frontiers in Neuroscience

SYSTEMATIC REVIEW
published: 30 April 2021
doi: 10.3389/fnins.2021.664740

Auricular Vagus Neuromodulation—A Systematic Review on Quality of Evidence and Clinical Effects

Nishant Verma[1,2], *Jonah D. Mudge*[1,2], *Maïsha Kasole*[1,2], *Rex C. Chen*[1,2], *Stephan L. Blanz*[1,2], *James K. Trevathan*[1,2], *Eric G. Lovett*[3], *Justin C. Williams*[1,2,4] *and Kip A. Ludwig*[1,2,4]*

[1] Department of Biomedical Engineering, University of Wisconsin – Madison, Madison, WI, United States, [2] Wisconsin Institute for Translational Neuroengineering (WITNe) – Madison, Madison, WI, United States, [3] LivaNova PLC, Minneapolis, MN, United States, [4] Department of Neurosurgery, University of Wisconsin – Madison, Madison, WI, United States

OPEN ACCESS

Edited by:
Arun Sridhar, Independent Researcher, Milton Keynes, United Kingdom

Reviewed by:
Stavros Zanos, Feinstein Institute for Medical Research, United States
Wouter De Jonge, Academic Medical Center, Netherlands
Matteo Donega, Consultant, Village-Neuf, France

***Correspondence:**
Kip A. Ludwig
kip.ludwig@wisc.edu

Specialty section:
This article was submitted to Autonomic Neuroscience, a section of the journal Frontiers in Neuroscience

Received: *05 February 2021*
Accepted: *25 March 2021*
Published: *30 April 2021*

Citation:
Verma N, Mudge JD, Kasole M, Chen RC, Blanz SL, Trevathan JK, Lovett EG, Williams JC and Ludwig KA (2021) Auricular Vagus Neuromodulation—A Systematic Review on Quality of Evidence and Clinical Effects. Front. Neurosci. 15:664740. doi: 10.3389/fnins.2021.664740

Background: The auricular branch of the vagus nerve runs superficially, which makes it a favorable target for non-invasive stimulation techniques to modulate vagal activity. For this reason, there have been many early-stage clinical trials on a diverse range of conditions. These trials often report conflicting results for the same indication.

Methods: Using the Cochrane Risk of Bias tool we conducted a systematic review of auricular vagus nerve stimulation (aVNS) randomized controlled trials (RCTs) to identify the factors that led to these conflicting results. The majority of aVNS studies were assessed as having "some" or "high" risk of bias, which makes it difficult to interpret their results in a broader context.

Results: There is evidence of a modest decrease in heart rate during higher stimulation dosages, sometimes at above the level of sensory discomfort. Findings on heart rate variability conflict between studies and are hindered by trial design, including inappropriate washout periods, and multiple methods used to quantify heart rate variability. There is early-stage evidence to suggest aVNS may reduce circulating levels and endotoxin-induced levels of inflammatory markers. Studies on epilepsy reached primary endpoints similar to previous RCTs testing implantable vagus nerve stimulation therapy. Preliminary evidence shows that aVNS ameliorated pathological pain but not evoked pain.

Discussion: Based on results of the Cochrane analysis we list common improvements for the reporting of results, which can be implemented immediately to improve the quality of evidence. In the long term, existing data from aVNS studies and salient lessons from drug development highlight the need for direct measures of local neural target engagement. Direct measures of neural activity around the electrode will provide data for the optimization of electrode design, placement, and stimulation waveform parameters to improve on-target engagement and minimize off-target activation. Furthermore, direct measures of target engagement, along with consistent evaluation of blinding success,

must be used to improve the design of controls—a major source of concern identified in the Cochrane analysis. The need for direct measures of neural target engagement and consistent evaluation of blinding success is applicable to the development of other paresthesia-inducing neuromodulation therapies and their control designs.

Keywords: auricular stimulation, systematic review, vagus nerve stimulation or VNS, auricular vagus nerve stimulation, transcutaneous vagus nerve stimulation, microneurography, target engagement, blinding (masking)

INTRODUCTION

Electrical stimulation of the nervous system, commonly known as neuromodulation, manipulates nervous system activity for therapeutic benefits. The wandering path of the vagus nerve, the tenth cranial nerve, and its communication with several visceral organs and brain structures makes it an attractive target to address many diseases. Vagus nerve stimulation (VNS) to treat epilepsy has been approved by the United States Food and Drug Administration (FDA) since 1997 (Wellmark, 2018). An implantable pulse generator (IPG) is implanted below the clavicle and delivers controlled doses of electrical stimulation through electrodes wrapped around the cervical vagus. Due to the safety vs. efficacy profile of the therapy, implantable VNS is currently a last line therapy after patients have been shown refractory to at least two appropriately dosed anti-epileptic drugs (American Association of Neurological Surgeons, 2021). Implantable VNS for epilepsy is purported to work through vagal afferents terminating in the nucleus of the solitary tract (NTS). NTS in turn has direct or indirect projections to the nuclei providing noradrenergic, endorphinergic, and serotonergic fibers to different parts of the brain (Kaniusas et al., 2019).

In a similar fashion, the auricular branch of the vagus also projects to the NTS, carrying somatosensory signals from the ear (Kaniusas et al., 2019). The superficial path of the nerve (Bermejo et al., 2017) in the ear means a low amplitude electrical stimulation applied at the surface of the skin can, in theory, generate electric field gradients at the depth of the nerve sufficient to alter its activity. Auricular vagus nerve stimulation (aVNS) delivered percutaneously or transcutaneously offers a method to modulate neural activity on the vagus nerve with the potential for a more favorable safety profile. **Figure 1** shows innervation of the auricle by four major nerve branches, overlapping regions of innervation in the auricle, and several electrode designs to deliver electrical stimulation at the ear.

Given aVNS can be implemented with minimally invasive approaches and has the potential to modulate vagal activity, there have been many early-stage clinical trials investigating a diverse range of potential therapeutic indications, including heart failure, epilepsy, depression, pre-diabetes, Parkinson's, and rheumatoid arthritis. Several companies are already developing aVNS devices, such as Parasym (London, UK), Cerbomed (Erlangen, Germany), Spark Biomedical (Dallas, Texas, USA), SzeleSTIM (Vienna, Austria), Ducest Medical (Ducest, Mattersburg, Germany), Innovative Health Solutions (Versailles, IN, USA), and Hwato (Suzhou, Jiangsu Province, China). Despite the large number of aVNS clinical studies, clinical evidence to support a specific therapeutic outcome is often mixed, with conflicting trial results for the same physiological outcome measure (Burger et al., 2020; Keute et al., 2021).

According to the Oxford Center for Evidence Based Medicine's (CEBM) Levels of Clinical Evidence Scale, the highest level of clinical evidence is a systematic review of multiple high-quality double-blinded, randomized, and controlled clinical trials (RCTs) with narrow confidence intervals, each homogeneously supporting the efficacy and safety of a therapy for a specific clinical outcome (Centre for Evidence-Based Medicine, 2009). However, reaching this level of evidence is costly and time-consuming. Years of precursor clinical studies with fewer number of subjects are needed to identify the most efficacious embodiment of the therapy that can be safely delivered. Data from these precursor studies are required to design more definitive clinical studies. The field of aVNS, being relatively new clinically, is understandably still in these early phases of clinical development.

We performed a systematic review of aVNS RCTs with two primary goals: (1) to provide an accessible framework for the aVNS community to review current studies for specific outcome measures as a resource to inform future study design and (2) to perform a qualitative assessment of the current level of clinical evidence to support aVNS efficacy for the most common outcome measures reported. To this end, the Cochrane Risk of Bias Tool (Sterne et al., 2019)—a framework previously used to identify risk of bias in RCT studies of epidural spinal cord stimulation (Duarte et al., 2020b) and dorsal root ganglion stimulation (Deer et al., 2020) to treat pain—was first used to assess the quality of evidence in individual aVNS RCTs. These data were aggregated to broadly assess the current level of clinical evidence, according to the Oxford CEBM scale, to support aVNS efficacy across the common physiological outcomes. Our efforts were not intended to provide a precise assessment of the current level of clinical evidence but to identify the most common gaps in clinical study design and reporting. These gaps were analyzed to identify systematic next steps that should be addressed before aVNS can move to a higher level of evidence for any specific clinical outcome.

METHODS

Search Method

Our literature search was designed to identify reports of clinical RCTs testing aVNS as an intervention. Two databases were searched systematically: PubMed and Scopus (includes MEDLINE and Embase databases). Additionally, two search strategies were used. The first strategy combined search terms

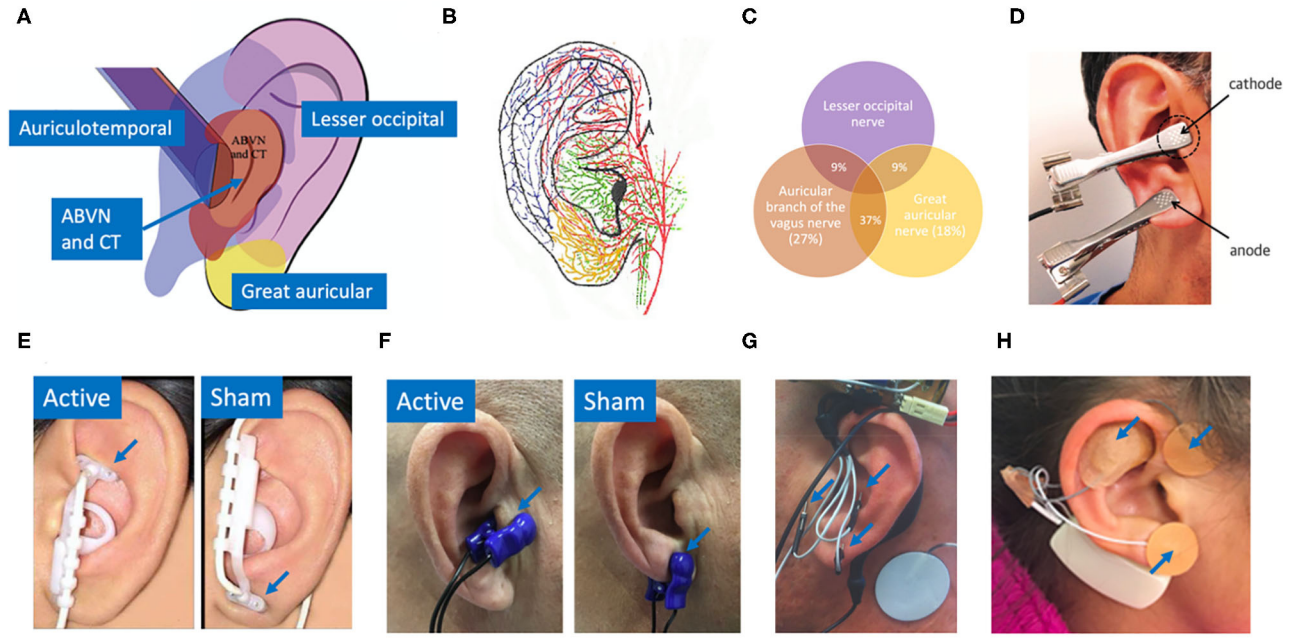

FIGURE 1 | (A) Innervation of the auricle by five nerves (Watanabe et al., 2016): auricular branch of the vagus nerve (ABVN), chorda tympani (CT) from the facial nerve, auriculotemporal nerve originating from the mandibular branch of the trigeminal nerve, great auricular nerve, and lesser occipital nerve. **(B)** Artist impression of auricular innervation (He et al., 2012). Refer to Peuker and Filler (2002) dissection study mapping the innervation of the human auricle performed in 7 cadavers for original photographs. Note, microdissection cannot trace the finest of nerve branches. **(C)** Overlapping regions of innervation reported between the auricular branch of the vagus nerve (ABVN), great auricular nerve (GAN), and lesser occipital nerve (Peuker and Filler, 2002). Here the ABVN and GAN overlap for 37% of the area on the medial dorsal middle third of the ear—setting the precedent for large overlaps in regions innervated by different nerves. Commonly used electrodes: **(D)** Clips transcutaneously targeting the tragus and earlobe simultaneously in the intervention group (Stavrakis et al., 2015). **(E)** NEMOS electrodes by Cerbomed (Erlangen, Germany) transcutaneously targeting cymba concha in the intervention group and ear lobe in the sham group (Frangos et al., 2015). **(F)** Parasym (London, UK) transcutaneously targeting tragus in the intervention group and ear lobe in the sham group (Stavrakis et al., 2020). **(G)** Percutaneously targeting intrinsic auricular muscles zones in the intervention group (Cakmak et al., 2017). **(H)** Innovative Health Solutions (Versailles, IN, USA) percutaneously targeting several cranial nerves in the auricular and periauricular region in the intervention group (Kovacic et al., 2017).

related to aVNS and RCT. The second search strategy focused on search terms related to commercial aVNS devices and their manufacturers. Complete search strings for both strategies are available in **Supplementary Material 1**. The search was last updated in July 2020. In addition, citations of all selected studies were searched to identify additional studies that met the inclusion criteria. The citations of relevant reviews (Murray et al., 2016; Yap et al., 2020) were also searched. Duplicate records were removed, and the remaining records were screened at a title and abstract level to check if a clinical RCT on auricular stimulation was reported.

As our primary goal was to assess the effects of auricular stimulation, studies using any stimulation modality from any field, including acupuncture and electroacupuncture were initially included as long as the intervention was at the auricle. When it became evident that a meta-analysis would not be possible due to incomplete reporting of information, we decided to exclude traditional Chinese medicine (TCM) studies, which typically used acupuncture, electroacupuncture, or acupuncture beads. Studies were considered TCM studies if acupoints were used to justify location of stimulation or if they were published in a TCM journal. This is captured in **Figure 2** adapted from PRISMA (Moher et al., 2009). All studies excluded at the end of the search after full-text review are listed in **Supplementary Material 2**.

Inclusion and Exclusion Criteria

In papers where more than one clinical trial was reported, each trial that was randomized and controlled was included in the systematic review; non-RCT portions of included publications were not analyzed. Only publications 1991 and after were included, with the cutoff marking the first time autonomic activity biomarkers were reportedly measured during auricular stimulation (Johnson et al., 1991).

Included studies had to report measurements of direct clinical significance. This exclusion partially relied on whether the study claimed direct clinical implications of their findings. Additionally, studies were excluded if the measurements did not have a well-established link to clinically significant outcomes. For instance, pupil size, functional magnetic resonance imaging (fMRI), electroencephalography (EEG), and somatosensory evoked potentials (SSEPs) are secondary physiological measures of target engagement (Burger et al., 2020). Although they may be useful to study the mechanisms of aVNS, they do not have well-established links to clinically significant outcomes. In comparison, heart rate variability (HRV), a measure of

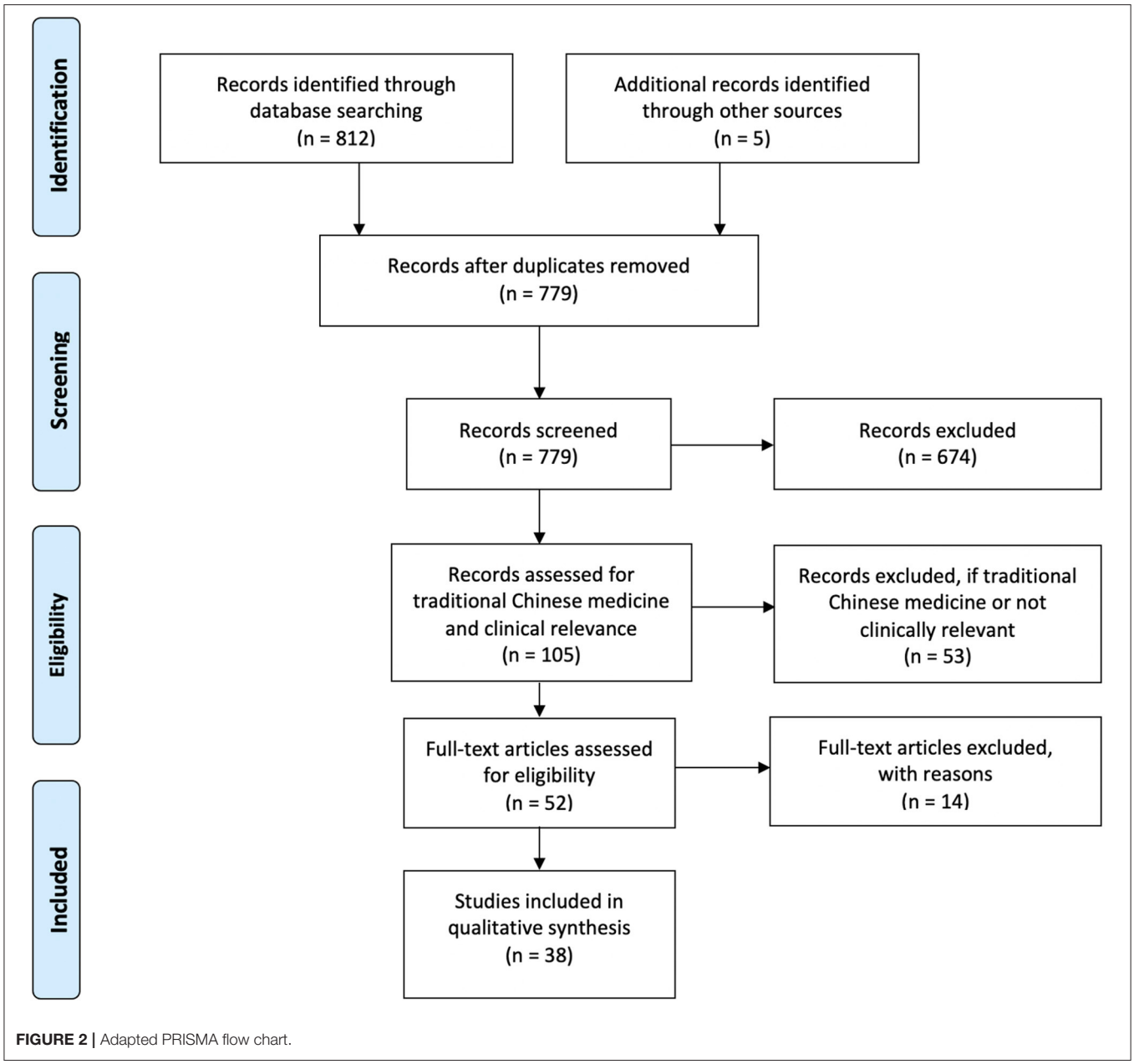

FIGURE 2 | Adapted PRISMA flow chart.

sympathovagal tone, is considered a measurement of direct clinical significance as sympathovagal imbalance is related to several disease states (Bootsma et al., 2003). Similarly, studies focused on cognitive neuroscience topics, such as behavior, learning, fear extinction, or executive functions were excluded. In contrast, psychological studies addressing addiction, depression, pain, and stress were included in the final qualitative review as they have had direct clinical significance.

