A Systematic Review and Meta-analysis on the Association Between Heartbeat Counting Task
Performance and Mental Disorders and Their Risk Factors Among Adults

Olivier Desmedt^{1,2}, Maaïke Van Den Houte^{1,2,3}, Marta Walentynowicz^{1,2,4}, Sarah Dekeyser¹,
Olivier Luminet^{1,2}, & Olivier Corneille¹

¹Psychological Science Research Institute, UCLouvain, Belgium

²Fund for Scientific Research – Belgium (FRS-FNRS)

³Translational Research Center for Gastrointestinal Disorders (TARGID), Department of Chronic
Diseases, Metabolism and Ageing, University of Leuven, Belgium

⁴Centre for the Psychology of Learning and Experimental Psychopathology, University of Leuven,
Leuven, Belgium

This is a preprint of a manuscript in preparation.

Author Note

The authors have no relevant interests to declare. Olivier Desmedt (Ph.D. student),
Maaïke Van Den Houte (PostDoc), Marta Walentynowicz (PostDoc), and Olivier Luminet
(Research Director) are funded by the Fund for Scientific Research – Belgium (FRS-FNRS).

Correspondence concerning this manuscript should be addressed to Olivier Desmedt,
Psychological Science Research Institute, UCLouvain, 10 Place du Cardinal Mercier, Louvain-
la-Neuve, B-1348, Belgium. Email: olivier.desmedt@uclouvain.be

Abstract

Interoceptive accuracy is thought to play a key role in mental health. However, the validity of its most frequently used measure (i.e., the Heartbeat Counting Task; HCT) has been questioned. This calls for a meta-analytic examination of associations between HCT performance and mental health. To this end, we performed a systematic review and meta-analysis on associations between HCT performance and mental disorders, symptoms, and their risk factors among adult participants. The final data set comprised 133 studies with 11 524 participants. The quality assessment indicates that research practice should be improved as, e.g., only a few studies reported sample size justification (6%), pre-registration (0.8%), and data in free access (6.8%). HCT performance was not significantly associated with trait measures of anxiety ($r = 0.03$), depression ($r = -0.04$), and alexithymia ($r = -0.01$). It was weakly and negatively associated with age ($r = -0.11$) after correction for publication bias, sex (male > female; $r = -0.14$), heart rate ($r = -0.17$), and body mass index ($r = -0.11$). That theoretically assumed associations between interoceptive accuracy and mental health indicators are not meta-analytically found raises further doubts about the validity, and even the very pragmatic value, of the HCT.

Keywords: interoception, interoceptive accuracy, heartbeat counting task, mental health

A Systematic Review and Meta-analysis on the Association Between Heartbeat Counting Task Performance and Mental Disorders and Their Risk Factors Among Adults.

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). Deficits in interoceptive 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, Quigley, & Hamilton, 2016), and problematic health conditions (e.g., obesity; Simmons & DeVille, 2017). Empirical studies, although inconsistent, suggest an association between interoceptive dimensions and mental/physical health indicators (Khalsa et al., 2017). Interventions are, therefore, implemented to improve interoceptive abilities (Khoury, Lutz, & Schuman-Olivier, 2018).

Interoceptive accuracy (IAcc), the objective capacity to detect internal signals, is considered a critical factor in interoception research (Garfinkel, Seth, Barrett, Suzuki, & Critchley, 2015), and it 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 on Figure 1, the HCT has been increasingly used and referred to in recent years.

Figure 1. Chart illustrating the number of times the HCT original paper (Schandry, 1981) has been cited each year since its publication.

The HCT has been criticized, however, for its lack of 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, Jones, & Hollandsworth, 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 during the task (Phillips, Jones, Rieger, & Snell, 1999; Ring & Brener, 1996; Ring, Brener, Knapp, & Mailloux, 2015, but see Meyerholz, Irzinger, Witthöft, Gerlach, & Pohl, 2019). Second, modifying participants' heart

1 rate with a pacemaker did not significantly change the reported number of heartbeats, indicating
2 that these patients did not rely on actual heartbeats to perform the task (Windmann, Schonecke,
3 Fröhlig, & Maldener, 1999). Third, increasing cardiac signal intensity (via body posture changes)
4 improved average HCT performance but only because actual heart rate decreased, as the number
5 of reported heartbeats did not significantly change (Ring & Brener, 1996). Finally, modifying
6 instructions in order to prompt participants to only count their felt heartbeats (rather than
7 guessing their heart rate) decreased average HCT performance by half (i.e., 50%) (Desmedt,
8 Luminet, & Corneille, 2018). These results strongly suggest that HCT performance largely
9 involves non-interoceptive guessing strategies.

