Distance
Closest (cm)

Unit-based parameters

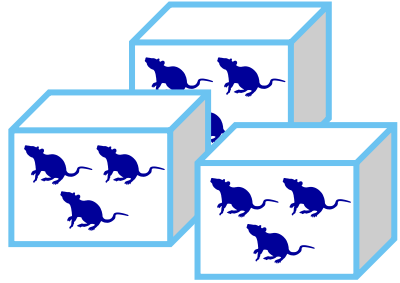
Animal Physiology

Respiratory rate
(breaths per min)

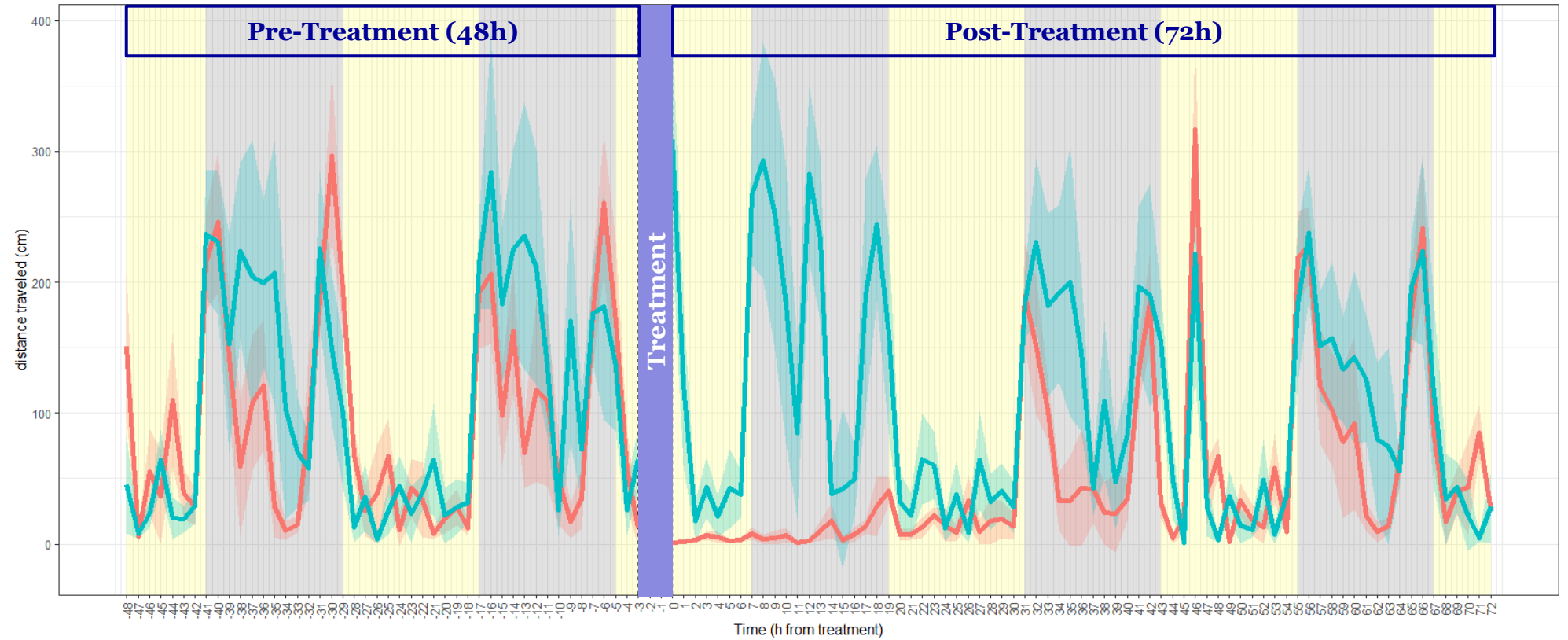
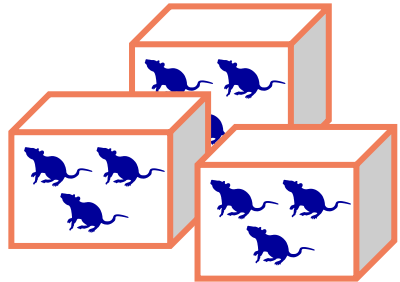


Study Design and Objective

Control (n=9)



Treatment (n=9)



