European Journal of Clinical Investigation

WILEY

ORIGINAL ARTICLE OPEN ACCESS

Sleep Apnoea Variability in Pacemaker Patients: A Women-Predominant Phenotype

Philipp Spitaler[1] | Valentin Bilgeri[1] | Gudrun Feuchtner[2] | Patrick Rockenschaub[3] | Fabian Barbieri[4,5,6] | Birkan Tugrul[1] | Agne Adukauskaite[1] | Markus Stühlinger[1] | Bernhard Erich Pfeifer[7,8] | Sabine Ludwig[9,10] | Peter Willeit[3,11] | Axel Bauer[1] | Wolfgang Dichtl[1]

[1]Department of Internal Medicine III, Medical University of Innsbruck, Innsbruck, Austria | [2]Department of Radiology, Medical University of Innsbruck, Innsbruck, Austria | [3]Institute of Clinical Epidemiology, Public Health, Health Economics, Medical Statistics and Informatics, Medical University of Innsbruck, Innsbruck, Austria | [4]Department of Cardiology, Angiology and Intensive Care Medicine, Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany | [5]Department of Cardiology, Angiology and Intensive Care Medicine, Deutsches Herzzentrum der Charité, Berlin, Germany | [6]Institute of Active Polymers and Berlin-Brandenburg Center for Regenerative Therapies, Helmholtz-Zentrum Hereon, Teltow, Germany | [7]Landesinstitut für Integrierte Versorgung Tirol, Innsbruck, Austria | [8]Division for Digital Medicine and Telehealth, University for Health Sciences, Medical Informatics and Technology (UMIT-Tirol), Hall in Tirol, Austria | [9]Institute of Diversity in Medicine, Medical University of Innsbruck, Innsbruck, Austria | [10]Charité Centre for Global Health, Global Engagement, Charité – Universitätsmedizin Berlin, Berlin, Germany | [11]Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

Correspondence: Valentin Bilgeri (valentin.bilgeri@i-med.ac.at)

Received: 18 February 2026 | **Revised:** 19 April 2026 | **Accepted:** 30 April 2026

Keywords: implantable devices | night-to-night variability | phenotyping | sex/gender differences | sleep apnoea

ABSTRACT

Study Objectives: Single-night polysomnography assumes stable sleep apnoea (SA) severity; night-to-night variability may identify clinically relevant phenotypes, particularly in women. Sex/gender-specific longitudinal phenotypes were characterised using device-detected respiratory disturbance index (RDI) and an instability metric independent of mean severity.

Methods: Pacemaker patients in the prospective ACaSA study with at least 180 valid nightly RDI measurements after a 30-day post-implant period were analysed. Severity was summarised by mean RDI. Variability was summarised by an Instability Index (residual from the log–log regression of RDI standard deviation on mean RDI). Phenotypes were defined by mean RDI (<20 vs. ≥ 20 events/h) and instability (above vs. below median).

Results: About 37,260 nights from 207 patients (median age 76 years, IQR [69–81]) were analysed. Four phenotypes were identified: *Mild-Stable* ($n=68$), *Mild-Unstable* ($n=67$), *Severe-Stable* ($n=35$), and *Severe-Unstable* ($n=37$). Women comprised 37.7% of the total cohort and they were disproportionately overrepresented in the *Mild-Unstable* versus in the *Mild-Stable* phenotype (61.2% vs. 33.8%; $p=0.002$). Among patients with mean RDI <20 ($n=135$), female sex/gender (adjusted OR 3.38, 95% CI 1.60–7.40; $p=0.002$) and obesity (BMI $\geq 30\,kg/m^2$; adjusted OR 3.80, 95% CI 1.31–12.53; $p=0.019$) were independent predictors of the *Mild-Unstable* phenotype. Mild-Unstable had more severe-range nights (RDI ≥ 20/h) than Mild-Stable (7.8% vs. 2.2%; $p<0.001$). Severe-Unstable recorded RDI <20/h on 33.9% of nights versus 13.9% in Severe-Stable ($p<0.001$).

Conclusion: Night-to-night variability identifies longitudinal SA phenotypes not captured by mean RDI. A female- and obesity-predominant *Mild-Unstable* phenotype shows intermittent severe-range nights despite low average burden, while instability in severe SA increases the likelihood of underestimating severity when classification is based on a single-night test.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2026 The Author(s). *European Journal of Clinical Investigation* published by John Wiley & Sons Ltd on behalf of Stichting European Society for Clinical Investigation Journal Foundation.

1 | Introduction

Sleep apnoea (SA) is a highly prevalent sleep disorder associated with substantial cardiovascular, metabolic, and neurocognitive morbidity, including obesity, hypertension, diabetes and cognitive impairment [1, 2]. Despite its clinical importance, diagnosis and severity classification remain largely based on a single-night polysomnographic assessment. This cross-sectional approach implicitly assumes that apnoea severity is stable over time—an assumption that is increasingly questioned by evidence demonstrating marked night-to-night variability in apnoea–hypopnoea index (AHI) or respiratory disturbance index (RDI) measurements [3].

Physiological factors such as sleep architecture, body position, alcohol intake, and nasal obstruction can vary between nights and contribute to fluctuations in AHI/RDI [3–6]. A growing body of literature demonstrates that such variability can lead to clinically meaningful misclassification when diagnosis is based on a single night, particularly in individuals with borderline or moderate disease severity [3, 7]. However, most prior studies rely on short-term repeat polysomnography or limited home sleep testing and therefore do not capture long-term disease dynamics under real-world conditions.

Sex/gender differences further complicate the diagnostic landscape. Women with symptoms of sleep-disordered breathing are less likely to be diagnosed and treated for sleep apnoea than men [8]. Rather than presenting with classic symptoms such as loud snoring and excessive daytime sleepiness, women more frequently report fatigue, insomnia, mood disturbance and morning headache [8–10]. At comparable AHI values, women often experience greater symptom burden and impaired quality of life [9, 10]. From a pathophysiological perspective, women more commonly exhibit REM-predominant or positional sleep apnoea, phenotypes that may amplify night-to-night variability and reduce the reliability of single-night severity assessment [4]. Despite these recognised differences, sex/gender-specific longitudinal patterns of sleep apnoea severity and variability remain poorly characterised.

Capturing these temporal dynamics requires repeated measurements beyond a single-night study. Recent advances in implantable cardiac devices permit continuous, long-term monitoring of respiratory disturbance using thoracic impedance-based algorithms. Although not a replacement for polysomnography, these systems enable observation of sleep apnoea behaviour across hundreds of nights under unconstrained, real-world conditions. Device-detected RDI has shown reasonable agreement with polysomnography for identifying severe sleep apnoea, as shown in the DREAM European study [11]. Such longitudinal data provide an opportunity to move beyond static severity classification toward phenotyping approaches that incorporate temporal dynamics.

In this study, long-term device-detected RDI data from patients enrolled in the prospective ACaSA study were used to develop a longitudinal phenotyping framework integrating both mean disease severity and night-to-night instability. It was hypothesised that (1) longitudinal analysis would identify clinically distinct sleep apnoea phenotypes not captured by average severity alone, and (2) instability would represent a defining feature of sleep apnoea expression in women, contributing to under-recognition when conventional single-night diagnostic approaches are applied.

