Supplemental Material

Machine learning models for temporally precise lapse prediction in alcohol use disorder

Kendra Wyant, Sarah J. Sant'Ana, Gaylen E. Fronk, John J. Curtin

2023-10-17

This file contains the supplemental materials for *Machine learning models for temporally precise lapse prediction in alcohol use disorder*. It includes a transparency report and all supplemental figures. Additional materials are made available on our study's OSF page (https://osf.io/w5h9y/).

Supplemental Methods

Lapse Labels

We predicted future lapses in three window widths: one week, one day, and one hour. Prediction windows were updated hourly. All classification models provide hour-by-hour predictions of future lapse probability for all three window widths.

For each participant, the start of the first prediction window for all three widths began at midnight on their second day of participation and ended one week, one day, or one hour later. By beginning at the end of the second day, we were assured that there would be at least 24 hours of past EMAs to use for future lapse prediction in these first windows. Subsequent windows for each participant were created for all three widths by repeatedly rolling the window start/end forward one hour until the end of their study participation was reached (i.e., each participant's last prediction window for each width ended at the date and hour of their last recorded EMA).

We labeled each prediction window as *lapse* or *no lapse* using participants' reports from the EMA item "Have you drank any alcohol that you have not yet reported?". If participants answered yes to this question, they were prompted to enter the hour and date of the start and end of the drinking episode. These reports were validated by study staff during monthly followup visits.

A prediction window was labeled *lapse* if the start date/hour of any drinking episode fell within that window. Conversely, a window was labeled *no lapse* if no alcohol use occurred within 24 hours of the window start/end. If no alcohol use occurred within the window but did occur within the 24 hours of the start or end of the window, the window was excluded. We used this conservative 24-hour fence for labeling windows as *no lapse* (vs. excluded) to increase the fidelity of these labels. Given that most windows were labeled *no lapse* (i.e., *no lapse* was the majority class, and the outcome was highly unbalanced), it was not problematic to exclude some *no lapse* events to further increase confidence in those labels.

Feature Engineering

Other generic feature engineering steps included: 1) imputation for missing data for features (median imputation for numeric features, mode imputation for nominal features); 2) dummy coding for nominal features; and 3) removal of any zero variance features. A sample feature engineering script (i.e., tidymodels recipe) containing all feature engineering steps is available on our OSF study page. Medians/mode for missing data imputation and identification of zero variance features were derived from training (held-in) data and applied to held out (validation and test) data to prevent issues associated with data leakage.

Bayesian Analyses

Bayesian analyses were accomplished using the tidyposterior [1] and rstanarm [2] packages in R. Following recommendations from the rstanarm team and others [3,4], we used the rstanarm default autoscaled, weakly informative, data-dependent priors that take into account the order of magnitude of the variables to provide some regularization to stabilize computation and avoid over-fitting. Specifically, the priors were set as follows: residual standard deviation ~ normal(location=0, scale=exp(2)), intercept (after centering predictors) ~ normal(location=2.3, scale=1.3), the two coefficients for window width contrasts ~ normal (location=0, scale=2.69), and covariance ~ decov(regularization=1, concentration=1, shape=1, scale=1).

Transparency Report 1.0 (full, 36 items)[5]

Manuscript Title: Personal sensing for temporally precise lapse risk prediction for alcohol use disorder
Authors: Kendra Wyant, Sarah Sant'Ana, Gaylen E. Fronk, John J. Curtin
Corresponding author's email address: jjcurtin@wisc.edu
Link to Project Repository: https://osf.io/w5h9y/

Preregistration Section

- Prior to analyzing the complete data set, a time-stamped preregistration was posted in an independent, third-party registry for the data analysis plan. No

Comments about your Preregistration:
Throughout this project, we iteratively improved machine learning methods that are rapidly evolving in the social sciences and used in this study. However, we restricted many researcher degrees of freedom via cross-validation procedures that can robustly guide decision-making. Replication is built into cross-validation; models are fit using held-in training sets, decisions are made using held-out validation sets, and final model performance is evaluated in a confirmatory manner using held-out test sets.

