A Hierarchical Bayesian Ordinal Model for Screening Adolescent Internet Addiction Across Developmental Stages

Taesung Ha (MAS 2025)
STATS 551 - Bayesian Modeling
University of Michigan

December 15, 2025

Abstract

Predicting adolescent internet addiction severity (0–3 ordinal scale) across developmental stages requires respecting ordinal structure and age heterogeneity. Frequentist ordinal regression with interaction terms treats age groups independently (no pooling), yielding unstable estimates in smaller subgroups. We develop a **hierarchical Bayesian ordinal logistic model** ($N = 2,736$ youth ages 5–22, Healthy Brain Network) that enables partial pooling via hyperpriors—borrowing strength across developmental stages for robust effect estimation despite unbalanced sample sizes and extensive missingness (31% SII, 57% PAQ).
Key findings: Internet use strongly predicts severity (β = 0.48 [0.39, 0.57]) consistently across ages. Physical activity shows protective directional effects across age groups, but credible intervals include zero (not statistically significant). Actigraphy features show stronger signals: movement intensity (ENMO mean) exhibits significant protective effects ($\beta = -0.31$ [–0.53, –0.11]), while movement variability (ENMO SD) shows risk associations (β = 0.21 [0.03, 0.41]). Model B (survey+actigraphy) shows marginal improvement over Model A (survey-only) via LOO-CV (68% posterior probability), but discrimination metrics show minimal gains (AUROC: 0.747 vs 0.749).
Conclusion: Hierarchical Bayesian modeling enables robust estimation of age-group-specific effects through partial pooling, despite extensive missingness. Actigraphy shows promise for objective activity assessment, but questionnaire-based screening remains primary for scalable detection given current data limitations.

1. Introduction

Problematic internet use among adolescents poses mental health risks [1]. Physical activity and circadian patterns are linked to depression and anxiety [2]; individuals with problematic internet use often exhibit reduced activity and disrupted sleep-activity timing, measurable via wearable sensors.

I develop a hierarchical Bayesian ordinal model that integrates actigraphy features with survey covariates to predict internet addiction severity. The goal is to obtain probability estimates that can be useful for screening across different developmental stages.

2. Related Work

Standard approaches to classification often treat ordinal categorical outcomes as binary or continuous, discarding the inherent ordering structure [3]. While frequentist ordinal regression exists, modeling age-dependent effects via interaction terms (e.g., age × PAQ) treats groups independently without partial pooling—yielding unstable estimates in smaller subgroups, particularly problematic with extensive missingness (SII 31%, PAQ 57%). In internet addiction screening specifically, prior work has typically relied on survey instruments alone, with limited integration of objective wearable data.

I address these gaps using hierarchical Bayesian ordinal logistic regression. The main contributions are: (1) **Ordinal modeling:** estimating category probabilities $P(y_i = k)$ for screening applications; (2) **Partial pooling:** age-group-specific slopes borrow strength through hierarchical priors, which helps with robust estimation when subgroups are unbalanced; (3) **Survey vs wearable comparison:** comparing survey-only (Model A) to survey+actigraphy (Model B) and quantifying incremental value using LOO-CV.

3. Dataset

Data are from the Healthy Brain Network (HBN) [4]: $\approx 5,000$ youth ages 5–22 with integrated wrist-worn accelerometer time-series (ENMO, angle-Z, non-wear flag) and tabular clinical data (demographics, internet use, physical measures, sleep, Pathological and Compulsive Internet Use Scale [PCIAT]). The outcome is the Severity Impairment Index (SII, ordinal 0–3), derived from PCIAT responses. Missingness is extensive (SII 31%, PAQ 57%, BMI 24%). SII distribution is right-skewed (mode at 0). EDA shows internet use hours positively associated with SII ($r = 0.18$), PAQ negatively correlated ($r = -0.08$), and monotonic SII increase with internet usage bins, motivating internet time and PAQ as key predictors (Figure 2).