2 | Methods

2.1 | Study Cohort

Patients enrolled in the prospective observational ACaSA study (NCT05127720) and implanted with a Microport BOREA DR or TEO DR pacemaker formed the analysis cohort. Participants were recruited at the Medical University Innsbruck after informed consent, and the protocol was approved by the local Ethics Committee (EK No.: 1322/2020). We included patients with at least 180 valid nightly RDI measurements after a 30-day post-implant stabilisation period. For each patient, the first 180 valid nights following this stabilisation period were included in the analysis (Figure 1).

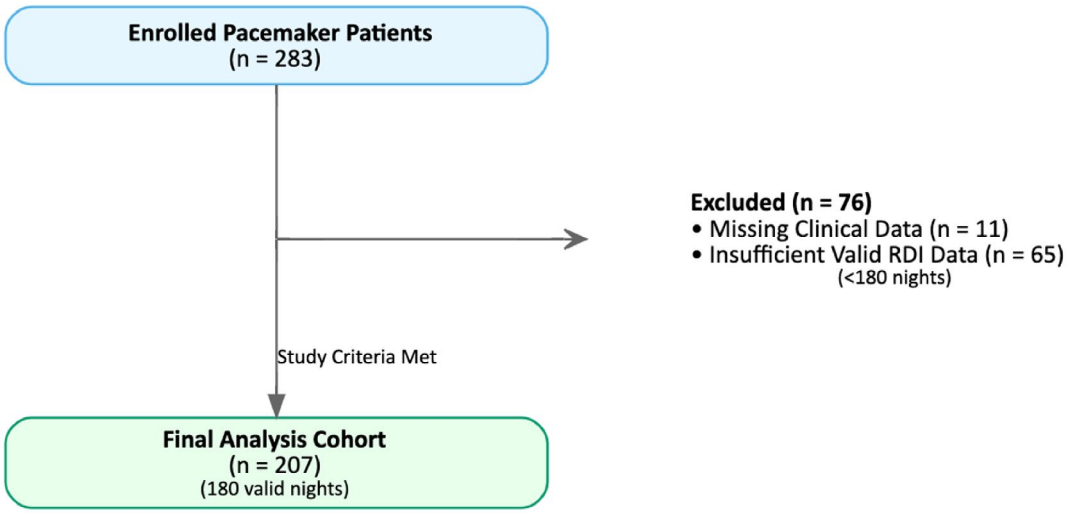


FIGURE 1 | Consort diagram.

2.2 | Data Collection

Device-detected sleep apnoea (DDSA) was defined by the device-measured RDI. The respiratory disturbance index (RDI) was measured from midnight to 5 o'clock in the morning, using thoracic impedance monitoring the patient's respiratory effort. The sleep apnoea algorithm in Microport pacemakers counts an apnoea event in the absence of a significant respiratory cycle for > 10 s and a hypopnoea event upon a sustained (> 10 s) reduction of the respiratory amplitude by at least 50% compared to the mean minute ventilation of preceding validated respiratory cycles. The RDI correlates with the apnoea-hypopnoea index (AHI) measured by polysomnography; an optimal cutoff value for detecting severe SA was found to be RDI > 20/h in the DREAM study [11]. Additionally, the measurements are labelled as invalid by the pacemaker software and not included in the analysis if there are more than 400 unstable or noisy signals per hour (due to movements while sleeping or active periods of the patients).

Because the device algorithm automatically invalidates and excludes nights with excessive signal noise (e.g., due to patient movement), the 180 valid nights were not strictly consecutive. In our cohort, the median chronological time required to accumulate 180 valid RDI measurements was 201 days (interquartile range: 189–224 days).

Baseline clinical data included demographics, comorbidities, medication use, laboratory parameters (including NT-proBNP), echocardiographic measurements, and pulmonary function testing collected at study enrollment.

Demographic data, including sex/gender, were retrospectively extracted from routine administrative electronic health records collected at the time of study inclusion. Operationally, this variable was recorded strictly as a binary administrative category (female/male).

2.3 | Longitudinal Feature Engineering

Two complementary dimensions of sleep apnoea behaviour were quantified over the 180-night window:

Severity (Mean RDI): calculated as the mean of nightly RDI values for each patient.

Instability (Instability Index): night-to-night variability was initially quantified as the standard deviation (SD) of nightly RDI values. However, SD alone was not used as the final instability metric because patients with higher mean RDI inherently exhibit higher absolute variability (Spearman's $\rho = 0.86$, $p < 0.001$) even without disproportionate night-to-night instability. The coefficient of variation (SD/mean) was also not used because in our cohort the relationship between mean RDI and SD was nonlinear; dividing by the mean would therefore have disproportionately inflated variability estimates in patients with lower mean RDI values.

Because variability scaled nonlinearly with average severity, we modelled the relationship between SD and mean RDI using log–log regression to linearise the data and stabilise the variance. The appropriateness of this transformation was formally assessed: homoscedasticity was confirmed using the Breusch–Pagan test ($p = 0.875$), demonstrating that the transformation successfully decoupled variance from mean severity, while visual inspection of residual plots confirmed a linear overall fit ($R^2 \approx 0.80$). The Instability Index was defined as the residual from this regression (Figure 2), representing whether a patient's variability was greater or lower than expected based on average severity.

$$\text{Instability_Index}_i = \log(\text{SD}_i) - \left[\beta_0 + \beta_1 \times \log(\text{Mean}_i)\right]$$

An equal-length observation window of 180 nights was used to avoid bias from unequal follow-up and to ensure comparability across participants. A median split was applied to create balanced groups for interpretability; continuous and quartile-based analyses were performed as sensitivity analyses.

2.4 | Phenotype Classification

Patients were classified into longitudinal sleep apnoea phenotypes using a two-dimensional framework incorporating average disease severity and night-to-night instability, both derived from the 180-night observation window. Severity was defined by mean nightly RDI and dichotomised at 20 events/h (mild: < 20; severe: ≥ 20), consistent with prior validation of device-detected RDI for identifying severe sleep apnoea (9). Instability was defined by the Instability Index (residual from the cohort-level log–log regression of SD on mean RDI) and dichotomised at the cohort median to yield balanced groups. Combining severity and instability created four phenotypes: Mild-Stable, Mild-Unstable, Severe-Stable, and Severe-Unstable.

2.5 | Statistical Analysis

Baseline characteristics were compared using Fisher's exact test for categorical variables and Wilcoxon rank-sum tests for continuous variables. Primary comparisons were performed within severity strata (Stable vs. Unstable within the Mild group; Stable vs. Unstable within the Severe group).

To evaluate whether the sex/gender-association with instability persisted after accounting for potential confounding, we conducted a multivariable logistic regression restricted to patients with mild disease (Mean RDI < 20 events/h). The dependent variable was Unstable phenotype membership (Mild-Unstable vs. Mild-Stable). Covariates were selected a priori based on biological plausibility and clinical relevance: sex, age, obesity (BMI ≥ 30 kg/m^2), heart failure, arterial hypertension and diabetes. Results are reported as odds ratios (OR) with 95% confidence intervals (CI).