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

19 Validity issues with the HCT are also apparent when considering that HCT performance
20 is associated with heart rate estimation strategies (Desmedt et al., 2020) and physiological
21 variables (e.g., heart rate; Rouse, Jones, & Jones, 1988; Zamariola et al., 2018). As a further
22 significant concern, a recent meta-analysis showed that HCT performance is weakly associated
23 ($r = 0.21$, $p < .001$, $R^2 = 4.4\%$) with another common measure of cardiac IAcc (i.e., the

Heartbeat Discrimination Task; Hickman, Seyedsalehi, Cook, Bird, & Murphy, 2020). Finally, scores on this task vary in test-retest (< 6 months) reliability ($r = \sim 0.41$ to 0.81 ; Murphy et al., 2019).

The Present Review

Considering that the HCT is facing such important validity issues, a fair and important question is: Why should significant associations be found between HCT performance and mental health and its risk factors among adults? We anticipated three possibilities: (1) contrary to widespread claims, no such associations are found across studies, (2) associations, when found, reflect a publication bias, and (3) reliable associations are found that remain to be theorized. Informing these questions is of primary importance considering that human and financial resources are invested in studies administering this task (Horváth et al., 2020; Jakubczyk et al., 2021; Ogishima, Maeda, Tanaka, & Shimada, 2020), theoretical models are partly built on HCT outcomes (Harshaw, 2015; Herbert & Pollatos, 2012; Verdejo-Garcia, Clark, & Dunn, 2012), and HCT research keeps inspiring health interventions (Bornemann & Singer, 2017; Sugawara, Terasawa, Katsunuma, & Sekiguchi, 2020).

To answer these questions, we performed a pre-registered systematic review and meta-analysis on the association between the HCT performance and mental disorders, symptoms, and risk factors.

Method

This systematic review and meta-analysis were conducted in accordance with PRISMA¹ statements (Moher, Liberati, Tetzlaff, Altman, & Group, 2009). The protocol was registered in

---

¹ The PRISMA checklist can be found at the following address:
https://osf.io/3cwt4/?view_only=7e466bd994134dc787cc6a778f1d0723.

PROSPERO (registration number: CRD42019142176). The Covidence platform (www.covidence.org) was used to perform every step of the systematic review.

Eligibility Criteria

We included studies with adult participants (18+) and investigating the association between HCT performance and mental disorders (e.g., major depressive disorder, substance use disorder) as well as their risk factors (e.g., alexithymia, age, sex) and symptoms (e.g., apathy, body mass index, body dissatisfaction). We chose to rely on the DSM-V criteria (American Psychiatric Association, 2013) for including studies in our systematic review in order to have objective selection criteria allowing for high 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 9th, 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 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 meet 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, 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. Other types of effect sizes (e.g., *Cohen's d*) were converted.

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

We performed random-effects model meta-analyses² as the true effect could vary between studies. To do so, we used meta, metaphor, and dmetar R packages. As 130 outcomes were correlated 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

---

² 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.

effect size, and (ii) this is the minimum number of studies for performing funnel plot asymmetry tests (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 three broad categories: mental health (trait anxiety, depression, and alexithymia), physical health (heart rate and BMI), and sociodemographic variables (age and sex). 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.

Heterogeneity was assessed by the Cochran Q test and the I^2 test (Borenstein, Hedges, Higgins, & Rothstein, 2011; 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 population and measurement 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. 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. The final data set comprised of 133 studies with 11 524 participants.

Selection Process

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

Quality Assessment and Research Practice

Most studies adequately reported the research question and objectives (100%), inclusion and exclusion criteria when recruiting participants (75.2%), definition of the independent variable(s) (91.7%), and 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%).

Meta-analysis

Results are organized into three categories: mental health, physical health, and sociodemographic variables. The last two categories are considered risk factors of mental disorders in the DSM-V.

Mental health.

HCT performance was not significantly associated with trait anxiety ($r = 0.03$, 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 influential case biasing 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.

1

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

1

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

Figure 5. Aggregated correlation between HCT performance and alexithymia.

