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.

Social determinants and disparities of past-year panic attacks in US young adults: a nationally representative NHANES analysis

A. T. Martins
University of Algarve
A. Gomes
University of Algarve
A. Ros
University of Algarve
J. Santos
University of Algarve
J. Rodrigues
University of Algarve
Filipe Ricardo Carvalho
frcarvalho@ualg.pt

University of Algarve

Article

Keywords: Panic attacks, NHANES, Epidemiology, Complex survey design, Risk factors, Health disparities

Posted Date: September 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10434634/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

Panic attacks are common and disabling, yet their socioeconomic, physical and behavioural correlates at the population level remain poorly characterised in young adults. We estimated the national prevalence and correlates of past-year panic attacks in US young adults and assessed whether routinely available markers can predict panic at the individual level. We analysed NHANES 1999–2004 (adults aged 20–39, n = 2,222), in which past-year panic attacks were assessed with the DSM-IV Composite International Diagnostic Interview. Prevalence and associations were estimated accounting for the complex survey design (sampling weights, strata and primary sampling units). Multivariable survey-weighted logistic regression provided adjusted odds ratios (OR) with 95% confidence intervals (CI). Individual-level prediction was evaluated with leakage-free nested cross-validation, comparing logistic regression with random forest, gradient boosting and support vector machines, and assessing calibration. Past-year panic attacks affected 14.1% (95% CI 12.2–16.1) of US young adults, and were more common in women than men (17.8% vs 10.5%) and in low-income individuals (18.6%). In adjusted models, panic attacks were independently associated with low back pain (OR 1.42, 95% CI 1.05–1.91), lifetime marijuana use (OR 1.57, 1.04–2.37), cigarette smoking (OR 1.28, 1.00-1.64) and poorer self-rated health (OR 1.31 per SD, 1.08–1.60). At the individual level, discrimination was limited (ROC-AUC approximately 0.60–0.64) and no machine-learning model outperformed logistic regression. Panic attacks in US young adults are common and socially patterned, and are associated with pain comorbidity, substance use and poorer self-rated health. However, these routinely available markers have limited ability to identify affected individuals, arguing against their use as a standalone screening tool and underscoring the gap between population-level associations and individual-level prediction.

INTRODUCTION

Panic attacks—abrupt surges of intense fear accompanied by somatic and cognitive symptoms such as palpitations, shortness of breath, dizziness, trembling, and fear of losing control or dying—are among the most common anxiety-related experiences in the general population (American Psychiatric Association, 2013; Bandelow and Michaelis, 2015). Whereas panic disorder, defined by recurrent unexpected attacks together with persistent concern or maladaptive behavioural change, affects a minority of adults, isolated and recurrent panic attacks are considerably more prevalent and cut across diagnostic boundaries (Kessler et al., 2006; de Jonge et al., 2016). Cross-national evidence indicates that panic attacks affect a substantial share of the population and predict elevated risk of subsequent psychiatric disorders, underscoring their public-health relevance (de Jonge et al., 2016; Javaid et al., 2023).

Panic is associated with marked functional impairment, reduced quality of life, and high health-service use (Strine et al., 2008). It frequently co-occurs with chronic physical conditions—including cardiovascular disease, respiratory illness, obesity, and chronic pain—and this comorbidity compounds disability and complicates recognition, because the somatic symptoms of panic overlap with those of medical illness (Gariépy et al., 2010; Celano et al., 2016; Costello et al., 2019; Meuret et al., 2020).

Contemporary models emphasise the catastrophic misinterpretation of bodily sensations and heightened anxiety sensitivity, whereby benign somatic cues are appraised as signs of imminent catastrophe, fuelling anticipatory anxiety and avoidance (Clark, 1986; Reiss et al., 1986; Schmidt et al., 1997; Boswell et al., 2013). These cognitive processes are paralleled by neurobiological accounts implicating autonomic dysregulation and altered fear-circuit function (Gorman et al., 2000; Fonzo and Etkin, 2017; Thayer et al., 2012; LeDoux and Daw, 2018). Although informative, such mechanisms are not readily captured by the routinely collected markers available in population surveys and are considered here only as background.

Beyond individual-level mechanisms, mental health is powerfully shaped by the social determinants of health—the conditions in which people are born, grow, live, and work (Marmot et al., 2008). Socioeconomic disadvantage, financial strain, and limited access to care are consistently associated with a higher burden of anxiety, and adverse experiences across the life course, including childhood adversity, increase vulnerability to anxiety and panic in adulthood (Strine et al., 2008; Assari and Bazargan, 2019; Juruena et al., 2020; Kessler et al., 2010). Situating panic within this framework is central to an equity-oriented account of population mental health (Vos et al., 2020).

Anxiety disorders, including panic, are consistently more prevalent and more disabling in women than in men, a disparity attributed to interacting biological, psychological, and social factors (McLean et al., 2011; Moreno-Peral et al., 2014). Characterising such disparities at the population level is a prerequisite for equitable prevention.

The growing availability of large, richly phenotyped population datasets has spurred interest in machine-learning approaches to mental health, both for deriving multidimensional indicators and for predicting outcomes at the individual level (Dwyer et al., 2018; Dadi et al., 2021; Chekroud et al., 2021; Olender et al., 2023; Zaboski et al., 2024). Yet enthusiasm for prediction has often outpaced attention to whether population-level associations translate into accurate individual-level classification, and physiological reactivity, for instance, is only modestly associated with—and does not reliably predict—panic (Robinaugh et al., 2019; Chekroud et al., 2021).

A prominent strand of research locates vulnerability to panic in interoception—the perception of internal bodily states—and in interoceptive conditioning, whereby internal cues become learned signals of threat (De Cort et al., 2017; Van Diest, 2019). On this account, heightened sensitivity to, and catastrophic appraisal of, bodily sensations can convert ordinary physiological fluctuations into full-blown attacks, integrating the cognitive and autonomic features of panic. Familial and genetic factors further contribute to susceptibility, with meta-analytic evidence indicating moderate heritability of anxiety disorders (Hettema et al., 2001); such mechanisms, however, are not directly captured by the routine markers available in national health surveys.

At the population level, panic and its correlates have most often been examined through descriptive epidemiology and clinical cohorts, whereas recent work has begun to apply machine learning to psychiatric prediction more broadly (Dwyer et al., 2018; Rony et al., 2025). Such predictive efforts are

seldom combined with the design-based methods required for valid inference from complex national surveys, and the social patterning of panic attacks—as distinct from their clinical mechanisms—has received comparatively little quantitative attention, despite evidence that adverse life events and limited social support cluster in the period preceding onset (Batinić et al., 2009).

Despite extensive clinical characterisation of panic, its social and behavioural patterning in the general young-adult population, and the extent to which routinely available markers can identify affected individuals, remain insufficiently examined using nationally representative data and appropriate survey methods. Using the National Health and Nutrition Examination Survey (NHANES 1999–2004), we therefore (i) estimate the population prevalence of past-year panic attacks in US young adults and its variation by sex and socioeconomic position; (ii) identify the socioeconomic, physical, and behavioural factors independently associated with panic attacks using survey-weighted models; and (iii) assess, as a secondary aim, whether these routinely available markers can predict panic attacks at the individual level. Throughout, we distinguish population-level association from individual-level prediction and consider the implications for equity-oriented prevention.