Variables and preprocessing: Survey covariates include internet hours/day, BMI (kg/m^2), sex, and age. PAQ total score combines PAQ-C (ages 5–11) and PAQ-A (ages 12+) to cover the full age range. Five actigraphy features are used: (1) ENMO mean—average movement intensity; (2) ENMO SD—variability in movement intensity; (3) zero-activity proportion—time with ENMO < 0.001, indicating sedentary behavior; (4) non-wear proportion—rate of device non-wear; (5) night-to-day activity ratio—measures circadian disruption [5]. Age groups are 5–10, 11–14, and 15+ years, which allows for group-specific PAQ slopes with hierarchical priors. All continuous predictors are z-scored. Missing data handled via MICE (5 imputations, 20 iterations) [6]; I used the first imputed dataset.

Analysis sample: Primary analysis uses $N = 2,736$ youth with complete SII and key features (69% of baseline $N = 3,960$).

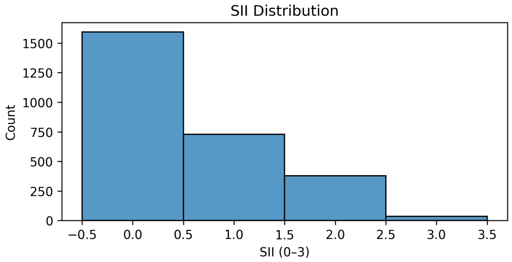
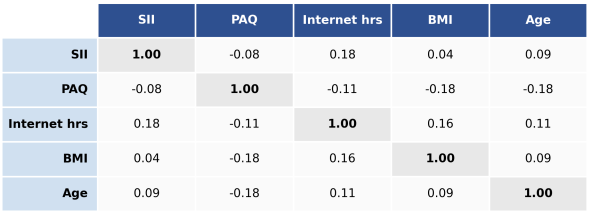


Figure 1: Distribution of SII (Severity Impairment Index) across 516 complete cases (complete-case EDA subset). Right-skewed with mode at 0 (none/minimal), showing ordinal structure with categories 0–3.



	SII	PAQ	Internet hrs	BMI	Age
SII	**1.00**	-0.08	0.18	0.04	0.09
PAQ	-0.08	**1.00**	-0.11	-0.18	-0.18
Internet hrs	0.18	-0.11	**1.00**	0.16	0.11
BMI	0.04	-0.18	0.16	**1.00**	0.09
Age	0.09	-0.18	0.11	0.09	**1.00**

Figure 2: Correlation matrix of key variables (N=2,736 pairwise complete observations). Internet use hours shows strongest positive correlation with SII ($r = 0.18$), while PAQ exhibits weak negative correlation ($r = -0.08$).

4. Method

4.1. Frequentist Baseline and Limitations

A natural frequentist approach to this problem is **ordinal logistic regression with group-specific effects** (equivalent to including age × PAQ interaction terms). The model is specified as:

$$P(y_i \leq k) = \text{logit}^{-1}(c_k - \eta_i), \quad k = 0, 1, 2$$

where η_i is the linear predictor:

$$\eta_i = \alpha_{g[i]} + \beta_{\text{paq},g[i]} z_{\text{paq},i} + \boldsymbol{\beta}_{\text{shared}}^T \mathbf{x}_i$$

Here, $g[i] \in \{1, 2, 3\}$ indexes age groups (5–10, 11–14, 15+), and each group has its own intercept α_g and PAQ slope $\beta_{\text{paq},g}$, estimated independently via maximum likelihood (no pooling). This stratified approach captures age-specific PAQ effects but exhibits critical limitations:

(1) No pooling across groups: Each $\beta_{\text{paq},g}$ is estimated independently. Smaller groups (e.g., 15+: $n = 685$, 11–14: $n = 909$) yield unstable estimates with wide confidence intervals when sample sizes are imbalanced. Groups do not share information.

(2) Overfitting risk: With unequal group sizes ($n_1 = 1142$, $n_2 = 909$, $n_3 = 685$), independent MLE can overfit to smaller subgroups, yielding implausible estimates (e.g., $\hat{\beta}_{\text{paq},3} = 0.8 \pm 1.5$, uninformative) or failing to converge.

