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

Nishant Verma[1], Jonah D Mudge[1], Maïsha S E Kasole[1], Rex C Chen[1], Stephan L Blanz[1], James K Trevathan[1], Kip A Ludwig[1,3]

[1]Department of Biomedical Engineering, University of Wisconsin – Madison, Madison, WI, USA
[2]Department of Neurosurgery, University of Wisconsin – Madison, Madison, WI, USA

ABSTRACT

Background: The auricular branch of the vagus runs superficial to the surface of the skin, making it a favorable target for non-invasive techniques to modulate vagal activity. For this reason, there have been many early-stage clinical trials on a diverse range of conditions. Unfortunately, often with conflicting results.

Methods: To investigate the conflicting results, we conducted a systematic review of auricular vagus nerve stimulation (aVNS) randomized controlled trials (RCTs) using the established Cochrane Risk of Bias tool as a framework. The Risk of Bias tool is intended to identify deviations from an ideal RCT that may cause the effect of an intervention to be overestimated or underestimated. As is common for early-stage studies, the majority of aVNS studies were assessed as having 'some' or 'high' risk of bias, which makes interpreting their results in a broader context problematic.

Results: The reported trial outcomes were qualitatively synthesized across studies. There is evidence of a modest decrease in HR during higher stimulation current amplitudes. Findings on heart rate variability (HRV) conflicted between studies and were hindered by trial design including inappropriate washout periods and multiple methods used to quantify HRV. There is early-stage evidence to suggest aVNS may reduce circulating levels or 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. aVNS showed preliminary evidence of ameliorating pathological pain but not induced pain.

Discussion: Drawing on the fundamentals of neuromodulation, we establish the need for direct measures of neural target engagement in aVNS. Firstly, for the optimization of electrode design, placement, and stimulation waveform parameters to improve on-target engagement and minimize off-target engagement. Secondly, direct measures of target engagement, along with consistent evaluation of the double blind, must be used to improve the design of controls in the long term - a major source of concern identified in the Cochrane analysis. Lastly, we list common improvements for the reporting of results that can be addressed in the short term.

Conclusion: The need for direct measures of neural target engagement and consistent evaluation of the double blind is applicable to other paresthesia-inducing neuromodulation therapies and their control designs. We intend for this review to contribute to the successful translation of neuromodulation therapies such as aVNS.

1. INTRODUCTION

Electrical stimulation of the nervous system, commonly known as neuromodulation, aims to manipulate 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 FDA approved 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 versus efficacy profile of the therapy, VNS is currently a last line therapy after patients have been shown refractory to at least two anti-epileptic drugs dosed appropriately. 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 create electric fields 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 to efficacy profile. Fig. 1 shows innervation of the auricle by four major nerve branches and several electrode designs to deliver electrical stimulation at the ear.

