The following publication Yeung, M. K., & Lin, J. (2021). Probing depression, schizophrenia, and other psychiatric disorders using fNIRS and the verbal fluency test: A systematic review and meta-analysis. Journal of Psychiatric Research, 140, 416-435 is available at https://doi.org/10.1016/j.jpsychires.2021.06.015.

Probing depression, schizophrenia, and other psychiatric disorders using fNIRS and the verbal fluency test: A systematic review and meta-analysis

Highlights

- We reviewed 121 studies using fNIRS and VFT to probe psychiatric disorders
- Many psychiatric disorders show reduced HbO increases in frontotemporal regions
- HbO is more sensitive than HbR and task performance to detect psychopathologies
- Major depression and schizophrenia have partially distinct hypoactivation patterns
- Some methodological issues may confound the hypoactivation observed in patients

1. Introduction

Accumulating evidence suggests that structural and functional abnormalities in the prefrontal cortex (PFC) and its connected regions characterize a wide variety of psychiatric disorders (Whitfield-Gabrieli & Ford, 2012; Xia et al., 2018). Thus, accessible neuroimaging tools that can identify specific frontal lobe dysfunction associated with different psychiatric disorders could be useful for improving disease diagnosis and prognosis and treatment development (Klöppel et al., 2012; Singh & Rose, 2009). Functional near-infrared spectroscopy (fNIRS) has become increasingly popular in the field of psychiatry, due to its low costs and the ease of application in ecologically valid environments (Ehlis et al., 2014). This technique leverages the principle of neurovascular coupling and the optical properties of hemoglobin to estimate changes in cerebral hemoglobin concentrations in response to changes in neuronal activity (Pinti et al., 2020; Quaresima & Ferrari, 2019). fNIRS is most commonly paired with the verbal fluency test (VFT) because this test is easy to administer and requires little time, equipment, or space. During the VFT, individuals are asked to generate as many words from a category (usually phonemic or semantic) as possible within a given time limit (usually 60 seconds; Crowe, 1988). Because the VFT taps the strategic access to lexical-semantic information, especially during the later period of the task, VFT performance activates and crucially relies on the superior medial frontal cortex, ventrolateral prefrontal cortex (vlPFC), and anterior temporal lobe (e.g., superior temporal cortex; STC), especially in the left hemisphere (Henry & Crawford, 2004; Robinson et al., 2012; Wagner et al., 2014).

The combined use of fNIRS with the VFT has been applied to psychiatric research for many years (Ehlis et al., 2014; Yeung & Chan, 2020). One common finding has been that many

psychiatric disorders, including major depressive disorder (MDD) and schizophrenia (SCZ), are associated with hypoactivation in the frontal or temporal subregions during VFT performance (Ohi et al., 2017; Takizawa et al., 2014; Wei et al., 2020). However, the topographical distribution of hypoactivation, even within a single class of disorder, has often been inconsistent and unclear. For example, the extent and foci of hypoactivation in MDD patients have varied across fNIRS studies examining similar sample sizes (Akiyama et al., 2018; Pu et al., 2012; Tsujii et al., 2017). In addition, although the ventral frontotemporal region, particularly in the left hemisphere, is consistently activated during VFT performance in healthy individuals (Wagner et al., 2014), whether this region is the locus of hypoactivation in psychiatric patients, characterized by more pronounced hypoactivation in the left hemisphere and inferior prefrontal regions compared with the right hemisphere and superior prefrontal regions, remains unclear.

Relatedly, although the information provided by fNIRS during VFT performance appears to be associated with the presence of psychopathology, the specificity and uniqueness of the observed patterns remains unclear. Impaired verbal fluency is common among psychiatric patients, and is often used to inform diagnosis (Bokat & Goldberg, 2003; Henry & Crawford, 2005; Monsch et al., 1992; Raucher-Chéné et al., 2017). However, whether fNIRS measurements can provide additional value for the detection of psychiatric disorders remains unknown. In addition, psychiatric disorders are highly comorbid, with large overlaps in the behavioral and neurobiological features associated with varying types of psychopathology (Buckley et al., 2009; Matson & Goldin, 2013). Whether each psychiatric disorder is associated with a signature pattern of brain activation that is measurable by fNIRS during VFT performance has yet to be determined. Advancing knowledge in these areas would shed light on the utility of using fNIRS

for differential diagnosis, which is timely and clinically important because each form of psychopathology requires specialized treatments and management strategies.

When near-infrared lights are delivered to the scalp, they penetrate superficial layers (skin, skull, cerebrospinal fluid) before reaching the brain and as they return to the scalp. For continuous-wave fNIRS, which is the most common fNIRS method, the estimated changes in the concentrations of oxyhemoglobin (HbO) and deoxyhemoglobin (HbR) depend on the differential pathlength factor (DPF), which is influenced by the brain-scalp distance or scattering-to-absorption ratio (Beauchamp et al., 2011; Whiteman, Santosa, Chen, Perlman, & Huppert, 2017). Because altered cortical structures (e.g., gray matter abnormalities) are common features in many psychiatric disorders (Bora et al., 2012; Duerden et al., 2012; Yüksel et al., 2012), group differences in HbO and HbR changes may be confounded by group differences in brain structures. This problem can be mitigated by using unitless metrics (e.g., effect sizes) that are DPF-independent (Schroeter et al., 2003). Although fNIRS has a relatively high motion tolerance, the jaw movements and head motions that occur during overt word production might contaminate fNIRS signals, reducing the statistical power of analyses or introducing biases. Task-induced systemic artifacts may also obscure the true estimation of brain activity (Scholkmann et al., 2014). These problems can be alleviated by applying motion correction (Brigadoi et al., 2014). Multichannel fNIRS can be used to characterize the topographical distribution of brain activity, which cannot be achieved when using single-channel fNIRS; however, the risks of Type I errors increase when the number of comparisons increases, although this can be mitigated by applying multiple comparison correction.

To clarify the clinical utility of the fNIRS-VFT paradigm and enhance the practical applications of this paradigm during psychiatric research, we conducted a systematic review of fNIRS studies performed while using the VFT to probe psychiatric disorders, with a focus on two objectives: (1) to determine the specificity and uniqueness of frontal activation patterns associated with different psychiatric disorders during the VFT; and (2) to evaluate the fNIRS methods that have been applied during past VFT studies. We also took advantage of the relatively homogeneous task designs and fNIRS acquisition methods used across studies, undertaking two complementary meta-analytic approaches to synthesize the findings of these studies. To perform the meta-analyses, we first mapped the individual studies' findings (e.g., the location of single channels or clusters showing significant reductions in HbO increases observed in patients) to a sensor template, which comprised eight frontotemporal regions. For the first approach, vote counting based on the direction of effect was performed (McKenzie & Brennan, 2020), and the proportion of channels that exhibited significant reductions in HbO changes between patients and controls in each region was analyzed as the dependent variable. For the second approach, meta-analyses based on the effect sizes of group differences in HbO changes were conducted. These two approaches were complementary—the first approach made few assumptions about the outcome variables but did not consider the sample size of each study, whereas the second approach considered the sample size but was based on some assumptions about the outcome variables (due to incomplete reporting of results in many studies).

1. Methods

1.1. Search Strategy and Study Selection

This study was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines (Moher et al., 2009). The review protocol was not registered in any registry. The first author performed a literature search using PubMed and PsycINFO on October 27, 2020, to identify fNIRS studies that used the VFT to examine psychiatric disorders. The keywords used were "((verbal fluency) OR (semantic fluency) OR (category fluency) OR (phonemic fluency) OR (letter fluency)) AND (fNIRS OR NIRS OR (near-infrared spectroscopy))." No limits were set for the search.

We screened the titles and abstracts of the articles identified by the search engines. A study was included if it (1) applied fNIRS, (2) included psychiatric patients diagnosed based on the DSM, ICD, or other established diagnostic criteria, (3) used the VFT as the activation task, (4) enrolled healthy controls, and (5) reported original data. In addition, a study was excluded if it (1) was a review paper, protocol, or case study, (2) was not written in English, or (3) used the VFT as a secondary task (e.g., dual-task walking). We did not set a limit on the version of DSM and ICD to include as many studies as possible for each class of mental disorder and to avoid selection biases. In addition, we did not list handedness as an inclusion criterion because we had no hypothesis that handedness would influence the difference in cortical activation between patients and controls. The full texts of all screened articles were then retrieved for eligibility assessment, based on the same set of criteria. The two authors independently screened the search result, and discrepancies were resolved by consensus.

2.2. Data Extraction and Coding

A spreadsheet was used to document all data extracted from the studies. In addition to the first authors' names and publication dates of all studies, the following information regarding the

research design, experimental paradigm, fNIRS measurement, data preprocessing, and data analysis was extracted from each identified article: (a) sample characteristics (sample size, matching variables, and neurological, psychological, and behavioral features); (b) task design (VFT type, test stimuli, and block duration); (c) signal acquisition (fNIRS instrument, sampling rate, source-detector distance, and the number and location of optodes/channels); (d) data preprocessing (data/channel rejection, frequency filtering, moving average, motion/systemic correction, detrending, other artifact removal methods, and the conversion to unitless metrics); (e) data analysis (fNIRS variables, HbO and HbR results, and multiple comparison correction methods); and (f) factors affecting the level of HbO increases (the onset and duration of illness, drug dose, symptom level, and other factors). No study authors were contacted for additional information.

For HbO and HbR results, '+' and '−' were used to indicate significantly larger (e.g., greater activation or longer latency) and smaller (e.g., weaker activation or shorter latency) values in patients relative to healthy controls, respectively, whereas 'o' was used to indicate no significant effect. We primarily considered results that were corrected for multiple comparisons.

2.3. Quantitative Synthesis of the Level of HbO Changes

We took advantage of the substantial homogeneity in the task designs and fNIRS measurement methods used across studies, including the complete reporting of significant channel locations in most studies. Two meta-analytic approaches, each with its pros and cons, were adopted. First, meta-analyses were conducted based on the proportion of single channels or channel clusters that exhibited significantly reduced VFT-induced HbO increases (i.e., hypoactivation) in patients compared to healthy controls. This approach is illustrated in Figure 1. Specifically, the diagnosis

effect (i.e., patients > controls) observed in each channel or cluster was converted to a 5 × 21 matrix space, which represented the bilateral frontotemporal region. Five anchor points, corresponding to Fpz, Fp1, Fp2, T3, and T4, and any two adjacent cells corresponded to a 1.5-cm separation. This layout was based on the 52-channel ETG-4000 system (Hitachi Medical Co., Tokyo, Japan), in which a 3 × 11 probe with a 3-cm source–detector separation is symmetrically placed on the forehead, and the center of the bottom probe is placed at Fpz. Based on the literature (Akiyama et al., 2018; Takizawa et al., 2014), two channels are located near Fp1 and Fp2, and the two bottom, outermost optodes are located near T3 and T4. Each channel or cluster was coded as either '+' or '−' if the patient showed significantly larger or smaller HbO increases, respectively, compared with those in the controls. The activity was coded as "o" if patients exhibited similar HbO increases as controls.

Based on previous studies (Tsujii et al., 2017; Tsuzuki et al., 2007), the frontotemporal region was segmented into eight regions: (1) right dorsolateral PFC (dlPFC); (2) dorsal frontopolar cortex (dFPC); (3) left dlPFC; (4) right STC; (5) right vlPFC; (6) ventral FPC (vFPC); (7) left vlPFC; and (8) left STC. For each region, the proportion of channels that exhibited hypoactivation in patients compared with controls was calculated by dividing the number of channels or clusters showing a significant diagnosis effect by the total number of channels or clusters being tested in that region. A positive or negative value was obtained if patients showed significantly smaller or larger overall HbO increases than controls, respectively. No region included a mixture of channels that showed both hyperactivation and hypoactivation. A zero value was obtained when patients and controls did not differ significantly in any of the channels or clusters. Because the vlPFC and STC comprised the brain regions directly inferior to the

dlPFC (Figure 1), these two regions were combined when addressing the inferior vs. superior dimension of hypoactivation in the lateral frontotemporal regions in patients.

Next, the proportions of channels that exhibited hypoactivation in patients, both within and between regions, were evaluated using one-sample and paired-sample sign tests, respectively. The variables that were derived from testing clusters of channels that covered more than one region were excluded due to uncertainty regarding the locus of the effects. Studies that used either the phonemic or semantic VFT were aggregated because both VFT versions induce similar PFC activation (Wagner et al., 2014). For studies that included two groups with the same diagnosis or two versions of the VFT, the group with the poorer clinical symptoms or the group with fewer comorbid symptoms and phonemic VFT were chosen for analysis. Additionally, for studies that made a distinction between early and late task periods, the late task period was selected. Conceptually, this analytic approach is similar to vote counting based on the direction of effect (McKenzie & Brennan, 2020): the proportions of channels that exhibit significant hypoactivation in patients in one region can differ from zero or that in another region in either direction. A single-channel-based meta-analysis was not performed because the head size varies across individuals, and the test-retest reliability of fNIRS signals has been shown to be acceptable at the cluster level but unsatisfactory at the single-channel level (Plichta et al., 2006; Schecklmann, Ehlis, Plichta, & Fallgatter, 2008).

We also performed effect-size-based, random-effects meta-analyses to enhance the robustness of the results. For each study, we first calculated the Cohen's *d* of the group difference in HbO changes (controls > patients) for each individual channel or channel cluster (Thalheimer & Cook, 2002). A conservative approach was adopted to deal with missing values—a nonsignificant

channel was assigned a Cohen's *d* of 0 (i.e., assuming $p = 1.00$), whereas a significant channel was assigned a Cohen's *d* that corresponded to the lower limit of the reported *p*-value range (e.g., assuming $p = .05$) (Fox et al., 2016). Next, Cohen's *d* was averaged across channels for each region, and the mean Cohen's *d* was converted to Hedges' *g* to correct for small sample size biases (Hedges & Olkin, 1985).

Hedges' *g* was pooled across studies for each region to compare the difference in regional activation between controls and patients. Also, the differences in Hedges' *g* between two regions were also pooled across studies to compare the inferior vs. superior and the left vs. right dimensions of activation between groups, assuming zero correlation for the pooled standard error (i.e., larger variance and thus more conservative) (Borenstein et al., 2009). The meta-analyses were done using Meta-analysis with R (Schwarzer et al., 2015). For each meta-analysis, we first identified and excluded outliers using the "find.outliers" function included in the R package (Viechtbauer & Cheung, 2010), and then estimated the pooled effects using the Hartung-Knapp-Sidik-Jonkman (HKSJ) method (IntHout et al., 2014).

