

Heartbeat counting task performance in alcohol use disorder: a systematic review and meta-analysis

Nicolás Guillermo González ^{1,*}, Luisa Göde ², Lucía Alba-Ferrara ^{3,4}

¹Faculty of Psychology, National University of Tucumán, Av. Benjamín Aráoz 800, 4000 San Miguel de Tucumán, Tucumán, Argentina

²Wilhelm Wundt Institute for Psychology, Faculty of Life Sciences, Leipzig University, Neumarkt 9, 04109 Leipzig, Germany

³INCYT (Instituto de Neurociencia Cognitiva y Traslacional), CONICET–Fundación Favaloro–Fundación INECO, Marcelo T. de Alvear 1632, C1060AAF Buenos Aires, Argentina

⁴Faculty of Psychology and Human Relations, Center for Advanced Studies in Human Development and Psychology (CAEDHPSI), Universidad Abierta Interamericana, San Juan 960, C1147AAU Buenos Aires, Argentina

*Corresponding author. E-mail: nicolas.gonzalez@psicologia.unt.edu.ar

Abstract

Aims: Interoception has been proposed as a relevant process in alcohol use disorder (AUD), and studies using the heartbeat counting task (HCT) have often reported poorer performance in individuals with AUD than in healthy controls (HC). However, concerns about the construct validity of the HCT complicate interpretation of these findings. This systematic review and meta-analysis quantified the magnitude and consistency of group differences in HCT performance between adults with AUD and HC, without assuming that the task provides a process-pure measure of cardiac interoceptive accuracy.

Methods: Eligible studies compared adults diagnosed with AUD with HC and reported sufficient HCT data to compute standardized mean differences. PubMed/MEDLINE, PsycINFO, Embase, and Web of Science were searched up to 4 November 2025. Study quality was assessed using an adapted Newcastle–Ottawa Scale. Hedges' g values were pooled using a random-effects model. The review received no specific external funding and was preregistered in PROSPERO [ID = CRD420251173904].

Results: Six studies including 314 individuals with AUD and 324 HC were included. The pooled effect was moderate-to-large and negative, indicating lower HCT performance in AUD than HC, $g = -0.79$, 95% CI $[-0.96, -0.63]$. All included studies showed negative effect sizes, although the small evidence base limits certainty.

Conclusions: Lower HCT performance appears to represent a preliminary but consistent task-level behavioral difference in AUD. These findings should not be interpreted as direct evidence of reduced cardiac interoceptive accuracy, but they support further investigation of whether HCT performance relates to clinically relevant AUD features.

Keywords interoception, alcohol use disorder, meta-analysis, heartbeat counting task, cardiac perception

Introduction

Alcohol use disorder (AUD) is a chronic and relapsing condition associated with substantial personal, social, and economic costs worldwide (World Health Organization 2024). Despite the availability of evidence-based interventions, relapse rates remain high, and many individuals experience persistent difficulties in regulating craving and affect even after treatment (Alba-Ferrara et al. 2014, Nguyen et al. 2020, Beaulieu et al. 2021, Zaiser et al. 2025). Traditional models of AUD have primarily emphasized cognitive control deficits, reinforcement learning processes, and social or environmental factors (Goldstein and Volkow 2011, Berridge and Robinson 2016). More recently, theoretical frameworks in cognitive neuroscience have highlighted the potential role of interoception, defined as the perception and interpretation of internal bodily signals (Chen et al. 2021), in addictive behavior and

self-regulation (Paulus and Stewart 2013, Lovelock et al. 2021, Wiśniewski et al. 2021).

Within predictive coding and active inference frameworks, interoception is conceptualized as the brain's ongoing attempt to infer bodily states by integrating noisy sensory input with prior expectations (Seth 2013). In this view, affective experience and adaptive regulation depend on how interoceptive prediction errors are weighted relative to prior expectations and used to update bodily-state beliefs (Seth and Friston 2016). From this perspective, disrupted interoceptive processing may contribute to alcohol-related decision-making, including continued use despite negative consequences and difficulty regulating craving, withdrawal, or stress-related bodily states (Paulus and Stewart 2013). In AUD, chronic alcohol consumption has been proposed to alter both peripheral physiological signals and their central processing, potentially leading to distorted bodily awareness and

Received: 23 June 2026. Revised: 20 August 2026. Accepted: 21 August 2026

© The Author(s) 2026. Published by Oxford University Press on behalf of the Medical Council on Alcohol and Oxford University Press. All rights reserved. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

reduced capacity for adaptive regulation (Paulus and Stewart 2013, Lovelock et al. 2021).

Empirically, interoception has been investigated across multiple domains and levels of measurement, including behavioral tasks, self-report questionnaires, and neurophysiological indices (Chen et al. 2021, Desmedt et al. 2023). Following the influential framework proposed by Garfinkel et al. (2015), interoception is commonly decomposed into dissociable dimensions. These include interoceptive accuracy (IAcc), defined as performance on tasks requiring detection or estimation of internal bodily signals; interoceptive sensibility, referring to self-reported beliefs or perceptions about one's internal bodily sensations; and interoceptive awareness, reflecting metacognitive correspondence between objective performance and subjective confidence.

Research on AUD has predominantly focused on cardiac interoception (Lovelock et al. 2021), most commonly assessed using the heartbeat counting task (HCT) developed by Schandry (1981). In this task, participants are asked to silently count their heartbeats during predefined intervals, with performance typically quantified by comparing reported and recorded beats. Owing to its simplicity and widespread use, the HCT has generated a substantial literature across healthy and clinical populations, including AUD (Lovelock et al. 2021, Wiśniewski et al. 2021). Clarifying whether task-based indices of bodily-signal processing show reliable alterations in AUD is therefore relevant for addiction research, given the proposed involvement of interoceptive processes in craving, withdrawal, affect regulation, and relapse vulnerability.

Several studies using the HCT have reported lower HCT performance in individuals with AUD compared to HC, suggesting the presence of systematic group differences in task performance (e.g. Jakubczyk et al. 2020, Avcu Meriç and Sönmez 2022). However, the HCT has been subject to increasing methodological scrutiny, and its interpretation remains contested, as task performance appears to be influenced not only by interoceptive perceptual processes but also by non-interoceptive factors such as time estimation, prior knowledge of heart rate, and response strategies (Brener and Ring 2016, Desmedt et al. 2018, Zamariola et al. 2018, Corneille et al. 2020, Desmedt et al. 2020). As a result, it remains unclear whether observed group differences in HCT performance can be straightforwardly interpreted as reflecting differences in IAcc per se. In the present study, we therefore treat HCT performance as a task-based behavioral index that may reflect a composite of interoceptive, cognitive, inferential, and estimation-related processes, rather than as a process-pure measure of cardiac IAcc.

