


## Clinical Psychology

# How Does Heartbeat Counting Task Performance Relate to Theoretically-Relevant Mental Health Outcomes? A Meta-Analysis

Olivier Desmedt<sup>1</sup> <sup>a</sup>, Maaike Van Den Houte<sup>2</sup>, Marta Walentynowicz<sup>3</sup>, Sarah Dekeyser<sup>4</sup>, Olivier Luminet<sup>1</sup>, Olivier Corneille<sup>4</sup>

<sup>1</sup> Psychological Science Research Institute, UCLouvain, Louvain-la-Neuve, Belgium; Fund for Scientific Research (FRS-FNRS), Brussels, Belgium, <sup>2</sup> Psychological Science Research Institute, UCLouvain, Louvain-la-Neuve, Belgium; Fund for Scientific Research (FRS-FNRS), Brussels, Belgium; Translational Research Center for Gastrointestinal Disorders (Targid), Department of Chronic Diseases, Metabolism and Ageing, KU Leuven, Leuven, Belgium, <sup>3</sup> Psychological Science Research Institute, UCLouvain, Louvain-la-Neuve, Belgium; Fund for Scientific Research (FRS-FNRS), Brussels, Belgium; Centre for the Psychology of Learning and Experimental Psychopathology, KU Leuven, Leuven, Belgium, <sup>4</sup> Psychological Science Research Institute, UCLouvain, Louvain-la-Neuve, Belgium

Keywords: Interoception, Interoceptive accuracy, Heartbeat counting task, Mental health, Depression, Anxiety

<https://doi.org/10.1525/collabra.33271>

## Collabra: Psychology

Vol. 8, Issue 1, 2022

The Heartbeat Counting Task (HCT) was designed and is intended to measure the objective ability to detect cardiac signals (also called cardiac *interoceptive accuracy*). Because interoceptive accuracy is thought to play a key role in biological (e.g., body mass index) and psychological (e.g., trait anxiety) risk factors and indicators of mental health, HCT scores should be associated with these outcomes. In order to examine this question, we performed a meta-analysis on these associations among adult participants. The final data set comprised 133 studies with 11,524 participants. We focused here on the seven most studied outcomes (i.e., outcomes that were studied in at least ten published studies). HCT performance was not significantly associated with trait anxiety, depression, and alexithymia. It was weakly and negatively associated with age after correction for publication bias, sex (male > female), heart rate, and body mass index. In addition, the quality assessment indicates that only a few studies reported sample size justification (6%), pre-registration (0.8%), and data in free access (6.8%). Theoretically expected associations between HCT performance and psychological indicators of mental health were not meta-analytically found. We discuss the implications of these findings for practice and theory.

Interoception, the processing of internal bodily states by the nervous system, is thought to facilitate homeostasis by allowing adaptive autonomous physiological changes and behavioral reactions (Craig, 2015). Interoceptive accuracy (IAcc), the objective capacity to detect internal signals, is considered a critical dimension of interoception (Garfinkel et al., 2015). Deficits in interoceptive accuracy, among other processes, are thought to cause or maintain dysfunctional personality traits (e.g., alexithymia; Craig, 2004), psychological symptoms (e.g., anxiety; Paulus & Stein, 2006), mental disorders (e.g., major depressive disorder; Barrett et al., 2016), and problematic health conditions (e.g., obesity; Simmons & DeVille, 2017). Empirical studies, although inconsistent, suggest an association between interoceptive accuracy and mental/physical health indicators (Khalsa et al., 2018).

Interoceptive accuracy has been extensively measured with the Heartbeat Counting Task (HCT; Dale & Anderson, 1978; Schandry, 1981). In the HCT, participants are asked to

count their heartbeats without taking their pulse during different time intervals. The absolute proportional difference between the number of reported and actual heartbeats (i.e., HCT score) is computed to quantify cardiac IAcc. Due to its ease of use, this task has been administered in several hundreds of studies. As can be seen in [Figure 1](#), the HCT has been increasingly used and referred to in recent years. Resources are currently invested in studies administering this task (Horváth et al., 2020; Jakubczyk et al., 2021), theoretical models are partly built on HCT outcomes (e.g., Herbert & Pollatos, 2012; Verdejo-Garcia et al., 2012), and HCT research keeps inspiring health interventions (Bornemann & Singer, 2017; Sugawara et al., 2020). However, the HCT has been recently heavily criticized for its lack of construct validity (see Discussion for details).

Given the importance of interoceptive accuracy in general and resources allocated to HCT research in particular, it is unfortunate that no integrative meta-analysis has been conducted yet on the associations between HCT perfor-

<sup>a</sup> [olivier.desmedt@uclouvain.be](mailto:olivier.desmedt@uclouvain.be)  
<https://www.researchgate.net/profile/Olivier-Desmedt>

mance and mental health outcomes. Likewise, one may regret that no overall quality assessment of these studies has been reported. In the present research, we sought to contribute to this gap by examining associations between HCT performance and a set of highly studied biological (e.g., body mass index) and psychological (e.g., trait anxiety) risk factors and indicators of mental health. On a theoretical level, these variables are strongly expected to correlate with IAcc. If this is the case and if HCT scores are valid indicators of IAcc, these associations should be found. To examine this question, we performed a pre-registered systematic review and meta-analysis.

## Method

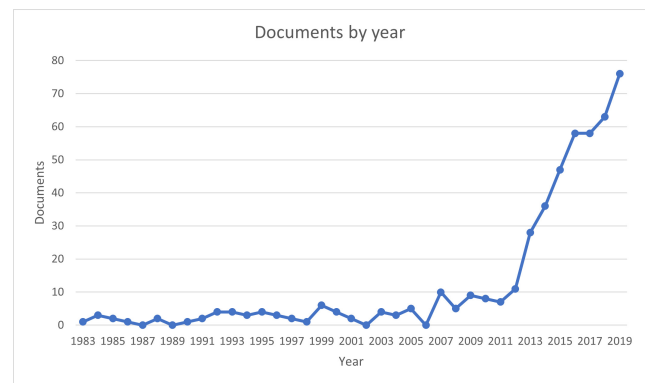
This systematic review and meta-analysis were conducted in accordance with PRISMA statements (Moher et al., 2009). The PRISMA checklist can be found in the Supplementary Materials. The protocol was registered in PROSPERO (registration number: CRD42019142176) before the review was conducted. The Covidence platform ([www.covidence.org](http://www.covidence.org)) was used to perform every step of the systematic review.

## Eligibility Criteria

We included studies with adult participants (18+) that investigated the association between HCT performance and mental disorders (e.g., major depressive disorder and anorexia nervosa), their associated symptoms (e.g., apathy) or indicators (e.g., body mass index), and their risk factors (e.g., alexithymia and sex) cited in the DSM-V. We chose to rely on the DSM-V criteria (American Psychiatric Association, 2013) for including studies in our systematic review, as most theoretically relevant outcomes are cited in the DSM-V and it allows having objective selection criteria to maximize inter-rater agreement. We considered zero-order correlation, simple regression, and mean difference, without covariate inclusion because substantial heterogeneity characterizes the type of covariate considered (e.g., heart rate, body mass index (BMI), and time estimation). We excluded (1) moderation effects as we were interested in main effects, (2) associations computed over aggregated samples coming from different population types (e.g., healthy vs. clinical) as we expected potentially different associations depending on the population type, (3) single case studies as it does not provide enough statistical power for association test, and (4) books, chapters, dissertations, and reviews as they generally do not provide original data.

## Search Strategy

A systematic literature search was performed by OD on PubMed, Scopus, PsychINFO, and ScienceDirect on July 9<sup>th</sup>, 2019 by restricting results to English papers. Unpublished studies were not sought. The following keywords were used: “Heart beat perception and emotional experience” or “Information variables in voluntary control and classical conditioning of heart rate: Field dependence and heart-rate perception” or “mental tracking task\*” or “heartbeat counting task\*” or “heartbeat perception task\*” or “heartbeat detection task\*” or “heart beat perception task\*” or “heart



**Figure 1. Yearly number of citations of the original HCT publication (Schandry, 1981) since its publication**

Note. Source: Scopus.

beat detection task\*” or “heart beat counting task\*” or “heartbeat tracking task\*” or “heart beat tracking task\*” or “mental tracking method\*”. The first two sentences correspond to the title of the two original papers of the HCT. Introducing them in the search allows finding all papers that cited these references. In case these references were not cited whereas the HCT was used, we also searched for the terms used to describe the HCT.

## Study Selection

Titles, abstracts, and full texts of studies were screened independently by OD & MW to identify studies that met the inclusion/exclusion criteria. Any disagreement between them over the eligibility of particular studies was resolved through discussion with MV. The percentage of agreement was 93.66%. Cohen’s Kappa indicates an almost perfect agreement ( $k = 0.83$ ).

## Data Collection Process

A standardized form was used to extract data from the included studies. The complete form can be found in our PROSPERO registration (CRD42019142176) in the “Published protocol” section. This protocol was pre-piloted on 20 papers by MV and SD. OD extracted 100% of included studies and MV extracted 50% of them, these papers being selected randomly. Discrepancies were identified and resolved with MW. The extracted information was: title, paper number, references, year of publication, the sample size of the study, population type, sex ratio, mean age and standard deviation, the country in which data were collected, university of the first author, the ethnicity of participants, HCT procedural details (i.e., instructions, heart rate device, the presence of a training phase, time intervals, and the formula used to compute HCT scores), variable(s) (e.g., alexithymia) that were associated with the HCT, measure and association tests (e.g.,  $r$ ,  $t$ -test,  $F$  value,  $R^2$  or Cohen’s  $d$ ). For mean differences, we computed the mean difference when no statistics were reported. When both mean difference and correlation were reported for the same hypothesis, we chose correlation as it contains more information about variance.

The correlation coefficient ( $r$ ) was the chosen effect size type. When other types of effect sizes (e.g., *Cohen's d*) were reported, we converted them in correlation coefficient ( $r$ ).

Inter-rater reliability was assessed by variable type (continuous vs. categorical) after resolving typos. For continuous variables, intra-class correlation coefficient indicates a perfect agreement ( $ICC = 1.00$ ). Although some disagreements were noticed and resolved, the large number of observations could have inflated this coefficient. For categorical variables, Cohen's kappa indicates an almost perfect agreement ( $k = 0.92$ ).

## Quality Assessment

Two modified quality assessment tools from NIH (i.e., Quality Assessment Tool for Observational Cohort and Cross-sectional Studies (National Institutes of Health, 2014a) and Quality Assessment of Case-control Studies (National Institutes of Health, 2014b)) were combined. These were completed by an additional question more related to our research question and questions related to research practice. The complete protocol can be found in our PROSPERO registration (CRD42019142176) in the "Published protocol" section. OD assessed 100% of included studies and MV assessed 50%. Discrepancies were identified and resolved with MW. Inter-rater reliability was assessed with Cohen's kappa and indicates an almost perfect agreement ( $k = 0.81$ ).