Sensitivity analyses repeated phenotyping across observation windows ranging from 90 to 365 nights to evaluate robustness. To evaluate potential confounding by atrial fibrillation on respiratory sensing, sensitivity analyses were repeated restricting the dataset to nights with 0% device-detected AF burden and excluding patients with single chamber (VVI) devices lacking atrial sensing.

To validate the Instability Index, sensitivity analyses were performed treating instability as a continuous variable (per

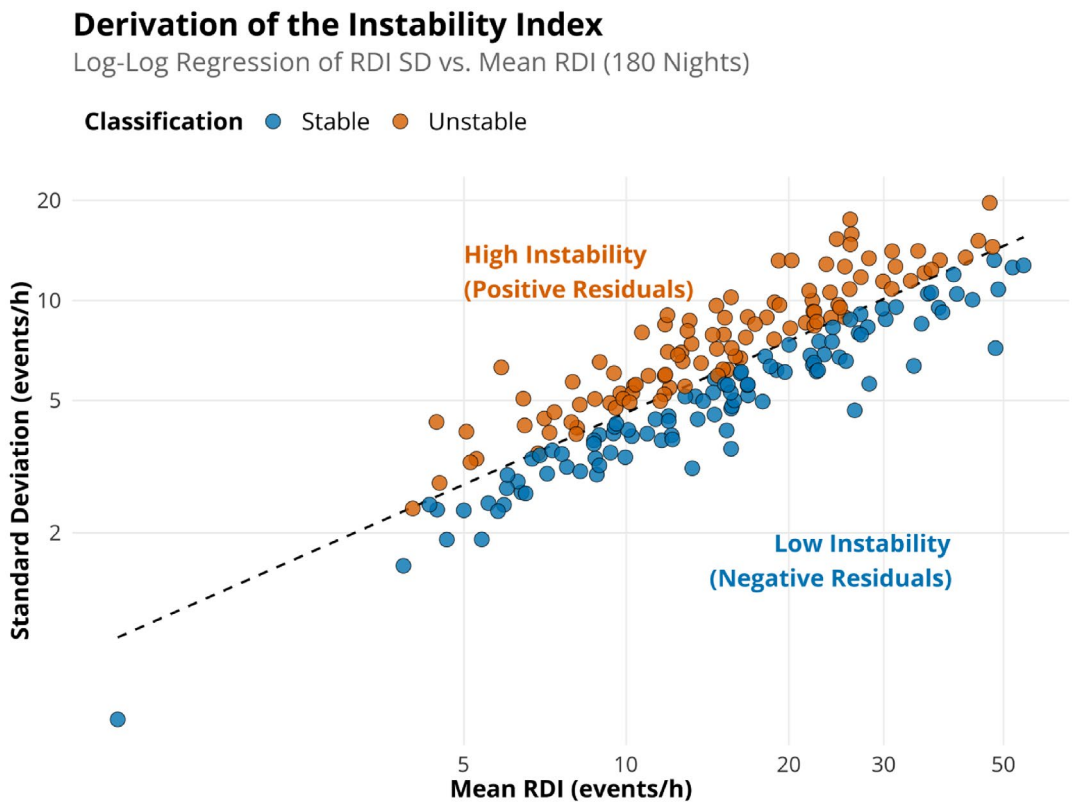


FIGURE 2 | Derivation of the instability index. Log–log regression of RDI standard deviation (*y*-axis) versus mean RDI (*x*-axis). The dashed line represents the expected variability for a given severity ($R^2 \approx 0.8$). The instability index is defined as the residual (vertical distance) from this line. Patients above the line (positive residuals, orange) are classified as 'Unstable', while those below (blue) are 'Stable'.

standard deviation increase) and categorising patients by quartiles rather than the median. The appropriateness of the log–log transformation was assessed using residual plots and the Breusch–Pagan test.

Analyses were performed using R version 4.5.1. A $p < 0.05$ was considered significant.

3 | Results

3.1 | Cohort Characteristics

Of 283 enrolled patients, a total of 207 patients met the inclusion criteria for the analysis (Figure 1). The median age was 76 years, and 37.7% were women. Across individuals, mean RDI and RDI SD showed a nonlinear relationship (Figure 2). Overall, women ($n = 78$) had lower mean RDI (11.4 [8.0–16.7] vs. 16.8 [12.1–26.2] events/h in men ($n = 129$); $p < 0.001$) but higher Instability Index (0.1 [−0.1–0.2] vs. −0.1 [−0.2–0.2]; $p = 0.017$). Baseline characteristics by sex/gender are summarised in Table 1.

3.2 | Longitudinal Phenotypes Defined by Severity and Instability

Using the combined framework of Mean RDI and Instability Index, four longitudinal phenotypes were identified (Figure 3A): *Mild-Stable* ($n = 68$), *Mild-Unstable* ($n = 67$), *Severe-Stable* ($n = 35$), and *Severe-Unstable* ($n = 37$). Notably, patients with similar mean RDI values frequently exhibited distinct longitudinal patterns, indicating that average severity alone did not capture disease dynamics. These phenotype-specific temporal patterns are illustrated in representative patient heatmaps (Figure 4), which highlight the stable low-burden profile of *Mild-Stable*, the intermittent 'speckled' pattern of *Mild-Unstable*, the persistently high-burden profile of *Severe-Stable*, and the fluctuating high-to-mild pattern of *Severe-Unstable*.

3.3 | Sex/Gender-Specific Distribution of Phenotypes

Women were significantly overrepresented in the *Mild-Unstable* phenotype. They comprised 61.2% of *Mild-Unstable* patients compared with 33.8% of *Mild-Stable* patients ($p = 0.002$; Table 2, Figure 3B). In contrast, severe phenotypes were predominantly male, and sex/gender distribution did not differ significantly between *Severe-Stable* and *Severe-Unstable* groups ($p = 0.768$). Baseline medication use was comparable across these four distinct groups (Table S3). Sensitivity analyses demonstrated that the association between female sex and instability was consistent across observation windows exceeding 90 nights (Table S1).

3.4 | Multivariable Predictors of Instability in Mild Disease

To assess whether the observed sex/gender-associated instability could be explained by differences in age, adiposity, or

TABLE 1 | Baseline clinical characteristics by sex/gender.