MATERIAL AND METHODS

Study design and analytic sample

We analysed the National Health and Nutrition Examination Survey (NHANES), a continuous, nationally representative survey of the US civilian non-institutionalised population conducted by the National Center for Health Statistics using a complex, multistage, stratified probability design. We pooled the three cycles (1999–2000, 2001–2002 and 2003–2004) in which a structured diagnostic assessment of panic was administered, as part of the Composite International Diagnostic Interview (CIDI) mental-health subsample offered to adults aged 20–39 years. The analytic dataset was reconstructed from the original public-use component files (demographics, examination, questionnaire and dietary) rather than from any pre-compiled extract, and records were linked at the participant level using the unique respondent sequence number (SEQN), producing exactly one row per participant. After restriction to respondents aged 20–39 with a valid panic assessment, the final analytic sample comprised 2,222 participants; the construction is summarised in the participant-flow diagram (Fig. 4).

Outcome ascertainment

Outcome ascertainment

The primary outcome was a past-year panic attack, ascertained by the DSM-IV CIDI administered in NHANES 1999–2004 (Kessler and Üstün, 2004): respondents were classified as cases if they reported an attack of fear or panic in the preceding 12 months. Isolated and recurrent panic attacks are both more common and more heterogeneous than the full disorder, and this definition captures a clinically meaningful, public-health-relevant phenotype while preserving statistical power. As a sensitivity analysis,

all models were repeated using the stricter CIDI diagnosis of DSM-IV panic disorder, defined by recurrent unexpected attacks together with persistent concern or maladaptive behavioural change.

Candidate predictors

We considered socioeconomic, physiological, behavioural and general-health markers that are routinely available in population and primary-care settings and that do not require specialised psychiatric assessment: diastolic blood pressure, pulse rate, body mass index, the poverty-income ratio, low back pain, family history of hypertension or stroke, cardiovascular fitness examination status, regular alcohol use, time since the last blood-pressure measurement, cigarette smoking (having smoked at least 100 cigarettes in life), lifetime cannabis use, self-rated general health, physical-activity level, asthma, arthritis, headache/migraine, neck pain and dietary caffeine intake. Variables from the concurrent CIDI depression and generalised-anxiety modules were deliberately excluded as predictors: because they are ascertained in the same interview as the outcome, their inclusion would introduce circularity and spuriously inflate apparent performance. NHANES 'refused' and 'don't know' codes, together with a small number of non-physiological placeholder values identified during data cleaning, were recoded as missing before analysis.

Preprocessing and encoding

Continuous variables were standardised to zero mean and unit variance; binary items were coded 0/1; the nominal cardiovascular-examination-status variable was one-hot encoded, thereby avoiding the artificial ordinality that integer (label) encoding imposes on nominal categories; and the ordinal 'time since last blood-pressure check' was retained on its ordinal scale. Missing values were imputed within cross-validation folds only—using the median for continuous and ordinal variables and the mode for categorical variables—so that no information from held-out observations could influence model development or evaluation.

Complex-survey association analysis (primary)

The primary analyses accounted for the NHANES complex survey design. Because panic was assessed only in the CIDI subsample, we used the corresponding CIDI subsample examination weights together with the design strata and primary sampling units; a combined six-year weight was constructed for the pooled cycles following NHANES analytic guidance (Johnson et al., 2013). The population prevalence of past-year panic attacks, overall and within demographic domains (sex, income, education and race/ethnicity), was estimated by Taylor-series linearisation, which yields design-based 95% confidence intervals that appropriately reflect weighting, stratification and clustering. Associations between each candidate marker and past-year panic attacks were estimated by multivariable survey-weighted logistic regression, reporting adjusted odds ratios (OR) with 95% confidence intervals and design-consistent, cluster-robust variance. We compared weighted with unweighted estimates and fitted an additional

model excluding the poverty-income ratio, both as sensitivity analyses. Bivariate differences between participants with and without a past-year panic attack were examined using independent-samples t-tests for continuous variables and chi-square tests for categorical variables.

Individual-level prediction (secondary)

As a secondary aim we assessed whether the markers could identify affected individuals. All modelling was implemented as a single pipeline in which imputation, encoding, standardisation, feature handling, hyperparameter tuning and decision-threshold selection were performed strictly within training folds; held-out observations were used only once, for final evaluation. Discrimination was estimated by leakage-free nested cross-validation (five outer by five inner folds), reporting the area under the receiver-operating-characteristic curve (ROC-AUC) and the area under the precision–recall curve (PR-AUC) with their variability across folds. Threshold-dependent operating characteristics—sensitivity, specificity, positive and negative predictive value and balanced accuracy—were derived from out-of-fold predictions, and 95% confidence intervals were obtained by bootstrap resampling; the decision threshold was selected on training out-of-fold predictions to favour sensitivity, never on the evaluation data. Under identical conditions we compared logistic regression with random forest, gradient boosting (XGBoost) and a radial-basis-function support-vector machine. We further assessed temporal transportability by training on the 1999–2002 cycles and testing on 2003–2004; feature-selection stability across bootstrap resamples using the Nogueira index; multicollinearity using variance inflation factors (VIF); and calibration using the Brier score, calibration curves, the Hosmer–Lemeshow test and the calibration slope and intercept, before and after Platt and isotonic recalibration. Finally, to translate discrimination into interpretable clinical terms, we estimated the expected number of true- and false-positive results per 10,000 individuals screened at the observed sample prevalence.

Software and reproducibility

Analyses were performed in Python using scikit-learn and statsmodels, with survey estimation implemented through Taylor-series linearisation. No artificial class balancing (over- or under-sampling) was applied at any stage; the natural outcome frequency was preserved in both development and evaluation. All estimates are reported with 95% confidence intervals, and the complete data-construction and analysis code is openly available to permit full reproduction.

Data Availability Statement

All data underlying this study are publicly available. The raw data are from the National Health and Nutrition Examination Survey (NHANES), cycles 1999–2000, 2001–2002 and 2003–2004, freely distributed by the US Centers for Disease Control and Prevention, National Center for Health Statistics (https://www.cdc.gov/nchs/nhanes/). The derived analytic dataset (data_final.csv), together with the complete code that reconstructs it from the original NHANES source files and reproduces every analysis,

table and figure, is openly available at https://github.com/frpcarvalho/panicprediction. No restrictions apply: NHANES public-use files are de-identified US government data in the public domain.

RESULTS

Sample characteristics and prevalence

The analytic sample comprised 2,222 US adults aged 20–39 years (45.9% male), of whom 289 (13.0% unweighted) reported a past-year panic attack; the proportion was stable across cycles (11.9%, 13.1% and 14.0% in 1999–2000, 2001–2002 and 2003–2004, respectively). Accounting for the complex survey design, the weighted population prevalence of past-year panic attacks was 14.1% (95% CI 12.2–16.1), confirming that recent panic attacks are common in this age group.

Prevalence was strongly patterned by sex and socioeconomic position (Fig. 1). It was substantially higher in women than in men (17.8% vs 10.5%) and followed a clear income gradient, being highest among low-income participants (18.6%) and lower in the middle- and high-income groups (13.0% and 12.8%, respectively). Variation by educational attainment was more modest and non-monotonic (from 10.9% among those with less than a 9th-grade education to 16.8% among those with 9–11 years of schooling), and prevalence by race/ethnicity ranged from 10.5% among other Hispanic participants to 14.9% among non-Hispanic White participants.

