Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Passive In-Home Sensing for Predicting Next-Day Agitation in Dementia: A Temporal Analysis of In-Home Movement Patterns in the Technology Integrated Health Management (TIHM) Dataset

Yoo-Jeong Jang
Wonkwang University Hospital
Sang-Yeol Lee
Wonkwang University Hospital
Chan-Mo Yang
Wonkwang University Hospital
Dae-Jin Kim
Wonkwang University Hospital
Yeong-Dae Jo
Wonkwang University Hospital
Sung-Hoon Yoon
ysh92614@naver.com

Wonkwang University Hospital

Research Article

Keywords: dementia, psychomotor agitation, activities of daily living, circadian rhythm, sleep, REM, caregivers, telemedicine, digital technology, markov chains, logistic models

Posted Date: June 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9874325/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Agitation is a clinically significant behavioral and psychological symptom of dementia (BPSD), associated with increased patient distress, caregiver burden, and healthcare utilization. Existing approaches to monitoring and predicting agitation often rely on intermittent assessments or self-managed wearable devices, which may limit temporal resolution, adherence, and scalability in real-world dementia care. Passive in-home sensing technologies offer a promising alternative by enabling continuous, non-intrusive monitoring of behavioral patterns without requiring active user engagement.

Objective

We aimed to evaluate the feasibility of predicting next-day agitation in dementia using passive in-home activity sensing and to identify behavioral digital biomarkers of agitation risk.

Methods

We conducted a secondary analysis of the Technology Integrated Health Management (TIHM) dataset, including 56 participants with dementia (2,610 patient-days; 114 agitation events). Daily behavioral features were derived from passive infrared (PIR) motion sensors, including activity volume, location entropy, low-stimulation room ratio, evening activity ratio, and a Markov chain–based self-loop ratio (SLR). Associations with next-day agitation were assessed using generalized estimating equations. Predictive performance was evaluated using leave-one-patient-out cross-validation, with PRAUC as the primary metric. A sleep-sensor subgroup analysis examined the incremental value of sleep features among participants with available sleep data (n = 17).

Results

In generalized estimating equations(GEE) analyses, higher activity volume (odds ratio [OR] 2.16, 95% CI 1.44–3.24; P<.001) and higher evening activity ratio (OR 1.54, 95% CI 1.11–2.15; P=.01) were independently associated with increased next-day agitation risk, whereas higher SLR was strongly associated with reduced risk (OR 0.25, 95% CI 0.18–0.33; P<.001). In LOPO-CV, the model combining PIR features and SLR showed the best performance (AUROC 0.848, 95% CI 0.776–0.901; PRAUC 0.263, 95% CI 0.090–0.395), corresponding to approximately 6.1-fold improvement over chance. Performance remained consistent in the agitation-present subgroup (PRAUC 0.273). In the sleep-sensor sub-cohort, adding sleep features did not improve predictive performance (PRAUC 0.225 vs 0.244). Only REM ratio was independently associated with next-day agitation (OR 1.50, 95% CI 1.20–1.87; P<.001).

Conclusions

Passive in-home activity sensing enables next-day agitation prediction using low-burden behavioral digital biomarkers. Activity-based features, particularly overall activity level, evening activity patterns, and spatial behavioral stability, provide consistent predictive signals, whereas sleep architecture features offer limited incremental value in this setting. These findings support passive sensing as a scalable approach for early identification of agitation risk in dementia care.

Background

Dementia is a progressive disorder that not only impairs cognitive function but also substantially deteriorates daily functioning and overall quality of life [1]. Clinically, behavioral and psychological symptoms of dementia (BPSD) are highly prevalent [2]. A substantial proportion of people with dementia experience diverse BPSD manifestations, including sleep disturbance, agitation, and apathy [3]. These symptoms increase patient distress while also amplifying caregiver stress and anxiety and increasing healthcare utilization [4]. In particular, agitation has a considerable clinical impact by worsening quality of life for both patients and caregivers and by increasing emergency department visits, hospitalizations, and medication use [5]. Accordingly, there is a growing need for prediction and monitoring frameworks that enable early identification of BPSD and proactive intervention during high-risk periods [1].

The burden of BPSD and access to remote monitoring technologies are not uniformly distributed; older adults from socioeconomically disadvantaged backgrounds, ethnic minority groups, and rural communities face well-documented disparities in dementia diagnosis, caregiver support, and access to digital health infrastructure. Predictive monitoring systems must therefore be evaluated for performance and feasibility across diverse sociodemographic groups to avoid widening existing care gaps.

Previous studies have explored factors associated with BPSD, including sleep disturbances, social isolation, and instability of daily-life rhythms [6–8]. However, several limitations hinder translation into real-world clinical practice. First, many studies rely on questionnaires or rating scales assessed at specific time points or on relatively low-resolution observational indicators, making it difficult to capture the timing of symptom onset and short-term fluctuations with sufficient precision [3]. Second, although wearable and smartphone sensing has enabled continuous digital biomarker collection, studies that rely primarily on participant-managed devices may face practical barriers—such as variable adherence, missingness, and selection bias —which can limit long-term feasibility and real-world scalability in dementia populations [9, 10]. Third, much of the literature remains framed around contemporaneous episode monitoring or retrospective characterization, often without explicit short-horizon forecasting designs or operationally usable alert criteria for real-world deployment. [11].

In this context, in-home sensing technologies—particularly smart-home/IoT(Internet-of-Things)–based monitoring—have attracted attention as a practical approach for BPSD management in real-world settings [12, 13]. Ambient in-home IoT sensors can passively and non-invasively capture longitudinal

signals related to daily activity and sleep without requiring active user input, making this approach suitable for populations in which self-reporting and device management are challenging. While sleep and nocturnal physiological signals have been proposed as complementary prodromal markers for agitation, their systematic evaluation has been constrained by incomplete sensor data coverage in existing studies [6, 7]. Ambient IoT-based monitoring is therefore well positioned to complement wearable approaches, particularly in populations such as people with dementia where adherence to wearable devices is structurally challenging.

The Technology Integrated Health Management (TIHM) for Dementia dataset is a representative resource developed for “real-world remote monitoring” and includes in-home passive activity sensors (e.g., PIR), door sensors, and an under-mattress sleep mat (sleep staging and nocturnal physiological signals) [14]. Agitation events — along with five other behavioral alert types — are verified through a structured clinical workflow in which a monitoring team confirms events with participants and caregivers, yielding clinically validated binary labels that enable direct linkage between sensor signals and meaningful clinical outcomes.

Nevertheless, TIHM-based research on agitation prediction remains limited. Prior studies have been constrained by one or more of the following: a monitoring and anomaly-detection framing without explicit short-horizon prediction targets [15], substantial missingness in key sensor modalities limiting systematic evaluation of multimodal feature integration [16], or insufficient cohort size for reliable cross-patient validation [17]. Collectively, these gaps highlight the need to build on TIHM's core strengths — real-world longitudinal sensor data with clinically verified event labels — while advancing toward deployment-aligned short-horizon prediction with validation designs that reflect cross-patient generalizability.

In the present study, we used the TIHM dataset to predict “next-day agitation occurrence” based on passive behavioral digital biomarkers derived from PIR motion sensors, including Markov-chain-based movement pattern features, and conducted validation using Leave-One-Patient-Out Cross-Validation (LOPO-CV) to assess cross-patient generalizability. This study aims to evaluate the feasibility of predicting next-day agitation from passive in-home sensing and to identify behavioral features capable of supporting deployment-oriented early warning systems in real-world dementia care. In the subset of participants with concurrent sleep sensor data (n = 17), we additionally examined whether sleep architecture features provide incremental predictive value beyond PIR-based behavioral monitoring.

Methods

Study Design and Dataset

We conducted a secondary analysis of the Technology Integrated Health Management (TIHM) dataset, a prospective, naturalistic, remote monitoring study of people living with Alzheimer's disease and related dementias in the United Kingdom [14]. Participants were monitored continuously in their own homes

using a network of passive infrared (PIR) motion detectors, physiological sensors, and wearable actigraphy. The study received ethical approval from the NHS Health Research Authority (IRAS: 224526), and all participants provided written informed consent. Participant data were collected during the second phase (feasibility study) of the TIHM project (approximately 2019–2021); full enrollment and follow-up details are reported in Palermo et al. [14].

The analytical dataset comprised 56 participants with dementia (mean monitoring duration 50.9 ± 24.2 days, range 3–91 days), yielding 2,610 patient-days. Agitation events were identified by trained nursing staff based on clinical assessment and caregiver report. We defined a binary outcome, *agitation-next-day*, indicating whether an agitation event occurred on the calendar day following the feature observation day ($t + 1$). This predictive framing tests whether today's passive behavioral signals indicate impending agitation, consistent with an early warning use case. A total of 114 agitation events were recorded across 27 of 56 participants (48.2%); 29 participants (51.8%) experienced no agitation. Patient-day prevalence was 4.4% (full cohort) and 7.4% (agitation-present sub-cohort, $n = 27$).