Additionally, recent meta-analytic and systematic review work has begun to evaluate the empirical and conceptual status of cardiac interoception, particularly as indexed by the HCT. For instance, large-scale meta-analyses have reported weak and inconsistent associations between HCT performance and theoretically relevant mental health outcomes, raising concerns about its construct validity as a measure of IAcc (Desmedt et al. 2022). In parallel, meta-analytic evidence suggests that commonly used cardiac interoception tasks, such as the HCT and the heartbeat discrimination task (Whitehead et al. 1977), show only limited convergence, indicating that they may not index a common underlying ability (Hickman et al. 2020). More broadly, meta-analyses examining interoception across domains (e.g. alexithymia, stress) highlight substantial heterogeneity in both measurement and effect sizes, suggesting that interoception may not constitute a unitary construct (e.g. Trevisan et al. 2019, Irgoras

Izagirre et al. 2026). Consistent with these concerns, a recent systematic review of interoception in substance use disorders (SUDs) found that the literature remains heavily focused on IAcc and interoceptive sensibility and relies substantially on cardiac measures such as the HCT. The review identified reduced IAcc as the most consistent pattern in AUD, although interpretive caution is warranted because of unresolved measurement limitations (Im et al. 2026).

In this context, an important question is whether HCT performance reveals a consistent and reproducible behavioral pattern across studies comparing individuals with AUD and HC, despite ongoing concerns about task validity. Crucially, if such a group-level difference were found to be robust and directionally consistent across studies, it would constitute a non-trivial empirical regularity, even while its construct validity remains contested, as it would suggest that HCT performance systematically tracks clinically AUD–HC group distinctions.

The present study addresses this gap by conducting a meta-analysis of studies comparing HCT performance between individuals with AUD and HC participants. Beyond estimating the magnitude of this difference, the primary aim was to determine whether the observed pattern is sufficiently consistent across studies to be interpreted as reproducible. Rather than adjudicating the construct validity of the HCT or endorsing a single theoretical account of interoception, the present analysis focuses on establishing whether a reliable and interpretable behavioral regularity emerges in AUD, one that may constrain and inform subsequent models of bodily-signal processing in addiction and guide more precise assessment of clinically relevant interoceptive mechanisms in AUD.

Materials and methods

Protocol and reporting standards

The study protocol was preregistered in PROSPERO prior to data extraction (ID = CRD420251173904). The meta-analysis was conducted and reported in accordance with PRISMA 2020 guidelines (Page et al. 2021). A completed PRISMA 2020 checklist is provided as Supplementary Material.

Search strategy

A systematic search was conducted in PubMed/MEDLINE, PsycINFO, Embase, and Web of Science to identify peer-reviewed studies reporting HCT performance in individuals with AUD. Search terms combined interoception-related keywords (e.g. interoception, heartbeat counting) with AUD-related terms (e.g. AUD, alcohol dependence). Searches were restricted to English-language articles published from 1981 onward, following the introduction of the HCT. The final search was conducted on 4 November 2025. In accordance with the preregistered protocol, backward citation chasing of the reference lists of included studies and forward citation chasing using Google Scholar were also conducted. Complete search strategies are reported in [Supplementary Table S1](#).

Eligibility criteria

Studies were included if they: (1) enrolled adults diagnosed with AUD according to DSM or ICD criteria; (2) included a healthy adult

control group without reported current SUDs; (3) reported performance on the HCT or a closely related heartbeat-counting variant, allowing for variations in task instructions, counting intervals, and the equipment used to record heartbeats; and (4) provided, or allowed derivation of, sufficient data to compute standardized effect sizes. Studies were excluded if they involved children or adolescents, lacked a control group, did not report quantitative HCT outcomes, or were not original empirical studies.

Study selection and data extraction

Search records were imported into Rayyan, which was used to facilitate the identification of potential duplicates and screening management (Ouzzani et al. 2016). No automation tools were used to determine study eligibility. After duplicate removal, two reviewers independently screened the remaining records in Rayyan and classified them as Include, Exclude, or Maybe. Any disagreements or uncertain classifications during title/abstract and full-text screening were resolved by consensus. Data were independently extracted by the two reviewers using standardized spreadsheets and included sample characteristics, diagnostic criteria, task characteristics, and HCT-derived outcomes. Discrepancies in data extraction were resolved by consensus. When multiple interoception measures were reported, only HCT-derived data were extracted. For studies from overlapping research groups, recruitment periods were checked and only non-overlapping samples were retained.

Data handling and effect size computation

Group differences in HCT performance between AUD and HC groups were quantified using Hedges' g , with negative values indicating lower HCT performance in AUD. Effect size variances were computed using standard formulas for independent groups. When studies reported SEMs, these were converted to SDs using sample sizes. When multiple eligible HCT conditions were available, one effect size was selected to preserve statistical independence. For Billaux et al. (2025), the standard "count felt heartbeats" condition was used in the primary analysis, and alternative instructions ("count felt heartbeats and do not guess") were examined in sensitivity analyses. Published summary statistics were supplemented with author-provided raw data when needed.

Statistical analysis

The primary meta-analysis pooled Hedges' g values using a random-effects framework to account for between-study variability. The pooled effect size and its 95% confidence interval were estimated using inverse-variance weighting. Between-study variance (τ^2) was estimated using the DerSimonian-Laird method. Statistical heterogeneity was assessed using Cochran's Q statistic and the I^2 index. To improve interpretability on the original HCT scale, we additionally calculated unstandardized mean differences, defined as AUD minus HC, and pooled them using the same random-effects framework.

Given the small number of included studies, estimates of between-study heterogeneity should be interpreted with caution, as they are likely to be imprecise. Accordingly, the absence

of detectable heterogeneity was not taken to indicate true homogeneity across studies.

All statistical analyses were conducted in Python using standard scientific and meta-analytic libraries, rather than in R as specified in the preregistered protocol; this change did not affect the analytic strategy or effect size estimation.