**Single
Parameters
Analysis**



Study Objective

assess the treatment effect on animal behavior over time



**Integrated
behavior
analysis**

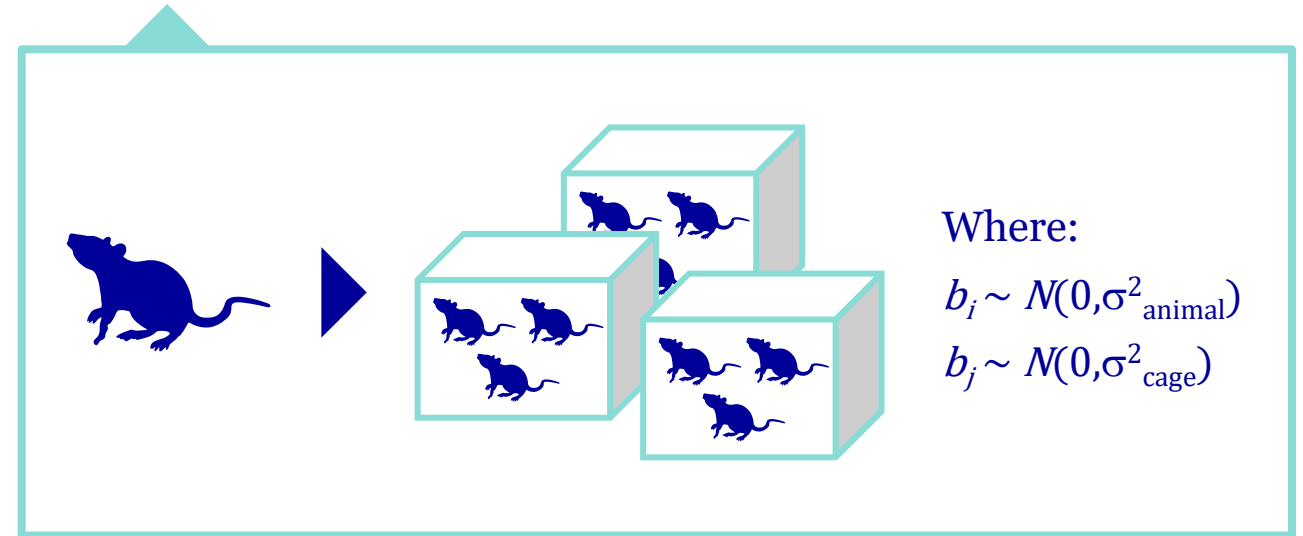
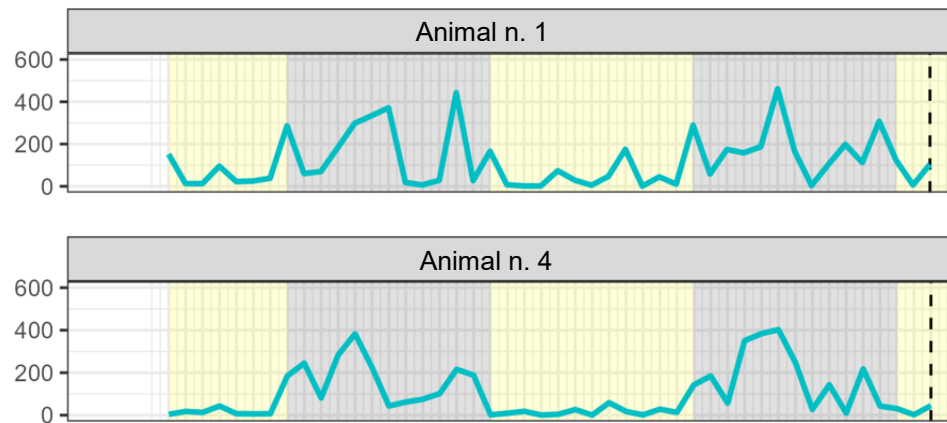


Statistical Model

Generalized Mixed effect model

Let Y_{ijkt} be the response for animal i in cage j , group k , at time t , with $E(Y_{ijkt} | b_i, b_j) = \mu_{ijkt}$
 $g(\mu_{ijkt}) = \beta_0 + \beta_1 \text{BaselineHour}_{ijkt} + \beta_2 \text{Group}_k + \beta_3 \text{Time}_t + \beta_4 (\text{Group}_k \times \text{Time}_t) + b_i + b_j$

**Pre-Treatment
(48h)**



Unit-based parameters

$h(Y_{ijkt}) | b_i, b_j \sim N(\cdot), g(\cdot) = I(\cdot)$,
 where h is the identity or a response transformation (e.g., log, sqrt, ...)

Percentage-based parameters

$Y_{ijkt} | b_i, b_j \sim \text{Beta}(\cdot), g(\cdot) = \text{logit}(\cdot)$
 with $0 < Y_{ijkt} < 1$



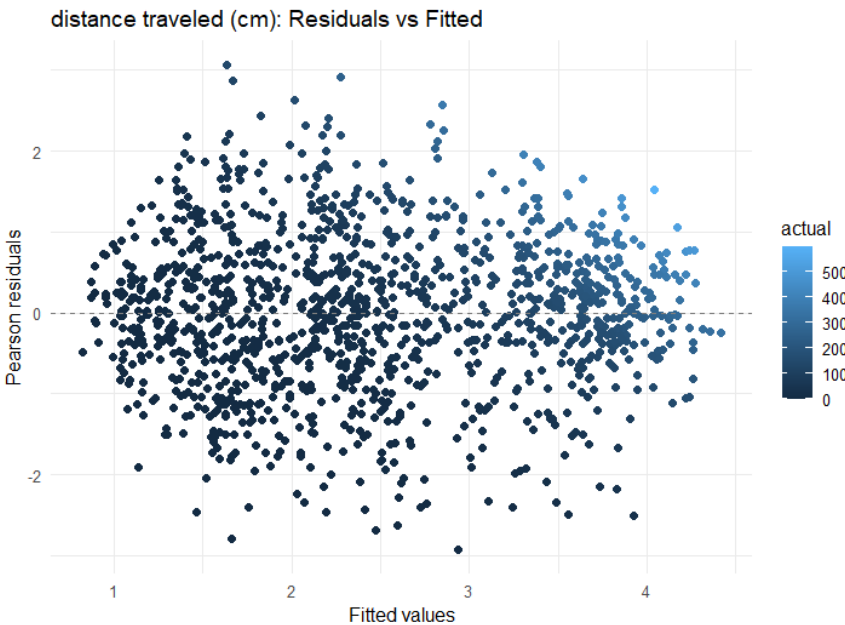
Results of unit-based parameters

Distance traveled (cm)