Cochrane Risk of Bias 2.0 Tool to Assess Quality of Evidence

We used the Cochrane risk of bias 2.0 tool (RoB), an established tool to assess bias in clinical RCTs (Sterne et al., 2019), which has been cited over 40,000 times in Google Scholar, to evaluate the quality of evidence in each study. The RoB tool assesses bias in five subsections intended to capture the most common sources of possible bias in clinical studies. It is important to note a rating of "some" or "high" risk of bias does not mean that researchers conducting the study were themselves biased, or that the results they found are inaccurate. Deviations from ideal practice frequently occur due to a variety of potentially uncontrollable reasons. These deviations from the ideal just increase the chance that any stated result is a "false negative" or a "false positive" beyond the stated statistical convention used in the study.

For each study, the tool provides a suggested algorithm to rate bias through a series of guiding questions across the following five sections. Each section ends with a bias assignment of

"low," "some concerns," or "high." Template rubrics provided by Cochrane with the answers to these guiding questions have been included for every study evaluated in **Supplementary Material 9**. At several instances, the suggested algorithm was overridden by the reviewer with justification annotated on the individual rubric found in **Supplementary Material 9**. Below is an explanation of how each subsection was evaluated with respect to aVNS, see (Higgins et al., 2019) for more information on the recommended implementation of the Cochrane assessment tool.

Bias Arising From the Randomization Process

Randomization is important in a clinical study to ensure that differences in the outcome measure between the treatment and control groups were related to the intervention as opposed to an unintended difference between the two groups at baseline. To obtain a rating of "low" risk of bias, the study had to (1) randomize the allocation sequence, (2) conceal the randomized sequence from investigators and subjects till the point of assignment, and (3) test for baseline differences even after randomization. The latter is essential as even in a truly randomized design, it is conceivable that randomization yields an unequal distribution of a nuisance variable across the two groups. This is more likely to occur in studies with a smaller number of participants (Kang et al., 2008), which is the sample size found in many aVNS studies. Even in studies with a crossover design—meaning participants may receive a treatment and then, after an appropriate wash-out period, receive a sham therapy—it is important to test for baseline differences and that an equal number of subjects be presented with sham or therapy first (Nair, 2019).

Bias Due to Deviations From Intended Interventions

During the implementation of a clinical trial it is foreseeable for several subjects who were randomized to a given group to not receive the intended intervention or for blinding to be compromised. Compromised blinding is especially pertinent in neuromodulation studies, including aVNS studies, where there may be a difference in paresthesia or electrode location between the intervention and control group (Robbins and Lipton, 2017). Marked visual or perceptual differences between intervention groups can clue investigators and subjects to become aware of the treatment or control arm assignments, thereby violating the principle of blinding and deviating from the intended intervention. In order to receive a "low" risk of bias score, the study must have minimized and accounted for deviations from intended intervention due to unblinding, lack of adherence, or other failures in implementation of the intervention.

Bias Due to Missing Outcome Data

In conducting a clinical trial, being unable to record all intended outcome measures on all subjects is common. This can happen for a variety of reasons, including participant withdrawal from the study, difficulties in making a measurement on a given day, or records being lost or unavailable for other reasons (Higgins et al., 2019). In assessing how missing data may lead to bias it is important to consider the reasons for missing outcome data as well as the proportions of missing data. In general, if data was available for all, or nearly all participants, this measure was given "low" risk of bias. If there was notable missing data that was disproportionate between the treatment and control group, or the root cause for missing data suggested there may be a systemic issue, this measure was rated "some concerns" or "high" risk of bias depending on severity.

Bias in Measurement of the Outcome

How an outcome was measured can introduce several potential biases into subsequent analyses. Studies in which the assessor was blinded, the outcome measure was deemed appropriate, and the measurement of the outcome was performed consistently between intervention and control groups were generally considered "low" risk of bias. If the outcome assessor was not blinded but the outcome measure was justified as unlikely to be influenced by knowledge of intervention the study was also generally considered "low" risk of bias.

Bias in Selection of the Reported Result

An important aspect in reporting of clinical trial results is to differentiate if the data is exploratory or confirmatory (Hewitt et al., 2017). Exploratory research is used to generate hypotheses and models for testing and often includes analyses that are done at least in part retrospectively and are therefore not conclusive. Exploratory research is intended to minimize false negatives but is more prone to false positives. Confirmatory research is intended to rigorously test the hypothesis and is designed to minimize false positives. An important aspect of confirmatory research is pre-registration of the clinical trial before execution, including outlining the hypothesis to be studied, the data to be collected, and the analysis methods to be used. This is necessary to ensure that the investigators did not (1) collect data at multiple timepoints but report only some of the data, (2) use several analysis methods on the raw data in search for statistical significance, or (3) evaluate multiple endpoints without appropriate correction for multiple comparisons. Each of these common analysis errors violates the framework by which certain statistical methods are intended to be conducted—introducing an additional chance of yielding a false positive result. Studies that pre-registered their primary outcomes and used the measurements and analyses outlined in pre-registration generally scored "low" risk of bias in this category.

Information Extraction

Each paper was read in its entirety and a summary table was completed capturing study motivation, study design, study results, and critical review. Study motivation outlined the study hypothesis and hypothesized therapeutic mechanism of action if mentioned in the paper. We also noted whether implantable VNS had achieved the hypothesized effect in humans. Study design encapsulated subject enrollment information (diseased or healthy, the power of the study, and the inclusion and exclusion criteria), type of control and blinding, group design (crossover or parallel), stimulation parameters, randomization, baseline comparison, and washout periods. Lastly, study results included primary and secondary endpoints, adverse effects,

excluded and missing data, and statistical analysis details (pre-registered, handling of missing and incomplete data, multiple group comparison, etc.). Study results also analyzed if the effect was due to a few responders or improvements across the broad group, worsening of any subjects, control group effect size, and clinical relevance and significance of findings. Where sufficient information was reported, standardized effect size was calculated using Hedges' g (Turner and Bernard, 2006).

Each publication had a primary reviewer, and an additional secondary reviewer went through all papers. Any concerns raised by either reviewer were discussed as a group. If crucial basic information (e.g., which ear was stimulated, electrode used, etc.) was not reported (NR), an attempt was made to reach out to the author and if unsuccessful, to infer the information from similar studies by the group. Inferred or requested information is annotated as such. This effort helped highlight incomplete reporting of work while maximizing available information for the review to conduct an informed analysis.

A sortable table summarizing the design and result features of every reviewed study has been included as an excel file in **Supplementary Material 3** to allow viewing based on specific features of interest. Design and result features have been reduced to common keywords in this spreadsheet to facilitate sorting; however, this means specific details of outcome measures have been reduced to general categories in some cases.

RESULTS

A total of 38 articles were reviewed totaling 41 RCTs—two each in the publications by Hein et al. (2013), Cakmak et al. (2017), and Badran et al. (2018). In an initial review of the RCTs, it was apparent that a wide variety of electrode designs, stimulation parameters, study methodologies, and clinical indications were tested. As a framework by which to organize this multifaceted problem in the results below, we first discuss the electrode designs and stimulation parameters used across aVNS studies with the goal of identifying the most common aVNS implementation strategies and rationale for selection. Next, we discuss the study design features across all studies, again with the goal of identifying the most common practices. We then provide an assessment of all studies, regardless of clinical indication, using the Cochrane risk of bias (RoB) tool. Finally, we discuss the commonly measured outcomes based on treatment indication to identify which findings were most consistent across studies—couching the synthesis in results from the RoB analysis.

aVNS Electrode Designs, Configurations, and Stimulation Parameters Across Studies

Upon initial review, it was immediately evident that implementation of both active and sham varied greatly across studies. **Table 1** details the electrode design, configuration (monopolar or bipolar), target location, and stimulation parameters for the active and control arms of the study.

Table 1 is organized by indication type, then primary endpoints, then RoB score. Data is organized in the following columns:

Primary Endpoints

The main result of clinical interest. Studies are grouped by endpoints measured within their respective indications. For example, within the cardiac diseases indication the studies investigating inflammatory cytokine levels are located adjacent to each other.

Active Waveform and Location

Frequency, pulse width (PW), on/off cycle duration (duty cycle), and stimulation location.

Active Amplitude and Electrode Type

Current amplitude, titration method used to reach that amplitude, and electrode type and stimulator model when available. Titration methods are denoted as sub-sensory, first sensory, strong sensory (not painful), painful, or set at a particular amplitude. These terms reflect the cues that investigators used (Badran et al., 2019) to determine the stimulation amplitude for each subject:

Sub-sensory Titration

Stimulation was kept just below the threshold of paresthesia sensation.

First Sensory Titration

Subject is barely able to feel a cutaneous sensation.

Strong Sensory Titration

Subject feels a strong, but not painful or uncomfortable sensation from the stimulation.

Pain Titration

Stimulation amplitude is increased until the subject feels a painful or uncomfortable sensation.

Set Stimulation

Fixed amplitude across all subjects—resulting in different levels of sensation due to the individual's unique anatomy and perception.

Control

Control group stimulation amplitude, control design (sham or placebo), and stimulation location. Following Duarte et al. (2020b), we defined sham as when the control group experience from the subject perspective is identical to the active group experience—including paresthesia and device operating behavior. Conversely, placebo control is defined when the control group subjects do not experience the same paresthesia, device operation, or clinician interaction as the active group subjects.

Table 1 details the electrode design, configuration (monopolar or bipolar), target location, and stimulation parameters for the active and control arms of the study and shows that implementation of both active and sham varied greatly across studies. **Figure 3** shows a box plot presenting the distribution of pulse widths, stimulation current amplitudes, and frequencies

TABLE 1 | Electrode design, configuration, target location, and stimulation parameters (sorted by indication type, primary endpoints, and color coded RoB score).

Author (y) indication*	Primary endpoints	Active waveform** and location	Active amplitude and electrode type***	Control
		Cardiac		
Andreas et al. (2019) Postoperative atrial fibrillation	Postoperative atrial fibrillation assessed on ECG	1 Hz, PW NR, 40 min on 20 min off Side NR Triangular fossa	Sub-sensory (1 mA), Ducest Neurostimulator V	Sham: same location, low level stimulation (mA NR)
Stavrakis et al. (2015) Atrial fibrillation	Atrial fibrillation cycle length and duration, TNF-a, CRP	20 Hz, 1 ms PW, on/off cycle NR Right tragus	50% of heart sinus rate slowing current threshold (mA NR), Grass S88 stimulator clip electrodes	Placebo: no current, same location
Stavrakis et al. (2020) Atrial Fibrillation	Atrial fibrillation burden assessed on ECG	20 Hz, 200 μs PW, on/off cycle NR Right tragus	Strong sensory (1 mA below mild pain threshold, mean 16.8 mA), Parasym clip electrodes	Sham: diff amplitude (mean 19.9 mA), diff location (right earlobe)
Badran et al. (2018) Heart rate	ΔHR during stimulation	9 waveforms (1, 10, 25 Hz) × (100, 200, 500 μs PW), on/off cycle NR Left tragus	2x sensory threshold (at 100 μs PW: 9.28 ± 2.56 mA, at 200 μs PW: 5.32 ± 1.60 mA, at 500 μs PW: 3 ± 0.93 mA), custom clip electrodes	Sham: At 100 μs PW: 6.57 ± 1.83 mA. At 200 μs PW: 3.64 ± 1.26 mA. At 500 μs PW: 1.97 ± 0.71 mA. diff location (earlobe)
Afanasiev et al. (2015) Coronary Insufficiency and LV dysfunction	Heart rate and 6 min walk distance	Frequency NR, PW NR, on/off cycle NR Side NR concha	Titration method NR (mean 0.05–0.15 mA), electrode NR	Placebo: same location (concha)
Tobaldini et al. (2019) Orthostatic stress	ΔHR, LF/HF, systolic arterial BP variance, RR interval pattern, respiratory rate	25 Hz, 200 ms PW (reported 200 ms in methods and 200 us in discussion), on/off cycle NR, phase NR Left cymba concha	First sensory (1–6 mA), NEMOS ball contact electrodes	Placebo: No stimulation, same location.
Fisher et al. (2018) Hypertension	Percentage decrease in median systolic blood pressure (SBP)	25 Hz, 15 ms PW, 1 s duration (gated to exhalation), biphasic Left cymba concha and beneath antihelix	Strong sensory (mA NR), stimulator NR, surface electrodes	Placebo: no current, same location
Stowell et al. (2019) Hypertension	Arterial blood pressure	2, 10, 25, or 100 Hz. 300 μs PW, 1 s on/off Left cymba concha	Strong sensory (mA NR) Urostim device, custom-built ergonomic electrodes	Placebo: no current, same location
Zamotrinsky et al. (2001) coronary artery disease	HR, BP, LV diastolic function, LV filling	3 Hz, 1,500 μs PW, on/off cycle NR Bilateral cavum concha	Titration method NR (0.2–1.25 mA), acupuncture needles	No intervention
Antonino et al. (2017) Baroreflex sensitivity	cBRS from systolic blood pressure and RR interval HRV (LF/HF)	30 Hz, 200 μs PW, on/off cycle NR Bilateral tragus	Sensory threshold (10–50 mA device range), ear clip electrodes	1 Placebo 1 Sham: same waveform (mA NR), diff location (bilateral earlobe)
Bretherton et al. (2019) HRV and baroreflex sensitivity	cBRS and HRV	30 Hz, 200 μs PW, on/off cycle NR Side NR inner & outer tragus	Sensory threshold (2–4 mA), custom TENS electrodes	Placebo: same location (inner and outer tragus)
Clancy et al. (2014) HRV and sympathetic activity	HRV (LF/HF)	30 Hz, 200 μs PW, continuous, Side NR inner and outer tragus	Sensory threshold (10–50 mA device range), V-TENS Plus with modified surface electrodes	Placebo: no current, same location
De Couck et al. (2017) HRV	ECG with HRV	25 Hz, 250 μs PW, on/off cycle NR Bilateral cymba concha	Strong sensory (mean ~0.7 mA), NEMOS ball contact electrodes	Placebo: no current, same location
Borges et al. (2019) Cardiac vagal activity	HRV	25 Hz, PW of 200–300 μs, 30 s on/off Left cymba concha	Set stimulation (1 mA) and strong sensory stimulation (2.5 ± 0.93 mA), NEMOS ball contact electrodes	Sham: same set stimulation, different strong sensory stimulation (2.76 ± 1.01 mA), diff location (earlobe)
Tran et al. (2019) LV Strain and autonomic tone	LV global longitudinal strain	20 Hz, 200 μs PW, on/off cycle NR Right tragus	Strong sensory (1 mA below pain threshold, mean 22.6 mA active), Parasym earclip electrodes	Sham: diff current (mean 21.8 mA), diff location (right earlobe)
Yu et al. (2017) Myocardial ischemia-reperfusion Injury	Ventricular premature beat incidence	20 Hz, 1 ms PW, 5 s on/off Right tragus	50% of heart sinus rate slowing threshold, clip electrodes (S20 stimulator, Jinjiang, Chengdu City, China)	Placebo: no current, same location (right tragus)
		Epilepsy		
Bauer et al. (2016) Epilepsy	Reduction in seizure frequency (per 28 days)	25 Hz, 250 μs PW, on/off cycle NR Left concha	Sensory threshold (0.50 ± 0.47 mA), NEMOS ball contact electrode	Sham: diff waveform 1 Hz and 1.02 ± 0.83 mA, same location (left concha)

(Continued)

TABLE 1 | Continued

Author (y) indication*	Primary endpoints	Active waveform** and location	Active amplitude and electrode type***	Control
Aihua et al. (2014) Epilepsy	Reduction in seizure frequency (per month)	20 Hz, 200 ms PW, on/off cycle NR Bilateral concha and external ear canal	Pain threshold (mA NR), electrode NR	Sham: same waveform (mA NR), diff location (bilateral earlobe)
Rong et al. (2014) Epilepsy	Seizure frequency (per 4 weeks)	20–30 Hz, ≤ 1 ms PW Ear side NR cymba concha and cavum concha	Set stim (1 mA), electrode with 3 carbon-impregnated silicone tips (Suzhou Medical Appliance Co. Ltd.)	Sham: same waveform, diff location; contacts at scapha and antihelical fold
		Pain		
Straube et al. (2015) Chronic migraine	Decrease in headache per 28 days	25 Hz, 250 μs PW, 30 s on/off Left concha	Strong sensory (mA NR), NEMOS ball contact electrodes	Sham: 1 Hz frequency, same location
Janner et al. (2018) Pain	Perceived pain intensity and temporal summation of pain	100 Hz, 200 μs PW, 0.01 s on/0.49 s off Bilateral cymba concha	Strong sensory (mA NR), custom earplug electrodes wrapped in NaCl-soaked wool	Sham: same waveform, diff location (earlobe). Placebo: same location, no current
Kovacic et al. (2017) GI pain	Change in max abdominal pain intensity and composite of Pain-Frequency-Severity-Duration scale	Alternating 1 Hz and 10 Hz every 2 s, 1 ms PW, 2 hrs on/off Side NR Earlobe, triangular fossa, ventral periauricular tragus	Sub-sensory (mA NR), 2 mm titanium percutaneous electrodes (monopolar)	Sham: no current (sub sensory like active), same location
Kutlu et al. (2020) Fibromyalgia	Visual analog sale, beck depression scale, beck anxiety scale, fibromyalgia impact questionnaire, short form-36 for life quality	10 Hz, <500 μs PW, biphasic asymmetrical Bilateral tragus and concha	First sensory (mA NR), custom designed surface electrodes	Only exercise (active group is exercise and aVNS)
Busch et al. (2013) Pain	Thermal, mechanical, and pressure pain thresholds	25 Hz, 250 μs PW, on/off cycle NR Left concha	First sensory (1.6 ± 1.5 mA), STV02 (Cerbomed) electrodes	Placebo: no current, same location
Juel et al. (2017) Pain and GI motility	ECG and PPG, mechanical pain threshold, cold pressor test, and drink test (ultrasound imaging)	30 Hz, 250 μs PW, continuous Left concha	Pain/uncomfortable (intensity increased to counteract habituation, device rated between 0.1 and 10 mA), NEMOS ball contact electrodes	Sham: same waveform, diff location (left earlobe)
Frokjaer et al. (2016) Gastroduodenal motility and pain threshold	ECG and PPG, mechanical pain threshold, cold pressor test, and drink test (ultrasound imaging)	30 Hz, 250 μs PW, continuous Left concha	Strong sensory (1.07 mA), NEMOS ball contact electrode	Sham: diff amplitude (mean 1.57 mA), diff location (left earlobe)
Napadow et al. (2012) Pain	Mechanical deep-tissue pain intensity rating, and temporal pain summation	30 Hz, 450 μs PW, 0.5 s on/off, gated to exhalation phase of respiration Left cymba concha and antihelix/cavum concha slope	Strong sensory (mA NR), modified press-tack electrodes (0.20 × 1.5 mm)	Sham: same waveform, diff location (left earlobe)
Johnson et al. (1991) Pain threshold and autonomic function	Electrical pain threshold and autonomic function	100 Hz pulses every 2.4 Hz, 10 ms PW, 10 ms on/off, Right concha	Strong sensory (mA NR), carbon rubber electrodes	Sham: same waveform, diff location (antitragus)
Laqua et al. (2014) Pain	Electrical pain threshold	Alternating between 2 and 10 Hz, 200 μs PW, on/off cycle NR Bilateral concha (anode) and mastoid (cathode)	Strong sensory (mA NR), silver EEG electrode at anode and PECG electrode at cathode	Placebo: subsensory stim, same location
		Psychological		
Hein et al. (2013) Depression	Hamilton depression rating scale, Beck's depression inventory	1.5 Hz, PW NR, on/off cycle NR Bilateral concha	Sub-sensory (0–600 μA device range in study 1 and 130 μA in study 2), TENS-2000 (study 1) and TENS-1000 (study 2) Auri-Stim Medical, Inc. with corresponding electrodes (4 contacts)	Placebo: no current, same location
Hasan et al. (2015) Schizophrenia	Positive and negative schizophrenia symptom scale	25 Hz, 30 s on/180 s off, 250 μs PW, phase NR Left outer ear canal	Strong sensory (mA NR), CM02 (Cerbomed) titan electrodes	Placebo: no current, same location