Category	Characteristics	Overall ($N = 207$)	Men ($n = 129$)	Women ($n = 78$)	p
Demographics	Age, years	76 [69–81]	76 [70–81]	75 [69–81]	0.464
	BMI, kg/m^2	25.7 [22.8–28.4]	26.2 [23.8–28.4]	23.8 [21.3–28.6]	**0.018**
Device indication	SSS	114 (56.2%)	68 (54.0%)	46 (59.7%)	0.467
	AVB II (Wenckebach)	4 (2.0%)	4 (3.2%)	0 (0.0%)	0.300
	AVB II (Mobitz)	23 (11.3%)	10 (7.9%)	13 (16.9%)	0.067
	AVB III	46 (22.7%)	30 (23.8%)	16 (20.8%)	0.730
	Bradyarrhythmia	16 (7.9%)	14 (11.1%)	2 (2.6%)	0.032
Comorbidities	Arterial hypertension	136 (66.3%)	89 (70.1%)	47 (60.3%)	0.172
	Diabetes mellitus	56 (27.2%)	39 (30.5%)	17 (21.8%)	0.199
	AF history	98 (47.3%)	63 (48.8%)	35 (44.9%)	0.667
	CAD	72 (35.0%)	54 (42.2%)	18 (23.1%)	**0.007**
	Previous MI	18 (8.7%)	12 (9.4%)	6 (7.7%)	0.802
Medications	Beta blocker	61 (29.5%)	32 (24.8%)	29 (37.2%)	0.062
	Amiodarone	5 (2.4%)	3 (2.3%)	2 (2.6%)	1.000
	Flecainide	4 (1.9%)	3 (2.3%)	1 (1.3%)	1.000
	Digitalis	5 (2.4%)	4 (3.1%)	1 (1.3%)	0.652
	CCB	46 (22.2%)	30 (23.3%)	16 (20.5%)	0.731
	Statin	124 (59.9%)	82 (63.6%)	42 (53.8%)	0.189
	Loop diuretic	33 (15.9%)	15 (11.6%)	18 (23.1%)	**0.033**
	HCT	46 (22.2%)	30 (23.3%)	16 (20.5%)	0.731
	MRA	24 (11.6%)	18 (14.0%)	6 (7.7%)	0.188
	ACE inhibitor or ARB	111 (53.6%)	72 (55.8%)	39 (50.0%)	0.473
	Aspirin	62 (30.0%)	42 (32.6%)	20 (25.6%)	0.348
	NOAC	84 (40.6%)	55 (42.6%)	29 (37.2%)	0.468
Clinical parameters	LVEF, %	58.0 [53.5–63.2]	57.5 [53.3–63.0]	58.3 [54.1–63.6]	0.326
	NT-proBNP (ng/L)	359 [141–796]	352 [125–686]	391 [156–1036]	0.350
	LDL cholesterol (mg/dL)	83.5 [62.0–121.8]	80.0 [60.8–111.0]	93.5 [63.2–132.8]	0.079
	FEV1/FVC z-score	0.3 [−0.4–0.8]	0.3 [−0.3–0.8]	0.3 [−0.5–0.9]	0.865
Sleep apnoea	Mean RDI (180-day)	15.2 [9.5–24.0]	16.8 [12.1–26.2]	11.4 [8.0–16.7]	**< 0.001**
	Instability Index	0.0 [−0.2–0.2]	−0.1 [−0.2–0.2]	0.1 [−0.1–0.2]	**0.017**

Note: Data are median [IQR] or n (%). p-values compare men versus women using Wilcoxon rank-sum tests for continuous variables and Fisher's exact tests for categorical variables. Significant associations ($p < 0.05$) are highlighted in bold.

Abbreviations: ACE/ARB, angiotensin-converting enzyme inhibitor/angiotensin receptor blocker; AF, atrial fibrillation; AVB, atrioventricular block; BMI, body mass index; CAD, coronary artery disease; CCB, calcium channel blocker; FEV1/FVC, forced expiratory volume in 1 s/forced vital capacity; HCT, hydrochlorothiazide; IQR, interquartile range; LDL, low density lipoprotein; LVEF, left ventricular ejection fraction; MI, myocardial infarction; MRA, mineralocorticoid receptor antagonist; NOAC, non-vitamin K antagonist oral anticoagulant; NT-proBNP, N-terminal pro-B-type natriuretic peptide; RDI, Respiratory Disturbance Index; SSS, sick sinus syndrome.

comorbidities, we performed a multivariable logistic regression restricted to patients with mild mean severity (Mean RDI < 20 events/h; $n = 135$). In this expanded model, female sex/gender remained a strong independent predictor of being classified as *Mild-Unstable* (adjusted OR 3.38, 95% CI 1.60–7.40; $p = 0.002$).

We also identified obesity (BMI ≥ 30 kg/m^2) as a second, independent driver of instability (adjusted OR 3.80, 95% CI 1.31–12.53; $p = 0.019$). Age retained a borderline association (OR 1.27 per 5 years; $p = 0.049$). Importantly, other cardiovascular comorbidities including heart failure, arterial hypertension and diabetes were not significantly associated with the unstable

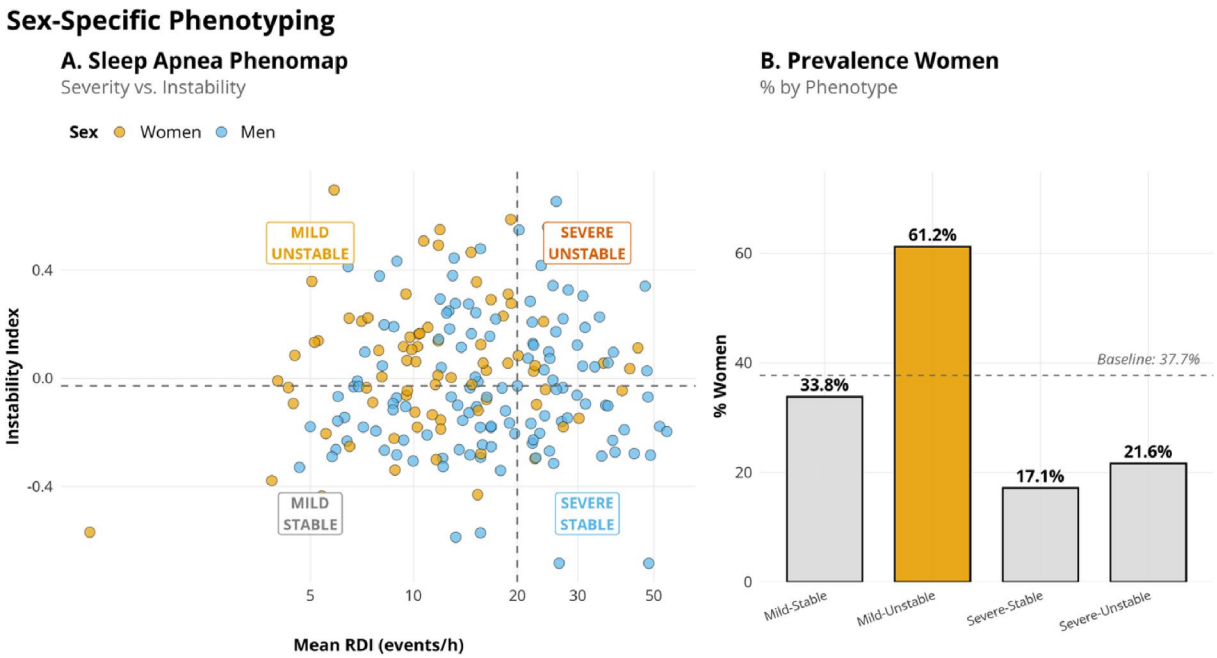


FIGURE 3 | Sex/gender-specific phenotyping. (A) Sleep apnoea phenomap: patients plotted by mean RDI (x-axis) and instability index (y-axis). Women (orange) significantly cluster in the 'Mild-Unstable' quadrant. (B) Prevalence of women: percentage of women within each phenotype. The 'Mild-Unstable' group is 61.2% women, significantly higher than all other groups ($p < 0.001$).

phenotype (Table 3). Conversely, in the severe disease cohort ($n = 72$), neither female sex ($p = 0.701$) nor obesity ($p = 0.982$) predicted instability.