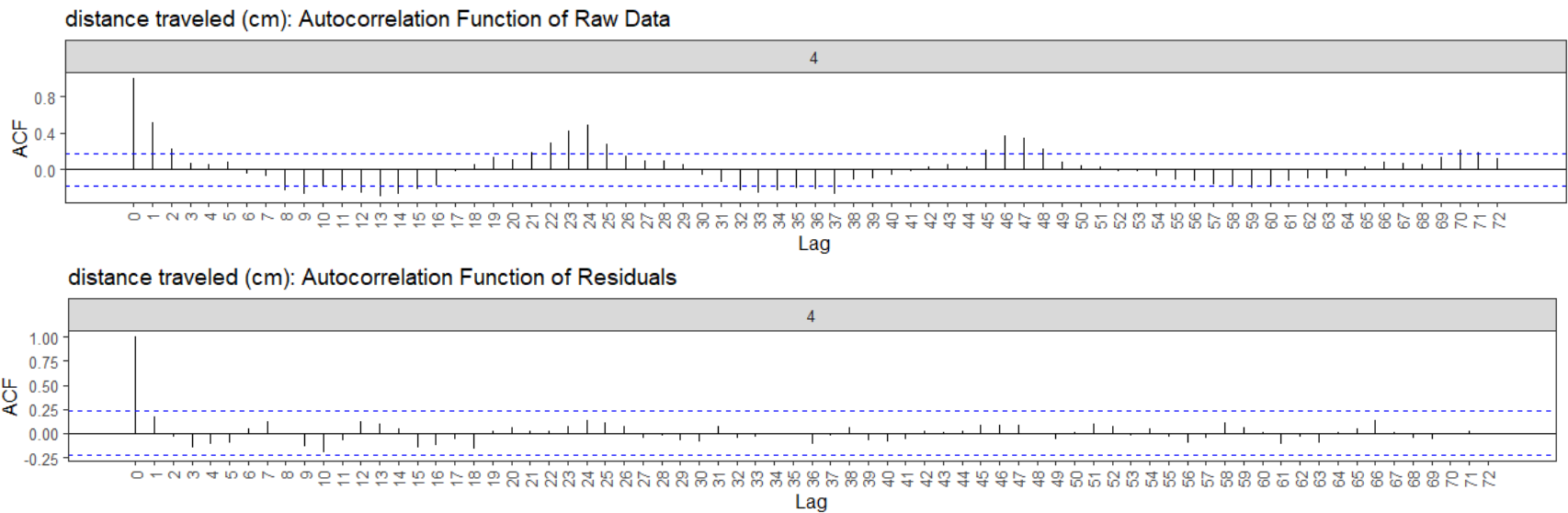
ANOVA table and model diagnostics

Parameter	Num d.f.	Den d.f.	F-value	P-value	Parameter	% of variability	Marginal R ²	Conditional R ²
(Intercept)	1	1126	3127.19	<.0001	Cage	2.63	0.68	0.69
BaselineHour	1	1126	850.42	<.0001	Animal	0.09		
Group	1	4	45.08	0.0026	Residual	97.28		
Time	72	1126	11.81	<.0001				
Group : Time	72	1126	11.17	<.0001				

Residuals



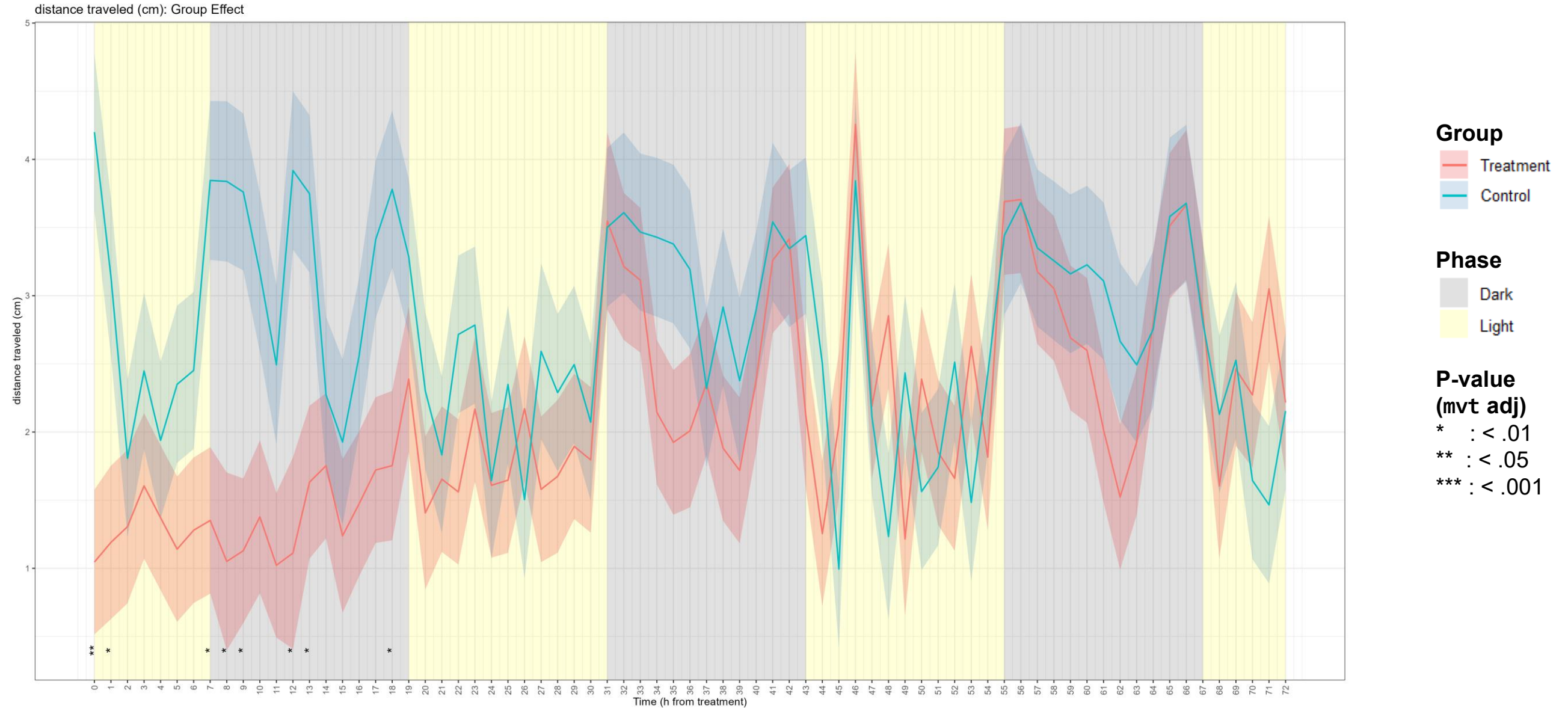
Autocorrelation





Results of unit-based parameters

Distance traveled (cm)