Methods Section

The manuscript fully describes…

- the rationale for the sample size used (e.g., an a priori power analysis). Yes
- how participants were recruited. Yes
- how participants were selected (e.g., eligibility criteria). Yes
- what compensation was offered for participation. Yes
- how participant dropout was handled (e.g., replaced, omitted, etc). Yes
- how participants were assigned to conditions. N/A. There are no conditions.
- how stimulus materials were randomized. N/A.
- whether (and, if so, how) participants, experimenters, and data-analysts were kept naive to potentially biasing information. N/A. This is an observations study that does not include analysis of group or manipulations. There were no study conditions to blind.
- the study design, procedures, and materials to allow independent replication. Yes
- the measures of interest (e.g., friendliness). Yes
- all operationalizations for the measures of interest (e.g., a questionnaire measuring friendliness). Yes

Results and Discussion Section

The manuscript…

- distinguishes explicitly between “confirmatory” (i.e., prespecified) and “exploratory” (i.e., not prespecified) analyses. There were no pre-registered, “confirmatory” analyses in this study. The analyses in the study are primarily descriptive.

- describes how violations of statistical assumptions were handled. No

- justifies all statistical choices (e.g., including or excluding covariates; applying or not applying transformations; use of multi-level models vs. ANOVA). Yes

- reports the sample size for each cell of the design. Yes

- reports how incomplete or missing data were handled. Yes

- presents protocols for data preprocessing (e.g., cleaning, discarding of cases and items, normalizing, smoothing, artifact correction). Yes

Data, Code, and Materials Availability Section

The following have been made publicly available…

- the (processed) data, on which the analyses of the manuscript were based. Yes

- all code and software (that is not copyright protected). Yes

- all instructions, stimuli, and test materials (that are not copyright protected). Yes

- Are the data properly archived (i.e., would a graduate student with relevant background knowledge be able to identify each variable and reproduce the analysis)? Yes

- The manuscript includes a statement concerning the availability and location of all research items, including data, materials, and code relevant to the study. Yes

References

[1] Kuhn M. Tidyposterior: Bayesian Analysis to Compare Models using Resampling Statistics 2022.

[2] Goodrich B, Gabry J, Ali I, Brilleman S. Rstanarm: Bayesian Applied Regression Modeling via Stan 2023.

[3] RStudio Team. RStudio: Integrated Development for R 2020.

[4] Gabry J, Goodrich B. Prior Distributions for rstanarm Models. CRAN R-Project 2023.

[5] Aczel B, Szaszi B, Sarafoglou A, Kekecs Z, Kucharský Š, Benjamin D, et al. A consensus-based transparency checklist. Nature Human Behaviour 2019:1–3. https://doi.org/10.1038/s41562-019-0772-6.