Bivariate characteristics

Participants with a past-year panic attack differed from those without on several characteristics (Table 1). They were more likely to be female (65.7% vs 54.4%, $p < 0.001$) and had a lower mean poverty-income ratio (2.30 vs 2.52, $p = 0.041$). They also reported markedly more low back pain (50.2% vs 36.5%, $p < 0.001$), regular alcohol use (19.4% vs 11.5%, $p = 0.001$), cigarette smoking (52.9% vs 39.0%, $p < 0.001$) and lifetime cannabis use (29.4% vs 17.5%, $p < 0.001$), and poorer self-rated health (mean 2.70 vs 2.27 on a 1–5 scale, $p < 0.001$). By contrast, age, body mass index, diastolic blood pressure and pulse rate did not differ significantly between the two groups, indicating that the crude physiological markers favoured in earlier work carried little discriminating information.

Factors independently associated with panic attacks

In the multivariable survey-weighted model (Table 2, Fig. 2), past-year panic attacks were independently associated with low back pain (OR 1.42, 95% CI 1.05–1.91), lifetime cannabis use (OR 1.57, 1.04–2.37), poorer self-rated health (OR 1.31 per standard deviation, 1.08–1.60) and cigarette smoking (OR 1.28, 1.00–1.64); headache/migraine showed a borderline association (OR 1.35, 0.97–1.88). Diastolic blood pressure, pulse rate, body mass index, the poverty-income ratio, family history of hypertension or stroke, regular alcohol use, asthma, arthritis, neck pain, physical activity and caffeine intake were not independently associated after adjustment.

These associations were robust to analytical choices. Weighted and unweighted odds ratios were closely concordant, indicating that the findings were not driven by the survey weighting, and the model excluding the poverty-income ratio left both the associations and the predictive discrimination essentially unchanged (change in ROC-AUC < 0.01), arguing against undue dependence on the socioeconomic variable and mitigating the ethical concerns that its inclusion might raise. In a sensitivity analysis using the stricter DSM-IV panic disorder diagnosis, the direction of associations was similar, although estimates were less precise given the much smaller number of cases.

Individual-level predictive performance

In contrast to these coherent population-level associations, the ability of the markers to identify affected individuals was limited. Under leakage-free nested cross-validation, logistic regression achieved a ROC-AUC of 0.60 (95% CI 0.57–0.65) with the original nine markers and 0.64 with the extended marker set, and a PR-AUC of only 0.20 (95% CI 0.16–0.25). At a sensitivity-oriented operating point (sensitivity 0.71), specificity was 0.41, positive predictive value 0.15 and negative predictive value 0.90, yielding a balanced accuracy of 0.56. No machine-learning model outperformed logistic regression under identical conditions: ROC-AUC was 0.60 for logistic regression, 0.59 for gradient boosting, 0.58 for random forest and 0.50 for the support-vector machine (Table 3), and none achieved a PR-AUC above 0.21.

Two further diagnostics reinforced the modest and non-robust nature of individual-level prediction. Feature selection was unstable (Nogueira stability index 0.29), meaning that the particular subset of 'important' markers varied substantially from one resample to another, so that any single ranking of predictors should be interpreted cautiously. Variance inflation factors were uniformly close to 1.0 (range 1.02–1.28), indicating negligible multicollinearity and confirming that the weak performance reflected a genuine lack of individual-level signal rather than redundancy among predictors. Exploratory subgroup analysis further suggested that discrimination was not uniform across sex, with a higher apparent ROC-AUC in men than in women (0.75 vs 0.49) and a correspondingly higher missed-case rate in women, although subgroup sizes were small and the associated confidence intervals wide.

Calibration

The uncalibrated, class-weighted model was poorly calibrated and systematically over-predicted risk (Brier score 0.24; Hosmer–Lemeshow $p < 0.001$; calibration slope 0.83, intercept − 1.91). Both Platt scaling and isotonic recalibration substantially improved agreement between predicted and observed probabilities (Brier score 0.11 in each case; Hosmer–Lemeshow $p = 0.10$ and $p = 0.47$, respectively; Fig. 3). This indicates that any probabilistic use of such a model would require explicit recalibration, and that headline discrimination metrics alone would give a misleading impression of readiness for use.

Temporal validation and expected screening burden

Discrimination degraded further under temporal validation: a model trained on the 1999–2002 cycles and tested on 2003–2004 achieved a ROC-AUC of only 0.57, indicating limited transportability even across adjacent survey cycles of the same programme. The practical implications of this level of

performance are made concrete by the expected screening burden. Applied at the observed sample prevalence (13.0%), screening 10,000 young adults would be expected to correctly flag approximately 923 of the 1,301 individuals with a past-year panic attack, while generating about 5,117 false-positive results—roughly 5.5 false positives for every true case detected, corresponding to a positive predictive value of only about 15%. Such a burden would be difficult to justify operationally and reinforces that these markers, though informative in aggregate, are unsuitable as a standalone individual screening tool.

For transparency, we note that an earlier version of this analysis—affected by duplicated participant records and by selection of the decision threshold on the test set—had reported an apparent balanced accuracy of approximately 82% and a ROC-AUC of 0.82. After reconstruction of the dataset with one record per participant and adoption of leakage-free, survey-appropriate evaluation, these figures fell to those reported above, quantifying the magnitude of the earlier optimistic bias and illustrating how readily such biases can arise in applied predictive modelling.

Table 1

Demographic, anthropometric, socioeconomic and behavioural characteristics of the study sample (N = 2,222; adults aged 20–39, NHANES 1999–2004), stratified by past-year panic attack. Values are mean (SD) or n (%). P values are from t-tests (continuous) or chi-square tests (categorical); estimates are unweighted.

Characteristic	**Total (N = 2222)**	**No panic attack (N = 1933)**	**Past-year panic attack (N = 289)**	**p**
Age, years — mean (SD)	29.2 (5.8)	29.2 (5.7)	29.2 (6.1)	0.83
Female — n (%)	1242 (55.9)	1052 (54.4)	190 (65.7)	< 0.001
Race/ethnicity — n (%)				**0.24**
Mexican-American	566 (25.5)	501 (25.9)	65 (22.5)	
Other Hispanic	116 (5.2)	104 (5.4)	12 (4.2)	
Non-Hispanic White	1003 (45.1)	863 (44.6)	140 (48.4)	
Non-Hispanic Black	440 (19.8)	386 (20.0)	54 (18.7)	
Other	97 (4.4)	79 (4.1)	18 (6.2)	
Education — n (%)				**0.08**
<9th grade	167 (7.5)	151 (7.8)	16 (5.5)	
9–11th grade	394 (17.7)	335 (17.3)	59 (20.4)	
High school/GED	567 (25.5)	478 (24.7)	89 (30.8)	
Some college	626 (28.2)	547 (28.3)	79 (27.3)	
College graduate or above	465 (20.9)	419 (21.7)	46 (15.9)	
Poverty–income ratio — mean (SD)	2.50 (1.61)	2.52 (1.60)	2.30 (1.69)	0.04
Body mass index, kg/m² — mean (SD)	28.0 (6.7)	27.9 (6.7)	28.1 (7.0)	0.73
Diastolic blood pressure, mmHg — mean (SD)	68.1 (12.4)	68.1 (12.3)	68.1 (13.0)	0.99
Pulse, bpm — mean (SD)	74.9 (12.5)	74.7 (12.6)	75.8 (12.3)	0.18
Low back pain — n (%)	851 (38.3)	706 (36.5)	145 (50.2)	< 0.001
Regular alcohol use — n (%)	278 (12.5)	222 (11.5)	56 (19.4)	0.001
Ever smoked ≥ 100 cigarettes — n (%)	906 (40.8)	753 (39.0)	153 (52.9)	< 0.001

Characteristic	Total (N = 2222)	No panic attack (N = 1933)	Past-year panic attack (N = 289)	p
Ever used marijuana — n (%)	423 (19.0)	338 (17.5)	85 (29.4)	< 0.001
Self-rated general health (1–5) — mean (SD)	2.33 (0.99)	2.27 (0.98)	2.70 (0.99)	< 0.001

Table 2

Adjusted associations between candidate markers and past-year panic attacks, from multivariable survey-weighted logistic regression (odds ratios with 95% confidence intervals). Continuous and ordinal predictors are expressed per standard deviation (SD).