2.4. Risk of Bias Assessment

The risk of bias was assessed throughout, similar to previously performed systematic reviews (Yeung, 2021; Yeung & Chan, 2021). Because differences in fNIRS variables between patients and healthy controls could be confounded by demographic or intellectual differences, we surveyed whether the two groups in each study were comparable in these aspects. Multiple comparison correction is typically necessary for most multichannel fNIRS studies because *a priori* hypotheses are often not made at the channel or cluster level. Therefore, we examined whether such corrections were applied in these studies. Brain structure or the brain-scalp

distance, which may differ between patients and controls due to pathology (Bora et al., 2012; Duerden et al., 2012; Yüksel et al., 2012), could confound group difference in fNIRS variables. Therefore, we examined whether previous studies used unitless metrics that were DPF-independent (Schroeter et al., 2003). Finally, although fNIRS studies typically analyze HbO only because this index has often, but not always, been shown to have a higher signal-to-noise ratio and to correlate better with blood-oxygen-level-dependent (BOLD) signals than HbR (Cui et al., 2011), some researchers also analyzed HbR because only this index yielded significant results. Thus, we also examined whether the studies compared groups for both HbO and HbR changes.

3. Results

3.1. Flow of the Literature Search and Publication Trends

Figure 2 illustrates the flow of the literature search. Initially, 242 and 182 articles were identified via PubMed and PsycINFO, respectively. After duplication removal, 271 articles were selected for screening and eligibility assessment. Finally, 121 articles were included in the qualitative synthesis, and 65 articles were included in the quantitative synthesis. Meta-analyses were only performed for MDD and SCZ because these were the only two groups of disorders for which information regarding the proportions of channels associated with significant differences in HbO increases between patients and controls were available for each of the eight regions in > 10 studies. A brief summary of the included studies is shown in Table 1. No studies reported using an exact duplicate sample from any earlier studies. As shown in Figure 3, the annual number of articles published in scholarly journals has been increasing since the early 2000s.

3.2. Summary of Study Characteristics

3.2.1. Sample Characteristics

Supplementary Table 1 presents the sample characteristics and the version of DSM and ICD used in each included study. Most studies that applied the combined use of fNIRS and VFT were performed to study SCZ (n = 42), MDD (n = 36), or bipolar disorder (BP; n = 20). Some studies investigated neurocognitive (e.g., mild cognitive impairment [MCI] and dementia; n = 14), neurodevelopmental (e.g., autism spectrum disorder and attention-deficit/hyperactivity disorder; n = 9), anxiety (e.g., social anxiety, panic, obsessive-compulsive, and posttraumatic stress disorders; n = 9), or other disorders (e.g., alcohol dependence, or eating, somatoform, sleep, or personality disorders; n = 9). Notably, some studies combined several disorders into a single group or examined more than one type or subtype of psychiatric disorder.

Among the 121 studies, except for those examining neurodevelopmental or neurocognitive disorders, many, if not most, patients were on psychotropic medications on the testing day. In addition, 83 (69%) of the studies reported enrolling only those patients who were free from any neurological conditions, including traumatic head injury, and 95 (79%) of the studies reported excluding patients who showed certain forms of psychiatric comorbidities, particularly substance/alcohol abuse or dependence. A total of 84 (69%) of the studies excluded patients for other reasons, including a history of electroconvulsive therapy, physical illness (e.g., endocrinological disease), and left-handedness.

Very few studies reported significant differences in age (8/121; 7%), sex (3/120; 3%), or handedness (0/103; 0%) between patients and controls. A handful of studies reported

significantly lower values for education (19/83; 23%) or intellectual ability or global cognition (20/60; 33%) among patients relative to controls, which were driven primarily by studies involving SCZ (*n* = 13) patients or SCZ (*n* = 11) and MCI/dementia (*n* = 6) patients, respectively. Notably, these results do not inform the effects of age, sex, handedness, and intellectual ability on cortical activation during the VFT in psychiatric disorders, which are beyond the scope of the present study. Readers interested in these aspects are referred to empirical studies using a large sample size (e.g., Chou et al., 2015; Koike et al., 2020).

3.2.2. Task Design

Supplementary Table 2 details the task design used by each study. Among the 121 studies, the VFT was administered primarily in Japanese (*n* = 78; 64%), Chinese (*n* = 18; 15%), or German (*n* = 13; 11%); the VFT was rarely administered in other languages. Excluding seven studies administered in either Chinese or Japanese that should be considered simultaneously phonemic and semantic, the phonemic VFT (*n* = 98; 86%) was more commonly administered than the semantic VFT (*n* = 34; 30%). Two separate VFT tasks, one phonemic and one semantic, were used in 18 (16%) of the 114 studies. An additional study used a phonemic VFT and an idea VFT. The number of blocks ranged from 1 to 5 per task, with a median of one. The task block duration ranged from 12.3 to 180 s, with a median of 60 s. After excluding three studies that reported an incomplete block structure, each VFT block consisted of either more than one trial (most often three 20-s trials), as in 84 (71%) studies, or only one trial (either 30 or 60 s), as in the remaining studies.

Among all studies, 100 (83%) included a control task, 88 (88%) of which involved repeating syllables, letters, or numbers. The remaining studies involved repeating weekdays or nonsense

words. Most, if not all, studies used a task that began and ended with a rest or control task period. If more than one VFT block was employed, the VFT blocks were alternated with rest or control task periods. In addition, the VFT paradigm was computerized in 51 (42%) studies and not computerized in 15 (12%); this information was missing from the remaining studies. Only nine (7%) of the studies mentioned having the participants practice the task before the data recording began.

3.2.3. Signal Acquisition

Supplementary Table 3 details the fNIRS signal acquisition process used in each study. Across studies, fNIRS data were acquired primarily using one of the following three systems: the Hitachi (e.g., ETG-4000; n = 86), the Shimadzu (e.g., FOIRE-300; n = 11), or the Hamamatsu (e.g., NIRO-200; n = 9) systems. Other systems were used in less than five studies each. Notably, a 52-channel Hitachi system involving the use of a 3 × 11 probe, in which the center of the lowest probe was anchored at Fpz, was employed in 60 (50%) of the studies. The sampling rate ranged from 0.5 to 24 Hz, with a median of 10 Hz. The source–detector separation ranged from 8 to 50 mm, with a median of 30 mm. The number of measurement channels ranged from 1 to 84, with a median of 52. The channels covered the frontal region in all studies, 80 (66%) of which extended to non-frontal regions (most often the STC). All studies examined both sides of the brain, with the exception of five (3%) studies that only examined the left hemisphere.

3.2.4. Data Preprocessing

Supplementary Table 4 presents the preprocessing steps taken in each study. Only eight (7%) of the 121 studies indicated the software used to preprocess fNIRS data. Channel or data removal

was performed in 57 (47%) of the studies, 36 (63%) of which implemented the algorithm described by Takizawa et al. (2008, 2014). Among the 57 studies, only 12 (21%) compared the number of available channels between patients and controls; the number was comparable between the two groups in 11 studies. A total of 65 (54%) studies removed spike artifacts and slow drifts by using linear fitting, in combination with either a moving average (n = 61) or low-pass filtering (n = 4). Other studies applied bandpass filtering alone (n = 11), precoloring with either a discrete cosine transform (n = 1) or wavelet minimum description length detrending (n = 1), or used a moving average, combined with either a bandpass (n = 1) or a cosine filter (n = 1). Some studies removed high-frequency noise while neglecting slow drifts by applying a low-pass filter only (n = 3), whereas others removed slow drifts while ignoring high-frequency noise by applying linear fitting (n = 10), a cosine filter (n = 1), or wavelet minimum description length detrending (n = 1).

Among the 63 studies that applied a moving average, almost all (n = 60; 95%) used a 5-s window, whereas others used a 1- or 2.2-s window. Among the 75 studies that performed linear fitting, most (n = 58; 77%) used the 10-s pre-VFT period and the 5- or 10-s period starting 50–65 s after the end of the VFT block. Among all 121 studies, only seven (6%) used motion/systemic correction, which included spline interpolation (n = 3), correlation-based signal improvement (n = 3), common average reference (n = 2), principal component analysis (n = 1), or short-channel regression (n = 1). Only ten (8%) of the 121 studies converted the fNIRS data into a unitless metric, a DPF-independent approach, prior to statistical analysis.

3.2.5. Data Analysis

Supplementary Table 5 presents the analytic method and results for each study. The HbO data were analyzed in all 121 studies, and the HbR data were analyzed in approximately one-third (*n* = 41; 34%) of the studies. The mean level of HbO or HbR changes was examined in almost all (*n* = 118; 98%) studies; other aspects of HbO or HbR changes, most notably laterality (*n* = 11) and latency (*n* = 8), were examined in a total of only 26 (21%) studies. Multiple comparison correction was applied in 68 (68%) of the 100 studies performed using more than one channel- or region-wise comparison between patients and controls, most of which did not specify *a priori* hypotheses regarding the exact foci of the group effect. Among these 68 studies, the correction was primarily performed using the false discovery rate (*n* = 48; 72%), although the Bonferroni correction (*n* = 12) and other methods (*n* = 8) were also used.

3.3. Summary of Study Findings

3.3.1. Qualitative Synthesis

As can be seen from Table 1, 118 (98%) of the 121 studies included in the review compared HbO changes during the VFT between patients and healthy controls. Many of the studies used the phonemic VFT, with a 60-s block comprising three 20-s trials, focused on the bilateral (ventral) frontotemporal region, and corrected the fNIRS results using the false discovery rate. Among the 118 studies that analyzed HbO changes, 97 (82%) found significantly reduced HbO increases in patients compared with those in controls for at least one region during VFT performance (SCZ: 39/41; MDD: 32/35; BP: 13/20; Neurocognitive disorders: 6/13; Neurodevelopmental disorders: 6/9; Anxiety disorders: 6/9; Others: 9/9). Only 19 (16%) studies

reported no significant differences in HbO increases between patients and controls for any region. Four (3%) studies found a significantly larger HbO increase in patients compared with controls for at least one region (BP: n = 2; Social anxiety disorder, frontotemporal dementia: n = 1 each).

Next, we investigated the specificity of the diagnosis effect associated with HbO increases. Among the 97 studies that reported significantly reduced HbO increases in patients relative to controls, three did not allow the investigation of HbR changes because HbO and HbR were integrated as a result of motion correction. In the remaining 94 studies, only 31 (33%) also examined HbR decreases during the VFT. Among these 31 studies, only 12 (39%) and 5 (16%) found significantly smaller and larger decreases in HbR in patients relative to controls, respectively, whereas 14 (45%) found no significant differences in HbR changes between patients and controls. Among the 21 studies that did not report any significantly reduced HbO increases in any region for patients relative to controls, only eight (38%) examined HbR decreases. However, only two (25%) studies, both of which found no significant group differences in HbO increases, reported significantly smaller HbR decreases in patients compared with controls. The remaining six studies (75%) reported no significant differences in HbR decreases between the two groups. When considering all the 39 studies that examined HbR decreases, the proportion of studies that reported significantly reduced HbR decreases in patients was fairly large for SCZ but small for other disorders (SCZ: 9/13; MDD: 2/8; BP: 2/9; Neurocognitive: 2/6; Neurodevelopmental: 0/4; Anxiety: 0/1; Others: 0/4).

In addition, among all 121 studies, patients and healthy controls were matched according to VFT performance in three studies, and the behavioral data were not reported in nine studies. In the

remaining 109 studies, 53 (49%) found significantly impaired VFT performances in patients compared with controls, and the significant results were driven by studies involving patients with SCZ or neurocognitive disorders (SCZ: 29/38; MDD: 12/35; BP: 7/19; Neurocognitive: 10/10; Neurodevelopmental: 2/8; Anxiety: 1/8; Others: 1/8). Taken together, these results suggested that frontal HbO increases were more likely than either frontal HbR decreases or task performance to yield a significant diagnosis effect for all of the psychiatric disorders studied, with the notable exception of neurocognitive disorders. Task performance seems more sensitive than frontal HbO increases for the detection of MCI/dementia.

3.3.2. Meta-Analysis based on the Proportion of Channels that Exhibited Significant Reductions in HbO Changes in Patients

Supplementary Table 6 lists the proportion of single channels or clusters that showed significantly reduced HbO increases in patients relative to controls for each study (see Appendix for a visual presentation of the individual studies' results). For both MDD and SCZ, information regarding this variable could be extracted for each of the eight regions from at least 19 studies. Due to the considerable number of reports that would yield reliable results, one-sample sign tests (i.e., meta-analyses) were separately performed for these two groups to determine whether patients exhibited hypoactivation in any of these eight regions during the VFT. The tests revealed significantly reduced HbO increases in all eight regions for both the MDD and SCZ groups ($ps < .001$, Figure 4a).

Next, six paired-sample sign tests were performed for MDD and SCZ studies, separately, to contrast the left vs. right and the inferior vs. superior dimensions of hypoactivation: (1) right vlPFC+STC vs. right dlPFC; (2) bilateral dFPC vs. bilateral vFPC; (3) left vlPFC+STC vs. left

dlPFC; (4) left dlPFC vs. right dlPFC; (5) left vlPFC vs. right vlPFC; and (6) left STC s. right STC. We found that MDD patients exhibited significantly greater hypoactivation in all inferior regions compared with superior regions: right vlPFC+STC compared with the right dlPFC (p = .002); vFPC compared with the dFPC (p = .001); and left vlPFC+STC compared with the left dlPFC (p = .013). No other results were significant, with ps ranging from 0.30 to 0.77. In addition, we found that SCZ patients also exhibited significantly greater hypoactivation in some inferior regions compared with superior regions: vFPC compared with the dFPC (p = .007); and left vlPFC+STC compared with the left dlPFC (p = .004). Interestingly, in contrast to MDD patients, SCZ patients also showed significantly greater hypoactivation in the left STC compared with the corresponding right regions (p = .039). No other results were significant (ps values ranging from 0.065 to 0.092, Figure 4b).

3.3.3. Meta-Analysis based on the Effect Sizes of Group Differences in HbO changes

Supplementary Tables 7 and 8 list the means and standard errors of Hedges' g in individual MDD and SCZ studies, respectively. In addition, Tables 2 and 3 present the meta-analytic results for MDD and SCZ studies, respectively. The meta-analyses involved 18–31 studies including 584–1039 MDD or 1069–1712 SCZ patients and 774–1391 healthy controls. Overall, the results generated from this approach were comparable, if not identical, to those derived from the previous approach. Specifically, we found significant hypoactivation in all regions for both MDD and SCZ, gs from 0.20 to 0.51, ps < .001. For both groups, hypoactivation was more pronounced in inferior compared with superior regions, gs from 0.09 to 0.15, ps < .020. In addition, a significant laterality effect (i.e., greater hypoactivation on the left side) was found in

the STC only for SCZ, $g = 0.09$, $p = .018$ (other regions: ps > .30), whereas no significant laterality effect was found for MDD, ps > .22.