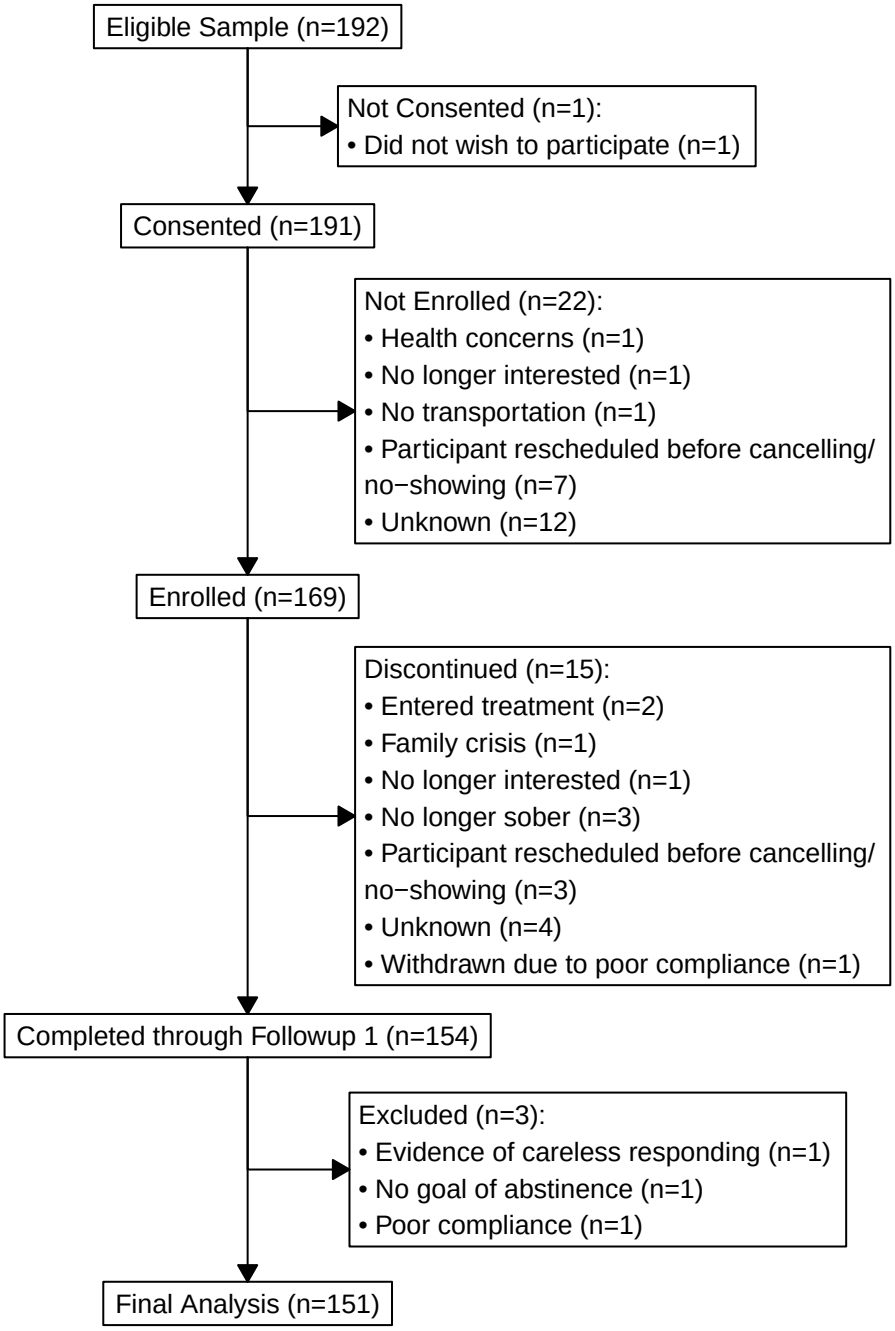

Figure S1 : CONSORT diagram. The diagram depicts participant retention at each study milestone. It also displays reasons for discontinuation when known and reasons for data exclusions.

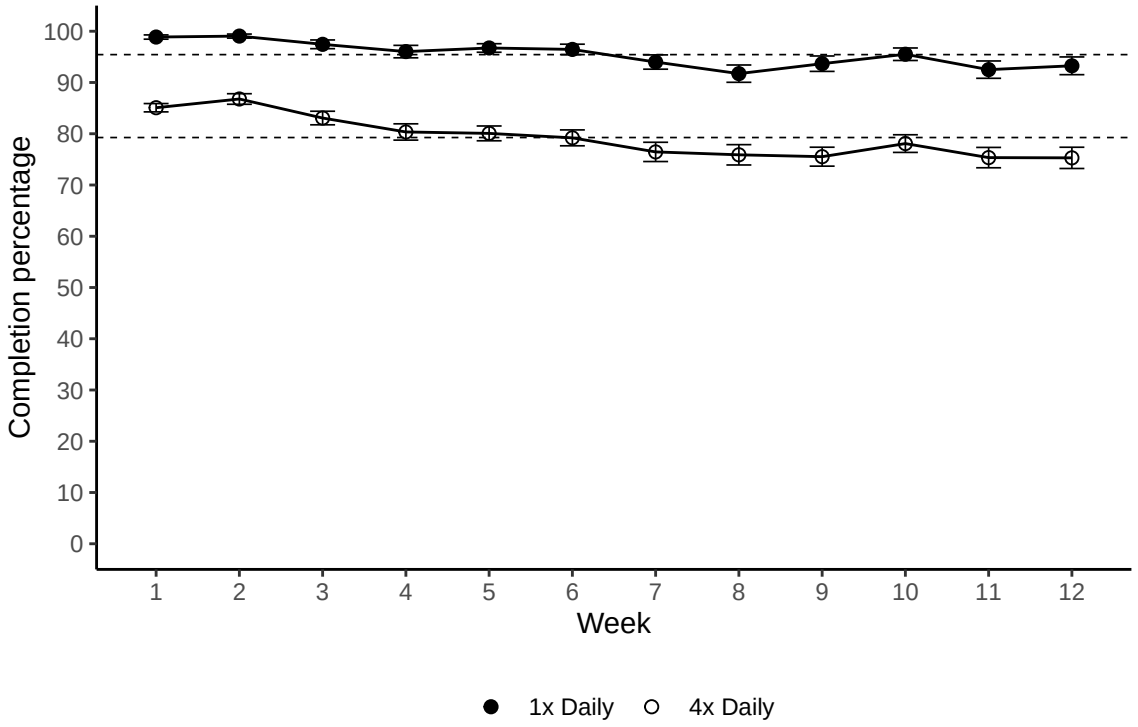

Figure S2 : EMA completion by week. The plot depicts completion percentages over time (by week) across the study period for 1x (closed circles) and 4x (open circles) daily EMA. Dashed lines represent mean EMA completion over entire study period for 1x and 4x daily EMA. Error bars represent the standard error for each completion percentage by week.