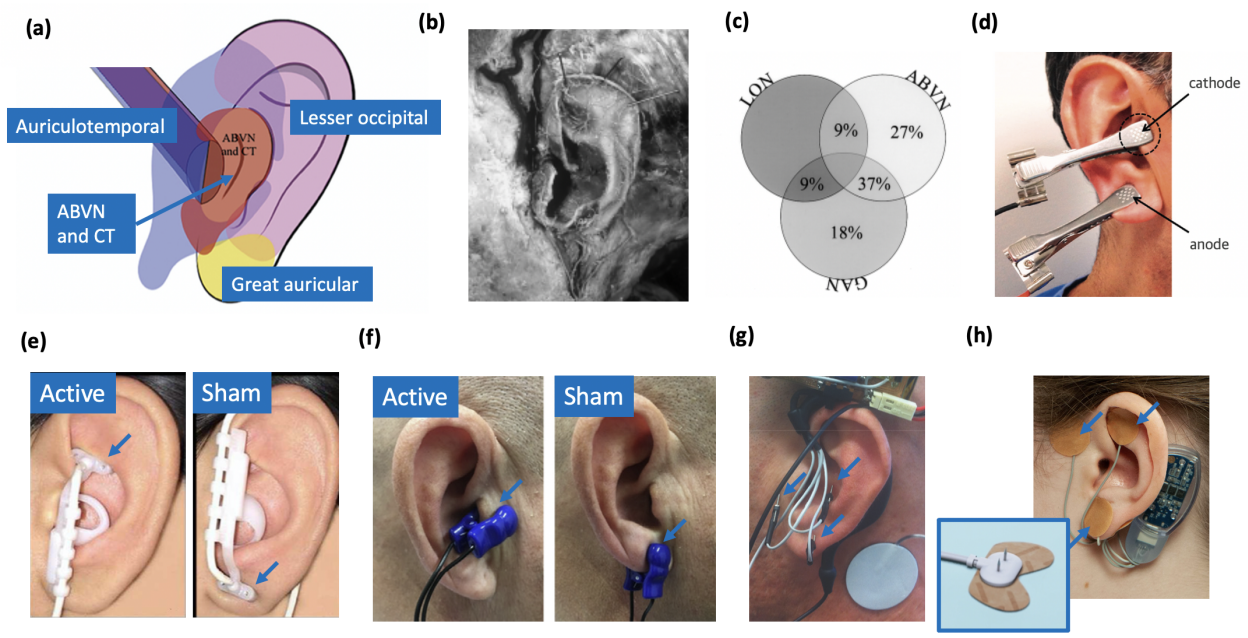

Figure 1: **(a)** Innervation of the auricle by five nerves (Watanabe et al., 2016): auricular branch of the vagus nerve (ABVN), chorda tympani (CT) from the facial nerve, auriculotemporal nerve originating from the mandibular branch of the trigeminal nerve, great auricular nerve, and lesser occipital nerve. **(b)** The only primary study (Peuker and Filler, 2002) mapping the innervation of the human auricle in 7 cadavers

by dissection. 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 (LON) (Peuker and Filler, 2002). Note microdissection may have missed the finer nerve branches and hence may be underestimating the overlapping regions of innervation. Peuker and Filler note the higher density of nerve fibers in the human auricle, which is suggestive of their increased overlap in innervation. 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)** Cerbomed (Erlangen, Germany) transcutaneously targeting cymba concha in the intervention group and ear lobe in the sham group. **(f)** Parasym (London, UK) transcutaneously targeting tragus in the intervention group and ear lobe in the sham group. (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 intervention group.

Given aVNS can be implemented non-invasively 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, the quality of clinical evidence is not strong as the trial results are often conflicting for the same physiological outcome measure.

The highest level of clinical evidence is a systematic review of multiple high-quality double-blinded, randomized, and controlled clinical trials (RCT) with narrow confidence intervals, each homogeneously supporting the efficacy and safety of a therapy for a specific desired clinical outcome, according to the widely utilized Oxford Centre for Evidence Based Medicine's (CEBM) Levels of Clinical Evidence Scale (Centre for Evidence-Based Medicine, 2009). However, reaching this level of evidence is a costly and time-consuming endeavor. Years of lower quality precursor clinical studies with smaller numbers of patients that require less resources to perform are needed to identify the most efficacious embodiment of the therapy that can be safely delivered to properly design more definitive clinical studies. The field of aVNS, being relatively new clinically, is understandably still in these early phases of clinical development.

Here, we perform a systematic review of aVNS RCTs with two primary goals: 1) to provide an accessible framework to 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 – a common framework (Sterne et al., 2019) previously used to identify risk of bias in RCT studies of epidural spinal cord stimulation (Duarte et al., 2020) 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 then 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 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 clinical evidence for its safety and efficacy for any specific physiological outcome.

2. METHODS

2.1 Search method

The literature search was designed to identify reports of clinical RCTs testing aVNS. Two databases were searched: PubMed and Scopus (includes MEDLINE and Embase databases). Two search strategies were used. The first, a combination of two concepts: aVNS and RCT. The second search strategy focused on commercial aVNS devices and their manufacturers. Complete search strings for both strategies are available in supplementary material. The search was last updated in July 2020. In addition, citations of all selected studies were hand searched to identify additional studies. The citations of relevant reviews (Murray et al., 2016; Yap et al., 2020) were also searched to identify aVNS clinical RCTs. Through these citation searches, 5 additional studies were found.

With the intention 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 acubeads. Studies were considered TCM studies if acupoints were used to justify location of stimulation or published in a TCM journal. This is captured in Fig. 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.