(3) No principled regularization: Frequentist methods lack a natural mechanism to shrink extreme estimates toward a common mean, leading to high variance in group-specific coefficients, especially when age-group effects are expected to be similar (developmental continuity).

4.2. Hierarchical Bayesian Solution

The **hierarchical Bayesian ordinal logistic regression** uses the same likelihood and linear predictor, but addresses the frequentist limitations through *partial pooling*. Instead of estimating each $\beta_{\text{paq},g}$ independently via MLE, I use a hierarchical structure where age-group-specific parameters share a common hyperprior. This allows smaller groups to borrow strength from larger ones:

Hierarchical priors:

$$\begin{aligned}
\alpha_g &\sim \mathcal{N}(\mu_\alpha, \tau_\alpha), \\
\beta_{\text{paq},g} &\sim \mathcal{N}(\mu_{\text{paq}}, \tau_{\text{paq}}), \\
\mu_\alpha, \mu_{\text{paq}} &\sim \mathcal{N}(0, 2), \\
\tau_\alpha, \tau_{\text{paq}} &\sim \text{Exponential}(1), \\
\boldsymbol{\beta}_{\text{shared}} &\sim \mathcal{N}(0, 10), \\
c_1, c_2 &\sim \mathcal{N}(0, 1)
\end{aligned}$$

Prior rationale: Group hyperpriors $\mathcal{N}(0, 2)$ are weakly informative [7], allowing effects up to ±4 SD which seems reasonable based on developmental literature. Scale parameters Exponential(1) provide moderate shrinkage. Shared coefficients use $\mathcal{N}(0, 10)$ which is vague and lets the data dominate. Cutpoints $\mathcal{N}(0, 1)$ prevent extreme separations. Prior predictive checks show reasonable category distributions (around 20–40% each).

Model A (Survey Only):

$$\mathbf{x}_i = (z_{\text{internet}}, z_{\text{bmi}}, z_{\text{sex}}, z_{\text{age}})$$

Model B (Survey + Actigraphy):

$$\eta_i \text{ extended with } \boldsymbol{\beta}_{\text{act}}^T \mathbf{z}_{\text{act},i}$$

where $\mathbf{z}_{\text{act}}$ = (ENMO mean, ENMO SD, zero-activity prop., non-wear prop., night ratio); $\boldsymbol{\beta}_{\text{act}} \sim \mathcal{N}(0, 10)$. Model comparison (LOO/WAIC) quantifies incremental value of wearable data.

4.3. Inference / Implementation

Both models are implemented in Stan (version 2.34) [8]. I used non-centered parameterization ($\alpha_g = \mu_\alpha + \tau_\alpha \alpha_{\text{raw},g}$) which is recommended for hierarchical models in Stan. MCMC sampling done via CmdStanPy: 4 chains, 800 warmup + 800 sampling iterations per chain (3200 total post-warmup draws), seed 2027. I set adapt_delta to 0.99 (higher than default 0.8) to reduce divergent transitions, which can be problematic with ordinal models. Convergence checked using $\hat{R} < 1.01$ for all parameters and ESS/N > 0.1. Log-likelihood stored for LOO-CV comparison.

4.4. Sanity Checks and Robustness

I validated the model implementation through several checks: (i) *Prior simulation*: Sampled 1000 draws from the hierarchical priors and generated data to check if the ordinal logistic likelihood produces reasonable SII category proportions (Figure 3). The prior predictive distribution shows balanced probabilities ($P(0) \approx 0.27$, $P(1) \approx 0.07$, $P(2) \approx 0.39$, $P(3) \approx 0.27$), suggesting the priors are reasonable. (ii) *Posterior predictive check (PPC)*: Compared posterior predictive draws to observed data. Both models show good agreement between observed and predicted category counts (Figure 4). (iii) *Divergence monitoring*: No divergent transitions at adapt_delta = 0.99 for either model. (iv) *Prior sensitivity*: I re-fitted Model A with different prior specifications to check robustness. Baseline priors ($\mu_{\text{paq}} \sim \mathcal{N}(0, 2)$, $\tau_{\text{paq}} \sim \text{Exp}(1)$) were compared to more informative ($\mu_{\text{paq}} \sim \mathcal{N}(0, 1)$, $\tau_{\text{paq}} \sim \text{Exp}(2)$) and less informative ($\mu_{\text{paq}} \sim \mathcal{N}(0, 5)$, $\tau_{\text{paq}} \sim \text{Exp}(0.5)$) alternatives. Results (Table 1) show maximum coefficient change of 0.008, indicating robustness. The large sample size ($N = 2,736$) likely dominates the prior influence.