By design, the TIHM clinical monitoring team responsible for verifying agitation events was not exposed to the sensor-derived behavioral features used in the present analysis; all features were computed post hoc from raw sensor logs after outcome labels had been recorded and clinically verified, providing structural blinding of outcome assessment to predictor values.

Data on concomitant pharmacological treatments (e.g., antipsychotics, benzodiazepines, cholinesterase inhibitors, memantine, and PRN sedatives), non-pharmacological interventions, and acute intercurrent medical events (e.g., urinary tract infection, constipation, pain) were not systematically available in the publicly released TIHM dataset and were therefore not included as covariates in the present analysis. These unmeasured factors are well-recognized triggers of agitation in dementia and are discussed further in the Limitations section.

Ethics Approval

The original TIHM study received ethical approval from the NHS Health Research Authority (IRAS: 224526), and all participants provided written informed consent. The present secondary analysis of deidentified publicly available data was reviewed and granted exemption by the Institutional Review Board of Wonkwang University Hospital (IRB No. WKUH 2026-03-052).

Behavioral Feature Engineering

Six daily behavioral features were specified a priori based on published behavioral phenotypes of dementia-related agitation and on prior work in unobtrusive in-home behavioral sensing. The features fall into two complementary families: aggregate summaries of daily PIR activity, and Markov chain-based features that characterize the sequential dynamics of room-to-room movement.

Daily Aggregate PIR Features

Activity volume (total PIR events per day) indexed overall daily motor activity. Excessive motor activity is recognized as a core behavioral component of agitation in dementia and is one of the three required domains in the International Psychogeriatric Association (IPA) consensus clinical and research definition of agitation in cognitive disorders [18, 19].

Location entropy (normalized Shannon entropy of room-use distribution between 07:00 and 22:00) characterized how broadly indoor movement was distributed across rooms. Daily activity patterns derived from in-home sensors have been used as digital markers of cognitive and behavioral change in older adults [20].

Low-stimulation room ratio (proportion of PIR events recorded in bedroom or bathroom) reflected time spent in low-arousal indoor environments. Environmental stimulation has long been proposed to modulate behavioral expression in dementia, with deviations from preferred stimulation levels associated with behavioral disturbance [21].

Evening activity ratio (proportion of PIR events occurring after 18:00) captured behavioral patterns associated with sundowning, a temporally bound pattern of behavioral worsening in the late afternoon and evening that is well described in people with dementia [18, 22, 23].

Markov Chain-Based Transition Features

Location-to-location transitions were modeled as a first-order Markov chain, with a transition matrix computed for each participant-day from consecutive PIR event pairs within the same calendar day. Whereas location entropy summarizes the marginal distribution of room use across the day, the Markov-based features below capture the sequential dynamics of how rooms are traversed, providing complementary rather than overlapping information.

Self-loop ratio (SLR) is the proportion of transitions in which the destination room equals the origin room — that is, the probability that two consecutive PIR events occur in the same room. SLR thus indexes within-room temporal persistence, distinct from the breadth-of-room-use information captured by location entropy. Routinized and stereotyped behavioral patterns have been described as correlates of frontal-lobe dysfunction in Alzheimer's disease, including in relation to agitation [24, 25].

Personalized SLR deviation (SLR-dev) captures individual-level behavioral change by computing the departure of each day's SLR from that participant's personal baseline. To prevent temporal leakage, the personal baseline was computed exclusively from the strict past: for each observation day t, the baseline was the mean SLR of all 'clean' days strictly before day t for that participant. A minimum of five prior clean days was required; days with fewer than five prior clean observations were excluded. Clean days were defined as days on which no agitation event occurred and which were not the day immediately preceding an agitation event (pre-agitation buffer). This expanding-window approach

ensures no future information contaminates the predictor, satisfying the requirements of prospective deployment. Days before sufficient history had accumulated were excluded from Model C analyses (2,367 of 2,610 patient-days retained).

Statistical Analysis

Primary associations were estimated using Generalized Estimating Equations (GEE) with a Binomial family and exchangeable correlation structure (statsmodels v0.14). GEE yields population-average (marginal) effect estimates while accounting for within-subject correlation in unbalanced longitudinal data. All predictors were standardized (zero mean, unit standard deviation) prior to fitting; ORs therefore reflect the change in agitation odds per 1 standard deviation increase. Robust sandwich standard errors were used throughout.

Three pre-specified nested models were evaluated: **Model A** (PIR features only), **Model B** (PIR + SLR), and **Model C** (PIR + SLR-dev). The event-to-predictor ratio was 22.8 for Model A and 19.0 for Models B and C; both are above the commonly cited minimum of 10, though this threshold was derived for standard logistic regression and should be interpreted with caution in repeated-measures GEE. Statistical significance was set at $\alpha = 0.05$, two-tailed.

As a supplementary sensitivity analysis, all three models were re-estimated using a Bayesian Binomial Mixed Generalized Linear Model (GLMM; statsmodels BinomialBayesMixedGLM) with a random intercept per patient. GLMM yields subject-specific estimates and directly partitions within-person from between-person variance. Convergence of the Laplace approximation (MAP fitting) was verified for all models. Agreement between GEE (population-average) and GLMM (subject-specific) effect directions and significance was interpreted as evidence of consistency across marginal and subject-specific modeling frameworks.

Predictive Modeling and Performance Evaluation

To assess prognostic utility, we applied Leave-One-Patient-Out Cross-Validation (LOPO-CV) with logistic regression (L2 penalty, balanced class weights; scikit-learn v1.3). The inverse regularization strength was fixed at $C = 1.0$ (scikit-learn default). No hyperparameter optimization or grid search was performed because the modest sample size and event count favored a pre-specified model configuration to reduce overfitting risk. Balanced class weighting was used to address the low agitation prevalence (4.4–7.4%). LOPO-CV trains on all patients except one and evaluates on the held-out patient, providing a conservative estimate of cross-patient generalization. Features were standardized using training-fold statistics; missing values were imputed with training-fold medians.

Performance was evaluated with three complementary metrics. The area under the precision-recall curve (PRAUC) was designated the primary metric, as it is insensitive to the large class imbalance. The area under the ROC curve (AUROC) was reported for comparability with the literature. The Brier score (mean

squared error between predicted probabilities and binary outcomes) assessed overall probabilistic accuracy. The model output was the predicted probability of next-day agitation rather than a binary classification. No fixed classification threshold was predefined because the intended use case is risk stratification. In future deployment, threshold selection should be determined according to the clinical context and desired trade-off between sensitivity (recall) and alert burden (precision). All metrics were supplemented with patient-level cluster-bootstrap 95% confidence intervals (1000 iterations, percentile method), resampling whole patients to preserve the within-patient dependence inherent in repeated daily observations. Receiver Operating Characteristic (ROC) curves were plotted to visualize discrimination across models. As an illustration of clinical utility, alert burden — sensitivity, positive predictive value (PPV), number needed to screen (NNS), and projected alerts per patient-month — was computed for the primary model (Model B) at percentile-based risk thresholds (top 5%–30% of out-of-fold risk scores).

Sensitivity Analyses

All analyses were repeated in the 27 agitation-present participants (7.4% prevalence) to test whether findings replicate in the clinically relevant sub-cohort and are not driven by the 29 agitation-free participants.

A subset of 17 participants (30.4%) had concurrent sleep sensor data available, providing minute-level recordings of sleep state (AWAKE, LIGHT, DEEP, REM), heart rate, respiratory rate, and snoring. Five daily sleep features were derived: total sleep duration (minutes of LIGHT + DEEP + REM per night), sleep efficiency (sleep duration / total recording time), deep sleep ratio (DEEP minutes / total sleep), REM ratio (REM minutes / total sleep), and nighttime awakenings (minutes in AWAKE state). Features were standardized using training-fold statistics in LOPO-CV. An augmented model (Model B+Sleep: all four PIR features [activity volume, location entropy, low-stimulation room ratio, evening activity ratio], SLR, and all five sleep architecture features) was evaluated on this sub-cohort. Because Models A, B, and B+Sleep do not use SLR-dev, no minimum-history requirement was applied, and all participants with concurrent sleep recording were retained. After merging with PIR feature days and excluding each participant's first and last observation day and patient-days lacking complete sleep recording, the sleep analytic sample comprised 17 participants and 767 patient-days (31 agitation events; prevalence 4.0%). Of the 17 participants, 11 had no agitation events during the observation period; consistent with the primary analysis, all participants were retained in LOPO-CV and contributed out-of-fold predictions to the pooled performance estimates. A sensitivity analysis restricting the sub-cohort to the 15 participants with at least seven days of monitoring (Additional file 2) confirmed that results were not materially altered by the two short-observation participants.