## Meta-analytic Procedure

A total of 130 outcomes were analyzed in association with the HCT. We selected those that were investigated in at least 10 studies. We chose this criterion as (i) more observations lead to higher accuracy when estimating the global effect size, and (ii) this is the minimum number of studies for performing funnel plot asymmetry tests (J. P. Higgins et al., 2019). Endorsing this criterion can lead to *positivity* biases as effects may be more studied/reported if they have a higher chance to be significant. This selection led us to consider trait anxiety ( $N = 41$ ), depression ( $N = 31$ ), alexithymia ( $N = 23$ ), BMI ( $N = 29$ ), heart rate ( $N = 40$ ), age ( $N = 20$ ), and sex ( $N = 14$ ). For the sake of clarity, we organize the result section into two broad categories: psychological (trait anxiety, depression, and alexithymia) and biological (heart rate, BMI, age, and sex) risk factors/indicators of mental health. Other associations (e.g., systolic blood pressure, health anxiety, autism) may be investigated using the data set we made public on [https://osf.io/3cwt4/?view\\_only=7e466bd994134dc787cc6a778f1d0723](https://osf.io/3cwt4/?view_only=7e466bd994134dc787cc6a778f1d0723).

We performed random-effects model meta-analyses<sup>1</sup> as the true effect could vary between studies. In these meta-analyses, more weight was given to larger sample sizes than smaller ones. To do so, we used meta, metaphor, and dmetar R packages. Heterogeneity was assessed by the Cochrane Q test and the  $I^2$  test (Borenstein et al., 2011; J. P. T. Higgins &

Thompson, 2002). Then, outlier detection analyses (Viechtbauer & Cheung, 2010) were performed and a second meta-analysis was conducted by omitting them. Influence analyses (i.e., Leave-One-Out methods; Viechtbauer & Cheung, 2010) allowed us to detect effect sizes that significantly impacted the overall effect size and overly contributed to heterogeneity. Next, moderation effects of the population (pre-registered) and measurement (not pre-registered) type were tested. While some studies used modified instructions prompting participants to only count their felt heartbeats and avoid guessing their heart rate, we could not validly test the moderation effect of instruction type as too few studies have used these modified instructions compared to original ones. We did not contact the authors when the data were not reported. However, when effects were detected, we investigated the presence of a publication bias by using two methods: the small sample bias methods (Borenstein et al., 2011) with funnel plot and Eger's test, and the p-curve analysis. Duval & Tweedie's trim and fill procedure (Duval & Tweedie, 2000) was used to adjust the overall effect size if the asymmetry was significant.

## Results

R script of analyses, all extracted data, and quality assessment can be found at the following address: [https://osf.io/3cwt4/?view\\_only=7e466bd994134dc787cc6a778f1d0723](https://osf.io/3cwt4/?view_only=7e466bd994134dc787cc6a778f1d0723). The final data set comprised of 133 studies (the complete reference list can be found in Supplementary Materials) with 11 524 participants for the associations between HCT and the 130 outcomes.

## Selection

The complete selection process is illustrated in Figure 2. The complete reference list of included studies can be found in Supplementary Materials.

## Quality Assessment and Research Practice

Most studies adequately reported the research question and objectives (100%), inclusion and exclusion criteria when recruiting participants (75.2%), the definition of the independent variable(s) (91.7%), and the measure of the independent variable(s) (99.2%). About half of the studies reported the validity and reliability of independent variable measures (49.6%), a clear description of the procedure of the HCT (49.1%), statistical assumptions (44.4%), and conflict of interest (43.6%). Only a few studies, however, reported a complete description of the group of people from which the study participants were recruited (7.5%), sample size justification (6%), pre-registration (0.8%), and data in free access (6.8%). No study disclosed having fully reported their findings (0%).

<sup>1</sup> Very similar effect sizes and confidence intervals are found when three-level meta-analyses, with study as a third-level variable, are performed to account for dependence between effect sizes.

## Meta-analysis

Results are organized into two categories: psychological and biological risk factors/indicators of mental health.

### Psychological variables

HCT performance was not significantly associated with trait anxiety ( $N = 40$ ,  $r = 0.03$ , 95% CI  $[-0.04, 0.11]$ ,  $p = 0.40$ ; see Figure 3), depression ( $N = 31$ ,  $r = -0.04$ ; 95% CI  $[-0.16, 0.08]$ ,  $p = 0.53$ ; see Figure 4), and alexithymia ( $N = 23$ ,  $r = -0.01$ , 95% CI  $[-0.17, 0.15]$ ,  $p = 0.90$ ; see Figure 5). There was low-to-moderate ( $I^2 = 43\%$ ,  $Q(39) = 68.38$ ,  $p = 0.003$ ), moderate-to-substantial ( $I^2 = 67.9\%$ ,  $Q(30) = 93.34$ ,  $p < 0.001$ ), and substantial ( $I^2 = 80.8\%$ ,  $Q(22) = 114.32$ ,  $p < 0.001$ ) heterogeneity between studies for these outcomes, respectively. Removing outliers did not substantially influence the associations between HCT performance, trait anxiety and, alexithymia. However, when removing one outlier, the association between HCT performance and depression approached significance ( $r = -0.08$ , 95% CI  $[-0.18, 0.01]$ ,  $p = 0.09$ ,  $I^2 = 35.4\%$ ,  $Q(29) = 44.92$ ,  $p = 0.03$ ). Terhaar and colleagues' study (2012) was an outlier and an influential case that could substantially bias the overall effect size.

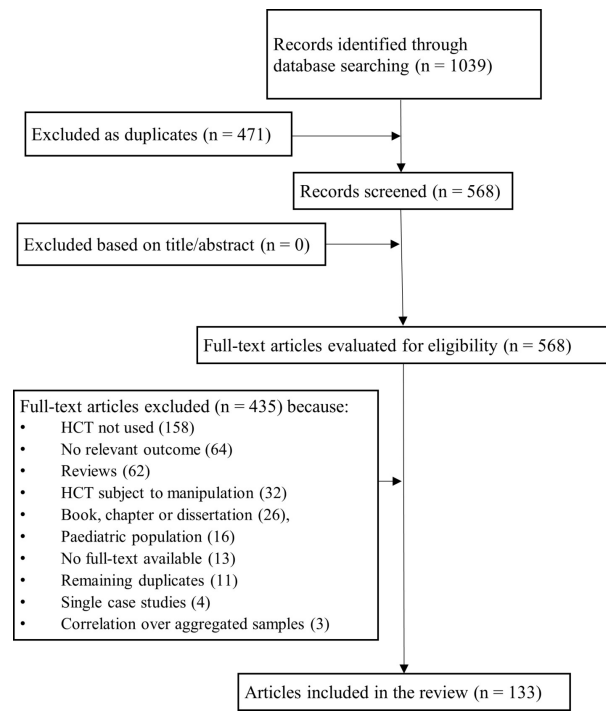
Regarding moderators, we performed three-level meta-analyses with study as the third-level variable to account for dependence between effect sizes. Population type (healthy vs clinical) did not influence the association between HCT performance and any of the three mental health indicators ( $F < 0.63$ ,  $p > 0.43$ ). Measurement type could also represent a moderator. However, each outcome was most of the time investigated by the same questionnaire. Specifically, trait anxiety, depression and, alexithymia were mostly measured by the State-Trait Anxiety Inventory (STAI;  $N = 32$ , 78%), Beck Depression Inventory-II (BDI-II;  $N = 17$ , 55%), and Toronto Alexithymia Scale-20 (TAS-20;  $N = 21$ , 91%), respectively. When the other questionnaires were removed from the meta-analysis, the conclusions remained the same.

Finally, we tested the presence of publication bias. Funnel plots and p-curves did not present any asymmetry suggesting that small studies were not missing more often than larger ones, and *P*-hacking was not common practice.

### Biological variables

HCT performance was significantly and negatively associated with heart rate ( $N = 40$ ,  $r = -0.17$ , 95% CI  $[-0.22, -0.10]$ ,  $p < 0.001$ ; see Figure 6), BMI ( $N = 29$ ,  $r = -0.11$ , 95% CI  $[-0.18$  to  $-0.04]$ ,  $p = 0.002$ ; see Figure 7), and sex (male > female;  $N = 14$ ,  $r = -0.14$ , 95% CI  $[-0.20, -0.07]$ ,  $p < 0.001$ ; see Figure 8) but not significantly associated with age ( $N = 20$ ,  $r = -0.06$ , 95% CI  $[-0.14, 0.02]$ ,  $p = 0.12$ ; see Figure 9). However, when performing three-level meta-analysis accounting for dependence between effect sizes extracted from the same study, the association between HCT performance and age became significant ( $N = 20$ ,  $r = -0.07$ , 95% CI  $[-0.14, -0.003]$ ,  $p = 0.04$ ).

There was moderate between-study heterogeneity for heart rate and BMI ( $I^2 = 48.5\%$ ,  $Q(39) = 75.75$ ,  $p < 0.001$  and  $I^2 = 59.2$ ,  $Q(28) = 68.63$ ,  $p < 0.001$ , respectively), but very low ( $I^2 = 0\%$ ,  $Q(13) = 5.90$ ,  $p = 0.95$ ) and low ( $I^2 = 27\%$ ,  $Q(19)$