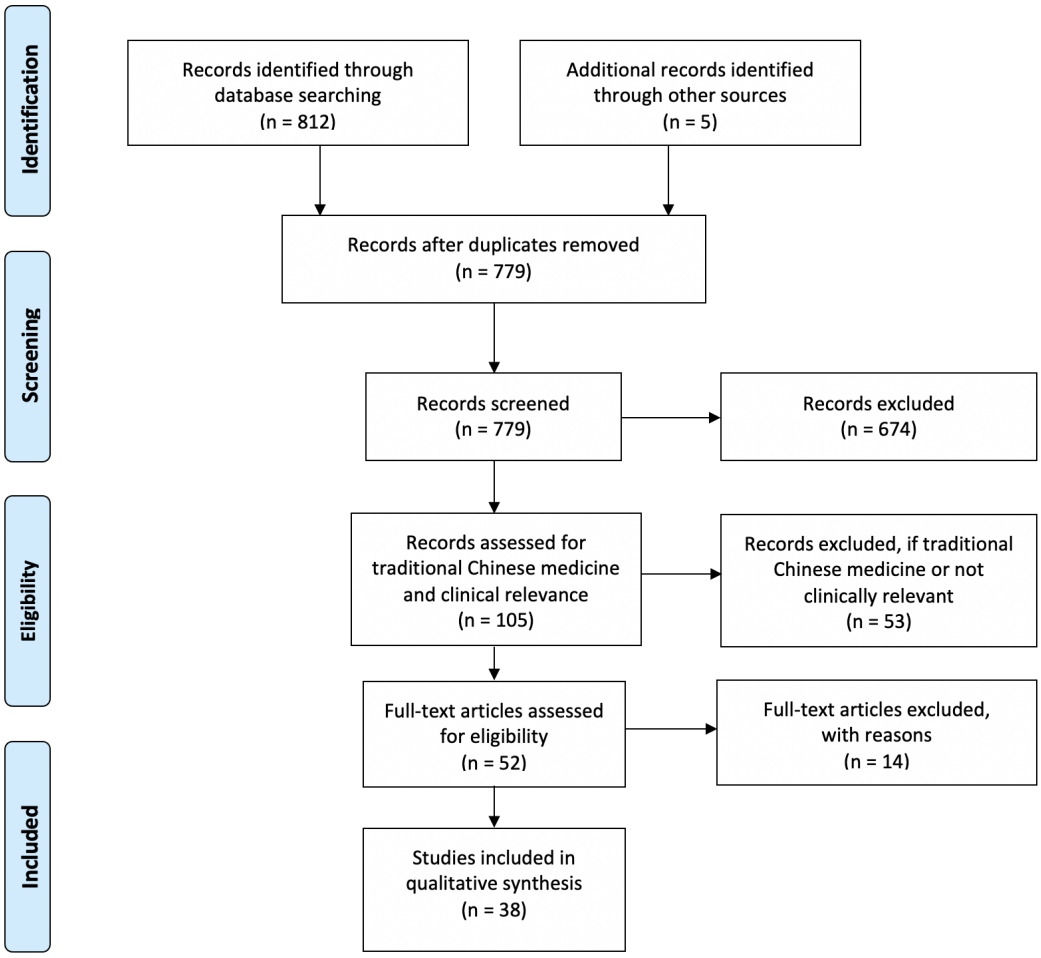

Figure 2: Adapted PRISMA Flow Chart.

2.2 Inclusion and exclusion criteria

Publications were included if their design randomized and controlled at least one trial; non-RCT portions of included publications were not analyzed. Only publications 1991 and after were included, with the 1991 cutoff marking the first time autonomic activity measurement was attempted during auricular stimulation in (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, fMRI, EEG, and somatosensory evoked potential are secondary physiological measures of target engagement. Although they are useful to study the mechanisms behind aVNS, they do not have well-established links to clinically significant outcomes. In comparison, heart rate variability (HRV), a measure of sympathovagal tone, is considered a measurement of direct clinical significance as sympathovagal imbalance is related to several disease states. Similarly, studies focused on cognitive neuroscience topics, such as behavior, learning, fear extinction, or executive functions were excluded. On

the other hand, psychological studies addressing addiction, depression, pain, and stress were included in the final qualitative review as they had direct clinical significance.

2.3 Cochrane risk of bias 2.0 tool to assess quality of evidence

To evaluate the quality of evidence, we used the Cochrane risk of bias 2.0 tool (RoB), an established tool to assess bias in randomized clinical trials (Sterne et al., 2019), which has been cited over 40,000 times in Google Scholar. 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 for each study evaluated have been included in supplementary material. At several instances, the suggested algorithm was overridden by the reviewer with justification annotated on the individual rubrics found in supplementary material. 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 was 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. 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 have an equal number of subjects being presented with sham or therapy first (Nair, 2019). Studies must perform an assessment of baseline differences between the intervention groups for known nuisance variables that may confound interpretation of the outcome results.