Results

Participant Characteristics

The analytical dataset comprised 56 participants monitored for a mean of 50.9 ± 24.2 days (range 3–91 days), contributing 2,610 patient-days (Fig. 1, Table 1). A total of 114 agitation events were recorded across 27 participants (48.2%); 29 participants (51.8%) had no agitation events. Patient-day prevalence was 4.4% (full cohort) and 7.4% (agitation-present). Event-to-predictor ratios were 22.8 (Model A) and 19.0 (Models B and C).

Table 1
Participant and monitoring characteristics.

Characteristic	**All Patients (n = 56)**	**Agitation-Present (n = 27)**
Monitoring duration (days), mean ± SD	50.9 ± 24.2	55.7 ± 22.8
Range	3–91	10–91
Total patient-days analyzed	2,610	1,534
Agitation events, n	114	114
Agitation prevalence (patient-day)	4.4%	7.4%
Patients with ≥ 1 agitation event, n (%)	27 (48.2%)	27 (100%)
Patients with zero agitation events, n (%)	29 (51.8%)	—
Events per variable (EPV)		
Model A (5 features)	22.8	22.8
Models B & C (6 features)	19.0	19.0

SD, standard deviation; EPV, events per variable. Values are mean ± SD or n (%) unless stated.

GEE Associations with Next-Day Agitation

Table 2 presents GEE results. In Model A (PIR only), two features were independently associated with next-day agitation: activity volume (OR = 2.16, 95% CI: 1.44–3.24, P = .0002) and evening activity ratio (OR = 1.54, 95% CI: 1.11–2.15, P = .010). Location entropy and low-stimulation room ratio were not independently significant in any model.

In Model B (PIR + SLR), self-loop ratio was strongly negatively associated with next-day agitation (OR = 0.25, 95% CI: 0.18–0.33, P<.001). At the population level, days with higher movement stereotypy (more same-room re-entries) were associated with lower agitation risk the next day. The PIR feature associations were directionally stable and their ORs minimally changed relative to Model A, indicating absence of clinically meaningful multicollinearity.

In **Model C** (PIR + SLR-dev, leakage-corrected rolling baseline), personalized SLR deviation was significantly associated with next-day agitation in GEE (OR = 0.73, 95% CI: 0.54–0.98, *P*=.038): a day on

which the patient's SLR fell below their own habitual baseline predicted agitation the following day. This represents an individual-level routine disruption signal, independent of between-person differences in baseline movement patterns. However, this association was attenuated and did not reach significance in the GLMM supplementary analysis (OR = 0.77, *P*=.108), suggesting that Model C's signal may be partly attributable to residual between-patient heterogeneity. These findings are presented as exploratory and warrant replication in larger independent cohorts.

Figure 2. **Dumbbell plot comparing GEE (filled circles) and GLMM (open circles) odds ratios for next-day agitation across Models A, B, and C (per 1 SD increase). For each predictor, the two estimates are connected by a horizontal line; line length visualises the magnitude of difference between marginal (GEE, population-average) and subject-specific (GLMM) effect estimates. Faint horizontal bars show 95% confidence intervals for the GEE estimate. The vertical dashed line at OR = 1 denotes the null. Significance markers refer to GEE estimates. * P<.05; ** P<.01; *** P<.001.**

Table 2
GEE results for next-day agitation (n = 56 full cohort; n = 27 agitation-present). All predictors standardized; ORs per 1 SD increase.

Feature	Model A (PIR only)	Model B (PIR + SLR)	Model C (PIR + SLR-dev)	Sensitivity B (agit-present)
Activity volume	2.16***	2.17***	1.88***	1.98***
	1.44–3.24	1.64–2.88	1.31–2.70	1.52–2.59
Location entropy	0.85 (ns)	0.90 (ns)	0.79 (ns)	0.81 (ns)
	0.45–1.63	0.56–1.44	0.52–1.20	0.50–1.32
Low-stim room ratio	1.19 (ns)	1.33 (ns)	1.21 (ns)	1.16 (ns)
	0.75–1.88	0.81–2.18	0.75–1.95	0.77–1.76
Evening act. ratio	1.54*	1.53***	1.43**	1.53***
	1.11–2.15	1.22–1.90	1.11–1.84	1.24–1.88
Self-loop ratio (SLR)	—	0.25***	—	0.38***
		0.18–0.33		0.29–0.50
SLR-dev (personalized)	—	—	0.73*	—
			0.54–0.98	
			[GLMM: 0.77, *P*=.108]	

*OR, odds ratio (95% CI); SLR, self-loop ratio; SLR-dev, personalized SLR deviation (rolling past-only baseline). ***P<.001; **P<.01; *P<.05; ns, not significant. Model C bracket shows GLMM result for SLR-dev.*

Leave-One-Patient-Out Cross-Validation Performance

Table 3 summarizes LOPO-CV performance. In the full cohort (chance PRAUC = 0.044), Model B achieved the highest discrimination and the lowest Brier score: AUROC = 0.848 (95% CI: 0.776–0.901), PRAUC = 0.263 (6.1× chance; 95% CI: 0.090–0.395), and Brier score = 0.161. Model A reached AUROC = 0.762 and PRAUC = 0.207 (4.8× chance; Brier = 0.195). Following leakage correction of the SLR-dev baseline, Model C demonstrated substantially reduced discrimination (AUROC = 0.716, PRAUC = 0.090, 2.1× chance; Brier = 0.216). Model B was therefore retained as the primary predictive model in subsequent analyses. The wider patient-level confidence intervals reflected uncertainty due to the modest number of participants and agitation events, but Model B retained the highest point estimates across cohort definitions.

In the agitation-present sub-cohort (chance PRAUC = 0.074), Model B retained the highest performance (AUROC = 0.780, PRAUC = 0.273, 3.7× chance; Brier = 0.229), consistent with the full-cohort findings. Model B's Brier score was lower than Models A and C in both cohorts, indicating lower overall prediction error; predicted probabilities, however, were generated under class-balanced weighting and should be interpreted as risk scores rather than calibrated absolute risks. Figure 3 presents precision-recall curves; Fig. 4 presents ROC curves. Per-patient discrimination varied substantially across patients (Additional file 3, Figure A3-1), consistent with the small number of agitation events per patient.

Table 3
LOPO-CV performance (logistic regression with balanced class weights).

Model	AUROC	AUROC 95%CI	PRAUC	PRAUC 95%CI	PRAUC/Chance	Brier
All patients (n = 56; chance PRAUC = 0.044)						
Model A	0.762	0.660–0.843	0.207	0.051–0.387	4.8×	0.195
Model B	**0.848**	**0.776–0.901**	**0.263**	**0.090–0.395**	**6.1×**	**0.161**
Model C (leakage-corrected)	0.716	0.646–0.784	0.090	0.054–0.137	2.1×	0.216
Agitation-present (n = 27; chance PRAUC = 0.074)						
Model A	0.704	0.592–0.804	0.234	0.077–0.418	3.2×	0.237
Model B	**0.780**	**0.692–0.848**	**0.273**	**0.108–0.404**	**3.7×**	**0.229**
Model C (leakage-corrected)	0.649	0.581–0.726	0.114	0.079–0.165	1.5×	0.260

AUROC, area under the ROC curve; PRAUC, area under the precision-recall curve; CI, patient-level cluster-bootstrap 95% CI (1000 iterations, resampling whole patients); Brier, Brier score (lower = better overall

probabilistic accuracy). Model B highlighted in yellow as best-performing. Model C results reflect leakage-corrected rolling baseline.

Sensitivity Analyses

Restricting analyses to the 27 agitation-present participants produced consistent GEE findings: activity volume, evening activity ratio, and SLR (Model B) maintained significant associations in the same direction as the primary analysis (Table 2). SLR-dev remained directionally consistent in the sensitivity GEE (OR = 0.73, P=.038) and sensitivity GLMM (OR = 0.74, P=.057), though the latter remained marginally non-significant. Location entropy and low-stimulation room ratio were non-significant in all sensitivity analyses, confirming that the primary results are not attributable to the inclusion of agitation-free participants and that Model B findings are consistent across both cohort definitions.

Exploratory Subgroup Analysis: Sleep Sensor Features

Within the sleep-sensor sub-cohort (n = 17 patients, 31 agitation events, chance PRAUC = 0.040), PIR-based models retained substantial discriminative performance (Table 4). Model A achieved AUROC = 0.770 and PRAUC = 0.124 (3.1× chance); Model B achieved AUROC = 0.881 and PRAUC = 0.244 (6.0× chance). Critically, adding five sleep architecture features in Model B+Sleep did not improve predictive performance (AUROC = 0.858, PRAUC = 0.225, 5.6× chance), suggesting that PIR behavioral features capture predictive signal that is not substantively supplemented by sleep architecture metrics at the single-day resolution available in this dataset.