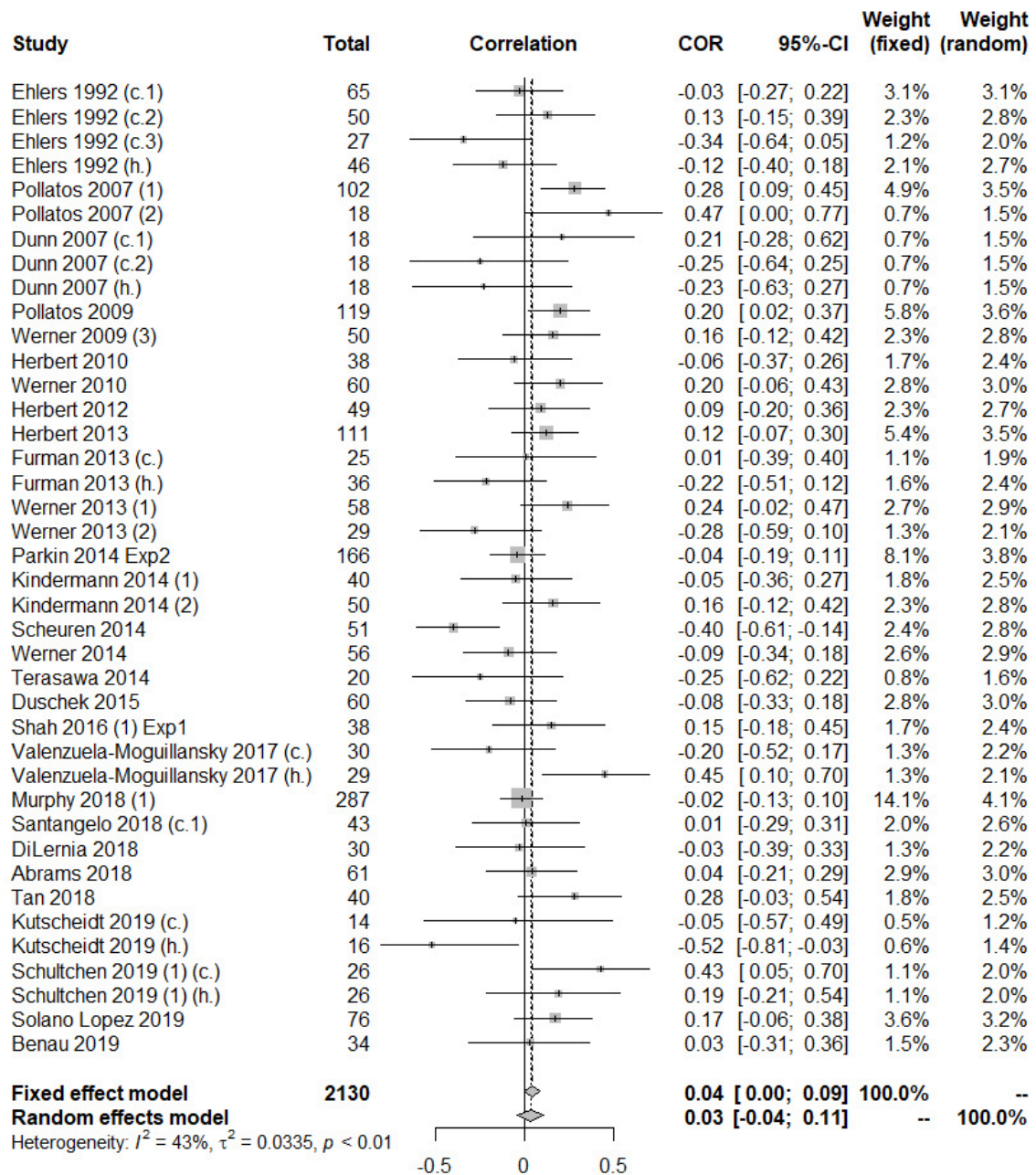
**Figure 2. Flowchart illustrating the literature search process in accordance with PRISMA guidelines**

*Note.* We directly selected the papers based on full-text as associations between HCT performance and outcomes could be included in the articles as secondary analyses not mentioned in the title or abstract.

$= 26.02$ ,  $p = 0.13$ ) between-study heterogeneity for sex and age, respectively. Removing outliers did not substantially influence the results. We did not find any outlier for the association with sex and age. For the association with heart rate and BMI, we tested moderation effects under three-level meta-analysis to account for dependence between effect sizes. There was no significant effect of population type ( $F_{1,38} = 0.15$ ,  $p = 0.70$ ) and measurement type (ECG vs other devices;  $F_{1,38} = 3.46$ ,  $p = 0.07$ ). The moderation effects of population and measurement type on the HCT-BMI association could not be tested given that too few studies ( $N = 3$ ) used a clinical sample and measurement type was often not reported. For sex and age, we could not test the effect of population type given that most studies used healthy samples.

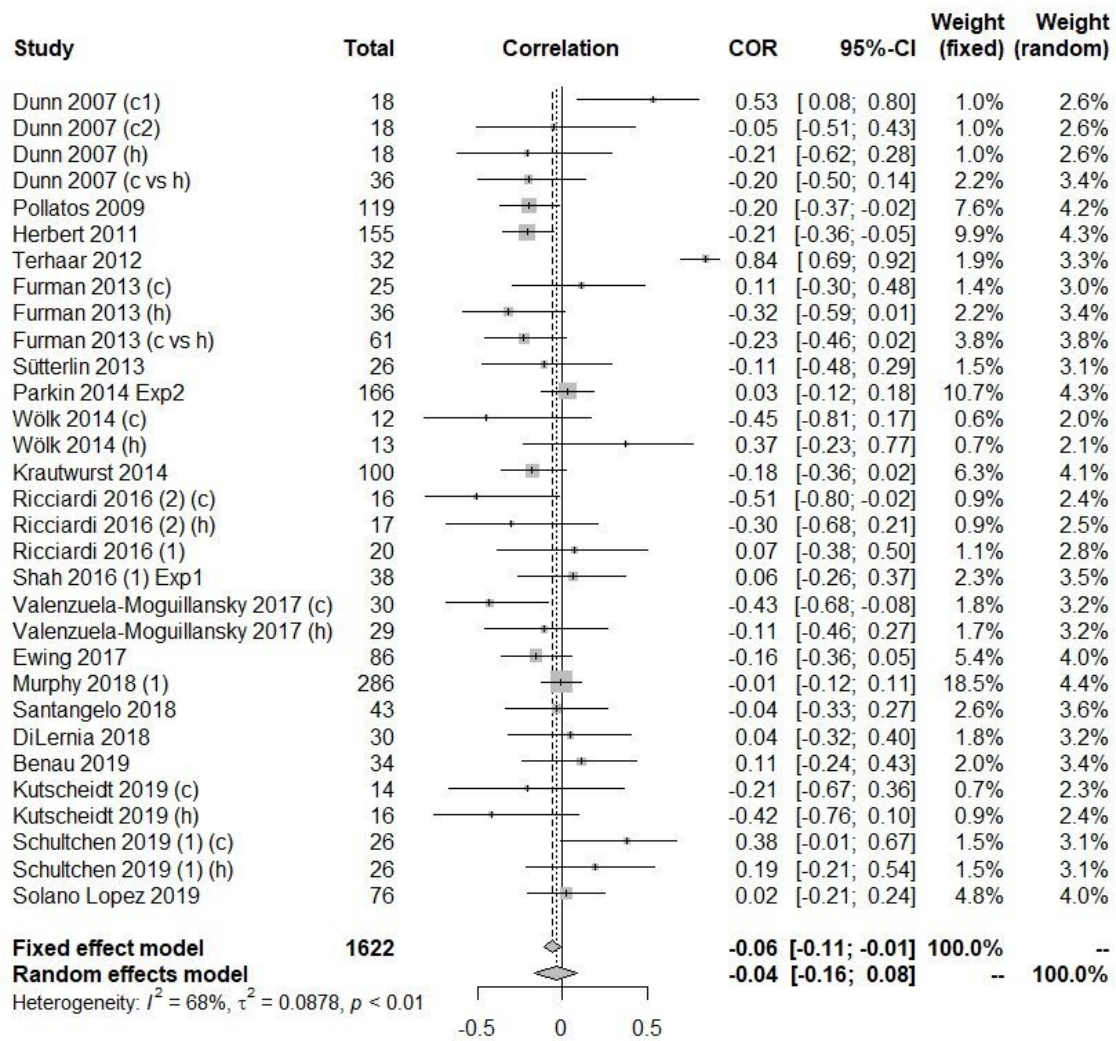
Finally, we tested the presence of a publication bias. Funnel plots and p-curves did not present any significant asymmetry for heart rate and BMI. However, funnel plots showed a significant ( $t = 2.45$ ,  $p = 0.02$ ) and marginally significant ( $t = 1.98$ ,  $p = 0.06$ ) distribution asymmetry for sex and age, respectively. This suggests that a publication bias may exist where small studies with non-significant findings are less often published. When missing studies were imputed with the trim and fill method, we observed a significant negative association with sex ( $N = 19$ ,  $r = -0.16$ , 95% CI  $[-0.24, -0.10]$ ,  $p < 0.001$ ) and age ( $N = 25$ ,  $r = -0.11$ , 95% CI  $[-0.20, -0.03]$ ,  $p = 0.009$ ). P-curve analysis could not be performed for sex given that only two or less significant ( $p < 0.05$ ) effect sizes were detected. For age, the p-curve graph does not present a left asymmetry which indicates that *p*-hacking was not common practice.





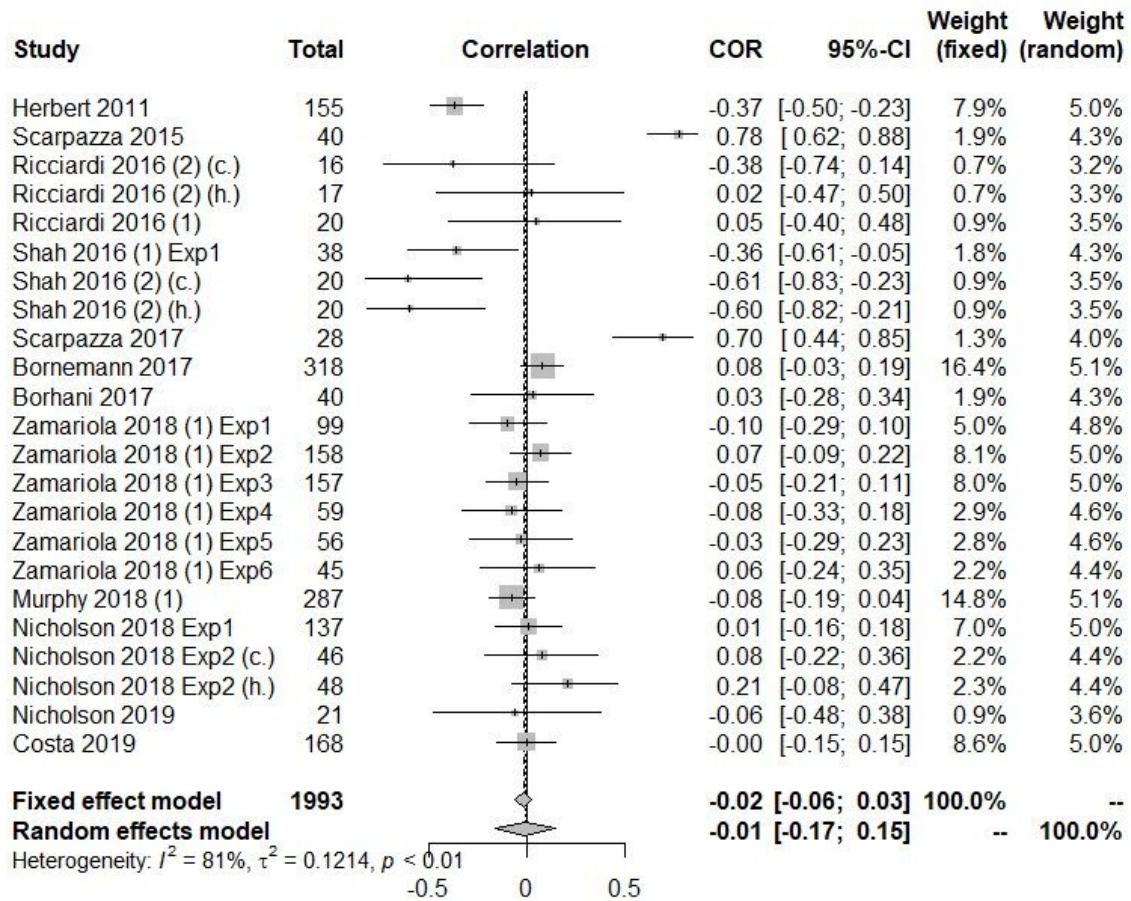
**Figure 3. Aggregated correlation between HCT performance and trait anxiety**

Note. (1) = first study by the referred author; (2) = second study by the referred author; (c.) = clinical sample; (h.) = healthy sample; (c.1) = first clinical sample; (c.2) = second clinical sample; (c.3) = third clinical sample; Exp 1 = first experiment in the paper; Exp 2 = second experiment in the paper.