3.5 | Single-Night Classification Vulnerability Driven by Instability

Instability was associated with marked vulnerability to single-night severity misclassification (Table 4). In mild disease, *Mild-Unstable* patients spent a significantly greater proportion of nights in the severe range (7.8% vs. 2.2% in *Mild-Stable*; $p < 0.001$), indicating that a low mean RDI concealed frequent severe-range excursions. In severe disease, instability was associated with frequent nights below the severe threshold: *Severe-Unstable* patients recorded RDI < 20 events/h on 33.9% of nights versus 13.9% in *Severe-Stable* patients ($p < 0.001$), indicating increased susceptibility to under-classification when severity is inferred from a single-night assessment.

3.6 | Validation of the Instability Index

To ensure the findings were not artefacts of dichotomisation, the core relationships were re-evaluated treating the Instability Index as a continuous variable (Z-score). In a continuous multivariable linear regression model restricted to the mild cohort, female sex remained a robust, independent predictor of higher instability ($\beta = 0.43$ SD units, 95% CI [0.10–0.77]; $p = 0.012$), after adjusting for age, obesity, and above-mentioned comorbidities. Obesity demonstrated a positive trend with continuous instability ($\beta = 0.38$ SD units; $p = 0.089$). Furthermore, the proportion of women showed a significant increasing trend across quartiles of instability (p-trend = 0.008), rising from 32.4% in the lowest quartile to 54.5% and 69.7% in the upper quartiles.

The Instability Index also predicted night-to-night threshold crossings: in mild disease, every 1-SD increase in instability was associated with 1.31-fold higher odds of a night exceeding the severe threshold (OR 1.31, 95% CI 1.24–1.39; $p < 0.001$), independent of the mean RDI.

3.7 | Sensitivity Analysis

To confirm that the observed instability was not driven by atrial fibrillation or artefacts, we performed a sensitivity analysis restricted to confirmed sinus rhythm (nights with 0% AF burden, excluding patients with single chamber [VVI] devices). The association between female sex and the *Mild-Unstable* phenotype remained robust and statistically significant across all observation windows of 90 days or longer (Table S2).

4 | Discussion

In this longitudinal analysis of device-detected sleep apnoea, distinct phenotypes emerged when disease severity was characterised jointly by mean burden and instability. This framework identifies clinically important heterogeneity that is not apparent when sleep apnoea is summarised by a single-night AHI/RDI or by average severity alone. The results extend prior observations of night-to-night variability by demonstrating that instability forms a reproducible dimension that stratifies patients

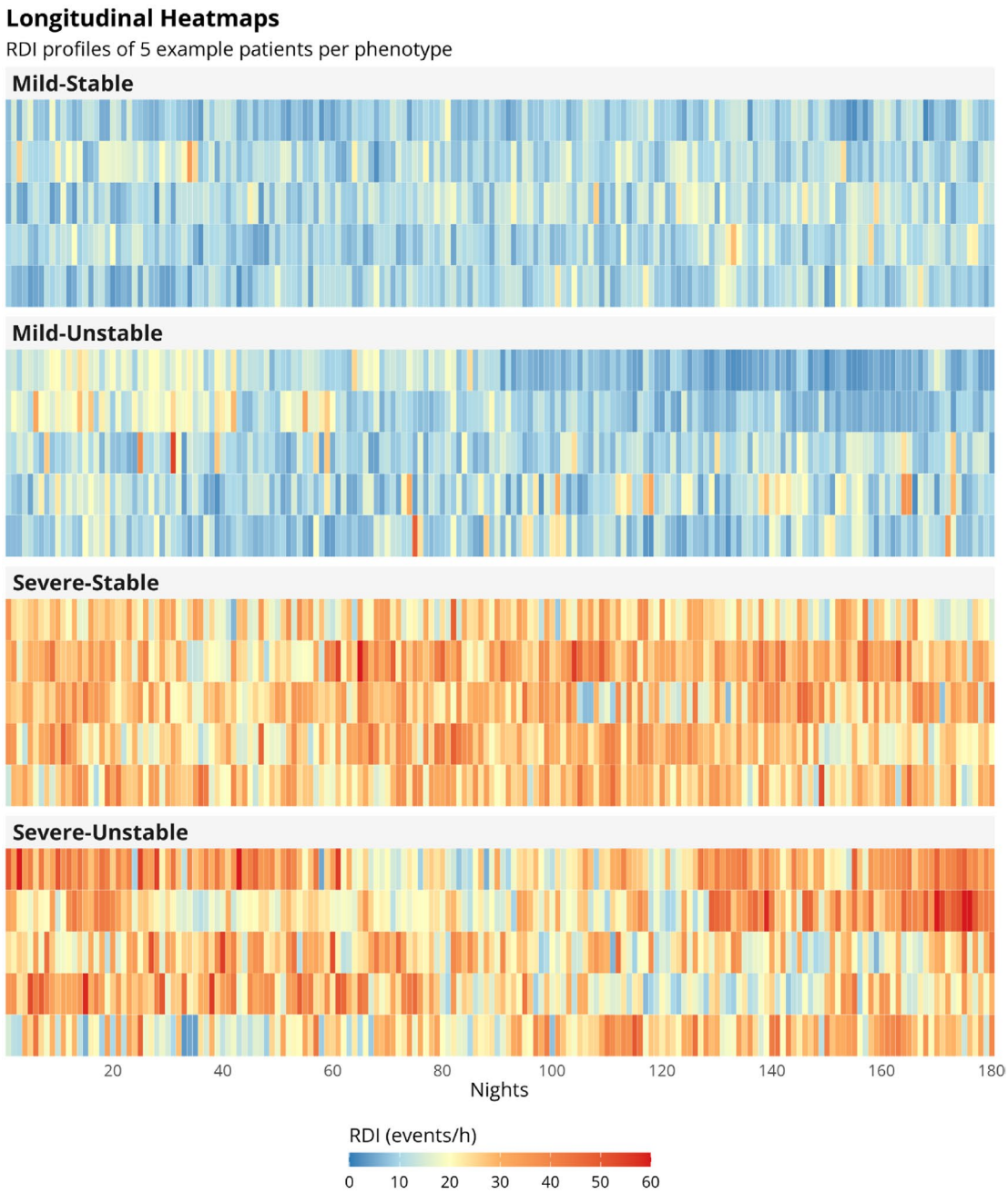


FIGURE 4 | Longitudinal heatmaps of example patients. Heatmaps displaying the night-to-night respiratory disturbance index (RDI) for five representative patients from each of the four identified phenotypes over their first 90 valid nights of monitoring. The colour scale indicates daily RDI severity, ranging from blue (low/healthy) to red (severe). Note the distinct 'speckled' pattern in the Mild-Unstable group (predominantly women), where nights of low burden are frequently interspersed with spikes of moderate-to-severe disease, contrasting with the consistent blue of the Mild-Stable group. Conversely, the Severe-Unstable group shows periods of high burden interrupted by nights of mild values, illustrating vulnerability to under-classification when severity is inferred from single-night testing.

into phenotypes with markedly different vulnerability to single-night severity misclassification [3].