In the GEE analysis of the sleep sub-cohort (Model B+Sleep), the three primary PIR features retained significant independent associations with next-day agitation after adjusting for all sleep variables: activity volume (OR = 2.83, 95% CI: 2.02–3.98, P<.001), evening activity ratio (OR = 1.88, 95% CI: 1.59–2.23, P<.001), and self-loop ratio (OR = 0.11, 95% CI: 0.03–0.41, P=.001). Among sleep features, only REM ratio was independently associated with next-day agitation (OR = 1.50, 95% CI: 1.20–1.87, P<.001); total sleep duration, sleep efficiency, deep sleep ratio, and nighttime awakenings were non-significant. Location entropy reached nominal significance in this small sleep sub-cohort (OR = 0.47, 95% CI: 0.24–0.92, P=.027) in a direction not observed in the full cohort; given the limited events (n = 31) and the instability of secondary estimates at this sample size, this isolated finding was not interpreted further. These findings should be interpreted cautiously given the small sample size (31 events) and the exploratory nature of this sub-cohort analysis.

Table 4
LOPO-CV performance in sleep-sensor sub-cohort (n = 17 patients, 767 patient-days, 31 agitation events; chance PRAUC = 0.040).

Model	AUROC	PRAUC	PRAUC/Chance	Brier	Sleep features
Model A (PIR only)	0.770	0.124	3.1×	0.194	No
Model B (PIR + SLR)	**0.881**	**0.244**	**6.0×**	**0.151**	**No**
Model B+Sleep (PIR + SLR + Sleep)	0.858	0.225	5.6×	0.155	Yes

AUROC, area under the ROC curve; PRAUC, area under the precision-recall curve; Brier, Brier score (lower = better). Model B highlighted in yellow. Sleep features: total sleep duration, sleep efficiency, deep sleep ratio, REM ratio, nighttime awakenings.

Table 5
GEE results for next-day agitation in sleep-sensor sub-cohort (Model B+Sleep: all four PIR features (activity volume, location entropy, low-stimulation room ratio, evening activity ratio), SLR, and five sleep architecture features; n = 17, 767 patient-days, 31 agitation events). All participants with concurrent sleep recording were retained because Model B+Sleep does not use SLR-dev. All predictors standardized; ORs per 1 SD increase.

Feature	OR (95% CI)	z	P value
— PIR Features —			
Activity volume	**2.83 (2.02–3.98)**	6.03	**< .001**
Location entropy	0.47 (0.24–0.92)	−2.21	.027
Low-stim room ratio	1.51 (0.79–2.89)	1.26	.209
Evening activity ratio	**1.88 (1.59–2.23)**	7.24	**< .001**
Self-loop ratio (SLR)	**0.11 (0.03–0.41)**	−3.28	**.001**
— Sleep Features —			
Total sleep duration	0.73 (0.47–1.16)	−1.34	.182
Sleep efficiency	1.02 (0.52–2.00)	0.06	.950
Deep sleep ratio	1.29 (0.51–3.23)	0.54	.593
REM ratio*	**1.50 (1.20–1.87)**	3.56	**< .001**
Nighttime awakenings	1.08 (0.76–1.55)	0.43	.667

*OR, odds ratio (95% CI); GEE with Binomial family, exchangeable correlation, robust sandwich SE. *REM ratio highlighted (only significant sleep feature).*

Clinical Utility: Alert Burden

To illustrate operating characteristics relevant to deployment, percentile-based alert thresholds were applied to the Model B out-of-fold risk scores in the full cohort (Table 6). At a threshold flagging the top 20% of daily risk scores, Model B detected 69% of next-day agitation events (sensitivity 0.690) at a PPV of 0.149, corresponding to approximately six alerts per patient per month (NNS 6.7). A more conservative top-5% threshold reduced the alert burden to approximately 1.5 alerts per patient-month (sensitivity 0.336, PPV 0.290). Because predicted probabilities were generated under class-balanced weighting, these percentile-based thresholds are reported as risk-ranking operating points rather than calibrated absolute-risk cut-offs; in deployment, threshold selection should reflect the desired sensitivity–alert-burden trade-off in the local clinical context.

Table 6
Alert burden for Model B (PIR + SLR) at percentile-based risk thresholds (full cohort; 2,610 patient-days, 114 events). Thresholds flag the top K% of out-of-fold daily risk scores.

Risk threshold	Sensitivity	PPV	NNS	Alerts/patient-month
Top 5%	0.336	0.290	3.4	1.5
Top 10%	0.469	0.203	4.9	3.0
Top 20%	0.690	0.149	6.7	6.0
Top 30%	0.805	0.116	8.6	9.0

Discussion

Principal Results

This study demonstrates that passive in-home activity monitoring can support clinically promising prediction of next-day agitation in dementia. Across association analyses and cross-patient validation, activity volume and evening activity ratio were consistently associated with increased agitation risk, whereas self-loop ratio (SLR) showed a strong inverse association. In predictive modeling, the combination of PIR-based behavioral features and SLR achieved the highest point estimates among candidate models, with PRAUC values above baseline prevalence in both the full cohort and the agitation-present subgroup. Patient-level confidence intervals indicated uncertainty consistent with the modest number of participants and events. In contrast, personalized SLR deviation (SLR-dev) showed weaker and less consistent associations after accounting for temporal leakage, and sleep architecture features did not add predictive value beyond activity-based measures. These findings suggest that passive activity-based behavioral features may provide useful short-horizon signals for agitation risk in dementia.

Comparison with Prior Work

Our findings indicate that daytime behavioral activity patterns may carry stronger short-horizon predictive signals for next-day agitation than sleep architecture features. This interpretation is consistent with prior literature showing that agitation and other BPSD are closely linked to disruption of rest–activity rhythms, circadian dysregulation, and late-day behavioral worsening in dementia [26–28]. Increased activity volume may capture early behavioral manifestations of agitation, such as restlessness, wandering, pacing, or purposeless motor activity, which are core behavioral expressions frequently observed in agitated dementia patients [25]. The relatively large effect size observed for activity volume suggests that increased overall activity may reflect clinically meaningful escalation of prodromal agitation-related behaviors rather than a nonspecific movement signal. Evening activity may further reflect circadian misalignment or sundowning-related behavioral escalation, a phenomenon characterized by increased agitation, confusion, restlessness, and wandering in the late afternoon or evening [25, 26, 28]. Thus, PIR-derived activity features may serve as low-burden proxies of diurnal behavioral dysregulation, integrating both overall activity escalation and late-day behavioral instability.

In contrast, sleep architecture features showed limited incremental predictive value in this dataset. Sleep disruption is common in dementia and has been associated with cognitive impairment, circadian rhythm disturbance, caregiver burden, and neuropsychiatric symptoms [3, 6, 29]. In this dataset, however, our results suggest that sleep architecture measures may have been less proximal, less consistently captured, or less informative for next-day agitation risk than daytime activity and spatial behavioral patterns in this cohort. Within the sleep sub-cohort, REM ratio was the only sleep architecture feature independently associated with next-day agitation. Consumer-grade and non-polysomnographic sleep-stage estimation remains imperfect, particularly for REM detection, which may limit the precision of REM-related findings in this study [30]. However, given the small number of events (n = 31) and the exploratory nature of this analysis, this finding should be interpreted cautiously and considered hypothesis-generating. Replication in larger cohorts with more stable and direct sleep-stage measurements is needed [31].

The inverse association between SLR and next-day agitation can be interpreted in light of prior literature emphasizing the role of routine, environmental familiarity, and contextual stability in dementia-related behavioral symptoms [32]. Non-pharmacological intervention studies and reviews have shown that person-centered care, individualized activities, and environmental approaches can reduce agitation and other neuropsychiatric symptoms in dementia [33]. Environmental design and home-based sensing studies further suggest that spatial context, room-level location, and environmental triggers may influence agitation risk [11, 15]. From this perspective, SLR should not be viewed as an established agitation biomarker, but rather as a sensor-derived operational measure of spatial behavioral regularity. A higher SLR may reflect a more stable and familiar pattern of room use, whereas lower SLR may indicate disruption of habitual spatial routines or increased room-to-room transitions. This interpretation is hypothesis-generating and should be validated in future studies incorporating direct behavioral annotation, caregiver-reported routines, and environmental context data.

The association of SLR-dev differed between modeling frameworks, reaching statistical significance in GEE but not in GLMM. This divergence likely reflects differences in estimands. GEE provides population-average estimates incorporating both within- and between-patient variation, whereas GLMM more explicitly accounts for patient-level baseline differences through random intercepts [34, 35]. The attenuation of SLR-dev in GLMM suggests that deviations from habitual spatial patterns may be less stable as within-patient predictors than as population-level markers. These findings highlight the importance of distinguishing between population-level associations and patient-specific dynamics when interpreting behavioral digital biomarkers. In this context, stable spatial patterning captured by SLR appeared more stable than personalized short-term deviation from baseline SLR.