**Figure 4. Aggregated correlation between HCT performance and depression**

*Note.* (1) = first study by the referred author; (2) = second study by the referred author; (c) = clinical sample; (h) = healthy sample; (c vs h) = effect size based on the comparison between clinical and healthy samples; Exp 1 = first experiment in the paper; Exp 2 = second experiment in the paper.



**Figure 5. Aggregated correlation between HCT performance and alexithymia**

Note. (1) = first study by the referred author; (2) = second study by the referred author; (c.) = clinical sample; (h.) = healthy sample; Exp 1, 2, 3, 4, 5, 6 = experiment number in the paper.



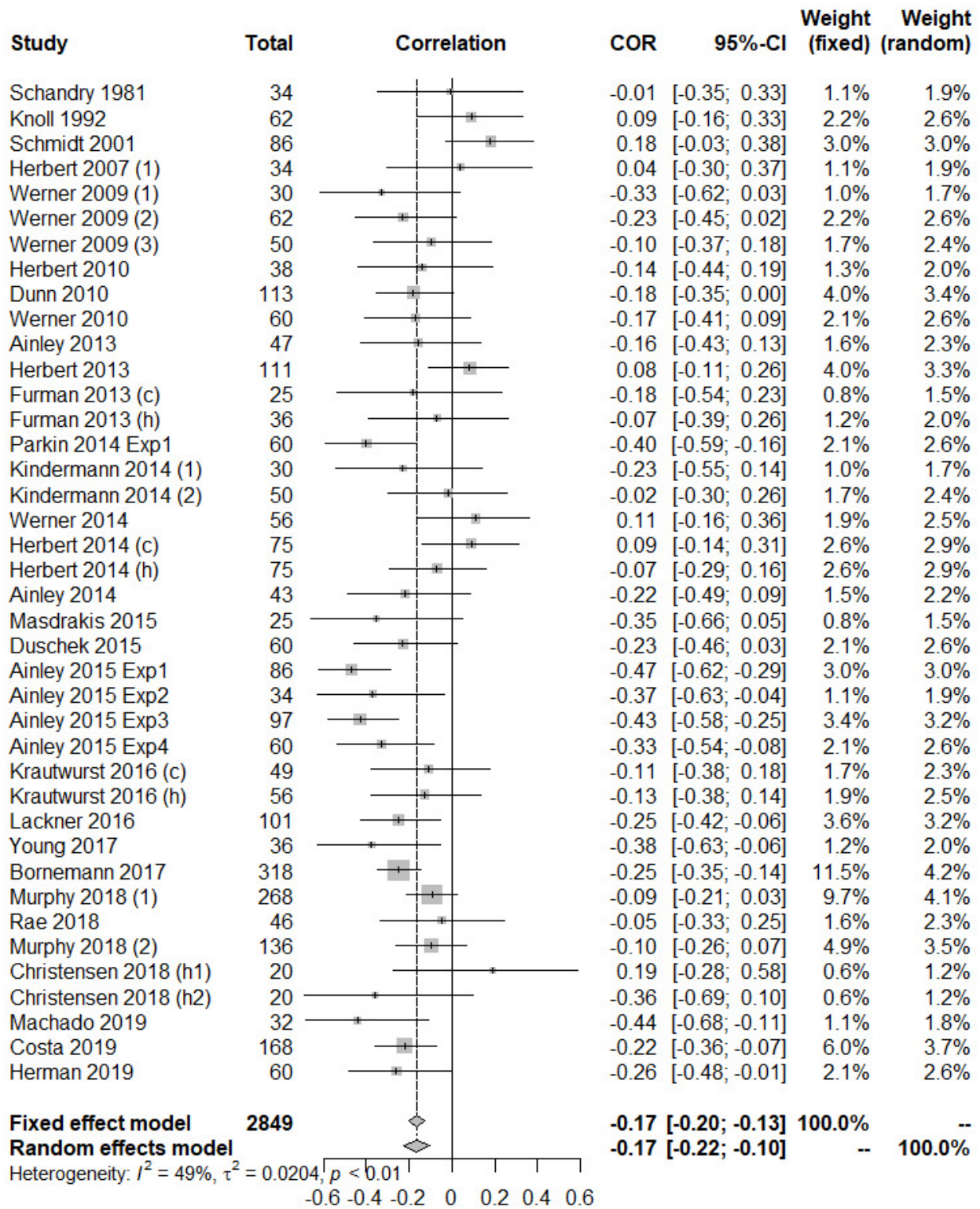
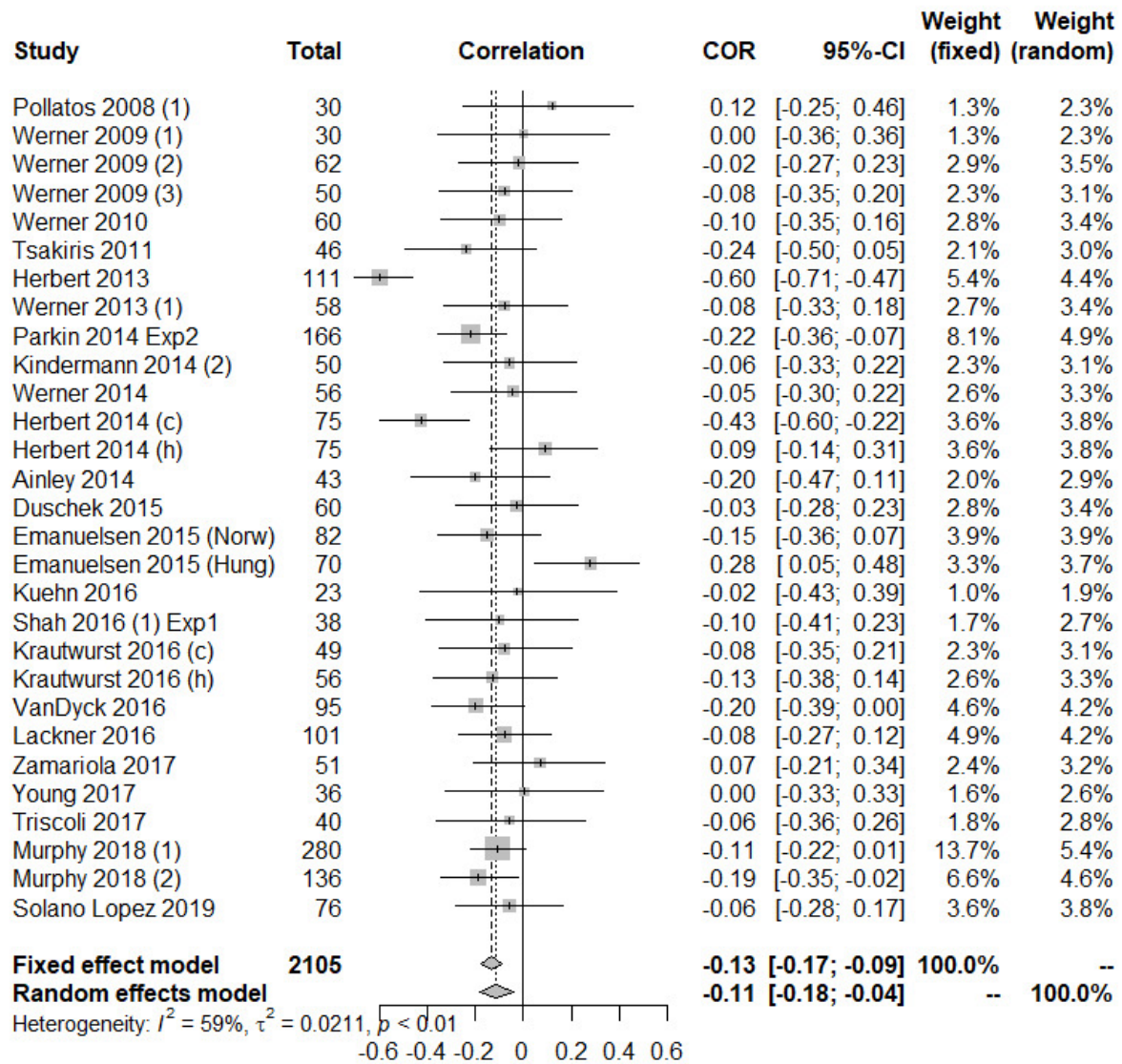


Figure 6. Aggregated correlation between HCT performance and heart rate

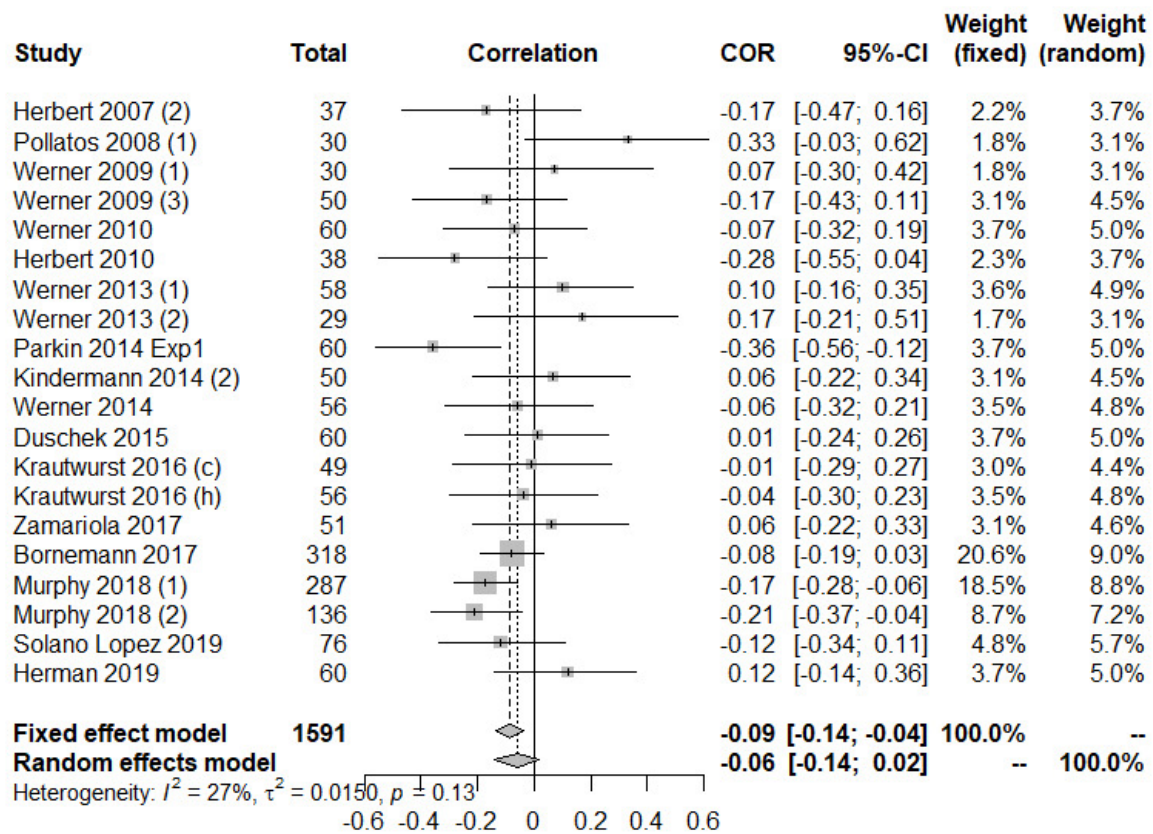
Note. (1) = first study by the referred author; (2) = second study by the referred author; (3) third study by the referred author; (c) = clinical sample; (h) = healthy sample; (h1) = first healthy sample; (h2) = second healthy sample; Exp 1, 2, 3, 4 = experiment number in the paper.





