Running head: SEX DIFFERENCES IN INTEROCEPTION

Word count: 12,664 words

Submission date: Friday 17th September; Revision date: Tuesday 26th October; Revision date: Thursday 18th November

Sex differences in interoceptive accuracy: A meta-analysis

Freya Prentice[1*] & Jennifer Murphy[2]

[1]Developmental Neurosciences Department, UCL Great Ormond Street Institute of Child Health

[2]Department of Psychology, Royal Holloway, University of London

*Freya Prentice

Developmental Neurosciences Department

UCL Great Ormond Street Institute of Child Health

London

WC1N 1EH

freya.prentice.19@ucl.ac.uk

Abstract

Interoceptive accuracy, the ability to correctly perceive internal signals arising from the body, is thought to be disrupted in numerous mental and physical health conditions. Whilst evidence suggests poorer interoceptive accuracy in females compared to males, raising the possibility that interoceptive differences may relate to sex differences in mental and physical health, results concerning sex differences in interoceptive accuracy are mixed. Given such ambiguity, this meta-analysis aimed to establish the presence or absence of sex differences in interoceptive accuracy across cardiac, respiratory, and gastric domains. A review of 7956 abstracts resulted in 93 eligible studies. Results demonstrated superior accuracy in males across cardiac, but not gastric, tasks, while findings on respiratory tasks were mixed. Effect sizes were consistent across cardiac tasks, but instability and/or moderate heterogeneity was observed across other domains, likely due to the small number of eligible studies. Despite such limitations, results indicate the possibility of sex differences across interoception tasks and domains. Methodological limitations concerning the influence of physiological factors, and directions for future research are discussed.

Keywords: interoceptive accuracy, cardiac interoception, respiratory interoception, gastric interoception, sex differences, meta-analysis

1. Introduction

Interoception is defined as the perception of the body's internal state (Craig, 2002), including cardiac, respiratory, gastric, and other bodily sensations. Interoceptive accuracy is typically defined as the ability to correctly perceive internal signals on objective tests (e.g., heartbeat counting or discrimination procedures; Garfinkel et al., 2015) and is thought to be disrupted in numerous psychological and physical health conditions (for reviews see Barrett & Simmons, 2015; Khalsa et al., 2018; Murphy et al., 2017); for example, atypically low interoceptive accuracy has been reported in those with anorexia nervosa (Pollatos et al., 2008), bulimia nervosa (Klabunde et al., 2013), depression (Terhaar et al., 2012; but see Dunn et al., 2007), schizophrenia (Ardizzi et al., 2016) and substance use disorders (Jakubczyk et al., 2019; Sönmez et al., 2017). Impaired interoceptive accuracy has also been observed in physical health conditions such as obesity (Herbert & Pollatos, 2014) and diabetes (Pauli et al., 1991). Conversely, individuals who are highly socially anxious (Stevens et al., 2011) and those with panic disorder (Ehlers & Breuer, 1992) have been reported to perform better on certain tasks of interoceptive accuracy, with studies also reporting a positive association between interoceptive accuracy and anxiety symptoms in adults (Dunn et al., 2007; but see Garfinkel et al., 2016). The observation that interoception is atypical across numerous psychiatric conditions has led to the suggestion that atypical interoception may constitute a common risk factor for psychopathology or the 'p factor' (Caspi et al., 2014), representing lesser-to-greater severity of psychopathology with associated disruption in neural circuitry (for a review see Brewer et al., 2021).

From puberty onwards, many common mental health conditions are more prevalent in females, with sex differences in risk, symptoms, course, and influencing factors (Kuehner, 2017; Riecher-Rössler, 2017). Importantly sex differences in interoception have also been reported; compared to males, females often exhibit poorer interoceptive accuracy across a range of internal bodily signals (e.g., cardiac, respiratory, and gastric domains: Bornemann & Singer, 2017; Grabauskaitė et al., 2017; Harver et al., 1993; Suschinsky & Lalumière, 2012; Whitehead & Drescher, 1980) and yet, report more attention to internal signals (for a review see Murphy, Viding et al., 2019). Given the aforementioned links between interoception and mental health, it has recently been proposed that sex differences in interoceptive ability may partly explain sex differences in the prevalence and presentation of common mental health conditions (Murphy, Viding et al., 2019).

If the theory that sex differences in interoception relate to sex differences in mental health is correct, it is likely to be of considerable importance for understanding individual risk, resilience, and treatment of mental ill-health. It is therefore crucial to establish whether sex differences in interoception are consistently observed across tasks, domains, and studies of interoceptive ability. In the domain of cardiac interoception, whilst a recent preprint reporting a meta-analysis by Desmedt et al. (2020) highlighted a small yet significant male advantage on the heartbeat counting task (HCT; Schandry, 1981; see Appendix C for abbreviations table) – where participants are required to count their heartbeats over a series of intervals – numerous methodological concerns regarding this task prevent strong conclusions from being drawn regarding sex differences in interoception (Desmedt et al., 2018; Desmedt et al., 2020; Murphy, Brewer, et al., 2018; Murphy, Millgate, et al., 2018; Ring & Brener, 1996; Ring et al., 2015). Indeed, given that performance on the HCT is only weakly related to heartbeat discrimination variants (HDT; Katkin et al., 1983; Whitehead et al., 1977) – where participants are required to judge whether an external stimulus is synchronous with their heartbeat (Hickman et al., 2020) – it is crucial to establish whether sex differences in cardiac interoception are routinely observed or are instead a potential artefact attributable to methodological confounds inherent in the HCT.

Whilst evidence in the domain of cardiac interoception highlights a need to consider dissociations across tasks of interoceptive ability, the same is true when considering interoception across other domains. Indeed, whilst some studies do suggest relationships across interoceptive domains (e.g., correlations have been reported across cardiac and gastric domains; Herbert et al., 2012; Whitehead & Drescher, 1980), the majority of evidence suggests that performance in one domain of interoception does not relate to performance in another (Ferentzi et al., 2019; Garfinkel et al., 2016; Harver et al., 1993; Pollatos et al., 2016; Steptoe & Vögele, 1992). This is particularly apparent when considering the relationship between tasks of respiratory and cardiac interoceptive ability, where relationships have not been reported (Garfinkel et al., 2016; Harver et al., 1993). Whilst sex differences have been reported across both gastric and respiratory tasks (Harver et al., 1993; Whitehead & Drescher, 1980), like cardiac interoception, sex differences in these domains have not been invariably observed. To establish the clinical relevance of sex differences in interoception, it is therefore important to establish both whether sex differences in interoception are only observed for the HCT, or whether sex differences are observed both for other measures of cardiac interoceptive accuracy and interoceptive accuracy across other domains.

As such, the aim of this systematic review and meta-analysis was to provide a comprehensive review of sex differences in interoception across the domains of cardiac, respiratory, and gastric interoception, considering also the influence of task in the domain of cardiac interoceptive accuracy. We review the methodological differences across studies within each domain and examine the impact of the domain (cardiac, respiratory, gastric), task (HCT, HDT), clinical group membership (typical, mixed clinical groups), and age (adults, children) on the pooled effect size.

2. Method

2.1 Search strategy

The systematic literature search was conducted following the 2020 Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) guidelines (Page et al., 2021). The PRISMA checklists and extracted data can be found at https://tinyurl.com/9r8z8vb4. Cardiac, respiratory, and gastric interoceptive accuracy were focused on as they are consistently included in definitions of interoception (Ceunen et al., 2016; Craig, 2002; Khalsa & Lapidus, 2016; Khalsa, Rudrauf, Feinstein, et al., 2009; Sherrington, 1906), and most research has focused on these domains. We excluded interoceptive domains that do not fall into traditional definitions of interoception (e.g., temperature, pain, itch, cough), those where there are distinct anatomical differences between males and females (e.g., urination, sexual arousal), or where there is no objective measure or problems measuring accuracy (e.g., defecation, hunger, thirst). We searched PubMed for studies that were available online before June 17th 2020 using the following search terms:

> (“interoception” OR “visceroception” OR (“interoceptive” AND (“awareness” OR “sensation” OR “accuracy” OR “sensitivity” OR “perception” OR “recognition”))) OR (“heartbeat” AND (“awareness” OR “perception” OR “discrimination” OR “detection” OR “tracking” OR “counting”)) OR (“respiratory resistance” OR “respiratory load” OR “resistance load” OR “inspiratory resistance” OR ((“respiration” OR “respiratory”) AND “perception”)) OR ((“gastric” OR “electrogastrography” AND (“perception”)) OR “water load”)

These search terms were selected to identify studies assessing cardiac, respiratory, or gastric interoceptive accuracy even when they did not explicitly refer to interoception or interoceptive accuracy, ensuring that we comprehensively reviewed the literature. A smaller selection of more specific terms was used to search for respiratory and gastric studies, compared to cardiac, to minimise the number of irrelevant physiological studies that arose from the search. PubMed filters were also applied to remove articles that did not present empirical data (e.g., review studies) and articles that were not in English. With filters applied, this search returned 7542 articles. Following the removal of two duplicates, 7540 articles remained.

As only one database was searched given the large number of papers identified, a separate citation search was conducted to ensure no studies were missed (conducted by examining citations on relevant tasks of interoception: Daubenmier et al., 2013; Garfinkel et al., 2016; Harver et al., 1993; Katkin et al., 1981; Schandry, 1981; van Dyck et al., 2016; Whitehead & Drescher, 1980; Whitehead et al., 1977). This search identified 416 studies.

2.2 Study selection

The remaining articles were screened in two phases by two reviewers (FP and JM). First, both reviewers assessed the titles and abstracts for relevance to the meta-analysis. If the article was deemed potentially relevant, it was included for full text screening. Ambiguous studies were resolved by discussion between FP and JM. The initial abstract screening removed a total of 6766 articles (339 from citation search), leaving 1190 articles (77 from the citation search) for full text screening. Studies were removed if they were not conducted on humans, were not in English, were conducted with participants of one sex only, used self-report measures of interoception only, did not present empirical data, were deemed not relevant (e.g., they did not focus on cardiac, respiratory, or gastric interoception) or the full text could not be retrieved.

The full text screening resulted in the removal of a further 645 studies, leaving 394 relevant studies. Full studies were screened by FP and checked by JM where there was uncertainty. Studies were removed if they were conducted with participants of one sex only, included data duplicated in another paper, included no subjective-objective comparison (e.g., they measured subjective scores over time providing no measure of interoceptive accuracy) or included group scores only (e.g., examined subjective-objective relationships at the group level, rather than examining the within-subject correlation).

Respiratory studies were removed if they measured perceived exertion with no mention of breathlessness or dyspnoea because participants may have used other signals (e.g., muscular fatigue) to gauge exertion. Sensations of cough or nasal congestion were also removed due to a lack of measures of objective accuracy. Respiratory studies measuring Borg scores as a percentage fall in forced expiratory volume were also removed as they measure the participants' subjective, rather than objective, perception of respiratory change over time. For gastric studies, perception of the stomach was focused on. Studies were removed if they used oesophageal, colonic, or rectal distention due to a greater focus on somatic or painful sensations and a dearth of research on interoceptive accuracy in these domains. Gastric studies were also removed if they measured perception of nausea, baseline satiation, discomfort, or fullness only, given the absence of objective measures for these signals. Studies using the original one-step water load task were also removed as the measure does not control for stomach capacity and may be influenced by individual differences in physiology (Cox, 1945).

From the PubMed and citation searches, 394 studies were deemed relevant, and authors were contacted for data where it was not readily available (e.g., the data was not reported in the article, supplement or in online depositories). We were unable to obtain data from 299 studies due to the authors being uncontactable, not responding to requests for data or data no longer being available. Data were available from 95 studies, 68 from the PubMed search and 27 from the citation search. At the request of Reviewers, two studies identified by the citation search were removed; one because summary statistics were not reported separately for clinical and typical groups (Failla et al., 2020) and one because the paper focused on children with a clinical diagnosis (Palser et al., 2020).

Of the studies included, 87 of these measured cardiac interoceptive accuracy, 6 measured respiratory interoceptive accuracy, and 4 measured gastric interoceptive accuracy. 4 papers included both cardiac and respiratory (Harver et al., 1993; Van Den Houte et al., 2021) or cardiac and gastric tasks (Ferentzi et al., 2019; Whitehead & Drescher, 1980). Within the cardiac domain, 78 papers employed the HCT and 17 employed the HDT, with an overlap of 7 papers that included both. Of the 93 studies included, 17 reported the means and standard deviations of male and female groups or an F- or T-score in the paper or supplement, 5 provided open access data, and data from the remaining 71 papers were requested directly from the authors. The number of studies excluded at each stage of the screening process and the reasons for exclusion are illustrated in Figure 1.

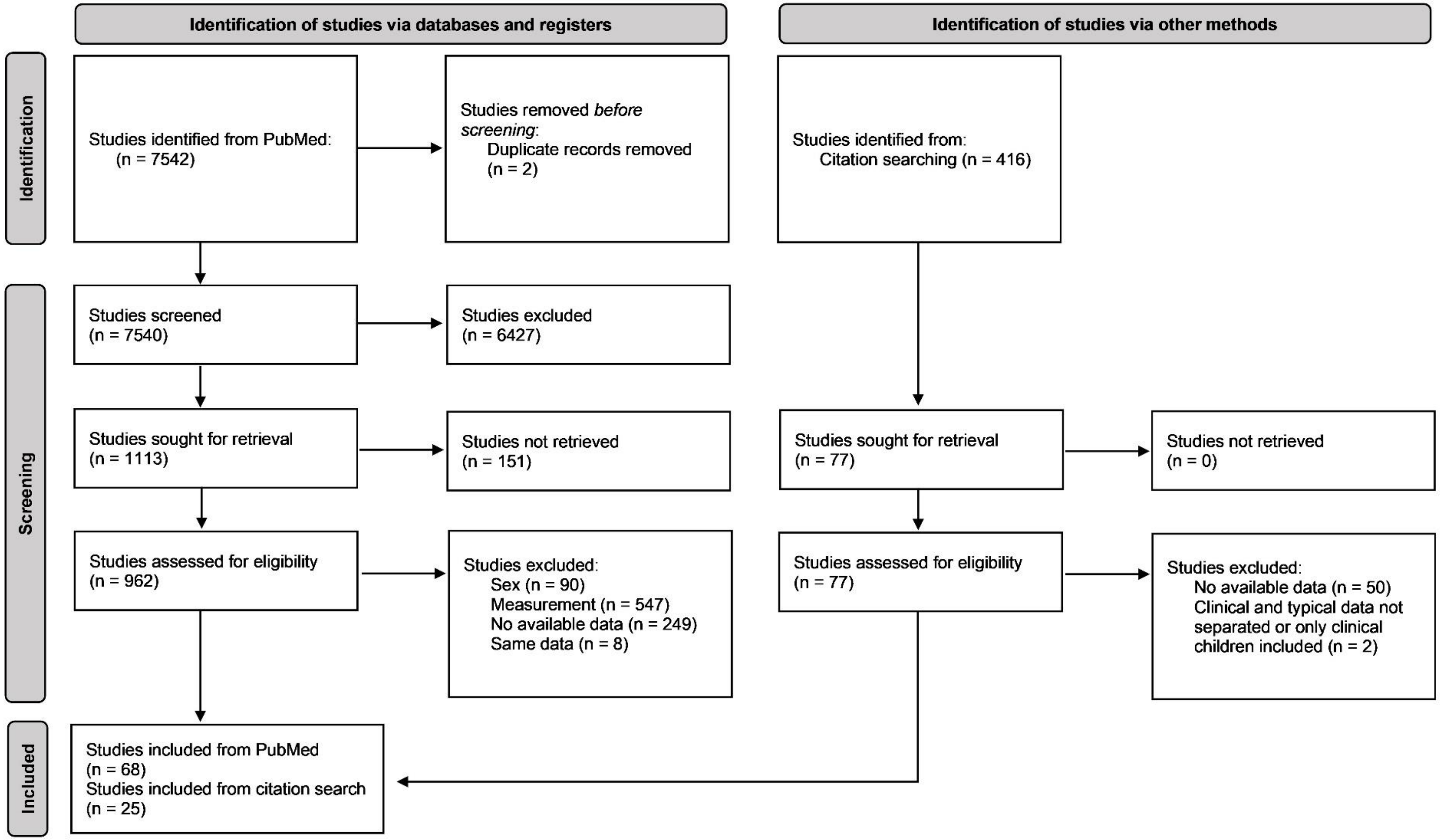

Figure 1. A flow chart depicting the number of papers identified by the database and citation searches and how many were excluded at each stage of the screening process.

2.3 Data extraction

Relevant data were extracted by one reviewer, including details of the interoceptive domain, task, scoring of the accuracy measure, mean and standard deviations for male and female groups, or F- or T-scores when means and standard deviations were not included. Where multiple conditions were utilised in a single study, the baseline (or lowest perturbation level) was extracted. A proportion of extracted data were subsequently checked by the second reviewer and no errors were noted. Where data were not included in the paper, supplement or in online depositaries, authors were contacted by email to request the raw participant data or means, standard deviations and sample sizes of both groups and information on sex and age. The data extracted for each study are presented in Tables 1-4 according to the domain or task used for cardiac interoceptive accuracy.

2.4 Measures used

2.4.1 HCT and variants

The HCT (Schandry, 1981) requires participants to count the number of heartbeats they can feel during specific time intervals, and this is compared to the participant's objective heart rate to determine accuracy. Most studies including the HCT calculated an accuracy score where a higher value indicates more accurate performance, but some studies calculated an error score where a higher value indicates less accurate performance. Heartbeat tapping tasks (e.g., McFarland, 1975), where participants are required to tap or press a computer key following each heartbeat instead of counting, were included in this category 'HCT and variants' as they are generally considered to involve similar processes (Brener & Ring, 2016). One single study conducted on children involved a slightly modified 'jumping jack' paradigm (Schaan et al., 2019) whereby children estimated their heartbeats before and after physical perturbation. This was also included in the HCT and variants category as scoring was based on discrepancies between self-reported and objectively assessed heart rate in a similar manner to the HCT. Data and scoring methods for all studies using the HCT and variants can be found in Table 1.