Predictor	Adjusted OR	95% CI	p
Time since last BP check (per SD)	0.81	0.72–0.93	0.002
Worse self-rated health (per SD)	1.31	1.08–1.60	0.007
Low back pain	1.42	1.05–1.91	0.022
Ever used marijuana	1.57	1.04–2.37	0.031
Smoked ≥ 100 cigarettes	1.28	1.00–1.64	0.048
Headache/migraine	1.35	0.97–1.88	0.079
Arthritis	1.40	0.86–2.28	0.171
CV exam status = 3	1.22	0.89–1.68	0.213
Physical activity level (per SD)	0.90	0.77–1.06	0.224
Regular alcohol use	1.23	0.79–1.90	0.365
Diastolic blood pressure (per SD)	0.94	0.81–1.09	0.388
Vigorous physical activity	1.12	0.85–1.49	0.420
Poverty–income ratio (per SD)	0.93	0.78–1.11	0.422
Body mass index (per SD)	0.93	0.78–1.11	0.422
Neck pain	1.14	0.81–1.61	0.458
CV exam status = 2	0.81	0.43–1.49	0.495
Pulse rate (per SD)	1.03	0.91–1.18	0.617
Asthma	1.02	0.71–1.44	0.931
Family history of HTN/stroke	0.99	0.74–1.33	0.942
Caffeine intake (per SD)	1.00	0.88–1.14	0.990

Table 3
Individual-level predictive performance of four classifiers under identical conditions (leakage-free nested cross-validation), evaluated on the natural outcome frequency. No machine-learning model outperformed logistic regression. Threshold-dependent metrics (sensitivity, specificity and positive predictive value) reflect each model's own sensitivity-oriented operating point selected within cross-validation and therefore differ slightly from the primary pipeline's operating point reported in the text; ROC-AUC and PR-AUC are threshold-independent.

Model	ROC-AUC	PR-AUC	Balanced acc.	Sensitivity	Specificity	PPV
Logistic regression	0.605	0.194	0.562	0.778	0.347	0.150
Random forest	0.581	0.185	0.505	0.736	0.275	0.131
Gradient boosting (XGBoost)	0.590	0.204	0.482	0.764	0.200	0.124
Support vector machine (RBF)	0.504	0.148	0.498	0.806	0.190	0.129

DISCUSSION

In this nationally representative sample of US young adults, past-year panic attacks were common (weighted prevalence 14.1%) and clearly socially patterned, being more frequent in women and in those of lower socioeconomic position, and were independently associated with pain comorbidity, substance use, and poorer self-rated health. Crucially, however, the same routinely available markers had only limited ability to identify affected individuals, illustrating the well-recognised gap between population-level association and individual-level prediction (Chekroud et al., 2021).

The estimated prevalence is consistent with, though at the upper end of, previous epidemiological estimates of past-year panic attacks (Kessler et al., 2006; de Jonge et al., 2016; Javaid et al., 2023), reflecting our focus on young adults, among whom anxiety-related conditions peak. It reinforces that panic attacks are far more common than panic disorder and constitute a substantial, often under-recognised, component of the population burden of anxiety (Moreno-Peral et al., 2014).

The markedly higher prevalence in women (17.8% vs 10.5%) is consistent with the established female preponderance of anxiety and with evidence that anxiety is not only more common but more disabling in women (McLean et al., 2011). This disparity has implications for equitable screening and prevention, which should avoid systematically under-detecting or over-pathologising particular groups.

The socioeconomic gradient—with the highest prevalence among low-income participants—echoes the social-determinants-of-health framework (Marmot et al., 2008) and evidence linking financial strain, disadvantage, and adverse life-course experiences to anxiety (Strine et al., 2008; Assari and Bazargan, 2019; Kessler et al., 2010; Evans-Lacko et al., 2018). That panic is more common where material and psychosocial resources are scarcer situates it within broader patterns of health inequity rather than as a purely individual clinical phenomenon.

Panic attacks were associated with low back pain and, at borderline significance, with headache/migraine, in keeping with the well-documented co-occurrence of chronic pain and panic/anxiety (Smitherman et al., 2013; Meuret et al., 2020). Shared mechanisms—including autonomic arousal, interoceptive hypersensitivity, and central sensitisation—may underlie this comorbidity, and the overlap of somatic symptoms can complicate diagnosis in both directions.

Associations with cigarette smoking and lifetime cannabis use are consistent with literature linking substance use to panic vulnerability (Breslau et al., 1991; Crippa et al., 2009). Because our design is cross-sectional, these associations cannot establish whether substance use precipitates panic, reflects self-medication, or shares an underlying vulnerability.

Poorer self-rated health was the single strongest correlate of panic attacks. Rather than a specific causal pathway, this most likely reflects a general dimension of distress and somatic burden, consistent with the broad comorbidity of anxiety with physical ill-health and conditions such as obesity (Gariépy et al., 2010; Milaneschi et al., 2019). Self-rated health is therefore best interpreted as a marker of overall burden rather than a modifiable target.

Despite these coherent population-level associations, individual-level discrimination was modest (ROC-AUC ≈ 0.60–0.64), declined further under temporal validation, and was not improved by any machine-learning model relative to logistic regression. This dissociation between robust group-level associations and weak individual prediction is increasingly recognised in psychiatry (Robinaugh et al., 2019; Chekroud et al., 2021) and argues against using these routinely available markers as a standalone screening tool for panic.

Our analysis also carries a methodological lesson. An earlier iteration inadvertently duplicated participant records and selected the decision threshold on the test set, producing markedly inflated performance. After reconstructing the dataset from source with one record per participant, incorporating the complex survey design, and adopting leakage-free nested cross-validation, performance was far more modest. This underscores the importance of participant-level data integrity, design-appropriate analysis, and strict separation of model development from evaluation—prerequisites easily overlooked in applied machine-learning studies (Dwyer et al., 2018).

Strengths of this study include a large, nationally representative sample with a structured DSM-IV diagnostic instrument, appropriate incorporation of the complex survey design, transparent and fully reproducible dataset construction from source, and leakage-free evaluation with calibration and uncertainty quantification.

Several limitations should be acknowledged. The cross-sectional design precludes causal or temporal inference. The NHANES cycles analysed (1999–2004) are the only ones containing the structured panic module, which limits contemporaneity, and the number of cases, while adequate for association analysis, remains modest for individual-level prediction. Several exposures rely on self-report, and the analysis lacks external validation in an independent cohort. Finally, self-rated health and examination-

status variables should be interpreted with caution, as they may capture process or general-burden information rather than specific constructs.