4. Discussion

The fNIRS-VFT paradigm has the potential to serve as a low-cost and easy-to-administer probe for understanding PFC function in psychiatric patients. In this review, we aimed to clarify the clinical utility of fNIRS measurements and the soundness of the applied methods in previous studies. As anticipated, most of the fNIRS studies using the VFT found significantly reduced HbO increases (i.e., hypoactivation) in some frontal and temporal regions during VFT performance in a variety of psychiatric disorders. However, studies of neurocognitive disorders tended to report no significant reductions in the levels of frontal HbO increases, such as in patients with MCI or dementia. Instead, amnestic MCI and Alzheimer's disease patients were found to exhibit parietal hypoactivation or reduced left-lateralized frontal activation during the VFT relative to controls (Arai et al., 2006; Hock et al., 1997; Yeung et al., 2016). Hypoactivation may be a manifestation of neural inefficiency or represent pathological changes in neurovascular coupling. When VFT performance is compromised, hypoactivation may also be associated with task disengagement, a lack of motivation, or both. Because only 3% of studies reported hyperactivation, neural compensation during the VFT, at least in the regions measured on the frontal cortical surface, appears to be uncommon among patients.

In contrast, the findings of reduced HbR decreases in psychiatric patients relative to controls were relatively rare. This discrepancy may suggest the increased sensitivity of HbO for the detection of brain activity or the selective dysregulation of neurovascular coupling due to psychopathology. In addition, interestingly, slightly less than half of the studies reported

impaired VFT performance in patients, with very few reporting poor VFT performances, without altered brain activation in patients. These findings suggested that for most psychiatric disorders, fNIRS measurements, specifically HbO increases, are more sensitive than task performance variables for the detection of psychopathology in the context of VFT, supporting the unique value of fNIRS.

Due to the considerable number of reports, meta-analyses were performed for MDD and SCZ separately, both of which are characterized by frontal lobe pathology. Interestingly, although these two commonly comorbid disorders exhibited reductions in HbO increases across frontotemporal regions during VFT performance, they each showed partially distinct hypoactivation patterns. Specifically, although both MDD and SCZ patients exhibited greater hypoactivation in inferior than in superior prefrontal regions, only SCZ patients showed greater hypoactivation in the left STC relative to the right homologous region. This laterality effect was specific to the STC, which may indicate a functional difference between the STC and the lateral PFC, as well as a disturbance in lexical semantic processes subserved by the left anterior temporal pole in SCZ (Tsapkini et al., 2011). In contrast, no similar laterality effect was observed in MDD patients. Four VFT studies compared MDD and SCZ, but none of these found any significant difference in HbO increases for any single channel or cluster between these two groups (Kinou et al., 2013; Ohi et al., 2017; Takizawa et al., 2014; Wei et al., 2020). Therefore, when using fNIRS and the VFT, the lateralization of hypoactivation (i.e., difference in the extent of hypoactivation between the two hemispheres) may be a more discriminative biomarker than the level of hypoactivation in one particular region for differentiating between MDD and SCZ, which may be necessary when prescribing individualized treatment.

We undertook two complementary meta-analytic approaches as a first step toward the synthesis of the existing fNIRS literature. Both approaches have their strengths and drawbacks. Specifically, although the meta-analysis based on the significance of activity made little assumptions about the outcome variables, it considered neither the sample size nor the effect size. In contrast, while the meta-analysis based on the effect size took into account the sample size, it was based on some assumptions about the outcome variables (due to incomplete reporting of statistical test results in many studies). As such, caution must be taken when interpreting the magnitude of the pooled effects. Despite methodological differences, the meta-analytic results derived from the two approaches were comparable, if not identical, suggesting that converging evidence was obtained.

Based on the evaluation of the methods used in these studies, the findings of frontotemporal hypoactivation in psychiatric patients generally cannot be solely explained by differences in demographic or intellectual features between patients and healthy controls. For SCZ, however, a discernible proportion of studies reported lower educational or intellectual levels among patients compared with controls. No SCZ studies reported the effects of education on HbO increases, and four SCZ studies reported the effects of cognitive functioning on HbO increases (Itakura et al., 2017; Noda et al., 2017; Pu, Nakagome, Itakura et al., 2015; Yamamuro et al., 2018). For these four studies, one reported a positive association between cognitive functioning (verbal memory) and HbO increases in SCZ patients (Yamamuro et al., 2018). Thus, whether the hypoactivation findings may have been confounded by educational or intellectual factors remains unclear.

At least two-thirds of the studies applied multiple comparison corrections when comparing patients and controls. Thus, the group differences in fNIRS measurements could not be

attributable to inflated false positives due to multiple testing. The frequent use of multiple comparison corrections in fNIRS studies using the VFT to probe psychiatric disorders is a notable strength of this body of literature. Neurological illness, major physical illness, and electroconvulsive therapy are also unlikely to be confounding factors because many studies only included patients who were free from any history of these factors. Alcohol/substance abuse or dependence is also an implausible confounding factor for most of these studies, except for those that specifically examined alcohol dependence.

Few studies analyzed metrics that did not rely on the assumed DPF, which is strongly dependent on the brain-scalp distance (or scattering-to-absorption ratio). Alterations in brain structures (e.g., gray matter abnormalities) involving the frontal lobes are common in psychiatric disorders and may alter tissue optical properties in the prefrontal regions of psychiatric patients (Belleau et al., 2019; Whitford et al., 2006). Consequently, whether the observed reductions in HbO changes during the VFT in patients, including the widespread reductions in HbO increases across the frontotemporal regions in MDD and SCZ patients, are attributable to group differences in the prefrontal cortical structure is unclear. In addition, very few studies applied motion or systemic correction methods, which have been previously shown to be effective for reducing non-brain-evoked activity artifacts (Brigadoi et al., 2014). Motions lower the signal-to-noise ratio, and evoked systemic activity inflates HbO increases. Although the number of available channels that remain after channel rejection can provide some indication of signal quality, and most of the 12 studies that considered this index reported no differences between patients and controls, the evidence is limited in size and, thus, inconclusive. Thus, whether the observed reductions in HbO increases in patients were due to more head or body motions or less systemic activity (i.e., less arousal or task engagement) in patients than in controls remains unclear.

Most studies employed only one 60-s block of the phonemic VFT, which consisted of three successive 20-s word generation trials. Although this task design was useful for achieving homogeneity and facilitating comparisons between studies, it did not allow any investigations of the effects of retrieval demand (e.g., task duration and the types and categories of VFT), which may provide a more complete picture of the neural processing that occurs during access to long-term memory. Specifically, studies using the standard 60-s VFT have shown that executive strategic processes do not become primarily engaged until after 15–20 s of word generation (Crowe, 1998; Hurks et al., 2010). Lesion work has suggested that the dorsomedial PFC, but not other PFC subregions, is necessary for sustaining VFT performance (Stuss et al., 1998). In addition, the phonemic and semantic VFTs are mediated by overlapping but distinct vlPFC subregion (BA 44 during the phonemic VFT and BA 45 and 47 during the phonemic and semantic VFTs; Wagner et al., 2014). The retrieval of words that belong to different semantic categories (e.g., living and non-living things) also depends on different subregions within the frontal or temporal lobes (Pobric et al., 2010; Yeung et al., 2019).

Accordingly, we have several recommendations for the field. First, unitless metrics (e.g., effect sizes) should be used to minimize the potential confounding effects of brain structures that may differ between psychiatric patients and healthy controls. Second, motion correction methods, especially wavelet filtering, should be applied to enhance signal quality (Brigadoi et al., 2014). Third, HbR data should be reported whenever possible because this data provides important information regarding the basis of HbO changes (Cui et al., 2010). Fourth, the lateralization of activation and the effects of retrieval effort and switching demand on cortical activation should be examined to further our understanding of the neural mechanisms underlying VFT performance in psychiatric disorders. The functional difference between the STC and the lateral

PFC should also be considered in light of the regional specificity regarding the laterality of hypoactivation in SCZ patients. Finally, we recommend future studies to fully report effect sizes, preferably in a table, to facilitate a finer-grain quantitative synthesis in the future.

In summary, our review provides supportive evidence to suggest that the combined use of fNIRS and the VFT has a unique and valuable role in psychiatry. It also critically evaluates different possibilities that may explain the previous fNIRS findings of VFT studies involving psychiatric patients. We highlight the strengths and weaknesses of the existing literature and provide recommendations to enhance future research practices in the field. This review constitutes an important first step toward a quantitative synthesis of the findings involving the use of the VFT as an fNIRS probe for psychiatric disorders. Our findings encourage the expanded application of the fNIRS-VFT paradigm to improve the identification, differentiation, and management of psychiatric disorders and stress the importance of taking rigorous steps to improve the strength and clarity of evidence derived using the fNIRS method.

5. Acknowledgment

This work was supported by the Departmental Start-up Fund awarded by the Hong Kong Polytechnic University to MKY (P0034754) and JL (P0035194).

6. Conflict of Interest

No authors had a conflict of interest to disclose.

7. References

Abbott, C. C., Jaramillo, A., Wilcox, E. C., & Hamilton, A. D. (2013). Antipsychotic drug effects in schizophrenia: a review of longitudinal FMRI investigations and neural interpretations. *Current medicinal chemistry*, *20*(3), 428-437.

Akiyama, T., Koeda, M., Okubo, Y., & Kimura, M. (2018). Hypofunction of left dorsolateral prefrontal cortex in depression during verbal fluency task: A multi-channel near-infrared spectroscopy study. *Journal of affective disorders*, *231*, 83-90.

Arai, H., Takano, M., Miyakawa, K., Ota, T., Takahashi, T., Asaka, H., & Kawaguchi, T. (2006). A quantitative near-infrared spectroscopy study: a decrease in cerebral hemoglobin oxygenation in Alzheimer's disease and mild cognitive impairment. *Brain and cognition*, *61*(2), 189-194.

Beauchamp, M. S., Beurlot, M. R., Fava, E., Nath, A. R., Parikh, N. A., Saad, Z. S., ... & Oghalai, J. S. (2011). The developmental trajectory of brain-scalp distance from birth through childhood: implications for functional neuroimaging. *PloS one*, *6*(9), e24981.

Belleau, E. L., Treadway, M. T., & Pizzagalli, D. A. (2019). The impact of stress and major depressive disorder on hippocampal and medial prefrontal cortex morphology. *Biological psychiatry*, *85*(6), 443-453.

Bokat, C. E., & Goldberg, T. E. (2003). Letter and category fluency in schizophrenic patients: a meta-analysis. *Schizophrenia research*, *64*(1), 73-78.

Bora, E., Fornito, A., Pantelis, C., & Yücel, M. (2012). Gray matter abnormalities in major depressive disorder: a meta-analysis of voxel based morphometry studies. *Journal of affective disorders*, *138*(1-2), 9-18.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis (1st Ed.)*. John Wiley & Sons.

Brigadoi, S., Ceccherini, L., Cutini, S., Scarpa, F., Scatturin, P., Selb, J., ... & Cooper, R. J. (2014). Motion artifacts in functional near-infrared spectroscopy: a comparison of motion correction techniques applied to real cognitive data. *Neuroimage*, *85*, 181-191.

Buckley, P. F., Miller, B. J., Lehrer, D. S., & Castle, D. J. (2009). Psychiatric comorbidities and schizophrenia. *Schizophrenia bulletin*, *35*(2), 383-402.

Crowe, S. F. (1998). Decrease in performance on the verbal fluency test as a function of time: Evaluation in a young healthy sample. *Journal of clinical and experimental neuropsychology*, *20*(3), 391-401.

Cui, X., Bray, S., & Reiss, A. L. (2010). Functional near infrared spectroscopy (NIRS) signal improvement based on negative correlation between oxygenated and deoxygenated hemoglobin dynamics. *Neuroimage*, *49*(4), 3039-3046.

Cui, X., Bray, S., Bryant, D. M., Glover, G. H., & Reiss, A. L. (2011). A quantitative comparison of NIRS and fMRI across multiple cognitive tasks. *Neuroimage*, *54*(4), 2808-2821.

Duerden, E. G., Mak-Fan, K. M., Taylor, M. J., & Roberts, S. W. (2012). Regional differences in grey and white matter in children and adults with autism spectrum disorders: An activation likelihood estimate (ALE) meta-analysis. *Autism Research*, *5*(1), 49-66.

Ehlis, A. C., Schneider, S., Dresler, T., & Fallgatter, A. J. (2014). Application of functional near-infrared spectroscopy in psychiatry. *Neuroimage*, *85*, 478-488.

Fox, N. A., Bakermans-Kranenburg, M. J., Yoo, K. H., Bowman, L. C., Cannon, E. N., Vanderwert, R. E., ... & Van IJzendoorn, M. H. (2016). Assessing human mirror activity with EEG mu rhythm: A meta-analysis. *Psychological bulletin*, *142*(3), 291.

Hafeman, D. M., Chang, K. D., Garrett, A. S., Sanders, E. M., & Phillips, M. L. (2012). Effects of medication on neuroimaging findings in bipolar disorder: an updated review. *Bipolar disorders*, *14*(4), 375-410.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis (1st Ed.).* Academic press.

Henry, J. D., & Crawford, J. R. (2004). A meta-analytic review of verbal fluency performance in patients with traumatic brain injury. *Neuropsychology*, *18*(4), 621.

Henry, J. D., & Crawford, J. R. (2005). A meta-analytic review of verbal fluency deficits in depression. *Journal of clinical and experimental neuropsychology*, *27*(1), 78-101.

Hock, C., Villringer, K., Müller-Spahn, F., Wenzel, R., Heekeren, H., Schuh-Hofer, S., ... & Villringer, A. (1997). Decrease in parietal cerebral hemoglobin oxygenation during performance of a verbal fluency task in patients with Alzheimer's disease monitored by

means of near-infrared spectroscopy (NIRS)—correlation with simultaneous rCBF-PET measurements. *Brain research*, *755*(2), 293-303.

Hurks, P. P., Schrans, D., Meijs, C., Wassenberg, R., Feron, F. J. M., & Jolles, J. (2010). Developmental changes in semantic verbal fluency: Analyses of word productivity as a function of time, clustering, and switching. *Child Neuropsychology*, *16*(4), 366-387.

IntHout, J., Ioannidis, J. P., & Borm, G. F. (2014). The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC medical research methodology*, *14*(1), 1-12.