Table 1

Characteristics and findings of studies using the HCT and variants

Author	Group	Condition	Participant overlap	Extraction	Male N	Female N	Mean age (SD)	Task	Dependent variable	High score indicates	Male mean (SD)	Female mean (SD)	T or F value
Abrams et al. (2018)	Typical	Baseline		Author	26	35	21.95 (3.73)	HCT	IAc score	High accuracy	0.6165 (0.14676)	0.6476 (0.15677)	
Abrevaya et al. (2020)	Typical & Clinical (NP & CP)			Author	41	99	57.44 (15.55)	HCT (tapping)	Mean distance index	High accuracy	0.4239 (0.26662)	0.477 (0.27487)	
	Typical			Author	16	49	55.11 (15.42)	HCT (tapping)	Mean distance index	High accuracy	0.4213 (0.23947)	0.4533 (0.27861)	
	Clinical			Author	25	50	59.45 (15.48)	HCT (tapping)	Mean distance index	High accuracy	0.4256 (0.28745)	0.5002 (0.27195)	
Ainley et al. (2012)	Typical	Baseline		Author	55	74	28.69 (13.53)	HCT	IAc score	High accuracy	0.6801 (0.17339)	0.6159 (0.20234)	
Ainley et al. (2013)	Typical	Baseline		Author	20	21	21.71 (1.99)	HCT	IAc score	High accuracy	0.5879 (0.21397)	0.5664 (0.1623)	
Ainley et al. (2014)	Typical	Baseline		Author	9	36	19.53 (4.84)	HCT	IAc score	High accuracy	0.6725 (0.192)	0.5782 (0.17293)	
Ainley et al. (2015)	Typical			Author	14	72	20.43 (6.50)	HCT	IAc score	High accuracy	0.698 (0.17682)	0.6233 (0.17958)	
Azevedo et al. (2016)	Typical			Author	8	19	28.56 (3.32)	HCT	IAc score	High accuracy	0.8313 (0.10412)	0.71 (0.20309)	
Azevedo et al. (2016) 2 Expt 2	Typical			Author	14	27	25.76 (8.76)	HCT	IAc score	High accuracy	0.7579 (0.14973)	0.6904 (0.1439)	
Azevedo et al. (2017)	Typical			Author	8	22	21.97 (4.49)	HCT	IAc score	High accuracy	0.7409 (0.17249)	0.6665 (0.15805)	
Babo-Rebelo et al. (2016)	Typical			Author	8	8	24.23 (2.53)	HCT	IAc score	High accuracy	0.8004 (0.10035)	0.7718 (0.10346)	

Author	Group	Condition	Participant overlap	Extraction	Male N	Female N	Mean age (SD)	Task	Dependent variable	High score indicates	Male mean (SD)	Female mean (SD)	T or F value
Babo-Rebelo et al. (2019)	Typical			Author	11	12	23.74 (2.65)	HCT	IAc score	High accuracy	0.8073 (0.12327)	0.7148 (0.1414)	
Bornemann & Singer (2017)	Typical			Author	130	187	40.77 (9.26)	HCT	IAc score	High accuracy	0.6652 (0.23763)	0.5689 (0.26216)	
Chua & Bliss-Moreau (2016)	Typical			Author	4	26	21.86 (5.19)	HCT	IAc score	High accuracy	0.7342 (0.224)	0.6677 (0.16954)	
de La Fuente et al. (2019)	Typical & Clinical (SCD & CCD)			Author	66	3	20.17 (2.69)	HCT	Mean distance index	Low accuracy	0.376 (0.22668)	0.6785 (0.10141)	
	Typical			Author	22	2	19.75 (2.61)	HCT	Mean distance index	Low accuracy	0.4067 (0.19696)	0.6493 (0.12429)	
Eley et al. (2004)	Typical Children			Author	123	156	8.47 (0.18)	HCT	Error score	Low accuracy	66.5276 (28.35309)	70.9639 (24.48499)	
Emanuelsen et al. (2015)	Typical			Author	50	102	19.24 (2.95)	HCT	IAc score	High accuracy	0.6442 (0.21363)	0.5287 (0.25234)	
Erle et al. (2019)	Typical			Author	18	78	24.6 (4.59)	HCT	IAc score	High accuracy	0.6376 (0.39151)	0.6646 (0.1637)	
Ferentzi et al. (2018)	Typical			Author	30	72	23.34 (4.34)	HCT	IAc score	High accuracy	0.576 (0.261792)	0.440833 (0.258712)	
Ferentzi et al. (2019)	Typical			Author	54	53	21.72 (3.01)	HCT	IAc score	High accuracy	0.57 (0.29)	0.51 (0.25)	
Fittipaldi et al. (2020)	Typical			Paper	55	59	40.81 (20.54)	HCT (tapping)	Absolute difference score	Low accuracy	0.41 (0.23)	0.46 (0.27)	
Garcia-Cordero et al. (2016)	Typical & Clinical (dementia & stroke)	Baseline		Author	37	62	66.67 (8.86)	HCT (tapping)	IAc score	High accuracy	0.4356 (0.21729)	0.4254 (0.22046)	

Author	Group	Condition	Participant overlap	Extraction	Male N	Female N	Mean age (SD)	Task	Dependent variable	High score indicates	Male mean (SD)	Female mean (SD)	T or F value
	Typical	Baseline		Author	16	26	66.9 (7.45)	HCT (tapping)	IAc score	High accuracy	0.574 (0.15074)	0.5478 (0.19358)	
	Clinical	Baseline		Author	21	36	66.49 (9.84)	HCT (tapping)	IAc score	High accuracy	0.3302 (0.20233)	0.337 (0.19693)	
Georgiou et al. (2015)	Typical Children (main sample)			Paper	29	20	9.72 (0.56)	HCT	IAc score	High accuracy	0.62 (0.19)	0.55 (0.17)	
Godefroid et al. (2016)	Typical	Averages from hand visible & hand covered conditions		Author	7	35	22.19 (2.86)	HCT	IAc score	High accuracy	0.6799 (0.201)	0.6713 (0.17069)	
Grabauskaitė et al. (2017)	Typical			Author	18	16	24.24 (2.37)	HCT	IAc score	High accuracy	79.9094 (15.27235)	53.855 (17.75356)	
Herbert et al. (2007)	Typical			Paper	15	19	26.4 (5.8)	HCT	IAc score	High accuracy	0.81 (0.17)	0.73 (0.21)	
Herbert et al. (2010)	Typical			Paper	19	19	26.3 (5.8)	HCT	IAc score	High accuracy	0.78 (0.17)	0.71 (0.24)	
Herbert et al. (2011)	Typical			Paper	67	88	28.65 (7.29)	HCT	IAc score	High accuracy	0.74 (0.2)	0.72 (0.19)	
Herman et al. (2019)	Typical		Same as HDT	Author	16	44	22.33 (3.75)	HCT	Iac score	High accuracy	64.5 (25.20846)	56.5 (26.38049)	
Hina & Aspell (2019)	Typical & clinical (smokers)		Same as HDT	Author	40	60	25.67 (8.72)	HCT	IAc score	High accuracy	0.737 (0.18425)	0.6707 (0.15067)	
	Typical		Same as HDT	Author	21	28	25.98 (8.07)	HCT	IAc score	High accuracy	0.6919 (0.20012)	0.6207 (0.15019)	
	Clinical (smokers)		Same as HDT	Author	19	32	25.37 (9.37)	HCT	IAc score	High accuracy	0.7868 (0.15514)	0.7144 (0.13905)	
Imafuku et al. (2020)	Typical			Paper	45	35	24.5 (9.53)	HCT	IAc score	High accuracy	71.25 (18)	63.44 (20.65)	

Author	Group	Condition	Participant overlap	Extraction	Male N	Female N	Mean age (SD)	Task	Dependent variable	High score indicates	Male mean (SD)	Female mean (SD)	T or F value
Isomura & Watanabe (2020)	Typical	Baseline		Author	31	26	Not reported	HCT	IAc score	High accuracy	0.6487 (0.2321)	0.6892 (0.19545)	
Jakubczyk et al. (2019)	Typical & Clinical (AUD)			Author	172	38	42.61 (10.19)	HCT	IAc score	High accuracy	0.6011 (0.25061)	0.6403 (0.19617)	
	Typical			Author	79	25	40.57 (8.16)	HCT	IAc score	High accuracy	0.7222 (0.08775)	0.727 (0.07496)	
	Clinical			Author	93	13	44.03 (11.21)	HCT	IAc score	High accuracy	0.4983 (0.29486)	0.4737 (0.24841)	
Kennedy et al. (2019)	Typical & Clinical (GD)			Author	40	40	40.87 (13.07)	HCT	IAc score	High accuracy	0.5683 (0.29695)	0.5341 (0.3147)	
	Typical			Author	16	17	38.11 (13.16)	HCT	IAc score	High accuracy	0.554 (0.36045)	0.4502 (0.33728)	
	Clinical			Author	24	23	42.46 (12.85)	HCT	IAc score	High accuracy	0.5779 (0.25407)	0.596 (0.28877)	
Kinnaird et al. (2020)	Typical & Clinical (AN)			Author	4	70	26.07 (7.64)	HCT	IAc score	High accuracy	0.5975 (0.17115)	0.6623 (0.25759)	
	Typical			Author	2	35	26.05 (7.31)	HCT	IAc score	High accuracy	0.51 (0.22627)	0.6311 (0.27229)	
	Clinical			Author	2	35	26.08 (8.05)	HCT	IAc score	High accuracy	0.685 (0.07778)	0.6934 (0.24192)	
Koch & Pollatos (2014)	Typical Children			Paper	657	693	8.39 (0.94)	HCT	IAc score	High accuracy	0.56 (0.26)	0.53 (0.26)	
Koeppel et al. (2020)	Typical			Author	12	19	24.26 (3.71)	HCT	IAc score	High accuracy	0.6725 (0.14858)	0.7379 (0.1818)	
Koteles et al. (2020)	Typical			Author	19	26	21.41 (1.62)	HCT	Bias score	High accuracy	0.547 (0.33082)	0.5455 (0.22919)	
Krautwurst et al. (2014)	Typical			Author	21	79	23.66 (3.39)	HCT	IAc score	High accuracy	0.6739 (0.26721)	0.5383 (0.21179)	

Author	Group	Condition	Participant overlap	Extraction	Male N	Female N	Mean age (SD)	Task	Dependent variable	High score indicates	Male mean (SD)	Female mean (SD)	T or F value
Krautwurst et al. (2016)	Typical & Clinical (HA)			Author	41	64	38.1 (12.3)	HCT	IAc score	High accuracy	0.5951 (0.23317)	0.5673 (0.23505)	
	Typical			Author	19	37	38.41 (12.66)	HCT	IAc score	High accuracy	0.5856 (0.23455)	0.5447 (0.26891)	
	Clinical (HA)			Author	22	27	37.76 (12.01)	HCT	IAc score	High accuracy	0.6034 (0.23718)	0.5982 (0.17896)	
Lischke et al. (2020)	Typical (Schandry equation)		Same as below	Paper	41	41	26.37 (0.74)	HCT	IAc scores	High accuracy	0.7 (0.1921)	0.61 (0.1921)	
	Typical (Hart equation)		Same as above	Paper	41	41	26.37 (0.74)	HCT	IAc scores	High accuracy	0.62 (0.2561)	0.49 (0.2561)	
Lugo et al. (2018)	Typical (CDC)			Author	24	3	21.7 (0.71)	HCT	IAc score	High accuracy	0.6717 (0.23284)	0.6467 (0.25166)	
Lukowska et al. (2018)	Typical			OSF	12	14	20.04 (4.67)	HCT	IAc score	High accuracy	0.7214 (0.22419)	0.5992 (0.27846)	
Marshall et al. (2017) Expt 1	Typical			Author	10	15	25.28 (3.95)	HCT	IAc score	High accuracy	0.444 (0.41045)	0.4773 (0.37829)	
Marshall et al. (2017) Expt 2	Typical			Author	9	16	26.68 (4.26)	HCT	IAc score	High accuracy	0.4789 (0.31251)	0.5575 (0.30176)	
Marshall et al. (2018) Expt 1	Typical			Author	15	10	25.6 (4.97)	HCT	IAc score	High accuracy	0.364 (0.32863)	0.567 (0.20249)	
Marshall et al. (2018) Expt 2	Typical			Author	11	14	26.72 (4.63)	HCT	IAC score	High accuracy	0.4873 (0.25542)	0.5614 (0.29247)	
Marshall et al. (2019)	Typical			Author	17	10	27.04 (4.65)	HCT	IAc score	High accuracy	0.56 (0.26789)	0.695 (0.23839)	
Michael et al. (2015)	Typical			Author	7	19	21.85 (2.77)	HCT	IAc score	High accuracy	0.7367 (0.27983)	0.7278 (0.19115)	

Author	Group	Condition	Participant overlap	Extraction	Male N	Female N	Mean age (SD)	Task	Dependent variable	High score indicates	Male mean (SD)	Female mean (SD)	T or F value
Moeini-Jazani et al. (2017)	Typical	Control		Author	20	22	24.07 (4.05)	HCT	IAc score	High accuracy	0.5206 (0.27021)	0.5476 (0.22754)	
Motyka et al. (2019)	Typical			Author	16	16	25.75 (4.01)	HCT	IAc score	High accuracy	0.7793 (0.12835)	0.6693 (0.15235)	
Murphy, Brewer et al. (2018)	Typical			Author	86	201	38.07 (21.09)	HCT	IAc score	High accuracy	214.5111 (118.31)	183.4698 (116.8397)	
Murphy, Brewer et al. (2019) Expt 1	Typical			Author	38	29	23.15 (6.05)	HCT	IAc score	High accuracy	42.08221 (31.67911)	33.23378 (27.43321)	
Murphy, Brewer et al. (2019) Expt 2	Typical			Author	12	19	32.49 (11.4)	HCT	IAc score	High accuracy	38.36038 (28.14163)	39.72213 (28.12411)	
Palser et al. (2018)	Typical & Clinical Children (ASD)			Author	48	27	11.56 (3.04)	HCT	IAc score	High accuracy	0.3581 (0.55671)	0.3656 (0.50043)	
	Typical Children			Author	23	22	11.07 (3.07)	HCT	IAc score	High accuracy	0.6139 (0.27445)	0.4341 (0.44833)	
	Clinical Children			Author	25	5	12.3 (2.88)	HCT	IAc score	High accuracy	0.1228 (0.64614)	0.064 (0.65805)	
Pollatos et al. (2005)	Typical			Paper	16	28	25.5 (4.5)	HCT	IAc score	High accuracy	0.8 (0.14)	0.77 (0.21)	
Pollatos, Traut-Mattausch et al. (2007)	Typical			Paper	35	67	26.9 (6.3)	HCT	IAc score	High accuracy	0.8 (0.17)	0.77 (0.17)	
Pollatos, Matthias et al. (2007)	Typical			Paper	10	26	27.6 (7.2)	HCT	IAc score	High accuracy	0.82	0.8	F = 0.559

Author	Group	Condition	Participant overlap	Extraction	Male N	Female N	Mean age (SD)	Task	Dependent variable	High score indicates	Male mean (SD)	Female mean (SD)	T or F value
Rae et al. (2018)	Typical		Same as HDT	Author	21	25	22.98 (3.35)	HCT	IAc score	High accuracy	0.6398 (0.28737)	0.6479 (0.19635)	
Rae et al. (2019)	Typical & Clinical (TS)		Same as HDT	OSF	24	19	34.42 (11.01)	HCT	IAc score	High accuracy	0.7367 (0.19838)	0.6268 (0.25126)	
	Typical		Same as HDT	OSF	12	10	34.45 (11.52)	HCT	IAc score	High accuracy	0.7933 (0.14054)	0.696 (0.16893)	
	Clinical		Same as HDT	OSF	12	9	34.38 (10.74)	HCT	IAc score	High accuracy	0.68 (0.23564)	0.55 (0.31197)	
Rae et al. (2020)	Typical			OSF	17	24	21.54 (2.46)	HCT	IAc score	High accuracy	0.6396 (0.2269)	0.6039 (0.32385)	
Ricciardi et al. (2016)	Typical & Clinical (FMD)			Author	9	26	38.91 (10.41)	HCT	IAc score	High accuracy	0.6756 (0.2209)	0.5557 (0.18943)	
	Typical			Author	5	13	37.22 (9.47)	HCT	IAc score	High accuracy	0.7293 (0.23908)	0.6219 (0.1707)	
	Clinical			Author	4	13	40.71 (11.34)	HCT	IAc score	High accuracy	0.6086 (0.20768)	0.4895 (0.19006)	
Ring & Brener (2018)	Typical			Author	18	30	18.69 (0.78)	HCT	IAc score	High accuracy	0.72 (0.26)	0.55 (0.29)	
Schaan et al. (2019)	Typical Children			Author	27	22	4.86 (0.677)	HCT (jumping jack paradigm)	Error score	Low accuracy	0.8674 (0.69931)	0.9455 (0.76069)	
Schauder et al. (2015)	Typical & Clinical Children (ASD)			Author	39	6	Not reported	HCT	IAc score	High accuracy	0.6765 (0.17731)	0.6484 (0.22754)	
	Typical Children			Author	20	4	Not reported	HCT	IAc score	High accuracy	0.6648 (0.18762)	0.5982 (0.26931)	
	Clinical Children			Author	19	2	Not reported	HCT	IAc score	High accuracy	0.6888 (0.17001)	0.7488 (0.10505)	

Author	Group	Condition	Participant overlap	Extraction	Male N	Female N	Mean age (SD)	Task	Dependent variable	High score indicates	Male mean (SD)	Female mean (SD)	T or F value
Schroeder et al. (2015)	Typical & Clinical (CCP & NCCP)		Same as HDT	Author	68	58	52.73 (10.61)	HCT	Error score	Low accuracy	51.1969 (35.81919)	55.8985 (29.48492)	
	Typical		Sample of HDT	Author	19	32	49.08 (9.61)	HCT	Error score	Low accuracy	41.0915 (36.34319)	53.8951 (26.02518)	
	Clinical		Same as HDT	Author	49	36	55.21 (10.61)	HCT	Error score	Low accuracy	55.1153 (35.20608)	58.3644 (33.6276)	
Schultchen et al. (2019)	Typical & Clinical (OCD)	Baseline		OSF	28	24	27.54 (6.46)	HCT	IAc score	High accuracy	0.6793 (0.18347)	0.6726 (0.1811)	
	Typical	Baseline		OSF	14	12	26.5 (5.6)	HCT	IAc score	High accuracy	0.7249 (0.18356)	0.7803 (0.14449)	
	Clinical	Baseline		OSF	14	12	28.58 (7.19)	HCT	IAc score	High accuracy	0.6337 (0.17818)	0.565 (0.14967)	
Schulz et al. (2013)	Typical	Baseline	Same as HDT	Author	13	29	22.95 (2.52)	HCT	IAc score	High accuracy	0.7395 (0.19445)	0.8294 (0.14843)	
Stern et al. (2020)	Typical			Author	11	9	26.48 (5.04)	HCT	IAc score	High accuracy	0.632 (0.27677)	0.6806 (0.14713)	
Sueyoshi et al. (2014)	Typical			Author	5	15	20.25 (2.15)	HCT	IAc score	High accuracy	0.705 (0.23443)	0.6772 (0.17775)	
Sutterlin et al. (2013)	Typical			Paper	20	20	27.6 (6.7)	HCT	IAc score	High accuracy	0.53 (0.23)	0.45 (0.21)	
Todd, Hina et al. (2020)	Typical			Author	20	29	26.08 (6.73)	HCT	IAc score	High accuracy	0.6069 (0.20261)	0.5833 (0.20164)	
Ueno et al. (2020)	Typical (older adults)			Author	8	19	77.29 (6.24)	HCT	IAc score	High accuracy	89.7881 (10.80199)	72.9837 (18.41486)	
Van Den Houte et al. (2021)	Typical		Sample of respiratory	Author	21	54	23.63 (4.99)	HCT	IAc score	High accuracy	0.3907 (0.2706)	0.2809 (0.1699)	
van't Wout et al. (2013)	Typical			Author	10	21	25.36 (6.85)	HCT (tapping)	IAc score	High accuracy	0.712 (0.21207)	0.6414 (0.2136)	

Author	Group	Condition	Participant overlap	Extraction	Male N	Female N	Mean age (SD)	Task	Dependent variable	High score indicates	Male mean (SD)	Female mean (SD)	T or F value
Villani et al. (2019)	Typical	Sham VNS	Same as HDT	OSF	14	32	21.17 (3.14)	HCT	IAc score	High accuracy	0.7077 (0.12549)	0.6313 (0.19317)	
Wegner et al. (2015)	Typical	Placebo		Paper	20	20	27.6 (0.5)	HCT	IAc score	High accuracy	0.73 (0.15)	0.74 (0.17)	
Weineck et al. (2020)	Typical			OSF	14	45	22.79 (3.79)	HCT	IAc score	High accuracy	0.7523 (0.17461)	0.6373 (0.17476)	
Wiersema & Godefroid (2018)	Typical & Clinical (ADHD)	Baseline		Author	25	22	23.51 (3.86)	HCT	IAc score	High accuracy	0.807 (0.16715)	0.7973 (0.16336)	
	Typical	Baseline		Author	13	10	23.51 (3.86)	HCT	IAc score	High accuracy	0.813 (0.16413)	0.8 (0.19412)	
	Clinical	Baseline		Author	12	12	23.46 (4.48)	HCT	IAc score	High accuracy	0.8005 (0.17742)	0.7949 (0.14178)	
Wolk et al. (2014)	Typical & Clinical (PD)			Author	16	14	39.06 (12.99)	HCT	IAc score	High accuracy	0.7706 (0.12615)	0.4986 (0.19066)	
	Typical			Author	8	7	36.53 (12.61)	HCT	IAc score	High accuracy	0.8013 (0.10232)	0.5343 (0.2355)	
	Clinical			Author	8	7	41.59 (13.26)	HCT	IAc score	High accuracy	0.74 (0.14658)	0.4629 (0.14256)	
Zamariola et al. (2019) Expt 1	Typical	Inclusion & exclusion conditions in cyberball paradigm		Author	49	50	Not reported	HCT	IAc score	High accuracy	0.6869 (0.14749)	0.6171 (0.20714)	
Zamariola et al. (2019) Expt 2	Typical	Inclusion & exclusion conditions in cyberball paradigm		Author	40	118	21.85 (3.52)	HCT	IAc score	High accuracy	0.672 (0.16834)	0.6352 (0.1776)	