The pattern of associations we observed is coherent with interoceptive and autonomic models of panic. Comorbid pain and poorer self-rated health may heighten the salience of aversive bodily sensations which, through interoceptive conditioning, can precipitate or maintain attacks (De Cort et al., 2017; Van Diest, 2019), while substance use may modulate autonomic arousal and stress reactivity. That such markers are associated at the population level yet contribute little to individual prediction is consistent with panic emerging from the interaction of diffuse cognitive, autonomic and contextual vulnerabilities rather than from any single measurable risk factor.

Our findings temper the optimism of some applied machine-learning studies in psychiatry, which have reported high predictive accuracy but frequently without adequate attention to data leakage, participant-level independence, calibration, or survey design (Chekroud et al., 2021; Rony et al., 2025). When these safeguards are enforced, discrimination for panic attacks from routinely available markers is modest and does not exceed that of a simple logistic model, mirroring a broader reckoning in which impressive cross-validated performance often fails to survive rigorous, leakage-free evaluation or to transport to new settings.

The social patterning we describe likely reflects cumulative exposure to stressors and constrained material and psychosocial resources across the life course, including adverse life events and limited social support in the period preceding onset (Batinić et al., 2009; Kessler et al., 2010). Situating panic within this life-course and social-determinants framework reframes it, in part, as a marker of population-level disadvantage rather than solely an individual pathology, and points to the upstream conditions where preventive effort may be most effective.

For clinical practice and policy, these results counsel caution about deploying opportunistic risk-screening for panic based on routine physiological or socioeconomic data, given the substantial false-positive burden and the attendant risk of stigmatising already-disadvantaged groups. They instead support universal, low-threshold access to anxiety assessment and care, together with population-level strategies that address the social conditions associated with panic; any future individual-level screening application should be contemplated only after external validation in contemporary, independent cohorts and demonstration of adequate calibration and net clinical benefit.

Finally, the high population prevalence we document, combined with the well-documented impairment that panic and agoraphobic avoidance impose on quality of life and functioning (Preti et al., 2021), underscores the public-health importance of panic attacks even where individual prediction is weak: interventions that modestly reduce the aggregate burden across the population may deliver meaningful benefit precisely because the condition is common and socially concentrated.

In conclusion, past-year panic attacks are common and socially patterned among US young adults, and are associated with pain comorbidity, substance use, and poorer self-rated health, yet routinely available

markers have limited value for identifying affected individuals. From a public-health perspective, the pronounced patterning of panic by sex and socioeconomic position supports population-level, equity-oriented approaches to anxiety prevention rather than reliance on individual risk screening; future work should prioritise longitudinal designs and richer psychological measures over prediction based on these markers.

Declarations

Ethics approval

This study is a secondary analysis of publicly available, de-identified data from the National Health and Nutrition Examination Survey (NHANES), administered by the National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention, United States. The original NHANES data collection was reviewed and approved by the NCHS Institutional Review Board (Protocol #98-12, 1999–2004) in accordance with the Declaration of Helsinki. All participants in the original surveys provided written informed consent. As this study involved no primary data collection, no contact with human participants, and no possibility of re-identification, no additional institutional ethics approval was required.

Consent to participate

Not applicable. This study is a secondary analysis of fully de-identified, publicly available data. No direct contact with human participants took place.

Consent to publish

Not applicable.

Competing interests

The authors have no competing interests to declare that are relevant to the content of this article.

Funding

This work was supported by the project AI.INSIGHTS — AI-Powered Real-Time Anonymous User Echo Monitoring (ALGARVE-FEDER 02964500, ref. 24298), co-financed by ALGARVE 2030, Portugal 2030, and the European Union.

AI.INSIGHTS, developed by SPIC and the University of Algarve, is a privacy-first Edge AI platform that uses multi-sensor fusion (cameras, LiDAR, radar, UWB, microphone arrays) to anonymously analyse visitor behaviour and satisfaction in event stands and small retail stores. It delivers real-time metrics such as engagement, sentiment, dwell time and crowd dynamics, without storing personal data, through dashboards, predictive analytics and AR overlays. The project will produce three commercial offerings, fully aligned with the GDPR and the EU AI Act, and targets the global market to improve return on investment and operational efficiency.

Author contributions

Martins, A.T.: Conceptualization, Investigation, Writing – Original Draft. Gomes, A.: Methodology, Data Curation, Software. Ros, A.: Formal Analysis, Visualization. Santos, J.: Investigation, Writing – Review & Editing. Rodrigues, J.: Software. Carvalho, F.R.: Conceptualization, Methodology, Supervision, Writing – Review & Editing, Funding Acquisition.

References

1. American Psychiatric Association (2013) Diagnostic and statistical manual of mental disorders, 5th edn. American Psychiatric Publishing, Washington, DC. https://doi.org/10.1176/appi.books.9780890425596
2. Assari S, Bazargan M (2019) Unequal associations between educational attainment and occupational stress across racial and ethnic groups. Int J Environ Res Public Health 16:3539. https://doi.org/10.3390/ijerph16193539
3. Bandelow B, Michaelis S (2015) Epidemiology of anxiety disorders in the 21st century. Dialogues Clin Neurosci 17:327–335. https://doi.org/10.31887/DCNS.2015.17.3/bbandelow
4. Batinić B, Trajković G, Duisin D et al (2009) Life events and social support in a 1-year preceding panic disorder. Psychiatr Danub 21:33–40
5. Boswell JF, Thompson-Hollands J, Farchione TJ, Barlow DH (2013) Intolerance of uncertainty: a common factor in the treatment of emotional disorders. J Clin Psychol 69:630–645. https://doi.org/10.1002/jclp.21965
6. Breslau N, Kilbey MM, Andreski P (1991) Nicotine dependence, major depression, and anxiety in young adults. Arch Gen Psychiatry 48:1069–1074. https://doi.org/10.1001/archpsyc.1991.01810360045007
7. Celano CM, Daunis DJ, Lokko HN, Campbell KA, Huffman JC (2016) Anxiety disorders and cardiovascular disease. Curr Psychiatry Rep 18:101. https://doi.org/10.1007/s11920-016-0739-5
8. Chekroud AM, Bondar J, Delgadillo J et al (2021) The promise of machine learning in predicting treatment outcomes in psychiatry. World Psychiatry 20:154–170. https://doi.org/10.1002/wps.20882
9. Clark DM (1986) A cognitive approach to panic. Behav Res Ther 24:461–470. https://doi.org/10.1016/0005-7967(86)90011-2
10. Costello H, Gould RL, Abrol E, Howard R (2019) Systematic review and meta-analysis of the association between peripheral inflammatory cytokines and generalised anxiety disorder. BMJ Open 9:e027925. https://doi.org/10.1136/bmjopen-2018-027925
11. Crippa JA, Zuardi AW, Martín-Santos R et al (2009) Cannabis and anxiety: a critical review of the evidence. Hum Psychopharmacol 24:515–523. https://doi.org/10.1002/hup.1048
12. Dadi K, Varoquaux G, Houenou J et al (2021) Population modeling with machine learning can enhance measures of mental health. GigaScience 10:giab071.