Risk of bias and meta-bias

Study quality was assessed independently by two reviewers using an adapted Newcastle-Ottawa Scale covering selection, comparability, and exposure/outcome domains. Discrepancies were resolved by consensus. Small-study effects were explored using Egger's regression test and funnel-plot inspection, recognizing that these analyses were underpowered.

Sensitivity and additional analyses

Robustness was examined using leave-one-out analyses, a quality-restricted analysis, and an instruction-based sensitivity analysis replacing the primary Billaux et al. (2025) estimate with the alternative no-guessing condition. Potential moderator analyses, including task instruction type, were limited to descriptive and visual comparisons because of the small number of studies and incomplete reporting.

Results

Study selection

Six studies met all inclusion criteria and were included in the meta-analysis. Backward and forward citation chasing identified no additional eligible studies. The study selection process, including reasons for exclusion at each stage, is summarized in the PRISMA flow diagram (Fig. 1) using PRISMA 2020 (Haddaway et al. 2022).

Study characteristics

The included studies comprised adult participants diagnosed with AUD and HC groups, with diagnoses established according to DSM or ICD criteria, with the specific diagnostic system and version used in each study reported in Table 1. Across studies, HCT outcomes were used as the primary task-based index of cardiac interoception. Sample sizes ranged from 68 to 190 participants per study, with a total of 314 individuals with AUD and 324 HC included in the meta-analysis. Mean age was relatively consistent across studies, whereas sex distribution was markedly unbalanced, with most samples comprising predominantly male participants and only one study including a substantial proportion of women (Table 1).

Four of the six included estimates came from studies conducted at Trakya University School of Medicine or its associated addiction treatment center and shared overlapping authorship (Ateş Çöl et al. 2016, Sönmez et al. 2017, Sübay and Sönmez 2021, Avcu Meriç and Sönmez 2022). As shown in Table 1, reported sampling periods did not overlap across these studies, reducing concern about direct participant duplication; however, this concentration may still limit the external independence of the evidence.

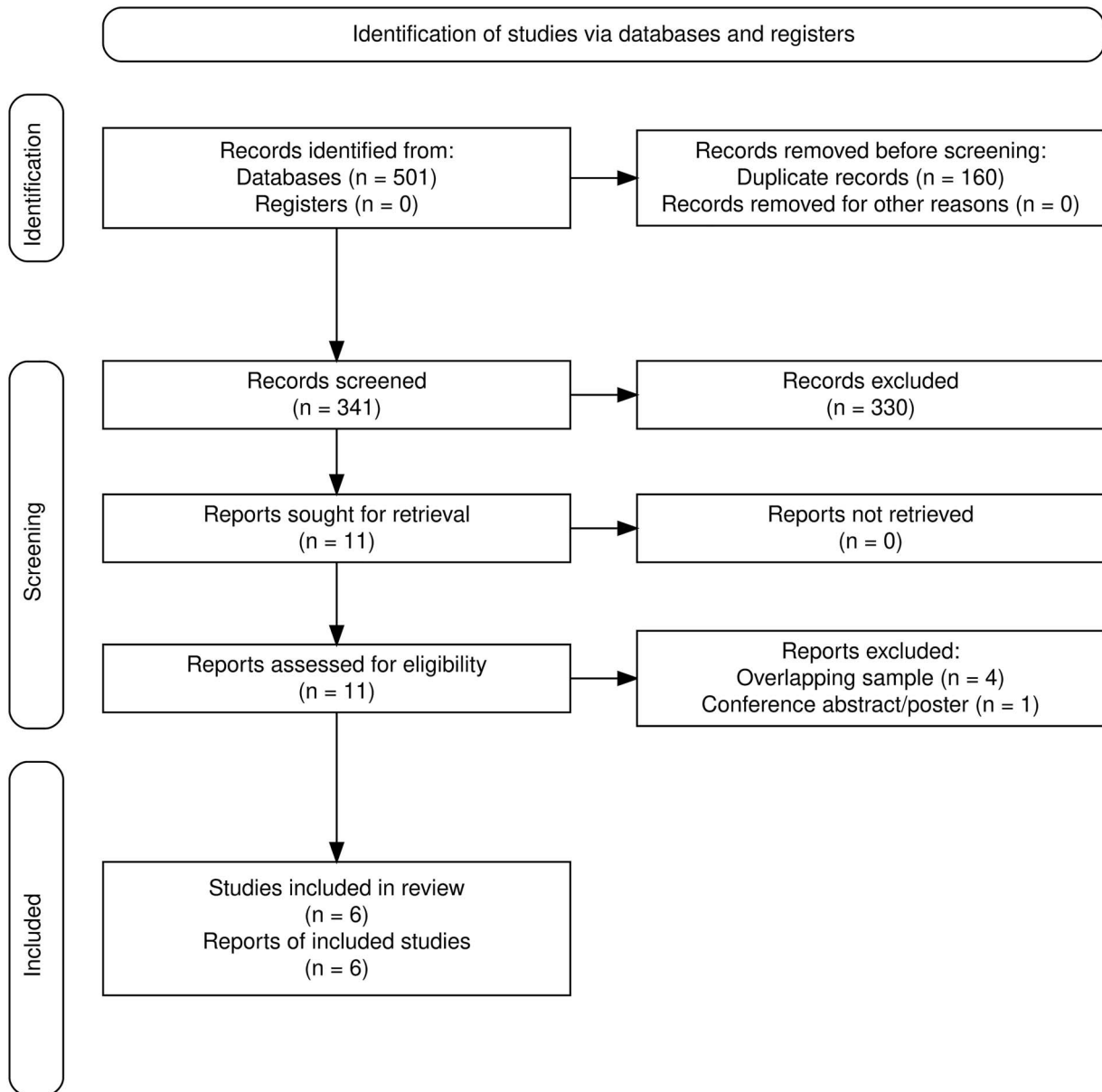


Figure 1 PRISMA flow diagram of study selection.

Task parameters varied across studies, including the number and duration of counting intervals, recording hardware (predominantly electrocardiographic devices), pauses between trials, and task instructions. Despite these variations, all studies used the same algebraic HCT scoring structure: the correspondence between counted and recorded heartbeats was calculated for each interval, and interval-specific scores were then averaged across the intervals administered. Task instructions were classified into three categories: (1) standard counting instructions without explicit restrictions on estimation strategies (“count heartbeats”); (2) instructions emphasizing perception of consciously felt heartbeats (“count felt heartbeats”); and (3) instructions explicitly discouraging guessing or estimation (“count felt heartbeats and do not guess”). Most studies used either standard or perception-focused instructions, whereas one study (Billaux et al. 2025) included two instruction conditions. Reporting of clinical and demographic variables was inconsistent, limiting

comparability across secondary measures such as craving and affective processing (Table S4).