A central finding of this study is the identification of a distinct Mild-Unstable phenotype, which multivariable analysis revealed is driven by two powerful, independent pathways: female sex/gender and obesity. Women were significantly over-represented in this volatile phenotype, and female sex/gender remained a robust predictor (adjusted OR 3.38) even after adjusting for obesity and comorbidities. Simultaneously, obesity emerged as a distinct, independent driver of instability (adjusted OR 3.80). While obesity likely contributes to variability via anatomical mechanisms—such as positional dependence—the persistence of the female signal implies distinct non-anatomical mechanisms. Recent data further suggest that women may be more susceptible to environmental influences on sleep-disordered breathing, which could also contribute to the greater night-to-night variability observed in women [10]. Because our study lacks hormonal assays, sleep-stage scoring, and body position data, the physiological drivers of this female-predominant

TABLE 2 | Baseline clinical characteristics by phenotype.

Category	Characteristics	Overall ($n = 207$)	Mild-Stable ($n = 68$)	Mild-Unstable ($n = 67$)	p (Mild)	Severe-Stable ($n = 35$)	Severe-Unstable ($n = 37$)	p (Severe)
Demographics	Age, years	76 [69–81]	74 [67–78]	77 [69–81]	**0.049**	80 [74–83]	75 [70–81]	**0.040**
	Women	78 (37.7%)	23 (33.8%)	41 (61.2%)	**0.002**	6 (17.1%)	8 (21.6%)	0.768
	BMI, kg/m^2	25.7 [22.8–28.4]	24.8 [22.4–26.9]	26.0 [22.7–30.0]	0.144	26.2 [24.4–28.1]	27.1 [23.8–29.4]	0.407
Device indications	SSS	114 (56.2%)	39 (57.4%)	35 (54.7%)	0.861	21 (60.0%)	19 (52.8%)	0.634
	AVB II (Wenckebach)	4 (2.0%)	1 (1.5%)	2 (3.1%)	0.611	0 (0.0%)	1 (2.8%)	1.000
	AVB II (Mobitz)	23 (11.3%)	11 (16.2%)	7 (10.9%)	0.452	3 (8.6%)	2 (5.6%)	0.674
	AVB III, *n* (%)	46 (22.7%)	14 (20.6%)	19 (29.7%)	0.237	3 (8.6%)	10 (27.8%)	0.063
	Bradyarrhythmia	16 (7.9%)	3 (4.4%)	1 (1.6%)	0.620	8 (22.9%)	4 (11.1%)	0.220
Comorbidities	Arterial hypertension	136 (66.3%)	40 (58.8%)	44 (66.7%)	0.376	24 (70.6%)	28 (75.7%)	0.789
	Diabetes	56 (27.2%)	23 (33.8%)	11 (16.7%)	**0.029**	11 (31.4%)	11 (29.7%)	1.000
	AF history	98 (47.3%)	24 (35.3%)	30 (44.8%)	0.294	23 (65.7%)	21 (56.8%)	0.476
	CAD	72 (35.0%)	26 (38.2%)	20 (30.3%)	0.367	10 (28.6%)	16 (43.2%)	0.227
	Previous MI	18 (8.7%)	7 (10.3%)	5 (7.6%)	0.764	2 (5.7%)	4 (10.8%)	0.675
Clinical parameters	LVEF, %	58.0 [53.5–63.2]	59.6 [55.3–64.5]	59.4 [54.0–63.3]	0.497	57.6 [54.3–63.2]	54.1 [45.0–58.3]	**0.007**
	NT-proBNP (ng/L)	359 [141–796]	265 [90–528]	349 [144–660]	0.075	595 [335–1102]	686 [176–1515]	0.797
	LDL cholesterol (mg/dL)	83.5 [62.0–121.8]	78.0 [60.0–120.5]	96.5 [66.0–122.5]	0.231	79.0 [62.0–130.0]	86.0 [57.8–108.8]	0.723
	FEV1/FVC Z	0.3 [−0.4 to 0.8]	0.3 [−0.5 to 0.8]	0.5 [−0.4 to 1.0]	0.197	0.4 [−0.3 to 0.7]	0.2 [−0.1 to 0.6]	0.858

Note: Phenotypes were defined by mean respiratory disturbance index (RDI) (< 20 vs. ≥ 20 events/h) and instability (Instability Index below vs. above the cohort median). Data are median [IQR] or *n* (%). *p*-values compare Stable versus Unstable within severity strata (Mild and Severe). Significant associations ($p < 0.05$) are highlighted in bold.

Abbreviations: AF, atrial fibrillation; AVB, atrioventricular block; BMI, body mass index; CAD, coronary artery disease; FEV1/FVC, forced expiratory volume in 1 s/forced vital capacity; IQR, interquartile range; LDL, low-density lipoprotein; LVEF, left ventricular ejection fraction; MI, myocardial infarction; NT-proBNP, N-terminal pro-B-type natriuretic peptide; RDI, respiratory disturbance index; SSS, sick sinus syndrome.

TABLE 3 | Predictors of the unstable phenotype (univariate and multivariable).

Cohort	Variable	Univariate OR (95% CI)	*p*	Adjusted OR (95% CI)	*p*
Mild disease (Mild-Stable vs. Mild-Unstable)	Women	**3.35 (1.66–6.94)**	**< 0.001**	**3.38 (1.60–7.40)**	**0.002**
	Obesity (BMI ≥ 30)	**2.56 (1.04–6.74)**	**0.046**	**3.80 (1.31–12.53)**	**0.019**
	Age (per 5 years)	1.22 (1.00–1.52)	0.060	**1.27 (1.01–1.61)**	**0.049**
	Heart failure	1.71 (0.54–5.94)	0.370	1.42 (0.40–5.41)	0.588
	Hypertension	1.35 (0.67–2.76)	0.406	1.27 (0.57–2.88)	0.564
	Diabetes	0.51 (0.17–1.43)	0.212	0.30 (0.08–1.02)	0.066
Severe disease (Severe-Stable vs. Severe-Unstable)	Women	1.29 (0.40–4.36)	0.675	1.27 (0.38–4.41)	0.701
	Obesity (BMI ≥ 30)	1.35 (0.39–5.03)	0.637	1.02 (0.26–4.02)	0.982
	Age (per 5 years)	0.82 (0.60–1.09)	0.180	0.79 (0.56–1.07)	0.139
	Heart failure	1.25 (0.26–6.78)	0.779	1.39 (0.24–8.90)	0.711
	Hypertension	1.30 (0.45–3.78)	0.629	1.72 (0.54–5.73)	0.360
	Diabetes	1.03 (0.36–2.99)	0.958	1.00 (0.31–3.20)	0.993