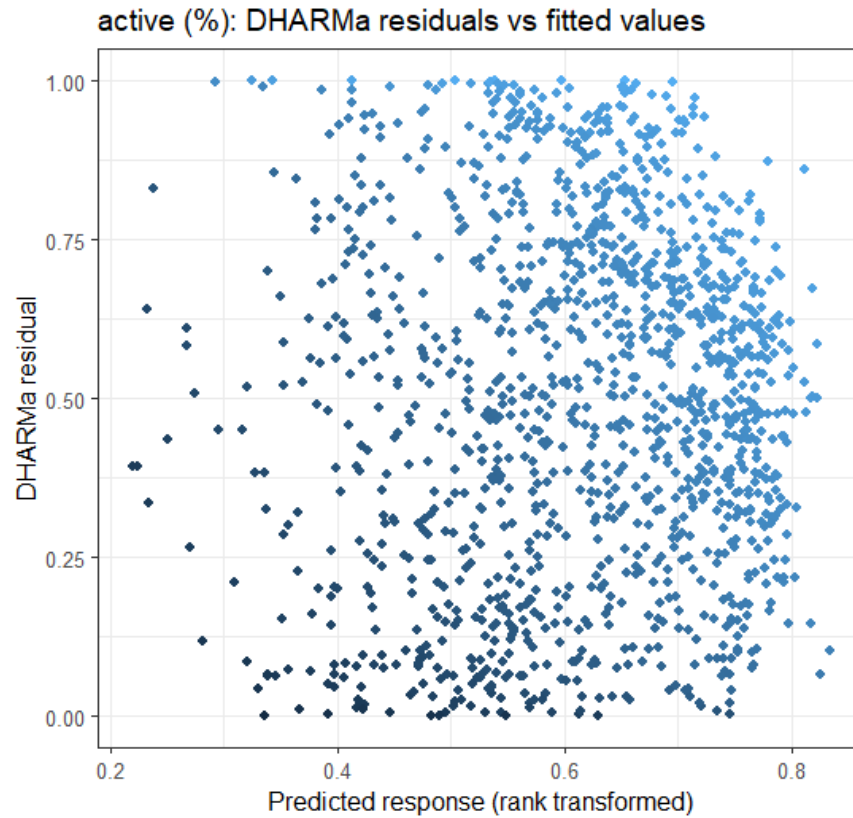
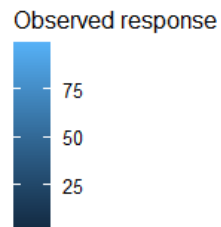


Results of percentage-based parameters

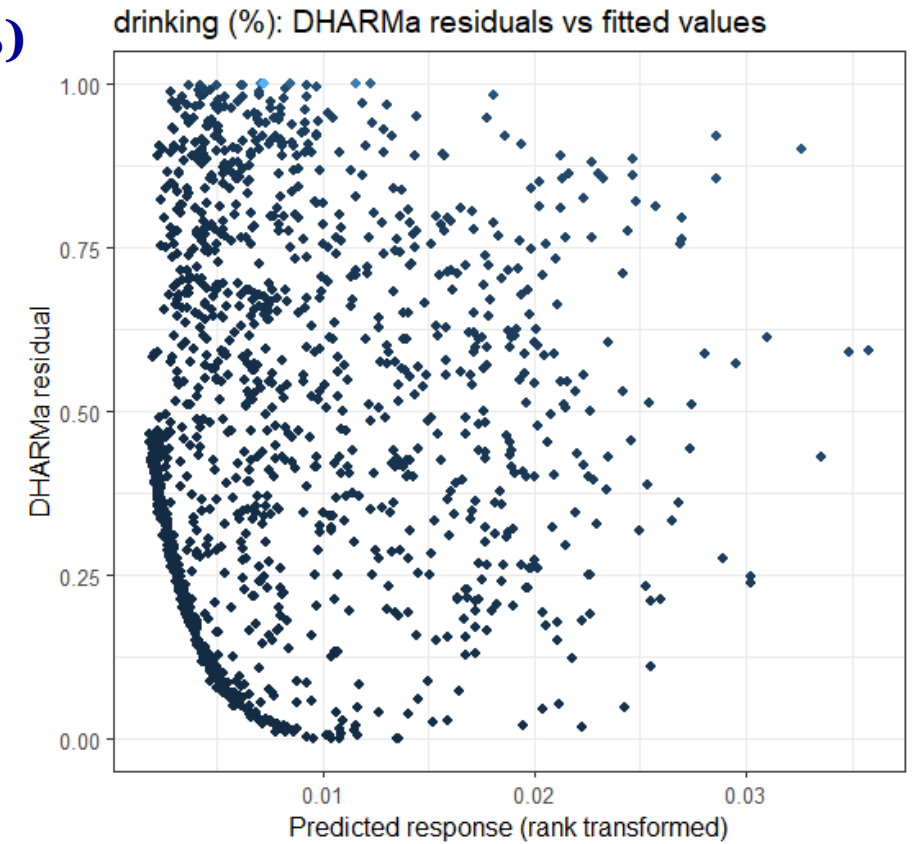
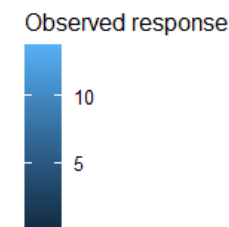
Active (%) and Drinking (%)

- **Data Transformation:** From [0%, 100%] range to (0, 1) range using Smithson-Verkuilen transformation
- **DHARMA residuals:**