Physical health.

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) and BMI ($N = 29$, $r = -0.11$, 95% CI $[-0.18$ to $-0.04]$, $p = 0.002$; see Figure 7). There was moderate between-study heterogeneity for both outcomes ($I^2 = 48.5\%$, $Q(39) = 75.75$, $p < 0.001$ and $I^2 = 59.2$, $Q(28) = 68.63$, $p < 0.001$, respectively). Removing outliers did not substantially influence the results. We then 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

1 measurement type (ECG vs other devices; $F_{1,38} = 3.46, p = 0.07$) on the association between
2 HCT performance and heart rate. The moderation effects of population and measurement type on
3 the HCT-BMI association could not be tested given that too few studies ($N = 3$) used a clinical
4 sample and measurement type was often not reported. Finally, we tested the presence of a
5 publication bias. Funnel plots and p-curves did not present any significant asymmetry.

1

2 *Figure 6. Aggregated correlation between HCT performance and heart rate.*

Figure 7. Aggregated correlation between HCT performance and BMI.

Sociodemographic variables.

HCT performance was significantly and negatively associated with 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 very low ($I^2 = 0\%$, $Q(13) = 5.90$, $p = 0.95$) and low ($I^2 = 27\%$, $Q(19) = 26.02$, $p = 0.13$) between-study heterogeneity for sex and age, respectively. We did not find any outlier. Regarding moderators, we could not test the effect of population type given that most studies used healthy samples.

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 published. When missing studies are imputed with the trim and fill method, we observe 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.

1

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

Figure 9. Aggregated correlation between HCT performance and sex.

Discussion

This research addressed a simple but important question: if the Heartbeat Counting Task (HCT) faces important validity issues, why significant associations are found and claimed with mental health and its risk factors? The results of this meta-analysis provide a straightforward answer to this question: important theoretically expected associations are actually not reliably found and, when some are found, they represent very small effect sizes that can be explained by non-interceptive processes. This is explained in the remainder of the discussion.

HCT performance was not significantly associated with mental health variables, i.e., trait anxiety, depression, and alexithymia. This is inconsistent with several theoretical assumptions. First, interoception is thought to play a central role in emotional experiences (Barrett, 2006a;

Damasio, 1994; Prinz, 2004). Relatedly, low access to internal signals would 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, would predispose individuals to experiment anxiety (Domschke, Stevens, Pfleiderer, & Gerlach, 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 interoceptive accuracy, and emotion and interoception in psychopathology” (Quadt, Critchley, Garfinkel, Tsakiris, & De Preester, 2018, p.138). The present meta-analysis indicates that, when using the HCT, these theoretical assumptions are not supported. This questions the validity of the theories or that of the HCT (possibly, both).

Regarding physical health, 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, Tsakiris, Pollatos, Schulz, & Herbert, 2020) or as additional evidence that the task lacks validity (Corneille et al., 2020). Specifically, as individuals with higher heart rate (Schandry, Bestler, & Montoya, 1993) and BMI (Rouse et al., 1988) receive cardiac signal 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

1 to) perceptual abilities (Corneille et al., 2020). If individuals perform better when the task
2 becomes easier (i.e., when the signal gets stronger), this ambiguously speaks to their abilities.
3 The association between HCT and heart rate may also be spurious. A recent study showed that
4 the correlation between HCT scores and heart rate becomes non-significant when time estimation
5 and knowledge about heart rate are controlled for (Desmedt et al., 2020). A similar rationale
6 could apply to the association between HCT performance and BMI.

7   Regarding sociodemographic variables, male participants had significantly higher HCT
8 scores than female participants, which is consistent with previous suggestions (Rouse et al.,
9 1988). No significant association was found between HCT performance and age. However, when
10 publication bias was corrected or a third-level meta-analysis was performed, a significant
11 negative association was observed. This is consistent with many empirical studies showing that
12 perceptual abilities (e.g., vision, hearing, proprioception, and pain perception) decrease with age
13 (Heft & Robinson, 2017). However, although these results could be explained by true inter-
14 individual differences in cardiac IAcc, they can also be due to differences in heart rate estimation
15 abilities and cardiac signal intensity.