Itakura, M., Pu, S., Ohdachi, H., Matsumura, H., Yokoyama, K., Nagata, I., ... & Kaneko, K. (2017). Association between social functioning and prefrontal cortex function during a verbal fluency task in schizophrenia: A near-infrared spectroscopic study. *Psychiatry and Clinical Neurosciences*, *71*(11), 769-779.

Kinou, M., Takizawa, R., Marumo, K., Kawasaki, S., Kawakubo, Y., Fukuda, M., & Kasai, K. (2013). Differential spatiotemporal characteristics of the prefrontal hemodynamic response and their association with functional impairment in schizophrenia and major depression. *Schizophrenia research*, *150*(2-3), 459-467.

Klöppel, S., Abdulkadir, A., Jack Jr, C. R., Koutsouleris, N., Mourão-Miranda, J., & Vemuri, P. (2012). Diagnostic neuroimaging across diseases. *Neuroimage*, *61*(2), 457-463.

Koike, S., Sakakibara, E., Satomura, Y., Sakurada, H., Yamagishi, M., Matsuoka, J., ... & Kasai, K. (2020). Shared functional impairment in the prefrontal cortex affects symptom

severity across psychiatric disorders. *Psychological Medicine*, 1-10. Advance online publication. https://doi.org/10.1017/S0033291720004742

Matson, J. L., & Goldin, R. L. (2013). Comorbidity and autism: Trends, topics and future directions. *Research in Autism Spectrum Disorders*, *7*(10), 1228-1233.

McKenzie, J. E., & Brennan, S. E (2020). Chapter 12: Synthesizing and presenting findings using other methods. In Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions* version 6.1 (updated September 2020). Available from www.training.cochrane.org/handbook.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Prisma Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS med*, *6*(7), e1000097.

Monsch, A. U., Bondi, M. W., Butters, N., Salmon, D. P., Katzman, R., & Thal, L. J. (1992). Comparisons of verbal fluency tasks in the detection of dementia of the Alzheimer type. *Archives of neurology*, *49*(12), 1253-1258.

Noda, T., Nakagome, K., Setoyama, S., & Matsushima, E. (2017). Working memory and prefrontal/temporal hemodynamic responses during post-task period in patients with schizophrenia: A multi-channel near-infrared spectroscopy study. *Journal of Psychiatric Research*, *95*, 288-298.

Ohi, K., Shimada, T., Kihara, H., Yasuyama, T., Sawai, K., Matsuda, Y., ... & Kawasaki, Y. (2017). Impact of familial loading on prefrontal activation in major psychiatric disorders: a near-infrared spectroscopy (NIRS) study. *Scientific reports*, *7*, 44268.

Pinti, P., Tachtsidis, I., Hamilton, A., Hirsch, J., Aichelburg, C., Gilbert, S., & Burgess, P. W. (2020). The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience. *Annals of the New York Academy of Sciences*, *1464*(1), 5.

Plichta, M. M., Herrmann, M. J., Baehne, C. G., Ehlis, A. C., Richter, M. M., Pauli, P., & Fallgatter, A. J. (2006). Event-related functional near-infrared spectroscopy (fNIRS): are the measurements reliable?. *Neuroimage*, *31*(1), 116-124.

Pobric, G., Jefferies, E., & Ralph, M. A. L. (2010). Category-specific versus category-general semantic impairment induced by transcranial magnetic stimulation. *Current biology*, *20*(10), 964-968.

Pu, S., Nakagome, K., Itakura, M., Yamanashi, T., Sugie, T., Miura, A., ... & Kaneko, K. (2015). Self-reported social functioning and prefrontal hemodynamic responses during a cognitive task in schizophrenia. *Psychiatry Research: Neuroimaging*, *234*(1), 121-129.

Pu, S., Nakagome, K., Yamada, T., Yokoyama, K., Matsumura, H., Mitani, H., ... & Kaneko, K. (2012). The relationship between the prefrontal activation during a verbal fluency task and stress-coping style in major depressive disorder: a near-infrared spectroscopy study. *Journal of psychiatric research*, *46*(11), 1427-1434.

Quaresima, V., & Ferrari, M. (2019). Functional near-infrared spectroscopy (fNIRS) for assessing cerebral cortex function during human behavior in natural/social situations: a concise review. *Organizational Research Methods*, *22*(1), 46-68.

Raucher-Chéné, D., Achim, A. M., Kaladjian, A., & Besche-Richard, C. (2017). Verbal fluency in bipolar disorders: A systematic review and meta-analysis. *Journal of affective disorders*, *207*, 359-366.

Robinson, G., Shallice, T., Bozzali, M., & Cipolotti, L. (2012). The differing roles of the frontal cortex in fluency tests. *Brain*, *135*(7), 2202-2214.

Rubia, K., Alegria, A. A., Cubillo, A. I., Smith, A. B., Brammer, M. J., & Radua, J. (2014). Effects of stimulants on brain function in attention-deficit/hyperactivity disorder: a systematic review and meta-analysis. *Biological psychiatry*, *76*(8), 616-628.

Schecklmann, M., Ehlis, A. C., Plichta, M. M., & Fallgatter, A. J. (2008). Functional near-infrared spectroscopy: a long-term reliable tool for measuring brain activity during verbal fluency. *Neuroimage*, *43*(1), 147-155.

Schecklmann, M., Ehlis, A. C., Plichta, M. M., Boutter, H. K., Metzger, F. G., & Fallgatter, A. J. (2007). Altered frontal brain oxygenation in detoxified alcohol dependent patients with unaffected verbal fluency performance. *Psychiatry Research: Neuroimaging*, *156*(2), 129-138.

Schecklmann, M., Ehlis, A. C., Plichta, M. M., Romanos, J., Heine, M., Boreatti-Hümmer, A., ... & Fallgatter, A. J. (2008). Diminished prefrontal oxygenation with normal and

above-average verbal fluency performance in adult ADHD. *Journal of psychiatric research*, *43*(2), 98-106.

Scholkmann, F., Kleiser, S., Metz, A. J., Zimmermann, R., Pavia, J. M., Wolf, U., & Wolf, M. (2014). A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. *Neuroimage*, *85*, 6-27.

Schroeter, M. L., Zysset, S., Kruggel, F., & Von Cramon, D. Y. (2003). Age dependency of the hemodynamic response as measured by functional near-infrared spectroscopy. *Neuroimage*, *19*(3), 555-564.

Schwarzer, G., Carpenter, J. R., & Rücker, G. (2015). *Meta-analysis with R* (Vol. 4784). New York: springer.

Singh, I., & Rose, N. (2009). Biomarkers in psychiatry. *Nature*, *460*(7252), 202-207.

Stuss, D. T., Alexander, M. P., Hamer, L., Palumbo, C., Dempster, R., Binns, M., ... & Izukawa, D. (1998). The effects of focal anterior and posterior brain lesions on verbal fluency. *Journal of the International Neuropsychological Society*, *4*(3), 265-278.

Takizawa, R., Fukuda, M., Kawasaki, S., Kasai, K., Mimura, M., Pu, S., ... & Joint Project for Psychiatric Application of Near-Infrared Spectroscopy (JPSY-NIRS) Group. (2014). Neuroimaging-aided differential diagnosis of the depressive state. *Neuroimage*, *85*, 498-507.

Takizawa, R., Kasai, K., Kawakubo, Y., Marumo, K., Kawasaki, S., Yamasue, H., & Fukuda, M. (2008). Reduced frontopolar activation during verbal fluency task in schizophrenia: a

multi-channel near-infrared spectroscopy study. *Schizophrenia research*, *99*(1-3), 250-262.

Thalheimer, W., & Cook, S. (2002). How to calculate effect sizes from published research: A simplified methodology. *Work-Learning Research*, *1*, 1-9.

Tsapkini, K., Frangakis, C. E., & Hillis, A. E. (2011). The function of the left anterior temporal pole: evidence from acute stroke and infarct volume. *Brain*, *134*(10), 3094-3105.

Tsujii, N., Mikawa, W., Tsujimoto, E., Adachi, T., Niwa, A., Ono, H., & Shirakawa, O. (2017). Reduced left precentral regional responses in patients with major depressive disorder and history of suicide attempts. *PLoS One*, *12*(4), e0175249.

Tsuzuki, D., Jurcak, V., Singh, A. K., Okamoto, M., Watanabe, E., & Dan, I. (2007). Virtual spatial registration of stand-alone fNIRS data to MNI space. *Neuroimage*, *34*(4), 1506-1518.

Viechtbauer, W., & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Research synthesis methods*, *1*(2), 112-125.

Wagner, S., Sebastian, A., Lieb, K., Tüscher, O., & Tadić, A. (2014). A coordinate-based ALE functional MRI meta-analysis of brain activation during verbal fluency tasks in healthy control subjects. *BMC neuroscience*, *15*(1), 19.

Wei, Y., Chen, Q., Curtin, A., Tu, L., Tang, X., Tang, Y., ... & Wang, J. (2020). Functional near-infrared spectroscopy (fNIRS) as a tool to assist the diagnosis of major psychiatric

disorders in a Chinese population. *European archives of psychiatry and clinical neuroscience*. Advance online publication. https://doi.org/10.1007/s00406-020-01125-y

Whiteman, A. C., Santosa, H., Chen, D. F., Perlman, S. B., & Huppert, T. (2017). Investigation of the sensitivity of functional near-infrared spectroscopy brain imaging to anatomical variations in 5-to 11-year-old children. *Neurophotonics*, *5*(1), 011009.

Whitfield-Gabrieli, S., & Ford, J. M. (2012). Default mode network activity and connectivity in psychopathology. *Annual review of clinical psychology*, *8*, 49-76.

Whitford, T. J., Grieve, S. M., Farrow, T. F., Gomes, L., Brennan, J., Harris, A. W., ... & Williams, L. M. (2006). Progressive grey matter atrophy over the first 2–3 years of illness in first-episode schizophrenia: a tensor-based morphometry study. *Neuroimage*, *32*(2), 511-519.

Xia, C. H., Ma, Z., Ciric, R., Gu, S., Betzel, R. F., Kaczkurkin, A. N., ... & Satterthwaite, T. D. (2018). Linked dimensions of psychopathology and connectivity in functional brain networks. *Nature communications*, *9*(1), 1-14.

Yamamuro, K., Kimoto, S., Iida, J., Kishimoto, N., Tanaka, S., Toritsuka, M., ... & Kishimoto, T. (2018). Distinct patterns of blood oxygenation in the prefrontal cortex in clinical phenotypes of schizophrenia and bipolar disorder. *Journal of affective disorders*, *234*, 45-53.

Yeung, M. K. (2021). An optical window into brain function in children and adolescents: A systematic review of functional near-infrared spectroscopy studies. *NeuroImage, 227,* 117672.

Yeung, M. K., & Chan, A. S. (2020). Functional near-infrared spectroscopy reveals decreased resting oxygenation levels and task-related oxygenation changes in mild cognitive impairment and dementia: A systematic review. *Journal of Psychiatric Research, 124,* 58-76.

Yeung, M. K., & Chan, A. S. (2021). A Systematic Review of the Application of Functional Near-Infrared Spectroscopy to the Study of Cerebral Hemodynamics in Healthy Aging. *Neuropsychology Review, 31***,** 139-166.

Yeung, M. K., Lee, T. L., & Chan, A. S. (2019). Right-lateralized frontal activation underlies successful updating of verbal working memory in adolescents with high-functioning autism spectrum disorder. *Biological psychology, 148,* 107743.

Yeung, M. K., Sze, S. L., Woo, J., Kwok, T., Shum, D. H., Yu, R., & Chan, A. S. (2016). Altered frontal lateralization underlies the category fluency deficits in older adults with mild cognitive impairment: a near-infrared spectroscopy study. *Frontiers in aging neuroscience, 8,* 59.

Yüksel, C., McCarthy, J., Shinn, A., Pfaff, D. L., Baker, J. T., Heckers, S., ... & Öngür, D. (2012). Gray matter volume in schizophrenia and bipolar disorder with psychotic features. *Schizophrenia research, 138*(2-3), 177-182.

Table 1

A brief summary of functional near-infrared spectroscopy (fNIRS) studies using the verbal fluency test (VFT) to probe psychiatric disorders

First author	Year	Disorder (criteria)	Healthy controls (*N*, mean age (*SD*) [range])	Patients (*N*, mean age (*SD*) [range])	VFT type	Paradigm flow	fNIRS instrument	Number of channels (region)	Aspect of signal change	HbO increases (patients > healthy controls)	HbR decreases (healthy controls > patients)
Fallgatter	1997	AD (NINCDS-ADRDA)	10, 30 (2)	10, 67 (11)	P, S	(>27R, 180T, >55R) per task	Critikon 2020 (two pieces) (Johnson and Johnson Medical)	2 (F)	Level, laterality	Phonemic+semantic: level: o (0/1 region); Left laterality: - (1/1 region)	Phonemic+semantic: level: o (0/1 region); Left laterality: o (0/1 region)
Hock	1997	AD (NINCDS-ADRDA)	Exp 1: 19, 67 (10); Exp 2: 8, 60 (16)	Exp 1: 19, 71 (10); Exp 2: 10, 65 (13)	P	120R, 120T, 120R	NIRO-500 (Exp 2: two pieces) (Hamamatsu)	Exp 1: 1 (P); Exp 2: 2 (F, P)	Level	Exp 1: - (1/1 ch); Exp 2: - (1/2 ch)	Exp 1: o (0/1 ch); Exp 2: o (0/2 ch)
Matsuo	2000	MDD+BP (8/9 MDD; DSM-IV)	10, 60 (6)	9, 66 (6)	P	180R, 60C, 60R, 60T, 60R	HEO-200 (Omron)	1 (F)	Level	- (1/1 ch)	+ (1/1 ch)
Matsuo	2002	MDD, BP (DSM-IV)	21, 50 (13)	MDD: 14, 56 (17); BP: 11, 48 (13)	P	180R, 60C, 60R, 60T, 60R	HEO-200 (Omron)	1 (F)	Level	MDD, BP: - (1/1 ch)	MDD, BP: o (0/1 ch)
Matsuo	2003	Post-traumatic stress disorder (PTSD; DSM-IV)	26, 43 (12)	8, 46 (17)	P	180R, 60C, 60R, 60T, 60R	ETG-100 (Hitachi)	24 (F)	Level	- (6/14 ch)	o (0/1 region)
Herrmann	2004	Depression (ICD-10)	9, 35 (6) [27–44]	9, 37 (14) [19–62]	P	60R, 180T	NIRO-300 (Hamamatsu)	2 (F)	Level, laterality	Level: - (1/1 region), still significant after controlling for age; Left laterality: o (0/1 region)	Level: o (0/1 region); Left laterality: o (0/1 region)
Matsuo	2004	BP (DSM-IV)	9, 47 (10)	9, 47 (15)	P	180R, 60C, 60R, 60T, 60R	ETG-100 (Hitachi)	24 (F, T)	Level	- (1/1 region)	o (0/1 region)
Suto	2004	MDD, SCZ (DSM-IV)	16, 43 (4) [36–52]	MDD: 48 [23–60]; SCZ: 38 [23–60]	P	30C, 60T, 60C	ETG-100 (Hitachi)	48 (F, T, P)	Level	MDD: early: - (16/48 ch), late, after: o (0/48 ch); SCZ: early: - (17/48 ch), late: o (0/48 ch), after + (1/48 ch)	n.r.
Watanabe	2004	SCZ (DSM-IV)	31, 36 (12)	62, 40 (12)	P	n.r.	HEO-200 (Omron)	1 (F)	Level	- (1/1 ch), still significant after controlling for VFT performance	+ (1/1 ch)
Kubota	2005	SCZ (DSM-IV)	19, 37 (14)	16, 38 (13)	P, S	(20C, 90T, 90T) per task	NIRO-300 (Hamamatsu)	2 (F)	Level	Phonemic, semantic: o; Smaller increases during	Phonemic, semantic: o; Similar changes during the phonemic