https://doi.org/10.1093/gigascience/giab071

13. De Cort K, Schroijen M, Hurlemann R et al (2017) Modeling the development of panic disorder with interoceptive conditioning. Eur Neuropsychopharmacol 27:59–69. https://doi.org/10.1016/j.euroneuro.2016.11.001
14. de Jonge P, Roest AM, Lim CCW et al (2016) Cross-national epidemiology of panic disorder and panic attacks in the World Mental Health Surveys. Depress Anxiety 33:1155–1177. https://doi.org/10.1002/da.22572
15. Dwyer DB, Falkai P, Koutsouleris N (2018) Machine learning approaches for clinical psychology and psychiatry. Annu Rev Clin Psychol 14:91–118. https://doi.org/10.1146/annurev-clinpsy-032816-045037
16. Evans-Lacko S, Aguilar-Gaxiola S, Al-Hamzawi A et al (2018) Socio-economic variations in the mental health treatment gap for people with anxiety, mood, and substance use disorders: results from the WHO World Mental Health (WMH) surveys. Psychol Med 48:1560–1571. https://doi.org/10.1017/S0033291717003336
17. Fonzo GA, Etkin A (2017) Affective neuroimaging in generalized anxiety disorder: an integrated review. Dialogues Clin Neurosci 19:169–179. https://doi.org/10.31887/DCNS.2017.19.2/gfonzo
18. Gariépy G, Nitka D, Schmitz N (2010) The association between obesity and anxiety disorders in the population: a systematic review and meta-analysis. Int J Obes 34:407–419. https://doi.org/10.1038/ijo.2009.252
19. Gorman JM, Kent JM, Sullivan GM, Coplan JD (2000) Neuroanatomical hypothesis of panic disorder, revised. Am J Psychiatry 157:493–505. https://doi.org/10.1176/appi.ajp.157.4.493
20. Hettema JM, Neale MC, Kendler KS (2001) A review and meta-analysis of the genetic epidemiology of anxiety disorders. Am J Psychiatry 158:1568–1578. https://doi.org/10.1176/appi.ajp.158.10.1568
21. Javaid SF, Hashim IJ, Hashim MJ et al (2023) Epidemiology of anxiety disorders: global burden and sociodemographic associations. Middle East Curr Psychiatry 30:44. https://doi.org/10.1186/s43045-023-00315-3
22. Johnson CL, Paulose-Ram R, Ogden CL et al (2013) National Health and Nutrition Examination Survey: analytic guidelines, 1999–2010. Vital Health Stat 2(161):1–24
23. Juruena MF, Eror F, Cleare AJ, Young AH (2020) The role of early life stress in HPA axis and anxiety. In: Kim YK (ed) Anxiety disorders, vol 1191. Springer, Singapore, pp 141–153. https://doi.org/10.1007/978-981-32-9705-0_9
24. Kessler RC, Üstün TB (2004) The World Mental Health (WMH) Survey Initiative version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). Int J Methods Psychiatr Res 13:93–121. https://doi.org/10.1002/mpr.168
25. Kessler RC, Chiu WT, Jin R et al (2006) The epidemiology of panic attacks, panic disorder, and agoraphobia in the National Comorbidity Survey Replication. Arch Gen Psychiatry 63:415–424. https://doi.org/10.1001/archpsyc.63.4.415

26. Kessler RC, McLaughlin KA, Green JG et al (2010) Childhood adversities and adult psychopathology in the WHO World Mental Health Surveys. Br J Psychiatry 197:378–385. https://doi.org/10.1192/bjp.bp.110.080499
27. LeDoux J, Daw ND (2018) Surviving threats: neural circuit and computational implications of a new taxonomy of defensive behaviour. Nat Rev Neurosci 19:269–282. https://doi.org/10.1038/nrn.2018.22
28. Marmot M, Friel S, Bell R et al (2008) Closing the gap in a generation: health equity through action on the social determinants of health. Lancet 372:1661–1669. https://doi.org/10.1016/S0140-6736(08)61690-6
29. McLean CP, Asnaani A, Litz BT, Hofmann SG (2011) Gender differences in anxiety disorders: prevalence, course of illness, comorbidity and burden of illness. J Psychiatr Res 45:1027–1035. https://doi.org/10.1016/j.jpsychires.2011.03.006
30. Meuret AE, Tunnell N, Roque A (2020) Anxiety disorders and medical comorbidity: treatment implications. Adv Exp Med Biol 1191:237–261. https://doi.org/10.1007/978-981-32-9705-0_15
31. Milaneschi Y, Simmons WK, van Rossum EFC, Penninx BW (2019) Depression and obesity: evidence of shared biological mechanisms. Mol Psychiatry 24:18–33. https://doi.org/10.1038/s41380-018-0017-5
32. Moreno-Peral P, Conejo-Cerón S, Motrico E et al (2014) Risk factors for the onset of panic and generalised anxiety disorders in the general adult population: a systematic review of cohort studies. J Affect Disord 168:337–348. https://doi.org/10.1016/j.jad.2014.07.003
33. Olender RT, Roy S, Nishtala PS (2023) Application of machine learning approaches in predicting clinical outcomes in older adults – a systematic review and meta-analysis. BMC Geriatr 23:561. https://doi.org/10.1186/s12877-023-04246-w
34. Preti A, Piras M, Cossu G et al (2021) The burden of agoraphobia in worsening quality of life in a community survey in Italy. Psychiatry Investig 18:277–283. https://doi.org/10.30773/pi.2020.0342
35. Reiss S, Peterson RA, Gursky DM, McNally RJ (1986) Anxiety sensitivity, anxiety frequency and the prediction of fearfulness. Behav Res Ther 24:1–8. https://doi.org/10.1016/0005-7967(86)90143-9
36. Robinaugh DJ, Millner AJ, McNally RJ (2019) Assessing vulnerability to panic: a systematic review of psychological and physiological responses to biological challenges as prospective predictors of panic attacks and panic disorder. Gen Psychiatry 32:e100202. https://doi.org/10.1136/gpsych-2019-100202
37. Rony MKK, Das DC, Khatun MT et al (2025) Artificial intelligence in psychiatry: a systematic review and meta-analysis of diagnostic and therapeutic efficacy. Digit Health 11:20552076251330528. https://doi.org/10.1177/20552076251330528
38. Schmidt NB, Lerew DR, Jackson RJ (1997) The role of anxiety sensitivity in the pathogenesis of panic: prospective evaluation of spontaneous panic attacks during acute stress. J Abnorm Psychol 106:355–364. https://doi.org/10.1037/0021-843X.106.3.355

39. Smitherman TA, Kolivas ED, Bailey JR (2013) Panic disorder and migraine: comorbidity, mechanisms, and clinical implications. Headache 53:23–45. https://doi.org/10.1111/head.12004
40. Strine TW, Chapman DP, Balluz L et al (2008) Health-related quality of life and health behaviors by social and emotional support. Soc Psychiatry Psychiatr Epidemiol 43:151–159. https://doi.org/10.1007/s00127-007-0277-x
41. Thayer JF, Åhs F, Fredrikson M et al (2012) A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health. Neurosci Biobehav Rev 36:747–756. https://doi.org/10.1016/j.neubiorev.2011.11.009
42. Van Diest I (2019) Interoception, conditioning, and fear: the panic threesome. Psychophysiology 56:e13421. https://doi.org/10.1111/psyp.13421
43. Vos T, Lim SS, Abbafati C et al (2020) Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 396:1204–1222. https://doi.org/10.1016/S0140-6736(20)30925-9
44. Zaboski BA, Wilens A, McNamara JPH, Muller GN (2024) Predicting OCD severity from religiosity and personality: a machine learning and neural network approach. J Mood Anxiety Disord 8:100089. https://doi.org/10.1016/j.xjmad.2024.100089