Author	Group	Condition	Participant overlap	Extraction	Male N	Female N	Mean age (SD)	Task	Dependent variable	High score indicates	Male mean (SD)	Female mean (SD)	T or F value
Zamariola et al. (2019) Expt 3	Typical	Neural & negative feedback conditions		Author	39	118	22.24 (2.94)	HCT	IAc score	High accuracy	0.6806 (0.20448)	0.5909 (0.18178)	
Zamariola et al. (2019) Expt 4	Typical	Inclusion & exclusion conditions in cyberball paradigm		Author	22	105	22.48 (5.81)	HCT	IAc score	High accuracy	0.6418 (0.18441)	0.5956 (0.17382)	
Zoellner & Craske (1999)	Typical & Clinical (IP)	Baseline		Author	30	28	18.8 (2.37)	HCT	Error score	Low accuracy	0.4555 (0.27295)	0.6087 (0.32774)	
	Typical	Baseline		Author	13	14	18.57 (1.04)	HCT	Error score	Low accuracy	0.4835 (0.31193)	0.7664 (0.24534)	
	Clinical	Baseline		Author	17	14	19 (3.11)	HCT	Error score	Low accuracy	0.4341 (0.25481)	0.4509 (0.33059)	

Abbreviations: ADHD attention deficit hyperactivity disorder *AN* anorexia nervosa *ASD* autism spectrum disorder *AUD* alcohol use disorder *CCD* clorhidrate cocaine dependence *CCP* cardiac chest pain *CP* cardiac patients *CDC* cyber defence cadets *FMD* functional movement disorder *GD* gambling disorder *HA* health anxiety *IP* infrequent panickers *NCCP* non-cardiac chest pain *NP* neurological patients *OCD* obsessive compulsive disorder *PD* panic disorder SCD smoked cocaine dependence *TS* Tourette's syndrome

2.4.2 HDT and other variants

The HDT requires participants to indicate whether an auditory or visual stimulus is synchronous with their heartbeat or not, over a series of trials. The delay at which the stimulus is presented with respect to the heartbeat is determined by which version of the HDT is employed. There are several different HDT variants: the two-alternative forced choice procedure (2AFC, e.g., Whitehead et al., 1977); the six-alternative forced choice procedure (6AFC; e.g., Brener & Kluvitse, 1988); and the method of constant stimuli (MCS, e.g., Brener et al., 1993; Yates et al., 1985). Most HDT studies included in this review used the 2AFC method, while one used the 6AFC design and one used the MCS (see Table 2). Scoring methods for HDTs vary across studies (see Table 2). In this category we also included measures where interoceptive accuracy was assessed during perturbations of the body's state (Khalsa et al., 2020; Khalsa, Rudrauf, Sandesara, et al., 2009). In these studies, participants rotated a dial during bolus infusions of isoproterenol, a non-selective beta-adrenergic agonist, to indicate changes in the intensity of physical sensations. Accuracy was inferred from zero-lag cross-correlations between perceived and actual heart rate change. As this task involved matching of internal (cardiac) and external (the dial) stimuli, it was included in the HDT and other variants category, despite methodological differences between these tasks. Data and scoring methods for all studies using the HDT and variants can be found in Table 2.

Table 2

Characteristics and findings of studies using the HDT and variants

Author	Group	Condition	Participant overlap	Extraction	Male N	Female N	Mean age (SD)	Task	Dependent variable	High score indicates	Male mean (SD)	Female mean (SD)	T or F value
Brener et al. (1993)	Typical		Same as below	Author	12	12	27.15 (7.09)	2AFC HDT	Accuracy (A')	High accuracy	0.68 (0.13)	0.62 (0.07)	
	Typical		Same as above	Author	12	12	27.15 (7.09)	2AFC HDT	Sensitivity (d')	High accuracy	0.71 (0.66)	0.36 (0.24)	
Harver et al. (1993)	Typical		Same as respiratory	Paper	12	13	19.18	2AFC HDT	Accuracy (A')	High accuracy			F = 3.6
Herman et al. (2019)	Typical		Same as HCT	Author	16	43	22.33 (3.75)	2AFC HDT	Sensitivity	High accuracy	48.3956 (16.4554)	49.3744 (14.50814)	
Hina & Aspell (2019)	Typical & clinical (smokers)		Same as HCT	Author	40	60	25.67 (8.72)	2AFC HDT	Sensitivity	High accuracy	0.6125 (0.15412)	0.5632 (0.16789)	
	Typical		Same as HCT	Author	21	28	25.98 (8.07)	2AFC HDT	Sensitivity	High accuracy	0.5938 (0.17414)	0.5382 (0.16789)	
	Clinical (smokers)		Same as HCT	Author	19	32	25.37 (9.37)	2AFC HDT	Sensitivity	High accuracy	0.6332 (0.13004)	0.585 (0.16747)	
Khalsa et al. (2008) + Khalsa, Rudrauf & Tranel (2009)	Typical (meditators & non meditators)			Author	23	50	47.77 (11.86)	2AFC HDT	Accuracy score (A')	High accuracy	2.17 (0.39)	2.04 (0.38)	
Khalsa, Rudrauf, Sandesara et al. (2009)	Typical	2µg dosage		Author	10	5	36.95 (12.79)	Bolus infusions of isoproterenol	Cross correlation at zero lag	High accuracy	0.5258 (0.18089)	0.1939 (0.23628)	
Khalsa et al. (2020)	Typical (meditators	2µg dosage		Author	20	10	44.35 (13.22)	bolus infusions of	Cross correlation at zero lag	High accuracy	0.52 (0.28)	0.33 (0.21)	

Author	Group	Condition	Participant overlap	Extraction	Male N	Female N	Mean age (SD)	Task	Dependent variable	High score indicates	Male mean (SD)	Female mean (SD)	T or F value
	& non-meditators)							isoproterenol					
Kleckner et al. (2015)	Typical	100 trials		Author	63	104	24.02 (6.99)	2AFC HDT	Sensitivity (d')	High accuracy accurate	0.5685 (0.08673)	0.5628 (0.09593)	
Rae et al. (2018)	Typical		Sample of HCT participants	Author	10	14	23.38 (3.7)	2AFC HDT	Sensitivity (d')	High accuracy	0.53 (0.17826)	0.5536 (0.12004)	
Rae et al. (2019)	Typical & Clinical (TS)		Same as HCT	OSF	24	19	34.42 (11.01)	2AFC HDT	Sensitivity (d')	High accuracy	0.5629 (0.14097)	0.5384 (0.137)	
	Typical		Same as HCT	OSF	12	10	34.45 (11.52)	2AFC HDT	Sensitivity (d')	High accuracy	0.5217 (0.07626)	0.57 (0.15312)	
	Clinical		Same as HCT	OSF	12	9	34.38 (10.74)	2AFC HDT	Sensitivity (d')	High accuracy	0.6042 (0.17896)	0.5033 (0.115)	
Saloman et al. (2016)	Typical			Author	10	13	23.87 (3.06)	2AFC HDT	Sensitivity (d')	High accuracy	54.6 (12.08488)	51.4615 (14.17519)	
Schroeder et al. (2015)	Typical & Clinical (CCP & NCCP)		Sample of HCT	Author	68	55	52.35 (10.32)	6AFC HDT	Sensitivity (d')	High accuracy	0.429204 (0.648066)	0.238925 (0.484075)	
	Typical		Same as HCT	Author	20	32	49.08 (9.51)	6AFC HDT	Sensitivity (d')	High accuracy	0.609365 (0.775822)	0.179728 (0.44684)	
	Clinical		Sample of HCT	Author	48	23	54.75 (10.29)	6AFC HDT	Sensitivity (d')	High accuracy	0.354137 (0.5795)	0.321285 (0.530692)	
Schulz et al. (2013)	Typical	Baseline	Same as HCT + V HDT	Author	13	29	22.95 (2.52)	Auditory 2AFC HDT	Sensitivity (d')	High accuracy	0.4923 (1.10217)	0.2845 (0.89834)	

Author	Group	Condition	Participant overlap	Extraction	Male N	Female N	Mean age (SD)	Task	Dependent variable	High score indicates	Male mean (SD)	Female mean (SD)	T or F value
	Typical	Baseline	Same as HCT + A HDT	Author	13	29	22.95 (2.52)	Visual 2AFC HDT	Sensitivity (d')	High accuracy	1.4354 (1.0219)	1.2176 (1.00787)	
Schwerdtfeger et al. (2019)	Typical			Author	58	54	37.36 (7.64)	MCS HDT	Sensitivity (d')	High accuracy	0.764 (1.58476)	0.4557 (1.31795)	
Suzuki et al. (2013)	Typical			Author	10	11	21.24 (3.06)	2AFC HDT	Sensitivity (d')	High accuracy	0.7815 (0.12233)	0.6706 (0.19195)	
Villani et al. (2019)	Typical	Sham VNS	Same as HCT	OSF	14	32	21.17 (3.14)	2AFC HDT	Sensitivity (d')	High accuracy	0.5457 (0.06676)	0.5506 (0.10096)	
Whitehead & Drescher (1980)	Typical			Paper	9	11	26.6	2AFC HDT	Sensitivity (d')	High accuracy	0.75	0.2	T = 2.48

Abbreviations: ASD autism spectrum disorder *CCP* cardiac chest pain *NCCP* non-cardiac chest pain *TS* Tourette’s syndrome

2.4.3 Respiratory

Most respiratory studies utilised respiratory loads or occlusions; for example, in one detection variant, participants were presented with loads of increasing volume and had to indicate the first detectable load (Benke et al., 2018). In a discrimination variant, participants were required to discriminate between two different respiratory loads (Axen et al., 1994). Other signal-detection variants required participants to detect whether a load (Harver et al., 1993) or obstruction (Harver & Smith, 1996) was present or absent, or to detect differences in the length of presented occlusions (Van Den Houte et al., 2021). One single study utilised a more invasive procedure where participants’ lungs were inflated to different degrees and participants were asked to detect and identify the lateral position of the inflation (e.g., right or left lung). Scoring methods varied across the different tasks (data and scoring methods for all respiratory studies can be found in Table 3).

Table 3

Characteristics and findings of respiratory studies

Author	Group	Condition	Participant overlap	Extraction	Male N	Female N	Mean age (SD)	Task	Dependent variable	High score indicates	Male mean (SD)	Female mean (SD)	T or F value
Axen et al. (1994) Elastic	Typical	Breathing condition	Same as below	Paper	12	12	23.35 (1.37)	Load Discrimination Task	Load discrimination score (%)	High accuracy	54.5 (2.7)	55.3 (3.8)	
Axen et al. (1994) Resistive	Typical	Breathing condition	Same as above	Paper	12	12	23.35 (1.37)	Load Discrimination Task	Load discrimination score (%)	High accuracy	58.8 (2.3)	57.5 (2.8)	
Banzett et al. (1997)	Typical	Small Lobe Only		Paper	3	3	37 (11.75)	Lung Inflation Detection Task	Correct response (%)	High accuracy	47.667 (11.015)	72.667 (12.503)	
Benke et al. (2018)	Typical (Control)			Author	5	23	22.65 (3.29)	Load Detection Task	Load detection threshold	Low accuracy	1.6 (0.55)	1.65 (0.88)	
	Typical (Premature Termination)			Author	5	23	22.46 (3.72)	Load Detection Task	Load detection threshold	Low accuracy	1.5 (0.87)	1.78 (1.6)	
Harver et al. (1993)	Typical		Same as HDT	Paper	12	13	19.18	Resistance Detection Task	Accuracy (A’)	High accuracy			F = 6.17
Harver & Smith (1996)	Typical			Paper	35	45	21.39	Load Detection Task	Sensitivity	Low accuracy	1.41	1.98	T = 2.3
Van Den Houte et al. (2021) Respiratory	Typical		Same as HCT	Author	28	69	23.33 (4.74)	Occlusion Discrimination Task	Just Noticeable Difference	Low accuracy	67.619 (30.951)	76.896 (39.21)	

2.4.4 Gastric

Three different types of gastric interoception task were used: a distention detection task; a stomach contraction discrimination task; and two modified water load tasks. In the distention detection task, a latex balloon was inserted in the fundus of the stomach and distended using a barostat (Bouin et al., 2004). Participants were instructed to identify their first sensation of distention, with sensations at higher volumes representing lower accuracy. In the discrimination task, participants were required to indicate whether a light was presented during or after the peak of a stomach contraction over a series of sessions (Whitehead & Drescher, 1980). In one modified water load task, participants were instructed to drink the same volume of water five times (adjusted for height), and rate changes in their subjective fullness following each drinking session (Ferentzi et al., 2019). A measure of gastric interoceptive accuracy was calculated by subtracting the fullness rating from the first drinking session from the fullness rating of the last drinking session, with higher values reflecting greater accuracy. Another study used the two-stage water load task (van Dyck et al., 2016), in which participants are instructed to drink to satiation within a five-minute period (Todd, Aspell et al., 2020). After completion of this period, participants were instructed to drink until their stomachs were full during a second five-minute period. To calculate accuracy, the volume of water consumed during the first period was divided by the total volume of water consumed, with higher values indicating lower accuracy. Data and scoring methods for all gastric studies can be found in Table 4.

Table 4

Characteristics and findings of gastric studies

Author	Group	Condition	Participant overlap	Extraction	Male N	Female N	Mean age (SD)	Task	Dependent variable	High score indicates	Male mean (SD)	Female mean (SD)	T or F value
Bouin et al. (2004)	Typical & Clinical (FD & IBS)			Author	14	25	42.95 (11.52)	Distension Detection Task	Perception score (volume at first sensation of distension)	Low accuracy	332.14 (170.53)	274 (118.25)	
	Typical			Author	6	6	44.75 (5.71)	Distension Detection Task	Perception score (volume at first sensation of distension)	Low accuracy	408.33 (128.13)	391.67 (86.12)	
	Clinical			Author	8	19	42.12 (13.41)	Distension Detection Task	Perception score (volume at first sensation of distension)	Low accuracy	275 (183.23)	236.84 (102.53)	
Ferentzi et al. (2019)	Typical		Sample of HCT participants	Author	47	42	21.61 (1.64)	Modified WLT	Gastric sensitivity	High accuracy	33.31 (14.624)	35.31 (29.846)	
Todd, Aspell et al. (2020)	Typical (British & Malaysian)			Paper	87	104	23.98 (6.4)	WLT-II	GIAcc (%)	Low accuracy	61.32 (14.61)	62.23 (17.16)	
Whitehead & Drescher (1980)	Typical		Same as HDT	Paper	9	11	26.6	Discrimination Task	Sensitivity (d’)	High accuracy	0.54	0.29	T = 1.87

Abbreviations: FD functional dyspepsia *GIAcc* gastric interoceptive accuracy *IBS* irritable bowel syndrome *WLT* water load task

3. Results

3.1 Meta-analyses

3.1.1 Analysis strategy

Results from studies utilising the HCT and HDT were analysed in separate meta-analyses due to findings that the two are not highly correlated (Hickman et al., 2020). Respiratory studies were grouped together despite different methodologies due to the small number of respiratory studies included. The same was done for gastric studies given the small number of studies identified. Meta-analyses focused on sex differences in 'typical' adult participants only, with separate subgroup analyses for the HCT (typical children, clinical adult groups) reported in Appendix B.

All test statistics were converted to Hedges g which was chosen as it is less biased than other effect sizes (e.g., Cohen's d), especially for small sample sizes (Hedges & Olkin, 1985). The standard error was calculated for Hedges g in R using the *esc* package (Lüdecke, 2019). All analyses were conducted using the *dmetar*, *meta* and *metafor* packages in R (Balduzzi et al., 2019; Harrer et al., 2019; Viechtbauer, 2010;). Random-effects models were used for all the meta-analyses due to suspected heterogeneity of effects (Field, 2001; Hunter & Schmidt, 2000) and the conservative Sidik-Jonkman estimator was applied (Sidik & Jonkman, 2007). Heterogeneity was investigated for each meta-analysis, and Q and I^2 statistics are reported. A Q statistic is calculated by summing the weighted squared differences between the effect sizes of each study and the fixed-effect estimate and is compared to the null hypothesis of homogeneity. It indicates whether variation is higher across studies or between participants within a study. I^2 describes the percentage of variation in effect sizes across studies that is due to heterogeneity rather than chance. Values of 25%, 50% and 75% indicate low, moderate, and high heterogeneity. Publication bias was also assessed. Funnel plots were produced to show the relationship between effect sizes and standard error. Egger's tests were used to assess the asymmetry of the funnel plot where there were sufficient studies included (k>10). Influence analysis using the leave-one-out method was used to assess the influence of individual studies on pooled effect sizes. Where indicated, Baujat plots were used to investigate the contribution of individual studies to the overall heterogeneity.

3.1.2 HCT and variants

Lischke et al. (2020) reported two different scoring methods for the HCT, the Schandry (Schandry, 1981) and Hart methods (Hart et al., 2013). The two scoring methods tend to be highly correlated (Murphy et al., 2020), so the Schandry method was selected as it is most commonly used and it provided the most conservative estimate of sex differences in the paper where the two different scoring methods were used. One participant with below chance performance was removed from the data downloaded from an online depository for Lukowska et al. (2018) due to there being a discrepancy of N=1 between the number of participants reported in the online depository and the published manuscript. Two participants were removed from data reported by Köteles et al. (2020) because no information on sex was provided for these participants.

For the HCT and variants, males performed significantly more accurately than females, with a pooled effect size of 0.241 (p<.0001; see Forest plot Fig 2). No evidence of publication bias was observed, as indicated by the asymmetry of the funnel plot (Fig. 3) and a non-significant Egger's test (p=.768). In terms of heterogeneity, a non-significant Q statistic (Q=92.46, p=.110) and an I^2 value of 16.7% indicated low heterogeneity in the total sample. Three outliers were identified (Grabauskaitė et al., 2017; Marshall et al., 2018 Experiment 1; Schulz et al., 2013 HCT) that significantly deviated from the 95% confidence interval of the pooled effect size; however, the pooled effect size was not substantially changed by the removal of these three outliers (SMD=0.25, p<.0001) and heterogeneity was substantially reduced after removal ($I^2 = 0\%$, Q=71.27, p=.568). Influence analysis in the total sample using the leave-one-out method indicated that effect size was not substantially influenced by individual studies with effect sizes ranging from 0.23 to 0.25 all of which indicated a significant sex effect.