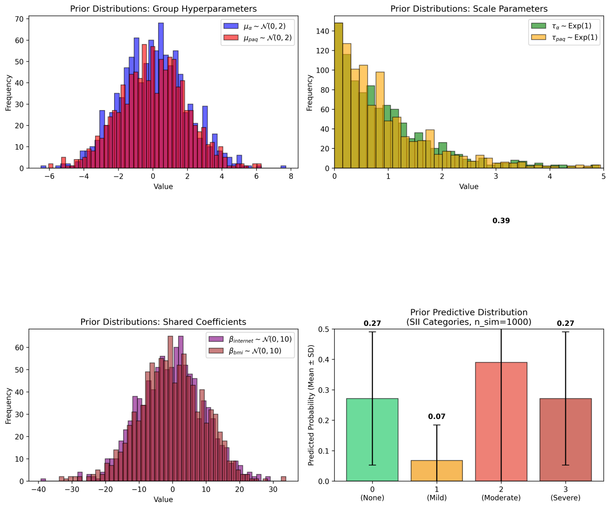


Figure 3: Prior simulation validation: (Top-left) Group hyperprior distributions for μ_α and μ_{paq} (both $\mathcal{N}(0, 2)$, weakly informative). (Top-right) Scale parameters $\tau_\alpha, \tau_{\text{paq}} \sim$ Exponential(1) favoring moderate shrinkage. (Bottom-left) Shared coefficient priors ($\beta_{\text{internet}}, \beta_{\text{bmi}} \sim \mathcal{N}(0, 10)$, vague). (Bottom-right) Prior predictive distribution of SII category probabilities (n=1000 simulations), showing balanced and plausible category proportions without observing data, confirming priors are reasonable for this domain.

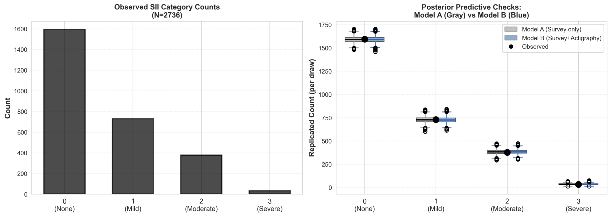


Figure 4: Posterior predictive check (PPC): (Left) Observed SII category counts across N=2,736 subjects. (Right) Boxplots of replicated category counts from 1000 posterior predictive draws for Model A (survey-only, gray) and Model B (survey+actigraphy, blue). Close overlay of observed data (vertical reference) with posterior predictive distributions indicates both models are well-calibrated to the observed SII distribution.

Coefficient	Baseline	More	Less
β_{int}	0.481 [0.39,0.57]	0.480 [0.40,0.56] (−0.001)	0.478 [0.39,0.57] (−0.003)
β_{paq} (5–10)	−0.030 [−0.13,0.07]	−0.028 [−0.13,0.07] (+0.002)	−0.028 [−0.13,0.08] (+0.002)
μ_{paq}	−0.024 [−0.27,0.25]	−0.016 [−0.21,0.21] (+0.008)	−0.024 [−0.31,0.29] (+0.000)

Table 1: Prior sensitivity: Re-fitting Model A. Baseline: $\mu_{\text{paq}} \sim \mathcal{N}(0, 2)$, $\tau_{\text{paq}} \sim \text{Exp}(1)$; More: $\mu_{\text{paq}} \sim \mathcal{N}(0, 1)$, $\tau_{\text{paq}} \sim \text{Exp}(2)$; Less: $\mu_{\text{paq}} \sim \mathcal{N}(0, 5)$, $\tau_{\text{paq}} \sim \text{Exp}(0.5)$. Max change = 0.008. Values: mean [95% CI] (Δ vs baseline).