Figures

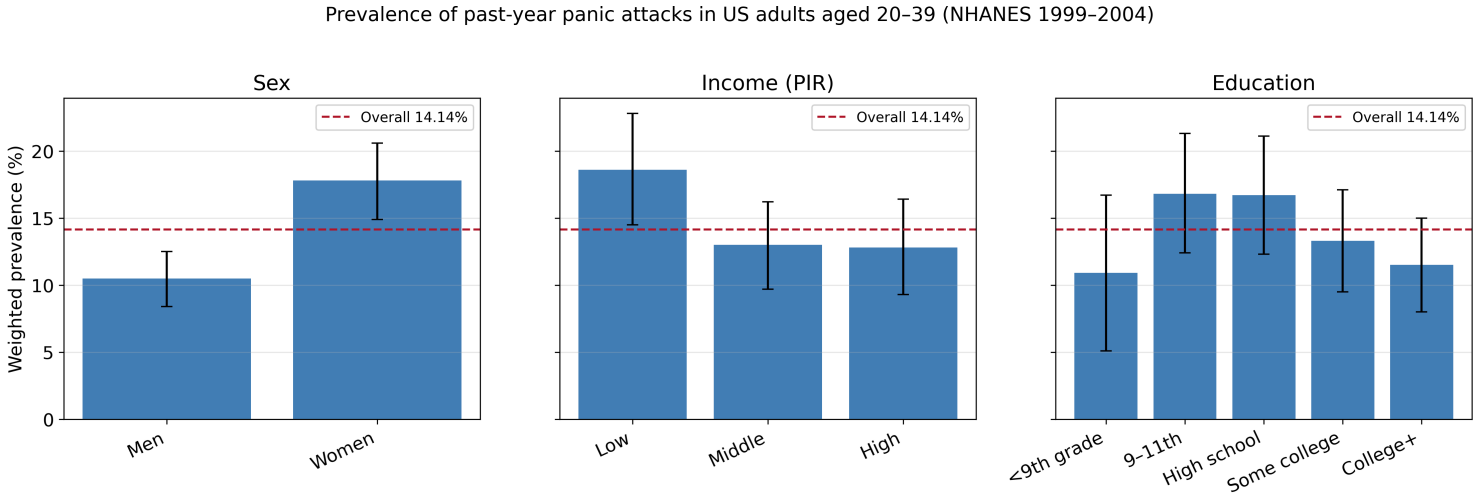


Figure 1

Weighted prevalence of past-year panic attacks in US adults aged 20-39 (NHANES 1999-2004), overall and by sex, income (poverty-income ratio) and education. Error bars are 95% confidence intervals; the dashed line is the overall weighted prevalence (14.1%).

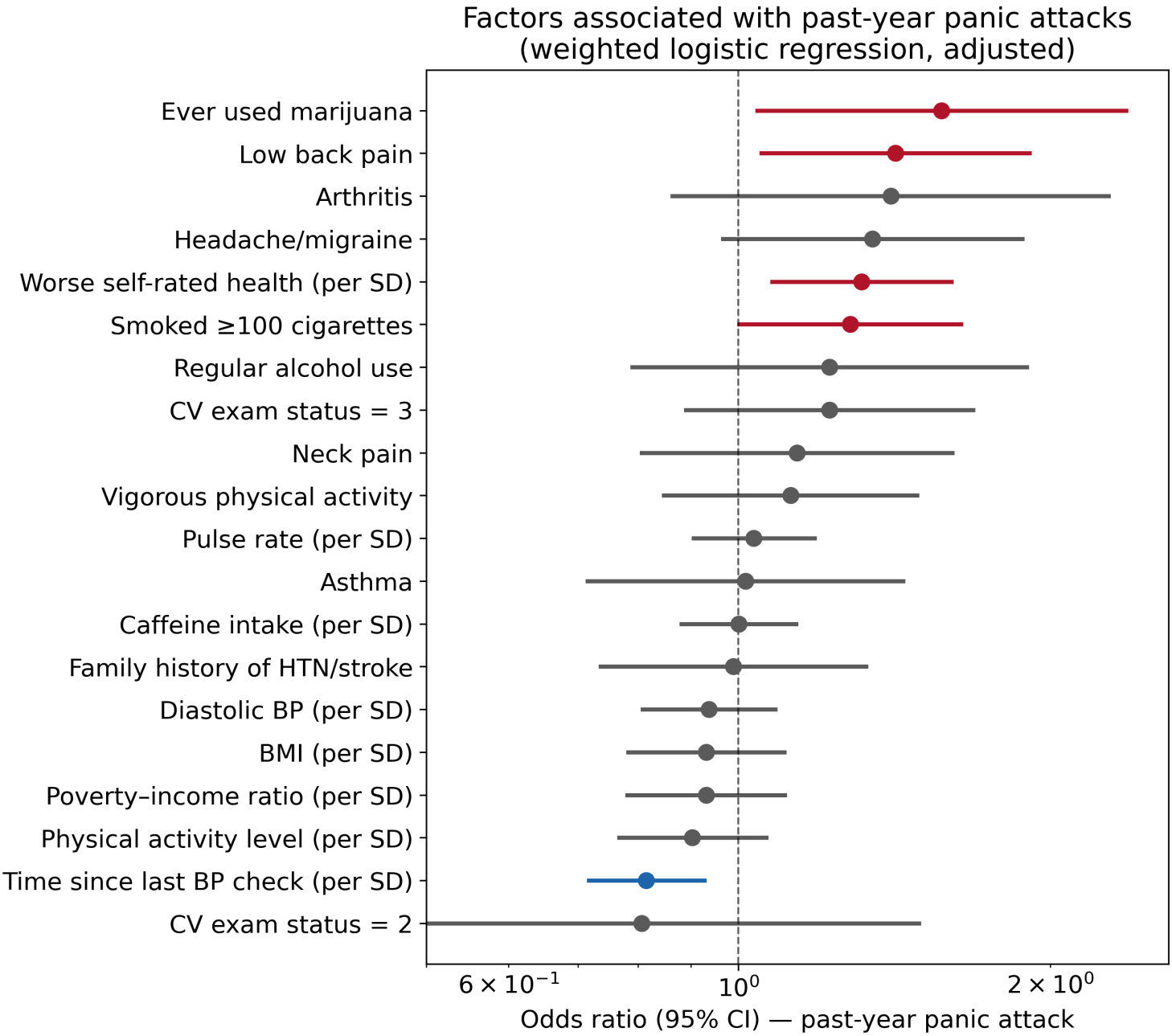


Figure 2

Adjusted odds ratios (95% CI) for the association between candidate markers and past-year panic attacks, from multivariable survey-weighted logistic regression. Estimates to the right of 1.0 indicate higher odds; red denotes statistically significant associations.