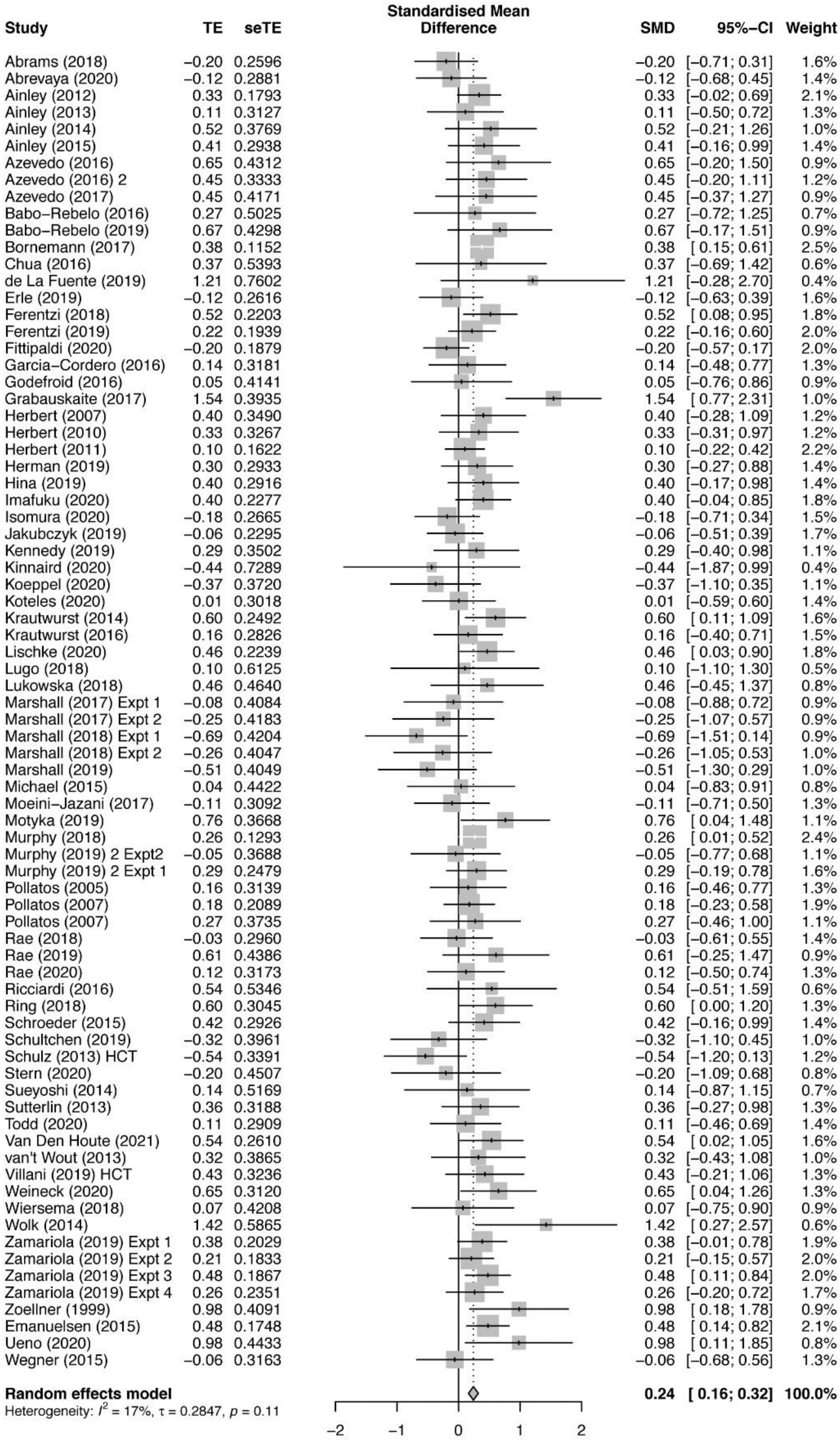

Figure 2. A forest plot showing the standardised mean difference, confidence intervals and weighting for typical participants in each HCT paper.

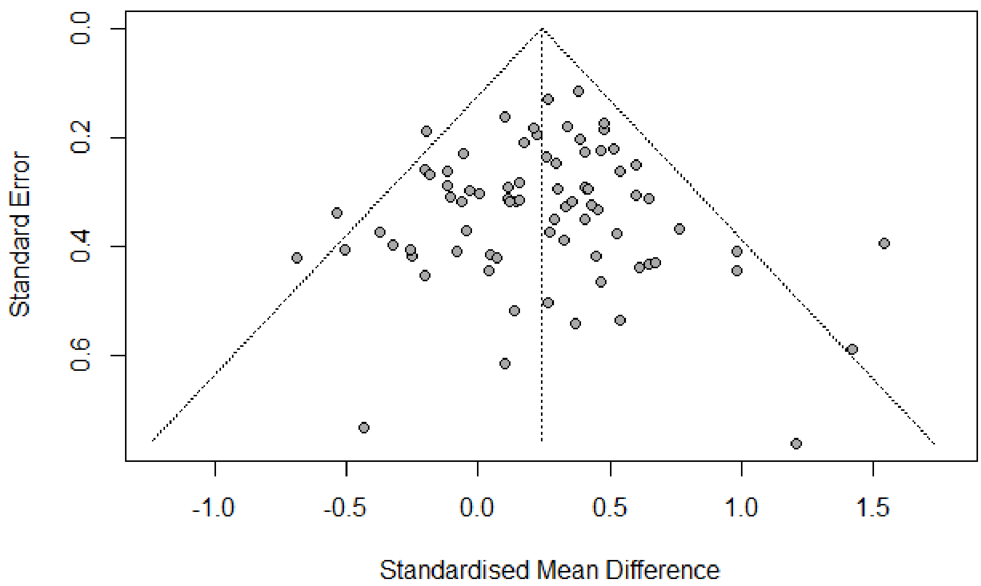

Figure 3. A funnel plot showing the relationship between the standardised mean difference and the standard error for typical participants in each HCT paper.

3.1.3 HDT and variants

Brener et al. (1993) reported both d' and A', and d' was selected as it is more commonly used. Schulz et al. (2013) administered both a visual and an auditory version of the HDT. As sex differences across the two were very similar the auditory task was selected as it is the more commonly used method. Meditators (Khalsa et al., 2020) were included as they were not considered a clinical group.

Males performed significantly more accurately on the HDT and variants than females, with a pooled effect size of 0.33 (p=.005). Individual effect sizes from each study are shown in Fig. 4. Evidence of publication bias was observed, as indicated by the funnel plot (Fig. 5) and significant Egger's test (p=.048). No outliers significantly outside of the 95% confidence interval were detected. Influence analysis using the leave-one-out method indicated that effect size was not substantially influenced by individual studies with effect sizes ranging from 0.29 to 0.36 all of which indicated a significant sex effect. A non-significant Q statistic ($Q=21.13$, $p=.17$) and an I^2 value of 24.3% both indicated low heterogeneity.

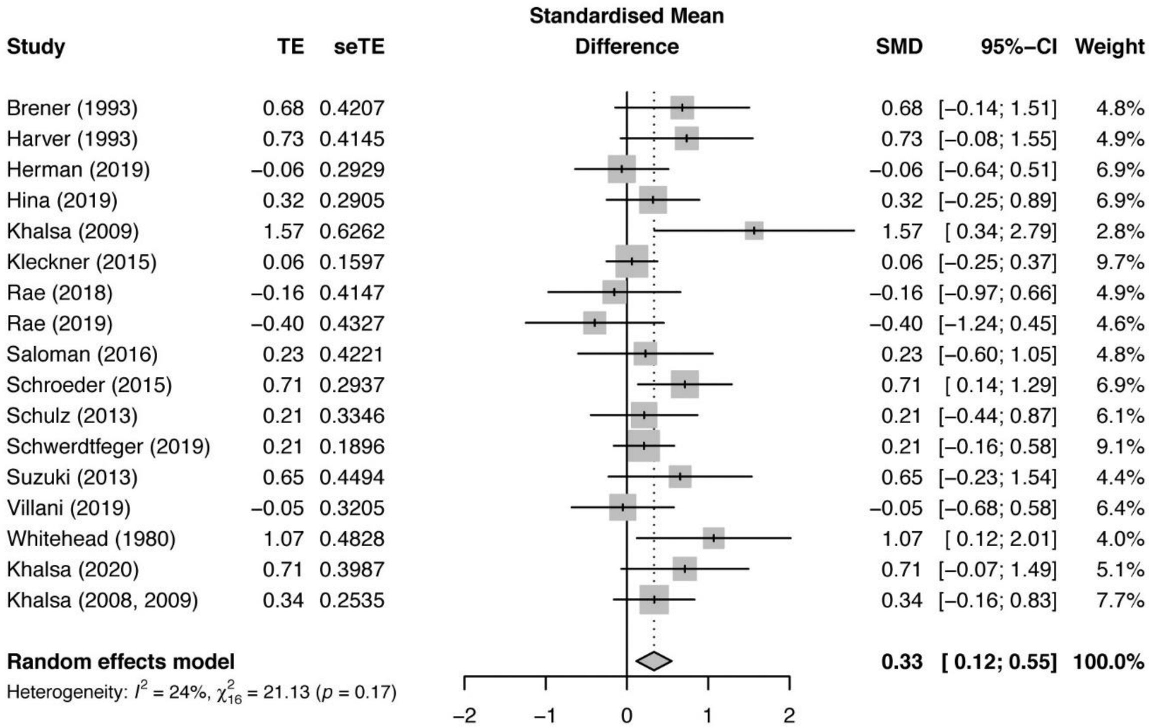

Figure 4. A forest plot showing the standardised mean difference, confidence intervals and weighting for typical participants in each HDT paper.

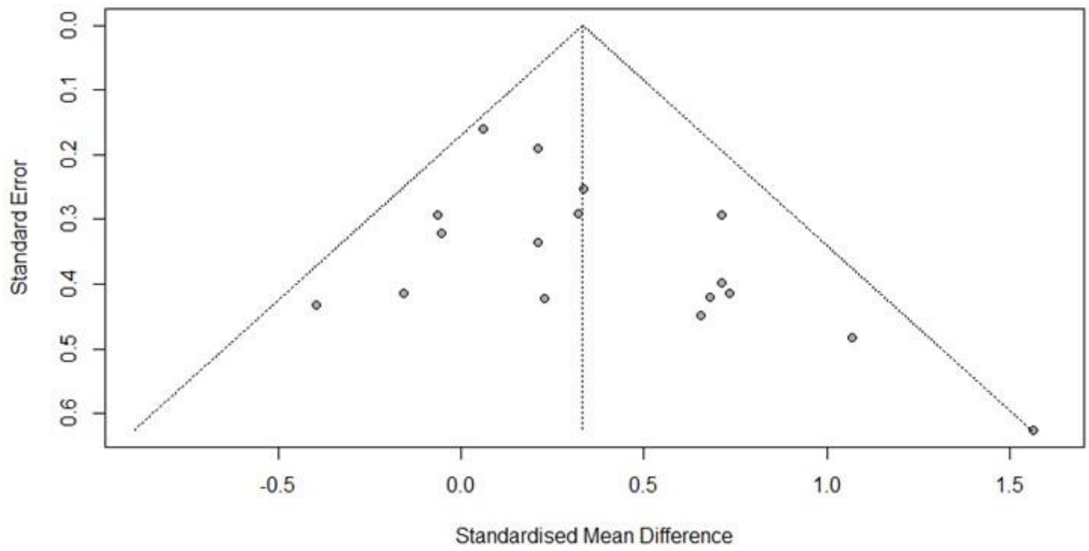

Figure 5. A funnel plot showing the relationship between the standardised mean difference and the standard error for typical participants in each HDT paper.

3.1.4 Respiratory

Preliminary analyses suggested that the results of Banzett et al. (1997) contributed towards some heterogeneity in the pooled effect size for respiratory interoceptive accuracy. As this paper also had a very small sample size (3 males, 3 females), the smallest of any of the included studies, which may contribute towards unreliability, the decision was taken to exclude this paper from all subsequent meta-analyses. Meta-analyses including Banzett et al. (1997) can be found in Appendix A. There was a discrepancy between the number of participants reported in Benke et al. (2018) and the number included in the summary statistics sent over by the authors, so the latter was used. The premature termination group in Benke et al. (2018) included those who terminated their exposure to a breathing occlusion multiple times, but did not include those with clinical diagnoses, so this group was included as typical participants. Axen et al. (1994) reported discrimination on resistive and elastic loads, and both sets of results were evaluated in two separate meta-analyses as there was no theoretical reason to select one measure over the other.

When the resistive load results were included, males performed significantly more accurately on tests of respiratory interceptive accuracy than females, with a pooled effect size of 0.41 (p=.014). Individual effect sizes from each study are shown in Fig. 6. No evidence of publication bias was observed as evidenced by the funnel plot (Fig. 7) and there was an insufficient number of studies to run an Egger’s test. No outliers significantly outside of the 95% confidence interval were detected. Influence analysis using the leave-one-out method demonstrated a range of effect sizes from 0.35 to 0.48 with only one non-significant relationship. A non-significant Q statistic (Q=3.17, p=.673) and an I^2 value of 0% both indicated low heterogeneity.

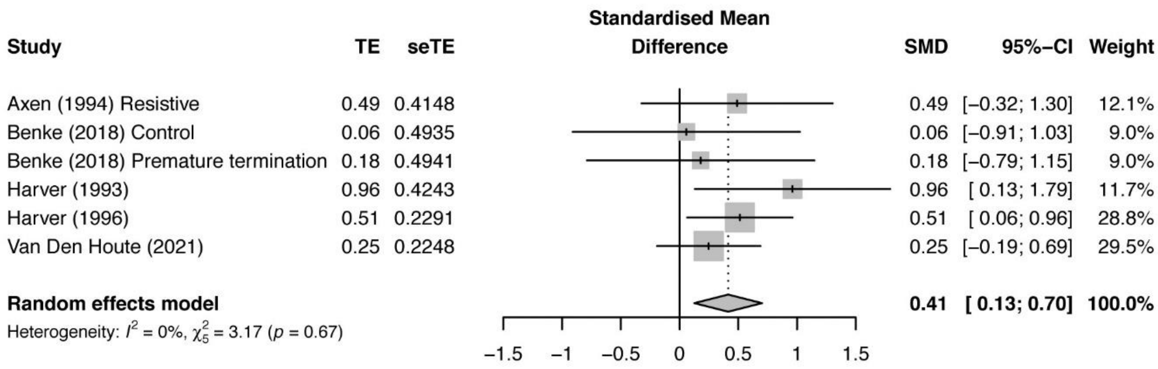

Figure 6. A forest plot showing the standardised mean difference, confidence intervals and weighting of each respiratory paper including the resistive task from Axen et al. (1994).

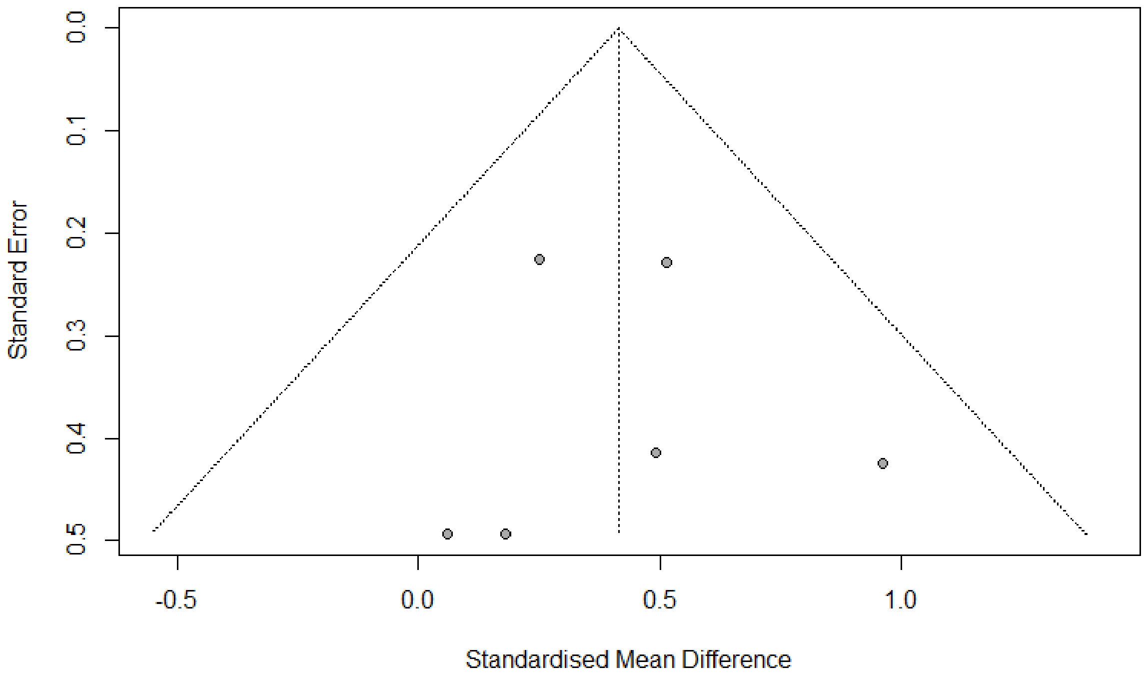

Figure 7. A funnel plot showing the relationship between the standardised mean difference and the standard error for each respiratory paper including the resistive task from Axen et al. (1994).

When the elastic load results were included, the difference between males and females was reduced to trend level (SMD=0.32, p=0.092). Individual effect sizes from each study are shown in Fig. 8. No evidence of publication bias was observed, as evidenced by the funnel plot (Fig 9.) and there was an insufficient number of studies to run an Egger's test. No outliers outside of the 95% confidence interval were detected. Influence analysis using the leave-one-out method demonstrated a range of effect sizes from 0.25 to 0.4, with all but one relationship being non-significant. When the elastic load results were removed the effect was significant with an effect size of 0.4. A non-significant Q statistic (Q=5.28, p=.383) and an I^2 value of 5.3% indicated low heterogeneity.

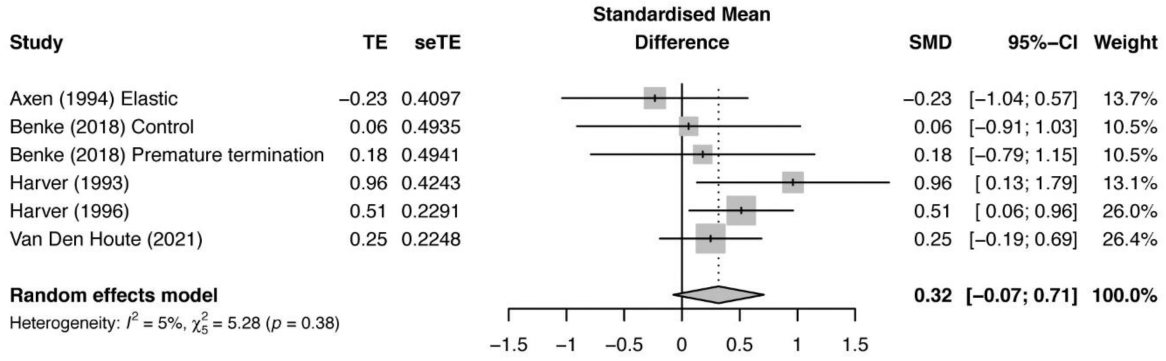

Figure 8. A forest plot showing the standardised mean difference, confidence intervals and weighting of each respiratory paper including the elastic task from Axen et al. (1994).

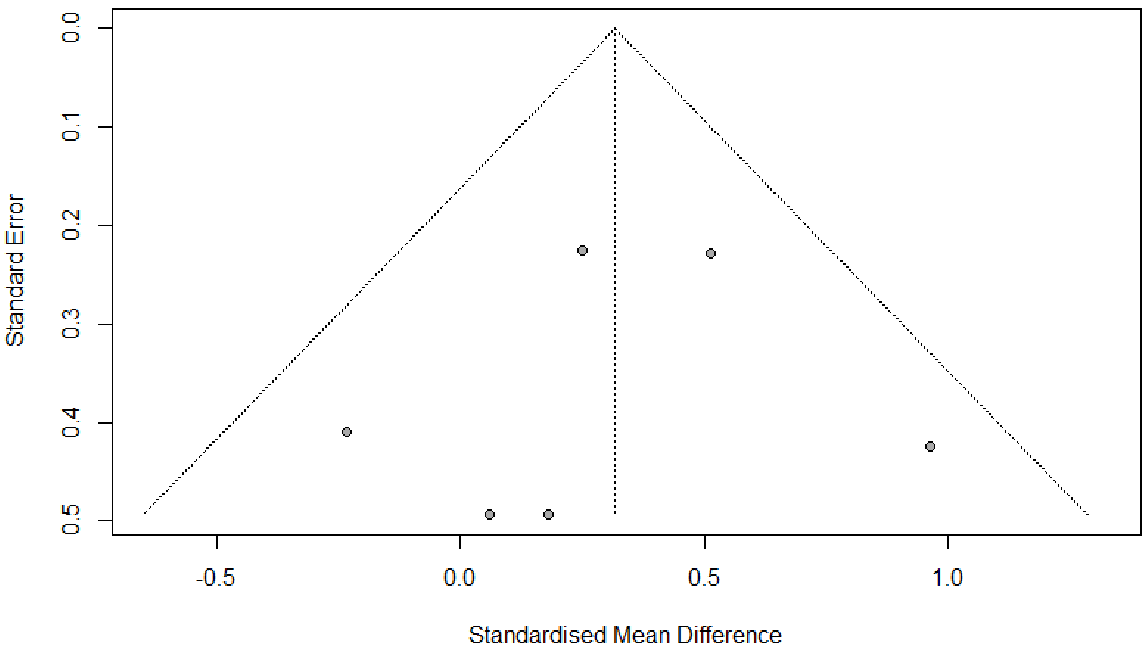

Figure 9. A funnel plot showing the relationship between the standardised mean difference and the standard error for each respiratory paper including the elastic task from Axen et al. (1994).