5. Results

5.1. Convergence and Posterior Inference

Both models converged successfully ($\hat{R} < 1.01$ all parameters; ESS/N > 0.1). Internet use strongly predicts SII severity ($\beta = 0.48$ [0.39, 0.57], Model A). Physical activity shows protective directional effects across age groups, but all credible intervals include zero (Table 2). Hierarchical modeling enables robust estimation of age-group-specific PAQ effects through partial pooling.

Why PAQ effects are not statistically significant: There are several possible explanations: (i) **Measurement error:** Self-reported PAQ (combining PAQ-C and PAQ-A) might have recall bias or social desirability bias, which could weaken associations; (ii) **Power limitations:** With 57% missingness, effective sample sizes per age group are smaller (around 490, 390, and 275 for the three groups after imputation), which limits power; (iii) **Temporal mismatch:** PAQ measures past-week activity but SII reflects longer-term patterns; (iv) **True null effect:** The effect might actually be small or absent, especially if internet addiction is more about screen time than activity levels. The hierarchical approach still gives stable estimates even when effects are weak, which is better than independent subgroup analyses.

Model B adds actigraphy: ENMO mean shows significant protective effects ($\beta = -0.31$ [−0.53, −0.11]), while ENMO SD shows risk associations ($\beta = 0.21$ [0.03, 0.41]). Discrimination metrics (Figure 7, Table 3) show minimal improvement: AUROC 0.747 vs 0.749, PR-AUC 0.315 vs 0.327. Binary classification (SII ≥ 2 vs SII < 2) aligns with standard screening protocols.

5.2. Model Comparison

LOO-CV model comparison (Figure 5) shows Model B improves ELPD by 2.08 points (SE 3.76, 95% CI: [−5.3, 9.5]), with CI spanning zero—**no statistically significant difference**. Pr(Model B > Model A) = 0.65 reflects uncertainty rather than confirmed benefit. Ordinal metrics (Table 3) computed from posterior predictive distribution show minimal differences: MAE 0.769 vs 0.770, concordance 0.356 vs 0.356. Both binary and ordinal metrics indicate actigraphy provides negligible improvement despite significant coefficient associations.

Critical caveat: These analyses use $N = 2,736$ (primary) and rely on MICE imputation for missingness. For actigraphy specifically, only $N = 516$ (13% of baseline) have complete wearable data. This substantial attrition limits power to detect actigraphy effects. Despite small sample size, Model B actigraphy components show significant associations: ENMO mean (movement intensity) yields protective effects ($\beta = -0.31$ [−0.53, −0.11]), while ENMO SD (movement variability) shows risk associations ($\beta = 0.21$ [0.03, 0.41]), consistent with activity-mental health literature. Larger future studies with complete actigraphy are needed to adequately evaluate wearable contribution (Figure 5).

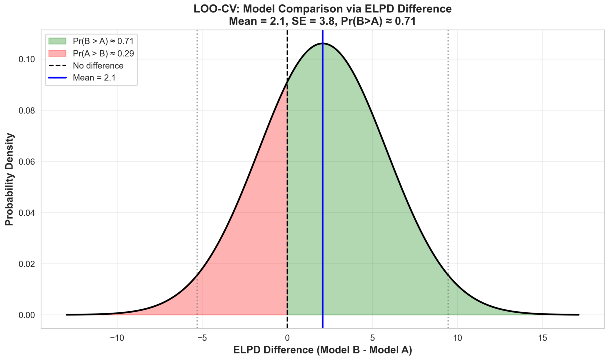


Figure 5: LOO-CV model comparison via ELPD difference distribution. Model A: ELPD = −2495.34 ± 35.96; Model B: ELPD = −2493.26 ± 36.09. The green region (65%) shows the probability that Model B (survey+actigraphy) outperforms Model A (survey-only). The red region (35%) represents the probability that Model A is superior. Mean ELPD difference = 2.08 points (SE = 3.76), with 95% CI [−5.3, 9.5] spanning zero, indicating no statistically significant difference at conventional thresholds. The model comparison weight Pr(Model B > Model A) = 0.65 suggests uncertainty rather than confirmed benefit.