Bias due to deviations from intended interventions – During the implementation of a clinical trial, it is foreseeable to have several subjects that were randomized to a given group not receive the intended intervention or for blinding to be compromised. Compromised blinding is especially pertinent in aVNS studies where there may be a difference in paresthesia or electrode location between the intervention and control group that clues subjects or investigators to become aware of the treatment or control arm assignments. This violates the principle of double blinding and deviates from the intended intervention.

In order to receive a low risk of bias score, the study must minimize and account 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, it is common not to be able to record all intended outcome measures on all subjects. 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. 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 were 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' or 'high' risk of bias depending on severity.

Bias in measurement of the outcome – How an outcome is 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 trials results is to differentiate if the data is "exploratory" versus "confirmatory" (Hewitt et al., 2017). Exploratory research is used to generate hypotheses and models for testing, often includes analyses that are done at least in part retrospectively and is 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 and only reported 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 and introduces 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.

2.4 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 hypothesis and hypothesized mechanism of action if mentioned in the paper. We also included notes on whether implantable VNS had achieved this effect in humans. Study design encapsulated ideas including 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 vs parallel), stimulation parameters, randomization, baseline comparison, and washout periods. Lastly, study results included: primary and secondary endpoints, adverse effects, excluded and missing data, as well as 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 group, worsening of any subjects, control group effect size, and clinical relevance and significance of findings. Where sufficient information was reported, standardized effect size using Hedges' g (Turner and Bernard, 2006) was calculated.

Each publication had a primary reviewer, concerns were discussed in a group, and an additional secondary reviewer went through all papers. If crucial basic information was missing (e.g. which ear was stimulated, electrode used, etc.), 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.

3. RESULTS

A total of 38 articles were reviewed totaling 41 RCTs – two each in the publications (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 included RCT studies, again with the goal of identifying 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.

A user sortable table summarizing every reviewed studies design and result features has been included as an excel file in supplementary material, to allow the readers to view based on their own features of interest. Design and result features have been reduced to common keywords in this spreadsheet to facilitate user sorting; however, this means specific details of outcome measures have been reduced to general categories.

3.1 aVNS Electrode Designs, Configurations, and Stimulation Parameters across Studies

What is immediately evident is that implementation of both active and sham varied greatly across studies. Table 1 below details the electrode design, configuration (monopolar/bipolar), target location and stimulation parameters for the active arm of the study and the control arm if present. Fig. 3 shows a box plot presenting the distribution of pulse widths, pulse amplitudes, and frequencies used across studies. In the active arm, the interquartile range (IQR) of pulse amplitude was 0.5 - 5 mA, pulse width was 250 - 500 µs, and frequency of stimulation was 20 - 25 Hz. It is notable that the commonly used aVNS waveform parameters are highly similar to the typical parameters used for the stimulation of the cervical vagus, at a pulsewidth of 250 us 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

This large variation in waveform parameters is indicative of the exploratory nature of aVNS studies and highlights the difficulties in comparisons across studies when 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 are discussed later in this section.

Table 1 presents neuromodulation device parameters and is organized by indication type, then primary endpoints, then RoB score. Data is organized by the following columns:
<u>Primary endpoints:</u> The main result of clinical interest. Studies are grouped by endpoints measured within their respective indications. For example, within the cardiac disease indication, the studies investigating inflammatory cytokine levels are located adjacent to each other.
<u>Active waveform & location:</u> frequency, pulse width, on/off cycle duration (duty cycle), and stimulation location.
<u>Active amplitude & electrode type:</u> current amplitude and the titration method used to reach that amplitude, as well as 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 cue that investigators used (Badran et al., 2019) to determine the stimulation amplitude for each subject, described as follows:

> *Sub-sensory titration:* stimulation was purposely kept just below the threshold of paresthesia sensation.
> *First sensory titration*: patient 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 patient feels a painful sensation.
> *Set stimulation*: fixed amplitude across all subjects – resulting in different levels of sensation due to the individual's unique anatomy and perception.

<u>Control:</u> control group stimulation amplitude, control design (sham vs. placebo), and stimulation location. Following (Duarte et al., 2020), we defined sham as when the control group's experience from the subject perspective is identical to the active group – including paresthesia and device indications and

operating behavior. Conversely, placebo control is defined as when the control group does not experience the same paresthesia, device operation, or clinician interaction.