An important implication of this study is that clinically promising risk stratification can be achieved using simple PIR-based activity monitoring, without requiring more complex physiological sensing systems. In clinical practice, BPSD assessment often relies on caregiver reports and intermittent clinical evaluations, which are subjective and limited in temporal resolution [3]. Passive sensing can complement these approaches by continuously and objectively capturing behavioral patterns without requiring active patient participation, an important advantage in dementia populations where cognitive impairment and reduced insight may limit reliable self-report [12, 36]. The ability to predict agitation on a next-day timescale may support proactive care strategies, including anticipatory caregiver support, environmental modification, and timely clinical intervention. Collaborative dementia care models incorporating remote monitoring and structured caregiver support have demonstrated improvements in patient quality of life, caregiver well-being, and health care utilization [37], and integration of predictive behavioral signals into such frameworks may further enable proactive, individualized BPSD management.

Compared with prior TIHM-based and sensor-based dementia studies that primarily focused on continuous monitoring, anomaly detection, contemporaneous symptom detection, or smaller multimodal cohorts, our approach extends this literature by focusing on prospective next-day risk stratification with cross-patient validation [11, 16, 17]. The use of LOPO-CV across the full cohort provides a conservative and deployment-relevant estimate of generalizability. In addition, the finding that sleep features did not improve prediction beyond PIR-derived activity and spatial regularity measures suggests that low-burden activity monitoring may be sufficient for scalable agitation risk prediction in real-world dementia care settings. Taken together, these findings support the potential role of passive in-home activity monitoring as a practical digital biomarker framework for short-horizon agitation risk prediction in dementia [38, 39].

Implementation Considerations

Several practical considerations follow from these findings. First, in real-world deployment, prolonged gaps in PIR sensor data due to sensor failure, network outage, or participant relocation would compromise daily feature extraction. Downstream applications should therefore implement a minimum-data-completeness threshold (for example, requiring sufficient valid PIR coverage and a minimum number of events per day) below which the system withholds a prediction and emits a data-quality flag rather than generating a potentially misleading low-confidence forecast.

Second, the intended deployment is passive from the patient's perspective: no active interaction or self-report is required. Daily predictions would be delivered through an alert dashboard accessible to community mental-health teams, dementia care coordinators, or designated family caregivers. The required technical infrastructure is modest—a small network of PIR sensors per home and a centralized data aggregation pipeline—comparable to existing telecare deployments. End users would not require technical expertise beyond routine dashboard navigation, and clinical interpretation of alerts and decisions about intervention would remain with the responsible clinician or caregiver.

Strengths

This study has several methodological strengths. First, to our knowledge, this is the first TIHM-based agitation study to explicitly frame agitation as a prospective next-day prediction task and to evaluate performance using Leave-One-Patient-Out Cross-Validation (LOPO-CV). Unlike contemporaneous monitoring or anomaly-detection approaches, LOPO-CV provides a deployment-oriented estimate of cross-patient generalizability by testing prediction in previously unseen individuals. Second, we evaluated associations using two complementary longitudinal modeling frameworks, Generalized Estimating Equations (GEE) and Generalized Linear Mixed Models (GLMM). The consistency of findings across population-average and subject-specific estimands strengthens confidence in the consistency of the identified activity-based behavioral markers. Third, we explicitly addressed temporal leakage in personalized behavioral baselines by implementing a leakage-corrected expanding-window approach for SLR-dev computation. By restricting baseline estimation to strictly prior observations and excluding pre-agitation periods, the analysis more closely reflects prospective real-world deployment conditions. Fourth, this study systematically evaluated multimodal sensing integration by jointly examining PIR-derived activity features and sleep architecture measures. Although sleep variables provided limited incremental predictive value, this comparison clarified the relative contribution of low-burden activity sensing versus more complex physiological monitoring systems and supports scalable implementation strategies for dementia care.

Limitations

Despite these methodological strengths, several limitations should be acknowledged. First, the sample size was relatively modest, particularly with respect to the number of agitation events. Although repeated-measures modeling and LOPO-CV were applied to improve statistical efficiency and generalizability, external validation in larger cohorts is necessary to confirm these findings [40, 41]. Second, the study cohort included individuals with heterogeneous dementia etiologies, and subtype-specific differences were not explicitly analyzed. Given that the clinical presentation and neurobiological mechanisms of BPSD may differ across dementia subtypes, this represents an area for future investigation [3, 42]. Nevertheless, the consistency of findings across multiple analytical frameworks (GEE, GLMM, and LOPO-CV) and sensitivity analyses supports the consistency of activity-based behavioral markers as predictors of agitation risk. Third, sleep sensor data were available in only a subset of participants (17 of 56, 30%; 767 patient-days, 31 agitation events), resulting in limited

statistical power for the sleep-related analyses. As such, the lack of incremental predictive value of sleep features should be interpreted with caution, and definitive conclusions regarding their role in next-day agitation prediction cannot be established from the present data [43]. Fourth, agitation was modeled as a binary outcome, without accounting for severity, duration, or temporal dynamics of individual episodes [3, 18]. These dimensions may carry important clinical information and could influence both prediction performance and clinical applicability. Future studies incorporating more granular outcome measures may provide a more complete understanding of agitation trajectories. Additionally, fairness analyses across sociodemographic subgroups could not be performed because the publicly available TIHM dataset provides limited demographic and socioeconomic information. Variables such as ethnicity, educational level, income, and digital access were not systematically available. Consequently, potential differences in model performance across vulnerable populations could not be evaluated and should be addressed in future validation studies. Finally, given the observational design of this study, causal relationships between sensor-derived behavioral features and agitation cannot be established [44]. The observed associations may reflect shared underlying neurobehavioral processes or common upstream factors, rather than direct causal effects. Prospective interventional studies are needed to clarify the causal role of behavioral digital biomarkers in agitation development [45]. In addition, the patient-level cluster-bootstrap confidence intervals were wide, indicating that the precision of the performance estimates is constrained by the number of patients and agitation events.

Conclusions

In conclusion, these findings highlight the clinical relevance of activity-based digital biomarkers for short-horizon agitation prediction and support their potential role as a practical and scalable approach for proactive management of BPSD in real-world dementia care. The strong association of SLR suggests that stability of spatial behavioral patterns, reflecting consistent daily routines and environmental familiarity, may be associated with reduced agitation risk. These findings raise the hypothesis that maintaining structured daily routines and spatial familiarity could be relevant intervention targets, but this hypothesis requires prospective validation. While these results are encouraging, they are derived from a single cohort and require validation in larger and more diverse populations. Future studies should focus on external validation and the integration of predictive models into real-time monitoring and alert systems, enabling timely intervention and translation into routine clinical practice.

Abbreviations

AUROC
area under the receiver operating characteristic curve
BPSD
behavioral and psychological symptoms of dementia
CI
confidence interval
EPV

events per variable
GEE
generalized estimating equations
GLMM
generalized linear mixed model
IPA
International Psychogeriatric Association
LOPO-CV
leave-one-patient-out cross-validation
OR
odds ratio
PIR
passive infrared
PRAUC
area under the precision-recall curve
REM
rapid eye movement
SLR
self-loop ratio
SLR-dev
personalized self-loop ratio deviation
TIHM
Technology Integrated Health Management

Declarations

Ethics approval and consent to participate

The original TIHM study received ethical approval from the NHS Health Research Authority (IRAS: 224526), and all participants provided written informed consent. The present secondary analysis of deidentified publicly available data was reviewed and granted exemption by the Institutional Review Board of Wonkwang University Hospital (IRB No. WKUH 2026-03-052).

Consent for publication

Not applicable. The TIHM dataset is publicly available and fully deidentified.

Availability of data and materials

The TIHM dataset analyzed in this study is publicly available at Zenodo (DOI: 10.5281/zenodo.7622128, Palermo et al., 2023). The analytical code (feature engineering, GEE/GLMM models, and LOPO-CV pipeline) is available from the corresponding author upon reasonable request.

Competing interests

None

Funding

None

Authors' contributions

SHY: Conceptualization, Methodology, Supervision, Writing — review & editing. YJJ: Methodology, Formal analysis, Investigation, Data curation, Visualization, Writing — original draft. SYL: Resources, Writing — review & editing. CMY: Validation, Writing — review & editing. DJK: Investigation, Writing — review & editing. YDJ: Investigation, Writing — review & editing. All authors read and approved the final manuscript.

Acknowledgements

This study used the publicly available Technology Integrated Health Management (TIHM) dataset. The authors acknowledge the Surrey and Borders Partnership NHS Foundation Trust and Howz for supporting the development and public release of the TIHM dataset and enabling its use for research purposes.