Meta-analytic results

Individual study effect sizes and 95% confidence intervals are shown in the forest plot (Fig. 2). The random-effects meta-analysis yielded a negative pooled effect size (Hedges’ $g = -0.79$, $SE = 0.08$, 95% CI $[-0.96, -0.63]$), indicating lower HCT performance in AUD relative to HC. All studies showed negative effect sizes, indicating a consistent directional pattern across samples, although precision varied across estimates (Table 2 and Table S5). Across individual studies, unstandardized AUD–HC mean differences in HCT scores ranged from -0.14 to -0.09 . The pooled unstandardized mean difference was -0.12 , 95% CI $[-0.14, -0.09]$, indicating that participants with AUD scored, on average, approximately 0.12 points lower than HC on the 0–1 HCT score

Table 1 Characteristics of the studies included in the meta-analysis.

Reference	Year	Sampling period	<i>n</i> (AUD)	<i>n</i> (HC)	Age (years)	% Female	Diagnostic criteria	HCT intervals(s)	Instructions
Ateş Çöl et al.	2016	Mar 2012–Aug 2012	55	52	42.39 ± 11.15	9.3	DSM-IV-TR	25, 35, 45, 100	1
Sönmez et al.	2017	Jan 2014–Jun 2015	33	35	41.24 ± 9.58	0.0	DSM-IV-TR	25, 35, 45, 100	2
Jakubczyk et al.	2020	N/R	86	104	42.64 ± 10.20 ^a	17.3 ^a	ICD-10	25, 35, 45	1
Sübay and Sönmez	2021	Jul 2016–Jul 2017	40	40	42.02 ± 12.49	0.0	DSM-5	25, 35, 45, 100	2
Avçu Meriç and Sönmez	2022	Aug 2018–Feb 2020	52	52	45.59 ± 10.85	0.0	DSM-5	25, 35, 45, 100	2
Billaux et al.	2025	N/R	48	41	45.66 ± 10.25	52.0	DSM-5	25, 35, 45	1 and 3
Total	–	–	314	324	43.20 ^b	14.4 ^b	–	–	–

Note. Age is reported in years as mean ± standard deviation. Sample sizes in the *n* (AUD) and *n* (HC) columns correspond to participants with available HCT data. Female percentages refer to the samples for which demographic characteristics were reported in the original studies. HCT intervals indicate the duration of each counting interval in seconds. Dashes indicate values that are not applicable to the total row. Instruction types were coded as follows: (1) standard counting instructions without explicit restrictions (“count heartbeats”); (2) instructions emphasizing the perception of consciously felt heartbeats (“count felt heartbeats”); (3) instructions explicitly discouraging guessing or estimation (“count felt heartbeats and do not guess”). AUD, alcohol use disorder; HC, healthy controls; DSM-IV-TR, Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision; DSM-5, Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition; ICD-10, International Classification of Diseases, Tenth Revision; N/R, not reported. ^aFor Jakubczyk et al. (2020), age and female proportion refer to the full study sample because demographic characteristics were not reported separately for the subsample with available HCT data. ^bThe overall mean age and female proportion were calculated from study-level demographic statistics weighted by the corresponding demographic sample sizes. The overall female proportion was estimated from reported percentages.

Table 2 Meta-analytic results for HCT performance in AUD versus HC.

Reference	<i>n</i> (AUD/HC)	HCT (AUD)	HCT (HC)	Hedges' <i>g</i>	SE <i>g</i>	95% CI	Weight (%)
Ateş Çöl et al.	55/52	0.58 ± 0.18	0.71 ± 0.16	−0.76	0.20	[−1.15, −0.36]	17.0
Sönmez et al.	33/35	0.61 ± 0.23	0.75 ± 0.12	−0.76	0.25	[−1.26, −0.27]	10.8
Jakubczyk et al.	86/104	0.61 ± 0.16	0.72 ± 0.08	−0.89	0.15	[−1.19, −0.59]	29.2
Sübay and Sönmez	40/40	0.64 ± 0.15	0.73 ± 0.16	−0.58	0.23	[−1.02, −0.13]	13.1
Avçu Meriç and Sönmez	52/52	0.60 ± 0.13	0.72 ± 0.09	−1.07	0.21	[−1.48, −0.65]	15.5
Billaux et al. (A)	48/41	0.50 ± 0.26	0.64 ± 0.22	−0.57	0.22	[−1.00, −0.15]	14.5
Total	314/324	–	–	−0.79	0.08	[−0.96, −0.63]	100.0

Note. HCT values are reported as mean ± standard deviation. Effect sizes are expressed as Hedges' *g*, with negative values indicating lower HCT performance in the AUD group relative to the HC group. Weight indicates each study's percentage contribution to the pooled effect estimate in the random-effects meta-analysis. Billaux et al. (A) refers to the statistics obtained with the classic HCT instruction. AUD, alcohol use disorder; HC, healthy controls; HCT, heartbeat counting task; SE, standard error; CI, confidence interval.

scale. Heterogeneity was not detectable in this complementary analysis, $Q(5) = 1.39$, $I^2 = 0\%$, $\tau^2 = 0$ (Supplementary Table S6).

Heterogeneity

Between-study heterogeneity was not detectable in the primary analysis ($Q = 4.11$, $df = 5$, $I^2 = 0.0\%$; Supplementary Table S11). Given the small number of included studies, these estimates should be interpreted cautiously.

Risk of bias and study quality

Study quality assessments are summarized in (Supplementary Tables S2 and S3). Overall, studies were rated as having moderate methodological quality according to the adapted Newcastle–Ottawa Scale, with adequate scores in selection and outcome domains but more variable comparability. Certainty of evidence was not formally assessed; instead, findings were interpreted narratively in relation to study quality, sample size, consistency, precision, and sensitivity analyses.

Assessment of meta-bias

Potential small-study effects were explored using Egger's regression test and visual inspection of funnel plots. Egger's test did not indicate asymmetry (intercept = 2.26, $P = .375$) (Supplementary Table S7), and visual inspection did not reveal clear deviations from symmetry (Supplementary Fig. S3). However, these analyses are underpowered given the small number of included studies and should be interpreted with caution.