(Continued)

TABLE 1 | Continued

Author (y) indication*	Primary endpoints	Active waveform** and location	Active amplitude and electrode type***	Control
Burger et al. (2019) Negative thought occurrence	Number of negative thought intrusions	25 Hz, 250 μs PW, 30 s on/off Left cymba concha	Set (0.5 mA), NEMOS ball contact electrodes	Sham: same waveform, diff location (left earlobe)
		Others		
Addorisio et al. (2019) Rheumatoid arthritis	Endotoxin-induced IL-6, IL-1β, and TNF	∼160 Hz vibrations, Right cymba concha	NA (vibratory device)	Sham: same waveform diff location (right gastrocnemius)
Salama et al. (2020) Acute inflammatory response after lung lobectomy	CRP, IL6, IL10, IL-1B, IL-18, TNF-a	1 Hz, 200 μs PW, 40 min on 20 min off Side NR Triangular Fossa	First sensory (230 nA), ∼2 mm needles	No intervention
Huang et al. (2014) Impaired glucose tolerance	2-h plasma glucose levels	20 Hz, ≤ 1 ms PW, phase NR Side NR concha	Set (1 mA, intensity adjusted based on tolerance of subjects), electrodes similar to Rong et al. (2014)	Sham: same waveform, diff location (superior scapha)
Cakmak et al. (2017) Parkinson's	Motor examination (part III of the unified Parkinson's disease rating scale)	130 Hz, 100 μs PW, continuous, biphasic Ipsilateral ear to dominant motor symptoms in tragicus, antitragicus, and helicis minor muscles	Strong sensory (100–130 μA), percutaneous electrodes	Placebo: no current, 2 percutaneous electrodes in upper helix
Maharjan et al. (2018) Olfactory function	Odor threshold test and supra-threshold test	80 or 10 Hz, 180 μs PW, on/off cycle NR Left internal (concha) and external ear	Pain threshold (0.1–10 mA), electrode NR	Sham: same waveform, diff location (left earlobe)
Tutar et al. (2020) Tinnitus	Tinnitus handicap inventory and depression anxiety stress scales	200 Hz, 1 ms PW, on/off cycle NR Unilateral and bilateral cymba concha	First sensory (10–30 mA), silver electrodes (Provile TENS stimulator) (monopolar)	Placebo: same location

Risk of Bias score indicated by box color: red for "high," orange for "some concerns," and green for "low" risk of bias.
Monophasic if phase not otherwise specified.
Bipolar if electrode polarity not otherwise specified. Configuration considered monopolar only if the return electrode is distant enough not to activate the target region.
BP, Blood pressure; CRP, c-reactive protein; diff, different; ECG, electrocardiogram; GI, gastrointestinal; HR, heart rate; HRV, heart rate variability; IL, interleukin; LV, left ventricular; LF/HF, low frequency/high frequency (LF/HF); NR, not reported; PPG, photoplethysmogram; PW, pulse width; Cbrs, spontaneous cardiac baroreflex sensitivity; TENS, transcutaneous electrical nerve stimulation; TNF-a, tumor necrosis factor alpha. A full list of abbreviations is found in ***Supplementary Material 3****.*

used across studies. In the active arm, the interquartile range (IQR) of stimulation current amplitudes was 0.2–5 mA, pulse width was 200–500 μs, and frequency of stimulation was 10–26 Hz. It is notable that the commonly used aVNS waveform parameters are similar to the parameters typically used for stimulation of the cervical vagus at a pulse width of 250 μs and frequency of 20 Hz (LivaNova, 2017), which uses surgically implanted epineural cuff electrodes. Outside of the IQR, the spread of the parameters is wide.

The large variation in waveform parameters is indicative of the exploratory nature of aVNS studies and underscores the difficulties in comparisons across studies where similar indications use widely different parameters. The variations in pulse width and stimulation frequency are due to the range of values chosen by investigators. The variations in stimulation amplitude are more nuanced and discussed next.

While the large variation in pulse width and frequency parameters can be explained as choices made by investigators, the sources of the large variation in stimulation current amplitude is not as trivial. It is important to consider differences in electrode design, material, area, and stimulation polarity when comparing stimulation current amplitudes across studies. This is because electrode geometry and contact area have the potential to impact target engagement of underlying nerves (Poulsen et al., 2020). Furthermore, nerve activation is a function of current density at the stimulating electrode (Rattay, 1999), and it is not possible to accurately estimate current density without knowing electrode geometry. Another source of variability arises from the fact that different studies used different titration methods to determine stimulation current amplitude. In the studies reviewed, current amplitude was often calibrated to different levels of paresthesia perception. The level of paresthesia subjects feel is related to current density, which is once again related to stimulation current through electrode geometry.

In order to determine an optimal stimulation paradigm, target engagement must be thoroughly quantified with respect to the aforementioned variables. Direct measures of target engagement of the nerve branches exiting the auricle will further our understanding of optimal stimulation parameters (see section Long Term Solutions—Target Engagement on directly measuring local neural target engagement).

aVNS Trial Designs Across Studies

The way in which studies were designed also varied greatly. Of the 41 RCTs reviewed, 20 used a crossover design while 21 opted for a parallel design. In terms of control group design, 19 studies used

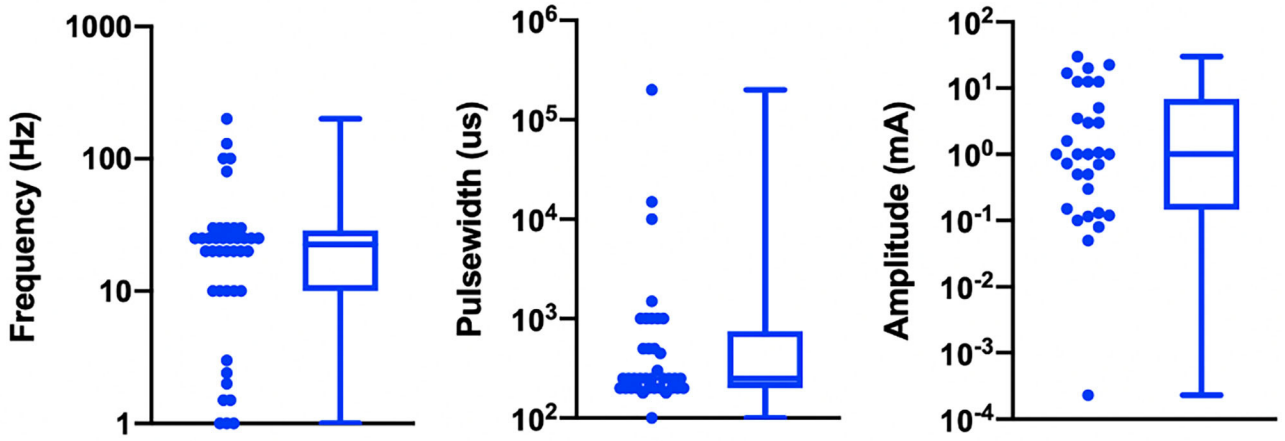

FIGURE 3 | Visualization of stimulation waveform parameters in 41 reviewed aVNS RCTs. Illustrated here are the interquartile ranges, maximums, minimums, and medians of the stimulation waveform parameters, including extreme cases. Median stimulation amplitude is 1.0 mA, pulse width is 250 us, and frequency is 22.5 Hz. Studies that report ranges for parameters are included as a single value representing the average of the boundaries of that range.

a sham, 17 a placebo, 2 used both sham and placebo, and 3 had no intervention as control. Studies also varied in duration: 12 were chronic and 29 were acute. The differences between these study design methods are important to emphasize and explored in section Long Term Solutions—Improvement in Control Design and Blinding.

Table 2 summarizes study designs and is organized by indication type, then primary endpoints, then RoB score. Data is organized in the following columns:

Primary Endpoints

The main result of clinical interest. Studies are grouped by endpoints measured within their respective indications. For example, within the cardiac diseases indication the studies investigating inflammatory cytokine levels are located adjacent to each other.

Subjects Analyzed

Sample size and whether subjects were healthy or diseased.

Control

Following Duarte et al. (2020b), we defined sham as when the control group experience from the subject perspective is identical to the active group experience—including paresthesia and device operating behavior. Conversely, placebo control is defined when the control group subjects do not experience the same paresthesia, device operation, or clinician interaction as the active group subjects.

Design

Study type (parallel or crossover), study time scale (acute or chronic), and intervention duration. Studies were classified as parallel if they randomized participants to intervention arms and each subject was assigned to only one intervention arm. Studies were classified as crossover if each subject group received every treatment but in a different order from the other subject groups (Nair, 2019). In some instances, the initial experimental group remained on the same intervention for the course of the study, while the control group was switched to the experimental intervention. These studies were classified as parallel, since not every subject received both interventions. Studies were classified as acute or chronic based on their duration being shorter or longer than 30 days, respectively.

Risk of Bias Tool to Assess Quality of Evidence

The Cochrane 2.0 Risk of Bias (RoB) assessment subscores for each study are summarized in **Table 3**. Explanations for each RoB subscore assignment [L = low (green), S = some concerns (orange), and H = high (red)] can be found generally explained in section Cochrane Risk of Bias 2.0 Tool to Assess Quality of Evidence and specifically explained for each study reviewed in **Supplementary Material 9**. The RoB tool provides a suggested algorithm to determine overall score based on the subscores of all sections. In several instances, the suggested algorithm was overridden by the reviewer with justification annotated on the individual rubric found in **Supplementary Material 9**.

Only two studies (Bauer et al., 2016; Maharjan et al., 2018) were assigned an overall "low" risk of bias. This is unsurprising, as the risk of bias assessment is rigorous and aVNS studies are in the less rigorous exploratory stages of investigation. Subsection and overall score percentages are illustrated in **Figure 4**.

The subsection "randomization process" was one of the best-scoring sections, in part because we assumed randomization was concealed from study investigators and subjects, even if the methodology to do so was not explicit. The studies that scored poorly in this section did not check for baseline imbalances between randomized groups or had baseline imbalances suggesting possible issues with the randomization method.

Notably, the "deviations from intended interventions" subsection tended to have the highest risk of bias. This

TABLE 2 | Study designs (sorted by indication type, primary endpoints, and color coded RoB score).

Author (y) indication*	Primary endpoints	Subjects analyzed, disease	Control	Design
Cardiac				
Andreas et al. (2019) Postoperative atrial fibrillation	Postoperative atrial fibrillation assessed on ECG	40, patients undergoing cardiac surgery	Sham	Parallel, acute (1 h)
Stavrakis et al. (2015) Atrial fibrillation	Atrial fibrillation cycle length and duration, TNF-a, CRP	40, paroxysmal atrial fibrillation ablation patients	Placebo	Parallel, acute (1 h)
Stavrakis et al. (2020) Atrial Fibrillation	Atrial fibrillation burden assessed on ECG	53, paroxysmal atrial fibrillation patients	Sham	Parallel, chronic (6 months)
Badran et al. (2018) Heart rate	ΔHR during stimulation.	35, healthy	Sham	Crossover, acute (duration NR)
Manasiev et al. (2015) Coronary insufficiency and LV dysfunction	Heart rate and 6 mins walk distance.	70, coronary insufficiency/LV dysfunction patients	Placebo	Parallel, acute (~1 h)
Tobaldini et al. (2013) Orthostatic stress	ΔHR, LF/HF, systolic arterial BP variance, RR interval pattern, respiratory rate	13, healthy	Placebo	Crossover, acute (~30 min)
Fisher et al. (2018) Hypertension	Percentage decrease in median systolic blood pressure (SBP)	10, hypertensive	Placebo	Crossover, acute (duration NR)
Stowell et al. (2019) Hypertension	Arterial blood pressure	12, diagnosed with primary hypertension	Placebo	Crossover, acute (5 days)
Zamotrinsky et al. (2001) Coronary artery disease	HR, BP, LV diastolic function, LV filling	18, stable angina pectoris class IV	No intervention	Parallel, acute (~10 days)
Antonino et al. (2017) Baroreflex sensitivity	cBRS from systolic blood pressure and RR interval HRV (LF/HF)	13, healthy	Both placebo and sham	Crossover, acute (duration NR)
Bretherton et al. (2019) HRV and baroreflex sensitivity	cBRS and HRV	14, healthy	Placebo	Crossover, acute (1 week)
Clancy et al. (2014) HRV and sympathetic activity	HRV (LF/HF)	48, healthy	Placebo	Parallel, acute (~15 min)
De Couck et al. (2017) HRV	ECG with HRV	30, healthy	Placebo	Crossover, acute (duration NR)
Borges et al. (2019) Cardiac vagal activity	HRV	60, healthy	Sham	Crossover, acute (duration NR)
Tran et al. (2019) LV strain and autonomic tone	LV global longitudinal strain	24, diagnosed with diastolic dysfunction by echocardiogram	Sham	Crossover, acute (duration NR)
Yu et al. (2017) Myocardial ischemia-reperfusion injury	Ventricular premature beat incidence	95, myocardial ischemia-reperfusion injury patients	Placebo	Parallel, acute (~2.5 h)
Epilepsy				
Bauer et al. (2016) Epilepsy	Reduction in seizure frequency (per 28 days)	58, epileptic	Sham	Parallel, chronic (20 weeks)
Aihua et al. (2014) Epilepsy	Reduction in seizure frequency (per month)	47, epileptic	Sham	Parallel, chronic (12 months)
Rong et al. (2014) Epilepsy	Seizure frequency (per 4 weeks)	144, epileptic	Sham	Parallel, chronic (24 weeks)
Pain				
Straube et al. (2015) Chronic migraine	Decrease in headache (per 28 days)	46, chronic migraine patients	Sham	Parallel, chronic (12 weeks)
Janner et al. (2018) Pain	Perceived pain intensity and temporal summation of pain	49, healthy	Both placebo and sham	Crossover, acute (8 days)
Kovacic et al. (2017) GI pain	Change in max abdominal pain intensity and composite of pain-frequency-severity-duration scale	104, children with GI pain	Sham	Parallel chronic (4 weeks)
Kutlu et al. (2020) Fibromyalgia	Visual analog sale, beck depression scale, beck anxiety scale, fibromyalgia impact questionnaire, short form-36 for life quality	52, fibromyalgia patients	No intervention	Parallel, chronic (4 weeks)
Busch et al. (2013) Pain	Thermal, mechanical, and pressure pain thresholds	48, healthy	Placebo	Crossover, acute (2 days)

(Continued)

TABLE 2 | Continued

Author (y) indication*	Primary endpoints	Subjects analyzed, disease	Control	Design
Juel et al. (2017) Pain and GI motility	ECG and PPG, mechanical pain threshold, cold pressor test, and drink test (ultrasound imaging)	20, chronic pancreatitis patients	Sham	Crossover, acute (7 days)
Frokjaer et al. (2016) Gastroduodenal motility and pain threshold	ECG and PPG, mechanical pain threshold, cold pressor test, and drink test (ultrasound imaging)	18, healthy	Sham	Crossover, acute (~6 days)
Napadow et al. (2012) Pain	Mechanical deep-tissue pain intensity rating, and temporal pain summation	15, chronic pelvic pain due to endometriosis patients	Sham	Crossover, acute (~1 week)
Johnson et al. (1991) Pain threshold and autonomic function	Electrical pain threshold and autonomic function	24, healthy	Sham	Parallel, acute (~15 min)
Laqua et al. (2014) Pain	Electrical pain threshold	21, healthy	Placebo	Crossover, acute (~1 week)
Psychological				
Hein et al. (2013) Depression	Hamilton Depression Rating Scale, Beck's Depression Inventory	37, majorly depressed	Placebo	Parallel, chronic (~2 weeks)
Hasan et al. (2015) Schizophrenia	Positive and negative schizophrenia symptom scale	17, schizophrenic	Placebo	Parallel, chronic (26 weeks)
Burger et al. (2019) Negative thought occurrence	Number of negative thought intrusions	97, high worriers	Sham	Parallel, acute (~20 min)
Others				
Addorisio et al. (2019) Rheumatoid arthritis	Endotoxin-induced interleukin (IL)-6, IL-1β, and TNF	19, healthy	Sham	Crossover, acute (duration NR)
Salama et al. (2020) Acute inflammatory response after lung lobectomy	CRP, IL6, IL10, IL-1B, IL-18, TNF-a	100, lobectomy via thoracotomy in patients with non-small cell lung cancer	No intervention	Parallel, acute (5 days)
Huang et al. (2014) Impaired glucose tolerance	2-hr plasma glucose levels	102, impaired glucose tolerance patients	Sham	Parallel, chronic (12 weeks)
Cakmak et al. (2017) Parkinson's	Motor symptoms	24, Parkinson's Hoehn and Yahr stage 2–3	Placebo	Crossover, acute (duration NR)
Maharjan et al. (2018) Olfactory function	Odor threshold test and supra-threshold test	18, healthy	Sham	Crossover, acute (7 days)
Tutar et al. (2020) Tinnitus	Tinnitus handicap inventory and depression anxiety stress scales	60, 20 per group, each with constant tinnitus >3-month duration	Placebo	Parallel, chronic (~1 month)

**Risk of Bias score indicated by box color: red for "high," orange for "some concerns," and green for "low" risk of bias.*
BP, Blood pressure; CRP, c-reactive protein; ECG, electrocardiogram; GI, gastrointestinal; HR, heart rate; HRV, heart rate variability; IL, interleukin; LV, left ventricular; LF/HF, low frequency/high frequency; NR, not reported; PPG, photoplethysmogram; cBRS, spontaneous cardiac baroreflex sensitivity; TNF-a, tumor necrosis factor alpha. A full list of abbreviations is found in ***Supplementary Material 3****.*

was mainly due to issues with potential subject unblinding as a result of easily perceptible differences between active and control groups. For example, in a crossover design, the subject experiences both the paresthesia-inducing active intervention and the non-paresthesia-inducing placebo control. The difference in paresthesia may unblind the subject to the identity of the active vs. control interventions.