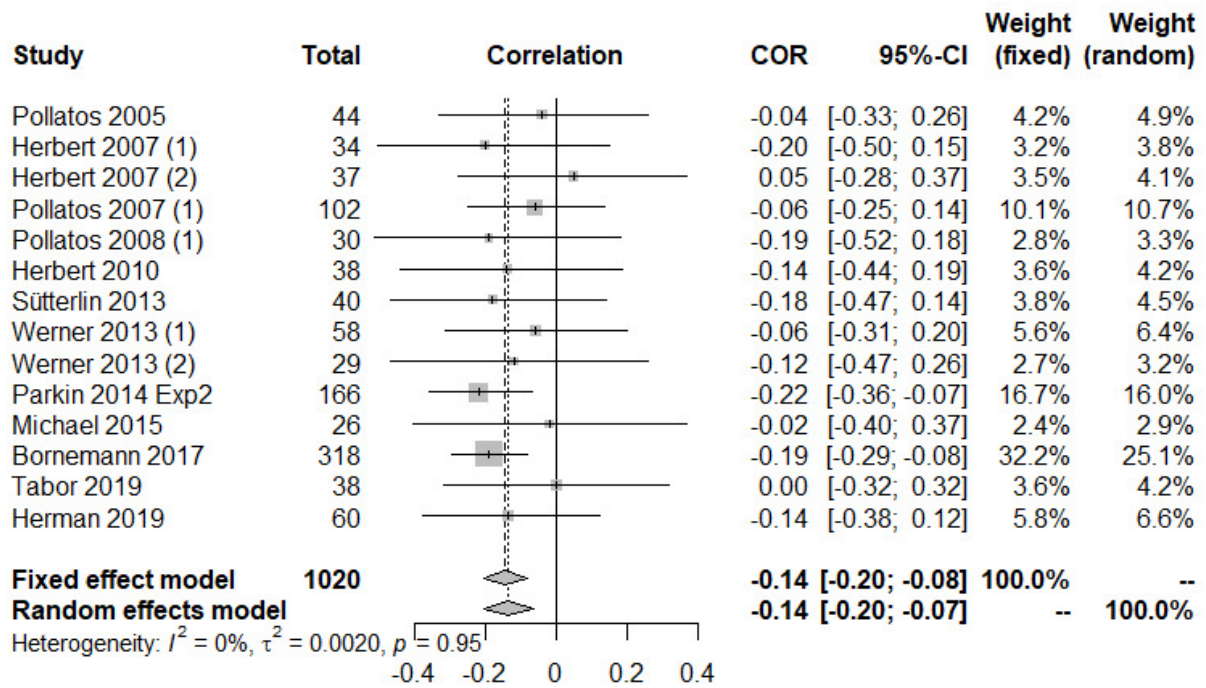
**Figure 7. Aggregated correlation between HCT performance and BMI**

*Note.* (1) = first study by the referred author; (2) = second study by the referred author; (3) third study by the referred author; (c) = clinical sample; (h) = healthy sample; (Norw) = Norwegian sample; (Hung) = Hungarian sample; Exp 1 = first experiment in the paper; Exp 2 = second experiment in the paper.



**Figure 8. Aggregated correlation between HCT performance and age**

*Note.* (1) = first study by the referred author; (2) = second study by the referred author; (3) = third study by the referred author; (c) = clinical sample; (h) = healthy sample; Exp 1 = first experiment in the paper.



**Figure 9. Aggregated correlation between HCT performance and sex**

Note. (1) = first study by the referred author; (2) = second study by the referred author; Exp 2 = second experiment in the paper.



## Discussion

This research addressed a simple but important question: is performance on the Heartbeat Counting Task (HCT) associated with risk factors and indicators of mental health? The results of this meta-analysis suggest a clear answer to this question for the (most studied) outcomes we considered here: some theoretically-expected associations were not reliably found and, when they were, they represented very small effect sizes. This is considered in the remainder of the discussion.

HCT performance was not significantly associated with psychological risk factors/indicators of mental health, i.e., trait anxiety, depression, and alexithymia. This finding is inconsistent with several theoretical assumptions. First, interoception is thought to play a central role in emotional experiences (Barrett, 2006; Damasio, 1994; Prinz, 2004). Relatedly, low access to internal signals may explain difficulties in identifying feelings (i.e., a key dimension of alexithymia), because the former is seen as a precondition for the latter (Craig, 2004). Conversely, higher awareness of bodily sensations, and catastrophic beliefs, are thought to predispose individuals to experiment anxiety (Domschke et al., 2010; Dunn et al., 2010). Finally, as depression is associated with dampened affective experience (e.g., anhedonia), it suggests the possibility that lower IAcc characterizes depressed individuals (Quadt, Critchley, & Garfinkel, 2018). Consistently with the above theoretical assumptions, the following statement was recently made: “links between interoception and emotion can be found in the relationship between interoceptive performance and experience of emotional intensity, the ability to recognize one’s own emotions (as measured in alexithymia scores) and IAcc, and emotion and interoception in psychopathology” (Quadt, Critchley, Garfinkel, et al., 2018, p. 138).

The present meta-analysis indicates that, when using the HCT, these theoretical assumptions are not supported. This could either suggest that current theoretical assumptions are incorrect (i.e., no relationship exists between IAcc and these mental health variables), that the HCT scores do not adequately capture individual differences in cardiac IAcc, or both. One way of answering this question is to examine previous studies that administered alternative IAcc tasks (e.g., the Heartbeat Discrimination Task or HDT; Whitehead & Drescher, 1980). However, as reported by others (Domschke et al., 2010; Eggart et al., 2019; Trevisan et al., 2019), the relationships between IAcc and depression, trait anxiety, and alexithymia have rarely been studied with other measures of IAcc than the HCT (e.g., Plans et al., 2021; Van Den Houte et al., 2021). Hence, it is currently not possible to know if the absence of significant association between these variables is due to the measurement limitations of the HCT (but see the section “implications for the HCT”). Future studies, using alternative tasks, are urgently needed to answer these crucial questions.

Regarding biological indicators/risk factors, HCT performance was significantly and negatively associated with heart rate and BMI. These results can either be seen as supporting the construct validity of the task (Ainley et al., 2020) or as additional evidence that the task lacks validity (Corneille et al., 2020). Specifically, as individuals with

higher heart rate (Schandry et al., 1993) and BMI (Rouse et al., 1988) receive cardiac signals with less intensity, heartbeats could be more difficult to detect for them, which could lower HCT performance; this would suggest that some participants perform the task by counting *felt* heartbeats (rather than guessing their heart rate), which supports HCT validity. However, this also indicates that HCT performance is influenced by signal intensity rather than (or in addition to) perceptual abilities (Corneille et al., 2020). If individuals perform better when the task becomes easier (i.e., when the signal gets stronger), this ambiguously speaks to their abilities. Please note, however, that other researchers may think otherwise, especially if an alternative definition of IAcc is endorsed (e.g., Schulz et al., 2021). The association between HCT and heart rate may also be spurious. A recent study showed that the correlation between HCT scores and heart rate becomes non-significant when time estimation and knowledge about heart rate are controlled for (Desmedt et al., 2020). A similar rationale could apply to the association between HCT performance and BMI.

Regarding sex, male participants had significantly higher HCT scores than female participants, which is consistent with previous suggestions (Rouse et al., 1988). This is also consistent with another recent meta-analysis showing that men have higher HCT, but also HDT, scores than females (Prentice & Murphy, 2022). Because the HDT is characterized by other methodological issues than the HCT, this suggests that sex differences observed with the HCT are not likely due to the measurement limitations of the HCT. However, a more recently developed measure of cardiac IAcc did not succeed to replicate these findings (Plans et al., 2021). Moreover, sex differences in other bodily domains (i.e., respiratory and gastro-intestinal) are far less compelling, especially given that few studies have tested them (Prentice & Murphy, 2022). This may suggest that, although a male advantage is observed in cardiac IAcc, this difference is not generalized to other bodily domains. Finally, current measures of IAcc conflate properties of the signal (e.g., the intensity of heartbeats) with the capacity to detect it (see Corneille et al., 2020). Men could be better at these cardiac tasks because the signal is easier to detect (e.g., they have higher blood pressure or lower body fat). Future studies are thus needed to examine if sex differences are (1) explained by measurement issues, (2) explained by physiological differences or the true capacity to detect internal signals, and (3) generalized to other bodily domains than the cardiac one.

A significant negative association was also found between HCT performance and age when publication bias was corrected or a third-level meta-analysis was performed. This result is consistent with a previous study that used the HDT (e.g., Khalsa et al., 2009). This is also consistent with many empirical studies showing that perceptual abilities (e.g., vision, hearing, taste, proprioception, pain perception, temperature perception, and thirst perception – notably, these three last perceptions are interoceptive under some definitions) decrease with age (Heft & Robinson, 2017; Murphy et al., 2018). However, the association observed between HCT performance and age in this meta-analysis could be explained by confounding variables such as heart rate estimation strategies and physiological char-

acteristics. This hypothesis has been tested by Murphy et al. (2018). They found that the relationship between HCT scores and age was partially mediated by BMI, but not by, e.g., resting heart rate, systolic blood pressure, time estimation, and beliefs about heart rate. These results suggest that age is associated with a decline in cardiac IAcc, but also an increase in BMI, the latter exacerbating differences in task performance. These authors, however, suggest that this relationship is not explained by individual differences in cardiac signal properties and estimation strategies.