Table 1: Electrode design, configuration, target location, and stimulation parameters.

Author (y) Indication	Primary Endpoints	Active waveform* & location	Active amplitude & electrode type**	Control
Cardiac				
Andreas (2019) postop atrial fibrillation	ECG post atrial fibrillation	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 (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
Salama (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
Stavrakis (2020) Atrial Fibrillation	Atrial fibrillation burden	20 Hz, 200 µs PW, on/off cycle NR Right tragus	Strong sensory (1 mA below mild discomfort threshold, mean 16.8 mA) Parasym clip electrodes	Sham: diff amplitude (mean 19.9 mA), diff location (right earlobe)
Badran (2018) Heart rate	ΔHR during stimulation.	9 waveforms (1,10,25 Hz)x(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 (2016) Coronary Insufficiency and left ventricular dysfunction	Heart rate and 6 minute walk distance.	Frequency NR, PW NR, on/off cycle NR Side NR concha	Titration method NR. Mean 0.05 mA for 1 min (procedure 1), 0.05 mA for 5 min (procedure 2 and 3), 0.08 mA for 10 min (procedures 4 and 5), 0.1 mA for 15 min (procedures 6 and 7), 0.12 mA for 20 min (procedure 8 and 9), and 0.15 mA for 30 min (procedure 10 to 15)	Placebo: same location (concha)
Tobaldini (2019) Orthostatic stress	ΔHR, LF/HF, systolic arterial BP variance, R-R interval pattern, respiratory rate	25 Hz, 200 ms PW, on/off cycle NR, phase NR Left cymba concha	First sensory (1-6 mA) NEMOS electrodes	Placebo: No stimulation, same location.

Fisher (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 (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 (2001) coronary artery disease	HR, BP, LV diastolic function, LV filling	3 Hz, 1500 µs PW, on/off cycle NR Bilateral cavum concha	Titration method NR (0.2-1.25 mA) Acupuncture needles	No intervention
Antonino (2017) Spontaneous cardiac baroreflex sensitivity (cBRS)	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 electrode	1 Placebo 1 Sham: same waveform (mA NR), diff location (bilateral earlobe)
Bretherton (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 & outer tragus)
Clancy (2014) HRV & Sympathetic Activity	HRV (LF/HF	30 Hz, 200 µs PW, continuous, Side NR inner & outer tragus	Sensory threshold (10-50 mA device range), V-TENS Plus with modified surface electrodes	Placebo: no current, same location
De Couck (2017) HRV	ECG with HRV	25Hz, 250 µs PW, on/off cycle NR Bilateral cymba concha	Strong sensory (mean ~0.7 mA), aVNS Cerbomed, ball contact electrodes	Placebo: no current, same location
Borges (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 electrodes	Sham: same set stimulation, different strong sensory stimulation (2.76 ± 1.01 mA), diff location (earlobe)
Tran (2019) Left Ventricular Strain and Autonomic Tone	LV global longitudinal strain (GLS)	20 Hz, 200 µs PW, on/off cycle NR Right tragus	Strong sensory (1mA below discomfort threshold, mean 22.6 mA active) Parasym earclip electrodes	Sham: diff current (mean 21.8 mA), diff location (right earlobe)
Yu (2017) Myocardial Ischemia Reperfusion Injury	Ventricular premature beat (VPB) 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 (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 electrode	Sham: 1 Hz with 219 ± 34 min duration and 1.02 ± 0.83 mA. same location (left concha)
Aihua (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 (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), TENS sm; Suzhou Medical Appliance Co. custom carbon-silicone electrodes	Sham: same waveform, diff location; contacts at scapha and antihelical fold
Pain				
Straube (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 electrodes	Sham: 1 Hz frequency, same location
Kutlu (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) specially designed surface electrodes	No intervention
Busch (2013) Pain	Thermal, mechanical, and pressure pain thresholds	25 Hz, 250 µs PW, on/off cycle NR Left concha	First sensory (mean 1.6 ± 1.5 mA), STV02 (Cerbomed) electrodes	Placebo: no current, same location
Janner (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, diff waveform (no current)
Kovacic (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 h 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), same location