										the phonemic relative to the semantic VFT in SCZ	and the semantic VFT in SCZ
Arai	2006	AD (NINCDS-ADRA), aMCI (Petersen et al., 2001)	32, 57 (6)	AD: 15, 59 (4); aMCI; 15, 63 (6)	P	R, 60T, n.r.	ETG-7000 (Hitachi)	84 (F, P, O)	Level	AD: - (3/4 regions); aMCI: - (1/4 regions)	n.r.
Kameyama	2006	BP, MDD (DSM-IV)	17, 43 (5) [36–52]	BP: 17, 41 (13) [20–62]; MDD: 11, 45 (13) [24–59]	P	30C, 60T, 60C	ETG-100 (two pieces) (Hitachi)	48 (F, T, P)	Level	BP: early: - (8/28 ch), late: + (4/28 ch); MDD: early: - (16/28 ch), late: - (3/28 ch)	BP: + (2/28 ch); MDD: o (0/28 ch)
Kuwabara	2006	ASD (DSM-IV)	10, 28 (4)	10, 27 (7) [19–37]	P	30C, 60T, 60C	ETG-100 (Hitachi)	24 (F)	Level, laterality	Level: - (13/24 ch); Left laterality: o (1/1 region)	Level: - (1/24 ch); Left laterality: o (0/1 region)
Nishimura	2006	Panic disorder (DSM-IV)	44, 30 (10)	109, 36 (9)	P	30C, 60T, 70C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	- (52/52 ch)	n.r.
Ehlis	2007	SCZ (ICD-10)	12, 34 (11)	12, 34 (10)	P, S	10R, (30T/30C, 30R, … x 6) per task	ETG-100 (two pieces) (Hitachi)	22 (F, T)	Level	Phonemic: - (8/22 ch); Semantic: - (2/22 ch), still significant after controlling for VFT performance or medication	Phonemic+semantic: - (1/1 region), not significant after controlling for VFT performance
Matsuo	2007	BP (DSM-IV)	13, 39 (13)	14, 39 (13)	P	(60C, 60T, … x 3)	ETG-100 (two pieces) (Hitachi)	24 (F)	Level, laterality	Level, left laterality: during, after task: o (0/1 region)	Level, left laterality: during, after task: o (0/1 region)
Richter	2007	Dementia (ICD-10)	12, 66 (3)	12, 69 (8)	P, S	(30T, 30R, … x 3) per task	ETG-100 (Hitachi)	24 (F, T)	Level	Phonemic: o (0/1 region); semantic: - (3/3 regions; female patients only)	Phonemic: o (0/2 regions); Semantic: - (2/2 regions; female patients only)
Schecklmann	2007	Alcohol dependence (DSM-IV)	17, 44 (10)	17, 44 (10)	P, S	10R, (30T/30C, 30R, … x 6) per task	ETG-4000 (Hitachi)	44 (F, T)	Level	Phonemic: - (4/22 ch pairs); Semantic: o (0/22 ch pairs)	Phonemic, semantic: o (0/22 ch pairs)
Uehara	2007	Eating disorders (DSM-IV)	11, 27 (2) [18–32]	11, 21 (6) [14–38]	P	60C, 60T, 60C	ETG-100 (two pieces) (Hitachi)	48 (F, T, P)	Level	During task: - (n.r.); After task: + (n.r.)	During task: - (n.r.); after task: o (n.r.)
Herrmann	2008	AD (ICD-10)	16, 70 (8)	16, 68 (5)	P, S	(30T, 30R, … x 3) per task	ETG-100 (two pieces) (Hitachi)	24 (F, T)	Level, laterality	Phonemic: level: - (22/24 ch); Left laterality: o (0/1 region); Semantic: level: - (n.r.); Left laterality: o (0/1 region), unchanged after controlling for VFT performance	Phonemic, semantic: level: o (0/1 region); Left laterality: o (0/1 region)
Ohta	2008	Panic disorder, MDD (DSM-IV)	24, 36 (17)	Panic disorder: 21, 35 (11); MDD: 17, 43 (18)	P	30C, 60T, 60C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	Panic disorder: - (13/52 ch); MDD: - (16/52 ch)	n.r.
Pu	2008	MDD (DSM-IV)	30, 72 (5)	24, 72 (6)	P	30C, 60T, 70C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	- (39/52 ch)	n.r.

Schecklmann	2008	ADHD (DSM-IV)	14, 41 (9)	13, 40 (11)	P, S	10R, (30T/30C, 30R, … x 6) per task	ETG-4000 (Hitachi)	44 (F, T)	Level	Phonemic: - (7/44 ch); semantic: - (7/44 ch)	Phonemic, semantic: o (0/44 ch)
Takizawa	2008	SCZ (DSM-IV)	70, 37 (14)	55, 40 (11)	P	30C, 60T, 70C	ETG-4000 (Hitachi)	52 (F, T, P)	Level, rate	Level: - (20/52 ch, still significant after covarying VFT performance or IQ); Slope: - (33/52 ch)	+ (2/52 ch)
Ikezawa	2009	SCZ (DSM-IV)	30, 37 (9)	30, 39 (12)	P, S	(30C, 60T, 60C) per task	NIRO-200 (Hamamatsu)	2 (F)	Level, laterality	Level: phonemic: - (1/1 region), semantic: o (0/1 region); Left laterality: phonemic, semantic: o (0/1 region)	Level: phonemic: o (0/1 region), semantic: + (1/1 region); Left laterality: phonemic, semantic: o (0/1 region)
Kawakubo	2009	ASD (DSM-IV)	Children: 14, 11 (3); adults: 13, 26 (5)	Children: 14, 13 (3); Adults: 13, 27 (6)	P	30R, 30T, 30R	NIRO-200 (Hamamatsu)	2 (F)	Level, laterality	Level: children: o (0/1 region), adults: - (1/1 region); Left laterality: children, adults: o (0/1 region)	Level, left laterality: children, adults: o (0/1 region)
Kubota	2009	BP (DSM-IV)	27, 40 (14) [18–65]	29, 41 (11) [18–65]	P, S	(30C, 90T, 90T) per task	NIRO-300 (Hamamatsu)	2 (F)	Level	Phonemic, semantic: o (0/2 ch)	Phonemic, semantic: o (0/2 ch)
Quaresima	2009	SCZ (DSM-IV)	9, 33 (16)	9, 32 (8)	P	120R, 120T	NIRO-300 (Hamamatsu)	2 (F)	Level	- (2/2 ch)	+ (2/2 ch)
Takizawa, Hashimoto	2009	SCZ (DSM-IV)	60, 31 (7)	40, 41 (11)	P	30C, 60T, 70C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	- (52/52 ch)	- (21/52 ch)
Takizawa, Tochigi	2009	SCZ (DSM-IV)	60, 37–38 (13–14)	45, 41–42 (10–12)	P	30C, 60T, 70C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	- (50/52 ch)	+ (18/52 ch)
Azechi	2010	SCZ (DSM-IV, ICD-10)	1st group: 30, 37 (9); 2nd group: 30, 40 (13)	1st group: 30, 39 (12); 2nd group: 30, 40 (13)	P, S	30C, 60T, 60C	NIRO-200 (Hamamatsu)	2 (F)	Level	1st, 2nd groups: phonemic, semantic: - (1/1 region)	n.r.
Suda	2010	Eating disorders (DSM-IV-TR)	27, 22 (2)	27, 24 (5)	P	30C, 60T, 70C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	- (16/31 ch)	o (0/31 ch)
Takeshi	2010	SCZ (ICD-10)	16, 25 (3)	18, 25 (6)	P, idea	30C, 60T/300T, 60C	OMM3000 (Shimadzu)	24 (F)	Level	Phonemic: - (17/24 ch); Idea: - (10/24 ch)	n.r.
Iwanami	2011	Asperger (DSM-IV)	18, 31 (5)	20, 27 (9) [18–60]	P, S	(30C, 60T, 70C) per task	ETG-4000 (Hitachi)	52 (F, T, P)	Level	Phonemic: - (1/1 region); Semantic: o (0/1 region)	n.r.
Koike	2011	SCZ (DSM-IV)	30, 24 (5)	First-episode: 27, 25 (7) [15–	P	30C, 60T, 70C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	First-episode: - (32/52 ch); Chronic: - (39/52 ch)	n.r.

				40]; Chronic: 38, 31 (6)							
Nagamitsu	2011	Anorexia nervosa (Great Ormond Street criteria)	12, 14 (1)	16, 14 (1)	P	20C, (15T, 15C, … x 5)	ETG-4000 (Hitachi)	24 (F)	Level	- (18/24 ch)	o (0/24 ch)
Ohi	2011	SCZ (DSM-IV)	216, 37 (12)	127, 37 (12)	P	30C, 60T, 60C	NIRO-200 (Hamamatsu)	2 (F)	Level	- (2/2 ch)	n.r.
Reif	2011	SCZ (ICD-10)	28, 39–47 (15)	26, 46–47 (10–11)	P	10R, (30T/30C, 30R, … x 6)	ETG-4000 (Hitachi)	44 (F, T, P)	Level	- (1/1 region)	+ (1/1 region)
Dresler	2012	Alcohol dependence (DSM-IV)	20, 49 (7)	Withdrawal: 20, 48 (7); Detoxified: 20, 49 (7); Abstinent: 20, 49 (6)	P, S	10R, (30T/30C, 30R, … x 6) per task	ETG-4000 (Hitachi)	44 (F, T, P)	Level	Phonemic: detoxified, withdrawal: - (2/2 regions), abstinent: o (0/2 regions); Semantic: withdrawal: - (1/2 regions), detoxified, abstinent: o (0/2 regions); Control: withdrawal, detoxified, abstinent: o (0/2 regions)	/ (combined with HbO after correlation-based signal improvement)
Noda	2012	MDD (DSM-IV)	30, 35 (9)	30, 37 (12)	P	30C, 60T, 70C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	- (22/31 ch)	n.r.
Pu	2012	MDD (DSM-IV-TR)	30, 51 (20)	26, 48 (19)	P	30C, 60T, 70C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	- (47/52 ch)	n.r.
Shimodera	2012	SCZ (DSM-IV-TR)	26, 41 (10)	31, 42 (16) [18–80]	P	30R, 60T, 70R	OMM-3000/16 (Shimadzu)	42 (F)	Level, latency, rate, rate, fluctuation, level	Level: - (19/42 ch, after controlling for VFT performance: 3/42 ch); Latency: o (0/42 ch); Up-slope: - (8/42 ch); Down-slope: o (0/42 ch); Fluctuation: - (9/42 ch); Average amplitude: - (9/42 ch)	n.r.
Hirosawa	2013	OCD (DSM-IV)	20, 37 (8)	20, 38 (11)	P	30C, 60T, 70C	FOIRE-3000 (Shimadzu)	42 (F)	Level	o (0/14 ch)	n.r.
Ikeda	2013	MDD (DSM-IV-TR)	16, 36 (5)	21, 41 (10)	P	30C, 60T	Pocket NIRS-NIY (DynaSense)	2 (F)	Level, laterality	Level: - (1/1 region); Difference between two sides: o (0/1 region)	n.r.
Kinou	2013	SCZ, MDD (DSM-IV)	32, 46 (14)	SCZ: 32, 42 (10); MDD: 32, 45 (10)	P	30C, 60T, 70C	ETG-4000 (Hitachi)	52 (F, T, P)	Level, rate	Level: SCZ: - (50/52 ch); MDD: - (46/52 ch); Slope: SCZ: - (26/52 ch; FPC, lPFC); MDD: o (0/52 ch)	Level: SCZ: + (8/52 ch), MDD: + (6/52 ch); Slope: SCZ, MDD: o (0/52 ch)
Pu	2013	SCZ (DSM-IV-TR)	30, 32 (11)	30, 32 (10)	P	30C, 60T, 70C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	- (47/52 ch), after controlling for VFT performance: - (31/52 ch)	n.r.
Chuang	2014	SCZ (DSM-IV)	46, 35 (11)	53, 34 (11)	P	30C, 60T, 70C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	- (6/52 ch)	n.r.