Notably, studies using HCT instructions type 2 (“count felt heartbeats”) appeared to cluster in regions of lower precision, suggesting that variation in task design may contribute to the observed distribution of effect sizes (Supplementary Fig. S3).

Additional analyses

Exploratory meta-regressions examined mean age, sex distribution, and task instruction type, but were underpowered and not interpreted. No clinical or psychological variable was consistently reported across all studies, precluding further quantitative analyses (Supplementary Table S4). Descriptive analyses suggested

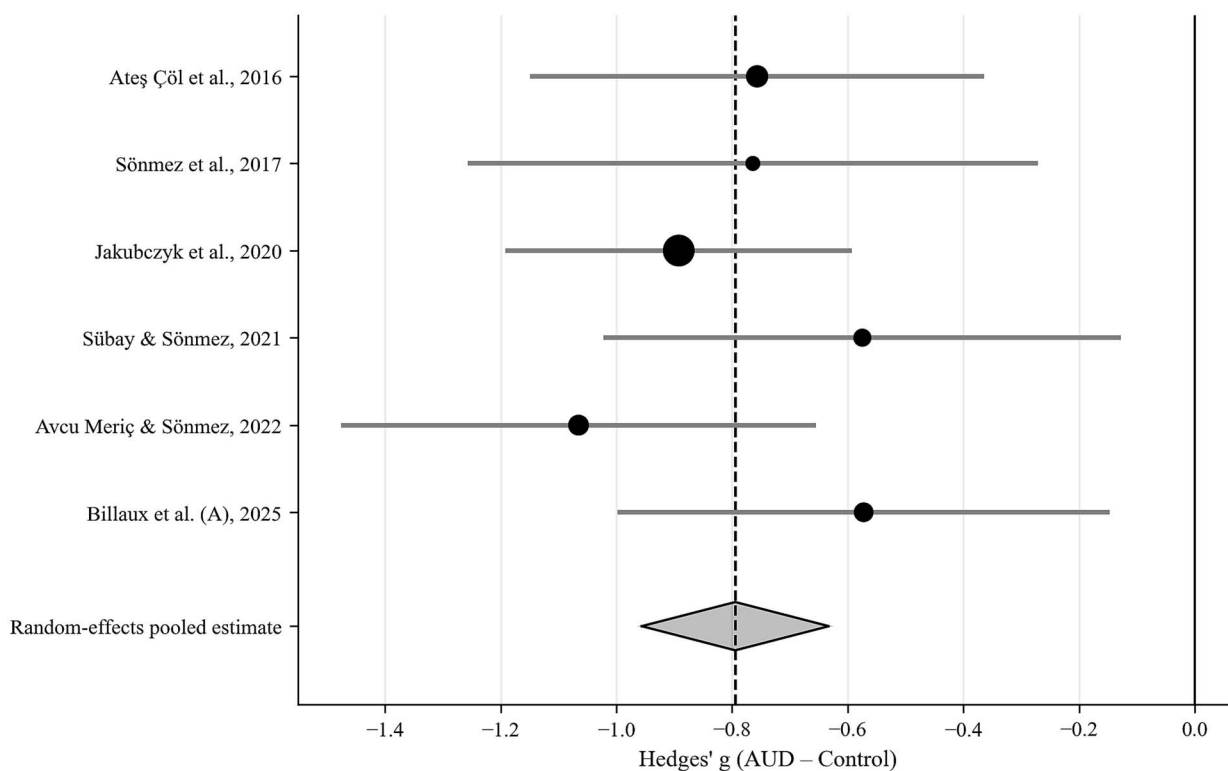


Figure 2 Forest plot of HCT performance differences between AUD and healthy controls. *Note.* Positive values indicate higher HCT performance in the AUD group than in the control subjects. Marker size represents each study's weight in the random-effects model. Horizontal lines indicate 95% confidence intervals, and the diamond represents the pooled random-effects estimate and its 95% confidence interval. Billaux et al. (A) refers to the statistics obtained with the classic HCT instruction.

consistently negative group differences across instruction categories (Supplementary Fig. S4), but no formal subgroup analyses were conducted.

Sensitivity analyses

Sensitivity analyses supported the robustness of the primary finding. Leave-one-out analyses yielded stable pooled estimates ranging from -0.83 to -0.75 , with changes not exceeding ± 0.05 , indicating that no single study disproportionately influenced the result (Supplementary Table S8 and Supplementary Fig. S5). Restricting the analysis to higher-quality studies produced a comparable estimate ($g = -0.80$, 95% CI $[-1.02, -0.57]$; Supplementary Table S9). Replacing the primary Billaux et al. (2025) estimate with the no-guessing condition also yielded a similar pooled effect ($g = -0.75$, 95% CI $[-0.95, -0.56]$; $\Delta g = 0.04$), although heterogeneity increased relative to the primary model ($I^2 = 27.8\%$ vs 0.0% ; Supplementary Tables S10–S11 and Supplementary Fig. S6).

Discussion

The present meta-analysis synthesizes evidence from six studies examining HCT performance in individuals with AUD. The pooled effect indicated lower HCT performance in AUD relative to HC ($g = -0.79$). On the original HCT scale, this corresponded to an average difference of about 0.12 points, with lower scores in AUD. The overall pattern remained similar across multiple sensitivity analyses, including leave-one-out procedures, restriction to

higher-quality studies, and substitution of the standard estimate from Billaux et al. (2025) with an alternative estimate obtained under instructions discouraging guessing. However, the apparent homogeneity should be interpreted cautiously given the small number of studies. In addition, four of the six studies originated from closely related institutional settings and shared overlapping authorship.

Together, these findings suggest a preliminary but consistent group-level difference in HCT performance, characterized by lower correspondence between reported and recorded heartbeats in AUD. The present meta-analysis cannot determine whether this pattern reflects a pre-existing vulnerability, a consequence of prolonged alcohol exposure, or their interaction. Importantly, even if concerns regarding HCT validity are accepted, the observed AUD–HC difference still calls for explanation.

These findings are aligned with a recent systematic review of interoception in SUDs, which identified reduced IAcc as the most consistent pattern in AUD (Im et al. 2026). Our study extends that conclusion by quantifying the magnitude, precision, and robustness of the specific AUD–HC difference in HCT performance, while explicitly treating HCT scores as task-based behavioral indices rather than direct measures of cardiac IAcc.