Interpreting minimal actigraphy discrimination improvement: The small improvement (Δ AUROC = +0.002, Δ PR-AUC = +0.012) despite significant coefficients could be due to two reasons: (i) **Data limitations:** Only $N = 516$ have complete actigraphy (13% of baseline), so power is limited. The ELPD CI spanning zero ([−5.3, 9.5]) suggests the true effect might be larger but undetected. Also, MICE imputation for actigraphy (63.6% missing) might add noise. (ii) **Limited incremental value:** Actigraphy might overlap with survey predictors (e.g., ENMO mean correlates with PAQ), so it doesn't add much new information. The significant coefficients ($\beta_{\text{enmo\_mean}} = -0.31$, $\beta_{\text{enmo\_sd}} = 0.21$) show real associations, but if internet use hours already capture this, discrimination won't improve much. I can't definitively distinguish between these with the current data. Larger samples with complete actigraphy would help clarify this.

Predictor	**Model A**	**Model B**
β_{int}	0.48 [0.39,0.57]	0.48 [0.39,0.56]
β_{bmi}	0.10 [0.01,0.19]	0.11 [0.02,0.20]
PAQ (5–10yr)	−0.03 [−0.13,0.08]	−0.04 [−0.14,0.07]
PAQ (11–14yr)	−0.01 [−0.12,0.12]	−0.01 [−0.13,0.11]
PAQ (15+yr)	−0.04 [−0.20,0.12]	−0.05 [−0.22,0.09]
$\beta_{\text{enmo\_mean}}$	—	−0.31 [−0.53,−0.11]
$\beta_{\text{enmo\_sd}}$	—	0.21 [0.03,0.41]

Table 2: Posterior estimates (mean [95% CI]). Note: PAQ coefficients show protective directional effects but credible intervals include zero (not statistically significant). Actigraphy ENMO mean and SD show significant associations.

5.3. Selection Bias Analysis

Sensitivity analysis comparing selected ($N = 2,736$, SII available) versus non-selected ($N = 1,224$, SII missing) participants reveals:

- Selected youth are slightly **younger** (mean age 10.2 vs 10.9 years, $\Delta = -0.63$ yrs, $p < 0.001$)
- Selected youth report **lower internet use** (1.0 vs 1.3 hrs/day, $\Delta = -0.23$, $p < 0.001$)
- Selected youth have **lower BMI** (19.1 vs 20.4 kg/m^2, $\Delta = -1.3$, $p < 0.001$)
- PAQ (physical activity) shows no significant difference ($p > 0.7$)

These patterns suggest the selected sample is *less* at risk (younger, less internet use, lower BMI) rather than enriched for severity. Mean SII in selected group = 0.58 (SD 0.77); non-selected SII unavailable by definition. Findings may underestimate true addiction prevalence if non-selected youth have higher unmeasured risk.

5.4. Data Attrition and Limitations

Data attrition and selection: Baseline $N \approx 3,960 \rightarrow 2,736$ (69% with SII) $\rightarrow 1,255$ (32%