Laqua (2014) Pain	Pain Threshold	Alternating between 2 Hz and 10 Hz, 200 µs PW, on/off cycle NR Bilateral concha (anoce) and mastoid (cathode)	Strong sensory (mA NR), silver EEG electrode at anode, PECG electrode at cathode	Placebo: subsensory stim, same location
Johnson (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)
Napadow (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) Cefar Acus II press-tack electrodes	Sham: same waveform, diff location (left earlobe)
Psychological				
Hein (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 study 1 or 130 µA study 2), TENS-2000 (study 1) or TENS-1000 (study 2) Auri-Stim Medical, Inc. with corresponding electrodes (4 contacts)	Placebo: no current, same location
Hasan (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
Burger (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 titanium electrodes on silicon earplug	Sham: same waveform, diff location (left earlobe)
Others				
Addorisio (2019) Rheumatoid Arthritis	Endotoxin induced interleukin (IL)-6, IL-1β, and TNF	~160 Hz vibrations, Right cymba concha	None; vibrotactile device	Sham: same waveform diff location (right gastrocnemius)
Huang (2014) Impaired glucose tolerance	2-hr plasma glucose levels	20 Hz, <= 1 ms PW, phase NR Side NR concha	Set (1 mA, intensity adjusted based on tolerance of patients), TENS-200 (Huatuo, Suzhou), electrodes similar to Rong (2014)	Sham: same waveform, diff location (superior scapha)
Juel (2017) Pain & Gastrointestinal Motility	ECG and PPG, 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 electrodes	Sham: same waveform, diff location (left earlobe)
Frøkjær (2016) Gastroduodenal motility and pain threshold	ECG and PPG, pain threshold, cold pressor test, and drink test (ultrasound imaging)	30 Hz, 250 µs PW, continuous Left concha	Strong sensory (1.07 mA), NEMOS electrode	Sham: diff amplitude (mean 1.57 mA), diff location (left earlobe)

Cakmak (2017) Parkinson's	Motor examination (part III of the unified Parkinson's disease rating scale)	130 Hz, 100 µs pulse width, continuous, biphasic Ipsilateral ear to dominant motor symptoms in tragicus, antitragicus, helicis minor muscles	Strong sensory (100-130 µA), self-designed stimulator, percutaneous electrodes	Placebo: no current, 2 percutaneous electrodes in upper helix
Maharjan (2018) olfactory function	Odor threshold test and supra-threshold test	80 Hz or 10 Hz, 180 µs PW, on/off cycle NR Left internal (concha) & external ear	Pain threshold (0.1 - 10 mA) Electrode NR	Sham: same waveform, diff location (left earlobe)
Tutar (2020) Tinnitus	Tinnitus Handicap Inventory and Depression Anxiety Stress Scales	200 Hz, 1 ms PW, on/off cycle NR Uni/bilateral cymba concha	First sensory (10-30 mA), silver electrodes (Provile TENS stimulator) (monopolar)	Placebo: same location
*Monophasic if phase not otherwise specified **Bipolar if electrode polarity not otherwise specified. Configuration considered monopolar only if the return electrode is distant enough and to not activate the target region.				

Fig. 3 presents the data from table 2 in box plot form. Illustrated here are the interquartile range, maximum, minimum, and median of each parameter's values, including extreme cases. Studies that report ranges for parameters are included as a single value representing the average of the boundaries of that range.

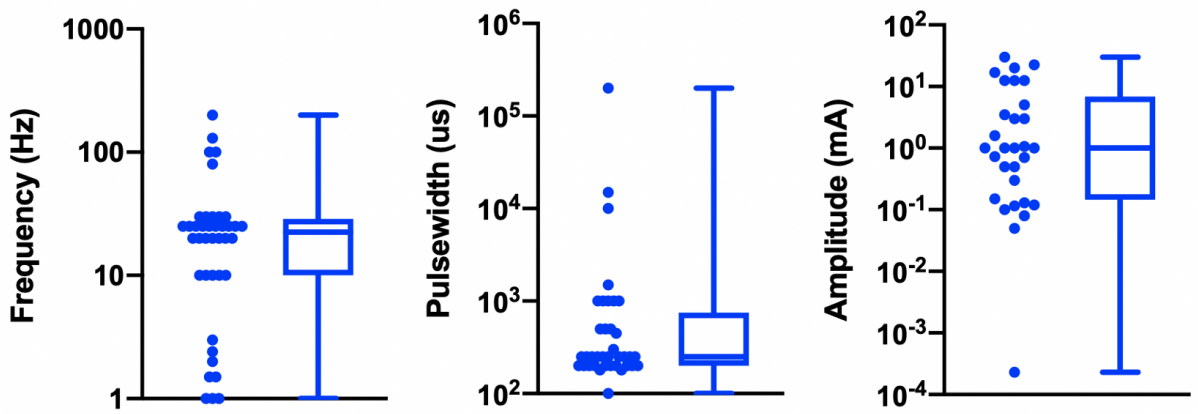

Figure 3: Visualization of reported stimulation waveform parameters in 41 reviewed aVNS RCTs.