Active (%)
0% of zeros



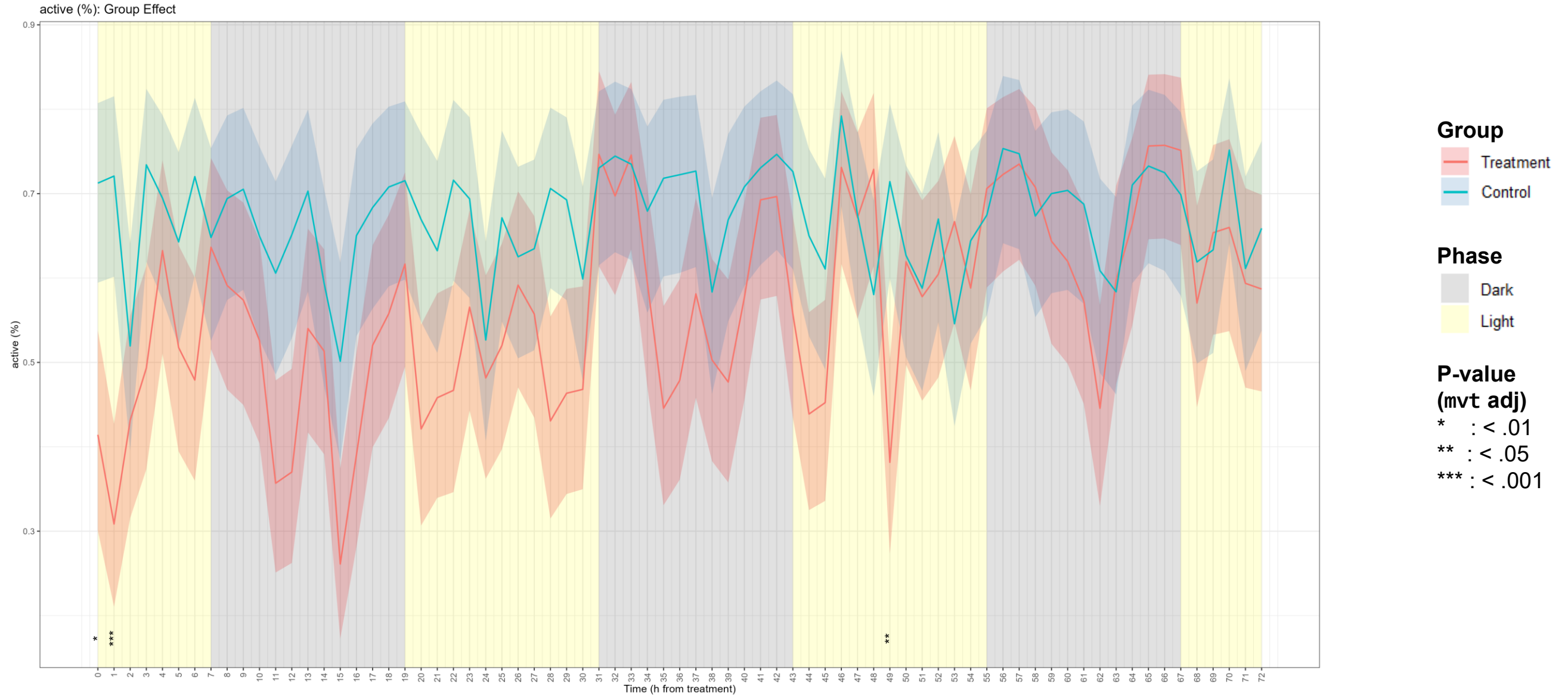
Drinking (%)
29% of zeros





Results of percentage-based parameters

Active (%)





Integrated approach

- Another way to look at the problem: **integrated approach** to study the treatment effect on animal behavior
- **Correlation** between measured parameters:

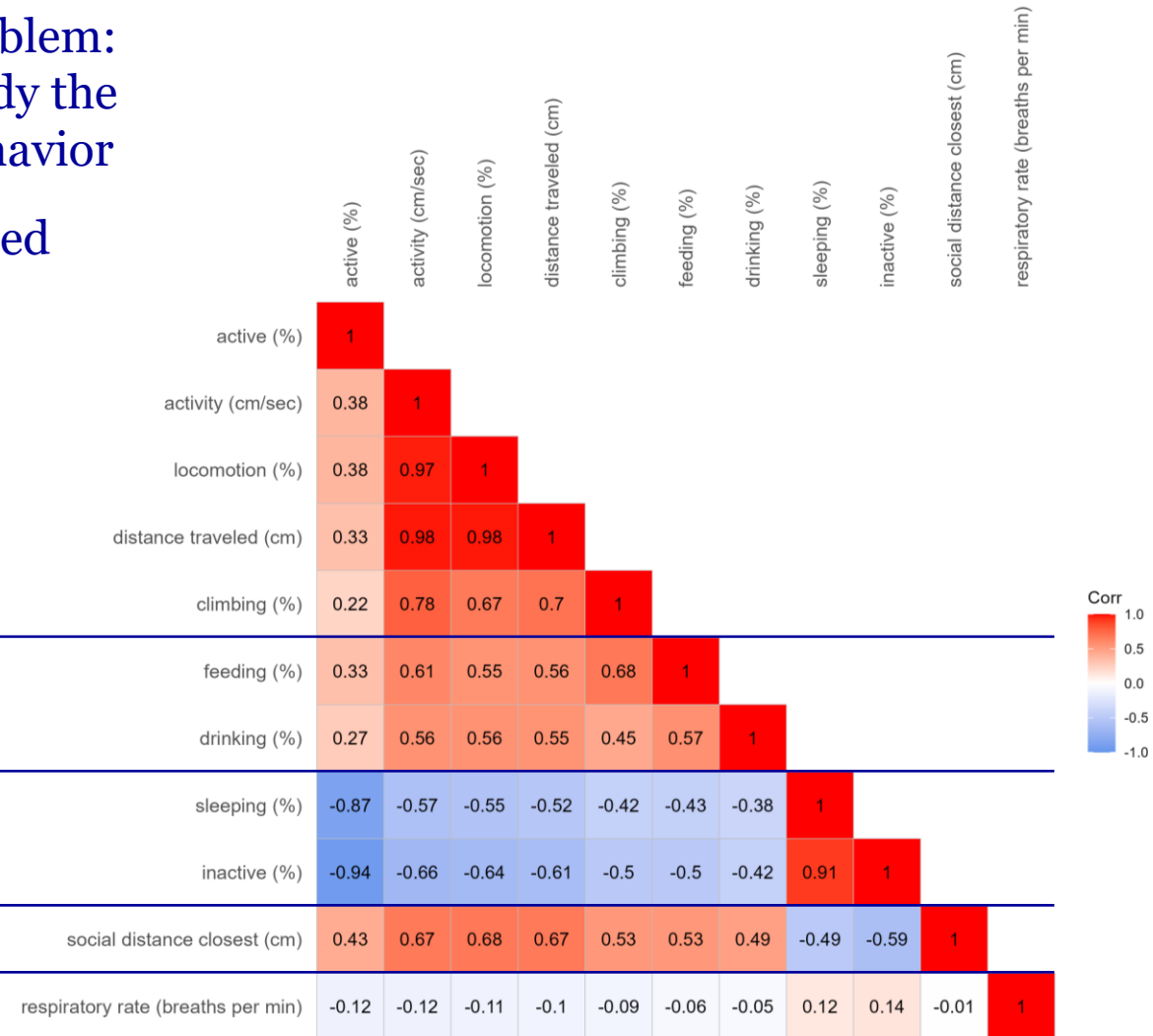
Activity

Ingestion

Rest

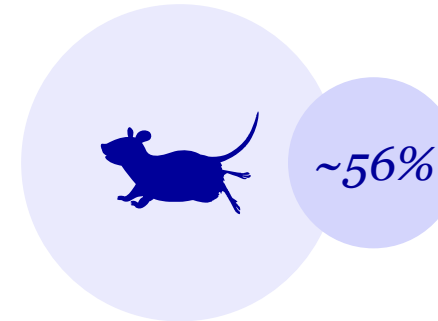
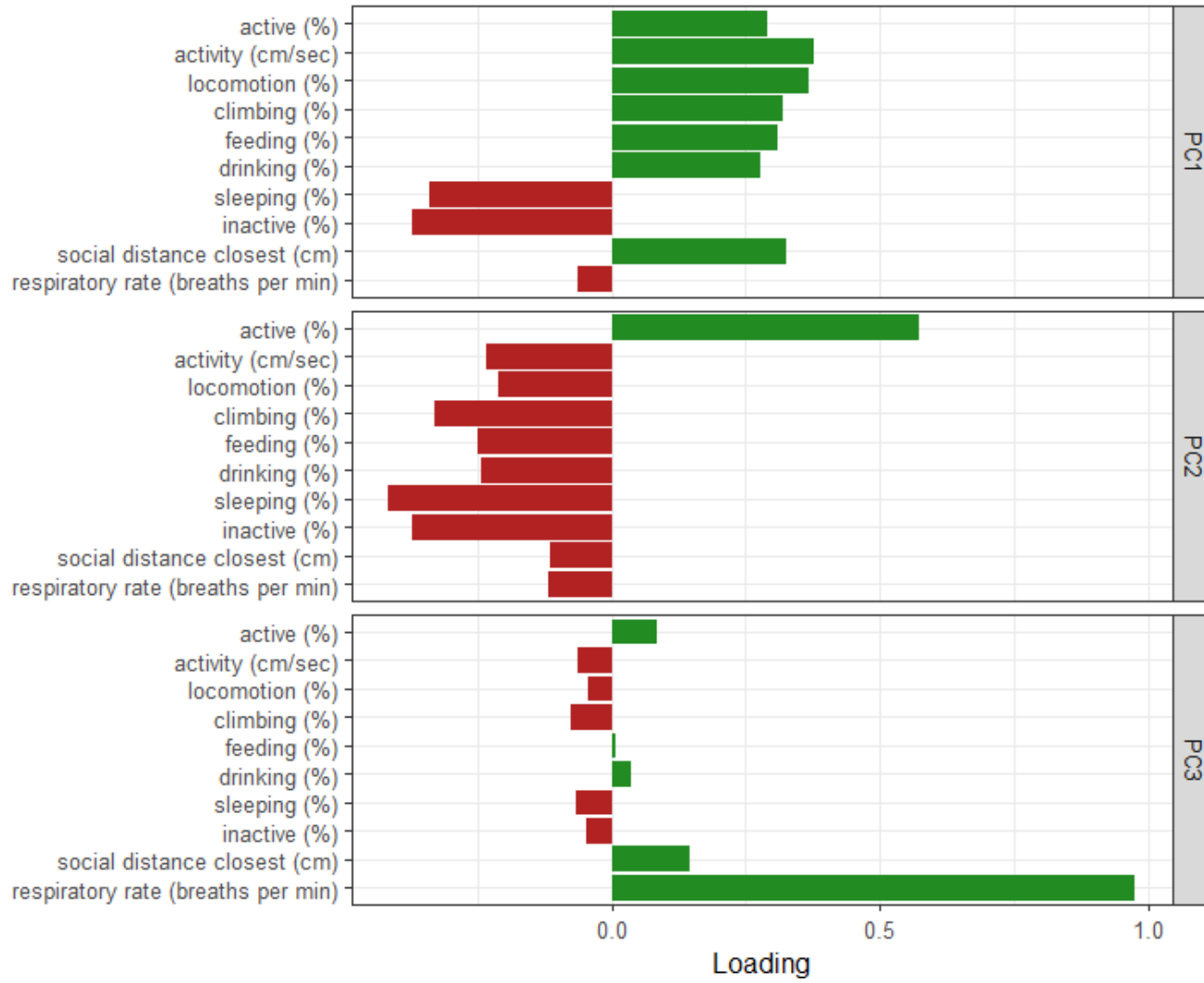
Social