The subsection "missing outcome data" was generally scored as "low" risk of bias across studies. Studies scoring "high" or "some concerns" had unreported outcome data with a non-trivial difference in the proportion of missing data between interventions.

The "measurement of the outcome" subsection also generally scored a "low" risk of bias across studies. In order to score "some concerns" or worse, outcome assessor blinding to subject intervention had to be compromised. For a "high" RoB score in this section, studies measured endpoints that could be influenced by investigator unblinding, such as investigator-assessed disease evaluation questionnaires. Empirical measurements such as heart rate and blood pressure were less susceptible to this kind of bias and hence scored better.

The subsection with the highest risk for bias ("high" or "some concerns" scores) was "selection of the reported results." This was primarily due to a lack of pre-registration in most studies. Suggestions to improve study reporting are listed in section Short Term Solutions—Guide to Reporting.

Summary of Outcome Measures Across Indications

The RoB analysis revealed potential for bias in the outcomes of the studies reviewed. Here, the studies are grouped by outcome measures—cardiac, inflammatory, epilepsy, and pain.

TABLE 3 | RoB overall and section scores (sorted by overall score).

Author (y)	Randomization process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported result	Overall risk of bias
Bauer et al. (2016)	L	L	L	L	L	L
Maharjan et al. (2018)	L	L	L	L	L	L
Addorisio et al. (2019)	L	S	L	L	S	S
Aihua et al. (2014)	S	H	L	S	S	S
Andreas et al. (2019)	L	S	L	L	L	S
Antonino et al. (2017)	L	L	S	L	S	S
Badran et al. (2018)	L	L	L	L	S	S
Bretherton et al. (2019)	L	L	L	L	S	S
Burger et al. (2019)	L	L	L	L	S	S
Busch et al. (2013)	L	S	L	L	S	S
Cakmak et al. (2017)	L	H	L	L	L	S
Clancy et al. (2014)	L	S	L	L	S	S
De Couck et al. (2017)	L	L	L	L	S	S
Hein et al. (2013)	L	L	L	S	S	S
Huang et al. (2014)	S	L	L	L	S	S
Janner et al. (2018)	L	L	L	L	S	S
Juel et al. (2017)	L	S	L	L	S	S
Kovacic et al. (2017)	L	S	L	L	L	S
Rong et al. (2014)	L	S	L	L	S	S
Stavrakis et al. (2015)	L	S	L	L	S	S
Stavrakis et al. (2020)	L	L	L	L	S	S
Tran et al. (2019)	L	S	L	L	S	S
Tutar et al. (2020)	L	S	L	S	S	S
Afanasiev et al. (2016)	H	S	H	L	S	H
Borges et al. (2019)	L	S	L	L	S	H
Fisher et al. (2018)	S	H	H	S	H	H
Frøkjaer et al. (2016)	S	S	L	L	S	H
Hasan et al. (2015)	L	H	L	L	S	H
Johnson et al. (1991)	S	H	L	S	S	H
Kutlu et al. (2020)	S	H	L	L	L	H
Laqua et al. (2014)	L	S	L	L	H	H
Napadow et al. (2012)	S	H	S	L	H	H
Salama et al. (2020)	L	H	L	S	S	H
Stowell et al. (2019)	L	S	L	L	S	H
Straube et al. (2015)	L	H	L	L	L	H
Tobaldini et al. (2019)	S	H	L	L	S	H
Yu et al. (2017)	L	H	L	L	S	H
Zamotrinsky et al. (2001)	L	S	L	S	S	H

L, low risk of bias (green); S, some concerns (yellow), and H, high risk of bias (red).
The Cochrane tools' suggested algorithms to determine subsection and overall scores did not always take into account certain caveats in study design. In such instances, the suggested algorithm is overridden, and a justification is provided in the study specific RoB rubric in ***Supplementary Material 9****.*

The outcome measures of the studies—regardless of indication—are qualitatively synthesized and couched in the findings of the RoB assessment. Findings from the RoB assessment are used to point out instances where trial design or reporting may have influenced interpretation of the trial outcomes. For reviews comprehensively synthesizing pre-clinical and both non-randomized and randomized clinical studies by indication, without emphasis on RoB, see Yap et al. (2020) and Jiang et al. (2020).

Cardiac Related Effects of aVNS

The most common cardiac effects assessed were changes to heart rate (HR) and sympathovagal balance. To measure sympathovagal balance, heart rate variability (HRV) was used.

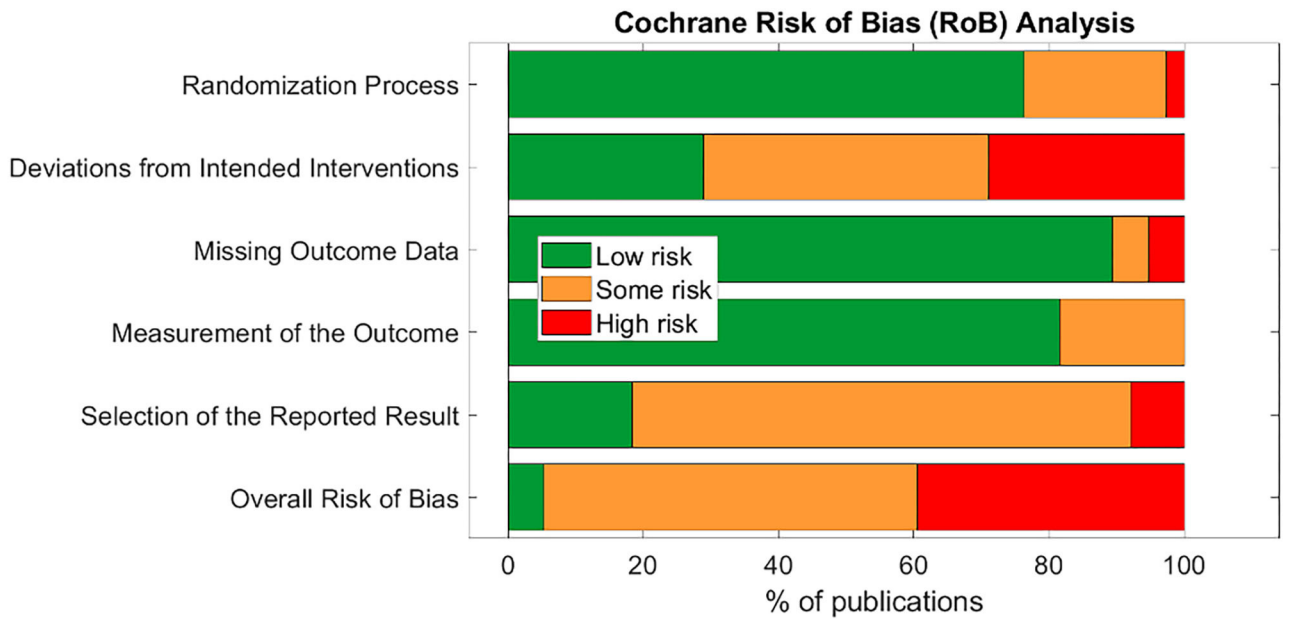

FIGURE 4 | Summary of Cochrane Risk of Bias overall and section scores for 38 publications reviewed.

Fifteen studies with cardiac, pain, or other indications reported HR or HRV measures and are summarized in **Table 4**. Results conflicted across studies for heart rate changes and sympathovagal balance but suggest aVNS may have an effect on both. However, there are concerns that these results may be attributed to trial design and inconsistent measurement methods.

Studies that reported a change in HR measured a modest mean drop of 2–3 beats per minute (BPM) in the active group. However, almost half of the 11 trials reporting HR effects reported no significant difference in effect between or within control and active stimulation. Stavrakis et al. (2015) and Yu et al. (2017) attained a consistent decrease in HR in every subject by increasing the stimulation amplitude until a decrease in HR was measured. They reported a mean stimulation threshold to elicit a HR decrease that was above the mean threshold for discomfort. In addition, Frøkjaer et al. (2016) and Juel et al. (2017) reported a significant decrease in HR during sham at the earlobe but not during active stimulation at the conchae and tragus. This observation suggests that decrease in HR may not be vagally mediated but perhaps mediated by trigeminal or cervical nerve branch afferents (see **Figure 1A**) or sympathetic efferents (Cakmak, 2019). See Cakmak (2019) for a comprehensive discussion on possible auricular stimulation pathways and mechanisms. Taken together, there is evidence for the effects of auricular stimulation on decreasing HR at high stimulation amplitudes, but it is uncertain if this decrease is mediated by the auricular branch of the vagus.

HRV, used as a measure of sympathovagal balance, was quantified inconsistently across studies and may not be an accurate indicator of whole body sympathovagal balance. Shown at the bottom of **Table 4** are multiple ways to analyze electrocardiogram (ECG) data for HRV. Different ways to quantify HRV enables multiple comparisons—which were often not appropriately corrected for—in search of statistical significance. HRV was calculated differently across studies making it difficult to uniformly draw conclusions across the aggregate of studies. Furthermore, HRV is not a measure of whole body sympathovagal tone, but of cardiac vagal activity—it relies on the physiological variance in HR with breathing. More variance in HR during breathing indicates more vagal control and a corresponding shift in cardiac sympathovagal balance to parasympathetic (Goldberger, 1999). Contradictory results on the parasympathetic effects of aVNS indicates that either the effects of aVNS on HRV are inconsistent, or that HRV is an unreliable measure of cardiac sympathovagal balance (Bootsma et al., 2003; Billman, 2013; Hayano and Yuda, 2019; Marmerstein et al., 2021), or both. Overall, the effects of aVNS on sympathovagal balance conflict between studies. Similarly, Wolf et al. (2021), in a meta-analysis pre-print, concluded that there was "no support for the hypothesis that HRV is a robust biomarker for acute [aVNS]."

Given that cardiac effects are closely related to a subject's comfort and stress levels, it is important to consider trial design influences such as subject familiarization. For example, a clinical trial visit could increase stress levels and blood pressure of the subjects (Wright et al., 2015) and mask any potential therapeutic effects of aVNS on blood pressure. Another example of subject familiarization is related to HRV. Borges et al. (2019) tried to accommodate for subject familiarization by delivering a 5 min "familiarization" stimulation at the cymba concha before the experimental intervention. Unfortunately, baseline HRV measurements were taken immediately after these

TABLE 4 | Summary of cardiac (heart rate and heart rate variability) studies.

Author (y) indication*	HR results	HRV results	Active stimulation level**	Active waveform	Subjects analyzed, disease
Antonino et al. (2017) Baroreflex sensitivity	Active: 2–3 BPM decrease Sham: not significant	Active: LF/HF parasympathetic Sham at earlobe: LF/HF parasympathetic Placebo: LF/HF sympathetic	First sensory	30 Hz, 200 μs PW	13, healthy
Badran et al. (2018) Heart rate	Active: ∼2.40 BPM decrease Sham: not significant	NA	2x sensory threshold	10 Hz, 500 μs PW	35, healthy
Bretherton et al. (2019) HRV and baroreflex sensitivity	NA	Active: LF/HF parasympathetic Placebo: not significant	Sensory threshold	30 Hz, 200 μs PW	14, healthy
Burger et al. (2019) Negative thought occurrence	NA	Active: not significant (RMSSD) Sham: not significant	Strong sensory	25 Hz, 250 μs PW	97, high worriers
Clancy et al. (2014) HRV and sympathetic activity	Active: HR decreased (data not reported) Placebo: not significant	Active: LF/HF parasympathetic Placebo: not significant	First sensory	30 Hz, 200 μs PW	48, healthy
De Couck et al. (2017) HRV	NA	Active: SDNN parasympathetic; no significant changes in LF/HF, LF, HF, RMSSD Placebo: SDNN parasympathetic; no significant changes in LF/HF, LF, HF, RMSSD	Strong sensory	25 Hz, 250 μs PW	30, healthy
Janner et al. (2018) Pain	No significant effects of aVNS on HR (data not reported)	NA	Strong sensory	100 Hz, 200 μs PW	49, healthy
Juel et al. (2017) Pain and GI motility	Active: not significant Sham: ∼2.8 BPM decrease, at earlobe	Active: RMSSD parasympathetic relative to sham, sympathetic relative to baseline Sham at earlobe: RMSSD sympathetic	Pain	30 Hz, 250 μs PW	20, chronic pancreatitis patients
Tran et al. (2019) LV strain and autonomic tone	No significant effects of aVNS on HR (data not reported)	Active: SDNN no change, RMSSD parasympathetic, pNN50 sympathetic, LF/HF parasympathetic Sham at earlobe: SDNN parasympathetic, RMSSD parasympathetic, pNN50 no change, LF/HF sympathetic	Strong sensory	20 Hz, 200 μs PW	24, diagnosed with diastolic dysfunction by ECG
Yu et al. (2017) Myocardial ischemia-reperfusion injury	HR decrease in every subject (values not reported)	NA	Sinus rate slowing threshold	20 Hz, 1 ms PW	95, myocardial ischemia-reperfusion injury patients
Borges et al. (2019) Cardiac vagal activity	NA	Active: RMSSD sympathetic between resting and second half stim Sham: not significant	Strong sensory	25 Hz, PW of 200– 300 μs	60, healthy
Frokjaer et al. (2016) Gastroduodenal motility and pain threshold	Active: not significant Sham: ∼2.2 BPM decrease, at earlobe	Active: RMSSD parasympathetic at 20 min and sympathetic at 25 min Sham at earlobe: RMSSD sympathetic at 20 min and parasympathetic at 25 min	Strong sensory	30 Hz, 250 μs PW	18, healthy
Johnson et al. (1991) Pain threshold and autonomic function	No significant effects of aVNS on HR (data not reported)	NA	Strong sensory	100 Hz pulses every 2.4 Hz, 10 ms PW	24, healthy
Laqua et al. (2014) Pain	No significant effects of aVNS on HR (data not reported)	NA	Strong sensory	Alternating between 2 and 10 Hz, 200 μs PW	21, healthy
Napadow et al. (2012) Pain	No significant effects of aVNS on HR (data not reported)	NA	Strong sensory	30 Hz, 450 μs PW	15, chronic pelvic pain patients

(Continued)

TABLE 4 | Continued

Author (y) indication*	HR results	HRV results	Active stimulation level**	Active waveform	Subjects analyzed, disease
HRV Measurements					
LF/HF	Ratio of low frequency over high frequency cardiac activity (frequency-domain measurement)				
RMSSD	Root mean square of successive differences between heartbeats (time-domain measurement)				
pNN50	The proportion of NN50 divided by the total number of NN (RR) intervals. NN50 is the number of times successive heartbeat intervals exceed 50ms (time-domain measurement)				
SDNN	Standard deviation of the NN (RR) intervals (time-domain measurement)				

*Risk of Bias score indicated by box color: red for "high," orange for "some concerns," and green for "low" risk of bias.
Refer to text above **Table 1 for definitions of all active stimulation levels (First sensory titration: Subject is barely able to feel a cutaneous sensation. Strong sensory titration: Subject feels a strong, but not painful or uncomfortable sensation from the stimulation. Pain titration: Stimulation amplitude is increased until the subject feels a painful or uncomfortable sensation). BP, Blood pressure; BPM, beats per minute; ECG, electrocardiogram; GI, gastrointestinal; HR, heart rate; HRV, heart rate variability; LV, left ventricular; NR, not reported; PPG, photoplethysmogram; PW, pulse width; cBRS, spontaneous cardiac baroreflex sensitivity. A full list of abbreviations is found in **Supplementary Material 3**.

"familiarization" stimulation sessions and could be affected by the stimulation delivered. Hence, there is no true baseline measurement—before any stimulation is applied—of HRV and casts doubt on the findings of the study, which claimed no significant effects of aVNS on HRV. In summary, conflicting results on the cardiac effects of aVNS could be attributed to trial design and measurement methods.

Inflammatory Related Effects of aVNS

Seven studies with cardiac or anti-inflammatory indications measured cytokine levels and are summarized in **Table 5**. Cytokine levels were either measured directly in drawn blood (circulating) or after an *in vitro* endotoxin-induced challenge on drawn blood. In the four studies that measured circulating cytokine levels, results are somewhat conflicting. Stavrakis et al. (2020) reported a significant decrease in tumor necrosis factor alpha (TNF-α) and no significant changes in IL-6, IL-1β, IL-10, and IL-17, consistent with subjects with moderate atrial fibrillation burden and not suffering from any inflammatory condition. TNF-α is one of the most abundant mediators in inflamed tissue and is present in the acute inflammatory response (Parameswaran and Patial, 2010). In subjects being treated for myocardial infarction, Yu et al. (2017) reported that the active group was significantly lower than the control group for all measured cytokine levels [TNF-α, IL-6, IL-1β, and high-mobility group-box 1 protein (HMGB1)]. Unlike Stavrakis et al. (2020) and Yu et al. (2017), Salama et al. (2020) reported no statistically significant change in TNF-α, along with lower levels of c-reactive protein (CRP) and IL-6, compared to control. CRP is also an acute phase protein whose release from the liver is stimulated by increased levels of IL-6 (Del Giudice and Gangestad, 2018). Lastly, Afanasiev et al. (2016) measured HSP60 and HSP70, which are heat shock proteins and responsible for preventing damage to proteins in response to stressors such as high temperature (Morimoto, 1993). Both HSP60 and HSP70 increased significantly in Afanasiev et al. (2016), indicating a potential anti-inflammatory effect.

The limited applicability of *in vitro* endotoxin-induced assays was discussed in Stoddard et al. (2010), Yang et al. (2011), and Thurm and Halsey (2005). Additionally, Broekman et al. (2015) provided an example where an *in vitro* assay was unsuccessful in identifying disease severity in patients with a quiescent autoimmune disorder. Nonetheless, we summarize the findings on anti-inflammatory effects of aVNS on *in vitro* endotoxin-induced assays. In Stavrakis et al. (2015), acute stimulation was delivered intraoperatively to subjects undergoing ablation treatment for atrial fibrillation. After 1 h of stimulation, there was a significant decrease in TNF-α and CRP levels in femoral vein draws. In Addorisio et al. (2019), vibrotactile stimulation was applied for only 2 min and showed a statistically significant decrease in endotoxin-induced cytokine levels of TNF-α, IL-6, and IL-1βin blood drawn 1 h after stimulation.

Overall, these studies provide evidence that aVNS may reduce circulating levels and endotoxin-induced levels of inflammatory markers and suggest a potential anti-inflammatory effect of aVNS. The clinical relevance of endotoxin-induced measures needs to be further explored and the implications of lowered circulating cytokine levels on disease burden needs to be further investigated in RCTs.