Deppermann	2014	Panic disorder (DSM-IV-TR)	23, 33 [19–64]	Sham: 22, 36 [22–56]; Verum: 22, 38 [19–63]	P, S	10R, (30T/30C, 30R, … x 9)	ETG-4000 (Hitachi)	52 (F, T, P)	Level	Phonemic: - (6/6 regions); Semantic: - (2/6 regions); Control: - (0/6 regions)	/ (combined with HbO after correlation-based signal improvement)
Fujiki	2014	SCZ (ICD-10)	35, 28 (7)	35, 29 (6)	P+S	(12.3T+C, … x 20–25)	ETG-4000 (Hitachi)	44 (F, T)	Level	- (5/44 ch)	n.r.
Ishii-Takahashi	2014	ASD, ADHD (DSM-IV)	21, 29 (6)	ASD: 21, 31 (7); ADHD: 19, 31 (7)	P	30C, 60T, 70C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	ASD, ADHD: During task: o (0/52 ch); During+after task: - (2/52 ch; right lPFC)	ASD, ADHD: during, during+after task: o (0/52 ch)
Katayama	2014	Eating disorders (DSM-IV-TR)	31, 29 (8)	20, 29 (8) [≥ 18]	P	30C, 60T, 70C	FOIRE-3000 (Shimadzu)	22 (F)	Level	During task: - (6/22 ch); After task: o (0/22 ch)	n.r.
Kito	2014	Depression (DSM-IV), AD	33, 70 (6) [≥ 60]	Depression: 30, 71 (7) [≥ 60]; AD: 28, 77 (7) [≥ 60]	P	30C, 60T, 30C	FOIRE-3000 (Shimadzu)	44 (F, P)	Level	Depression: - (6/44 ch); AD: o (0/44 ch)	n.r.
Liu	2014	MDD (DSM-IV-TR)	30, 33 (11) [18–65]	30, 38 (13) [18–65]	S	30R, (30T, 30R, … x 4)	FOIRE-3000 (Shimadzu)	45 (F)	Level	- (15/45 ch)	n.r.
Marumo	2014	SCZ (DSM-IV)	56, 41 (12)	56, 40 (11)	P, S	(30C, 60T, 70C) per task	ETG-4000 (Hitachi)	52 (F, T, P)	Level	Phonemic: - (19/52 ch); Semantic: - (13/52 ch)	Phonemic: o (0/52 ch); Semantic: + (4/52 ch)
Nishimura	2014	SCZ (DSM-IV-TR)	73, 38–41 (14–17)	73, 36–38 (11–14)	P	30C, 60T, 70C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	- (31/52 ch)	n.r.
Takizawa	2014	MDD, BP, SCZ (DSM-IV)	590, 44 (16)	MDD: 153, 33 (13); BP: 134, 44 (15); SCZ: 136, 44 (12)	P	10C, 60T, 55C	ETG-4000 (Hitachi)	52 (F, T, P)	Level, latency	Level: MDD+BP+SCZ > HC and BP > SCZ: - (2/2 regions); MDD > BP+SCZ: o (0/2 regions); Latency: MDD+BP+SCZ > HC and MDD > BP+SCZ and BP > SCZ: - (1/2 regions)	Level, latency: MDD, BP, SCZ: o (0/2 regions)
Tsujii	2014	MDD (DSM-IV)	24, 39 (9)	Melancholic MDD: 32, 41 (15); non-melancholic MDD: 28, 39 (12)	P	30C, 60T, 70C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	- (22/31 ch)	n.r.
Akashi	2015	MDD (DSM-IV)	48, 39 (10)	52, 42 (13)	P	30C, 60T, 55C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	- (45/52 ch)	n.r.
Chou	2015	SCZ (DSM-IV)	106, 32 (7) [16–59]	109, 33 (10) [16–59]	P	30C, 60T, 70C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	- (31/52 ch)	n.r.
Kinoshita	2015	SCZ (DSM-IV)	48, 33 (6)	31, 36 (8)	S	30C, 60T, 70C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	- (38/52 ch), after covarying IQ and task performance: o (0/52 ch)	n.r.

Mikawa	2015	BP (DSM-IV)	28, 37 (10)	Depressed: 30, 41 (14); Euthymic: 25, 42 (11)	P	30C, 60T, 60C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	- (18/52 ch)	n.r.
Nishimura, Takahashi, Ohtani, Ikeda-Sugita, Kasai	2015	BP (DSM-IV-TR)	12, 46 (7)	Depressed: 16, 45 (9); Hypomanic: 11, 44 (13)	P	30C, 60T, 70C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	Depressed: - (12/52 ch); Hypomanic: - (7/52 ch)	Depressed, hypomanic: o (0/52 ch)
Nishimura, Takahashi, Ohtani, Ikeda-Sugita, Okada	2015	BP (DSM-IV-TR)	65 [≥ 20]	33 [≥ 20]	P	30C, 60T, 70C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	- (38/52 ch)	+ (5/52 ch)
Ohtani	2015	MDD, BP (DSM-IV-TR)	14, 34 (8)	MDD: 10, 39 (12); BP: 18, 40 (9)	P	30C, 60T, 70C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	MDD, BP: - (2/3 regions)	n.r.
Ono	2015	BP (DSM-IV)	15, 33 (8) [23–43]	13, 38 (7) [19–45]	P	80C, (60T, 80C, … x 3)	ETG-4000 (Hitachi)	46 (F, T)	Level	o (0/6 regions)	n.r.
Pu, Nakagome, Itakura	2015	SCZ (DSM-IV-TR)	30, 32 (11)	33, 32 (9)	P	30C, 60T, 70C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	- (48/52 ch)	n.r.
Pu, Nakagome, Yamada	2015	MDD (DSM-IV)	67, 58 (18)	67, 58 (16)	P	30C, 60T, 70C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	- (50/52 ch)	n.r.
Quan	2015	SCZ (DSM-IV)	100, 34 (12)	140, 34 (12)	P+S	30C, 60T, 30C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	- (41/52 ch)	+ (24/52 ch)
Tomioka	2015	MDD (DSM-IV)	62, 52 (17)	25, 52 (17)	P	30C, 60T, 60C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	- (20/52 ch)	n.r.
Watanabe	2015	SCZ (DSM-IV-TR)	100, 44 (8)	199, 46 (14)	P	30C, 60T, 30C	ETG-4000 (Hitachi)	22 (F, T)	Level	- (1/1 region)	n.r.
Yokoyama	2015	Social anxiety disorder (DSM-IV-TR)	35, 37 (11)	24, 36 (13)	P	30C, 60T, 70C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	o (0/52 ch)	n.r.
Chou	2016	SCZ (DSM-IV-TR)	29, 30 (11)	28, 31 (6.1)	P, S	(30C, 60T, 70C) per task	ETG-4000 (Hitachi)	52 (F, T, P)	Level	Phonemic: - (30/52 ch); Semantic: o (0/52 ch)	n.r.
Iwashiro	2016	SCZ (DSM-IV)	16, 24 [16–36]	18, 25 [17–35]	P	30C, 60T, 70C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	- (1/2 regions)	n.r.
Kawashima	2016	Social anxiety disorder (DSM-IV)	152, 26 (6)	145, 27 (8)	P	30C, 60T, 70C	ETG-7100 (Hitachi)	47 (F, T)	Level	+ (4/47 ch)	n.r.
Koike	2016	SCZ (DSM-IV)	30, 26 (5)	31, 24 (7) [15–40]	P	30C, 60T, 70C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	- (44/52 ch)	n.r.
Metzger	2016	AD (McKhann, 2011), bvFTD (Rascovsky et al., 2011)	8, 66 (7)	AD: 8, 74 (5); bvFTD: 8, 68 (10)	P, S	(30T/30C, 30R, … x 9)	ETG-4000 (Hitachi)	44 (F, T, P)	Level	Phonemic, semantic: AD, bvFTD: ±	/ (combined with HbO after correlation-based signal improvement)

Tsujii	2016	MDD (DSM-IV)	68, 41 (11)	Melancholic: 30, 42 (12); non-melancholic: 52, 41 (12)	P	30C, 60T, 70C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	- (51/52 ch)	n.r.
Yeung	2016	MCI (Petersen et al., 2014; National Institute on Aging and Alzheimer's Association)	26, 69 (6) [69–91]	26, 69 (6) [69–91]	S	30C, (60T, 60C, ... x 2)	OEG-SpO2 (Spectratech)	16 (F)	Level, laterality	Level: o (0/2 regions); Left laterality: - (1/1 region)	n.r.
Chou	2017	SCZ (DSM-IV-TR)	33, 29 (10)	28, 30 (6)	P	30C, 60T, 70C	ETG-4000 (Hitachi)	52 (F, T, P)	Level, laterality	Level: - (1/2 regions); Left laterality: - (1/3 regions)	n.r.
Hirano	2017	MDD+BP (ICD-10)	108, 59 (13)	30, 59 (14)	P	30C, 60T, 60C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	- (40/52 ch)	n.r.
Itakura	2017	SCZ (DSM-IV-TR)	22, 36 (11) [21–58]	23, 42 (13) [20–65]	P	30C, 60T, 70C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	- (3/3 regions)	n.r.
Ma	2017	MDD (DSM-IV)	30, 35 (9) [18–60]	Menopausal: 30, 51 (6) [40–60]; Non-menopausal: 30, 38 (11) [18–60]	S	(30R, 30T, 30R, … x 4)	FOIRE-3000 (Shimadzu)	45 (F)	Level	Menopausal: - (4/45 ch); Non-menopausal: - (5/45 ch)	With, without menopausal symptoms: o (0/52 ch)
Masuda	2017	MDD (DSM-IV-TR)	63, 42 (1)	47, 49 (15)	P	30C, 60T, 70C	ETG-7100 (Hitachi)	47 (F, T)	Level	Responsive to SSRIs: - (7/47 ch); Unresponsive to SSRIs: - (6/47 ch)	n.r.
Nishida	2017	MDD (DSM-IV-TR)	15, 46 (11)	14, 46 (12)	P	30C, 60T, 70C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	- (13/52 ch)	n.r.
Noda	2017	SCZ (DSM-IV)	30, 32 (9) [18–60]	30, 33 (8) [18–60]	P	30C, 60T, 70C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	During task: - (37/52 ch); After task: - (14/52 ch)	n.r.
Ohi	2017	SCZ, MDD, BP (DSM-V)	51, 36 (12)	SCZ: 45, 35 (9); MDD: 26, 41 (13); BP: 22, 40 (13)	P	> 10C, 60T, > 55C	ETG-4000 (Hitachi)	52 (F, T, P)	Level, latency	Level: SCZ, MDD, BP: - (1/1 region); Latency: SCZ, MDD, BP: o (0/1 region)	n.r.
Ono	2017	BP (DSM-IV)	10, 15 (1)	10, 16 (0.5)	P	(60T, 80C, … x 3)	ETG-4000 (Hitachi)	24 (F)	Level	Initial 20 s: - (2/4 regions); Final 20 s: o (0/4 regions)	n.r.
Ren	2017	Somatoform pain disorder (DSM-IV)	24, 33 (11)	24, 37 (11)	S	30R, 30T, 30R	FOIRE-3000 (Shimadzu)	45 (F)	Level	- (12/45 ch)	n.r.
Sun	2017	Chronic insomnia disorder (International Classification	25, 41 (11) [21–59]	24, 41 (10) [23–62]	S	(30R, 30T, 30R, … x 4)	FOIRE-3000 (Shimadzu)	45 (F)	Level	- (20/45 ch)	n.r.

		of Sleep Disorders, 3rd version)									
Takeda	2017	OCD (DSM-V)	42, 35 (10)	42, 36 (10)	P	30C, 60T, 70C	ETG-4000 (Hitachi)	24 (F)	Level	SSRI responders: - (1/1 region)	n.r.
Tsujii	2017	MDD (DSM-IV)	40, 38 (11)	With suicidal attempts: 30, 38 (10); Without suicidal attempts: 38, 39 (10)	P	30C, 60T, 70C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	With suicidal attempts: - (23/52 ch); Without suicidal attempts: - (20/52 ch)	n.r.
Wang	2017	MDD (DSM-IV)	37, 36 (11)	First-episode: 36, 39 (14); Recurrent: 34, 43 (14)	P+S	30C, 60T, 30R	ETG-4000 (Hitachi)	52 (F, T, P)	Level	First-episode, recurrent: o (0/52 ch)	n.r.
Yamagata	2017	ADHD (DSM-IV)	38, 30 (5)	63, 31 (8)	P	30C, 60T, 60C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	Time windows of > 30 s: - (8/52 ch)	n.r.
Yap	2017	AD, MCI	31, 73 (9)	MCI: 12, 73 (8); AD: 18, 75 (10)	S	20R, (60T, 20R, … x 3)	OT-R40 (Hitachi)	52 (F, T, P)	Level, activation area, rate, latency	Level, activation area, slope, latency: MCI, AD: o (0/2 regions)	n.r.
Akiyama	2018	MDD (DSM-IV-TR)	50, 33 (8)	177, 47 (15)	P	30C, 60T, 60C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	- (31/31 ch)	n.r.
Fu	2018	BP (DSM-V)	32, 25 (2)	43, 27 (7) [16–50]	S	30C, (30T, 30 C, … x 4)	ETG-4000 (Psyche-Ark Science & Technology Development Co., Beijing)	40 (F)	Level	- (7/40 ch)	n.r.
Hirata	2018	ASD, SCZ (DSM-V)	18, 35 [28–39]	ASD: 13, 30 [23–39]; SCZ: 15, 36 [29–47]	P	30C, 60T, 60C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	ASD: - (3/3 regions); SCZ: - (1/3 regions)	n.r.
Katzorke	2018	MCI (Portet et al., 2006)	55, 74 (2) [70–77]	55, 74 (2) [70–77]	P, S	(30T/30C, 30R, … x 9)	ETG-4000 (Hitachi)	52 (F, T, P)	Level, laterality	Level, left laterality: phonemic, semantic: o (0/1 region)	Level: phonemic: + (1/1 region), semantic: o (0/1 region); Left laterality: phonemic, semantic: o (0/1 region)
Kono	2018	SCZ (DSM-IV)	10, 30 (9)	On olanzapine: 10, 27 (8); On risperidone: 10, 30 (9)	P	30C, 60T, 70C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	On olanzapine; o (0/52 ch); On risperidone: - (1/52 ch)	n.r.
Luo	2018	SCZ (DSM-IV)	17, 26 (6) [18–45]	16, 29 (8) [18–45]	S	30R, (30T,	CW5 (TechEN)	32 (F)	Level	Before treatment: - (28/32 ch)	n.r.