Physiology



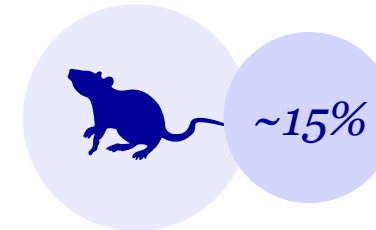


Principal Components Analysis



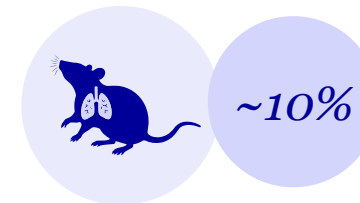
PC1

Overall Activity
vs Rest



PC2

Local activity
vs Other activities



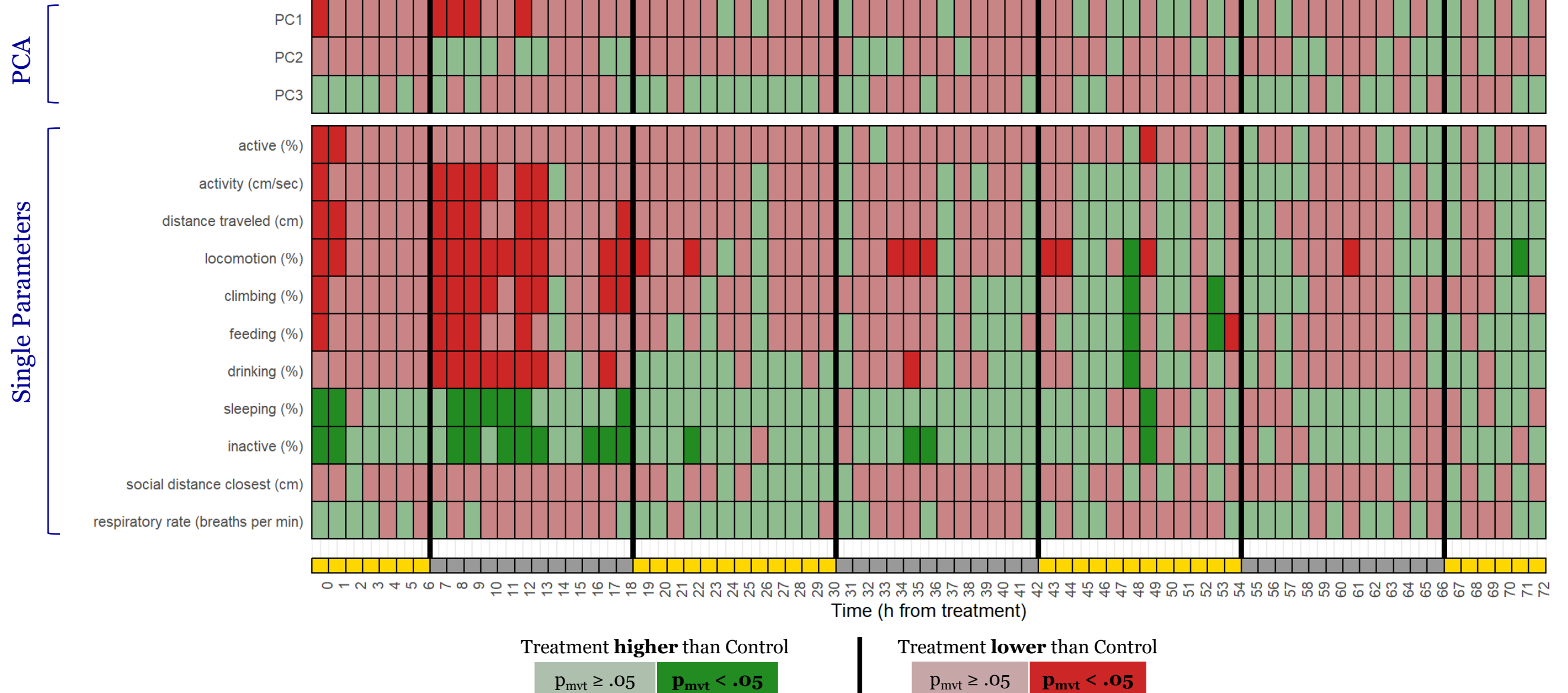
PC3

Respiratory
activity

~81%
of total
variability



Summary of treatment effect over time



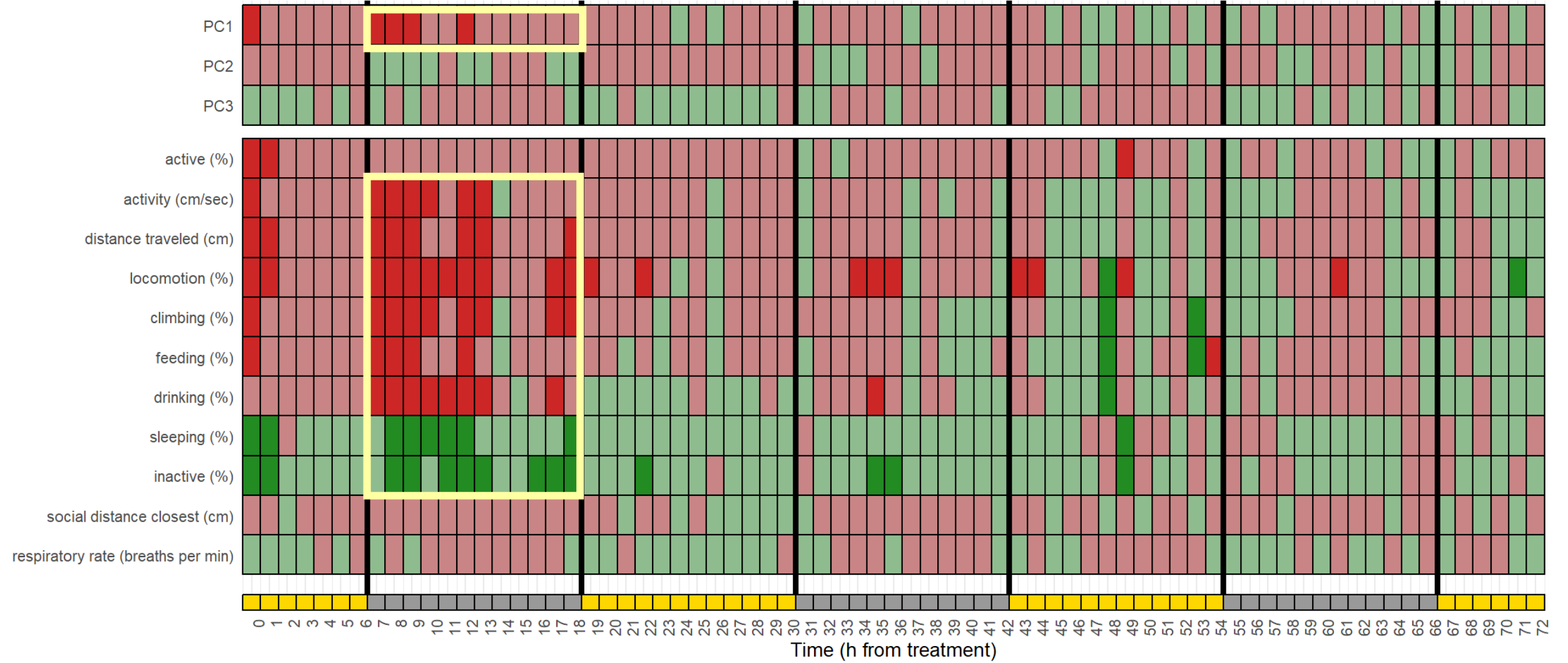


Summary of treatment effect over time

PC1: Overall Activity vs Rest

PCA

Single Parameters



Treatment **higher** than Control

$p_{mvt} \geq .05$

$p_{mvt} < .05$

Treatment **lower** than Control

$p_{mvt} \geq .05$

$p_{mvt} < .05$



Conclusions



Key Findings

- **Generalized mixed-effects models** captured treatment-related changes over time while accounting for baseline, repeated measurements and hierarchical structure of data.
- **Endpoint-specific** analyses allowed a focused assessment of treatment effects on parameters of particular interest in the study, while **PCA** provided a complementary integrated view of behavioural patterns.
- The first three **principal components** explained ~81% of the total variability. PC1 captured the overall activity vs rest dimension and showed the clearest agreement with the individual parameter results. PC2 and PC3 captured additional behavioural dimensions but showed no clear significant treatment effects in this study.



Next Steps

- **Refine models for challenging distributions**, in particular zero inflated models for endpoints with high percentage of zeros.
- **Incorporate additional endpoints**, such as huddling, and evaluate how they contribute to the overall behavioural profile.
- **Evaluate the robustness and reproducibility** of the identified PCs across studies, compounds and experimental settings, eventually exploring **alternative dimensionality reduction techniques**.



References

Papers

- Terry, Stéphane, et al. "Activity in group-housed home cages of mice as a novel preclinical biomarker in oncology studies." *Cancers* 15.19 (2023): 4798.
 - Bains, Rasneer S., et al. "Analysis of individual mouse activity in group housed animals of different inbred strains using a novel automated home cage analysis system." *Frontiers in behavioral neuroscience* 10 (2016): 106.
 - Sun, Rongwan, et al. "Accurate locomotor activity profiles of group-housed mice derived from home cage monitoring data." *Frontiers in Neuroscience* 18 (2024): 1456307.
-