Epilepsy Related Effects of aVNS

The three studies that investigated the antiepileptic effects of aVNS were all chronic studies and are summarized in **Table 6**. All used stimulation frequencies between 20 and 30 Hz similar to implantable VNS (LivaNova, 2017). However, other stimulation parameters varied widely. All studies reported a 20–40% decrease in seizure frequency from baseline and showed significance from baseline after a few weeks to months of daily prescribed stimulation.

Based on these three chronic studies, there is some evidence to support the anti-epileptic effects of aVNS. The primary outcomes of these non-invasive interventions are comparable to that of implantable VNS in studies of similar duration and sample size (Ben-Menachem et al., 1994; Handforth et al., 1998). However, due to concerns over unblinding and weaker evidence in between group analysis compared to within group analysis, it is possible that the effects may be attributed to placebo. The concern that this effect is placebo is exacerbated by the fact that the

TABLE 5 | Summary of inflammatory studies.

Author (y) indication*	aVNS group cytokine level results (compared to control)	Measurement method	Subjects analyzed, disease
Addorisio et al. (2019) Rheumatoid arthritis	TNF-α decrease ($p < 0.05$)	Endotoxin-induced	19, healthy
	IL-6 decrease ($p < 0.001$)		
	IL-1β decrease ($p < 0.01$)		
Andreas et al. (2019) Postoperative atrial fibrillation	IL-6 not significant	Endotoxin-induced	40, patients undergoing cardiac surgery
	CRP not significant		
Stavrakis et al. (2015) Atrial fibrillation	TNF-α decrease ($p < 0.05$)	Endotoxin-induced	40, paroxysmal atrial fibrillation ablation patients
	IL-6 not significant		
	IL-10 not significant		
	CRP decrease ($p < 0.05$)		
Stavrakis et al. (2020) Atrial fibrillation	TNF-α decrease ($p = 0.0093$)	Circulating	53, paroxysmal atrial fibrillation patients
	IL-6 not significant		
	IL-1β not significant		
	IL-17 not significant		
Yu et al. (2017) Myocardial ischemia-reperfusion injury	TNF-α lower increase than control ($p < 0.05$)	Circulating	95, myocardial ischemia-reperfusion injury patients
	IL-6 lower increase than control ($p < 0.05$)		
	IL-1β lower increase than control ($p < 0.05$)		
	HMGB1 lower increase than control ($p < 0.05$)		
Manasiev et al. (2016) Coronary insufficiency and LV dysfunction	Active group 1a HSP60 increase ($p < 0.05$); HSP70 increase ($p < 0.05$)	Circulating	70, coronary insufficiency/LV dysfunction patients
	Active group 2 HSP60 not significant; HSP70 increase ($p < 0.05$)		
Salama et al. (2020) Acute inflammatory response after lung lobectomy	TNF-α not significant	Circulating	100, lobectomy via thoracotomy in patients with non-small cell lung cancer
	IL-6 lower increase than control ($p = 0.02$)		
	IL-10 not significant		
	IL-1β not significant		
	IL-18 not significant		
	CRP lower increase than control ($p = 0.01$)		

**Risk of Bias score indicated by box color: red for "high," orange for "some concerns," and green for "low" risk of bias.*
CRP, C-reactive protein; IL, interleukin; LV, left ventricular; TNF-a, tumor necrosis factor alpha. A full list of abbreviations is found in ***Supplementary Material 3****.*

studies used widely varying stimulation parameters yet achieved similar results.

Pain Related Effects of aVNS

Ten studies investigated the effects of aVNS on the amelioration of pain and are summarized in **Table 7**. Five studies investigated the effects of aVNS on evoked pain threshold levels in healthy subjects but used varying pain-assessment methods. One study investigated the effects of aVNS on evoked pain threshold levels in chronic pancreatitis patients (Juel et al., 2017). The other four studies examined the effects of aVNS on self-reported pain scores in patients already suffering from pain due to endometriosis (Napadow et al., 2012), chronic migraine (Straube et al., 2015), fibromyalgia (Kutlu et al., 2020), and gastrointestinal (GI) disorders (Kovacic et al., 2017).

Across studies, the observed effects of aVNS for pain are highly varied. In studies that investigated evoked pain thresholds, results showed negligible changes in pain threshold levels due to aVNS therapy. For chronic migraines, Straube et al. (2015) reported a therapeutic effect in both the 1 and 25 Hz stimulation groups. Unexpectedly, the 1 Hz stimulation, originally designated as sham in the trial design, resulted in a reduction of headaches—comparable to medications used in migraine prevention—while the 25 Hz treatment had a smaller effect on the reduction of headaches (−7.0 episodes with 1 Hz vs. −3.3 episodes with 25 Hz over 28 days). Studies to isolate stimulation parameters that are therapeutic for pain are needed given that the 1 Hz stimulation, generally considered a sham due to its low frequency, was more effective than the 25 Hz intervention group. For gastrointestinal-related pain and chronic pelvic pain, pain was also significantly ameliorated by aVNS therapy. Overall,

TABLE 6 | Summary of epilepsy studies.

Author (y) indication*	Results	Active waveform	Study duration	Subjects analyzed, disease
Auricular VNS				
Bauer et al. (2016) Epilepsy	**Between group (not significant):** 2.9% ± 94.4 increase in 1 Hz sham group seizure frequency per 28 days and 23.4% ± 47.2 decrease in 25 Hz active group ($p = 0.146$)	25 Hz, 250 μs PW	20 weeks	58, epileptic
Aihua et al. (2014) Epilepsy	**Between group (S)**: not significant until 12 months; active median seizure frequency per month 4.0 (IQR 2.8–8.3) vs. sham median 8.0 (IQR 4.5–12.0) ($p < 0.001$)	20 Hz, 200 ms PW	12 months	47, epileptic
	Within group (S): Active seizure frequency per month reduced by 40% over 12 months; active baseline 6.0 (IQR I4.8–25.0) to 5.5 (IQR 3.0–12.0) at 6 months ($p < 0.001$) and 4.0 (IQR 2.8–8.3) at 12 months ($p < 0.001$)			
Rong et al. (2014) Epilepsy	**Between group (S)**: After 8 weeks, seizure frequency per 4 weeks decreased from 84.6 ± 145.5 to 48.6 ± 118.8 in active and 66.43 ± 85.5 to 58.8 ± 88 in sham ($p < 0.05$)	20–30 Hz, ≤1 ms PW	24 weeks	144, epileptic
Cervical implanted VNS				
Author (y)	**Results**			**Study Duration**
Ben-Menachem et al. (1994)	30.9% active seizure frequency decrease per 12 weeks, 11.3% sham decrease ($p = 0.036$ between, $p < 0.001$ active within, $p = 0.072$ sham within)			14 weeks
Handforth et al. (1998)	28% active seizure frequency decrease over 3 months, 15% sham decrease ($p = 0.04$ between, p <0.0001 active within, $p < 0.0001$ sham within)			3 months

**Risk of Bias score indicated by box color: red for "high," orange for "some concerns," and green for "low" risk of bias.*
IQR, Interquartile range; PW, pulse wdith. A full list of abbreviations is found in ***Supplementary Material 3****.*

these studies provide evidence that aVNS may be therapeutic for some pain conditions, but more studies are needed to further explore the effectiveness for specific medical conditions and rule out significant contributions from placebo effects.

Other Effects of aVNS

Other potential effects of aVNS investigated clinically included severity of motor symptoms in Parkinson's disease, depression, schizophrenia, obesity, impaired glucose tolerance, gastroduodenal motility, and tinnitus. Most of these effects were only investigated in a single RCT and there is not sufficient evidence to synthesize and evaluate across trials. The results of these individual studies are summarized in **Supplementary Material 3**. More studies are needed to further explore the effects of aVNS for these indications.

DISCUSSION

This is the first systematic review applying the Cochrane Risk of Bias (RoB) framework to auricular vagus nerve stimulation (aVNS) clinical trials. Our systematic review of 38 publications, totaling 41 RCTs, shows high heterogeneity in trial design and outcomes—even for the same indication. In the extreme, outcomes for heart rate effects of aVNS ranged from a consistent decrease in every subject in two studies to no heart rate effects in other studies. Findings on heart rate variability (HRV) conflict between studies and were hindered by trial designs including inappropriate washout periods and multiple methods used to quantify HRV. Early-stage evidence suggests aVNS reduces circulating levels and endotoxin-induced levels of inflammatory markers. Studies on epilepsy reached primary endpoints similar to previous RCTs on implantable VNS, albeit with concerns over quality of blinding. Clinical studies that tested aVNS for pain showed preliminary evidence of ameliorating pathological pain but not induced pain. Across the board, there are concerns on the extent of the contributions by placebo effects—especially since novel medical devices, such as aVNS devices, which also produce abnormal sensations (i.e., paresthesia), have a known larger placebo effect (Doherty and Dieppe, 2009).

The highest level of clinical evidence is multiple homogenous high quality RCTs as outlined in the Oxford CEBM Levels of Clinical Evidence Scale. The outcomes of these trials must consistently support the efficacy and safety of the therapy for a specific clinical indication. In the reviewed trials, several root causes—design of control, unblinding, and inconsistent reporting of results—raise the level of concern for bias in the outcomes. The current quality of evidence for aVNS RCTs supporting a particular clinical indication may generally be placed at grade 2, for "low quality" RCT, on the Oxford CEBM scale (Centre for Evidence-Based Medicine, 2009). An RCT is considered "low quality" for reasons including imprecise estimates, variability in results, indirect evidence, and presence of publication bias. For aVNS to reach the highest quality of clinical evidence for a particular indication, multiple RCTs must homogeneously support the safety and efficacy of the therapy for that indication.

In the following sections, we discuss gaps and improvements informed by our systematic review to aid the development of aVNS therapies. In the short term, we suggest improvements in the reporting of clinical trial results to allow meta-analysis of

TABLE 7 | Summary of pain studies.

Author (y) indication*	Results (compared to control)	Duration of study	Subjects analyzed, disease
Busch et al. (2013) Pain	**Not significant: fourteen** parameters were measured on both ipsilateral and contralateral. The pilot study did not correct for multiple comparisons. Mechanical and tonic heat pain parameters were marked for statistical significance and planned to be investigated further.	2 sessions (48 hrs apart)	48, healthy
Janner et al. (2018) Pain	**Not significant:** perceived pain intensity and temporal summation of pain not significant between active, placebo, and sham.	4 sessions (48 h apart)	49, healthy
Juel et al. (2017) Pain and GI motility	**Not significant:** no significance in evoked pain threshold between active and sham.	2 sessions (1 week apart)	20, chronic pancreatitis patients
Kovacic et al. (2017) GI pain	**Significant:** sham had significantly higher intensity pain at end of 3 weeks (median 7.0, IQR 5.0–9.0) compared to active (5.0, 4.0–7.0; $p = 0.003$). The composite score of the Pain-Frequency-Severity-Duration (PFSD) was significantly lower in the active group (8.4, IQR 3.2–16·2) than those in the sham group (15.2, 4.4–36.8) $p = 0.003$.	4 weeks	104, children with GI pain
Frokjaer et al. (2016) Gastroduodenal motility and pain threshold	**Not significant:** compared to sham, evoked pain threshold on bone increased ($p = 0.001$). Muscle evoked pain thresholds were significantly different at baseline ($p = 0.013$).	2 sessions (~6 days apart)	18, healthy
Johnson et al. (1991) Pain threshold and autonomic function	**Not significant:** no significance in evoked pain threshold or autonomic measures between any of the 3 active groups and control.	1 session	24, healthy
Kutlu et al. (2020) Fibromyalgia	**Not significant:** little improvement in pain visual analog scale ($p = 0.084$), physical role difficulty ($p = 0.496$), or emotional role difficulty ($p = 0.194$) between control (exercise) and active group (aVNS and exercise).	4 weeks	52, fibromyalgia patients
Laqua et al. (2014) Pain	**Not significant:** no significant evoked pain threshold difference between aVNS and placebo.	2 sessions (1 week apart)	21, healthy
Napadow et al. (2012) Pain	**Significant:** reduction in deep pain intensity rating ($p = 0.049$). Reduction in temporal summation ($p = 0.04$).	2 sessions (1 week apart)	15, chronic Pelvic Pain due to endometriosis patients
Straube et al. (2015) Chronic migraine	**Significant:** headache occurrence −7.0 per 28 days in 1 Hz group (36.4% reduction) and −3.3 in 25 Hz group (17.4% reduction) ($p = 0.035$).	12 weeks	46, chronic migraine patients

**Risk of Bias score indicated by box color: red for "high," orange for "some concerns," and green for "low" risk of bias.*
GI, Gastrointestinal; IQR, interquartile range. A full list of abbreviations is found in ***Supplementary Material 3****.*

results across aVNS studies. In the long term, we highlight the need for direct measures of target engagement as biomarkers to study therapeutic effects and therapy limiting side effects, and better translate learnings from animal models to humans. Also in the long term, we discuss the needs and associated challenges in careful design of controls and maintenance of blinding.

Short Term Solutions—Guide to Reporting

Several steps can be implemented immediately to increase the quality and consistency of reporting in aVNS studies, enabling comparison of results across studies.

Across the studies, the greatest risk of bias came from the Cochrane section "selection of reported results." A comprehensive guide to clinical trial reporting is found published by the CONSORT group along with detailed elaborations (Moher et al., 2010). See Kovacic et al. (2017) for an aVNS study that followed the CONSORT reporting recommendations. Pre-registration of trials, use of appropriate statistical analysis, justifying clinical relevance of outcome measures, and contextualizing clinical significance of results are discussed here. These ideas are summarized in **Table 8**. If followed across aVNS studies, these suggestions would reduce the risk of bias identified in the RoB section reporting of results and enable the synthesis of knowledge by making reporting more comparable across studies (Farmer et al., 2020).

Pre-registration

Pre-registration of planned enrollment, interventions, outcome measures and time points, and statistical plan to reach primary and secondary endpoints reduces risk of bias in the reporting of results. When the trial is reported, commentary should be made on adherence and deviations from the pre-registration with appropriate justifications. Exploratory analysis of the data may still be performed but needs to be denoted. Exploratory analysis can be used to suggest design of future investigations. The amount of exploratory analysis should be limited, and all non-significant exploratory analysis performed before reaching the significant results should also be reported. Of the 41 aVNS RCTs reviewed, 13 RCTs pre-registered, but only 5 had sufficient information to be considered a complete pre-registration. Pre-registration reduces risk of bias in reporting of results by preventing analysis of only select measures (section 5.1 of RoB rubric) and multiple analysis of data (section 5.2 of RoB rubric).

Appropriate Statistical Analysis

Data may be analyzed in many ways to claim the effect of an intervention. For example, studies may report a between group analysis comparing the change in the active arm to the change in the control arm or a within group analysis comparing the active arm after treatment to baseline. The more

TABLE 8 | Checklist for trial reporting.

	Item
✓	Pre-register
✓	Use appropriate statistical analysis
✓	Justify choice of between versus within group analysis
✓	Perform baseline comparison (even in crossover studies)
✓	Perform statistical test for carryover effects
✓	Report individual results
✓	Justify clinical relevance of outcome measures
✓	Discuss clinical significance of results
✓	Compare outcome to existing therapy
✓	Consider subject population when generalizing findings

appropriate method for a controlled study is a between group comparison of the active arm vs. the control arm. Several studies claimed statistically significant findings based just on the within group analysis even if the between group analysis was non-significant. An example illustrating this difference is found in **Supplementary Material 6**. Pre-registration of the planned statistical analysis will discourage unjustified multiple analysis of the data.

In crossover design studies, there was a major gap in the reporting of baseline comparison between randomized groups. Even in a crossover design where each subject receives all interventions, it is crucial to compare baseline differences between groups as one would do for a parallel study. This is especially pertinent in pilot studies with small sample sizes, where a baseline imbalance between groups is more likely to occur and affect the trial outcome (Kang et al., 2008). Additionally, if the order of intervention becomes pertinent, due to an incomplete washout period or compromised blinding, then it is essential that the baseline randomization between groups is balanced to enable further analysis.

Another concerning gap in crossover design studies was in the lack of reporting the statistical test for carryover effects. The test detects if the order of intervention received had an effect on the outcome (Shen and Lu, 2006). The test for carryover effects shows significance when there are incomplete washout effects, baseline imbalances, or compromise in blinding. It is perhaps the single most important gauge of the quality of a crossover design and should always be performed and reported—only 4 of 20 crossover design studies reported the carryover effects test. A baseline comparison between groups will ensure that baseline differences do not contribute to significance in the test for crossover effects—allowing effects from incomplete washout periods and compromised blinding to be isolated.

Reporting of individual results is a simple and effective way to convey the average and variance in outcomes, the fraction of responders, and worsening of symptoms (if any) in the non-responders. In Stavrakis et al. (2020) there is worsening of symptoms in the non-responders (53% of the active group) at the 3-month evaluation, which is also the only time point at which atrial fibrillation burden is measured concurrently during stimulation. This clinically relevant finding was evident during review because individual results were presented. Individual results were only presented in 10 of 41 aVNS studies reviewed. Reporting of individual results should be considered where allowed by clinical trial protocol.

Justify Clinical Relevance of Outcome Measures

The outcome measure itself may not be established as clinically relevant. For example, *in vitro* endotoxin-induced cytokine measurements were used to proxy *in vivo* immune response in several aVNS studies including Addorisio et al. (2019). While endotoxin-induced cytokine levels produce a stronger signal, they may not be clinically relevant in the case of an auto-immune disease such as Rheumatoid Arthritis tested in Addorisio et al. (2019).

The clinical accuracy of the measurement tool must also be considered. For example, aVNS studies often used photoplethysmography (PPG) based methods at the finger to measure blood pressure. Given the change in blood pressure signal during aVNS is already small, it is unnecessary to lose statistical power by using less accurate PPG based methods (Elgendi et al., 2019) to measure blood pressure. Clancy et al. (2014) used both a finger-based PPG, Finometer®, and a traditional arm sphygmomanometer to measure blood pressure and concluded that the increase in blood pressure measured using the Finometer® may be due to an artifact of the PPG measurement method. Discussion on clinical relevance of the outcome measure provides justification for the selection of reported results.

Contextualize Clinical Significance of Results

An outcome that is statistically significant does not necessarily indicate clinical significance. Contextualizing the trial results allows the reader to better understand the clinical significance of the findings. This may be done by comparing the study outcome to the outcome of the standard of care or another therapy. For example, Cakmak et al. (2017), in an aVNS trial for Parkinson's disease, showed a 5.3 points improvement on the UPDRS part 3 for motor symptoms (Goetz et al., 2008). Their result could be contextualized with the 18.4 points improvement in DBS (Kahn et al., 2019).