3.1.5 Gastric

There was no significant difference between males and females on tests of gastric interoceptive accuracy (SMD=0.09, p=.622). Individual effect sizes from each study are shown in Fig. 10. No evidence of publication bias was observed, as evidenced by the funnel plot (Fig. 11) and there was an insufficient number of studies to run an Egger's test. No outliers significantly outside of the 95% confidence interval were detected. Influence analysis using the leave-one-out method indicated some variation in effect sizes from -0.06 to 0.14, but none of these were significant. The Q statistic was non-significant (Q=3.11, p=.374) and there was an I^2 value of 3.7% suggesting low heterogeneity.

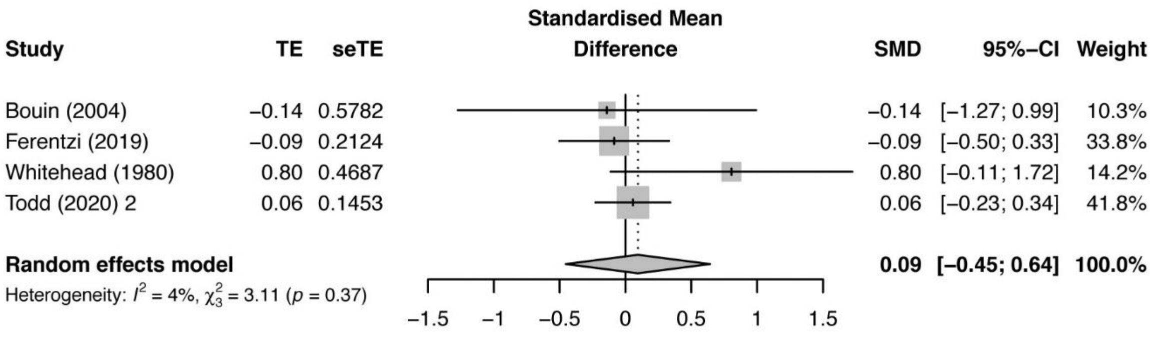

Figure 10. A forest plot showing the standardised mean differenced, confidence intervals and weighting for typical participants in each gastric paper.

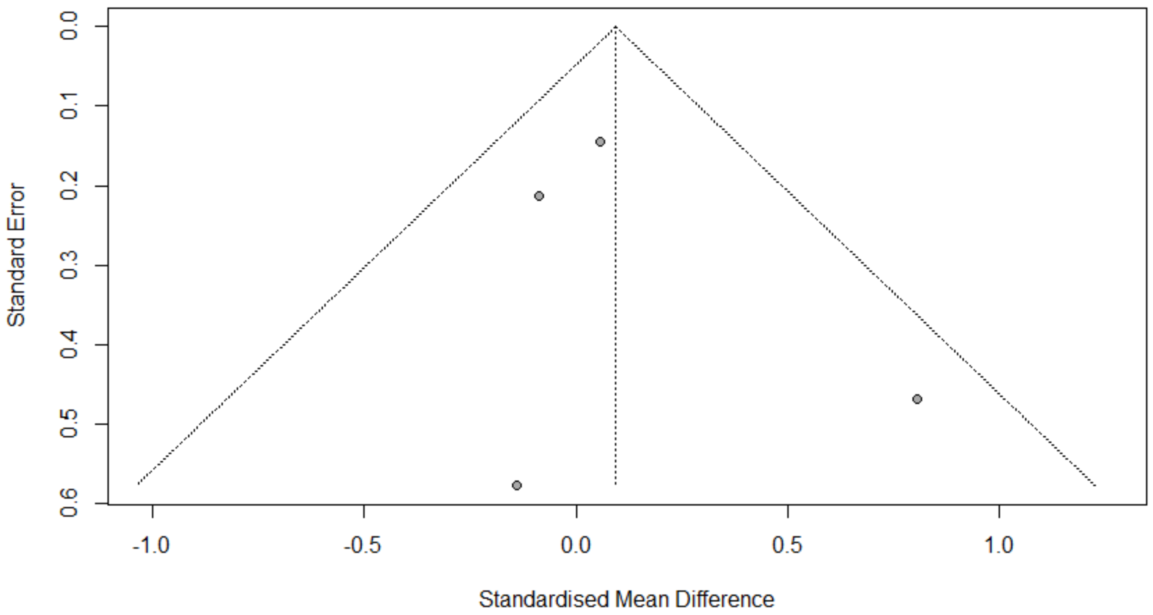

Figure 11. A funnel plot showing the relationship between the standardised mean difference and the standard error for each gastric paper.

3.1.6 Full sample comparison

As before, meditators (Khalsa et al., 2020) and the premature termination group (Benke et al., 2018) were included as neither were considered a clinical group. Banzett et al. (1997) and the outliers on the HCT meta-analysis were all excluded for the full sample comparison. Again, two meta-analyses were run to include the Axen et al. (1994) resistive and elastic load results separately. There was no significant difference in effect sizes across the different domains when including the resistive load results (Q=3.47, p =.325) and the elastic load results (Q=1.678, p=.642), indicating that the pooled effect sizes did not significantly vary across domains.

3.1.7 Subgroup analysis for cardiac interoception

We compared typical and clinical adult participant groups to see if sex differences in interoceptive accuracy were specific to one group. Outliers identified on the HCT meta-analyses were again excluded. Due to higher rates of mental health conditions in females, findings of sex differences in interoception could be driven by a higher proportion of females with mental health conditions, compared to men with mental health conditions, in the sample. Studies only reporting pooled effects across clinical and typical groups were excluded. There was no significant difference between typical and clinical adults on the HCT (Q=0.473, p=.492) and HDT (Q=0.22, p=.637) but the sex difference was found in typical (HCT: SMD=0.25, p<.001; HDT: SMD=0.33, p=.001), but not clinical, adults (HCT: SMD=0.168, p=.117; HDT: SMD=0.25, p=.088) (for forest plots see Appendix B).

Typical children and adults were also compared on the HCT but not the HDT due to an insufficient number of studies. Again, all studies previously identified as outliers were removed. There was no significant difference between groups (Q=2.15, p=0.149), with significant sex differences being present in both typical children (SMD=0.164, $p<.001$) and adults (SMD=0.25, $p<.001$).

4. Discussion

The aim of this systematic review and meta-analysis was to establish the presence or absence of sex differences in cardiac, respiratory, and gastric interoceptive accuracy. Meta-analyses revealed that males had greater interoceptive accuracy on the HCT and HDT and their variants. Sex differences were found on respiratory tasks when including the resistive load task from Axen et al. (1994), but not when the elastic load task was included, indicating that results of the respiratory analyses are less stable than those in the cardiac domain. There were no significant differences in interoceptive accuracy between males and females on gastric tasks, although there were only four studies included and substantial variability in effect sizes was noted. Overall, when considering effect sizes across domains, no significant differences were observed. In terms of subgroup analyses, males had greater interoceptive accuracy than females on the HCT and HDT for typical, but not clinical, adults. Sex differences on the HCT were found in both typical children and adult groups. As before however, no significant differences were observed when comparing subgroups.

The results of this meta-analysis replicate the findings of Desmedt et al. (2020) demonstrating consistent evidence for greater performance of males, compared to females, on the HCT. Importantly, the results of the present meta-analysis go further, confirming that despite dissociations between HCT and HDT tasks and their variants (Hickman et al., 2020), a male advantage is also observed on HDT tasks and their variants, though some evidence of publication bias was observed. Whilst only a small, but significant, pooled effect size was observed, the results of these two meta-analyses suggest that previously reported sex differences on the HCT are not likely due to methodological issues with this task specifically. Nevertheless, it should be noted that methodological concerns have also been raised regarding the use of the 2AFC version of the HDT as it does not control for individual differences in the delay at which an individual perceives an external stimulus to be synchronous with their heartbeat (Brener & Ring, 2016). Whilst sex differences in the delay at which an individual perceives a signal to be synchronous with their heartbeat have to our knowledge not been examined, given sex differences in cardiac parameters (Katkin, 1985; Pennebaker, 1982; Shephard & Miller, 1998), it is conceivable that males could have a physiological advantage on the 2AFC if the typical delays selected as synchronous and asynchronous (e.g., 200-250ms, 500-500ms) map more strongly onto the preferred delays of males, compared to females. However, when considering only studies that used a multi-level version of the HDT (Schroeder et al., 2015; typical sample) or perturbation methods (Khalsa et al., 2020; Khalsa, Rudrauf, Sandesara, et al., 2009), effect sizes ranged from (0.71-1.57) suggesting that it is unlikely that the sex differences observed here reflect a male advantage attributable to the 2AFC HDT format. Nevertheless, it should be noted that more recently developed measures of cardiac interoceptive accuracy have not observed sex differences in cardiac interoception (Plans et al., 2021). As such, further work is required to establish the presence of sex differences in cardiac interoception and to investigate the influence of task format on performance.

Compared to the findings from studies of cardiac interoception, the pooled results from studies of respiratory and gastric interoceptive accuracy were far less stable, likely due to the small number of studies resulting in each study exerting substantial influence on the pooled effect size. Unlike cardiac interoception, where the same tasks have been used across multiple studies, studies of gastric and respiratory interoception included a wide variety of methodologies, which likely contributed towards increased variability. Indeed, when considering respiratory interoception specifically, it is notable that the results of the meta-analysis were substantially altered when considering the resistive and elastic load tasks conducted in the same participant group (Axen et al., 1994), with a male advantage observed for the former and a female advantage observed for the latter. Whilst a significant male advantage was observed when including the results from resistive loads within the pooled effect size, the same was not observed when elastic loads were included. As such, caution is required when interpreting this result and further work is required to establish the presence or absence of sex differences for respiratory interoceptive accuracy. Nevertheless, such observations serve to highlight previous concerns that task format may considerably influence results obtained in tasks of cardiac interoceptive accuracy (Hickman et al., 2020) and go further to underscore that the same may be true for tasks of respiratory interoception.

Whilst some evidence for sex differences in interoceptive accuracy were observed for cardiac and respiratory interoception, no significant difference was observed for gastric interoception. Such a result may be surprising considering previous evidence that performance on tasks of gastric interoception is typically correlated with performance on tasks of cardiac interoception (e.g., Herbert et al., 2012; Whitehead & Drescher, 1980; but see Ferentzi et al., 2019). Like respiratory interoception, it is likely that differences in task methodology play a role. Two of the studies included in the gastric meta-analysis utilised a modified version of the water load task (Ferentzi et al., 2019; Todd, Aspell et al., 2020) which involved some manipulation of physiological states. The water load task measures the perception of satiation which may be influenced by beliefs about stomach capacity and feelings of fullness, and these could differ across the sexes. Although speculative, beliefs about reduced stomach capacity and appetite in females may lead females to stop drinking at an earlier point, which is then misinterpreted as greater accuracy.

Notably, the only study reporting a significant sex difference for gastric interoceptive accuracy required participants to report whether an external light was synchronous or asynchronous with a stomach contraction (Whitehead & Drescher, 1980), a task format similar to heartbeat discrimination methods (Whitehead et al., 1977). It is therefore possible that sex differences in interoception are most apparent where tasks involve the integration of internal and external stimuli. Whilst one possibility is that sex differences are the result of differences in multisensory integration or in other functions critical for task performance such as attention to and/or discrimination of stimuli, results from studies utilising control tasks are not consistent with this possibility; for example, where control tasks were used for tasks of respiratory interoceptive accuracy, including light-tone discrimination (Harver et al., 1993) or auditory discrimination (Van Den Houte et al., 2021), sex differences were not observed on control tasks. Nevertheless, the above concerns highlight the importance of including control tasks when examining group differences in tasks of interoceptive ability (Murphy, Brewer, et al., 2018) and underscore a need for further research examining sex differences in gastric interoceptive accuracy.

Despite the importance of these results for establishing the presence or absence of sex differences in interoception, these results cannot establish whether sex differences in interoception result from physiological differences between the sexes or differences in psychological processing of internal sensations. Although both possibilities would be clinically relevant as they would both result in differences in interoceptive ability in one’s everyday daily life, understanding the cause(s) of sex differences in interoception may aid understanding of where individual differences stem from, why different task formats influence results obtained, and may ultimately inform the selection of interventions where appropriate. Perhaps surprisingly, several studies have found that sex differences on the HCT remained after controlling for differences in body mass index (Koch & Pollatos, 2014) and several other parameters (including, heart rate, heart rate variability, systolic blood pressure, age, time perception ability etc.; Murphy, Brewer, et al., 2018) suggesting that the differences are not explained by physiological parameters alone. However, whether the same is true for other measures of cardiac interoception, or whether other physiological parameters (e.g., cardiac output) can explain sex differences, remains to be examined.

Like cardiac interoception, it is also conceivable that physiological differences between the sexes may explain differences in respiratory and gastric interoceptive accuracy, at least for certain tasks; for example, sex differences in respiratory physiology have been identified, most notably in lung size (Thurlbeck, 1982). Lung capacity has been shown to contribute to differences in perception of breathlessness, with more females reporting breathlessness than males due to reduced absolute volumes (Ekström et al., 2018). Van Den Houte et al. (2021) speculated that certain features of respiratory occlusions could be processed more readily by those who breathe more deeply. They found that performance on the respiratory interoceptive accuracy task was weakly correlated with breathing behaviour and explained 7% of the variance in task performance. It is conceivable that these kind of individual differences in breathing behaviour could vary by sex, potentially accounting for sex differences in interoceptive ability. Likewise, gastric tasks may also be influenced by sex differences in gastric physiology; for example, females have been found to have greater fluid retention and bloating (Camps et al., 2018; Jiang et al., 2008) and higher rates of functional gastrointestinal disorders than males (Lovell & Ford, 2012), which may relate to hormonal fluctuations (Heitkemper & Chang, 2009; Mulak et al., 2014). These differences could all conceivably contribute towards females demonstrating greater (or at least equivalent) gastric sensitivity on certain tasks. Whilst some studies included in this review attempted to control for some differences by adjusting for height (Ferentzi et al., 2019), it is possible that there are other physiological factors that may differ between males and females and influence performance on these tasks, and questions remain as to whether height and weight are adequate predictors of stomach capacity (Cox, 1945). Taken together, it is clear that further work is required to understand the influence of physiological differences on sex differences in interoception.

Finally, it is notable that sub-analyses examining performance on the HCT in typical adult and child groups revealed significant sex differences in both groups, suggesting that this difference is consistent across the lifespan. Whilst this result is somewhat at odds with previously reported suggestions that sex differences in interoception may arise from the increased amount of physical and hormonal change females experience across the lifespan (Murphy, Viding et al., 2019), it is notable that the majority of the child samples included also covered at least early adolescence, a period of substantial physical and hormonal change. Although only one study examined sex differences in interoception in a small sample of young children (aged 4-6 years), it is notable that sex differences were not observed (Schaan et al., 2019). As such, further work is required to establish the time point at which sex differences in interoception arise and the extent to which differences are attributable to physical change, psychological change, or social influence across development. In other subgroup analyses for cardiac interoception, it should also be noted that sex differences for the HCT and HDT were only observed when considering typical, but not clinical, adults (although, no significant difference was observed when comparing effect sizes). This is consistent with the theory that interoceptive accuracy may be disrupted in multiple mental and physical health conditions (Barrett & Simmons, 2015; Brewer et al., 2021; Khalsa et al., 2018; Murphy et al., 2017), as we would expect both males and females with those conditions to show similarly disrupted interoception. Nevertheless, it is of course possible that sex differences remain within certain conditions but are obscured by the inclusion of multiple conditions in the clinical meta-analysis.

Despite the importance of these results for understanding sex differences in interoception, it is important to consider limitations. One major limitation of this series of meta-analyses is the inclusion of only a small number of studies with a variety of different methodologies in the respiratory and gastric meta-analyses, contributing substantial heterogeneity and/or instability to the results. However, research into cardiac interoception has shown that the overreliance on one type of task can also be problematic due to the contribution of non-interoceptive factors to task performance and the lack of correlation between different interoceptive accuracy tasks within the same domain (Hickman et al., 2020). The recent development of several different novel tasks of respiratory and gastric interoceptive accuracy (Harrison et al., 2020; Van Den Houte et al., 2021; van Dyck et al., 2016), and tasks that involve the active manipulation of cardiac signals (Khalsa et al., 2020; Khalsa, Rudrauf, Sandesara, et al., 2009), may help to provide more bias-free measures of interoceptive accuracy. These new tasks, along with increased interest in research in the field (Ceunen et al., 2016), should allow for a more robust analysis of sex differences in these interoceptive domains in the future.

Second, one difficulty with studying sex differences in interoceptive accuracy is that we cannot establish whether findings reflect true differences in interoceptive accuracy, or whether differences result from non-interoceptive task factors such as the influence of physiology on task performance outwith the effect of physiology on interoceptive accuracy itself. More studies are required to examine the relative contribution of such factors to observed sex differences in interoceptive accuracy. Another limitation of this study was that due to the large number of studies identified by the PubMed search (7542) no other database was searched, although an additional citation search was done to ensure relevant studies were not missed. A further limitation is that it was not possible to assess publication bias for respiratory and gastric domains due to small study numbers (K<10) and analyses for some subgroup analyses were underpowered. Small study numbers are a limitation across other interoceptive domains and is an additional reason why rectal, urinary, and oesophageal domains, among others, were not examined. However, further research into these domains in the future may be promising for examining whether sex differences in interoceptive accuracy extend into other domains.

Despite these limitations, this meta-analysis provides initial evidence that sex differences in interoceptive accuracy may exist across different tasks and domains. However, it does not directly address the possible reasons for these sex differences, which may be physiological or psychological, nor does it address the consequences of such individual differences. Whilst physiological differences may contribute towards sex differences, the finding that males appear to have greater interoceptive accuracy than females across several different tasks and domains may suggest that a more general factor underlies sex differences. Given previously reported links between interoception and mental health, these results raise the possibility that sex differences in interoception may contribute, in part, to sex differences in mental health. Future research should therefore aim to establish the cause of the sex differences observed here, and the consequences of such differences for both mental and physical health.

Funding

Jennifer Murphy was supported by a BA/Leverhulme Small Research Grant [SRG20/200146].

Acknowledgements

We are ever so grateful to the authors who responded to our request for data, whether those data were available or not. Thank you for supporting our work.

Appendices

Appendix A: Respiratory meta-analyses including Banzett et al. (1997)

Meta-analyses were carried out with data from Banzett et al. (1997) included for both the respiratory resistive and elastic analyses.