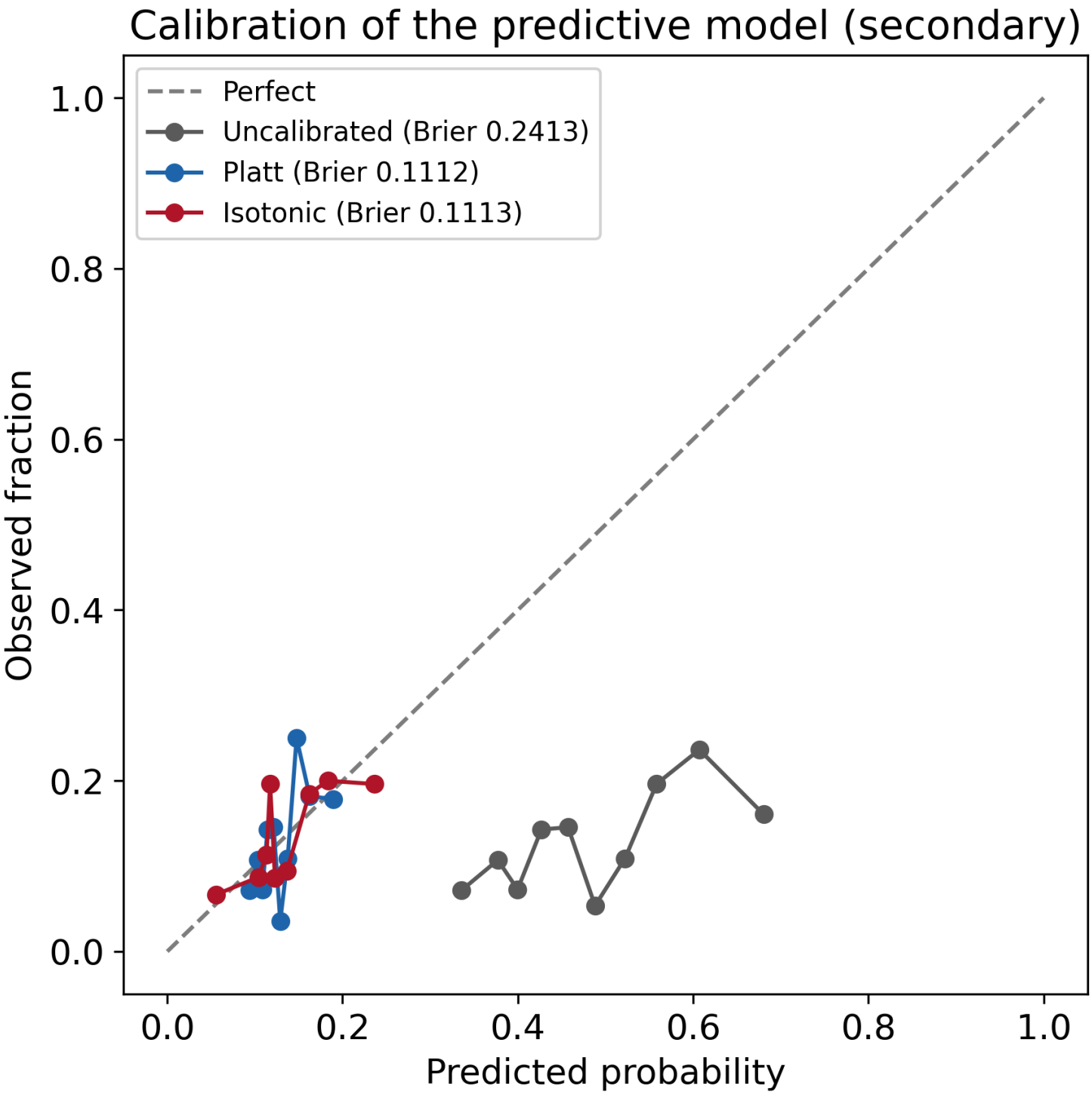
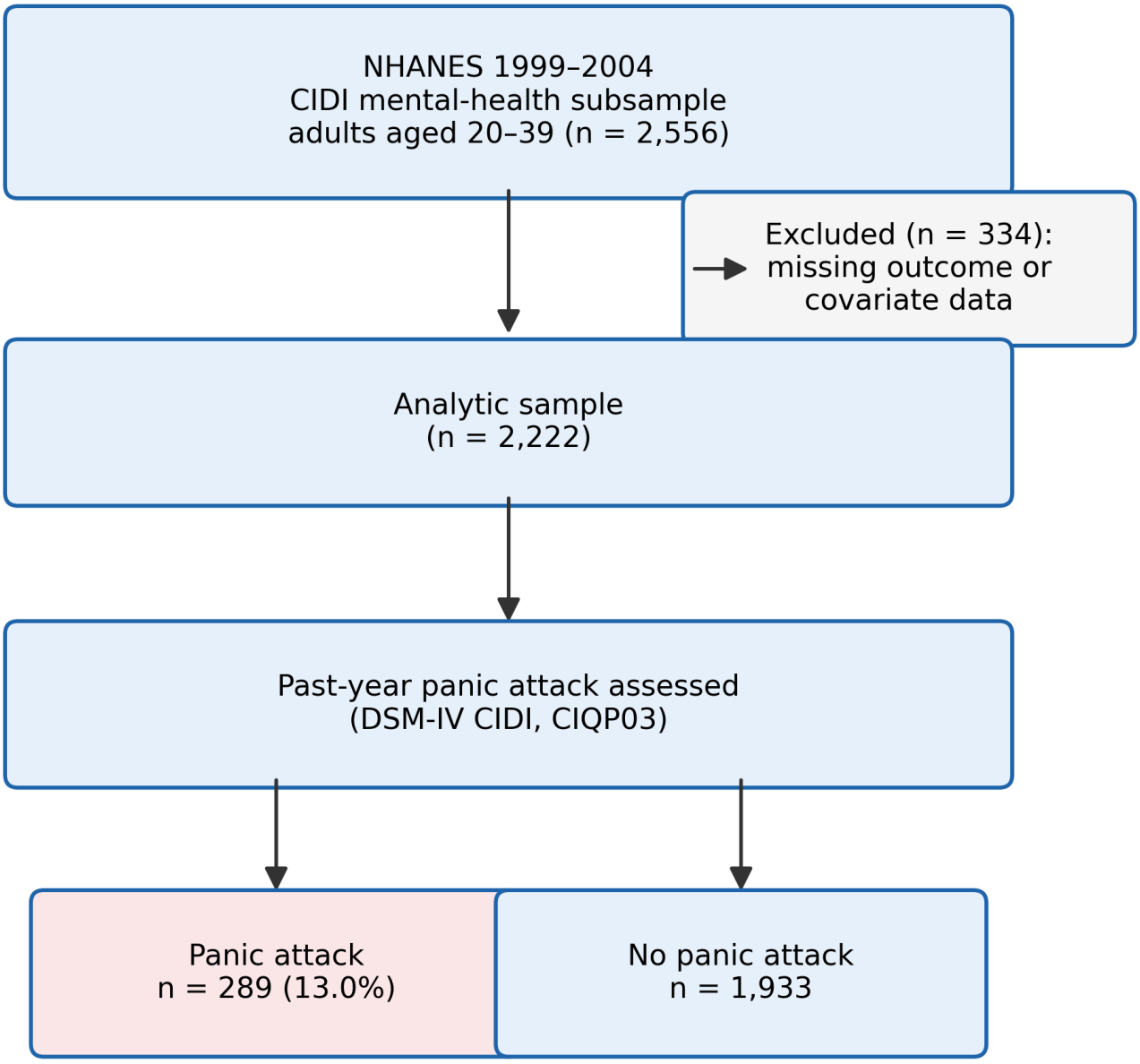


Figure 3

Calibration of the predictive model. The uncalibrated model was poorly calibrated; Platt and isotonic recalibration substantially improved agreement between predicted and observed probabilities (Brier score 0.11).



Weighted population prevalence: 14.1% (95% CI 12.2–16.1)

Figure 4

Participant flow and construction of the analytic sample, from the NHANES 1999-2004 CIDI mental-health subsample (adults aged 20-39) to the final analytic sample (n=2,222; 289 with and 1,933 without a past-year panic attack).

Supplementary Files

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

- SupplementaryMaterialHSSC.docx