As noted previously, frequency and pulse width are grouped around 0-35 Hz and 200-1000 µs, respectively. Stimulation amplitude values are concentrated around 1-5 mA. Further statistics on these parameters are presented in table x.

Table 2: Summary statistics of stimulation parameters

Stimulation Parameter	Median	IQR
Amplitude (mA)	1.0	0.19 - 4.7
Pulse Width (us)	250	200 - 500
Frequency (Hz)	22.5	10.0 - 26.3

While the large variation in pulse width and frequency parameters can be simply explained as choices made by investigators, the sources of the incongruities in stimulation amplitude are not as trivial. It is important to consider differences in electrode design, material, area, and stimulation polarity when comparing stimulation amplitude parameters across studies. This is because electrode geometry and contact area have the potential to impact target engagement of underlying nerves. Furthermore, nerve activation is a function of current density at the stimulating electrode, and it is impossible to accurately estimate current density without first knowing electrode geometry. Another source of variability arises from the fact that different studies used different titration methods. Stimulation amplitude was calibrated to different levels of paresthesia, and the level of paresthesia a subject feels is a direct result of current density magnitude, which is once again related closely to 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 4.2 on measuring direct target engagement).

3.2 aVNS Study Designs

The way in which studies were conducted and organized also varied greatly. Of the 41 RCTs reviewed, 20 used a crossover design while 21 opted for a parallel design. For control group design, 19 studies used 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 4.3.

Table 3 presents study design and is organized by indication type, then primary endpoints, then RoB score. Data is organized by 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 disease indication, the studies investigating inflammatory cytokine levels are located adjacent to each other.
Subjects analyzed: sample size and whether subjects were healthy or part of the disease group.
Control: Following (Duarte et al., 2020), we defined sham as when the control group's experience from the subject perspective is identical to the active group – including paresthesia and device indications and operating behavior. Conversely, placebo control is defined as when the control group does not experience the same paresthesia, device operation, or clinician interaction.
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 study arms, with different arms

receiving different treatments from each other. Studies were classified as crossover if each subject group received every treatment, but in a different order from the other 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 ultimately 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.

Table 3: Study designs sorted by indication type, primary endpoints, and RoB score.

Author (y) Indication	**Primary Endpoints**	**Subjects analyzed**	**Control**	**Design**
Cardiac				
Andreas (2019) postop atrial fibrillation	ECG post atrial fibrillation	40, patients undergoing cardiac surgery	Sham	Parallel, acute (1 hr)
Stavrakis (2015) Atrial fibrillation	Atrial fibrillation cycle length and duration, TNF-a, CRP	40, paroxysmal AF ablation patients	Placebo	Parallel, acute (1 hr)
Salama (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 interventio n	Parallel, acute (5 days)
Stavrakis (2020) Atrial Fibrillation	Atrial fibrillation burden	53, diagnosed with Paroxysmal AF	Sham	Parallel, chronic (6 months)
Badran (2018) Heart rate	ΔHR during stimulation.	35, Healthy	Sham	Crossover, acute (duration NR)
Afanasiev (2016) Coronary Insufficiency and left ventricular dysfunction	Heart rate and 6 minute walk distance.	70, coronary insufficiency/LV dysfunction patients	Placebo	Parallel. acute (~1 hr)
Tobaldini (2019) Orthostatic stress	ΔHR, LF/HF, systolic arterial BP variance, R-R interval pattern, respiratory rate	13, Healthy	Placebo	Crossover, acute (~30 min)
Fisher (2018) Hypertension	Percentage decrease in median systolic blood pressure (SBP)	10, Hypertensive	Placebo	Crossover, acute (duration NR)
Stowell (2019) Hypertension	Arterial blood pressure	12, diagnosed with Primary Hypertension	Placebo	Crossover, acute (5 days)
Zamotrinsky (2001) coronary artery disease	HR, BP, LV diastolic function, LV filling	18. patients with stable angina pectoris class IV	No interventio n	Parallel, acute (~10 days)
Antonino (2017) Spontaneous cardiac baroreflex sensitivity (cBRS)	cBRS from systolic blood pressure and RR interval HRV (LF/HF)	13, Healthy	Both	Crossover, acute (duration NR)
Bretherton (2019) HRV and baroreflex sensitivity	cBRS and HRV	14, Healthy	Placebo	Crossover, acute (1 week)