Additionally, our findings may appear to contrast with meta-analytic evidence showing no meaningful associations between HCT performance and mental health indicators such as anxiety, depression, and alexithymia (Desmedt et al. 2022). However, Desmedt et al. examined dimensional within-sample associations, whereas the present meta-analysis tests a categorical AUD–HC contrast. Thus, HCT performance may fail to track dimensional

mental health variation while still showing a consistent mean difference across the available studies. This pattern is compatible with the possibility that HCT captures a stable group-level performance shift rather than a valid dimensional marker of IACC or symptom severity.

This interpretation aligns with broader concerns regarding the construct validity of the HCT. Several studies argue that HCT performance is influenced by multiple factors, including time estimation, prior beliefs, attentional strategies, and response biases (Brener and Ring 2016, Desmedt et al. 2018, Zamariola et al. 2018, Corneille et al. 2020, Desmedt et al. 2020, Billaux et al. 2025). From this perspective, the observed AUD–HC difference may reflect systematic group differences in these perceptual and non-perceptual influences, rather than a specific interoceptive deficit.

In this regard, the present findings are compatible with accounts emphasizing the role of estimation-related processes in HCT performance. In the sensitivity analysis, replacing the standard estimate from Billaux et al. (2025) with the estimate obtained under instructions discouraging guessing slightly attenuated the pooled effect and increased heterogeneity (I^2 from 0% to 27.8%), while leaving the overall AUD–HC difference intact. Together with visual indications that effect sizes may vary across task instructions (Supplementary Figs S3–S4), this pattern suggests that instruction-dependent estimation or response strategies may influence HCT performance. Importantly, the persistence of a group difference under no-guessing instructions should not be interpreted as evidence that the residual effect reflects pure cardiac IACC. Rather, it indicates that the AUD–HC difference in HCT performance is not fully eliminated when explicit guessing is discouraged.

In light of these results, the HCT should be used cautiously and within a broader multimethod assessment of interoception in AUD. Interoception is a multidimensional construct encompassing perceptual sensitivity, metacognitive awareness, and affective interpretation, and no single task can fully capture all of these dimensions (Garfinkel et al. 2015, Khalsa et al. 2018). More broadly, these findings are consistent with evidence of substantial heterogeneity and limited cross-task convergence across interoceptive measures (e.g. Trevisan et al. 2019, Hickman et al. 2020, Irigoras Izagirre et al. 2026).

From this perspective, reduced HCT performance in AUD is unlikely to reflect a unitary deficit in cardiac perceptual accuracy and may instead arise from the combined contribution of perceptual, attentional, inferential, and estimation-related processes. Predictive coding accounts provide one possible framework for interpreting such differences, insofar as HCT performance may depend on the integration of bodily signals with prior expectations and response strategies (Seth 2013, Barrett and Simmons 2015, Ainley et al. 2016, Seth and Friston 2016). However, these mechanisms were not directly examined in the present meta-analysis and should therefore be regarded as hypotheses for future research rather than explanations established by the current findings.

To do this, future work should also further standardize the experimental protocol of the HCT (i.e. variations in the number and duration of intervals, pause length, practice trials) to reduce potential confounds. In addition, contextual conditions such as the acoustic testing environment remain insufficiently described and may influence whether participants rely on non-interoceptive cues or perceptual access to bodily signals.

A key challenge is to distinguish factors that introduce measurement bias from those that constitute genuine components of HCT performance. Future paradigms should explicitly model the interaction between sensory signals, prior knowledge, and cognitive processes (Smith et al. 2020, Allen et al. 2022, Legrand et al. 2022). Multimethod approaches integrating complementary interoceptive measures, including tasks grounded in signal detection frameworks and measures less susceptible to confounds, may further help disentangle these processes (Desmedt et al. 2023).

Clinically, our results should not be taken to support the use of HCT performance as a diagnostic marker or as a direct treatment target in AUD. Rather, they suggest that task-based differences in bodily-signal processing may be relevant for understanding clinically important alcohol-related processes such as craving, withdrawal, relapse vulnerability, and emotion regulation (Paulus and Stewart 2013, Lovelock et al. 2021, Wiśniewski et al. 2021). Future intervention studies could examine whether approaches targeting bodily awareness and the interpretation of internal states improve alcohol-related outcomes, while using complementary measures capable of separating perceptual, cognitive, attentional, and metacognitive components of interoceptive processing (Khalsa et al. 2018, Desmedt et al. 2023).

The present findings should be interpreted in light of several limitations. The small number of included studies limits statistical power and precludes robust moderator analyses, including the potential roles of abstinence duration, clinical severity, and sex differences. Moreover, four of the six included estimates came from closely related institutional settings and partially overlapping research groups. Although reported recruitment windows did not overlap, reducing concern about direct participant duplication, this concentration of evidence limits the external independence of the available literature and warrants caution when interpreting the apparent consistency of effects. Accordingly, independent replication studies using theoretically informed and methodologically standardized HCT protocols are needed to determine whether this emerging pattern generalizes beyond the currently available evidence base. Although several studies reported secondary measures relevant to interoception and addiction, these variables were inconsistently assessed across studies, limiting their inclusion in quantitative analyses.

In addition, the exclusive reliance on a single task limits the ability to generalize findings across different modalities of interoception and constrains the interpretation of the observed group-level difference. As such, the present results should be understood as specific to HCT performance rather than as providing a comprehensive account of interoceptive functioning in AUD.

Conclusion

Across six studies, this meta-analysis found lower HCT performance in individuals with AUD than in HC, with a moderate-to-large pooled effect ($g = -0.79$). Given the small evidence base and the concentration of four studies within closely related research settings, these findings should be regarded as preliminary but directionally consistent rather than as evidence of established reproducibility. They should also not be interpreted as direct evidence of reduced cardiac interoceptive accuracy. Given longstanding validity concerns, the observed pattern is better understood as a task-based difference that may reflect the combined influence

of perceptual sensitivity, prior expectations, attentional strategies, and estimation processes. Thus, although frameworks such as predictive coding offer plausible interpretations, the available data do not allow adjudication among competing mechanisms. Independent studies conducted across more diverse research settings are needed to determine whether this apparent AUD–HC difference generalizes beyond the currently available evidence base.

Acknowledgements

The authors thank Pauline Billaux for kindly providing additional study data to support data extraction. This acknowledgment does not imply endorsement of the present analyses or conclusions.