More generally, associations between HCT performance and outcomes were small and could lack clinical/theoretical significance. The shared variance ranged from 0.01 (for alexithymia) to 2.9% (for heart rate) which is negligible. The correlation between HCT performance and depression represented 0.6% of shared variance (when one outlier was removed). This fails to support the relevance of clinical interventions targeting interoceptive abilities to improve depressive symptoms (de Jong et al., 2016). It should also be noted that these weak correlations may be alternatively explained by non-interoceptive processes. For instance, depressed individuals generally tend to underestimate time (Bschor et al., 2004), which could explain their lower HCT performance if they estimate time for completing the task (Desmedt et al., 2020).

The quality assessment indicates that research practice should be improved in several regards. Studies should systematically (1) report the validity and reliability of measures, (2) report (and standardize) HCT procedures including full instructions, time intervals, heartbeat measurement, and the formula used to compute HCT scores, (3) report and verify statistical assumptions, (4) report potential conflicts of interest, (5) perform power calculation, (6) perform a pre-registration, (7) make their data publicly available, and (8) make disclosure of full reporting. On a more positive note, the absence of an overall association between HCT performance and indicators of mental health goes hand in hand with an absence of publication bias.

## Implications for the HCT

Prominent researchers in the field see the HCT as a valid task, pointing out that “It is a well-validated measure, it has a good test–retest reliability, and it discriminates well between individuals. [...]” (Christensen et al., 2021, p. 3), and that “a considerable and well-examined body of physiological evidence in support of the HCT’s construct and criterion validity” (Ainley et al., 2020, p. 6).

Tempering this encouraging view, however, several important concerns have accumulated about this task’s validity. Early research pointed out that participants could perform the task by guessing their heart rate (i.e., counting at a pace that approximates their heart rate) rather than counting their felt heartbeats (Flynn & Clemens, 1988; Montgomery et al., 1984; Pennebaker, 1981). Since then, several empirical studies have supported the relevance of this concern. First, multiple studies showed that providing feedback to participants about their heart rate (i.e., influencing their knowledge about their heart rate) before the task impacted their performance on the task (Meyerholz et al., 2019;

Phillips et al., 1999; Ring et al., 2015; Ring & Brener, 1996).

Second, modifying participants’ heart rate with a pacemaker did not significantly change the reported number of heartbeats, indicating that these patients did not rely on actual heartbeats to perform the task (Windmann et al., 1999). Third, increasing cardiac signal intensity (via body posture changes) improved average HCT performance but only because actual heart rate decreased, while the number of reported heartbeats did not significantly change (Ring & Brener, 1996). Fourth, modifying instructions in order to prompt participants to only count their felt heartbeats (rather than guessing their heart rate) decreased average HCT performance by half (i.e., 50%; Desmedt et al., 2018). Finally, Legrand et al. (2022) found that lower HCT scores were related to a lower threshold (i.e., the tendency to underestimate one’s heart rate) in a new psychophysical measure of beliefs about heart rate. Overall, these results strongly suggest that HCT performance largely involves guessing strategies.

A recent study suggests that HCT validity may be improved by using modified instructions that prompt participants to only count their felt heartbeats (Desmedt et al., 2020). Indeed, HCT performance was less contaminated by guessing strategies (based on time estimation and knowledge about heart rate) when modified (vs. original) instructions were used. However, in addition to the contamination by guessing strategies, the HCT is also influenced by response bias (i.e., the tendency to report more or fewer heartbeats; Corneille et al., 2020; Zamariola et al., 2018). Given that most participants underestimate their actual heart rate, those reporting more heartbeats, independently of their interoceptive abilities, will have better HCT performance.

Interpretational concerns with the meaning and validity of HCT scores are also apparent when considering that HCT performance is associated with heart rate estimation strategies (Desmedt et al., 2020) and physiological variables (e.g., heart rate; Rouse et al., 1988; Zamariola et al., 2018). As a further significant concern, a recent meta-analysis showed that HCT performance is weakly associated ( $r = 0.21$ ,  $p < .001$ ,  $R^2 = 4.4\%$ ) with another common measure of cardiac IAcc (i.e., the Heartbeat Discrimination Task; Hickman et al., 2020). Finally, scores on this task vary in test-retest ( $< 6$  months) reliability ( $r = -0.41$  to  $0.81$ ; Murphy et al., 2019).

The failure to observe associations between the HCT and theoretically relevant mental health indicators in the current meta-analysis does not in and by itself allow us to conclude that the HCT is an invalid measure of interoceptive accuracy. However, it adds to current concerns about the validity of this task and poses a question mark on its pragmatic relevance (e.g., for diagnostic purposes).

## Limitations

First, only a subset of outcomes was considered in this meta-analysis. As a result, one should avoid overgeneralizing conclusions to the relationship between HCT performance and other indicators of mental health. Indeed, we decided to focus on the most studied associations (i.e., associations examined in at least 10 published studies). This allowed us to maximize the estimation accuracy of effect

sizes and the theoretical relevance of these associations (as researchers may investigate more those associations that they think are the most theoretically or pragmatically relevant). In doing so, we increased the probability of finding true effect sizes. Moreover, this choice may have facilitated a positivity bias as effects may be more studied/reported/published if they have a higher chance to be significant. Importantly too, we made our complete dataset (including 130 outcomes) available in open access. Researchers are encouraged to use our accessible dataset to test other associations or moderation effects. These tests should, however, be theoretically justified and pre-registered to decrease the probability of finding false positives.

Second, some of our meta-analyses were characterized by moderate to substantial heterogeneity. This high heterogeneity can prevent strong conclusions. This can be explained by the differences in, e.g., clinical disorders of participants, the female/male ratio, HCT procedure, outcome measurement, and statistical analyses (i.e., correlation vs. mean difference) between studies. This limit has been partly solved by testing moderation effects for some outcomes. However, we were unable to test the effect of important moderators such as HCT instructions and outcome measurement for some or all outcomes. As we noted in the introduction, a simple change in HCT instructions can dramatically impact average HCT performance: asking participants to only count their felt heartbeats and avoid guessing their heart rate decreases HCT performance by half (Desmedt et al., 2018) and significantly decreases the correlation between heartbeat counting and estimation strategies (Desmedt et al., 2020). It would thus be interesting to test if HCT performance is associated with mental health outcomes when modified instructions are used. It is nevertheless important to note that these associations could still be explained by confounding variables such as response bias (Corneille et al., 2020) and cardiac signal intensity (Schandry et al., 1993).

Associations may also depend on outcomes measurement. Indeed, Carlson and Herdman (2012) showed that insufficient correlation between two measures of the same construct can lead to very heterogeneous findings (i.e., poor replicability) across measures. For instance, trait anxiety was assessed with different questionnaires such as State-Trait Anxiety Inventory (STAI), Beck Anxiety Inventory (BAI), and anxiety subscale of the Depression Anxiety and Stress Scale (DASS-A). While DASS-A and BAI highly correlate with each other ( $N = 717$ ,  $r = 0.81$ ; Lovibond & Lovibond, 1995), STAI only moderately correlates with BAI ( $N = 191$ ;  $r = 0.66$ ; Kohn et al., 2008) and DASS-A ( $N = 567$ ,  $r = 0.48$ ; Grös et al., 2007). This means that results may be relatively consistent between DAI-A and BAI, but inconsistent between DAI-A/BAI and STAI-T (Carlson & Herdman, 2012). However, our moderation analyses showed that our conclusions did not change depending on the questionnaire used (i.e., STAI vs. other anxiety measures, BDI-II vs. other depression measures, TAS-20 vs. other alexithymia measures).

Differences in HCT procedures may also be critical to observing associations between HCT performance and mental health outcomes. This question could not be investigated here, either due to small study samples or to a lack of

reporting in the procedural details. We also call for better standardization and reporting in the measurement procedures (see also Corneille et al., 2020).

In general, however, we urge caution when examining moderated associations. Specifically, we encourage interoception researchers to look at theoretically grounded (and ideally pre-registered) moderators and to correct for multiple tests. A wild examination of moderated associations between HCT and mental health outcomes is likely to prevent rather than advance research efforts (see Mierop et al., 2020, for a recent discussion in the context of intranasal oxytocin research). Likewise, although non-linear (e.g., quadratic, cubic) relationships might exist between HCT performance and some mental health indicators, we consider it an empirical question that should be pre-registered and tested in future studies.

Third, given the heterogeneity in the selection of covariates across studies, we did not (and actually could not) include covariates in our meta-analysis. This precluded finding associations that might have otherwise been found. For instance, a recent study has shown that HCT performance is only significantly associated with alexithymia when confounding variables are controlled for (Murphy et al., 2018). Unfortunately, theoretical consensus and empirical evidence are still lacking regarding which covariates should be considered in analyses. Of particular concern here is that spurious correlations may emerge when running a variety of complex multiple regression analyses. Moreover, while statistical control can help remove non-relevant variance, it also has the potential to remove variance of interest, making it difficult to interpret findings (Lynam et al., 2006).

Fourth, apart from the HCT, included studies contained methodological limitations that may hinder the possibility to find associations between HCT performance and mental health indicators/risk factors. In particular, some studies applied pre-screening that might have restricted the range of values of HCT scores and outcomes. Among these pre-screening procedures, we can cite the exclusion of participants with a history of mental disorders or medication. Besides this limit, we can also question the validity/reliability of some outcome measures. For instance, it is noteworthy that STAI was the most frequently used measure of trait anxiety whereas it may not represent the best choice for testing the relationship between HCT performance and trait anxiety. STAI may have low discriminant validity as it blends anxious and depressive symptomatology (Balsamo et al., 2013). Considering that IAcc is assumed to be negatively associated with depression but positively associated with anxiety, a measure mixing these constructs could prevent detecting more targeted associations. Future studies should examine the relationship between IAcc and trait anxiety with questionnaires that are less contaminated by depressive symptomatology.

## Conclusion

As it appears, there are various reasons why the present meta-analysis did not find meaningful associations between the HCT and the included indicator/risk factor of mental health: (1) associations are to be found with other indicators, (2) theories should be qualified (e.g., by assum-



ing non-linear associations), (3) relevant moderators should be considered, (4) the published studies suffered from methodological limitations (e.g., irrelevant pre-screening and low validity/reliability of some outcomes measurement), and (5) the HCT lacks validity. The current study cannot provide definite answers regarding which of these non-mutually exclusive factors account for the absence of effects. However, it highlights that theories or measures should be advanced to allow for strong a priori tests regarding any such associations.

### Contributions

Contributed to conception and design: OD, MV, MW, SD, OL, OC.

Contributed to acquisition of data: OD, MV, MW, SD.

Contributed to analysis and interpretation of data: OD, MV, MW, OL, OC.

Drafted and/or revised the article: OD, MV, MW, SD, OL, OC.