						30R, … x 4)					
Pu	2018	SCZ (DSM-IV)	32, 32 (11)	32, 31 (10)	P	30C, 60T, 70C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	- (3/3 regions)	n.r.
Sun	2018	BP with or without psychotic symptoms (DSM-IV)	23, 33 (10)	Without psychotic symptoms: 31, 31 (9); With psychotic symptoms: 29, 28 (7)	S	(15R, 30T, 13R, … x 4)	FOIRE-3000 (Shimadzu)	45 (F)	Level	Without psychotic symptoms: - (4/45 ch); With psychotic symptoms: - (11/45 ch)	n.r.
Yamamuro	2018	SCZ, BP (DSM-V)	26, 49 (8)	SCZ: 38, 46 (8); BP: 34, 50 (10)	P	40C, 60T, 90C	ETG-4000 (Hitachi)	24 (F)	Level	SCZ: - (13/24 ch); BP: - (1/24 ch)	n.r.
Baik	2019	MDD (DSM-V)	64, 33 (13) [19–65]	42, 38 (14) [19–65]	P	30R, (30C, 30T, 30R, … x 3)	NIRSIT (OBELAB)	48 (F)	Laterality	Left laterality: - (1/1 region)	Left laterality: o (0/1 region)
Downey	2019	MDD+BP (DSM-IV)	51, 56 (12)	18 (17 in analysis), 52 (16)	S	30R, (30T, 30R, … x 10)	NTS (Gowerlabs)	24 (F)	Level	o (0/2 regions)	o (0/2 regions)
Feng	2019	MDD (DSM-V)	15, 31 (10)	15, 31 (13)	S	(15R, 30T, 15R, … x 3)	FOIRE-3000 (Shimadzu)	45 (F)	Level	- (4/45 ch)	n.r.
Liao	2019	OCD (DSM-IV)	70, 30 (8) [16–55]	70, 28 (10) [16–55]	P+S	30C, 60T, 50R	ETG-4000 (Hitachi)	52 (F, T, P)	Level	- (16/52 ch)	n.r.
Nguyen	2019	MCI (National Institute on Aging and Alzheimer's Association and International Working Group)	42, 74 (4)	42, 76 (4)	P, S	30R, (30T, 30R, … x 6)	OE-MV7385-P (Opto ENG)	6 (F)	Connectivity	- (inter- but not intra-hemispheric ventral frontopolar cortex)	o
Tian	2019	SCZ (DSM-IV)	70, 37 (1)	100, 35 (1) [18–60]	P+S	30C, 60T, 70C	ETG-4000 (Hitachi)	52 (F, T, P)	Latency	+ (1/1 region)	n.r.
Yang	2019	MCI (n.r.)	9, 68 (5)	15, 69 (7)	S	30R, (60C, 60T, 30R, … x 3)	NIRSIT (OBELAB)	48 (F)	Level, rate, distribution, distribution	Level: o (0/3 regions); Rate: - (1/3 regions; left lPFC; especially during the 20–60 s task period); Distribution (skewness, kurtosis): o (0/3 regions)	Level: + (2/3 regions; during the 5–65 s task period); Slope: - (1/3 regions; during the 5–60 s task period); Skewness, kurtosis: o (0/1 region)

Yeung	2019	ASD (DSM-V)	22, 14 (2)	22, 14 (2)	S	30C, (60T, 60C, ... x 2)	OEG-SpO2 (Spectratech)	16 (F)	Level, localization	Level: o (0/3 regions); lPFC-FPC difference: - (1/2 categories; animals but not means of transportation)	n.r.
Yoon	2019	aMCI, naMCI (Petersen et al., 2014)	15, 68 (6) [> 65]	aMCI: 9, 67 (7) [> 65]; naMCI: 6, 68 (7) [> 65]	S	30R, (60T, 60R, ... x 3)	NIRSIT (OBELAB)	48 (F)	Level	aMCI, naMCI: o (0/2 regions)	n.r.
Devrimci-Ozguven	2020	BP (DSM-IV)	23, 35 (10)	30, 39 (12)	P	(30C, 30T, … x 2)	ETG-4000 (Hitachi)	24 (F)	Level	+ (3/24 ch)	n.r.
Husain, Tang	2020	MDD, borderline personality disorder (DSM-V)	31, 32 (11) [21–65]	MDD: 31, 32 (10) [21–65]; borderline personality disorder: 31, 32 (1) [21–65]	P	30C, 60T, 70C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	MDD: - (44/52 ch); Borderline personality disorder: - (43/52 ch)	n.r.
Husain, Yu	2020	MDD (DSM-V)	105, 36 (13) [21–65]	105, 36 (13) [21–65]	P	30C, 60T, 70C	ETG-4000 (Hitachi)	52 (F, T, P)	Level, latency	Level: - (2/2 regions); Latency: o (0/2 regions)	n.r.
Ji	2020	SCZ (DSM-IV-TR)	100, 34 (12) [18–78]	200, 34 (12) [17–62]	P+S	30C, 60T, 30C	ETG-4000 (Hitachi)	52 (F, T, P)	Level, connectivity	Level: - (13/52 ch); Connectivity: - (within frontotemporal regions)	n.r.
Kiriyama	2020	MDD (DSM-V)	22, 42 (11)	18, 44 (9)	P	30C, 60T, 70C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	- (15/52 ch)	n.r.
Ota	2020	ASD (DSM-V)	20, 29 (6)	20, 27 (4)	P	30C, 60T, 60C	ETG-4000 (Hitachi)	52 (F, T, P)	Level, latency	Level: o (0/24 ch); Latency: o (0/10 ch)	n.r.
Tsujii	2020	MDD (DSM-IV)	56, 41 (13)	45, 43 (13)	P	30C, 60T, 70C	ETG-4000 (Hitachi)	52 (F, T, P)	Level	- (12/52 ch)	n.r.
Wei	2020	SCZ, MDD, BP (DSM-IV, ICD-10)	101, 28 (6)	SCZ: 198, 38 (13); MDD: 54, 34 (13); BP: 64, 34 (13)	P+S	30C, 60T, 70C	ETG-4000 (Hitachi)	52 (F, T, P)	Level, latency	Level: SCZ, MDD, BP: - (2/2 regions); Latency: SCZ: + (1/2 regions); MDD, BP: o (0/2 regions)	n.r.
Yang	2020	MCI (n.r.)	9, 68 (5)	15, 69 (7)	S	30R, (30C, 60T, 30R, … x 3)	NIRSIT (OBELAB)	48 (F)	Level	- (especially during the 5–25 s task period)	n.r.

Note. Regarding the VFT type, the letters “P” and “S” refer to the phonemic and semantic VFTs, respectively. Regarding the paradigm flow, the letters “C”, “R”, and “T” refer to the control task, rest, and VFT periods, respectively. Regarding the measurement regions, the letters “F”, “O”, “P”, and “T” refer to the frontal, occipital, parietal, and temporal lobes, respectively. Regarding the hemoglobin concentration results, the symbols “+” and “-“ indicate significantly larger and smaller values in patients compared to healthy controls, respectively, whereas the symbol “o” indicates no significant difference between patients and controls. Reduced brain activation in patients compared to controls is indicated by a “-“ for the level of change in oxyhemoglobin concentration (HbO) and by a “+” for the level of change in deoxyhemoglobin concentration (HbR). AD =Alzheimer’s disease; ASD = autism spectrum disorder; BP = bipolar disorder; bvFTD = behavioral variant of frontotemporal dementia; FDR = false discovery rate; MCI = mild cognitive impairment (aMCI: amnestic subtype; naMCI: nonamnestic subtype); MDD = major depressive disorder; n.r. = not reported; SCZ = schizophrenia.

Table 2.

Random-effects meta-analyses on the differences between healthy controls (HC) and patients with major depressive disorder (MDD) in (a) regional activation and (b) the inferior vs. superior and the left vs. right dimensions of activation (reported in Section 3.3.3)

	Number of studies (*k*)	HC (total *N*)	MDD (total *N*)	Pooled Hedges' *g*	95% *CI*	*t*	*p*
a)							
Right dlPFC	20	841	651	0.23	[0.14; 0.31]	5.43	< .001***
Bilateral dFPC	22	874	710	0.20	[0.10; 0.29]	4.45	< .001***
Left dlPFC	20	841	651	0.28	[0.17; 0.38]	5.39	< .001***
Right STC	19	837	813	0.39	[0.29; 0.49]	8.50	< .001***
Right vlPFC[#]	22	895	741	0.35	[0.25; 0.45]	7.32	< .001***
Bilateral vFPC[#]	31	1246	1039	0.41	[0.31; 0.52]	8.01	< .001***
Left vlPFC[@]	21	847	689	0.40	[0.39; 0.50]	8.06	< .001***
Left STC	19	837	813	0.39	[0.29; 0.49]	8.50	< .001***
(b)							
Right dlPFC > Right vlPFC+STC	19	774	584	0.14	[0.05; 0.24]	3.06	.007**
Bilateral dFPC > bilateral vFPC	21	807	643	0.15	[0.07; 0.23]	4.04	< .001***
Left dlPFC > left vlPFC+STC	20	841	651	0.12	[0.04; 0.19]	3.34	.004**
Left dlPFC > right dlPFC	20	841	651	0.05	[-0.04; 0.15]	1.14	.27
Left vlPFC > Right vlPFC	23	945	918	0.05	[-0.03; 0.14]	1.25	.22
Left STC > right STC	19	837	813	-0.06	[-0.18; 0.06]	-1.06	.30

Note. One ([#]) or two ([@]) outliers were objectively identified and excluded from some meta-analyses. Asterisks indicate the significance levels of the pooled effects. **$p < .01$, ***$p < .001$.

Table 3.

Random-effects meta-analyses on the differences between healthy controls (HC) and patients with schizophrenia (SCZ) in (a) regional activation and (b) the inferior vs. superior and the left vs. right dimensions of activation (reported in Section 3.3.3)

	Number of studies (*k*)	HC (total *N*)	SCZ (total *N*)	Pooled Hedges' *g*	95% *CI*	*t*	*p*
a)							
Right dlPFC	20	988	1085	0.21	[0.13; 0.30]	5.25	< .001***
Bilateral dFPC	24	1058	1191	0.22	[0.12; 0.32]	4.64	< .001***
Left dlPFC	21	1000	1097	0.24	[0.16; 0.32]	6.07	< .001***
Right STC	19	971	1069	0.27	[0.17; 0.37]	5.45	< .001***
Right vlPFC	23	1256	1256	0.34	[0.26; 0.41]	9.35	< .001***
Bilateral vFPC[%]	28	1391	1712	0.51	[0.40; 0.62]	9.52	< .001***
Left vlPFC	26	1315	1348	0.36	[0.30; 0.43]	11.71	< .001***
Left STC	20	983	1081	0.32	[0.24; 0.40]	8.00	< .001***
(b)							
Right dlPFC > Right vlPFC+STC	20	988	1085	0.09	[0.02; 0.15]	2.90	.009**
Bilateral dFPC > bilateral vFPC	23	1032	1153	0.13	[0.02; 0.23]	2.51	.020*
Left dlPFC > left vlPFC+STC	21	1000	1097	0.11	[0.05; 0.17]	4.12	< .001***
Left dlPFC > right dlPFC	20	988	1085	0.02	[-0.02; 0.06]	1.07	.30
Left vlPFC > Right vlPFC	23	1256	1256	0.01	[-0.05; 0.07]	0.35	.73
Left STC > right STC[#]	18	971	1069	0.09	[0.02; 0.16]	2.61	.018*

Note. One ([#]) or three ([%]) outliers were objectively identified and excluded from some meta-analyses. Asterisks indicate the significance levels of the pooled effects. *$p < .05$, **$p < .01$, ***$p < .001$.

Figure Legends

Figure 1. The novel meta-analytic approach to synthesize the published findings of changes in oxyhemoglobin concentration (HbO) during verbal fluency test performance. (a) The virtual registration and (b) segmentation of channels based on the literature using the Hitachi ETG-4000 system. (c) Example of a study that shows significant reductions in HbO increases in patients compared with healthy controls for 50 out of 52 channels. The yellow marks indicate the five 10–20 locations (from the left to right of figure: T4, Fp2, Fpz, Fp1, T3). Figure (a) was taken from Tsujii et al. (2017) and reused under the Creative Commons Attribution (CC BY) license.

Figure 2. Flow of the literature search.

Figure 3. Publication trends. The (a) annual and (b) cumulative numbers of publications that involved the combined use of functional near-infrared spectroscopy and the verbal fluency test for examining psychiatric disorders.

Figure 4. Proportions of channels or channel clusters that exhibited significant reductions in oxyhemoglobin concentration increases in major depressive disorder (MDD) and schizophrenia (SCZ) patients, relative to healthy controls, during verbal fluency test performance (i.e., significant-channel-based meta-analyses reported in Section 3.3.2). Asterisks indicate the significance levels of (a) one-sample sign tests (proportion ≠ 0) and (b) paired-sample sign tests (proportion in one region ≠ proportion in another region). *p < .05, **p < .01, ***p < .001

Figure 1.

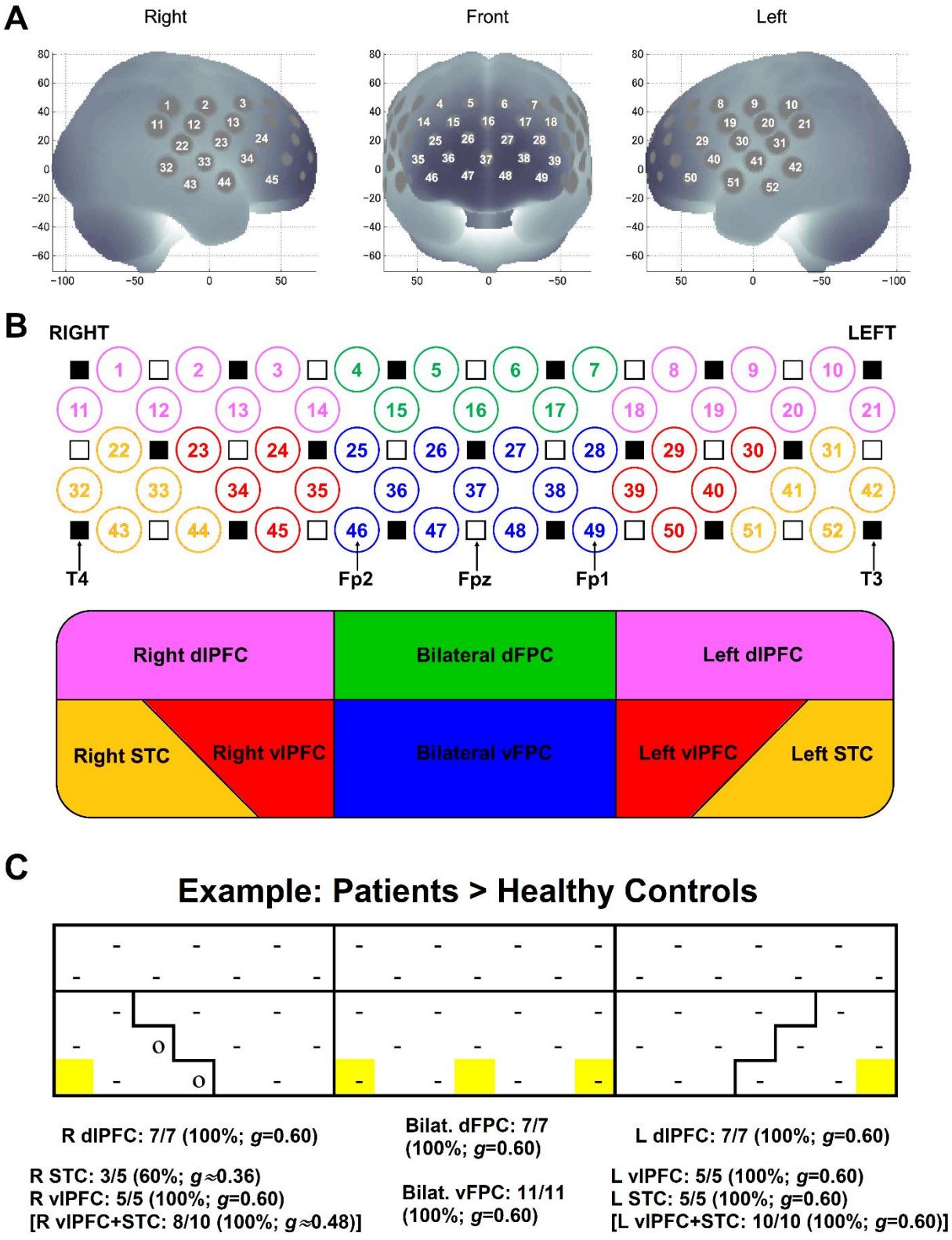

A
Right
Front
Left
B
RIGHT
LEFT
T4
Fp2
Fpz
Fp1
T3
Right dlPFC
Bilateral dFPC
Left dlPFC
Right STC
Right vlPFC
Bilateral vFPC
Left vlPFC
Left STC
C
Example: Patients > Healthy Controls
R dlPFC: 7/7 (100%; g=0.60)
Bilat. dFPC: 7/7 (100%; g=0.60)
L dlPFC: 7/7 (100%; g=0.60)
R STC: 3/5 (60%; g≈0.36)
R vlPFC: 5/5 (100%; g=0.60)
[R vlPFC+STC: 8/10 (100%; g≈0.48)]
Bilat. vFPC: 11/11 (100%; g=0.60)
L vlPFC: 5/5 (100%; g=0.60)
L STC: 5/5 (100%; g=0.60)
[L vlPFC+STC: 10/10 (100%; g=0.60)]

Figure 2.