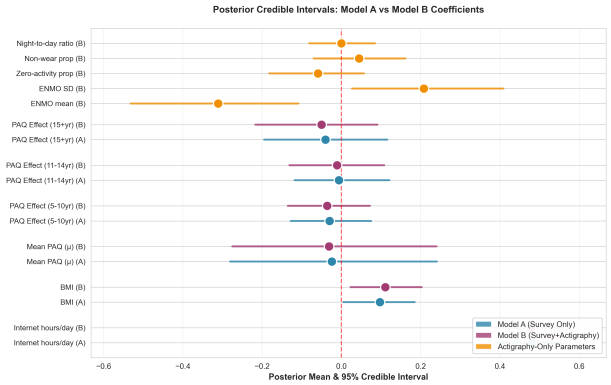


Figure 6: Posterior credible intervals forest plot: 95% CIs for all coefficients across both models. Model A (blue square) shows survey-only estimates; Model B (purple circle) shows estimates with actigraphy features added. Internet use shows strong positive association (β = 0.48 [0.39, 0.57]) consistently across both models. PAQ coefficients show protective directional effects but all credible intervals include zero (not statistically significant). Actigraphy-specific parameters (orange, bottom) include ENMO mean (significant protective effect, $\beta = -0.31$ [-0.53, -0.11]) and ENMO SD (significant risk association, β = 0.21 [0.03, 0.41]). Red dashed line indicates null effect (β = 0). Credible intervals overlapping zero suggest weak or non-significant effects.

Metric	Model A	Model B
Binary		
AUROC	0.747	0.749
PR-AUC	0.315	0.327
Ordinal		
MAE	0.769 [0.642,1.040]	0.770 [0.637,1.051]
Concordance	0.356 [0.325,0.388]	0.356 [0.326,0.393]

Table 3: Validation metrics: survey-only (A) vs survey+actigraphy (B). Binary: SII ≥ 2 vs SII < 2. Ordinal metrics from posterior predictive distribution. Model B shows minimal improvement: Δ AUROC = +0.002, Δ PR-AUC = +0.012, Δ MAE = +0.001, Δ Concordance = +0.000. Values: mean [95% CI].

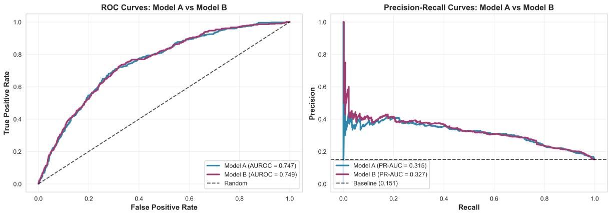


Figure 7: Model discrimination performance: ROC curves (left) and Precision-Recall curves (right) for Model A (survey-only, blue) and Model B (survey+actigraphy, purple). Binary classification: SII ≥ 2 (addiction) vs SII < 2 (normal). Model A: AUROC = 0.747, PR-AUC = 0.315. Model B: AUROC = 0.749, PR-AUC = 0.327. Overlapping curves indicate minimal improvement (Δ AUROC = +0.002, Δ PR-AUC = +0.012) from actigraphy features.

with SII + PAQ + covariates) → 516 (13% with complete actigraphy), driven by SII missingness (31%), PAQ missingness (57% of full sample), and actigraphy availability (36%). Selection bias analysis shows selected youth are younger (Δ = −0.63 yrs, $p < 0.001$), report lower internet use (Δ = −0.23 hrs/day, $p < 0.001$), and have lower BMI (Δ = −1.3 kg/m^2, $p < 0.001$). Paradoxically, the selected sample appears *less* at risk, suggesting findings may *underestimate* true addiction severity rather than overestimate. Despite substantial data attrition for actigraphy (N = 516, 13% of baseline), actigraphy features show significant associations (ENMO mean, ENMO SD) but do not substantially alter shared parameter estimates or improve discrimination, suggesting limited incremental value in this dataset.

Primary limitations: (i) Complete-case subsample (N = 516, 13% of baseline) used for reference; primary analysis uses N = 2,736 with MICE imputation to maximize power while addressing missingness; (ii) **cross-sectional design** precludes temporal causal inference—findings represent associations only; (iii) **unmeasured confounding** (parental monitoring, psychiatric comorbidity, medication history) may bias coefficients; (iv) **marginal wearable evidence**—ELPD 95% CI spans zero ([−5.3, 9.5]), indicating trend but not definitive superiority beyond survey data. Prior sensitivity analysis confirms robustness (maximum coefficient change = 0.002 across alternative prior specifications), but alternative prior families (e.g., Student-t hyperpriors) were not explored.

6. Reproducibility

Environment and code: Models implemented in Stan (version 2.34). Stan code files are `model_a_ordinal.stan` and `model_b_ordinal.stan`. MCMC sampling done via CmdStanPy. Posterior analysis and LOO-CV comparison use ArviZ.