Clinical significance of results should be discussed in relevance to the subject population—particularly with consideration to disease severity and heterogeneity. For example, Juel et al. (2017) repeated a study in the diseased population after Frøkjaer et al. (2016) first reported a similar trial in healthy subjects. While the study in healthy subjects concluded significant findings, the subsequent study in diseased subjects did not. They cited pathological neural circuitry as a possible reason. Whether the difference was due to pathophysiology or differences in trial design and analysis is uncertain. Regardless, inclusion and exclusion criteria often restrict the subjects enrolled in terms of disease severity and heterogeneity and consideration should be given to the study subject population when discussing clinical significance of the findings.

Lessons From Drug World

As the translation of drugs into clinical use is more established than neuromodulation therapies, it is instructive to review the translation of drugs for pitfalls in moving toward FDA market approved therapies. Less than 12% of drugs that received an FDA Investigational New Drug approval to begin human studies (the current stage of development of many aVNS based therapies) reached market approval (Paul et al., 2010; DiMasi et al., 2016). Gupta et al. (2011) identified several factors that hindered the successful translation of drug therapies from early-stage results to market approval, which are also relevant to aVNS therapies.

1. Lack of pharmacodynamic measures in early-stage clinical trials to confirm drug activity (Gallo, 2010). This is similar to the lack of evidence that aVNS is activating desired fiber types in the auricular branch of the vagus and not activating other fiber types including those within the great auricular, lesser occipital, facial, and trigeminal nerves which innervate the auricular and periauricular region.
2. Lack of validated biomarkers for on- and off-target engagement—impacting our ability to assess and confirm therapeutic activity vs. side effects (Institute of Medicine, 2014). Again, similar to the lack of biomarkers in aVNS trials to confirm on- and off-target nerve activation.
3. Lack of predictability of animal models for humans (Johnson et al., 2001). Relevant in aVNS to translatability of electrode configuration, dosing, and stimulation parameters given changes in size, neuroanatomy, and neurophysiology from animal models to humans. A concern confirmed by other neuromodulation therapies (De Ferrari et al., 2017).

A method to directly measure local neural target engagement will provide an immediate biomarker of on- and off-target activity and forestall some of the hurdles encountered in drug therapy development. A minimally invasive method to measure target engagement percutaneously could be deployed across preclinical models and early clinical studies (Ottaviani et al., 2020). Doing so would increase the translatability of findings by providing data to titrate electrode design, placement, and stimulation waveform parameters to optimize for on-target engagement.

Long Term Solutions—Target Engagement

On- and off-target nerve activation is especially relevant in the case of the auricle that is innervated by several nerves with uncertainty on the specific areas of innervation in literature and across subjects. Data from direct measures of on- and off-target engagement could be used (1) to titrate the therapy by adjusting electrode design, placement, and stimulation waveform parameters to optimize on-target engagement, (2) to scale pre-clinical animal doses to humans by preserving the fiber types activated, and (3) to help investigate fundamental mechanisms of action by isolating local neural pathways.

aVNS is commonly delivered at the cymba concha with the assumption that the cymba concha is innervated only by the auricular vagus. This is based on two pieces of evidence. Firstly, Peuker and Filler (2002) showed the cymba concha is innervated only by the auricular vagus in 7 of 7 cadavers. Notwithstanding, there may be variation in innervation or spread of the electric field, which could activate the neighboring auriculotemporal branch of the trigeminal nerve and even the great auricular nerve. Variations in peripheral nerve innervation has been well-studied for other regions of the body, such as the hand (Bas and Kleinert, 1999; Guru et al., 2015). The reliance on the Peuker and Filler study is concerning due to its small sample size. Given the importance of the claims in Peuker and Filler, a further dissection study with a larger sample size is called for to investigate whether the results hold over a larger population of ethnically diverse individuals. Secondly, functional magnetic resonance imaging (fMRI) evidence is also used to suggest vagal innervation of the conchae (Frangos et al., 2015). However, fMRI is a surrogate measure of target engagement and is especially problematic when imaging deep in the brainstem, as described below.

Target engagement is commonly established using secondary surrogates such as fMRI, somatosensory evoked potentials (SSEPs), and cardiac measures. Secondary surrogates of target engagement are often contaminated with physical and biological noise, leading to potential confounds. For example, Botvinik-Nezer et al. (2020) and Becq et al. (2020) showed that results of fMRI studies were highly dependent on data processing techniques applied. In addition, pathways starting from the trigeminal nerve in the auricle also connect to NTS (Chiluwal et al., 2017), and activate the same region in the brain when in fact the auricular vagus might not be recruited during stimulation. fMRI of the brainstem is further complicated (Napadow et al., 2019) as distance from the measurement coils increases, the effective resolution decreases (Gruber et al., 2018). Still further, novelty, such as being stimulated in the ear or being in an MRI scanner, activates the locus coeruleus (LC) (Wagatsuma et al., 2018), which has connections with NTS. Thereby confounding potential aVNS effects on NTS with LC induced activity due to novelty effects. Lastly, fMRI has non-standard results between subjects requiring individual calibration and making subject to subject comparisons challenging. Additionally, SSEP recordings are sometimes contaminated and misinterpreted due to EMG leakage (Usami et al., 2013). The common measures of target engagement are secondary surrogates and prone to confounds—creating a need for direct measures of local target engagement at the nerve trunks innervating the auricle.

Given the lack of direct measures of local target engagement, aVNS studies rely largely on stimulation parameters that are similar to those used for implantable VNS. The assumption that these stimulation parameters will result in similar target engagement and therapeutic effects may not hold due to the differences in target fiber type, fiber orientation, and electrode design and contact area—all of which affect neural recruitment. Cardiac effects of implantable VNS are thought to be mediated by activation of parasympathetic efferent B fibers innervating the heart (Sabbah et al., 2011) or aortic baroreceptor afferents depending on the stimulation parameters used. Studies of the baroreceptors at the aorta and carotid sinus bulb identified fiber types consistent with Aδ and C fibers (Seagard et al., 1990; Reynolds et al., 2006). Strikingly, consensus workshops have suggested that aVNS is mediated by activation of Aβ

fibers (Kaniusas et al., 2019). There is also no evidence of baroreceptors identified in the ear. These difference between the auricular and cervical vagus suggests a direct porting of stimulation parameters developed for implantable VNS would be insufficient to invoke cardiac responses, unless another yet unidentified mechanism for cardiac responses mediated through NTS is responsible, which can be targeted by Aβ fiber input that then indirectly modulates sympathetic or parasympathetic input to the heart. Unlike stimulation of the cervical vagus nerve trunk, where the electrode contacts are oriented parallel to the target axons, electrode contacts for aVNS do not have consistent orientation with respect to the target axons, which exist as a web of axons in the auricle. Orientation of fibers relative to the stimulation electrode have a large effect on fiber recruitment (Grill, 1999) and could potentially lead to preferential activation of nerve pathways oriented in a particular direction to the stimulation contacts, as well as inconsistent activation of specific fiber types across the auricle. Additionally, target fibers in the auricle transition to unmyelinated fibers as they approach sensory receptor cells (Provitera et al., 2007). For these reasons, while cathodic leading stimulation might have lower recruitment thresholds in implantable VNS, the principle may not hold for aVNS (Anderson et al., 2019). In aVNS, target fiber type, electrode design, electrode size, transcutaneous placement, and orientation of the target fiber relative to the electrode are different both compared to implantable VNS and across aVNS studies. Therefore, it is unsurprising that stimulation parameters ported from implantable VNS may not replicate the physiological effects or recruitment of fiber types that have been observed during implantable VNS.

In relation to electrode design, injected charge density, as opposed to current or voltage, is the most relevant metric of neural activation. This is because stimulation evoked action potentials occur in regions of the neural cell membrane where there is an elevated charge density (McNeal, 1976; Rattay, 1999). For effective comparison across studies using different electrodes, it is imperative to report on the electrode area, especially on the area as it makes contact with tissue, along with stimulation current.

The above discussion stresses a general lack of confidence in ascertaining which of several nerve trunks innervating the auricle are being activated during aVNS that may be generating the on- or off-target effects. Cakmak et al. (2017) further proposed that the therapeutic effects they reported for motor symptoms of Parkinson's diseases came from direct recruitment of the intrinsic auricular muscles instead of the auricular vagus nerve. The fundamentals of neural stimulation do not support directly porting stimulation parameters from implantable VNS. Therefore, it is important to understand which fiber types on the auricular vagus are being activated, if at all. This knowledge requires direct measures of local target engagement from the nerves innervating the auricle.

To further the development of aVNS, it will be essential to understand local target engagement of the nerve trunks innervating the ear. Ultrasound guided (Ritchie et al., 2016) percutaneous microelectrode recordings (Ottaviani et al., 2020) from the major nerve trunks innervating the ear, similar to the technique to measure muscle sympathetic nerve activity (MSNA), provides a way to directly measure local neural recruitment. Real-time data on neural target engagement would enable optimization of stimulation parameters, electrode, and control designs (Chang et al., 2020). These data would also improve translation of stimulation dosages from animal models to humans. Tsaava et al. (2020) showed that the anti-inflammatory effects of implantable VNS are stimulation dose dependent and could lead to the worsening of inflammation at certain dosages. Since target engagement in humans is currently unknown, there is no consistent means of determining dosage accurately—raising potential safety concerns. This minimally invasive method to record neural target engagement is already used clinically and could be rapidly translated to the clinic for use in titrating neuromodulation therapies.

Understanding primary target engagement at the ear will also further our understanding of aVNS mechanisms. For example, large animal recordings of evoked compound action potentials from the major nerve trunks innervating the ear may help in understanding the relationship between on- and off-target nerve engagement and corresponding physiological effects. Simultaneous recordings at the cervical vagus may allow differentiation of direct efferent vagal effects vs. NTS mediated effects, which would appear with a longer latency due to synaptic delay and longer conduction path length. Measuring neural target engagement at the auricle provides a first step to systematically studying aVNS mechanisms and optimizing clinical effects.

Long Term Solutions—Improvement in Control Design and Blinding

The design of an indistinguishable yet non-therapeutic control is central to maintaining the blinding in a clinical trial. Stemming from limited understanding of local target engagement and mechanism of action of aVNS, it is difficult to implement an active control (i.e., sham) that has similar perception to the therapeutic group but will not unknowingly engage a therapeutic pathway. This uncertainty in the therapeutic inertness of the control violates some of the basic premises for a RCT and makes it difficult to evaluate aVNS RCTs on the Oxford Scale for clinical evidence. Systematic effort must be made to design controls, which are key to maintaining blinding in aVNS RCTs.

Common control designs used in aVNS studies are summarized in **Figure 5**. A placebo is defined when the electrode placement and device are similar to the active intervention, but no stimulation is delivered. A waveform sham is defined when a different—non-therapeutic—waveform is delivered at the same location as active intervention. In a location sham, the same waveform as active intervention is delivered at a different location on the auricle and should not engage a therapeutic nerve. Lastly, no intervention or a pharmacological control may be used. Location sham was the most common control used in 16 of 41 RCTs reviewed. These different control designs are evaluated at length in **Supplementary Material 4** along with

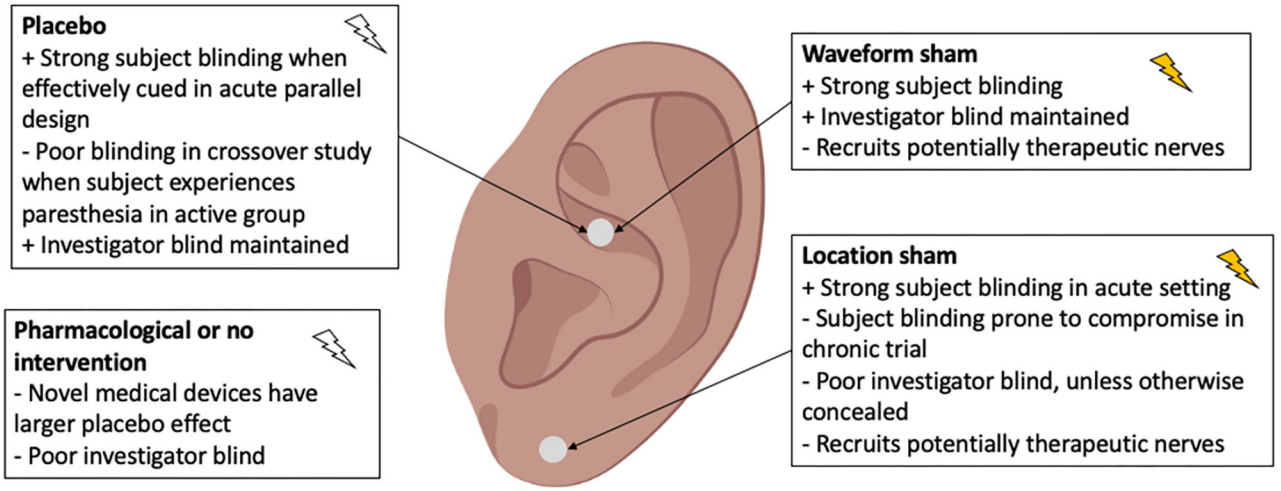

FIGURE 5 | Types of controls used in aVNS clinical trials. The ideal control is indistinguishable from active intervention, to both subject and investigator, yet therapeutically inert.

recommendations on appropriate control types depending on trial design.

Inappropriate implementation of the control group resulted in compromised blinding in many studies. Subject unblinding occurred when subjects were able to feel a paresthesia in the active intervention but not in the control. For example, in cross-over trials, where subjects undergo both the active and control intervention, *post-hoc* assessment of blinding becomes critical given the perception of sensation being unequal could easily break the blind. Investigator unblinding occurred when investigators were able to see differences in electrode placement or device operation. Unblinding due to inappropriate control design is the main contributor leading to risk of bias in the Cochrane section "deviation from intended intervention." The design of appropriate controls is difficult for trials testing non-pharmacological interventions—especially for paresthesia-inducing neuromodulation trials (Robbins and Lipton, 2017; Translating neuromodulation, 2019)—but is essential to establish a double blind.

To aid in the maintenance of the double blind, appropriate design of controls and consistent evaluation of blinding are required. Measurement of target engagement via microneurography of the major nerve trunks innervating the ear (Ottaviani et al., 2020) will enable understanding of neural recruitment occurring during active and sham stimulation and guide appropriate design of controls. Appropriate design of controls for non-invasive neuromodulation studies are discussed at length in **Supplementary Material 4**. In addition, *post-hoc* evaluation of blinding in subjects and investigators will gather knowledge on the blind quality setup by respective control designs. Methods to assess the quality of blinding in non-invasive neuromodulation studies are discussed in **Supplementary Material 5**. Three-armed trials (no treatment, placebo/sham, and treatment) will also generate data on extent of the placebo effect and quality of blinding (Howick et al., 2013). Over time, the consistent use of control designs and evaluation of blinding will grow our understanding of the concealability and therapeutic inertness of various control methods.

Applicability of Findings to Similar Neuromodulation Therapies

The discussion on lack of direct measures of target engagement and unknowns surrounding implementation of perceptually similar yet therapeutically inert controls to maintain the double blind are applicable to other neuromodulation therapies—especially paresthesia inducing therapies such as implantable VNS and spinal cord stimulation (SCS).

Study of local target engagement in neuromodulation therapies such as implantable VNS and SCS will inform stimulation parameters, possible mechanisms of action, and electrode design and placement to maximize on-target nerve recruitment and minimize therapy limiting off-target effects such as muscle contractions (Yoo et al., 2013; Nicolai et al., 2020). To investigate the central mechanisms of action we first have to establish local target engagement to determine which on- and off-target nerves are recruited during stimulation and which fiber types are recruited at therapeutically relevant levels of stimulation. Measurement of local target engagement in preclinical and early clinical studies provides a bottom-up approach to systematically develop and deploy neuromodulation therapies.

A systematic review and meta-analysis of SCS RCTs for pain showed that the quality of control used had an impact on the effect size of the outcome, (Duarte et al., 2020b). They concluded that thorough consideration of control design and consequent subject and investigator blinding is essential to improve the quality of evidence on SCS therapy for pain (Duarte et al., 2020a). A systematic review of dorsal root ganglion (DRG) stimulation for pain also showed serious concern for bias across all studies reviewed due to compromised subject and investigator blinding (Deer et al., 2020). The suggestions

laid forth on measuring local target engagement and consistent *post-hoc* evaluation of blinding in subjects and investigators will expand our knowledge of effective control design for paresthesia inducing neuromodulation therapies such as aVNS, SCS, and implantable VNS.

Limitations of This Review

This review is based on experience in other neuromodulation clinical trials and pre-clinical studies, existing frameworks for analysis such as the Cochrane Risk of Bias and Oxford clinical scale, and literature review motivated by an interest in conducting future aVNS clinical trials. However, at the writing of this article, none of the authors have conducted an aVNS clinical trial. Secondly, this is a systematic review but not a meta-analysis. Due to insufficient reporting in trials, a meta-analysis could not be conducted. The analysis was more qualitative with the intention of summarizing the quality of evidence in the field and making recommendations to improve clinical translatability. Thirdly, this review was not pre-registered, blinded, or formally randomized. Additionally, while the RoB tool provides a consistent method to evaluate trials where the shortcoming is stated explicitly, the ability to identify confounds is often reliant on the critical reading of the reviewer. This made it possible for a secondary reviewer to find additional risk of bias in several instances, which were initially missed by the primary reviewer but included upon identification and consensus. Lastly, numerous instances of missing information in trial reporting were identified and attempts were made to reach out to the authors for that information. These attempts were not always successful. Overall, the points made in the review are robust and withstand the limitations.

CONCLUSION

Based on our review of 38 publications, which reported on 41 aVNS clinical RCTs, aVNS shows physiological effects but has not yet shown strong clinically significant effects consistently supported by multiple studies. This review: (1) Identifies concerns in the design of trials, particularly control and blinding, and incomplete reporting of information using the Cochrane Risk of Bias analysis. (2) Finds aVNS studies are presently exploratory in nature, which is appropriate given the early stage of research of the aVNS field. (3) Qualitatively synthesizes study outcomes by clinical indications. (4) Proposes guidelines for the reporting of aVNS clinical trials, which can be implemented immediately to improve the quality of evidence. (5) Proposes progress in the field has been limited by lack of direct measures of neural target engagement at the site of stimulation. Measures of target engagement will inform therapy optimization, translation, and mechanistic understanding. (6) Proposes consistent *post-hoc* evaluation of subject and investigator blinding and direct measures of local neural target engagement to improve the design of controls for maintenance of blinding.

As a field, neuromodulation has not yet attained social normality or gained widespread adoption as a first-line therapy (Pagnin et al., 2004; Payne and Prudic, 2009; Li et al., 2020). To that end, our responsibility as pioneers is to move the field forward and build its credibility by thoroughly reporting on appropriately designed clinical trials. Given the conflicting data across aVNS studies, it is critical to implement high standards for rigor and quality of evidence to better assess the state of aVNS for a given indication.