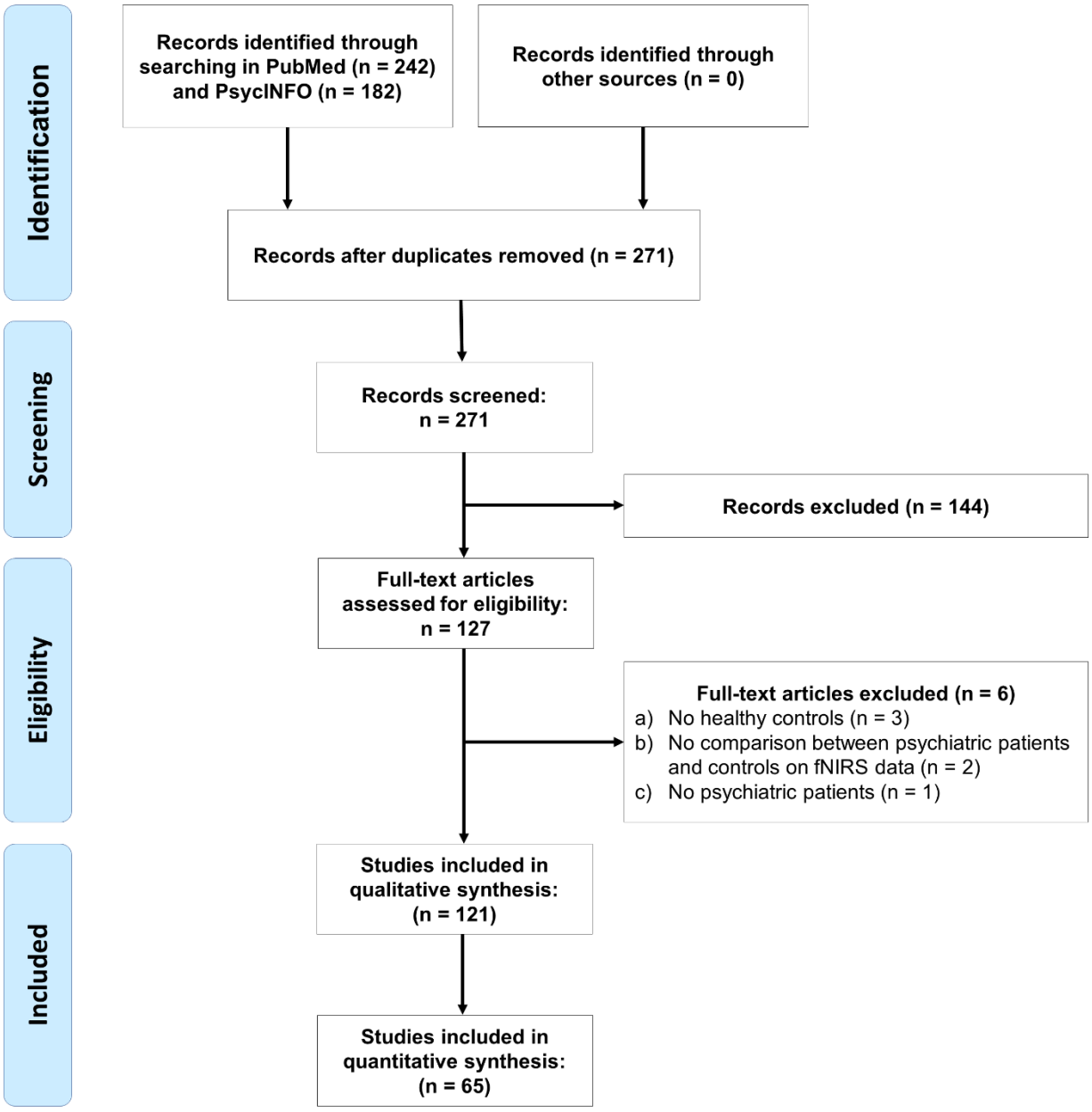

Identification
Screening
Eligibility
Included
Records identified through searching in PubMed (n = 242) and PsycINFO (n = 182)
Records identified through other sources (n = 0)
Records after duplicates removed (n = 271)
Records screened: n = 271
Records excluded (n = 144)
Full-text articles assessed for eligibility: n = 127
Full-text articles excluded (n = 6)
a) No healthy controls (n = 3)
b) No comparison between psychiatric patients and controls on fNIRS data (n = 2)
c) No psychiatric patients (n = 1)
Studies included in qualitative synthesis: (n = 121)
Studies included in quantitative synthesis: (n = 65)

Figure 3.

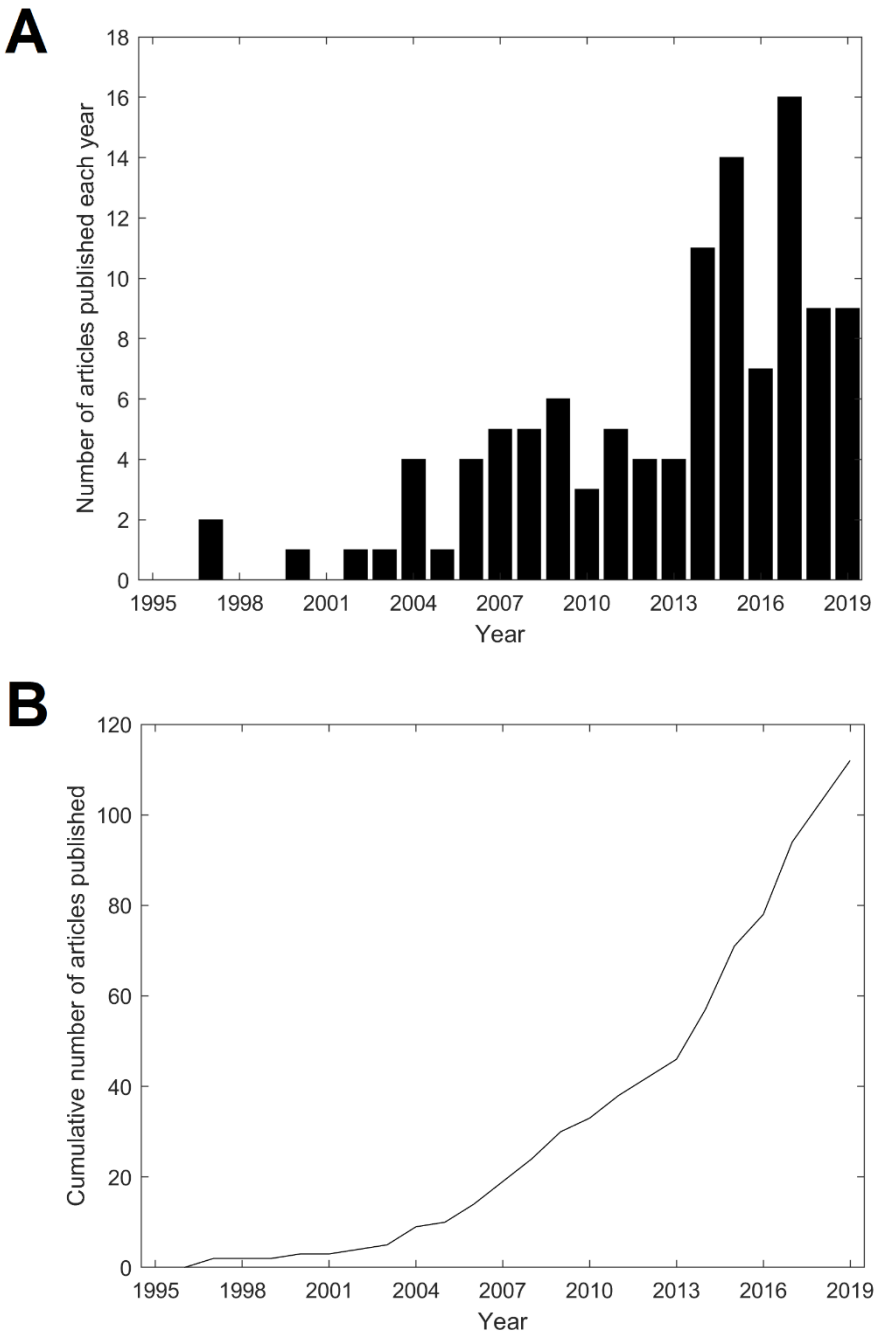
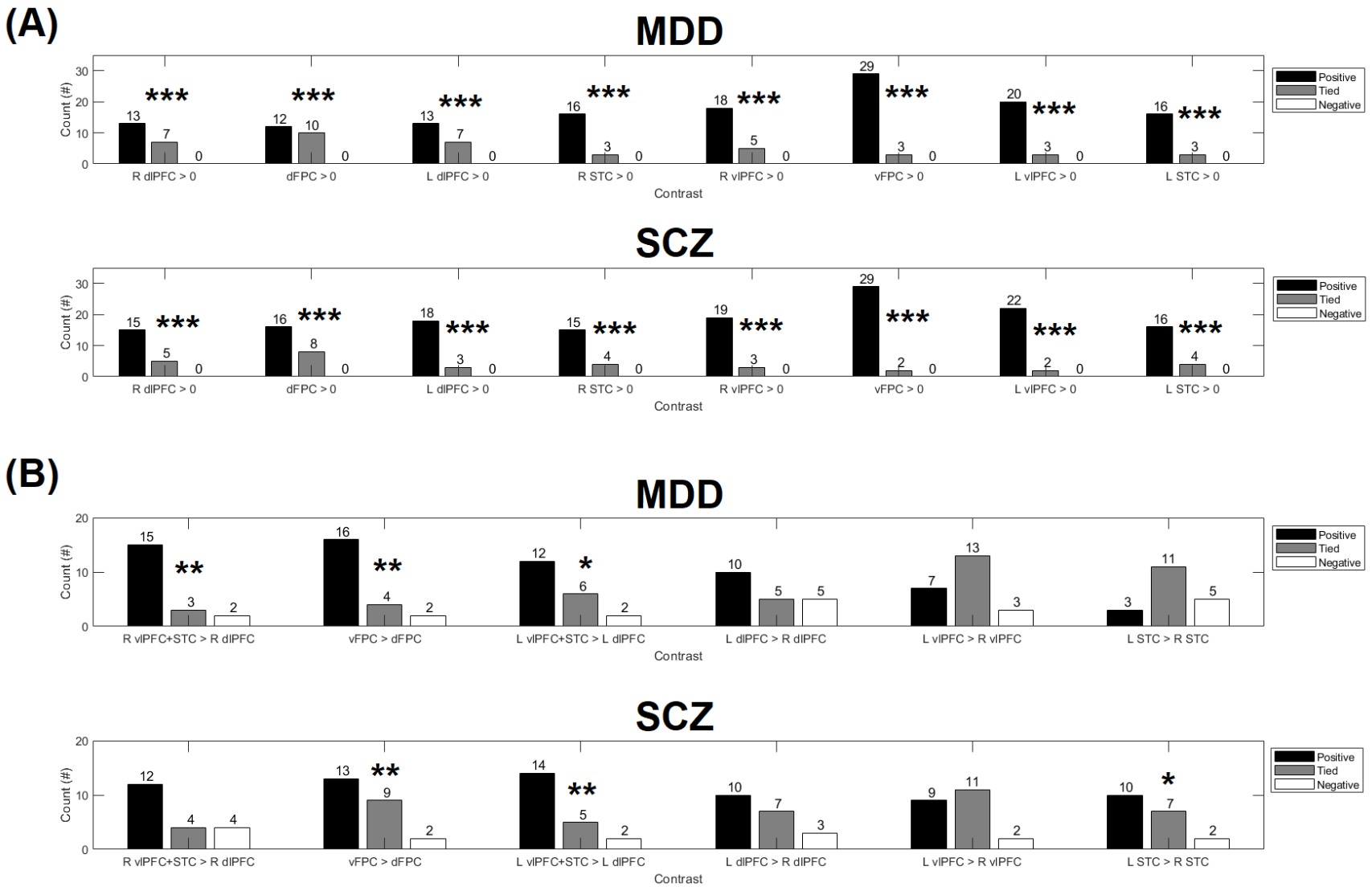

A
Number of articles published each year
0
2
4
6
8
10
12
14
16
18
1995
1998
2001
2004
2007
2010
2013
2016
2019
Year
B
Cumulative number of articles published
0
20
40
60
80
100
120
1995
1998
2001
2004
2007
2010
2013
2016
2019
Year

Figure 4.


(A)
MDD
Count (#)
Contrast
R dlPFC > 0
dFPC > 0
L dlPFC > 0
R STC > 0
R vlPFC > 0
vFPC > 0
L vlPFC > 0
L STC > 0
Positive
Tied
Negative
SCZ
(B)
MDD
R vlPFC+STC > R dlPFC
vFPC > dFPC
L vlPFC+STC > L dlPFC
L dlPFC > R dlPFC
L vlPFC > R vlPFC
L STC > R STC
SCZ