Reporting guideline

This manuscript was prepared in accordance with the TRIPOD+AI reporting guideline (Collins et al., BMJ 2024), and the completed checklist is provided as Additional file 1.

Protocol and registration

This study was a secondary analysis of a publicly available dataset; a prospective study protocol was not prepared. The analysis plan, including pre-specified behavioral features, statistical models, and performance metrics, was specified prior to model fitting and is fully described in the Methods section.

This secondary analysis was not prospectively registered with a public study registry.

Patient and public involvement

There was no direct patient or public involvement in the design, conduct, or interpretation of this secondary analysis. The original TIHM study (Palermo et al. [14]) involved people living with dementia and their caregivers as participants; readers are referred to that publication for details of patient and public engagement in the underlying cohort.

References

1. Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. lancet. 2020;396(10248):413–46.
2. Kwon C-Y, Lee B. Prevalence of behavioral and psychological symptoms of dementia in community-dwelling dementia patients: a systematic review. Front Psychiatry. 2021;12:741059.
3. Kales HC, Gitlin LN, Lyketsos CG. Assessment and management of behavioral and psychological symptoms of dementia. BMJ. 2015;350:h369.
4. Afonso-Argilés FJ, Meyer G, Stephan A, Comas M, Wübker A, Leino-Kilpi H, et al. Emergency department and hospital admissions among people with dementia living at home or in nursing homes: results of the European RightTimePlaceCare project on their frequency, associated factors and costs. BMC Geriatr. 2020;20(1):453.
5. McDermott CL, Gruenewald DA. Pharmacologic management of agitation in patients with dementia. Curr Geriatr Rep. 2019;8(1):1–11.
6. Kim S-S, Oh KM, Richards K. Sleep disturbance, nocturnal agitation behaviors, and medical comorbidity in older adults with dementia: relationship to reported caregiver burden. Res Gerontol Nurs. 2014;7(5):206–14.
7. Rose KM, Lorenz R. Sleep disturbances in dementia: What they are and what to do. J Gerontol Nurs. 2010;36(5):9–14.
8. Sun W, Matsuoka T, Oba H, Narumoto J. Importance of loneliness in behavioral and psychological symptoms of dementia. Int J Geriatr Psychiatry. 2021;36(4):540–6.
9. Peterson CM, St Louis RM, Flannagan C. Enhancing enrollment and adherence in long-term wearable research on dementia: qualitative systematic review and meta-synthesis. JMIR aging. 2025;8(1):e63768.
10. Ding H, Ho K, Searls E, Low S, Li Z, Rahman S, et al. Assessment of wearable device adherence for monitoring physical activity in older adults: pilot cohort study. JMIR aging. 2024;7:e60209.
11. Fletcher-Lloyd N, Soreq E, Wilson D, Nilforooshan R, Sharp DJ, Barnaghi P. Home monitoring of daily living activities and prediction of agitation risk in a cohort of people living with dementia. Alzheimers Dement. 2021;17(Suppl 12):e058614.
12. Tiersen F, Batey P, Harrison MJ, Naar L, Serban A-I, Daniels SJ, et al. Smart home sensing and monitoring in households with dementia: user-centered design approach. JMIR aging. 2021;4(3):e27047.
13. Lussier M, Lavoie M, Giroux S, Consel C, Guay M, Macoir J, et al. Early detection of mild cognitive impairment with in-home monitoring sensor technologies using functional measures: a systematic review. IEEE J biomedical health Inf. 2018;23(2):838–47.
14. Palermo F, Chen Y, Capstick A, Fletcher-Lloyd N, Walsh C, Kouchaki S, et al. Tihm: An open dataset for remote healthcare monitoring in dementia. Sci data. 2023;10(1):606. 10.1038/s41597-023-02519-y.
15. Enshaeifar S, Zoha A, Markides A, Skillman S, Acton ST, Elsaleh T, et al. Health management and pattern analysis of daily living activities of people with dementia using in-home sensors and

machine learning techniques. PLoS ONE. 2018;13(5):e0195605.

16. Abedi A, Chu CH, Khan SS. Benchmarking Early Agitation Prediction in Community-Dwelling People with Dementia Using Multimodal Sensors and Machine Learning. arXiv preprint arXiv:2506.06306. 2025.
17. Ramesh K, Yakoub A, Ghoneim Y, Al Korabi R, Ramesh J, Sagahyroon A, et al. Machine Learning Prediction of Agitation in Dementia Patients Using Sleep and Physiological Data. Appl Sci. 2025;15(18):9908.
18. Cohen-Mansfield J, Billig N. Agitated behaviors in the elderly: I. A conceptual review. J Am Geriatr Soc. 1986;34(10):711–21.
19. Cummings J, Mintzer J, Brodaty H, Sano M, Banerjee S, Devanand D, et al. Agitation in cognitive disorders: International Psychogeriatric Association provisional consensus clinical and research definition. Int Psychogeriatr. 2015;27(1):7–17.
20. Akl A, Taati B, Mihailidis A. Autonomous unobtrusive detection of mild cognitive impairment in older adults. IEEE Trans Biomed Eng. 2015;62(5):1383–94.
21. Hall GR, Buckwalter KC. Progressively lowered stress threshold: a conceptual model for care of adults with Alzheimer's disease. Arch Psychiatr Nurs. 1987;1(6):399–406.
22. Khachiyants N, Trinkle D, Son SJ, Kim KY. Sundown syndrome in persons with dementia: an update. Psychiatry Invest. 2011;8(4):275.
23. Canevelli M, Valletta M, Trebbastoni A, Sarli G, D'Antonio F, Tariciotti L, et al. Sundowning in dementia: clinical relevance, pathophysiological determinants, and therapeutic approaches. Front Med. 2016;3:73.
24. Tekin S, Cummings JL. Frontal–subcortical neuronal circuits and clinical neuropsychiatry: an update. J Psychosom Res. 2002;53(2):647–54.
25. Algase DL, Beattie ER, Antonakos C, Beel-Bates CA, Yao L. Wandering and the physical environment. Am J Alzheimer's Disease Other Dementias®. 2010;25(4):340–6.
26. Van Someren EJ, Hagebeuk EE, Lijzenga C, Scheltens P, De Rooij SE, Jonker C, et al. Circadian rest—activity rhythm disturbances in Alzheimer's disease. Biol Psychiatry. 1996;40(4):259–70.
27. Ancoli-Israel S, Cole R, Alessi C, Chambers M, Moorcroft W, Pollak CP. The role of actigraphy in the study of sleep and circadian rhythms. Sleep. 2003;26(3):342–92.
28. Musiek ES, Holtzman DM. Mechanisms linking circadian clocks, sleep, and neurodegeneration. Science. 2016;354(6315):1004–8.
29. Petit D, Gagnon J-F, Fantini ML, Ferini-Strambi L, Montplaisir J. Sleep and quantitative EEG in neurodegenerative disorders. J Psychosom Res. 2004;56(5):487–96.
30. De Zambotti M, Cellini N, Goldstone A, Colrain IM, Baker FC. Wearable sleep technology in clinical and research settings. Med Sci Sports Exerc. 2019;51(7):1538.
31. Camargos EF, Louzada FM, Nóbrega OT. Wrist actigraphy for measuring sleep in intervention studies with Alzheimer's disease patients: application, usefulness, and challenges. Sleep Med Rev.

2013;17(6):475–88.
32. Kales HC, Gitlin LN, Lyketsos CG. Management of neuropsychiatric symptoms of dementia in clinical settings: recommendations from a multidisciplinary expert panel. J Am Geriatr Soc. 2014;62(4):762–9.
33. Brodaty H, Arasaratnam C. Meta-analysis of nonpharmacological interventions for neuropsychiatric symptoms of dementia. Am J Psychiatry. 2012;169(9):946–53.
34. Hubbard AE, Ahern J, Fleischer NL, Van der Laan M, Lippman SA, Jewell N, et al. To GEE or not to GEE: comparing population average and mixed models for estimating the associations between neighborhood risk factors and health. Epidemiology. 2010;21(4):467–74.
35. Neuhaus JM, Kalbfleisch JD, Hauck WW. A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. Int Stat Review/Revue Int de Statistique. 1991:25–35.
36. Orfei MD, Blundo C, Celia E, Casini AR, Caltagirone C, Spalletta G, et al. Anosognosia in mild cognitive impairment and mild Alzheimer's disease: frequency and neuropsychological correlates. Am J Geriatric Psychiatry. 2010;18(12):1133–40.
37. Reuben DB, Tan ZS, Romero T, Wenger NS, Keeler E, Jennings LA. Patient and caregiver benefit from a comprehensive dementia care program: 1-year results from the UCLA Alzheimer's and Dementia Care Program. J Am Geriatr Soc. 2019;67(11):2267–73.
38. Kourtis LC, Regele OB, Wright JM, Jones GB. Digital biomarkers for Alzheimer’s disease: the mobile/wearable devices opportunity. NPJ Digit Med. 2019;2(1):9.
39. Piau A, Wild K, Mattek N, Kaye J. Current state of digital biomarker technologies for real-life, home-based monitoring of cognitive function for mild cognitive impairment to mild Alzheimer disease and implications for clinical care: systematic review. J Med Internet Res. 2019;21(8):e12785.
40. Riley RD, Hayden JA, Steyerberg EW, Moons KG, Abrams K, Kyzas PA, et al. Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. PLoS Med. 2013;10(2):e1001380.
41. Steyerberg EW. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. 2nd ed. Cham, Switzerland: Springer; 2019. p. 558.
42. Lyketsos CG, Carrillo MC, Ryan JM, Khachaturian AS, Trzepacz P, Amatniek J, et al. Neuropsychiatric symptoms in Alzheimer’s disease. Elsevier; 2011. pp. 532–9.
43. Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, et al. Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci. 2013;14(5):365–76.
44. Grimes DA, Schulz KF. Bias and causal associations in observational research. lancet. 2002;359(9302):248–52.
45. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ. 2015;350:g7594.