DATA AVAILABILITY STATEMENT

The original contributions generated for the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

NV, JT, EL, JW, and KL contributed to conception and design of the study. NV, JM, MK, and RC conducted the systematic search. NV, JM, MK, and RC reviewed the included articles. NV wrote the first draft of the manuscript. JM, MK, SB, and KL wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

The work presented here was funded by the Defense Advanced Research Projects Agency Biological Technologies Office (BTO) program title Targeted Neuroplasticity Training (TNT) under the auspices of Doug Weber and Tristan McClure-Begley through the Space and Naval Warfare Systems Command Pacific with cooperative agreement no. N66001-17-2-4010. The views, opinions, and/or findings expressed are those of the author and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

ACKNOWLEDGMENTS

Eric H. Chang for notes on endotoxin-induced cytokine assays. Carly Frieders, Arkaprabha Banerjee, Maria LaLuzerne, Megan Settell, Robbin Miranda, Rick Chappell, Carolyn Huff, and Harrell Huff for reviewing drafts of this publication. Members of the Wisconsin Institute for Translational Neuroengineering (WITNe) for feedback during group meetings.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2021.664740/full#supplementary-material

REFERENCES

Translating Neuromodulation (2019). *Translating Neuromodulation. Nat. Biotechnol.* 37, 967–967. doi: 10.1038/s41587-019-0263-3

Addorisio, M. E., Imperato, G. H., de Vos, A. F., Forti, S., Goldstein, R. S., Pavlov, V. A., et al. (2019). Investigational treatment of rheumatoid arthritis with a vibrotactile device applied to the external ear. *Bioelectron. Med.* 5:4. doi: 10.1186/s42234-019-0020-4

Afanasiev, S. A., Pavliukova, E. N., Kuzmichkina, M. A., Rebrova, T., Anfinogenova, Y., Likhomanov, K. S., et al. (2016). Nonpharmacological correction of hypersympatheticotonia in patients with chronic coronary insufficiency and severe left ventricular dysfunction: nonpharmacological correction of hypersympatheticotonia. *Ann. Noninvasive Electrocardiol.* 21, 548–556. doi: 10.1111/anec.12349

Aihua, L., Lu, S., Liping, L., Xiuru, W., Hua, L., and Yuping, W. (2014). A controlled trial of transcutaneous vagus nerve stimulation for the treatment of pharmacoresistant epilepsy. *Epilepsy Behav.* 39, 105–110. doi: 10.1016/j.yebeh.2014.08.005

American Association of Neurological Surgeons (2021). *Vagus Nerve Stimulation*. Available online at: https://www.aans.org/Patients/Neurosurgical-Conditions-and-Treatments/Vagus-Nerve-Stimulation (accessed January 12, 2021).

Anderson, D. N., Duffley, G., Vorwerk, J., Dorval, A. D., and Butson, C. R. (2019). Anodic stimulation misunderstood: preferential activation of fiber orientations with anodic waveforms in deep brain stimulation. *J. Neural. Eng.* 16:016026. doi: 10.1088/1741-2552/aae590

Andreas, M., Arzl, P., Mitterbauer, A., Ballarini, N. M., Kainz, F.-M., Kocher, A., et al. (2019). Electrical stimulation of the greater auricular nerve to reduce postoperative atrial fibrillation. *Circ. Arrhythm. Electrophysiol.* 12:e007711. doi: 10.1161/CIRCEP.119.008067

Antonino, D., Teixeira, A. L., Maia-Lopes, P. M., Souza, M. C., Sabino-Carvalho, J. L., Murray, A. R., et al. (2017). Non-invasive vagus nerve stimulation acutely improves spontaneous cardiac baroreflex sensitivity in healthy young men: a randomized placebo-controlled trial. *Brain Stimul.* 10, 875–881. doi: 10.1016/j.brs.2017.05.006

Badran, B. W., Mithoefer, O. J., Summer, C. E., LaBate, N. T., Glusman, C. E., Badran, A. W., et al. (2018). Short trains of transcutaneous auricular vagus nerve stimulation (taVNS) have parameter-specific effects on heart rate. *Brain Stimul.* 11, 699–708. doi: 10.1016/j.brs.2018.04.004

Badran, B. W., Yu, A. B., Adair, D., Mappin, G., DeVries, W. H., Jenkins, D. D., et al. (2019). Laboratory administration of Transcutaneous Auricular Vagus Nerve Stimulation (taVNS): technique, targeting, and considerations. *J. Vis. Exp.* 143:e58984. doi: 10.3791/58984

Bas, H., and Kleinert, J. M. (1999). Anatomic variations in sensory innervation of the hand and digits. *J. Hand Surg. Am.* 24, 1171–1184. doi: 10.1053/jhsu.1999.1171

Bauer, S., Baier, H., Baumgartner, C., Bohlmann, K., Fauser, S., Graf, W., et al. (2016). Transcutaneous Vagus Nerve Stimulation (tVNS) for treatment of drug-resistant epilepsy: a randomized, double-blind clinical trial (cMPsE02). *Brain Stimul.* 9, 356–363. doi: 10.1016/j.brs.2015.11.003

Becq, G. J.-P. C., Barbier, E. L., and Achard, S. (2020). Brain networks of rats under anesthesia using resting-state fMRI: comparison with dead rats, random noise and generative models of networks. *J. Neural Eng.* 17:045012. doi: 10.1088/1741-2552/ab9fec

Ben-Menachem, E., Manon-Espaillat, R., Ristanovic, R., Wilder, B. J., Stefan, H., Mirza, W., et al. (1994). Vagus nerve stimulation for treatment of partial seizures: 1. a controlled study of effect on seizures. *Epilepsia* 35, 616–626. doi: 10.1111/j.1528-1157.1994.tb02482.x

Bermejo, P., López, M., Larraya, I., Chamorro, J., Cobo, J. L., Ordóñez, S., et al. (2017). Innervation of the human cavum conchae and auditory canal: anatomical basis for transcutaneous auricular nerve stimulation. *Biomed. Res. Int.* 2017, 1–10. doi: 10.1155/2017/7830919

Billman, G. E. (2013). The LF/HF ratio does not accurately measure cardiac sympatho-vagal balance. *Front. Physio.* 4:26. doi: 10.3389/fphys.2013.00026

Bootsma, M., Swenne, C. A., Janssen, M. J. A., Cats, V. M., and Schalij, M. J. (2003). Heart rate variability and sympathovagal balance: pharmacological validation. *Neth. Heart J.* 11, 250–259.

Borges, U., Laborde, S., and Raab, M. (2019). Influence of transcutaneous vagus nerve stimulation on cardiac vagal activity: not different from sham stimulation and no effect of stimulation intensity. *PLoS ONE* 14:e0223848. doi: 10.1371/journal.pone.0223848

Botvinik-Nezer, R., Holzmeister, F., Camerer, C. F., Dreber, A., Huber, J., Johannesson, M., et al. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. *Nature* 582, 84–88. doi: 10.1038/s41586-020-2314-9

Bretherton, B., Atkinson, L., Murray, A., Clancy, J., Deuchars, S., and Deuchars, J. (2019). Effects of transcutaneous vagus nerve stimulation in individuals aged 55 years or above: potential benefits of daily stimulation. *Aging* 11, 4836–4857. doi: 10.18632/aging.102074

Broekman, M. M. T. J., Roelofs, H. M. J., Hoentjen, F., Wiegertjes, R., Stoel, N., Joosten, L. A., et al. (2015). LPS-stimulated whole blood cytokine production is not related to disease behavior in patients with quiescent Crohn's disease. *PLoS ONE* 10:e0133932. doi: 10.1371/journal.pone.0133932

Burger, A. M., D'Agostini, M., Verkuil, B., and Van Diest, I. (2020). Moving beyond belief: a narrative review of potential biomarkers for transcutaneous vagus nerve stimulation. *Psychophysiology* 57:e13571. doi: 10.1111/psyp.13571

Burger, A. M., Van der Does, W., Thayer, J. F., Brosschot, J. F., and Verkuil, B. (2019). Transcutaneous vagus nerve stimulation reduces spontaneous but not induced negative thought intrusions in high worriers. *Biol. Psychol.* 142, 80–89. doi: 10.1016/j.biopsycho.2019.01.014

Busch, V., Zeman, F., Heckel, A., Menne, F., Ellrich, J., and Eichhammer, P. (2013). The effect of transcutaneous vagus nerve stimulation on pain perception – an experimental study. *Brain Stimul.* 6, 202–209. doi: 10.1016/j.brs.2012.04.006

Cakmak, Y. O. (2019). Concerning auricular vagal nerve stimulation: occult neural networks. *Front. Hum. Neurosci.* 13:421. doi: 10.3389/fnhum.2019.00421

Cakmak, Y. O., Apaydin, H., Kiziltan, G., Gündüz, A., Ozsoy, B., Olcer, S., et al. (2017). Rapid alleviation of Parkinson's disease symptoms via electrostimulation of intrinsic auricular muscle zones. *Front. Hum. Neurosci.* 11:338. doi: 10.3389/fnhum.2017.00338

Centre for Evidence-Based Medicine (2009). *Oxford Centre for Evidence-Based Medicine: Levels of Evidence (March 2009)*. Available online at: https://www.cebm.ox.ac.uk/resources/levels-of-evidence/oxford-centre-for-evidence-based-medicine-levels-of-evidence-march-2009 (accessed November 26, 2020).

Chang, Y.-C., Cracchiolo, M., Ahmed, U., Mughrabi, I., Gabalski, A., Daytz, A., et al. (2020). Quantitative estimation of nerve fiber engagement by vagus nerve stimulation using physiological markers. *Brain Stimul.* 13, 1617–1630. doi: 10.1016/j.brs.2020.09.002

Chiluwal, A., Narayan, R. K., Chaung, W., Mehan, N., Wang, P., Bouton, C. E., et al. (2017). Neuroprotective effects of trigeminal nerve stimulation in severe traumatic brain injury. *Sci. Rep.* 7:6792. doi: 10.1038/s41598-017-07219-3

Clancy, J. A., Mary, D. A., Witte, K. K., Greenwood, J. P., Deuchars, S. A., and Deuchars, J. (2014). Non-invasive vagus nerve stimulation in healthy humans reduces sympathetic nerve activity. *Brain Stimul.* 7, 871–877. doi: 10.1016/j.brs.2014.07.031

De Couck, M., Cserjesi, R., Caers, R., Zijlstra, W. P., Widjaja, D., Wolf, N., et al. (2017). Effects of short and prolonged transcutaneous vagus nerve stimulation on heart rate variability in healthy subjects. *Autonomic Neurosci.* 203, 88–96. doi: 10.1016/j.autneu.2016.11.003

De Ferrari, G. M., Stolen, C., Tuinenburg, A. E., Wright, D. J., Brugada, J., Butter, C., et al. (2017). Long-term vagal stimulation for heart failure: eighteen month results from the NEural Cardiac TherApy foR Heart Failure (NECTAR-HF) trial. *Int. J. Cardiol.* 244, 229–234. doi: 10.1016/j.ijcard.2017.06.036

Deer, T. R., Hunter, C. W., Mehta, P., Sayed, D., Grider, J. S., Lamer, T. J., et al. (2020). A systematic literature review of dorsal root ganglion neurostimulation for the treatment of pain. *Pain Med.* 21, 1581–1589. doi: 10.1093/pm/pnaa005

Del Giudice, M., and Gangestad, S. W. (2018). Rethinking IL-6 and CRP: Why they are more than inflammatory biomarkers, and why it matters. *Brain Behav. Immun.* 70, 61–75. doi: 10.1016/j.bbi.2018.02.013

DiMasi, J. A., Grabowski, H. G., and Hansen, R. W. (2016). Innovation in the pharmaceutical industry: new estimates of R&D costs. *J. Health Econ.* 47, 20–33. doi: 10.1016/j.jhealeco.2016.01.012

Doherty, M., and Dieppe, P. (2009). The "placebo" response in osteoarthritis and its implications for clinical practice. *Osteoarthritis Cartilage* 17, 1255–1262. doi: 10.1016/j.joca.2009.03.023

Duarte, R. V., McNicol, E., Colloca, L., Taylor, R. S., North, R. B., and Eldabe, S. (2020a). Randomized placebo-/sham-controlled trials of spinal cord stimulation: a systematic review and methodological appraisal. *Neuromodulation* 23, 10–18. doi: 10.1111/ner.13018

Duarte, R. V., Nevitt, S., McNicol, E., Taylor, R. S., Buchser, E., North, R. B., et al. (2020b). Systematic review and meta-analysis of placebo/sham controlled randomised trials of spinal cord stimulation for neuropathic pain: *Pain* 161, 24–35. doi: 10.1097/j.pain.0000000000001689

Elgendi, M., Fletcher, R., Liang, Y., Howard, N., Lovell, N. H., Abbott, D., et al. (2019). The use of photoplethysmography for assessing hypertension. *npj Digit. Med*. 2:60. doi: 10.1038/s41746-019-0136-7

Farmer, D. F., Strzelczyk, A., Finisguerra, A., Gourine, A. V., Gharabaghi, A., Hasan, A., et al. (2020). International consensus based review and recommendations for minimum reporting standards in research on transcutaneous vagus nerve stimulation (Version 2020). *Front. Hum. Neurosci.* 14:568051. doi: 10.3389/fnhum.2020.568051

Fisher, H., Stowell, J., Garcia, R., Sclocco, R., Goldstein, J., Napadow, V., et al. (2018). "Acute effects of respiratory-gated auricular vagal afferent nerve stimulation in the modulation of blood pressure in hypertensive patients," in *2018 Computing in Cardiology Conference (CinC)* (Maastricht), 1–4. doi: 10.22489/CinC.2018.346

Frangos, E., Ellrich, J., and Komisaruk, B. R. (2015). Non-invasive access to the vagus nerve central projections via electrical stimulation of the external ear: fMRI evidence in humans. *Brain Stimul.* 8, 624–636. doi: 10.1016/j.brs.2014.11.018

Frøkjaer, J. B., Bergmann, S., Brock, C., Madzak, A., Farmer, A. D., Ellrich, J., et al. (2016). Modulation of vagal tone enhances gastroduodenal motility and reduces somatic pain sensitivity. *Neurogastroenterol. Motil.* 28, 592–598. doi: 10.1111/nmo.12760

Gallo, J. M. (2010). Pharmacokinetic/ pharmacodynamic-driven drug development: pharmacokinetic/pharmacodynamic-driven drug development. *Mt. Sinai J. Med.* 77, 381–388. doi: 10.1002/msj.20193

Goetz, C. G., Tilley, B. C., Shaftman, S. R., Stebbins, G. T., Fahn, S., Martinez-Martin, P., et al. (2008). Movement disorder society-sponsored revision of the unified Parkinson's disease rating scale (MDS-UPDRS): scale presentation and clinimetric testing results: MDS-UPDRS: clinimetric assessment. *Mov. Disord.* 23, 2129–2170. doi: 10.1002/mds.22340

Goldberger, J. J. (1999). Sympathovagal balance: how should we measure it? *Am. J. Physiol. Heart Circulat. Physiol.* 276, H1273–H1280. doi: 10.1152/ajpheart.1999.276.4.H1273

Grill, W. M. (1999). Modeling the effects of electric fields on nerve fibers: influence of tissue electrical properties. *IEEE Trans. Biomed. Eng.* 46, 918–928. doi: 10.1109/10.775401

Gruber, B., Froeling, M., Leiner, T., and Klomp, D. W. J. (2018). RF coils: a practical guide for nonphysicists: RF Coils. *J. Magn. Reson. Imaging* 48, 590–604. doi: 10.1002/jmri.26187

Gupta, U., Bhatia, S., Garg, A., Sharma, A., and Choudhary, V. (2011). Phase 0 clinical trials in oncology new drug development. *Perspect. Clin. Res.* 2:13. doi: 10.4103/2229-3485.76285

Guru, A., Kumar, N., Ravindra Shanthakumar, S., Patil, J., Nayak Badagabettu, S., Aithal Padur, A., et al. (2015). Anatomical study of the ulnar nerve variations at high humeral level and their possible clinical and diagnostic implications. *Anat. Res. Int.* 2015, 1–4. doi: 10.1155/2015/378063

Handforth, A., DeGiorgio, C. M., Schachter, S. C., Uthman, B. M., Naritoku, D. K., Tecoma, E. S., et al. (1998). Vagus nerve stimulation therapy for partial-onset seizures: a randomized active-control trial. *Neurology* 51, 48–55. doi: 10.1212/WNL.51.1.48

Hasan, A., Wolff-Menzler, C., Pfeiffer, S., Falkai, P., Weidinger, E., Jobst, A., et al. (2015). Transcutaneous noninvasive vagus nerve stimulation (tVNS) in the treatment of schizophrenia: a bicentric randomized controlled pilot study. *Eur. Arch. Psychiatry Clin. Neurosci.* 265, 589–600. doi: 10.1007/s00406-015-0618-9

Hayano, J., and Yuda, E. (2019). Pitfalls of assessment of autonomic function by heart rate variability. *J. Physiol. Anthropol.* 38:3. doi: 10.1186/s40101-019-0193-2

He, W., Wang, X., Shi, H., Shang, H., Li, L., Jing, X., et al. (2012). Auricular acupuncture and vagal regulation. *Evid. Based Complement. Alternat. Med.* 2012:786839. doi: 10.1155/2012/786839

Hein, E., Nowak, M., Kiess, O., Biermann, T., Bayerlein, K., Kornhuber, J., et al. (2013). Auricular transcutaneous electrical nerve stimulation in depressed patients: a randomized controlled pilot study. *J. Neural. Transm.* 120, 821–827. doi: 10.1007/s00702-012-0908-6

Hewitt, J. A., Brown, L. L., Murphy, S. J., Grieder, F., and Silberberg, S. D. (2017). Accelerating biomedical discoveries through rigor and transparency. *ILAR J.* 58, 115–128. doi: 10.1093/ilar/ilx011

Higgins, P. J., Savović, J., Page, J. M., Elbers, G. R., Sterne, A. J., (2019). *Cochrane Training Chapter 8: Assessing Risk of Bias in a Randomized Trial*. Available online at: https://training.cochrane.org/handbook/current/chapter-08 (accessed November 26, 2020).

Howick, J., Friedemann, C., Tsakok, M., Watson, R., Tsakok, T., Thomas, J., et al. (2013). Are treatments more effective than placebos? A systematic review and meta-analysis. *PLoS ONE* 8:e62599. doi: 10.1371/journal.pone.0062599

Huang, F., Dong, J., Kong, J., Wang, H., Meng, H., Spaeth, R. B., et al. (2014). Effect of transcutaneous auricular vagus nerve stimulation on impaired glucose tolerance: a pilot randomized study. *BMC Complement. Altern. Med.* 14:203. doi: 10.1186/1472-6882-14-203

Institute of Medicine. (2014). *Improving and Accelerating Therapeutic Development for Nervous System Disorders: Workshop Summary*. Washington, DC: The National Academies Press.