R packages

- **Nlme**: mixed effects model with normal distribution
- **glmmTMB**: mixed effects model with beta distribution
- **DHARMa**: DHARMa residuals
- **emmeans**: multivariate t adjustment

Q&A

Acknowledgments

Marco Pergher

Denise Federico

Giulia Calusi

Augusto Celon

Translational Sciences Department





Backup



Smithson-Verkuilen transformation

The Smithson-Verkuilen transformation is:

$$y^* = \frac{(N - 1)}{N}y + \frac{0.5}{N}$$

Where N is the size of data to transform.

It compresses the range slightly toward the center, **removing the problematic boundaries** while **preserving the ordering and almost all of the scale information**.

References

Smithson, Michael, and Jay Verkuilen. "A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables." *Psychological methods* 11.1 (2006): 54.



Multiplicity Adjustment: Multivariate t (mvt) Method

- P-values for each pairwise contrasts were computed using the **joint distribution of all the T contrasts**, one for each timepoint.
- This distribution is a multivariate t-distribution that takes into account the **correlation between contrasts**
- This method allow to control Family Wise Error Rate (**FWER**)

References

- Lenth R, Piaskowski J (2025). `_emmeans`: Estimated Marginal Means, aka Least-Squares Means_ doi:10.32614/CRAN.package.emmeans <https://doi.org/10.32614/CRAN.package.emmeans> R package version 2.0.1, <<https://CRAN.R-project.org/package=emmeans>>.
- Hothorn, Torsten, Frank Bretz, and Peter Westfall. “Simultaneous inference in general parametric models.” *Biometrical Journal: Journal of Mathematical Methods in Biosciences* 50.3 (2008): 346-363.



DHARMa residuals

- **Simulation-based** approach to evaluate model fitting for generalized mixed models
 - DHARMa residuals are obtained following these steps:
 - for each observation, **simulate** B responses from the fitted model
 - for each observation, compare the observed response with the distribution of the B simulated responses by computing the **quantile**. The quantile is the DHARMa residual.
 - The distribution of DHARMa residuals should be **Uniform(0, 1)** and they should show **no trend** vs fitted values.
-

References

- Hartig F (2026). `_DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models_`. doi:10.32614/CRAN.package.DHARMa <https://doi.org/10.32614/CRAN.package.DHARMa> R package version 0.5.0, <<https://CRAN.R-project.org/package=DHARMa>>.