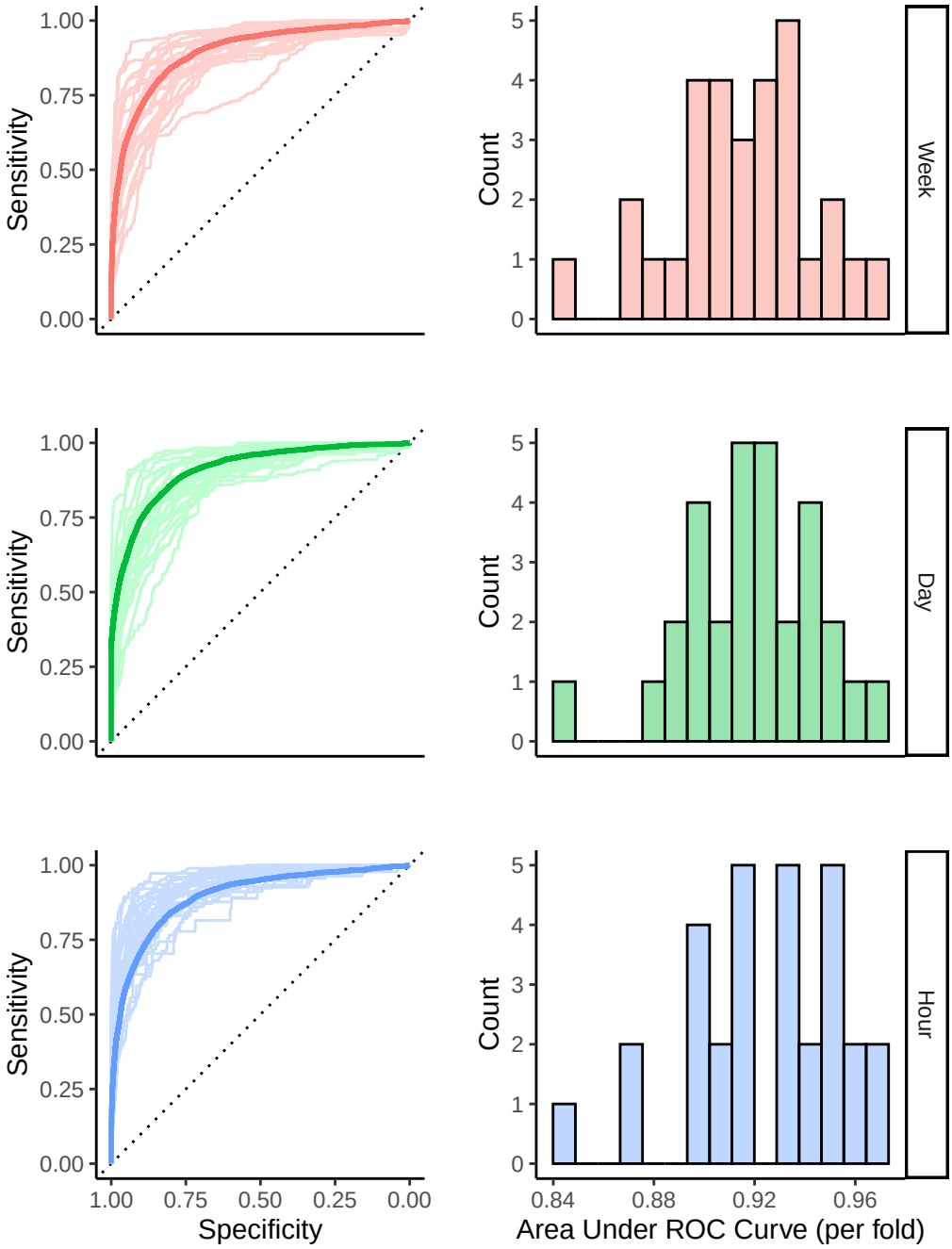

Figure S3 : ROC curves and auROCs. The plots on the left depict individual receiver operating characteristic (ROC) curves from each of the 30 test sets. The darker curves represent the aggregate ROC curve derived by concatenating all held out folds. The dotted line represents the expected ROC curve for a random classifier. The histograms on the right depict the distribution of areas under the ROC curves (auROCs) from the same 30 test sets. The rows are organized by model (week, day, hour).