Janner, H., Klausenitz, C., Gürtler, N., Hahnenkamp, K., and Usichenko, T. I. (2018). Effects of electrical transcutaneous vagus nerve stimulation on the perceived intensity of repetitive painful heat stimuli: a blinded placebo- and sham-controlled randomized crossover investigation. *Anesthesia Analgesia* 126, 2085–2092. doi: 10.1213/ANE.0000000000002820

Jiang, Y., Po, S. S., Amil, F., and Dasari, T. W. (2020). Non-invasive low-level tragus stimulation in cardiovascular diseases. *Arrhythm. Electrophysiol. Rev.* 9, 40–46. doi: 10.15420/aer.2020.01

Johnson, J. I., Decker, S., Zaharevitz, D., Rubinstein, L. V., Venditti, J. M., Schepartz, S., et al. (2001). Relationships between drug activity in NCI preclinical *in vitro* and *in vivo* models and early clinical trials. *Br. J. Cancer* 84, 1424–1431. doi: 10.1054/bjoc.2001.1796

Johnson, M. I., Hajela, V. K., Ashton, C. H., and Thompson, J. W. (1991). The effects of auricular transcutaneous electrical nerve stimulation (TENS) on experimental pain threshold and autonomic function in healthy subjects: *Pain* 46, 337–342. doi: 10.1016/0304-3959(91)90116-F

Juel, J., Brock, C., Olesen, S., Madzak, A., Farmer, A., Aziz, Q., et al. (2017). Acute physiological and electrical accentuation of vagal tone has no effect on pain or gastrointestinal motility in chronic pancreatitis. *JPR Volume* 10, 1347–1355. doi: 10.2147/JPR.S133438

Kahn, L., Mathkour, M., Lee, S. X., Gouveia, E. E., Hanna, J. A., Garces, J., et al. (2019). Long-term outcomes of deep brain stimulation in severe Parkinson's disease utilizing UPDRS III and modified Hoehn and Yahr as a severity scale. *Clin. Neurol. Neurosurg.* 179, 67–73. doi: 10.1016/j.clineuro.2019.02.018

Kang, M., Ragan, B. G., and Park, J.-H. (2008). Issues in outcomes research: an overview of randomization techniques for clinical trials. *J. Athl. Train.* 43, 215–221. doi: 10.4085/1062-6050-43.2.215

Kaniusas, E., Kampusch, S., Tittgemeyer, M., Panetsos, F., Gines, R. F., Papa, M., et al. (2019). Current directions in the auricular vagus nerve stimulation I – a physiological perspective. *Front. Neurosci.* 13:854. doi: 10.3389/fnins.2019.00854

Keute, M., Machetanz, K., Berelidze, L., Guggenberger, R., and Gharabaghi, A. (2021). Neuro-cardiac coupling predicts transcutaneous auricular Vagus Nerve Stimulation effects. *Brain Stimul.* 14, 209–216. doi: 10.1016/j.brs.2021.01.001

Kovacic, K., Hainsworth, K., Sood, M., Chelimsky, G., Unteutsch, R., Nugent, M., et al. (2017). Neurostimulation for abdominal pain-related functional gastrointestinal disorders in adolescents: a randomised, double-blind, sham-controlled trial. *Lancet Gastroenterol. Hepatol.* 2, 727–737. doi: 10.1016/S2468-1253(17)30253-4

Kutlu, N., Özden, A. V., Alptekin, H. K., and Alptekin, J. Ö. (2020). The impact of auricular vagus nerve stimulation on pain and life quality in patients with fibromyalgia syndrome. *Biomed Res. Int.* 2020, 1–10. doi: 10.1155/2020/8656218

Laqua, R., Leutzow, B., Wendt, M., and Usichenko, T. (2014). Transcutaneous vagal nerve stimulation may elicit anti- and pro-nociceptive effects under

experimentally-induced pain — a crossover placebo-controlled investigation. *Autonomic Neurosci.* 185, 120–122. doi: 10.1016/j.autneu.2014.07.008

Li, M., Yao, X., Sun, L., Zhao, L., Xu, W., Zhao, H., et al. (2020). Effects of electroconvulsive therapy on depression and its potential mechanism. *Front. Psychol.* 11:80. doi: 10.3389/fpsyg.2020.00080

LivaNova (2017). *VNS Therapy Dosing Guideline SenTiva*. Available online at: https://vnstherapy.com/healthcare-professionals/sites/vnstherapy.com.healthcare-professionals/files/SenTiva_Dosing_Guide_2019-DIGITAL.PDF (accessed November 26, 2020).

Maharjan, A., Wang, E., Peng, M., and Cakmak, Y. O. (2018). Improvement of olfactory function with high frequency non-invasive auricular electrostimulation in healthy humans. *Front. Neurosci.* 12:225. doi: 10.3389/fnins.2018.00225

Marmerstein, J. T., McCallum, G. A., and Durand, D. M. (2021). Direct measurement of vagal tone in rats does not show correlation to HRV. *Sci. Rep.* 11:1210. doi: 10.1038/s41598-020-79808-8

McNeal, D. R. (1976). Analysis of a model for excitation of myelinated nerve. *IEEE Trans. Biomed. Eng.* BME-23, 329–337. doi: 10.1109/TBME.1976.324593

Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gotzsche, P. C., Devereaux, P. J., et al. (2010). CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 340, c869–c869. doi: 10.1136/bmj.c869

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., and The PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: the prisma statement. *PLoS Med.* 6:e1000097. doi: 10.1371/journal.pmed.1000097

Morimoto, R. (1993). Cells in stress: transcriptional activation of heat shock genes. *Science* 259, 1409–1410. doi: 10.1126/science.8451637

Murray, A. R., Atkinson, L., Mahadi, M. K., Deuchars, S. A., and Deuchars, J. (2016). The strange case of the ear and the heart: The auricular vagus nerve and its influence on cardiac control. *Autonomic Neurosci.* 199, 48–53. doi: 10.1016/j.autneu.2016.06.004

Nair, B. (2019). Clinical trial designs. *Indian Dermatol. Online J.* 10, 193–201. doi: 10.4103/idoj.IDOJ_475_18

Napadow, V., Edwards, R. R., Cahalan, C. M., Mensing, G., Greenbaum, S., Valovska, A., et al. (2012). Evoked pain analgesia in chronic pelvic pain patients using respiratory-gated auricular vagal afferent nerve stimulation. *Pain Med.* 13, 777–789. doi: 10.1111/j.1526-4637.2012.01385.x

Napadow, V., Sclocco, R., and Henderson, L. A. (2019). Brainstem neuroimaging of nociception and pain circuitries: *PAIN Rep.* 4:*E*745. doi: 10.1097/PR9.0000000000000745

Nicolai, E. N., Settell, M. L., Knudsen, B. E., McConico, A. L., Gosink, B. A., Trevathan, J. K., et al. (2020). Sources of off-target effects of vagus nerve stimulation using the helical clinical lead in domestic pigs. *J. Neural. Eng.* 17:046017. doi: 10.1088/1741-2552/ab9db8

Ottaviani, M. M., Wright, L., Dawood, T., and Macefield, V. G. (2020). *In vivo* recordings from the human vagus nerve using ultrasound-guided microneurography. *J. Physiol.* 598, 3569–3576. doi: 10.1113/JP280077

Pagnin, D., de Queiroz, V., Pini, S., and Cassano, G. B. (2004). Efficacy of ECT in depression: a meta-analytic review. *J. ECT* 20:8. doi: 10.1097/00124509-200403000-00004

Parameswaran, N., and Patial, S. (2010). Tumor necrosis factor-α signaling in macrophages. *Crit. Rev. Eukar. Gene Expr.* 20, 87–103. doi: 10.1615/CritRevEukarGeneExpr.v20.i2.10

Paul, S. M., Mytelka, D. S., Dunwiddie, C. T., Persinger, C. C., Munos, B. H., Lindborg, S. R., et al. (2010). How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat. Rev. Drug Discov.* 9, 203–214. doi: 10.1038/nrd3078

Payne, N. A., and Prudic, J. (2009). Electroconvulsive therapy: Part I. a perspective on the evolution and current practice of ECT. *J. Psychiatr. Pract.* 15, 346–368. doi: 10.1097/01.pra.0000361277.65468.ef

Peuker, E. T., and Filler, T. J. (2002). The nerve supply of the human auricle. *Clin. Anat.* 15, 35–37. doi: 10.1002/ca.1089

Poulsen, A. H., Tigerholm, J., Meijs, S., Andersen, O. K., and Mørch, C. D. (2020). Comparison of existing electrode designs for preferential activation of cutaneous nociceptors. *J. Neural. Eng.* 17:036026. doi: 10.1088/1741-2552/ab85b1

Provitera, V., Nolano, M., Pagano, A., Caporaso, G., Stancanelli, A., and Santoro, L. (2007). Myelinated nerve endings in human skin. *Muscle Nerve.* 35, 767–775. doi: 10.1002/mus.20771

Rattay, F. (1999). The basic mechanism for the electrical stimulation of the nervous system. *Neuroscience* 89, 335–346. doi: 10.1016/S0306-4522(98)00330-3

Reynolds, P. J., Fan, W., and Andresen, M. C. (2006). Capsaicin- resistant arterial baroreceptors. *J. Negat. Results Biomed.* 5:6. doi: 10.1186/1477-5751-5-6

Ritchie, M. K., Wilson, C. A., Grose, B. W., Ranganathan, P., Howell, S. M., and Ellison, M. B. (2016). Ultrasound-guided greater auricular nerve block as sole anesthetic for ear surgery. *Clin. Pract.* 6:856. doi: 10.4081/cp.2016.856

Robbins, M. S., and Lipton, R. B. (2017). Transcutaneous and percutaneous neurostimulation for headache disorders. *Headache* 57, 4–13. doi: 10.1111/head.12829

Rong, P., Liu, A., Zhang, J., Wang, Y., He, W., Yang, A., et al. (2014). Transcutaneous vagus nerve stimulation for refractory epilepsy: a randomized controlled trial. *Clin. Sci.* doi: 10.1042/CS20130518. [Epub ahead of print].

Sabbah, H. N., Ilsar, I., Zaretsky, A., Rastogi, S., Wang, M., and Gupta, R. C. (2011). Vagus nerve stimulation in experimental heart failure. *Heart Fail. Rev.* 16, 171–178. doi: 10.1007/s10741-010-9209-z

Salama, M., Akan, A., and Mueller, M. R. (2020). Transcutaneous stimulation of auricular branch of the vagus nerve attenuates the acute inflammatory response after lung lobectomy. *World J. Surg.* 44, 3167–3174. doi: 10.1007/s00268-020-05543-w

Seagard, J. L., van Brederode, J. F., Dean, C., Hopp, F. A., Gallenberg, L. A., and Kampine, J. P. (1990). Firing characteristics of single-fiber carotid sinus baroreceptors. *Circ. Res.* 66, 1499–1509. doi: 10.1161/01.RES.66.6.1499

Shen, D., and Lu, Z. (2006). *Estimate Carryover Effect in Clinical Trial Crossover Designs*, 7. Available online at: https://www.lexjansen.com/pharmasug/2006/Posters/PO16.pdf

Stavrakis, S., Humphrey, M. B., Scherlag, B. J., Hu, Y., Jackman, W. M., Nakagawa, H., et al. (2015). Low-level transcutaneous electrical vagus nerve stimulation suppresses atrial fibrillation. *J. Am. Coll. Cardiol.* 65, 867–875. doi: 10.1016/j.jacc.2014.12.026

Stavrakis, S., Stoner, J. A., Humphrey, M. B., Morris, L., Filiberti, A., Reynolds, J. C., et al. (2020). TREAT AF (Transcutaneous Electrical Vagus Nerve Stimulation to Suppress Atrial Fibrillation). *JACC Clin. Electrophysiol.* 6, 282–291. doi: 10.1016/j.jacep.2019.11.008

Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., et al. (2019). RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 366:l4898. doi: 10.1136/bmj.l4898

Stoddard, M. B., Pinto, V., Keiser, P. B., and Zollinger, W. (2010). Evaluation of a whole-blood cytokine release assay for use in measuring endotoxin activity of Group B neisseria meningitidis vaccines made from Lipid A acylation mutants. *CVI* 17, 98–107. doi: 10.1128/CVI.00342-09

Stowell, J., Garcia, R., Staley, R., Sclocco, R., Fisher, H., Napadow, V., et al. (2019). "Dose-optimization of Respiratory-Gated Auricular Vagal Afferent Nerve Stimulation (RAVANS) for blood pressure modulation in hypertensive patients," in *2019 Computing in Cardiology (CinC)* (Singapore), 1–4. doi: 10.22489/CinC.2019.098

Straube, A., Ellrich, J., Eren, O., Blum, B., and Ruscheweyh, R. (2015). Treatment of chronic migraine with transcutaneous stimulation of the auricular branch of the vagal nerve (auricular t-VNS): a randomized, monocentric clinical trial. *J. Headache Pain* 16:63. doi: 10.1186/s10194-015-0543-3

Thurm, C. W., and Halsey, J. F. (2005). "Measurement of cytokine production using whole blood," in *Current Protocols in Immunology*, eds. J. E. Coligan, B. E. Bierer, D. H. Margulies, E. M. Shevach, and W. *Strober* (Hoboken, NJ: John Wiley & Sons, Inc.), *im*0718bs66.

Tobaldini, E., Toschi-Dias, E., Appratto de Souza, L., Rabello Casali, K., Vicenzi, M., Sandrone, G., et al. (2019). Cardiac and peripheral autonomic responses to orthostatic stress during transcutaneous vagus nerve stimulation in healthy subjects. *JCM* 8:496. doi: 10.3390/jcm8040496

Tran, N., Asad, Z., Elkholey, K., Scherlag, B. J., Po, S. S., and Stavrakis, S. (2019). Autonomic neuromodulation acutely ameliorates left ventricular strain in humans. *J. Cardiovasc. Trans. Res.* 12, 221–230. doi: 10.1007/s12265-018-9853-6

Tsaava, T., Datta-Chaudhuri, T., Addorisio, M. E., Masi, E. B., Silverman, H. A., Newman, J. E., et al. (2020). Specific vagus nerve stimulation parameters alter serum cytokine levels in the absence of inflammation. *Bioelectron Med.* 6:8. doi: 10.1186/s42234-020-00042-8

Turner, H. M. III, and Bernard, R. M. (2006). Calculating and synthesizing effect sizes. *CICSD* 33, 42–55. doi: 10.1044/cicsd_33_S_42

Tutar, B., Atar, S., Berkiten, G., Üstün, O., Kumral, T. L., and Uyar, Y. (2020). The effect of transcutaneous electrical nerve stimulation (TENS) on chronic subjective tinnitus. *Am. J. Otolaryngol.* 41:102326. doi: 10.1016/j.amjoto.2019.102326

Usami, K., Kawai, K., Sonoo, M., and Saito, N. (2013). Scalp-recorded evoked potentials as a marker for afferent nerve impulse in clinical vagus nerve stimulation. *Brain Stimul.* 6, 615–623. doi: 10.1016/j.brs.2012.09.007

Wagatsuma, A., Okuyama, T., Sun, C., Smith, L. M., Abe, K., and Tonegawa, S. (2018). Locus coeruleus input to hippocampal CA3 drives single-trial learning of a novel context. *Proc. Natl. Acad. Sci. U.S.A.* 115, E310–E316. doi: 10.1073/pnas.1714082115

Watanabe, K., Tubbs, R. S., Satoh, S., Zomorodi, A. R., Liedtke, W., Labidi, M., et al. (2016). Isolated deep ear canal pain: possible role of auricular branch of vagus nerve—case illustrations with cadaveric correlation. *World Neurosurg.* 96, 293–301. doi: 10.1016/j.wneu.2016.08.102

Wellmark (2018). *Vagus Nerve Stimulation (VNS) and Vagal Blocking Therapy*. Available online at: https://www.wellmark.com/Provider/MedpoliciesAndAuthorizations/MedicalPolicies/policies/Vagus_Nerve_Stimulation.aspx?fbclid=IwAR0Bq8iogosvLVR7p3xBW7J5fIziGxbyQk-2MnRRKADoQcJ0qJFotbOk_dM (accessed November 26, 2020).

Wolf, V., Kühnel, A., Teckentrup, V., Koenig, J., and Kroemer, N. B. (2021). Does non-invasive vagus nerve stimulation affect heart rate variability? A living and interactive Bayesian meta-analysis. *bioRxiv [preprint]*. doi: 10.1101/2021.01.18.426704

Wright, B. J., O'Brien, S., Hazi, A., and Kent, S. (2015). Increased systolic blood pressure reactivity to acute stress is related with better self-reported health. *Sci. Rep.* 4:6882. doi: 10.1038/srep06882

Yang, B., Pham, T., Goldbach-Mansky, R., and Gadina, M. (2011). Accurate and simple measurement of the pro-inflammatory cytokine IL-1β using a whole blood stimulation assay. *J. Vis. Exp.* 49:e2662. doi: 10.3791/2662

Yap, J. Y. Y., Keatch, C., Lambert, E., Woods, W., Stoddart, P. R., and Kameneva, T. (2020). Critical review of transcutaneous vagus nerve stimulation: challenges for translation to clinical practice. *Front. Neurosci.* 14:284. doi: 10.3389/fnins.2020.00284

Yoo, P. B., Lubock, N. B., Hincapie, J. G., Ruble, S. B., Hamann, J. J., and Grill, W. M. (2013). High-resolution measurement of electrically-evoked vagus nerve activity in the anesthetized dog. *J. Neural Eng.* 10:026003. doi: 10.1088/1741-2560/10/2/026003

Yu, L., Huang, B., Po, S. S., Tan, T., Wang, M., Zhou, L., et al. (2017). Low-level tragus stimulation for the treatment of ischemia and reperfusion injury in patients with ST-Segment elevation myocardial infarction. *JACC Cardiovasc. Intervent.* 10, 1511–1520. doi: 10.1016/j.jcin.2017.04.036

Zamotrinsky, A. V., Kondratiev, B., and de Jong, J. W. (2001). Vagal neurostimulation in patients with coronary artery disease. *Autonomic Neurosci.* 88, 109–116. doi: 10.1016/S1566-0702(01)00227-2

Conflict of Interest: EL is an employee of LivaNova PLC. JW and KL are scientific board members and have stock interests in NeuroOne Medical Inc., a company developing next generation epilepsy monitoring devices. JW also has an equity interest in NeuroNexus technology Inc., a company that supplies electrophysiology equipment and multichannel probes to the neuroscience research community. KL is also a paid member of the scientific advisory board of Cala Health, Blackfynn, Abbott and Battelle. KL also is a paid consultant for Galvani and Boston Scientific. KL is a consultant to and co-founder of Neuronoff Inc. None of these associations are directly relevant to the work presented in this manuscript.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.