Note: 'Mild Disease' refers to patients with Mean RDI < 20 events/h ($n = 135$); 'Severe Disease' refers to patients with Mean RDI ≥ 20 events/h ($n = 72$). The dependent variable is the 'Unstable' phenotype (vs. 'Stable') within each severity stratum. Adjusted odds ratios (aOR) are derived from a multivariable logistic regression model including all listed covariates. Obesity is defined as BMI ≥ 30 kg/m^2. Significant associations ($p < 0.05$) are highlighted in bold.
Abbreviations: BMI, body mass index; CI, confidence interval; OR, odds ratio; RDI, respiratory disturbance index.

instability remain speculative. However, we hypothesise that in our predominantly elderly cohort, postmenopausal hormonal changes may play a role [11]. The withdrawal of progesterone (a respiratory stimulant) and altered sleep architecture (lowered arousal threshold) could theoretically generate a 'physiological instability'. This hypothesised mechanism would result in an intermittent disease pattern consistent with known sex-linked phenotypes, such as REM-predominant or positional disease [4, 11], and mirrors the atypical symptom profiles reported by women that frequently contribute to clinical under-recognition [8, 9, 11]. Further research incorporating full polysomnography is required to test these hypotheses.

The clinical relevance of instability is underscored by the magnitude of diagnostic volatility. Within mild disease, instability identified individuals who spent substantially more nights above the severe threshold despite low mean RDI. This 'hidden severity' provides a plausible mechanism by which a single-night diagnostic assessment can underestimate true disease burden in intermittently severe patients, a limitation previously highlighted in studies of night-to-night variability [3]. In severe disease, the converse pattern was observed: *Severe-Unstable* patients fell below the severe threshold on approximately one-third of nights, indicating substantial vulnerability to under-classification if severity were inferred from a single-night assessment performed on a low-burden night. Together, these findings indicate that night-to-night variability, as captured by the Instability Index, is not simply measurement noise but a clinically meaningful trait with direct implications for diagnostic reliability.

These results suggest that reliance on a single-night assessment may be least reliable when sleep apnoea burden is intermittent. Multi-night evaluation (repeat testing or multi-night home sleep apnoea testing) may be most informative for symptomatic women with borderline/mild indices, suspected REM/positional disease, or discordant clinical suspicion and test results [3]. Incorporating instability metrics alongside mean severity could improve interpretation of sleep apnoea testing and reduce under-recognition in phenotypes characterised by night-to-night fluctuation.

While our primary analysis utilised a median split for clarity, the relationship between instability and sex was robust to continuous modelling and quartile-based sensitivity analyses. This suggests that the 'unstable' phenotype represents a biological continuum rather than a discrete category created by arbitrary cut-points.

Several limitations should be considered. First, device-derived RDI is a validated surrogate but has several technical limitations compared to polysomnography [12]. The thoracic impedance algorithm cannot reliably differentiate between obstructive and central sleep apnoea. Furthermore, because the device utilises a fixed time window (midnight to 05:00) rather than true sleep staging, periods of wakefulness occurring within this window may be misclassified as sleep, potentially underestimating the true RDI. While the algorithm demonstrates high sensitivity (89%) and specificity (85%) for severe disease, accuracy may be lower in patients with borderline or mild sleep apnoea. It also does not provide arousal scoring or desaturation metrics. Second, information on sleep apnoea treatment and nightly adherence (e.g., positive airway pressure) was not available and may influence night-to-night burden. Third, the cohort consisted entirely of elderly pacemaker recipients (median age 76 years) with high rates of cardiovascular comorbidity. This demographic differs substantially from the broader, generally younger sleep apnoea population evaluated in primary care or sleep clinics.

TABLE 4 | Longitudinal sleep apnoea features by phenotype.

Characteristics	Overall ($n = 207$)	Mild-Stable ($n = 68$)	Mild-Unstable ($n = 67$)	p (Mild)	Severe-Stable ($n = 35$)	Severe-Unstable ($n = 37$)	p (Severe)
Mean RDI (events/h)	15.2 [9.5–24.0]	10.2 [7.1–15.3]	11.8 [8.2–14.7]	0.472	28.2 [24.5–38.2]	25.9 [22.5–31.5]	0.051
Instability Index	0.0 [−0.2 to 0.2]	−0.2 [−0.3 to −0.1]	0.2 [0.1–0.3]	**< 0.001**	−0.2 [−0.3 to −0.1]	0.1 [0.1–0.3]	**< 0.001**
Severe night probability (% nights ≥ 20/h)	18.3 [2.8–60.6]	2.2 [0.0–15.1]	7.8 [3.6–20.0]	**< 0.001**	86.1 [70.8–97.8]	66.1 [54.4–81.1]	**< 0.001**
Mild night probability (% nights < 20/h)	81.7 [39.4–97.2]	97.8 [84.9–100.0]	92.2 [80.0–96.4]	**< 0.001**	13.9 [2.2–29.2]	33.9 [18.9–45.6]	**< 0.001**

Note: Mean RDI and Instability Index were calculated over the 180-night window. 'Severe-night probability' and 'mild-night probability' reflect the per-patient proportion of nights with RDI ≥ 20 and < 20 events/h, respectively. Data are median [IQR]. p-values compare Stable versus Unstable within severity strata. Significant associations ($p < 0.05$) are highlighted in bold.
Abbreviations: IQR, interquartile range; RDI, respiratory disturbance index.

Consequently, our conclusions regarding sex-specific phenotypes and instability must be interpreted with caution, and extrapolation to younger or less comorbid populations requires confirmation in future studies. Fourth, residual confounding cannot be excluded, and future work integrating sleep-stage information, body position, and concurrent polysomnography will be important to clarify physiological drivers of instability. The persistence of the sex–instability association in analyses restricted to sinus-rhythm nights argues against atrial fibrillation–related sensing artefact as the primary explanation for the observed phenotype distribution. The sinus-rhythm-restricted analyses also reduce concern that irregular rhythm or mode-specific sensing properties drive the observed mean–variance relationship. Together, these sensitivity analyses support that instability reflects a patient-level feature rather than a device- or rhythm-dependent artefact. Finally, regarding the operational definition of our demographic variables, sex/gender was acquired retrospectively as a binary administrative variable from electronic health records. Consequently, a strict retrospective differentiation between biological sex at birth and self-identified gender cannot be determined. While we utilise the composite term 'sex/gender' to reflect the entanglement of biological and sociocultural dimensions driving sleep apnoea phenotypes [13], the inability to distinguish these factors in the raw operational data remains a limitation.

In summary, longitudinal profiling based on mean severity and instability identifies sleep apnoea phenotypes with strong sex/gender differences. A female and obesity-predominant *Mild-Unstable* phenotype is characterised by frequent severe-night excursions despite low average burden, while instability in severe disease increases vulnerability to under-classification when severity is inferred from single-night severity assessment. These findings support consideration of longitudinal assessment strategies for sleep apnoea, particularly in patient groups where intermittency may undermine cross-sectional diagnostic approaches.