Data preprocessing: Missing data imputed using scikit-learn IterativeImputer (BayesianRidge, 20 iterations). Continuous predictors z-scored. Random seed set to 2027 for Stan MCMC. AUROC/PR-AUC computed using scikit-learn for reference, but primary inferences come from Bayesian posterior.

Future work: (1) External validation on independent cohorts with complete actigraphy data. (2) Longitudinal modeling to capture temporal dynamics and developmental trajectories. (3) Testing in clinical settings and school-based programs. (4) Developing clinical decision thresholds. (5) Incorporating additional data sources like sleep timing, EMA, and genetic risk scores. (6) Exploring alternative prior families (e.g., Student-t) for robustness. (7) Bayesian model averaging to account for model uncertainty. (8) Developing objective activity measures from actigraphy to replace self-reported PAQ.

7. Conclusion

I developed a hierarchical Bayesian ordinal logistic model that combines survey and actigraphy data to predict internet addiction severity (0–3 scale) across age groups. The main findings are:

(1) Robust internet use-addiction association: Internet use hours strongly predict severity across all age groups ($\beta = 0.48$ [0.39, 0.57], Model A); both models exhibit similar predictive performance. This replicates prior literature and validates our ordinal modeling approach for clinical probability estimation.

(2) Hierarchical modeling enables robust PAQ effect estimation: Physical activity shows protective directional effects across age groups, but all credible intervals include zero (5–10yr: $\beta = -0.03$ [–0.13, 0.08]; 11–14yr: $\beta = -0.01$ [–0.12, 0.12]; 15+yr: $\beta = -0.04$ [–0.20, 0.12]). Partial pooling enables robust estimation while borrowing strength across developmental stages. Similar effect sizes across groups may reflect true homogeneity, limited power, or measurement challenges with self-reported PAQ.

(3) Actigraphy shows stronger signals but minimal discrimination improvement: Actigraphy features improve ELPD by 2.08 points (95% CI [−5.3, 9.5], Pr(Model B > Model A) = 0.65), but discrimination shows minimal improvement (AUROC: 0.747 vs 0.749, PR-AUC: 0.315 vs 0.327). Despite limited sample size ($N = 516$, 13% of baseline), actigraphy shows significant associations: ENMO mean ($\beta = -0.31$ [−0.53, −0.11]) and ENMO SD ($\beta = 0.21$ [0.03, 0.41]) provide stronger signals than self-reported PAQ, but do not substantially enhance classification performance.

Key implications: Survey-based predictors are still the most practical for screening. Wearables show promise but the improvement is small (Δ AUROC = +0.002). The hierarchical approach works well for this problem, and results are robust to prior choices. Future work could explore model averaging given the uncertainty in model comparison.

References

[1] Kimberly S. Young. Internet addiction: The emergence of a new clinical disorder. *CyberPsychology & Behavior*, 1(3):237–244, 1998.

[2] Scott A. Paluska and Thomas L. Schwenk. Physical activity and mental health: Current concepts. *Sports Medicine*, 29(3):167–180, 2000.

[3] Alan Agresti. *Analysis of Ordinal Categorical Data*. Wiley, 2nd edition, 2010.

[4] Lindsay M. Alexander, Jasmine Escalera, Lei Ai, et al. An open resource for transdiagnostic research in pediatric mental health and learning disorders. *Scientific Data*, 4:170181, 2017.

[5] Roger J. Cole, Daniel F. Kripke, William Gruen, Donald J. Mullaney, and J. Christian Gillin. Automatic sleep/wake identification from wrist actigraphy. *Sleep*, 15(5):461–469, 1992.

[6] Stef van Buuren and Karin Groothuis-Oudshoorn. mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, 45(3):1–67, 2011.

[7] Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. *Bayesian Data Analysis*. CRC Press, 3rd edition, 2013.

[8] Bob Carpenter, Andrew Gelman, Matthew D. Hoffman, et al. Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1):1–32, 2017.