Approved the submitted version for publication: OD, MV, MW, SD, OL, OC.

### Competing Interests

The authors have no relevant interests to declare.

### Acknowledgments

The protocol was registered in PROSPERO (registration number: CRD42019142176) before the review was conducted. We thank Alexandre Heeren for his early feedback on the manuscript.

### Funding information

Olivier Desmedt (Ph.D. student, grant number: 34226579), Maaïke Van Den Houte (PostDoc; grant # T.0114.18.), Marta Walentynowicz (PostDoc; Grant # T.0114.18.), and Olivier Luminet (Research Director) are funded by the Fund for Scientific Research – Belgium (FRS-FNRS).

Submitted: September 23, 2021 PDT, Accepted: February 28, 2022 PDT



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at <http://creativecommons.org/licenses/by/4.0> and legal code at <http://creativecommons.org/licenses/by/4.0/legalcode> for more information.

## References

- Ainley, V., Tsakiris, M., Pollatos, O., Schulz, A., & Herbert, B. M. (2020). Comment on “Zamariola et al. (2018), Interoceptive Accuracy Scores are Problematic: Evidence from Simple Bivariate Correlations”—The empirical data base, the conceptual reasoning and the analysis behind this statement are misconceived and do not support the authors’ conclusions. *Biological Psychology*, 152, 107870. <https://doi.org/10.1016/j.biopsycho.2020.107870>
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*.
- Balsamo, M., Romanelli, R., Innamorati, M., Ciccarese, G., Carlucci, L., & Saggino, A. (2013). The state-trait anxiety inventory: Shadows and lights on its construct validity. *Journal of Psychopathology and Behavioral Assessment*, 35(4), 475–486. <https://doi.org/10.1007/s10862-013-9354-5>
- Barrett, L. F. (2006). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review*, 10(1), 20–46. [https://doi.org/10.1207/s15327957pspr1001\\_2](https://doi.org/10.1207/s15327957pspr1001_2)
- Barrett, L. F., Quigley, K. S., & Hamilton, P. (2016). An active inference theory of allostasis and interoception in depression. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1708), 20160011. <https://doi.org/10.1098/rstb.2016.0011>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2011). *Introduction to Meta-Analysis*. John Wiley & Sons.
- Bornemann, B., & Singer, T. (2017). Taking time to feel our body: Steady increases in heartbeat perception accuracy and decreases in alexithymia over 9 months of contemplative mental training. *Psychophysiology*, 54(3), 469–482. <https://doi.org/10.1111/psyp.12790>
- Bschor, T., Ising, M., Bauer, M., Lewitzka, U., Skerstupeit, M., Müller-Oerlinghausen, B., & Baethge, C. (2004). Time experience and time judgment in major depression, mania and healthy subjects. A controlled study of 93 subjects. *Acta Psychiatrica Scandinavica*, 109(3), 222–229. <https://doi.org/10.1046/j.0001-690x.2003.00244.x>
- Carlson, K. D., & Herdman, A. O. (2012). Understanding the Impact of Convergent Validity on Research Results. *Organizational Research Methods*, 15(1), 17–32. <https://doi.org/10.1177/1094428110392383>
- Christensen, J. F., Azevedo, R. T., & Tsakiris, M. (2021). Emotion matters: Different psychophysiological responses to expressive and non-expressive full-body movements. *Acta Psychologica*, 212, 103215. <https://doi.org/10.1016/j.actpsy.2020.103215>
- Corneille, O., Desmedt, O., Zamariola, G., Luminet, O., & Muraige, P. (2020). A heartfelt response to Zimprich et al. (2020), and Ainley et al. (2020)’s commentaries: Acknowledging issues with the HCT would benefit interoception research. *Biological Psychology*, 152, 107869. <https://doi.org/10.1016/j.biopsycho.2020.107869>
- Craig, A. D. (2004). Human feelings: Why are some more aware than others? *Trends in Cognitive Sciences*, 8(6), 239–241. <https://doi.org/10.1016/j.tics.2004.04.004>
- Craig, A. D. (2015). *How do you feel?: An interoceptive moment with your neurobiological self*. Princeton University Press. <https://doi.org/10.1515/9781400852727>
- Dale, A., & Anderson, D. (1978). Information Variables in Voluntary Control and Classical Conditioning of Heart Rate: Field Dependence and Heart-Rate Perception. *Perceptual and Motor Skills*, 47(1), 79–85. <https://doi.org/10.2466/pms.1978.47.1.79>
- Damasio, A. R. (1994). Descartes’ error and the future of human life. *Scientific American*, 271(4), 144–144. <https://doi.org/10.1038/scientificamerican1094-144>
- de Jong, M., Lazar, S. W., Hug, K., Mehling, W. E., Hölzel, B. K., Sack, A. T., Peeters, F., Ashih, H., Mischoulon, D., & Gard, T. (2016). Effects of Mindfulness-Based Cognitive Therapy on Body Awareness in Patients with Chronic Pain and Comorbid Depression. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.00967>
- Desmedt, O., Corneille, O., Luminet, O., Murphy, J., Bird, G., & Muraige, P. (2020). Contribution of Time Estimation and Knowledge to Heartbeat Counting Task Performance under Original and Adapted Instructions. *Biological Psychology*, 154, 107904. <https://doi.org/10.1016/j.biopsycho.2020.107904>
- Desmedt, O., Luminet, O., & Corneille, O. (2018). The heartbeat counting task largely involves non-interoceptive processes: Evidence from both the original and an adapted counting task. *Biological Psychology*, 138, 185–188. <https://doi.org/10.1016/j.biopsycho.2018.09.004>
- Domschke, K., Stevens, S., Pfleiderer, B., & Gerlach, A. L. (2010). Interoceptive sensitivity in anxiety and anxiety disorders: An overview and integration of neurobiological findings. *Clinical Psychology Review*, 30(1), 1–11. <https://doi.org/10.1016/j.cpr.2009.08.008>
- Dunn, B. D., Stefanovitch, I., Evans, D., Oliver, C., Hawkins, A., & Dalgleish, T. (2010). Can you feel the beat? Interoceptive awareness is an interactive function of anxiety- and depression-specific symptom dimensions. *Behaviour Research and Therapy*, 48(11), 1133–1138. <https://doi.org/10.1016/j.brat.2010.07.006>
- Duval, S., & Tweedie, R. (2000). A Nonparametric “Trim and Fill” Method of Accounting for Publication Bias in Meta-Analysis. *Journal of the American Statistical Association*, 95(449), 89–98. <https://doi.org/10.1080/01621459.2000.10473905>
- Eggart, M., Lange, A., Binser, M., Queri, S., & Müller-Oerlinghausen, B. (2019). Major depressive disorder is associated with impaired interoceptive accuracy: A systematic review. *Brain Sciences*, 9(6), 131. <https://doi.org/10.3390/brainsci9060131>
- Flynn, D. M., & Clemens, W. J. (1988). On the Validity of Heartbeat Tracking Tasks. *Psychophysiology*, 25(1), 92–96. <https://doi.org/10.1111/j.1469-8986.1988.tb00965.x>

- Garfinkel, S. N., Seth, A. K., Barrett, A. B., Suzuki, K., & Critchley, H. D. (2015). Knowing your own heart: Distinguishing interoceptive accuracy from interoceptive awareness. *Biological Psychology*, 104, 65–74. <https://doi.org/10.1016/j.biopsycho.2014.11.004>
- Grös, D. F., Antony, M. M., Simms, L. J., & McCabe, R. E. (2007). Psychometric properties of the State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA): Comparison to the State-Trait Anxiety Inventory (STAI). *Psychological Assessment*, 19(4), 369–381. <https://doi.org/10.1037/1040-3590.19.4.369>
- Heft, M. W., & Robinson, M. E. (2017). Somatosensory function in old age. *Journal of Oral Rehabilitation*, 44(4), 327–332. <https://doi.org/10.1111/joor.12488>
- Herbert, B. M., & Pollatos, O. (2012). The body in the mind: On the relationship between interoception and embodiment. *Topics in Cognitive Science*, 4(4), 692–704. <https://doi.org/10.1111/j.1756-8765.2012.01189.x>
- Hickman, L., Seyedsalehi, A., Cook, J. L., Bird, G., & Murphy, J. (2020). The relationship between heartbeat counting and heartbeat discrimination: A meta-analysis. *Biological Psychology*, 156, 107949. <https://doi.org/10.1016/j.biopsycho.2020.107949>
- Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21(11), 1539–1558. <https://doi.org/10.1002/sim.1186>
- Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. (Eds.). (2019). *Cochrane handbook for systematic reviews of interventions*. John Wiley & Sons.
- Horváth, Á., Ferentzi, E., Bogdány, T., Szolcsányi, T., Witthöft, M., & Köteles, F. (2020). Proprioception but not cardiac interoception is related to the rubber hand illusion. *Cortex*, 132, 361–373. <https://doi.org/10.1016/j.cortex.2020.08.026>
- Jakubczyk, A., Wiśniewski, P., Trucco, E. M., Kobyliński, P., Zaorska, J., Skrzyszewski, J., Suszek, H., Wojnar, M., & Kopera, M. (2021). The synergistic effect between interoceptive accuracy and alcohol use disorder status on pain sensitivity. *Addictive Behaviors*, 112, 106607. <https://doi.org/10.1016/j.addbeh.2020.106607>
- Khalsa, S. S., Adolphs, R., Cameron, O. G., Critchley, H. D., Davenport, P. W., Feinstein, J. S., Feusner, J. D., Garfinkel, S. N., Lane, R. D., Mehling, W. E., Meuret, A. E., Nemeroff, C. B., Oppenheimer, S., Petzschner, F. H., Pollatos, O., Rhudy, J. L., Schramm, L. P., Simmons, W. K., Stein, M. B., ... Zucker, N. (2018). Interoception and mental health: A roadmap. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 3(6), 501–513. <https://doi.org/10.1016/j.bpsc.2017.12.004>
- Khalsa, S. S., Rudrauf, D., & Tranel, D. (2009). Interoceptive awareness declines with age. *Psychophysiology*, 46(6), 1130–1136. <https://doi.org/10.1111/j.1469-8986.2009.00859.x>
- Kohn, P. M., Kantor, L., DeCicco, T. L., & Beck, A. T. (2008). The Beck Anxiety Inventory–Trait (BAIT): A Measure of Dispositional Anxiety Not Contaminated by Dispositional Depression. *Journal of Personality Assessment*, 90(5), 499–506. <https://doi.org/10.1080/0223890802248844>
- Legrand, N., Nikolova, N., Correa, C., Brændholt, M., Stuckert, A., Kildahl, N., Vejlbø, M., Fardo, F., & Allen, M. (2022). The heart rate discrimination task: A psychophysical method to estimate the accuracy and precision of interoceptive beliefs. *Biological Psychology*, 168, 108239. <https://doi.org/10.1016/j.biopsycho.2021.108239>
- Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy*, 33(3), 335–343. [https://doi.org/10.1016/0005-7967\(94\)00075-u](https://doi.org/10.1016/0005-7967(94)00075-u)
- Lynam, D. R., Hoyle, R. H., & Newman, J. P. (2006). The perils of partialling: Cautionary tales from aggression and psychopathy. *Assessment*, 13(3), 328–341. <https://doi.org/10.1177/1073191106290562>
- Meyerholz, L., Irzinger, J., Witthöft, M., Gerlach, A. L., & Pohl, A. (2019). Contingent biofeedback outperforms other methods to enhance the accuracy of cardiac interoception: A comparison of short interventions. *Journal of Behavior Therapy and Experimental Psychiatry*, 63, 12–20. <https://doi.org/10.1016/j.jbtep.2018.12.002>
- Mierop, A., Mikolajczak, M., Stahl, C., Béna, J., Luminet, O., Lane, A., & Corneille, O. (2020). How Can Intranasal Oxytocin Research Be Trusted? A Systematic Review of the Interactive Effects of Intranasal Oxytocin on Psychosocial Outcomes. *Perspectives on Psychological Science*, 15(5), 1228–1242. <https://doi.org/10.1177/1745691620921525>
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLOS Medicine*, 6(7), e1000097. <https://doi.org/10.1371/journal.pmed.1000097>
- Montgomery, W. A., Jones, G. E., & Hollandsworth, J. G., Jr. (1984). The effects of physical fitness and exercise on cardiac awareness. *Biological Psychology*, 18(1), 11–22. [https://doi.org/10.1016/0301-0511\(84\)90022-x](https://doi.org/10.1016/0301-0511(84)90022-x)
- Murphy, J., Cheesman, R., Gregory, A. M., Lau, J., Ehlers, A., Catmur, C., Bird, G., & Eley, T. C. (2019). Estimating the stability of heartbeat counting in middle childhood: A twin study. *Biological Psychology*, 148, 107764. <https://doi.org/10.1016/j.biopsycho.2019.107764>
- Murphy, J., Geary, H., Millgate, E., Catmur, C., & Bird, G. (2018). Direct and indirect effects of age on interoceptive accuracy and awareness across the adult lifespan. *Psychonomic Bulletin & Review*, 25(3), 1193–1202. <https://doi.org/10.3758/s13423-017-1339-z>