16   More generally, associations between HCT performance and the above-mentioned
17 outcomes are small and could lack clinical/theoretical significance. Shared variance was from
18 0.01 (for alexithymia) to 2.9% (for heart rate) which is negligible. The correlation between HCT
19 performance and depression represented 0.6% of shared variance (when one outlier was
20 removed). This fails to support the relevance of clinical interventions targeting interoceptive
21 abilities to improve depressive symptoms (de Jong et al., 2016). It should additionally be noted
22 that these weak correlations may be alternatively explained by non-interoceptive processes. For
23 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 procedural details 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.

Limitations

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 been found otherwise. For instance, a recent study has shown that HCT performance is only significantly associated with alexithymia when confounding variables are controlled for (Murphy, Brewer, Hobson, Catmur, & Bird, 2018). However, 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, Hoyle, & Newman, 2006).

We could also not test the effect of important moderators such as HCT instructions. As we noted in the introduction, a simple change in HCT instructions can dramatically impact

1 average HCT performance: asking participants to only count their felt heartbeats and avoid
2 guessing their heart rate decreases HCT performance by half (Desmedt et al., 2018) and
3 significantly decreases the correlation between heartbeat counting and estimation strategies
4 (Desmedt et al., 2020). It would thus be interesting to test if HCT performance is associated with
5 mental and physical health outcomes when modified instructions are used. It is nevertheless
6 important to note that these associations could still be explained by confounding variables such
7 as response bias (Corneille et al., 2020) and cardiac signal intensity (Schandry et al., 1993).

8   Associations may also depend on outcomes measurement. Indeed, Carlson and Herdman
9 (2012) showed that insufficient correlation between two measures of the same construct can lead
10 to very heterogeneous findings (i.e., poor replicability) across measures. For instance, trait
11 anxiety was assessed with different questionnaires such as State-Trait Anxiety Inventory (STAI),
12 Beck Anxiety Inventory (BAI) and anxiety subscale of the Depression Anxiety and Stress Scale
13 (DASS-A). While DASS-A and BAI highly correlate with each other ($N = 717$, $r = 0.81$;
14 Lovibond & Lovibond, 1995), STAI only moderately correlates with BAI ($N = 191$; $r = 0.66$;
15 Kohn, Kantor, DeCicco, & Beck, 2008) and DASS-A ($N = 567$, $r = 0.48$; Grös, Antony, Simms,
16 & McCabe, 2007). This means that results may be relatively consistent between DAI-A and BAI,
17 but inconsistent between DAI-A/BAI and STAI-T (Carlson & Herdman, 2012).

18   It is noteworthy that STAI was the most frequently used measure of trait anxiety whereas
19 it may not represent the best choice for testing the relationship between HCT performance and
20 trait anxiety. STAI may have low discriminant validity as it blends anxious and depressive
21 symptomatology (Balsamo et al., 2013). Considering that IAcc is assumed to be negatively
22 associated with depression but positively associated with anxiety, a measure mixing these
23 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.

Even though moderators could play a role in mental health correlates of HCT performance, we urge caution when examining these moderated associations. Specifically, we encourage researchers interested in interoception 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, and this is also true for a wild examination of covariates, 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).

Conclusion

Associations between IAcc and mental health indicators are strongly expected and frequently stated on a theoretical level. In contrast, we found no support for these associations at the meta-analytic level. Considering important concerns raised about the validity of the HCT, we believe interoception researchers should urgently design alternative measures of IAcc for addressing these important associations. On a positive note, several researchers are currently working on new promising paradigms. For the cardiac domain, we can cite the Phase Adjustment Task (Plans et al., 2020), the heartbeat tapping task combined with Bayesian computation modeling (Smith et al., 2020), and modified heartbeat tracking tasks (Fittipaldi et al., 2020; Larsson, Esposito, Critchley, Dienes, & Garfinkel, 2020). For the respiratory domain, we can cite the Filter Detection Task (Harrison et al., 2020) and the respiratory occlusion discrimination task (Van Den Houte, Vlemincx, Van Oudenhove, & Luminet, 2020). The development of these new measures may allow for a more sensitive test of mental health correlates of interoceptive accuracy and help detect any association that awaits to be found.