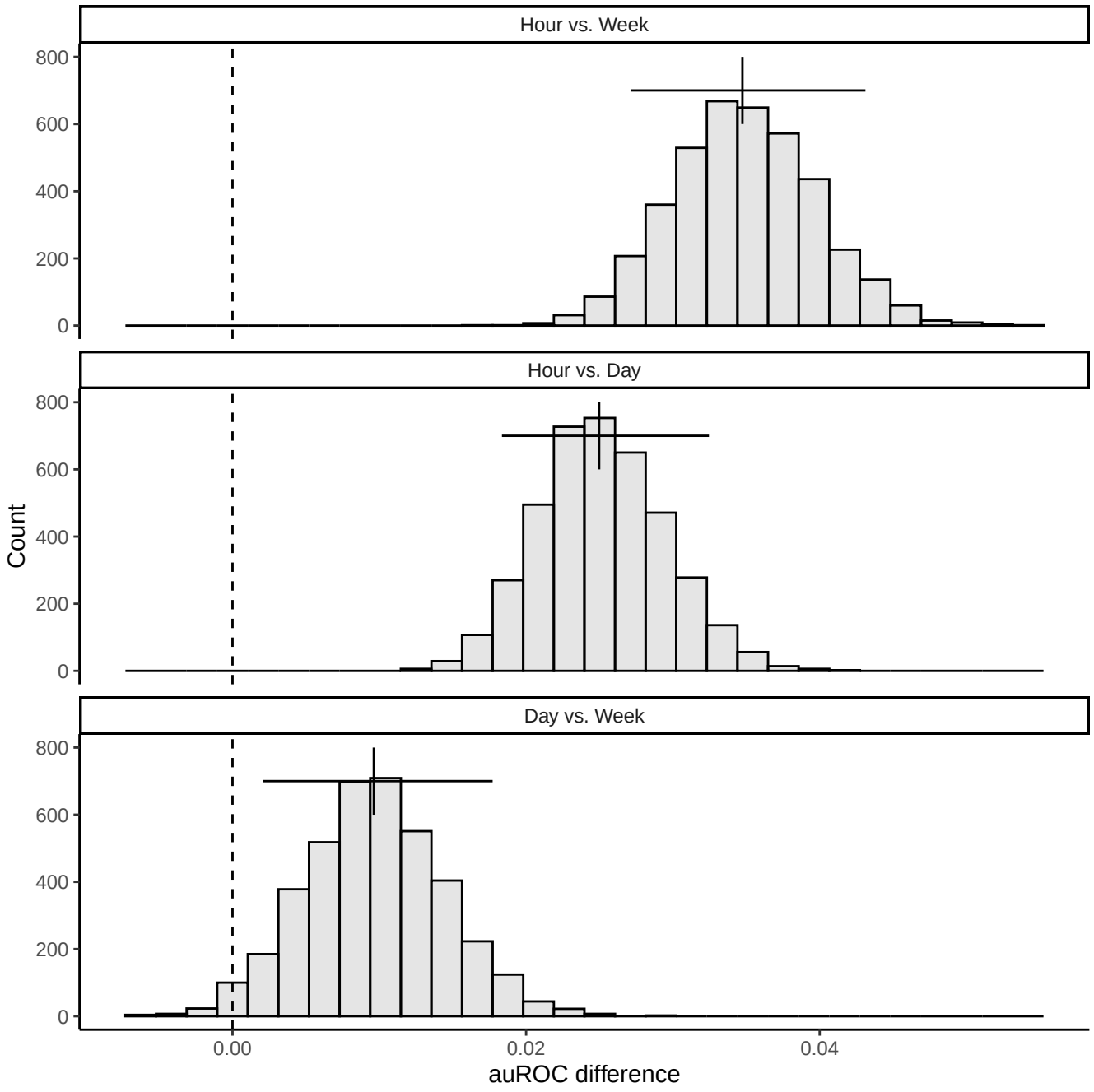

Figure S4 : Bayesian model contrasts. The plots above depict the posterior probabilities for the areas under the receiver operating characteristic curves (auROCs) of our model contrasts (i.e., the difference in auROC between the two models). Each row represents a model constrast (hour vs. week, hour vs. day, day vs. week). The solid vertical lines represent the median posterior probability. The horizontal lines represent the 95% credible interval. The dashed vertical lines indicate an auROC model difference of 0 which denotes no difference in performance.