5 | Conclusion

Night-to-night variability in SA identifies a female- and obesity-predominant *Mild-Unstable* phenotype characterised by substantial hidden disease burden. Night-to-night variability in severe SA also poses a diagnostic challenge by increasing susceptibility to under-classification. Recognising these specific volatility patterns could be critical for moving beyond single-night diagnostics and ensuring equitable care for patients with intermittent, yet clinically significant, sleep apnoea.

Author Contributions

Philipp Spitaler: conceptualisation, investigation, writing – original draft, writing – review and editing, validation, methodology, project administration, data curation, resources. **Valentin Bilgeri:** conceptualisation, investigation, writing – original draft, project administration, data curation, resources, methodology, validation, writing – review and editing. **Gudrun Feuchtner:** conceptualisation, investigation, writing – original draft, methodology, validation, writing – review and editing. **Patrick Rockenschaub:** conceptualisation, investigation, methodology, validation, formal analysis, software. **Fabian Barbieri:** investigation, resources, writing – review and editing. **Birkan Tugrul:** investigation, data curation, writing – review and editing. **Agne**

Adukauskaite: investigation, data curation, writing – review and editing. **Markus Stühlinger:** resources, methodology, writing – review and editing. **Bernhard Erich Pfeifer:** resources, software, writing – review and editing. **Sabine Ludwig:** methodology, validation, writing – review and editing. **Peter Willeit:** writing – review and editing, supervision, methodology, formal analysis. **Axel Bauer:** supervision, resources, methodology, writing – review and editing. **Wolfgang Dichtl:** conceptualisation, investigation, writing – original draft, writing – review and editing, validation, methodology, supervision, resources, project administration.

Acknowledgements

We thank Dr. Wilhelm Grander and Dr. Simone Gasser for their support during patient recruitment. "Graphical Abstract" created in BioRender. Bilgeri V. (https://BioRender.com/vq5w9ea) is licensed under CC BY 4.0. Open Access funding provided by Medizinische Universitat Innsbruck.

Funding

This research was conducted as part of the prospective ACaSA study (NCT05127720) at the Department of Internal Medicine III, Medical University of Innsbruck. The ACaSA study is supported by an unrestricted educational grant from Microport and a grant from the 'Österreichischer Herzfonds'.

Ethics Statement

The ACaSA study was conducted in accordance with the Declaration of Helsinki. The protocol was approved by the Ethics Committee of the Medical University of Innsbruck (EK No.: 1322/2020, approval date 11 March 2021). All participants provided written informed consent prior to enrolment.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data supporting the findings of the ACaSA study (NCT05127720) are available from the corresponding author, Valentin Bilgeri, upon reasonable request. Due to the sensitive nature of patient-level pacemaker data and institutional ethics committee regulations at the Medical University of Innsbruck, data sharing is subject to formal approval and data privacy agreements.

References

1. E. Tasali, S. Pamidi, N. Covassin, and V. K. Somers, "Obstructive Sleep Apnea and Cardiometabolic Disease: Obesity, Hypertension, and Diabetes," *Circulation Research* 137, no. 5 (2025): 764–787, https://doi.org/10.1161/CIRCRESAHA.125.325676.

2. S. Guo, C. Liang, J. Fei, et al., "Relationship Between Sleep Apnea Severity and Mild Cognitive Impairment in People With Obstructive Sleep Apnea," *BMC Psychiatry* 25, no. 1 (2025): 790, https://doi.org/10.1186/s12888-025-07244-x.

3. M. Roeder, M. Bradicich, E. I. Schwarz, et al., "Night-To-Night Variability of Respiratory Events in Obstructive Sleep Apnoea: A Systematic Review and Meta-Analysis," *Thorax* 75, no. 12 (2020): 1095–1102, https://doi.org/10.1136/thoraxjnl-2020-214544.

4. L. Ye, G. W. Pien, S. J. Ratcliffe, et al., "The Different Clinical Faces of Obstructive Sleep Apnoea: A Cluster Analysis," *European Respiratory Journal* 44, no. 6 (2014): 1600–1607, https://doi.org/10.1183/09031936.00032314.

5. E. Lindberg, B. Benediktsdottir, K. A. Franklin, et al., "Women With Symptoms of Sleep-Disordered Breathing Are Less Likely to Be Diagnosed and Treated for Sleep Apnea Than Men," *Sleep Medicine* 35 (2017): 17–22, https://doi.org/10.1016/j.sleep.2017.02.032.

6. A. Valipour, H. Lothaller, H. Rauscher, H. Zwick, O. C. Burghuber, and P. Lavie, "Gender-Related Differences in Symptoms of Patients With Suspected Breathing Disorders in Sleep: A Clinical Population Study Using the Sleep Disorders Questionnaire," *Sleep* 30, no. 3 (2007): 312–319, https://doi.org/10.1093/sleep/30.3.312.

7. A. J. Wimms, H. Woehrle, S. Ketheeswaran, D. Ramanan, and J. Armitstead, "Obstructive Sleep Apnea in Women: Specific Issues and Interventions," *BioMed Research International* 2016 (2016): 1764837, https://doi.org/10.1155/2016/1764837.

8. M. R. Bonsignore, E. Mazzuca, P. Baiamonte, B. Bouckaert, W. Verbeke, and D. A. Pevernagie, "REM Sleep Obstructive Sleep Apnoea," *European Respiratory Review* 33, no. 171 (2024): 230166, https://doi.org/10.1183/16000617.0166-2023.

9. P. Defaye, I. de la Cruz, J. Martí-Almor, et al., "A Pacemaker Transthoracic Impedance Sensor With an Advanced Algorithm to Identify Severe Sleep Apnea: The DREAM European Study," *Heart Rhythm* 11, no. 5 (2014): 842–848, https://doi.org/10.1016/j.hrthm.2014.02.011.

10. V. Bilgeri, P. Spitaler, P. Rockenschaub, et al., "Weather-Associated Variations of Device-Detected Severe Sleep Apnoea in Cardiac Pacemaker Patients," *Journal of Sleep Research* 34, no. 5 (2025): e70024, https://doi.org/10.1111/jsr.70024.

11. S. A. Landry, C. J. Beatty, L. D. J. Thomson, et al., "A Review of Supine Position Related Obstructive Sleep Apnea: Classification, Epidemiology, Pathogenesis and Treatment," *Sleep Medicine Reviews* 72 (2023): 101847, https://doi.org/10.1016/j.smrv.2023.101847.

12. E. Simou, J. Britton, and J. Leonardi-Bee, "Alcohol and the Risk of Sleep Apnoea: A Systematic Review and Meta-Analysis," *Sleep Medicine* 42 (2018): 38–46, https://doi.org/10.1016/j.sleep.2017.12.005.

13. M. J. Legato, ed., *Principles of Gender-Specific Medicine: Sex and Gender-Specific Biology in the Postgenomic Era*, 4th ed. (Academic Press, 2023).

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Table S1:** Sensitivity analysis (observation windows). **Table S2:** Sensitivity analysis (sinus rhythm). **Table S3:** Baseline medication use by phenotype.