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®)*. Arlington.
- Balsamo, M., Romanelli, R., Innamorati, M., Ciccarese, G., Carlucci, L., & Sagginò, 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., 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>
- Corneille, O., Desmedt, O., Zamariola, G., Luminet, O., & Maurage, 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.
- 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>
- de Jong, M., Lazar, S. W., Hug, K., Mehling, W. E., Hölzel, B. K., Sack, A. T., ... 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., & Maurage, 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>
- Fittipaldi, S., Abrevaya, S., Fuente, A. de la, Pascariello, G. O., Hesse, E., Birba, A., ... Ibáñez, A. (2020). A multidimensional and multi-feature framework for cardiac interoception. *NeuroImage*, *212*, 116677. <https://doi.org/10.1016/j.neuroimage.2020.116677>
- 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. Scopus.
<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>
- Harrison, O. K., Garfinkel, S. N., Marlow, L., Finnegan, S., Marino, S., Nanz, L., ... Fleming, S. M. (2020). *The Filter Detection Task for measurement of breathing-related interoception and metacognition* [Preprint]. Neuroscience. <https://doi.org/10.1101/2020.06.29.176941>
- Harshaw, C. (2015). Interoceptive dysfunction: Toward an integrated framework for understanding somatic and affective disturbance in depression. *Psychological Bulletin*, 141(2), 311–363. Scopus. <https://doi.org/10.1037/a0038101>
- 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.
- 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>
- 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., Kobylński, P., Zaorska, J., Skrzyszewski, J., ... 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., ... Mehling, W. E. (2017). Interoception and mental health: A roadmap. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 3, 501–503.
<https://doi.org/10.1016/j.bpsc.2017.12.004>
- Khoury, N. M., Lutz, J., & Schuman-Olivier, Z. (2018). Interoception in Psychiatric Disorders: A Review of Randomized, Controlled Trials with Interoception-Based Interventions. *Harvard Review of Psychiatry*, 26(5), 250–263. Scopus.
<https://doi.org/10.1097/HRP.0000000000000170>
- 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/10.1080/00223890802248844>
- Larsson, D., Esposito, G., Critchley, H., Dienes, Z., & Garfinkel, S. (2020). *Sensitivity to changes in rate of heartbeats as a measure of interoceptive ability*. PsyArXiv.
<https://doi.org/10.31234/osf.io/v34a7>
- 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.
- 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.
- 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., & Group, T. P. (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. (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., Brewer, R., Hobson, H., Catmur, C., & Bird, G. (2018). Is alexithymia characterised by impaired interoception? Further evidence, the importance of control variables, and the problems with the Heartbeat Counting Task. *Biological Psychology*, 136, 189–197. <https://doi.org/10.1016/j.biopsycho.2018.05.010>
- Murphy, J., Cheesman, R., Gregory, A. M., Lau, J., Ehlers, A., Catmur, C., ... 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>
- National Institutes of Health. (2014a). *Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies*. Retrieved from Available online at: <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*.

Retrieved from Available online at: <https://www.nhlbi.nih.gov/health-pro/guidelines/in-develop/cardiovascular-risk-reduction/tools/cohort>

Ogishima, H., Maeda, S., Tanaka, Y., & Shimada, H. (2020). Effects of Depressive Symptoms, Feelings, and Interoception on Reward-Based Decision-Making: Investigation Using Reinforcement Learning Model. *Brain Sciences*, 10(8), 508.

<https://doi.org/10.3390/brainsci10080508>

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., Ponzio, S., Bird, G., Ring, C., Morelli, D., Keating, C. T., ... Cairo, M. (2020).

Measuring Interoception: The Phase Adjustment Task. PsyArXiv.

<https://doi.org/10.31234/osf.io/sa953>

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*.

<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>
- 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>
- Smith, R., Kuplicki, R., Feinstein, J., Forthman, K. L., Stewart, J. L., Paulus, M. P., ... Khalsa, S. S. (2020). A Bayesian computational model reveals a failure to adapt interoceptive precision estimates across depression, anxiety, eating, and substance use disorders. *MedRxiv*, 2020.06.03.20121343. <https://doi.org/10.1101/2020.06.03.20121343>
- Sugawara, A., Terasawa, Y., Katsunuma, R., & Sekiguchi, A. (2020). Effects of interoceptive training on decision making, anxiety, and somatic symptoms. *BioPsychoSocial Medicine*, 14(1), 7. <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>
- Van Den Houte, M., Vlemincx, E., Van Oudenhove, L., & Luminet, O. (2020). The respiratory occlusion discrimination task: A new paradigm to measure respiratory interoceptive accuracy. *Manuscript Submitted for Publication*.
- 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>
- 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>