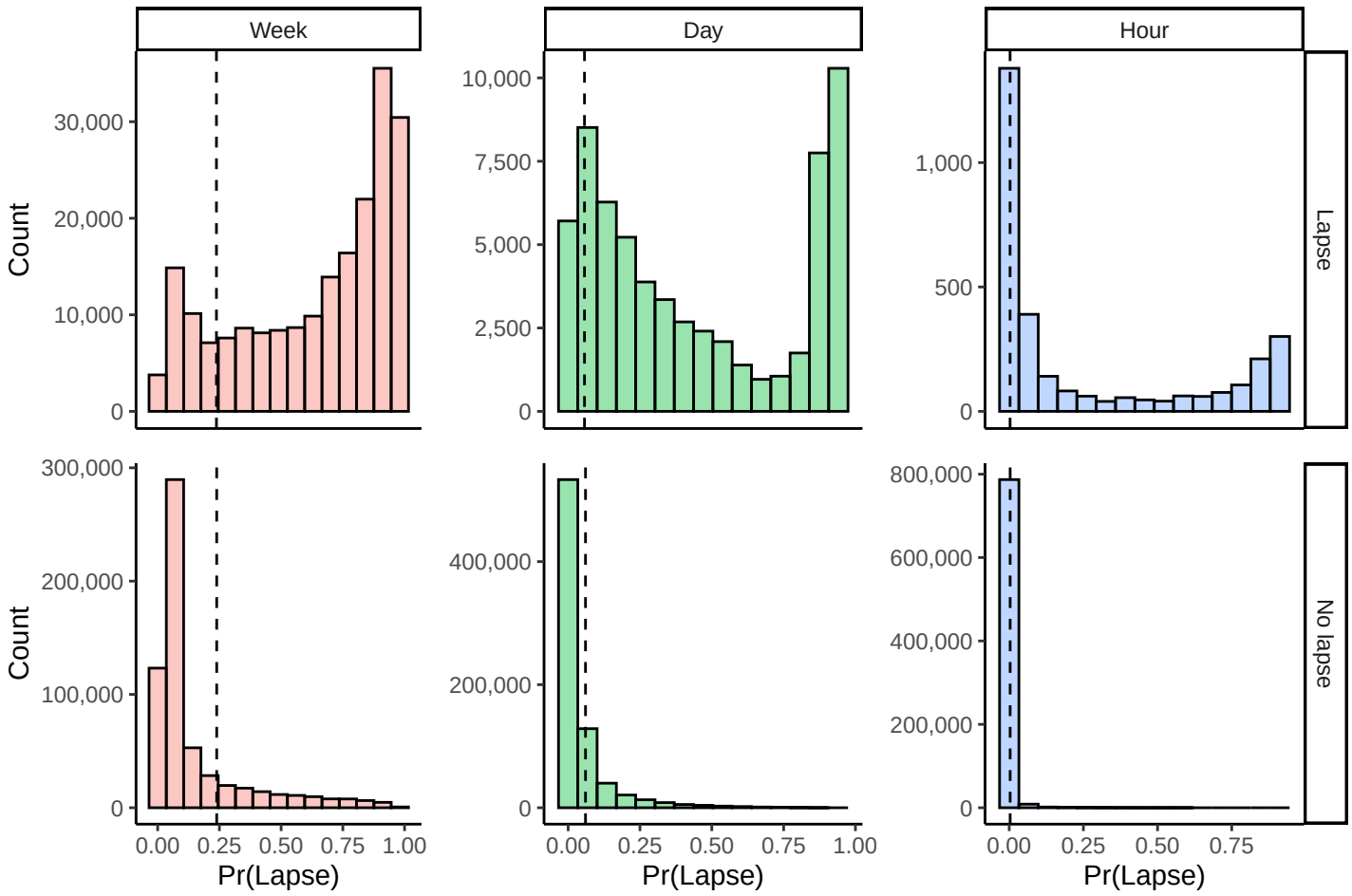

Figure S5 : Lapse probability predictions by model. The plots above depict predicted probabilities for all observations in the 30 test sets. The columns are organized by model (week, day, hour). The top row depicts estimated lapse probabilities for true lapses. The bottom row depicts estimated lapse probabilities for true no lapses. The dashed vertical lines represent the decision threshold for each model, determined using Youden's index.