When Banzett et al. (1997) and the resistive load results from Axen et al. (1994) were included, there was no significant difference between males and females on tests of respiratory interoceptive accuracy, with a pooled effect size of 0.30 (p=.228). Individual effect sizes from each study are shown in Fig A1. No evidence of publication bias was observed as evidenced by the funnel plot (Fig A2) and there was an insufficient number of studies to run an Egger's test. No outliers significantly outside of the 95% confidence interval were detected. Influence analysis using the leave-one-out method demonstrated a range of effect sizes from 0.20 to 0.41 with all but one relationship being non-significant. When Banzett et al. (1997) was removed the effect was significant with an effect size of 0.41. A non-significant Q statistic (Q=7.36, p=.289) and an I^2 value of 18.5% both indicated low heterogeneity.

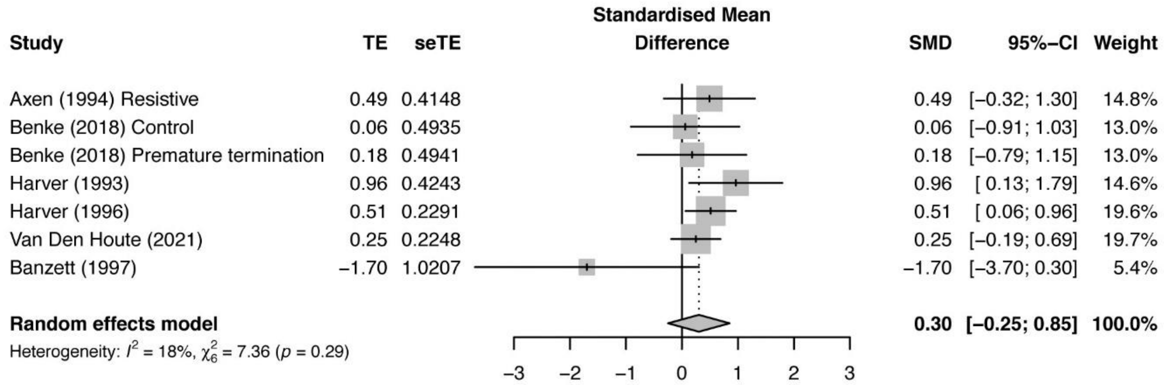

Figure A1. A forest plot showing the standardised mean difference, confidence intervals and weighting of each respiratory paper including the resistive task from Axen et al. (1994) and Banzett et al. (1997).

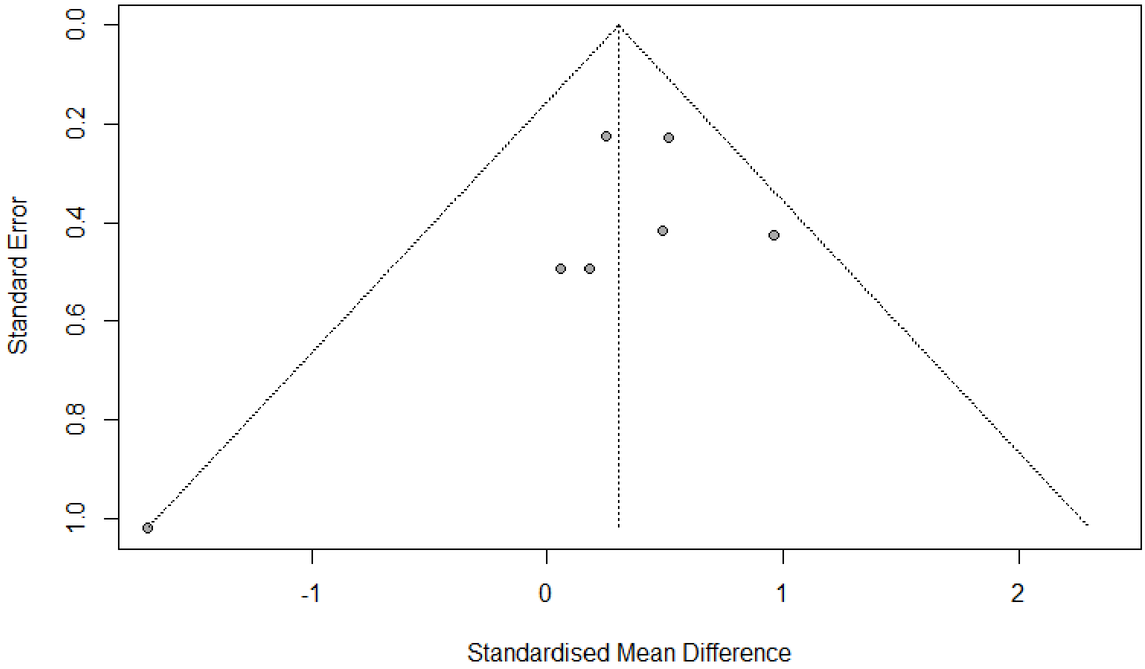

Figure A2. A funnel plot showing the relationship between the standardised mean difference and the standard error for each respiratory paper including the resistive task from Axen et al. (1994) and Banzett et al. (1997).

When Banzett et al. (1997) and the elastic load results from Axen et al. (1994) were included, there was again no significant difference between males and females (SMD=0.19, p=0.446). Individual effect sizes from each study are shown in Fig A3. No evidence of publication bias was observed, as evidenced by the funnel plot (Fig A4) and there was an insufficient number of studies to run an Egger's test. No outliers significantly outside of the 95% confidence interval were detected. Influence analysis using the leave-one-out method demonstrated a range of effect sizes from 0.08 to 0.32, with no significant relationships. The Q statistic was non-significant (Q=9.18, p=.164) and there was an I^2 value of 34.6% suggesting moderate heterogeneity.

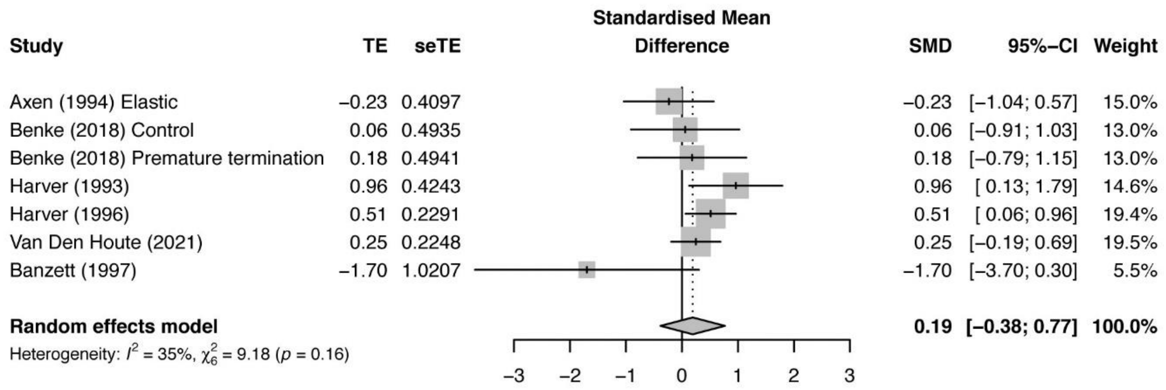

Figure A3. A forest plot showing the standardised mean difference, confidence intervals and weighting for each respiratory paper including the elastic task from Axen et al. (1994) and Banzett et al. (1997).

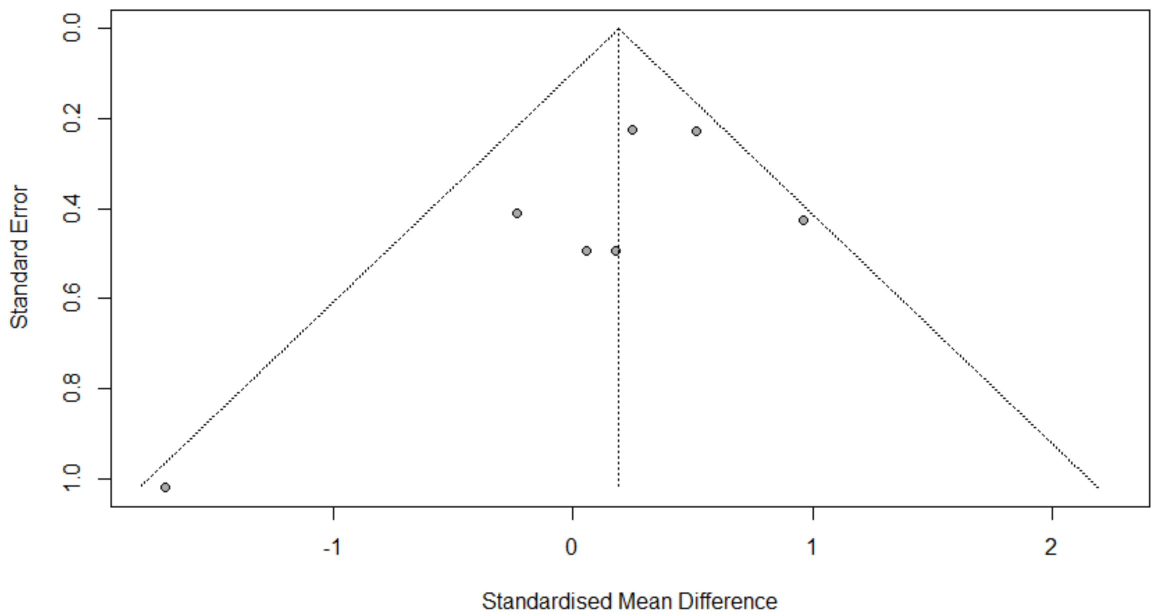

Figure A4. A funnel plot showing the relationship between the standardised mean difference and the standard error for each respiratory paper including the elastic task from Axen et al. (1994) and Banzett et al. (1997).

Appendix B: Forrest plots for cardiac subgroup analyses

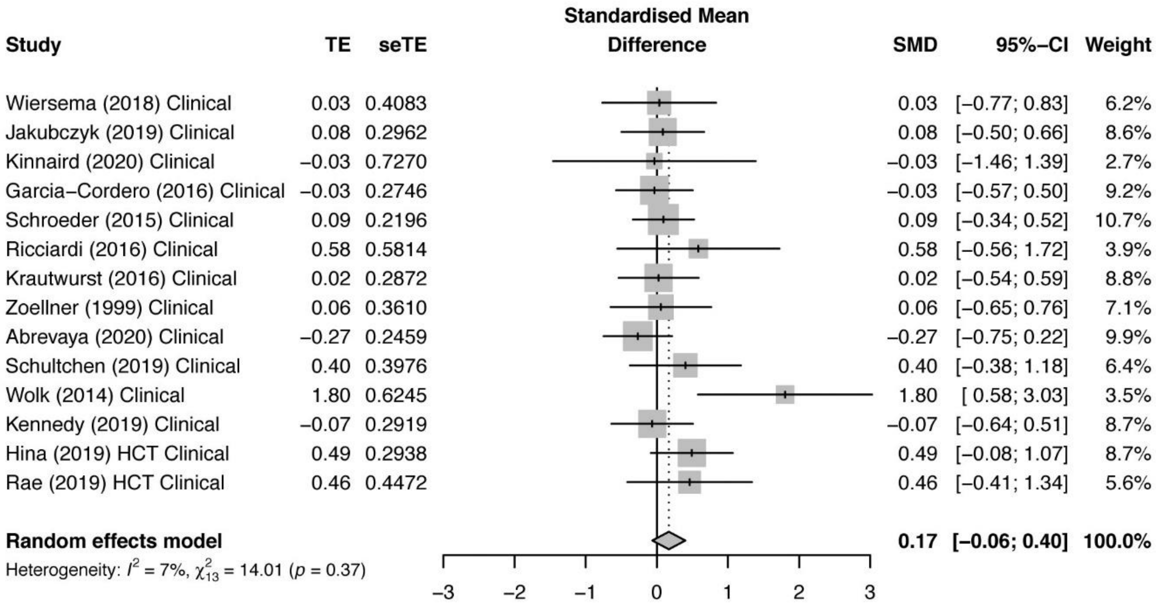

Figure B1. A forest plot showing the standardised mean difference, confidence intervals and weighting of the clinical adult subgroup for the HCT papers.

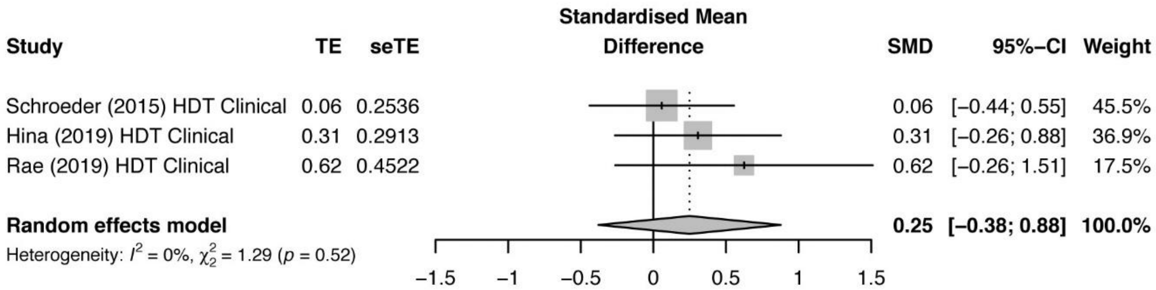

Figure B2. A forest plot showing the standardised mean difference, confidence intervals and weighting of the clinical adult subgroup for the HDT papers.

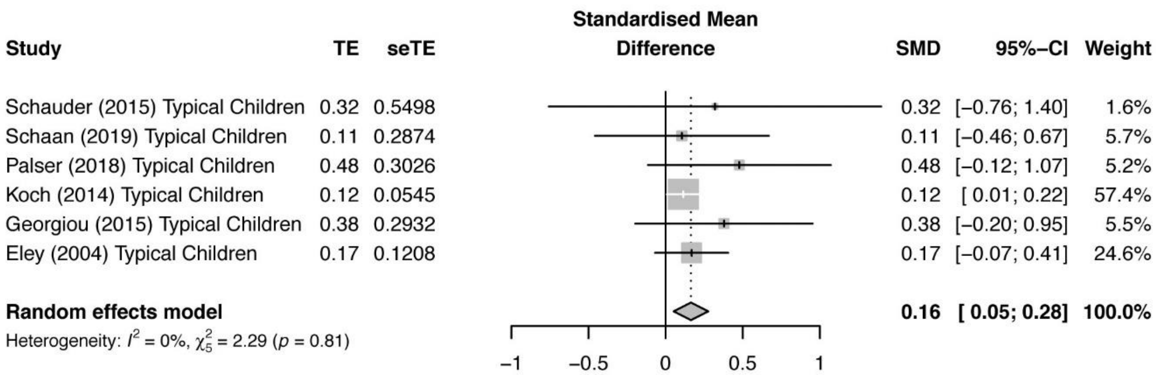

Figure B3. A forest plot showing the standardised mean difference, confidence intervals and weighting of the typical children subgroup for the HCT papers.

Appendix C: Abbreviations table

Abbreviation	***Description***
HCT	Heartbeat counting task
HDT	Heartbeat discrimination task
2AFC	Two-alternative forced choice
6AFC	Six-alternative forced choice
MCS	Method of constant stimuli

References

Abrams, K., Cieslowski, K., Johnson, S., Krimmel, S., La Rosa, G. B., Barton, K., & Silverman, P. (2018, Apr). The effects of alcohol on heartbeat perception: Implications for anxiety. *Addictive Behaviors, 79*, 151-158. https://doi.org/10.1016/j.addbeh.2017.12.023

Abrevaya, S., Fittipaldi, S., Garcia, A. M., Dottori, M., Santamaria-Garcia, H., Birba, A., Yoris, A., Hildebrandt, M. K., Salamone, P., De la Fuente, A., Alarco-Marti, S., Garcia-Cordero, I., Matorrel-Caro, M., Pautassi, R. M., Serrano, C., Sedeno, L., & Ibanez, A. (2020, Nov-Dec). At the heart of neurological dimensionality: Cross-nosological and multimodal cardiac interoceptive deficits. *Psychosomatic Medicine, 82*(9), 850-861. https://doi.org/10.1097/PSY.0000000000000868

Ainley, V., Brass, M., & Tsakiris, M. (2014, Jul). Heartfelt imitation: High interoceptive awareness is linked to greater automatic imitation. *Neuropsychologia*, *60*, 21-28. https://doi.org/10.1016/j.neuropsychologia.2014.05.010

Ainley, V., Maister, L., Brokfeld, J., Farmer, H., & Tsakiris, M. (2013, Dec). More of myself: manipulating interoceptive awareness by heightened attention to bodily and narrative aspects of the self. *Consciousness and Cognition*, *22*(4), 1231-1238. https://doi.org/10.1016/j.concog.2013.08.004

Ainley, V., Maister, L., & Tsakiris, M. (2015). Heartfelt empathy? No association between interoceptive awareness, questionnaire measures of empathy, reading the mind in the eyes task or the director task. *Frontiers in Psychology*, *6*, 554. https://doi.org/10.3389/fpsyg.2015.00554

Ainley, V., Tajadura-Jimenez, A., Fotopoulou, A., & Tsakiris, M. (2012, Nov). Looking into myself: Changes in interoceptive sensitivity during mirror self-observation.

Psychophysiology, 49(11), 1504-1508. https://doi.org/10.1111/j.1469-8986.2012.01468.x

Ardizzi, M., Ambrosecchia, M., Buratta, L., Ferri, F., Peciccia, M., Donnari, S., Mazzeschi, C., & Gallese, V. (2016). Interoception and positive symptoms in schizophrenia. *Frontiers in Human Neuroscience, 10,* 379. https://doi.org/10.3389/fnhum.2016.00379

Axen, K., Noble, D., Zimmer, S., Capponi, L., & Haas, F. (1994, Jan). Discrimination of transiently applied mechanical loads: breathing vs. pulling. *Journal of Applied Physiology, 76*(1), 14-25. https://doi.org/10.1152/jappl.1994.76.1.14

Azevedo, R. T., Aglioti, S. M., & Lenggenhager, B. (2016, May 23). Participants' above-chance recognition of own-heart sound combined with poor metacognitive awareness suggests implicit knowledge of own heart cardiodynamics. *Scientific Reports, 6*, 26545. https://doi.org/10.1038/srep26545

Azevedo, R. T., Garfinkel, S. N., Critchley, H. D., & Tsakiris, M. (2017, Jan 17). Cardiac afferent activity modulates the expression of racial stereotypes. *Nature Communications, 8,* 13854. https://doi.org/10.1038/ncomms13854

Babo-Rebelo, M., Buot, A., & Tallon-Baudry, C. (2019, May 1). Neural responses to heartbeats distinguish self from other during imagination. *Neuroimage, 191*, 10-20. https://doi.org/10.1016/j.neuroimage.2019.02.012

Babo-Rebelo, M., Richter, C. G., & Tallon-Baudry, C. (2016, Jul 27). Neural responses to heartbeats in the default network encode the self in spontaneous thoughts. *Journal of Neuroscience, 36*(30), 7829-7840. https://doi.org/10.1523/JNEUROSCI.0262-16.2016

Balduzzi, S., Rücker, G., & Schwarzer, G. (2019). How to perform a meta-analysis with R: a practical tutorial. *Evidence Based Mental Health, 22*(4), 153-160. https://doi.org/10.1136/ebmental-2019-300117

Banzett, R. B., Shea, S. A., Brown, R., Schwartzstein, R., Lansing, R., & Guz, A. (1997, Feb). Perception of inflation of a single lung lobe in humans. *Respiration Physiology, 107*(2), 125-136. https://doi.org/10.1016/s0034-5687(96)02510-8

Barrett, L. F., & Simmons, W. K. (2015, Jul). Interoceptive predictions in the brain. *Nature Reviews: Neuroscience, 16*(7), 419-429. https://doi.org/10.1038/nrn3950

Benke, C., Krause, E., Hamm, A. O., & Pané-Farré, C. A. (2018, Sep). Dynamics of defensive response mobilization during repeated terminations of exposure to increasing interoceptive threat. *International Journal of Psychophysiology, 131*, 44-56. https://doi.org/10.1016/j.ijpsycho.2017.09.013

Bornemann, B., & Singer, T. (2017, Mar). 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

Bouin, M., Lupien, F., Riberdy, M., Boivin, M., Plourde, V., & Poitras, P. (2004, Jun). Intolerance to visceral distension in functional dyspepsia or irritable bowel syndrome: An organ specific defect or a pan intestinal dysregulation? *Neurogastroenterology and Motility, 16*(3), 311-314. https://doi.org/10.1111/j.1365-2982.2004.00511.x

Brener, J., & Kluvitse, C. (1988, Sep). Heartbeat detection: judgments of the simultaneity of external stimuli and heartbeats. *Psychophysiology, 25*(5), 554-561. https://doi.org/10.1111/j.1469-8986.1988.tb01891.x

Brener, J., Liu, X., & Ring, C. (1993, Nov). A method of constant stimuli for examining heartbeat detection: Comparison with the Brener-Kluvitse and Whitehead methods. *Psychophysiology, 30*(6), 657-665. https://doi.org/10.1111/j.1469-8986.1993.tb02091.x

Brener, J., & Ring, C. (2016, Nov 19). Towards a psychophysics of interoceptive processes: The measurement of heartbeat detection. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 371*(1708). https://doi.org/10.1098/rstb.2016.0015

Brewer, R., Murphy, J., & Bird, G. (2021, Aug 3). Atypical interoception as a common risk factor for psychopathology: A review. *Neuroscience & Biobehavioural Reviews*. https://doi.org/10.1016/j.neubiorev.2021.07.036

Camps, G., de Graaf, K., & Smeets, P. A. M. (2018, Dec 1). Men and women differ in gastric fluid retention and neural activation after consumption of carbonated beverages. *Journal of Nutrition*, *148*(12), 1976-1983. https://doi.org/10.1093/jn/nxy230

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., Meier, M. H., Ramrakha, S., Shalev, I., Poulton, R., & Moffitt, T. E. (2014, Mar). The p factor: One general psychopathology factor in the structure of psychiatric disorders? *Clinical Psychological Science*, *2*(2), 119-137. https://doi.org/10.1177/2167702613497473

Ceunen, E., Vlaeyen, J. W., & Van Diest, I. (2016). On the origin of interoception. *Frontiers in Psychology*, *7*, 743. https://doi.org/10.3389/fpsyg.2016.00743

Chua, E. F., & Bliss-Moreau, E. (2016, Oct). Knowing your heart and your mind: The relationships between metamemory and interoception. *Consciousness and Cognition, 45*, 146-158. https://doi.org/10.1016/j.concog.2016.08.015

Cox, A. J. (1945, Dec). Variations in size of human stomach. *California and Western Medicine, 63*, 267.