- National Institutes of Health. (2014a). *Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies*. <https://www.nhlbi.nih.gov/health-pro/guidelines/in-develop/cardiovascular-risk-reduction/tools/cohort>
- National Institutes of Health. (2014b). *Quality Assessment Tool of Case-control Studies*. <https://www.nhlbi.nih.gov/health-pro/guidelines/in-develop/cardiovascular-risk-reduction/tools/cohort>
- Paulus, M. P., & Stein, M. B. (2006). An insular view of anxiety. *Biological Psychiatry*, 60(4), 383–387. <https://doi.org/10.1016/j.biopsych.2006.03.042>
- Pennebaker, J. W. (1981). Stimulus characteristics influencing estimation of heart rate. *Psychophysiology*, 18(5), 540–548. <https://doi.org/10.1111/j.1469-8986.1981.tb01824.x>
- Phillips, G. C., Jones, G. E., Rieger, E. J., & Snell, J. B. (1999). Effects of the presentation of false heart-rate feedback on the performance of two common heartbeat-detection tasks. *Psychophysiology*, 36(4), 504–510. <https://doi.org/10.1017/s0048577299980071>
- Plans, D., Ponzo, S., Morelli, D., Cairo, M., Ring, C., Keating, C. T., Cunningham, A. C., Catmur, C., Murphy, J., & Bird, G. (2021). Measuring interoception: The phase adjustment task. *Biological Psychology*, 165, 108171. <https://doi.org/10.1016/j.biopsycho.2021.108171>
- Prentice, F., & Murphy, J. (2022). Sex differences in interoceptive accuracy: A meta-analysis. *Neuroscience & Biobehavioral Reviews*, 132, 497–518. <https://doi.org/10.1016/j.neubiorev.2021.11.030>
- Prinz, J. J. (2004). *Gut reactions: A perceptual theory of emotion*. Oxford University Press.
- Quadt, L., Critchley, H. D., & Garfinkel, S. N. (2018). The neurobiology of interoception in health and disease. *Annals of the New York Academy of Sciences*, 1428(1), 112–128. <https://doi.org/10.1111/nyas.13915>
- Quadt, L., Critchley, H. D., Garfinkel, S. N., Tsakiris, M., & De Preester, H. (2018). Interoception and emotion: Shared mechanisms and clinical implications. *The Interoceptive Mind: From Homeostasis to Awareness*, 123.
- Ring, C., & Brener, J. (1996). Influence of beliefs about heart rate and actual heart rate on heartbeat counting. *Psychophysiology*, 33(5), 541–546. <https://doi.org/10.1111/j.1469-8986.1996.tb02430.x>
- Ring, C., Brener, J., Knapp, K., & Mailloux, J. (2015). Effects of heartbeat feedback on beliefs about heart rate and heartbeat counting: A cautionary tale about interoceptive awareness. *Biological Psychology*, 104, 193–198. <https://doi.org/10.1016/j.biopsycho.2014.12.010>
- Rouse, C. H., Jones, G. E., & Jones, K. R. (1988). The Effect of Body Composition and Gender on Cardiac Awareness. *Psychophysiology*, 25(4), 400–407. <https://doi.org/10.1111/j.1469-8986.1988.tb01876.x>
- Schandry, R. (1981). Heart beat perception and emotional experience. *Psychophysiology*, 18(4), 483–488. <https://doi.org/10.1111/j.1469-8986.1981.tb02486.x>
- Schandry, R., Bestler, M., & Montoya, P. (1993). On the relation between cardiodynamics and heartbeat perception. *Psychophysiology*, 30(5), 467–474. <https://doi.org/10.1111/j.1469-8986.1993.tb02070.x>
- Schulz, A., Back, S. N., Schaan, V. K., Bertsch, K., & Vögele, C. (2021). On the construct validity of interoceptive accuracy based on heartbeat counting: Cardiovascular determinants of absolute and tilt-induced change scores. *Biological Psychology*, 164, 108168. <https://doi.org/10.1016/j.biopsycho.2021.108168>
- Simmons, W. K., & DeVille, D. C. (2017). Interoceptive contributions to healthy eating and obesity. *Current Opinion in Psychology*, 17, 106–112. <https://doi.org/10.1016/j.copsyc.2017.07.001>
- Sugawara, A., Terasawa, Y., Katsunuma, R., & Sekiguchi, A. (2020). Effects of interoceptive training on decision making, anxiety, and somatic symptoms. *BioPsychoSocial Medicine*, 14(1), 13030–13020. <https://doi.org/10.1186/s13030-020-00179-7>
- Terhaar, J., Viola, F. C., Bär, K.-J., & Debener, S. (2012). Heartbeat evoked potentials mirror altered body perception in depressed patients. *Clinical Neurophysiology*, 123(10), 1950–1957. <https://doi.org/10.1016/j.clinph.2012.02.086>
- Trevisan, D. A., Altschuler, M. R., Bagdasarov, A., Carlos, C., Duan, S., Hamo, E., Kala, S., McNair, M. L., Parker, T., Stahl, D., Winkelman, T., Zhou, M., & McPartland, J. C. (2019). A meta-analysis on the relationship between interoceptive awareness and alexithymia: Distinguishing interoceptive accuracy and sensibility. *Journal of Abnormal Psychology*, 128(8), 765–776. <https://doi.org/10.1037/abn0000454>
- Van Den Houte, M., Vlemincx, E., Franssen, M., Van Diest, I., Van Oudenhove, L., & Luminet, O. (2021). The respiratory occlusion discrimination task: A new paradigm to measure respiratory interoceptive accuracy. *Psychophysiology*, 58(4), 13760. <https://doi.org/10.1111/psyp.13760>
- Verdejo-Garcia, A., Clark, L., & Dunn, B. D. (2012). The role of interoception in addiction: A critical review. *Neuroscience & Biobehavioral Reviews*, 36(8), 1857–1869. <https://doi.org/10.1016/j.neubiorev.2012.05.007>
- Viechtbauer, W., & Cheung, M. W.-L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, 1(2), 112–125. <https://doi.org/10.1002/jrsm.11>
- Whitehead, W. E., & Drescher, V. M. (1980). Perception of gastric contractions and self-control of gastric motility. *Psychophysiology*, 17(6), 552–558. <https://doi.org/10.1111/j.1469-8986.1980.tb02296.x>
- Windmann, S., Schonecke, O. W., Fröhlig, G., & Maldener, G. (1999). Dissociating beliefs about heart rates and actual heart rates in patients with cardiac pacemakers. *Psychophysiology*, 36(3), 339–342. <https://doi.org/10.1017/s0048577299980381>
- Zamariola, G., Maurage, P., Luminet, O., & Corneille, O. (2018). Interoceptive accuracy scores from the heartbeat counting task are problematic: Evidence from simple bivariate correlations. *Biological Psychology*, 137, 12–17. <https://doi.org/10.1016/j.biopsycho.2018.06.006>

## Supplementary Materials

### Peer Review History

Download: [https://collabra.scholasticahq.com/article/33271-how-does-heartbeat-counting-task-performance-relate-to-theoretically-relevant-mental-health-outcomes-a-meta-analysis/attachment/84513.docx?auth\\_token=N2GBbZK1maUb0D6rZoAG](https://collabra.scholasticahq.com/article/33271-how-does-heartbeat-counting-task-performance-relate-to-theoretically-relevant-mental-health-outcomes-a-meta-analysis/attachment/84513.docx?auth_token=N2GBbZK1maUb0D6rZoAG)

---

### References of Included Studies

Download: [https://collabra.scholasticahq.com/article/33271-how-does-heartbeat-counting-task-performance-relate-to-theoretically-relevant-mental-health-outcomes-a-meta-analysis/attachment/84514.docx?auth\\_token=N2GBbZK1maUb0D6rZoAG](https://collabra.scholasticahq.com/article/33271-how-does-heartbeat-counting-task-performance-relate-to-theoretically-relevant-mental-health-outcomes-a-meta-analysis/attachment/84514.docx?auth_token=N2GBbZK1maUb0D6rZoAG)

---

### PRISMA 2020 Checklist

Download: [https://collabra.scholasticahq.com/article/33271-how-does-heartbeat-counting-task-performance-relate-to-theoretically-relevant-mental-health-outcomes-a-meta-analysis/attachment/84515.docx?auth\\_token=N2GBbZK1maUb0D6rZoAG](https://collabra.scholasticahq.com/article/33271-how-does-heartbeat-counting-task-performance-relate-to-theoretically-relevant-mental-health-outcomes-a-meta-analysis/attachment/84515.docx?auth_token=N2GBbZK1maUb0D6rZoAG)

---