Figures

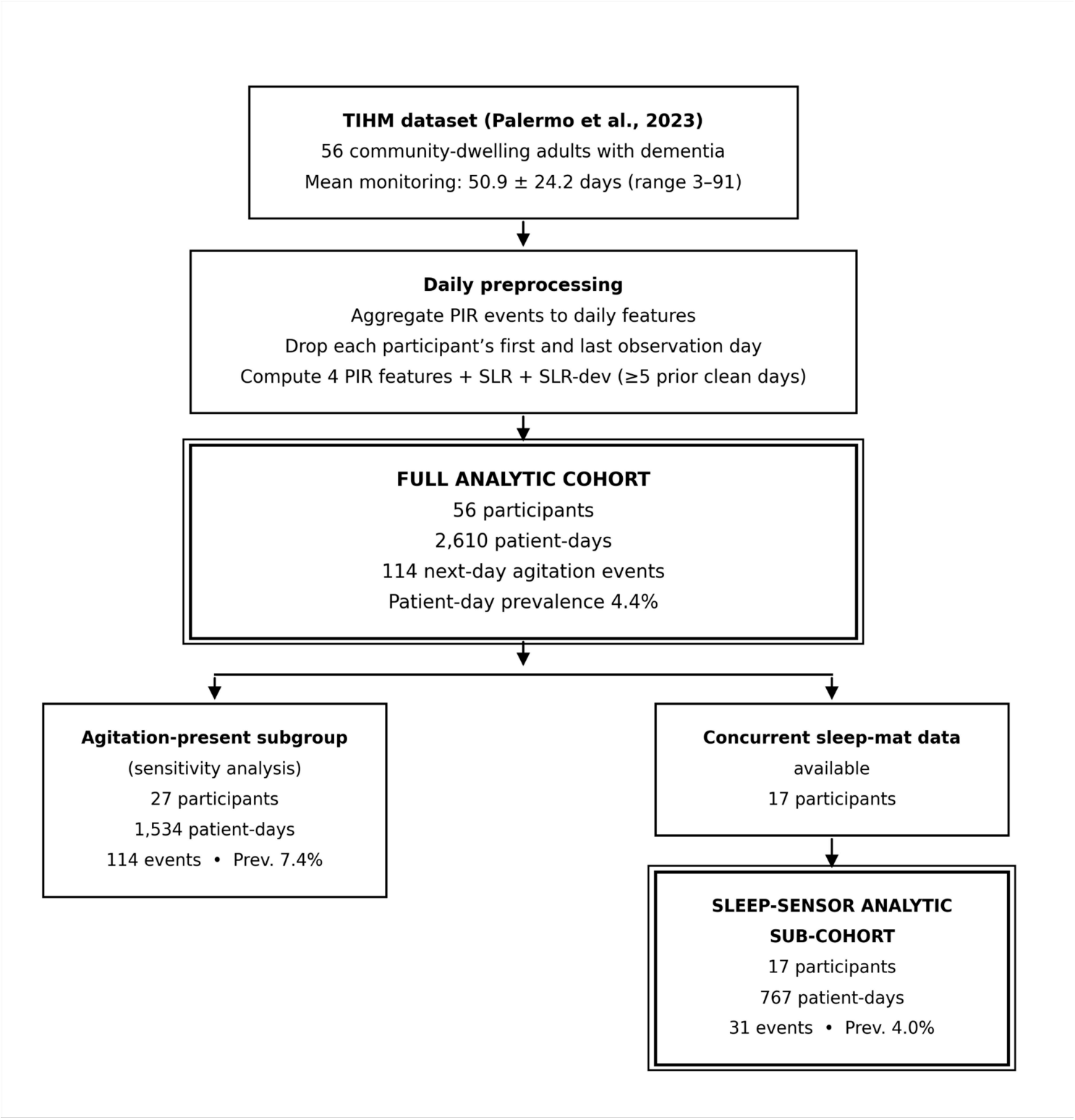


Figure 1

Analytic cohort derivation. The TIHM dataset (Palermo et al., 2023) comprised 56 community-dwelling adults with dementia. After daily aggregation of PIR sensor events and exclusion of each participant's first and last observation day, the full analytic cohort comprised 2,610 patient-days with 114 next-day agitation events. Sensitivity analyses examined the agitation-present subgroup (n = 27) and the

concurrent sleep-sensor sub-cohort (n = 17; 767 analytic patient-days, 31 events). All models were trained and evaluated using Leave-One-Patient-Out Cross-Validation.

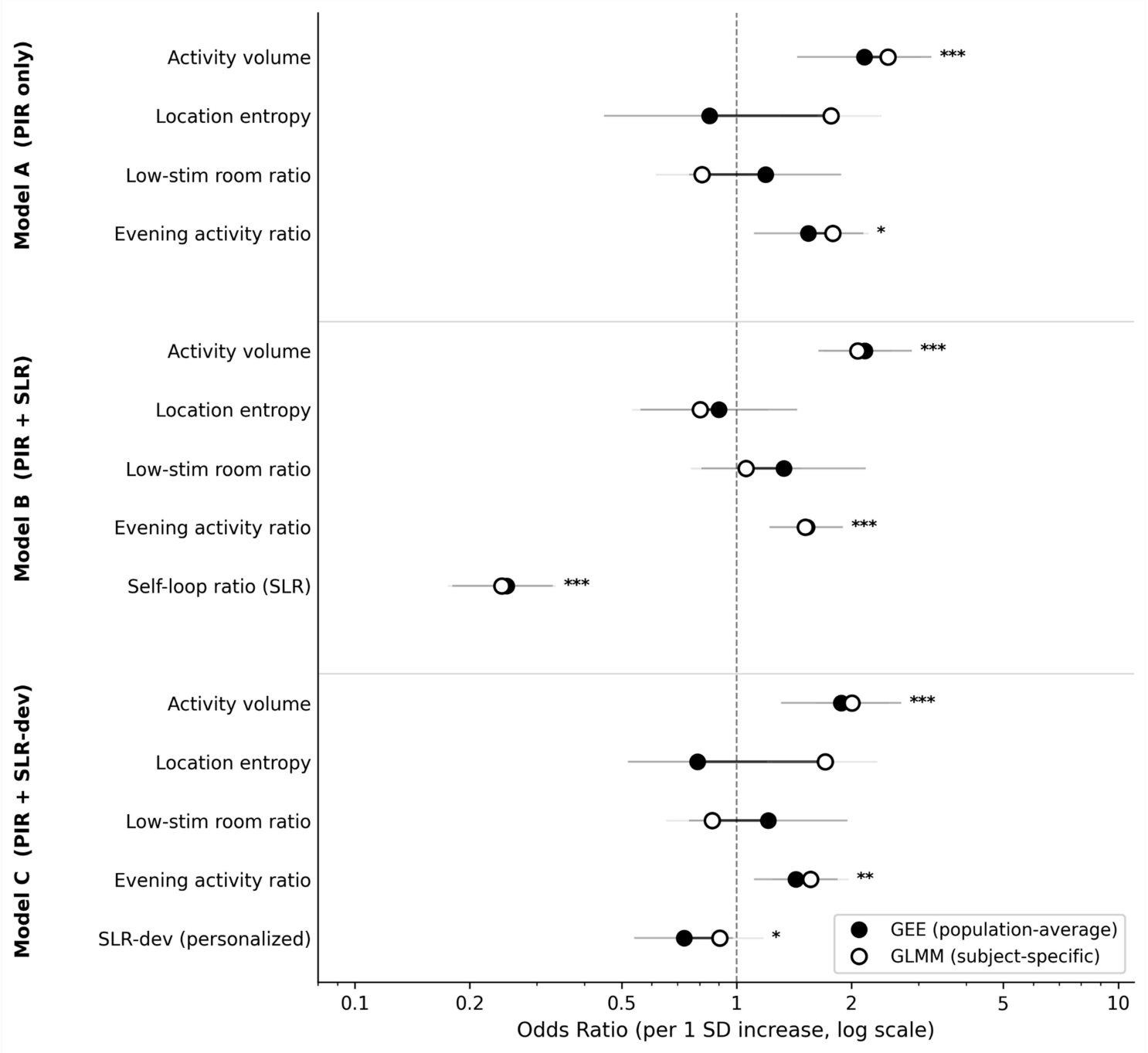


Figure 2

Dumbbell plot comparing GEE (filled circles) and GLMM (open circles) odds ratios for next-day agitation across Models A, B, and C (per 1 SD increase). For each predictor, the two estimates are connected by a horizontal line; line length visualises the magnitude of difference between marginal (GEE, population-average) and subject-specific (GLMM) effect estimates. Faint horizontal bars show 95% confidence intervals for the GEE estimate. The vertical dashed line at OR = 1 denotes the null. Significance markers refer to GEE estimates. * *P*<.05; ** *P*<.01; * *P*<.001.**