Craig, A. D. (2002, Aug). How do you feel? Interoception: the sense of the physiological condition of the body. *Nature Reviews: Neuroscience, 3*(8), 655-666. https://doi.org/10.1038/nrn894

Daubenmier, J., Sze, J., Kerr, C. E., Kemeny, M. E., & Mehling, W. (2013, Aug). Follow your breath: Respiratory interoceptive accuracy in experienced meditators. *Psychophysiology, 50*(8), 777-789. https://doi.org/10.1111/psyp.12057

de la Fuente, A., Sedeno, L., Vignaga, S. S., Ellmann, C., Sonzogni, S., Belluscio, L., Garcia-Cordero, I., Castagnaro, E., Boano, M., Cetkovich, M., Torralva, T., Canepa, E. T., Tagliazucchi, E., Garcia, A. M., & Ibanez, A. (2019, Jul). Multimodal neurocognitive markers of interoceptive tuning in smoked cocaine. *Neuropsychopharmacology, 44*(8), 1425-1434. https://doi.org/10.1038/s41386-019-0370-3

Desmedt, O., Luminet, O., & Corneille, O. (2018, Oct). 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

Desmedt, O., Van Den Houte, M., Walentynowicz, M., Dekeyser, S., Luminet, O., & Corneille, O. (2020). A systematic review and meta-analysis on the association between heartbeat counting task performance and mental disorders and their risk factors among adults. https://doi.org/10.31219/osf.io/h3by9

Dunn, B. D., Dalgleish, T., Ogilvie, A. D., & Lawrence, A. D. (2007, Aug). Heartbeat perception in depression. *Behavioural Research and Therapy, 45*(8), 1921-1930. https://doi.org/10.1016/j.brat.2006.09.008

Ehlers, A., & Breuer, P. (1992, Aug). Increased cardiac awareness in panic disorder. *Journal of Abnormal Psychology*, *101*(3), 371-382. https://doi.org/10.1037//0021-843x.101.3.371

Ekström, M., Sundh, J., Schiöler, L., Lindberg, E., Rosengren, A., Bergström, G., Angerås, O., Hedner, J., Brandberg, J., Bake, B., & Torén, K. (2018). Absolute lung size and the sex difference in breathlessness in the general population. *PloS One*, *13*(1), e0190876. https://doi.org/10.1371/journal.pone.0190876

Eley, T. C., Stirling, L., Ehlers, A., Gregory, A. M., & Clark, D. M. (2004, Apr). Heart-beat perception, panic/somatic symptoms and anxiety sensitivity in children. *Behavioural Research Therapy*, *42*(4), 439-448. https://doi.org/10.1016/s0005-7967(03)00152-9

Emanuelsen, L., Drew, R., & Koteles, F. (2015, Apr). Interoceptive sensitivity, body image dissatisfaction, and body awareness in healthy individuals. *Scandinavian Journal of Psychology*, *56*(2), 167-174. https://doi.org/10.1111/sjop.12183

Erle, T. M. (2019). Level-2 visuo-spatial perspective-taking and interoception - More evidence for the embodiment of perspective-taking. *PloS One*, *14*(6), e0219005. https://doi.org/10.1371/journal.pone.0219005

Failla, M. D., Bryant, L. K., Heflin, B. H., Mash, L. E., Schauder, K., Davis, S., Gerdes, M. B., Weitlauf, A., Rogers, B. P., & Cascio, C. J. (2020, Jun). Neural correlates of cardiac interoceptive focus across development: Implications for social symptoms in autism spectrum disorder. *Autism Research*, *13*(6), 908-920. https://doi.org/10.1002/aur.2289

Ferentzi, E., Drew, R., Tihanyi, B. T., & Köteles, F. (2018, Feb 1). Interoceptive accuracy and body awareness - Temporal and longitudinal associations in a non-clinical sample. *Physiology & Behaviour, 184*, 100-107. https://doi.org/10.1016/j.physbeh.2017.11.015

Ferentzi, E., Horváth, Á., & Köteles, F. (2019, Apr). Do body-related sensations make feel us better? Subjective well-being is associated only with the subjective aspect of interoception. *Psychophysiology, 56*(4), e13319. https://doi.org/10.1111/psyp.13319

Field, A. P. (2001, Jun). Meta-analysis of correlation coefficients: A Monte Carlo comparison of fixed- and random-effects methods. *Psychological Methods, 6*(2), 161-180. https://doi.org/10.1037/1082-989x.6.2.161

Fittipaldi, S., Abrevaya, S., Fuente, A., Pascariello, G. O., Hesse, E., Birba, A., Salamone, P., Hildebrandt, M., Martí, S. A., Pautassi, R. M., Huepe, D., Martorell, M. M., Yoris, A., Roca, M., García, A. M., Sedeño, L., & Ibáñez, A. (2020, May 15). A multidimensional and multi-feature framework for cardiac interoception. *Neuroimage, 212*, 116677. https://doi.org/10.1016/j.neuroimage.2020.116677

Garcia-Cordero, I., Sedeno, L., de la Fuente, L., Slachevsky, A., Forno, G., Klein, F., Lillo, P., Ferrari, J., Rodriguez, C., Bustin, J., Torralva, T., Baez, S., Yoris, A., Esteves, S., Melloni, M., Salamone, P., Huepe, D., Manes, F., Garcia, A. M., & Ibanez, A. (2016, Nov 19). Feeling, learning from and being aware of inner states: Interoceptive dimensions in neurodegeneration and stroke. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 371*(1708). https://doi.org/10.1098/rstb.2016.0006

Garfinkel, S. N., Manassei, M. F., Hamilton-Fletcher, G., In den Bosch, Y., Critchley, H. D., & Engels, M. (2016, Nov 19). Interoceptive dimensions across cardiac and respiratory axes. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 371*(1708). https://doi.org/10.1098/rstb.2016.0014

Garfinkel, S. N., Seth, A. K., Barrett, A. B., Suzuki, K., & Critchley, H. D. (2015, Jan). 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

Georgiou, E., Matthias, E., Kobel, S., Kettner, S., Dreyhaupt, J., Steinacker, J. M., & Pollatos, O. (2015). Interaction of physical activity and interoception in children. *Frontiers in Psychology, 6*, 502. https://doi.org/10.3389/fpsyg.2015.00502

Godefroid, E., Pourtois, G., & Wiersema, J. R. (2016, Feb). Joint effects of sensory feedback and interoceptive awareness on conscious error detection: Evidence from event related brain potentials. *Biological Psychology, 114*, 49-60. https://doi.org/10.1016/j.biopsycho.2015.12.005

Grabauskaitė, A., Baranauskas, M., & Griškova-Bulanova, I. (2017, Feb). Interoception and gender: What aspects should we pay attention to? *Consciousness and Cognition, 48,* 129-137. https://doi.org/10.1016/j.concog.2016.11.002

Harrer, M., Cuijpers, P., Furukawa, T. & Ebert, D. D. (2019). dmetar: Companion R package for the guide 'Doing Meta-Analysis in R'. R package version 0.0.9000. http://dmetar.protectlab.org/

Harrison, O. K., Garfinkel, S. N., Marlow, L., Finnegan, S., Marino, S., Nanz, L., Allen, M., Finnemann, J., Keur-Huizinga, L., Harrison, S. J., Stephan, K. E., Pattinson, K., & Fleming, S. M. (2020). The Filter Detection Task for measurement of breathing-related interoception and metacognition. *bioRxiv*, 2020.2006.2029.176941. https://doi.org/10.1101/2020.06.29.176941

Hart, N., McGowan, J., Minati, L., & Critchley, H. D. (2013, Aug). Emotional regulation and bodily sensation: interoceptive awareness is intact in borderline personality disorder.

Journal of Personality Disorders, 27(4), 506-518. https://doi.org/10.1521/pedi_2012_26_049

Harver, A., Katkin, E. S., & Bloch, E. (1993, May). Signal-detection outcomes on heartbeat and respiratory resistance detection tasks in male and female subjects. *Psychophysiology, 30*(3), 223-230. https://doi.org/10.1111/j.1469-8986.1993.tb03347.x

Harver, A., & Smith, D. M. (1996, Feb). Gender affects the detection of resistive loads. *Perceptual and Motor Skills, 82*(1), 301-302. https://doi.org/10.2466/pms.1996.82.1.301

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press.

Heitkemper, M. M., & Chang, L. (2009). Do fluctuations in ovarian hormones affect gastrointestinal symptoms in women with irritable bowel syndrome? *Gender Medicine, 6*, 152-167. https://doi.org/10.1016/j.genm.2009.03.004

Herbert, B. M., Herbert, C., & Pollatos, O. (2011, Oct). On the relationship between interoceptive awareness and alexithymia: is interoceptive awareness related to emotional awareness? *Journal of Personality, 79*(5), 1149-1175. https://doi.org/10.1111/j.1467-6494.2011.00717.x

Herbert, B. M., Muth, E. R., Pollatos, O., & Herbert, C. (2012). Interoception across modalities: On the relationship between cardiac awareness and the sensitivity for gastric functions. *PloS One, 7*(5), e36646. https://doi.org/10.1371/journal.pone.0036646

Herbert, B. M., & Pollatos, O. (2014, Aug). Attenuated interoceptive sensitivity in overweight and obese individuals. *Eating Behaviours, 15*(3), 445-448. https://doi.org/10.1016/j.eatbeh.2014.06.002

Herbert, B. M., Pollatos, O., Flor, H., Enck, P., & Schandry, R. (2010, Mar 1). Cardiac awareness and autonomic cardiac reactivity during emotional picture viewing and mental stress. *Psychophysiology, 47*(2), 342-354. https://doi.org/10.1111/j.1469-8986.2009.00931.x

Herbert, B. M., Ulbrich, P., & Schandry, R. (2007, Mar). Interoceptive sensitivity and physical effort: Implications for the self-control of physical load in everyday life. *Psychophysiology, 44*(2), 194-202. https://doi.org/10.1111/j.1469-8986.2007.00493.x

Herman, A. M., Rae, C. L., Critchley, H. D., & Duka, T. (2019, Jun). Interoceptive accuracy predicts nonplanning trait impulsivity. *Psychophysiology, 56*(6), e13339. https://doi.org/10.1111/psyp.13339

Hickman, L., Seyedsalehi, A., Cook, J. L., Bird, G., & Murphy, J. (2020, Oct). The relationship between heartbeat counting and heartbeat discrimination: A meta-analysis. *Biological Psychology, 156*, 107949. https://doi.org/10.1016/j.biopsycho.2020.107949

Hina, F., & Aspell, J. E. (2019, Aug). Altered interoceptive processing in smokers: Evidence from the heartbeat tracking task. *International Journal of Psychophysiology, 142*, 10-16. https://doi.org/10.1016/j.ijpsycho.2019.05.012

Hunter, J., & Schmidt, F. (2000). Fixed effects vs. random effects meta-analysis models: Implications for cumulative research knowledge. *International Journal of Selection and Assessment, 8*, 275-292.

Imafuku, M., Fukushima, H., Nakamura, Y., Myowa, M., & Koike, S. (2020, Nov 16). Interoception is associated with the impact of eye contact on spontaneous facial mimicry. *Scientific Reports*, *10*(1), 19866. https://doi.org/10.1038/s41598-020-76393-8

Isomura, T., & Watanabe, K. (2020, Feb). Direct gaze enhances interoceptive accuracy. *Cognition, 195*, 104113. https://doi.org/10.1016/j.cognition.2019.104113

Jakubczyk, A., Skrzeszewski, J., Trucco, E. M., Suszek, H., Zaorska, J., Nowakowska, M., Michalska, A., Wojnar, M., & Kopera, M. (2019, May 1). Interoceptive accuracy and interoceptive sensibility in individuals with alcohol use disorder - Different phenomena with different clinical correlations? *Drug and Alcohol Dependence*, *198*, 34-38. https://doi.org/10.1016/j.drugalcdep.2019.01.036

Jiang, X., Locke, G. R., 3rd, Choung, R. S., Zinsmeister, A. R., Schleck, C. D., & Talley, N. J. (2008, Jun). Prevalence and risk factors for abdominal bloating and visible distention: A population-based study. *Gut*, *57*(6), 756-763. https://doi.org/10.1136/gut.2007.142810

Katkin, E. S. (1985, Mar). Blood, sweat, and tears: individual differences in autonomic self-perception. *Psychophysiology, 22*(2), 125-137. https://doi.org/10.1111/j.1469-8986.1985.tb01573.x

Katkin, E. S., Blascovich, J., & Goldband, S. (1981, Jun). Empirical assessment of visceral self-perception: individual and sex differences in the acquisition of heartbeat discrimination. *Journal of Personality and Social Psychology, 40*(6), 1095-1101. https://doi.org/10.1037//0022-3514.40.6.1095

Katkin, S. D., Reed, C., & Deroo, A. (1983). A methodological analysis of 3 techniques for the assessment of individual-differences in heartbeat detection. *Psychophysiology, 20*(4), 452.

Kennedy, D., Goshko, C. B., Murch, W. S., Limbrick-Oldfield, E. H., Dunn, B. D., & Clark, L. (2019, Jun). Interoception and respiratory sinus arrhythmia in gambling disorder. *Psychophysiology*, *56*(6), e13333. https://doi.org/10.1111/psyp.13333

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., Stephan, K. E., Van den Bergh, O., Van Diest, I., von Leupoldt, A., & Paulus, M. P. (2018, Jun). 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., & Lapidus, R. C. (2016). Can Interoception Improve the Pragmatic Search for Biomarkers in Psychiatry? *Frontiers in Psychiatry, 7*, 121. https://doi.org/10.3389/fpsyt.2016.00121

Khalsa, S. S., Rudrauf, D., Damasio, A. R., Davidson, R. J., Lutz, A., & Tranel, D. (2008, Jul). Interoceptive awareness in experienced meditators. *Psychophysiology, 45*(4), 671-677. https://doi.org/10.1111/j.1469-8986.2008.00666.x

Khalsa, S. S., Rudrauf, D., Feinstein, J. S., & Tranel, D. (2009, Dec). The pathways of interoceptive awareness. *Nature Neuroscience, 12*(12), 1494-1496. https://doi.org/10.1038/nn.2411

Khalsa, S. S., Rudrauf, D., Hassanpour, M. S., Davidson, R. J., & Tranel, D. (2020, Feb). The practice of meditation is not associated with improved interoceptive awareness of the heartbeat. *Psychophysiology, 57*(2), e13479. https://doi.org/10.1111/psyp.13479

Khalsa, S. S., Rudrauf, D., Sandesara, C., Olshansky, B., & Tranel, D. (2009, Apr). Bolus isoproterenol infusions provide a reliable method for assessing interoceptive awareness. *International Journal of Psychophysiology, 72*(1), 34-45. https://doi.org/10.1016/j.ijpsycho.2008.08.010

Khalsa, S. S., Rudrauf, D., & Tranel, D. (2009, Nov). Interoceptive awareness declines with age. *Psychophysiology, 46*(6), 1130-1136. https://doi.org/10.1111/j.1469-8986.2009.00859.x

Kinnaird, E., Stewart, C., & Tchanturia, K. (2020). Interoception in Anorexia Nervosa: Exploring Associations With Alexithymia and Autistic Traits. *Frontiers in Psychiatry, 11*, 64. https://doi.org/10.3389/fpsyt.2020.00064

Klabunde, M., Acheson, D. T., Boutelle, K. N., Matthews, S. C., & Kaye, W. H. (2013, Dec). Interoceptive sensitivity deficits in women recovered from bulimia nervosa. *Eating Behaviours, 14*(4), 488-492. https://doi.org/10.1016/j.eatbeh.2013.08.002

Kleckner, I. R., Wormwood, J. B., Simmons, W. K., Barrett, L. F., & Quigley, K. S. (2015, Nov). Methodological recommendations for a heartbeat detection-based measure of interoceptive sensitivity. *Psychophysiology, 52*(11), 1432-1440. https://doi.org/10.1111/psyp.12503

Koch, A., & Pollatos, O. (2014, Sep). Cardiac sensitivity in children: sex differences and its relationship to parameters of emotional processing. *Psychophysiology, 51*(9), 932-941. https://doi.org/10.1111/psyp.12233

Koeppel, C. J., Ruser, P., Kitzler, H., Hummel, T., & Croy, I. (2020). Interoceptive accuracy and its impact on neuronal responses to olfactory stimulation in the insular cortex. *Human Brain Mapping, 41*(11), 2898-2908. https://doi.org/10.1002/hbm.24985

Köteles, F., Éliás, I., Szabolcs, Z., Körmendi, J., Ferentzi, E., & Szemerszky, R. (2020, Feb). Accuracy of reproduction of physical training load is not associated with resting heartbeat perception in healthy individuals. *Biological Psychology, 150*, 107831. https://doi.org/10.1016/j.biopsycho.2019.107831

Krautwurst, S., Gerlach, A. L., Gomille, L., Hiller, W., & Witthöft, M. (2014, Jun). Health anxiety--an indicator of higher interoceptive sensitivity? *Journal of Behavior Therapy and Experimental Psychiatry, 45*(2), 303-309. https://doi.org/10.1016/j.jbtep.2014.02.001

Krautwurst, S., Gerlach, A. L., & Witthoft, M. (2016, Nov). Interoception in pathological health anxiety. *Journal of Abnormal Psychology, 125*(8), 1179-1184. https://doi.org/10.1037/abn0000210

Kuehner, C. (2017, Feb). Why is depression more common among women than among men? *Lancet Psychiatry, 4*(2), 146-158. https://doi.org/10.1016/s2215-0366(16)30263-2