Author contributions

Nicolás Guillermo González: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Software, Visualization, Writing – Original Draft, Project administration. Luisa Göde: Conceptualization, Methodology, Investigation, Data curation, Validation, Writing – Review & Editing, Project administration. Lucía Alba-Ferrara: Conceptualization, Methodology, Resources, Supervision, Validation, Writing – Review & Editing.

Supplementary data

Supplementary data is available at *Alcohol and Alcoholism* online.

Conflict of interest

The authors declare no competing interests.

Funding

This review received no specific external funding. Non-financial institutional support was provided through the authors' affiliated institutions.

Data availability

The extracted datasets, analytic code, output files, supplementary materials, PRISMA 2020 checklists, data dictionary, and quality-assessment coding rules and rationale are available in the public OSF repository associated with this project: <https://osf.io/3b8jy/>. The repository contains materials corresponding to the different stages of the project and will be updated as necessary to reflect the final published version.

Deviations from the preregistered protocol

Several departures from the preregistered protocol should be noted (ID=CRD420251173904). First, although the protocol specified that analyses would be conducted in R, all analyses were ultimately implemented in Python. This change did not affect the analytic strategy, effect size metric, or interpretation framework. Second, a complementary analysis of raw HCT group differences was added post hoc to facilitate interpretation of the standardized meta-analytic effect. This analysis was descriptive and did not

alter the preregistered primary effect-size computation or meta-analytic model. Third, moderator analyses originally planned for variables such as age, sex, abstinence duration, and instruction type could not be meaningfully conducted because of the small number of included studies and inconsistent reporting across primary studies. Fourth, small-study bias correction methods originally specified in the preregistered protocol (trim-and-fill and PET-PEESE) were not applied, as these approaches are known to yield unstable estimates in meta-analyses with a limited number of studies. Additionally, meta-regression analyses were conducted only in an exploratory manner and not interpreted due to insufficient statistical power. No other deviations substantially affecting the main conclusions were introduced.

Declaration of generative AI and AI-assisted technologies

During the preparation of this work, the authors used ChatGPT (OpenAI, GPT-5.5 Thinking) to assist with language editing, stylistic refinement, and code review for error checking. The tool was not used to generate research hypotheses, determine eligibility decisions, extract data, conduct the statistical analyses, or interpret the findings. After using this tool, the authors reviewed and edited all outputs as needed and take full responsibility for the content of the published article.

References

- Ainley V, Apps MAJ, Fotopoulou A. *et al.* 'Bodily precision': a predictive coding account of individual differences in interoceptive accuracy. *Phil Trans R Soc B* 2016;**371**:20160003. <https://doi.org/10.1098/rstb.2016.0003>
- Alba-Ferrara L, Fernandez F, Salas R. *et al.* Transcranial magnetic stimulation and deep brain stimulation in the treatment of alcohol dependence. *Addict Disord Treat* 2014;**13**:159–69. <https://doi.org/10.1097/ADT.0b013e31829cf047>
- Allen M, Levy A, Parr T. *et al.* In the Body's eye: the computational anatomy of interoceptive inference. *PLoS Comput Biol* 2022;**18**:e1010490. <https://doi.org/10.1371/journal.pcbi.1010490>
- Ateş Çöl I, Sönmez MB, Vardar ME. Evaluation of interoceptive awareness in alcohol-addicted patients. *Arch Neuropsychiatry* 2016;**53**:17–22. <https://doi.org/10.5152/npa.2015.9898>
- Avcu Meriç I, Sönmez MB. Decision-making, interoceptive awareness and mindful attention awareness in male patients with alcohol use disorder. *Cogn Neuropsychiatry* 2022;**27**:35–48. <https://doi.org/10.1080/13546805.2021.2011183>
- Barrett LF, Simmons WK. Interoceptive predictions in the brain. *Nat Rev Neurosci* 2015;**16**:419–29. <https://doi.org/10.1038/nrn3950>
- Beaulieu M, Tremblay J, Baudry C. *et al.* A systematic review and meta-analysis of the efficacy of the long-term treatment and support of substance use disorders. *Soc Sci Med* 2021;**285**:114289. <https://doi.org/10.1016/j.socscimed.2021.114289>
- Berridge KC, Robinson TE. Liking, wanting, and the incentive-sensitization theory of addiction. *Am Psychol* 2016;**71**:670–9. <https://doi.org/10.1037/amp0000059>
- Billaux P, Desmedt O, Corneille O. *et al.* Validity concerns about the heartbeat counting task extend to alcohol use disorder:

- evidence from subclinical and clinical samples. *Addict Biol* 2025;**30**:e70032. <https://doi.org/10.1111/adb.70032>
- Brener J, Ring C. Toward a psychophysics of interoceptive processes: the measurement of heartbeat detection. *Philos Trans R Soc B* 2016;**371**:20160015. <https://doi.org/10.1098/rstb.2016.0015>
- Chen WG, Schloesser D, Arensdorf AM. *et al.* The emerging science of interoception: sensing, integrating, interpreting, and regulating signals within the self. *Trends Neurosci* 2021;**44**:3–16. <https://doi.org/10.1016/j.tins.2020.10.007>
- Corneille O, Desmedt O, Zamariola G. *et al.* A heartfelt response to Zimprich and Ainley (2019): comment on the measurement of interoceptive accuracy. *Biol Psychol* 2020;**152**:107869. <https://doi.org/10.1016/j.biopsycho.2020.107869>
- Desmedt O, Luminet O, Corneille O. The heartbeat counting task largely involves non-interoceptive processes: evidence from both the original and an adapted counting task. *Biol Psychol* 2018;**138**:185–8. <https://doi.org/10.1016/j.biopsycho.2018.09.004>
- Desmedt O, Corneille O, Luminet O. *et al.* Contribution of time estimation and knowledge to heartbeat counting task performance under original and adapted instructions. *Biol Psychol* 2020;**154**:107904. <https://doi.org/10.1016/j.biopsycho.2020.107904>
- Desmedt O, Van Den Houte M, Walentynowicz M. *et al.* How does heartbeat counting task performance relate to theoretically-relevant mental health outcomes? A meta-analysis. *Collabra Psychol* 2022;**8**. <https://doi.org/10.1525/collabra.33271>
- Desmedt O, Luminet O, Walentynowicz M. *et al.* The new measures of interoceptive accuracy: a systematic review and assessment. *Neurosci Biobehav Rev* 2023;**153**:105388. <https://doi.org/10.1016/j.neubiorev.2023.105388>
- Garfinkel SN, Seth AK, Barrett AB. *et al.* Knowing your own heart: distinguishing interoceptive accuracy from interoceptive awareness. *Biol Psychol* 2015;**104**:65–74. <https://doi.org/10.1016/j.biopsycho.2014.11.004>
- Goldstein RZ, Volkow ND. Dysfunction of the prefrontal cortex in addiction: neuroimaging findings and clinical implications. *Nat Rev Neurosci* 2011;**12**:652–69. <https://doi.org/10.1038/nrn3119>
- Haddaway NR, Page MJ, Pritchard CC. *et al.* PRISMA2020: an R package and shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and open synthesis. *Campbell Syst Rev* 2022;**18**:e1230. <https://doi.org/10.1002/cl2.1230>
- Hickman L, Seyedsalehi A, Cook JL. *et al.* The relationship between heartbeat counting and heartbeat discrimination: a meta-analysis. *Biol Psychol* 2020;**156**:107949. <https://doi.org/10.1016/j.biopsycho.2020.107949>
- Im JJ, So E, Kim K. *et al.* A systematic review of interoception in substance use disorders. *Curr Addict Rep* 2026;**13**. <https://doi.org/10.1007/s40429-026-00747-x>
- Irigoras Izagirre N, Palmer R, Schulz A. *et al.* Investigating the relationship between cardiac interoceptive accuracy and stress: a systematic review and meta-analysis. *Neurosci Biobehav Rev* 2026;**180**:106454. <https://doi.org/10.1016/j.neubiorev.2025.106454>
- Jakubczyk A, Trucco EM, Klimkiewicz A. *et al.* Association between interoception and emotion regulation in individuals with alcohol use disorder. *Front Psych* 2020;**10**:1028. <https://doi.org/10.3389/fpsy.2019.01028>
- Khalsa SS, Lapidus RC, Canales C. *et al.* Interoception and mental health: a roadmap. *Biol Psychiatry Cognit Neurosci Neuroimaging* 2018;**3**:501–13. <https://doi.org/10.1016/j.bpsc.2017.12.004>
- Legrand N, Nikolova N, Correa C. *et al.* The heart rate discrimination task: a psychophysical method to estimate the accuracy and precision of interoceptive beliefs. *Biol Psychol* 2022;**168**:108239. <https://doi.org/10.1016/j.biopsycho.2021.108239>
- Lovelock DF, Tyler RE, Besheer J. Interoception and alcohol: mechanisms, networks, and implications. *Neuropharmacology* 2021;**200**:108807. <https://doi.org/10.1016/j.neuropharm.2021.108807>
- Nguyen L-C, Durazzo TC, Dwyer CL. *et al.* Predicting relapse after alcohol use disorder treatment in a high-risk cohort: the roles of anhedonia and smoking. *J Psychiatr Res* 2020;**126**:1–7. <https://doi.org/10.1016/j.jpsychires.2020.04.003>
- Ouzzani M, Hammady H, Fedorowicz Z. *et al.* Rayyan—a web and mobile app for systematic reviews. *Syst Rev* 2016;**5**:210. <https://doi.org/10.1186/s13643-016-0384-4>
- Page MJ, McKenzie JE, Bossuyt PM. *et al.* The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;**372**:n71. <https://doi.org/10.1136/bmj.n71>
- Paulus MP, Stewart JL. Interoception and drug addiction. *Neuropharmacology* 2013;**76**:342–50. <https://doi.org/10.1016/j.neuropharm.2013.07.002>
- Schandry R. Heart beat perception and emotional experience. *Psychophysiology* 1981;**18**:483–8. <https://doi.org/10.1111/j.1469-8986.1981.tb02486.x>
- Seth AK. Interoceptive inference, emotion, and the embodied self. *Trends Cogn Sci* 2013;**17**:565–73. <https://doi.org/10.1016/j.tics.2013.09.007>
- Seth AK, Friston KJ. Active interoceptive inference and the emotional brain. *Philos Trans R Soc B* 2016;**371**:20160007. <https://doi.org/10.1098/rstb.2016.0007>
- Smith R, Kuplicki R, Feinstein JS. *et al.* A Bayesian computational model reveals a failure to adapt interoceptive precision estimates across depression, anxiety, eating, and substance use disorders. *PLoS Comput Biol* 2020;**16**:e1008484. <https://doi.org/10.1371/journal.pcbi.1008484>
- Sönmez MB, Kılıç EK, Ateş Çöl I. *et al.* Decreased interoceptive awareness in patients with substance use disorders. *J Subst Abuse* 2017;**22**:60–5. <https://doi.org/10.3109/14659891.2016.1143048>
- Sübay B, Sönmez MB. Interoceptive awareness, decision-making and impulsiveness in male patients with alcohol or opioid use disorder. *Subst Use Misuse* 2021;**56**:1275–83. <https://doi.org/10.1080/10826084.2021.1914108>
- Trevisan DA, Altschuler MR, Bagdasarov A. *et al.* A meta-analysis on the relationship between interoceptive awareness and alexithymia: distinguishing interoceptive accuracy and sensibility. *J Abnorm Psychol* 2019;**128**:765–76. <https://doi.org/10.1037/abn0000454>
- Whitehead WE, Drescher VM, Heiman P. *et al.* Relation of heart rate control to heartbeat perception. *Biofeedback Self Regul* 1977;**2**:371–92. <https://doi.org/10.1007/BF00998623>
- Wiśniewski P, Maurage P, Jakubczyk A. *et al.* Alcohol use and interoception: a narrative review. *Prog Neuro-Psychopharmacol Biol Psychiatry* 2021;**111**:110397. <https://doi.org/10.1016/j.pnpbp.2021.110397>

- World Health Organization. *Global Status Report on Alcohol and Health and Treatment of Substance Use Disorders* World Health Organization, 2024. <https://iris.who.int/handle/10665/377960>.
- Zaiser J, Hoffmann S, Zimmermann S. *et al.* Individual stress reactivity predicts alcohol craving and alcohol consumption in alcohol use disorder in experimental and real-life settings. *Transl Psychiatry* 2025;**15**:226. <https://doi.org/10.1038/s41398-025-03447-8>
- Zamariola G, Maurage P, Luminet O. *et al.* Interoceptive accuracy scores from the heartbeat counting task are problematic: evidence from simple assumptions and large-scale data. *Biol Psychol* 2018;**137**:12–7. <https://doi.org/10.1016/j.biopsycho.2018.06.006>