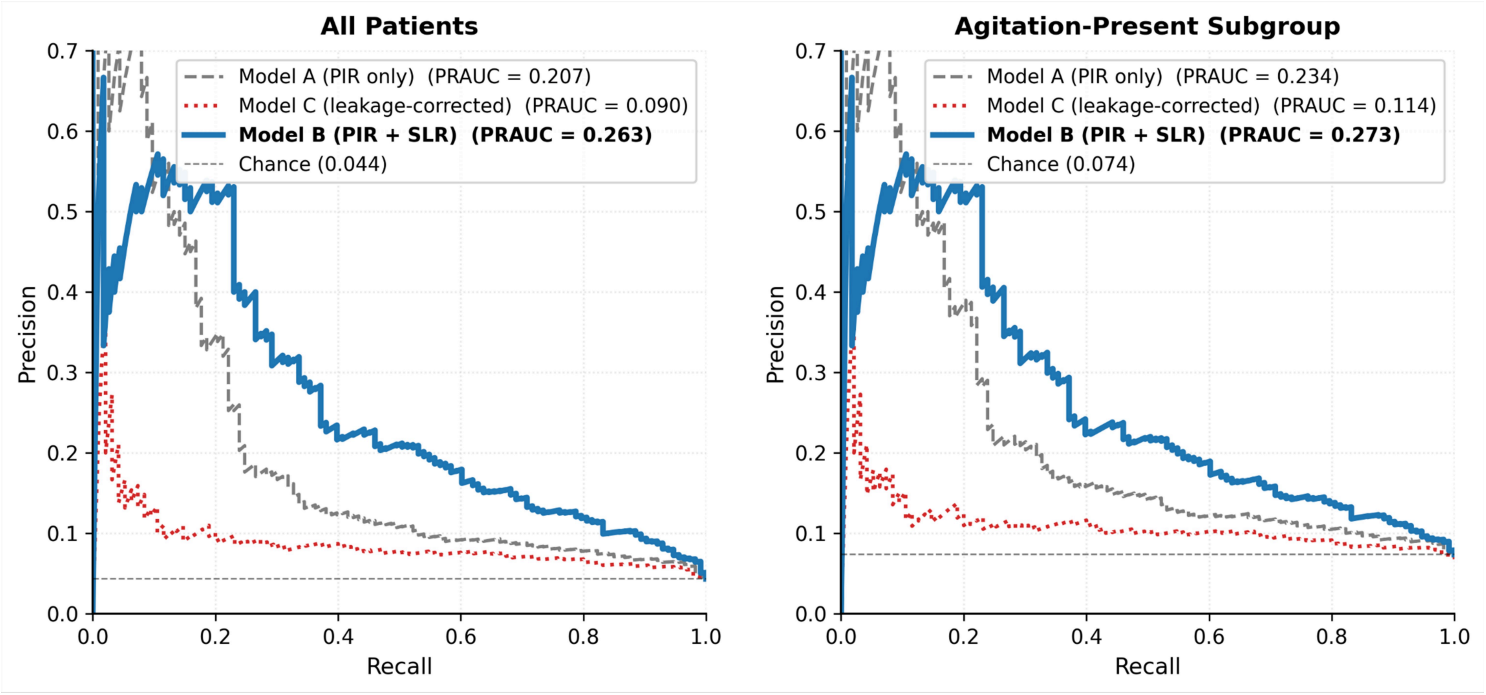


Figure 3

Precision-recall curves from LOPO-CV. Left: full cohort (n = 56; chance PRAUC = 0.044). Right: agitation-present sub-cohort (n = 27; chance PRAUC = 0.074). Dashed horizontal lines indicate chance-level PRAUC. Model C shown with leakage-corrected baseline.

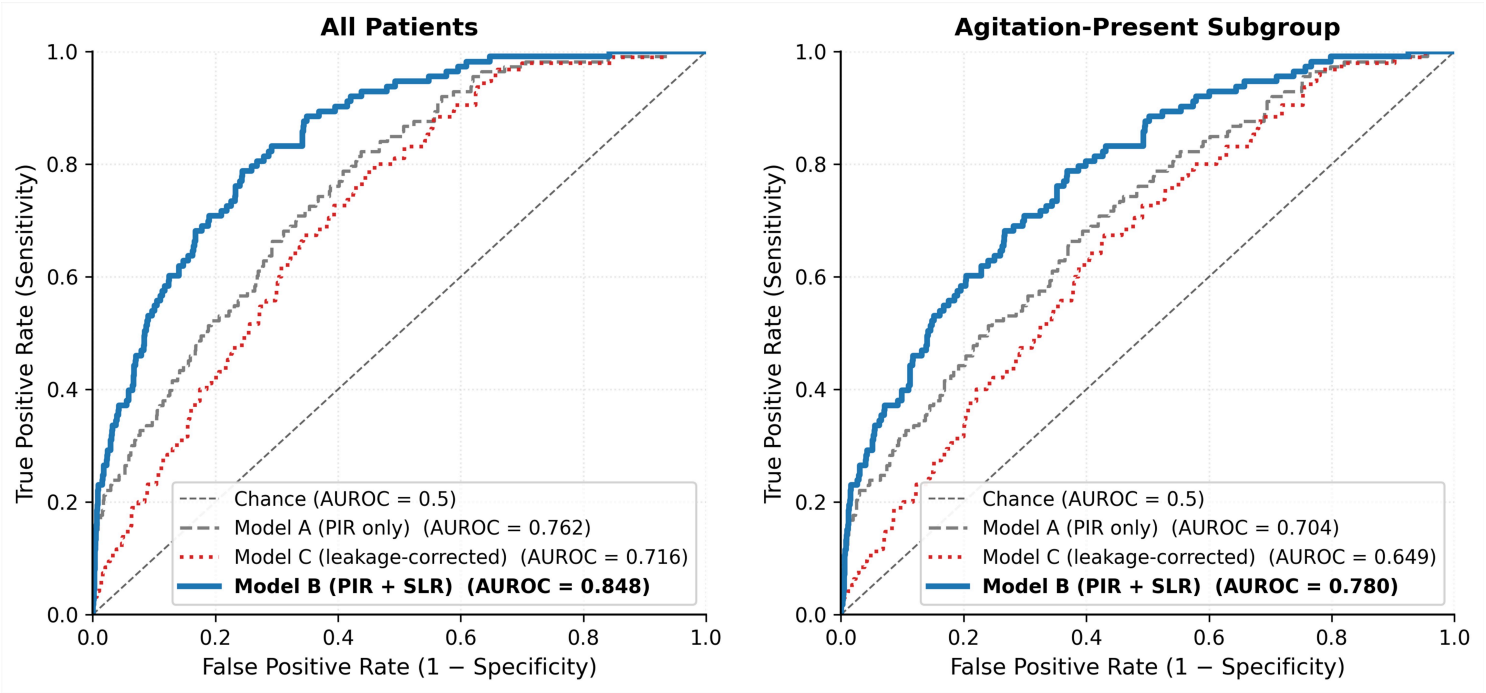


Figure 4

Receiver Operating Characteristic (ROC) curves from LOPO-CV. Left: full cohort (n = 56). Right: agitation-present sub-cohort (n = 27). Dashed diagonal line indicates chance-level performance

(AUROC = 0.5). AUROC values are shown in the legend. Model B (PIR + SLR) achieved the highest discrimination in both cohorts.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.TRIPODAIReportingChecklist.docx
- Additionalfile2.SleepSensitivityAnalysis.docx
- Additionalfile3.PerPatientPerformance.docx