Lischke, A., Weippert, M., Mau-Moeller, A., Jacksteit, R., & Pahnke, R. (2020, Apr). Interoceptive accuracy is associated with emotional contagion in a valence- and sex-dependent manner. *Social Neuroscience, 15*(2), 227-233. https://doi.org/10.1080/17470919.2019.1690573

Lovell, R. M., & Ford, A. C. (2012, Jul). Effect of gender on prevalence of irritable bowel syndrome in the community: Systematic review and meta-analysis. *American Journal of Gastroenterology, 107*(7), 991-1000. https://doi.org/10.1038/ajg.2012.131

Lüdecke, D. (2019). Esc: Effect Size Computation for Meta Analysis (Version 0.5.1). https://CRAN.R-project.org/package=esc. https://doi.org/10.5281/zenodo.1249218

Lugo, R. G., Helkala, K., Knox, B. J., Josok, O., Lande, N. M., & Sutterlin, S. (2018). Interoceptive sensitivity as a proxy for emotional intensity and its relationship with perseverative cognition. *Psychology Research and Behavior Management, 11*, 1-8. https://doi.org/10.2147/PRBM.S139790

Lukowska, M., Sznajder, M., & Wierzchon, M. (2018, Jan 18). Error-related cardiac response as information for visibility judgements. *Scientific Reports, 8*(1), 1131. https://doi.org/10.1038/s41598-018-19144-0

Marshall, A. C., Gentsch, A., Blum, A. L., Broering, C., & Schütz-Bosbach, S. (2019, May 1). I feel what I do: Relating interoceptive processes and reward-related behavior. *Neuroimage, 191*, 315-324. https://doi.org/10.1016/j.neuroimage.2019.02.032

Marshall, A. C., Gentsch, A., Schroder, L., & Schutz-Bosbach, S. (2018, Sep 4). Cardiac interoceptive learning is modulated by emotional valence perceived from facial expressions. *Social Cognitive and Affective Neuroscience, 13*(7), 677-686. https://doi.org/10.1093/scan/nsy042

Marshall, C. R., Hardy, C. J. D., Russell, L. L., Clark, C. N., Dick, K. M., Brotherhood, E. V., Bond, R. L., Mummery, C. J., Schott, J. M., Rohrer, J. D., Kilner, J. M., & Warren, J. D. (2017). Impaired interoceptive accuracy in semantic variant Primary Progressive Aphasia. *Frontiers in Neurology, 8*, 610. https://doi.org/10.3389/fneur.2017.00610

McFarland, R. A. (1975, Jul). Heart rate perception and heart rate control. *Psychophysiology, 12*(4), 402-405. https://doi.org/10.1111/j.1469-8986.1975.tb00011.x

Michael, G. A., Naveteur, J., Dupuy, M. A., & Jacquot, L. (2015, May 1). My heart is in my hands: The interoceptive nature of the spontaneous sensations felt on the hands. *Physiology & Behaviour, 143*, 113-120. https://doi.org/10.1016/j.physbeh.2015.02.030

Moeini-Jazani, M., Knoeferle, K., de Moliere, L., Gatti, E., & Warlop, L. (2017). Social power increases interoceptive accuracy. *Frontiers in Psychology, 8*, 1322. https://doi.org/10.3389/fpsyg.2017.01322

Motyka, P., Grund, M., Forschack, N., Al, E., Villringer, A., & Gaebler, M. (2019, Oct). Interactions between cardiac activity and conscious somatosensory perception. *Psychophysiology*, *56*(10), e13424. https://doi.org/10.1111/psyp.13424

Mulak, A., Taché, Y., & Larauche, M. (2014, Mar 14). Sex hormones in the modulation of irritable bowel syndrome. *World Journal of Gastroenterology, 20*(10), 2433-2448. https://doi.org/10.3748/wjg.v20.i10.2433

Murphy, J., Brewer, R., Catmur, C., & Bird, G. (2017, Feb). Interoception and psychopathology: A developmental neuroscience perspective. *Developmental Cognitive Neuroscience, 23*, 45-56. https://doi.org/10.1016/j.dcn.2016.12.006

Murphy, J., Brewer, R., Coll, M. P., Plans, D., Hall, M., Shiu, S. S., Catmur, C., & Bird, G. (2019, Nov). I feel it in my finger: Measurement device affects cardiac interoceptive accuracy. *Biological Psychology, 148*, 107765. https://doi.org/10.1016/j.biopsycho.2019.107765

Murphy, J., Brewer, R., Hobson, H., Catmur, C., & Bird, G. (2018, Jul). 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., Brewer, R., Plans, D., Khalsa, S. S., Catmur, C., & Bird, G. (2020, Jan). Testing the independence of self-reported interoceptive accuracy and attention. *Quarterly Journal of Experimental Psychology, 73*(1), 115-133. https://doi.org/10.1177/1747021819879826

Murphy, J., Millgate, E., Geary, H., Ichijo, E., Coll, M. P., Brewer, R., Catmur, C., & Bird, G. (2018, Mar). Knowledge of resting heart rate mediates the relationship between intelligence and the heartbeat counting task. *Biological Psychology, 133*, 1-3. https://doi.org/10.1016/j.biopsycho.2018.01.012

Murphy, J., Viding, E., & Bird, G. (2019, Oct). Does atypical interoception following physical change contribute to sex differences in mental illness? *Psychological Review, 126*(5), 787-789. https://doi.org/10.1037/rev0000158

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., & Moher, D. (2021, Mar 29). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71

Palser, E. R., Fotopoulou, A., Pellicano, E., & Kilner, J. M. (2018, Jul). The link between interoceptive processing and anxiety in children diagnosed with autism spectrum disorder: Extending adult findings into a developmental sample. *Biological Psychology, 136*, 13-21. https://doi.org/10.1016/j.biopsycho.2018.05.003

Palser, E. R., Fotopoulou, A., Pellicano, E., & Kilner, J. M. (2020). Dissociation in how core autism features relate to interoceptive dimensions: Evidence from cardiac awareness in children. *Journal of Autism and Developmental Disorders*, *50*(2), 572-582. https://doi.org/10.1007/s10803-019-04279-4

Pauli, P., Hartl, L., Marquardt, C., Stalmann, H., & Strian, F. (1991, May). Heartbeat and arrhythmia perception in diabetic autonomic neuropathy. *Psychological Medicine, 21*(2), 413-421. https://doi.org/10.1017/s0033291700020523

Pennebaker, J. W. (1982). *The psychology of physical symptoms*. Springer-Verlag.

Plans, D., Ponzo, S., Morelli, D., Cairo, M., Ring, C., Keating, C., Cunningham, A., Catmur, C., Murphy, J., & Bird, G. (2021). Measuring interoception: The phase adjustment task. *Biological Psychology*.

Pollatos, O., Herbert, B. M., Mai, S., & Kammer, T. (2016, Nov 19). Changes in interoceptive processes following brain stimulation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 371*(1708). https://doi.org/10.1098/rstb.2016.0016

Pollatos, O., Kirsch, W., & Schandry, R. (2005, Dec). On the relationship between interoceptive awareness, emotional experience, and brain processes. *Brain Research: Cognitive Brain Research, 25*(3), 948-962. https://doi.org/10.1016/j.cogbrainres.2005.09.019

Pollatos, O., Kurz, A. L., Albrecht, J., Schreder, T., Kleemann, A. M., Schöpf, V., Kopietz, R., Wiesmann, M., & Schandry, R. (2008, Dec). Reduced perception of bodily signals in anorexia nervosa. *Eating Behaviour, 9*(4), 381-388. https://doi.org/10.1016/j.eatbeh.2008.02.001

Pollatos, O., Matthias, E., & Schandry, R. (2007, Oct). Heartbeat perception and P300 amplitude in a visual oddball paradigm. *Clinical Neurophysiology, 118*(10), 2248-2253. https://doi.org/10.1016/j.clinph.2007.06.057

Pollatos, O., Traut-Mattausch, E., Schroeder, H., & Schandry, R. (2007). Interoceptive awareness mediates the relationship between anxiety and the intensity of unpleasant feelings. *Journal of Anxiety Disorders*, *21*(7), 931-943. https://doi.org/10.1016/j.janxdis.2006.12.004

Rae, C. L., Ahmad, A., Larsson, D. E. O., Silva, M., Praag, C., Garfinkel, S. N., & Critchley, H. D. (2020, Mar 6). Impact of cardiac interoception cues and confidence on voluntary decisions to make or withhold action in an intentional inhibition task. *Scientific Reports, 10*(1), 4184. https://doi.org/10.1038/s41598-020-60405-8

Rae, C. L., Botan, V. E., Gould van Praag, C. D., Herman, A. M., Nyyssonen, J. A. K., Watson, D. R., Duka, T., Garfinkel, S. N., & Critchley, H. D. (2018, Jun 14). Response inhibition on the stop signal task improves during cardiac contraction. *Scientific Reports, 8*(1), 9136. https://doi.org/10.1038/s41598-018-27513-y

Rae, C. L., Larsson, D. E. O., Garfinkel, S. N., & Critchley, H. D. (2019, Jan). Dimensions of interoception predict premonitory urges and tic severity in Tourette syndrome. *Psychiatry Research, 271,* 469-475. https://doi.org/10.1016/j.psychres.2018.12.036

Ricciardi, L., Demartini, B., Crucianelli, L., Krahé, C., Edwards, M. J., & Fotopoulou, A. (2016, Jan). Interoceptive awareness in patients with functional neurological symptoms. *Biological Psychology, 113,* 68-74. https://doi.org/10.1016/j.biopsycho.2015.10.009

Riecher-Rössler, A. (2017, Jan). Sex and gender differences in mental disorders. *Lancet Psychiatry, 4*(1), 8-9. https://doi.org/10.1016/s2215-0366(16)30348-0

Ring, C., & Brener, J. (1996, Sep). 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. (2018, Sep). Heartbeat counting is unrelated to heartbeat detection: A comparison of methods to quantify interoception. *Psychophysiology, 55*(9), e13084. https://doi.org/10.1111/psyp.13084

Ring, C., Brener, J., Knapp, K., & Mailloux, J. (2015, Jan). 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

Salomon, R., Ronchi, R., Donz, J., Bello-Ruiz, J., Herbelin, B., Martet, R., Faivre, N., Schaller, K., & Blanke, O. (2016, May 4). The insula mediates access to awareness of visual stimuli presented synchronously to the heartbeat. *Journal of Neuroscience, 36*(18), 5115-5127. https://doi.org/10.1523/JNEUROSCI.4262-15.2016

Schaan, L., Schulz, A., Nuraydin, S., Bergert, C., Hilger, A., Rach, H., & Hechler, T. (2019). Interoceptive accuracy, emotion recognition, and emotion regulation in preschool children. *International Journal of Psychophysiology, 138,* 47-56. https://doi.org/10.1016/j.ijpsycho.2019.02.001

Schandry, R. (1981, Jul). Heart beat perception and emotional experience. *Psychophysiology, 18*(4), 483-488. https://doi.org/10.1111/j.1469-8986.1981.tb02486.x

Schauder, K. B., Mash, L. E., Bryant, L. K., & Cascio, C. J. (2015, Mar). Interoceptive ability and body awareness in autism spectrum disorder. *Journal of Experimental Child Psychology, 131,* 193-200. https://doi.org/10.1016/j.jecp.2014.11.002

Schroeder, S., Gerlach, A. L., Achenbach, S., & Martin, A. (2015, Apr). The relevance of accuracy of heartbeat perception in noncardiac and cardiac chest pain. *International Journal of Behavioral Medicine, 22*(2), 258-267. https://doi.org/10.1007/s12529-014-9433-3

Schultchen, D., Zaudig, M., Krauseneck, T., Berberich, G., & Pollatos, O. (2019). Interoceptive deficits in patients with obsessive-compulsive disorder in the time course of cognitive-behavioral therapy. *PloS One, 14*(5), e0217237. https://doi.org/10.1371/journal.pone.0217237

Schulz, A., Lass-Hennemann, J., Sütterlin, S., Schächinger, H., & Vögele, C. (2013, Apr). Cold pressor stress induces opposite effects on cardioceptive accuracy dependent on assessment paradigm. *Biological Psychology, 93*(1), 167-174. https://doi.org/10.1016/j.biopsycho.2013.01.007

Schwerdtfeger, A. R., Heene, S., & Messner, E. M. (2019, Jul). Interoceptive awareness and perceived control moderate the relationship between cognitive reappraisal, self-esteem, and cardiac activity in daily life. *International Journal of Psychophysiology, 141*, 84-92. https://doi.org/10.1016/j.ijpsycho.2019.04.003

Shephard, R. J., & Miller, H. S. (1998). *Exercise and the health in health and disease (2nd ed.)*. CRC Press.

Sherrington, C. S. (1906). *The integrative action of the nervous system*. Yale University Press.

Sidik, K., & Jonkman, J. N. (2007, Apr 30). A comparison of heterogeneity variance estimators in combining results of studies. *Statistics in Medicine, 26*(9), 1964-1981. https://doi.org/10.1002/sim.2688

Sönmez, M. B., Kahyacı Kılıç, E., Ateş Çöl, I., Görgülü, Y., & Köse Çınar, R. (2017, 2017/01/02). Decreased interoceptive awareness in patients with substance use disorders. *Journal of Substance Use, 22*(1), 60-65. https://doi.org/10.3109/14659891.2016.1143048

Steptoe, A., & Vögele, C. (1992, Nov). Individual differences in the perception of bodily sensations: the role of trait anxiety and coping style. *Behavioural Research and Therapy, 30*(6), 597-607. https://doi.org/10.1016/0005-7967(92)90005-2

Stern, Y., Koren, D., Moebus, R., Panishev, G., & Salomon, R. (2020, Sep 11). Assessing the relationship between sense of agency, the bodily-self and stress: Four virtual-reality experiments in healthy individuals. *Journal of Clinical Medicine, 9*(9). https://doi.org/10.3390/jcm9092931

Sueyoshi, T., Sugimoto, F., Katayama, J., & Fukushima, H. (2014, Dec). Neural correlates of error processing reflect individual differences in interoceptive sensitivity. *International Journal of Psychophysiology, 94*(3), 278-286. https://doi.org/10.1016/j.ijpsycho.2014.10.001

Suschinsky, K. D., & Lalumière, M. L. (2012, Feb). Is sexual concordance related to awareness of physiological states? *Archives of Sexual Behavior, 41*(1), 199-208. https://doi.org/10.1007/s10508-012-9931-9

Sutterlin, S., Schulz, S. M., Stumpf, T., Pauli, P., & Vogele, C. (2013, Jul). Enhanced cardiac perception is associated with increased susceptibility to framing effects. *Cognitive Science, 37*(5), 922-935. https://doi.org/10.1111/cogs.12036

Suzuki, K., Garfinkel, S. N., Critchley, H. D., & Seth, A. K. (2013, Nov). Multisensory integration across exteroceptive and interoceptive domains modulates self-experience in the rubber-hand illusion. *Neuropsychologia, 51*(13), 2909-2917. https://doi.org/10.1016/j.neuropsychologia.2013.08.014

Terhaar, J., Viola, F. C., Bär, K. J., & Debener, S. (2012, Oct). 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

Thurlbeck, W. M. (1982, Aug). Postnatal human lung growth. *Thorax, 37*(8), 564-571. https://doi.org/10.1136/thx.37.8.564

Todd, J., Aspell, J. E., Barron, D., Toh, E. K. L., Zahari, H. S., Mohd Khatib, N. A., Laughton, R., & Swami, V. (2020, Sep). Greater gastric interoception is associated with more positive body image: Evidence from adults in Malaysia and the United Kingdom. *Body Image, 34*, 101-111. https://doi.org/10.1016/j.bodyim.2020.05.011

Todd, J., Hina, F., & Aspell, J. E. (2020, Mar). Reduction of auditory input improves performance on the heartbeat tracking task, but does not necessarily enhance interoception. *Experimental Brain Research, 238*(3), 621-629. https://doi.org/10.1007/s00221-020-05740-y

Ueno, D., Matsuoka, T., Kato, Y., Ayani, N., Maeda, S., Takeda, M., & Narumoto, J. (2020). Individual differences in interoceptive accuracy are correlated with salience network connectivity in older adults. *Frontiers in Aging Neuroscience, 12*, 592002. https://doi.org/10.3389/fnagi.2020.592002

Van Den Houte, M., Vlemincx, E., Franssen, M., Van Diest, I., Van Oudenhove, L., & Luminet, O. (2021, Apr). The respiratory occlusion discrimination task: A new paradigm to measure respiratory interoceptive accuracy. *Psychophysiology, 58*(4), e13760. https://doi.org/10.1111/psyp.13760 van Dyck, Z., Vögele, C., Blechert, J., Lutz, A. P., Schulz, A., & Herbert, B. M. (2016). The Water Load Test as a measure of gastric interoception: Development of a two-stage protocol and application to a healthy female population. *PloS One, 11*(9), e0163574. https://doi.org/10.1371/journal.pone.0163574

van 't Wout, M., Faught, S., & Menino, D. (2013). Does interoceptive awareness affect the ability to regulate unfair treatment by others? *Frontiers in Psychology, 4*, 880. https://doi.org/10.3389/fpsyg.2013.00880

Viechtbauer, W. (2010, 08/05). Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software, 36*(3), 1-48. https://doi.org/10.18637/jss.v036.i03

Villani, V., Tsakiris, M., & Azevedo, R. T. (2019, Nov). Transcutaneous vagus nerve stimulation improves interoceptive accuracy. *Neuropsychologia*, *134*, 107201. https://doi.org/10.1016/j.neuropsychologia.2019.107201

Wegner, A., Elsenbruch, S., Rebernik, L., Roderigo, T., Engelbrecht, E., Jager, M., Engler, H., Schedlowski, M., & Benson, S. (2015, Oct). Inflammation-induced pain sensitization in men and women: does sex matter in experimental endotoxemia? *Pain*, *156*(10), 1954-1964. https://doi.org/10.1097/j.pain.0000000000000256

Weineck, F., Schultchen, D., Hauke, G., Messner, M., & Pollatos, O. (2020). Using bodily postures to reduce anxiety and improve interoception: A comparison between powerful and neutral poses. *PloS One*, *15*(12), e0242578. https://doi.org/10.1371/journal.pone.0242578

Whitehead, W. E., & Drescher, V. M. (1980, Nov). 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

Whitehead, W. E., Drescher, V. M., Heiman, P., & Blackwell, B. (1977, Dec). Realtion of heart rate control to heartbeat perception. *Biofeedback and Self Regulation, 2*(4), 317-392.

Wiersema, J. R., & Godefroid, E. (2018). Interoceptive awareness in attention deficit hyperactivity disorder. *PloS One, 13*(10), e0205221. https://doi.org/10.1371/journal.pone.0205221

Wolk, J., Sutterlin, S., Koch, S., Vogele, C., & Schulz, S. M. (2014, Mar). Enhanced cardiac perception predicts impaired performance in the Iowa Gambling Task in patients with panic disorder. *Brain and Behavior, 4*(2), 238-246. https://doi.org/10.1002/brb3.206

Yates, A. J., Jones, K. E., Marie, G. V., & Hogben, J. H. (1985, Sep). Detection of the heartbeat and events in the cardiac cycle. *Psychophysiology, 22*(5), 561-567. https://doi.org/10.1111/j.1469-8986.1985.tb01651.x

Zamariola, G., Luminet, O., Mierop, A., & Corneille, O. (2019, Dec). Does it help to feel your body? Evidence is inconclusive that interoceptive accuracy and sensibility help cope with negative experiences. *Cognition & Emotion, 33*(8), 1627-1638. https://doi.org/10.1080/02699931.2019.1591345

Zoellner, L. A., & Craske, M. G. (1999, Dec). Interoceptive accuracy and panic. *Behavioural Ressearch and Therapy, 37*(12), 1141-1158. https://doi.org/10.1016/s0005-7967(98)00202-2