Clancy (2014) HRV & Sympathetic Activity	HRV (LF/HF	48, Healthy	Placebo	Parallel, acute (~15 min)
De Couck (2017) HRV	ECG with HRV	30, Healthy	Placebo	Crossover, acute (duration NR)
Borges (2019) cardiac vagal activity	HRV	60, Healthy	Sham	Crossover, acute (duration NR)
Tran (2019) Left Ventricular Strain and Autonomic Tone	LV global longitudinal strain (GLS)	24, diagnosed with diastolic dysfunction by echocardiogram	Sham	Crossover, acute (duration NR)
Yu (2017) Myocardial Ischemia Reperfusion Injury	Ventricular premature beat (VPB) incidence	95, MIRI patients	Placebo	Parallel, acute (~2.5 hr)
Epilepsy				
Bauer (2016) Epilepsy	Reduction in seizure frequency (per 28 days)	58, Epileptic	Sham	Parallel, chronic (20 weeks)
Aihua (2014) Epilepsy	Reduction in seizure frequency (per month)	47, Epileptic	Sham	Parallel, chronic (12 months)
Rong (2014) Epilepsy	Seizure frequency (per 4 weeks)	144, Epileptic	Sham	Parallel, chronic (24 weeks)
Pain				
Straube (2015) Chronic Migraine	Decrease in headache per 28 days	46, chronic migraine patients	Sham	Parallel, chronic (12 weeks)
Kutlu (2020) Fibromyalgia	Visual Analog Sale, Beck Depression Scale, Beck Anxiety Scale, Fibromyalgia Impact Questionnaire, Short Form-36 for life quality	52, Fibromyalgia patients	No interventio n	Parallel, chronic (4 weeks)
Busch (2013) Pain	Thermal, mechanical, and pressure pain thresholds	48, Healthy	Placebo	Crossover, acute (2 days)
Janner (2018) Pain	Perceived pain intensity and temporal summation of pain	49, Healthy	Both	Crossover, acute (8 days)
Kovacic (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)
Laqua (2014) Pain	Pain Threshold	21, Healthy	Placebo	Crossover, acute (~1 week)
Johnson (1991) Pain threshold and autonomic function	Electrical pain threshold and autonomic function	24, Healthy	Sham	Parallel, acute (~15 min)
Napadow (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)
Psychological				
Hein (2013) Depression	Hamilton Depression Rating Scale, Beck's Depression Inventory	37, Majorly depressed	Placebo	Parallel, chronic (~2 weeks)

Hasan (2015) Schizophrenia	Positive and negative schizophrenia symptom scale	17, Schizophrenic	Placebo	Parallel, chronic (26 weeks)
Burger (2019) Negative Thought Occurrence	Number of negative thought intrusions	97, High worriers	Sham	Parallel, acute (~20 min)
Others				
Addorisio (2019) Rheumatoid Arthritis	Endotoxin induced interleukin (IL)-6, IL-1β, and TNF	19, Healthy	Sham	Crossover, acute (duration NR)
Huang (2014) Impaired glucose tolerance	2-hr plasma glucose levels	102, IGT patients	Sham	Parallel, chronic (12 weeks)
Juel (2017) Pain & Gastrointestinal Motility	ECG and PPG, pain threshold, cold pressor test, and drink test (ultrasound imaging)	20, chronic pancreatitis patients	Sham	Crossover, acute (7 days)
Frøkjær (2016) Gastroduodenal motility and pain threshold	ECG and PPG, pain threshold, cold pressor test, and drink test (ultrasound imaging)	18, Healthy	Sham	Crossover, acute (~6 days)
Cakmak (2017) Parkinson's	Motor symptoms	24, Parkinson's Hoehn & Yahr stage 2-3	Placebo	Crossover, acute (duration NR)
Maharjan (2018) olfactory function	Odor threshold test and supra-threshould test	18, Healthy	Sham	Crossover, acute (7 days)
Tutar (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)

3.3 Risk of bias tool to assess quality of evidence

The Cochrane 2.0 Risk of Bias assessment subscores for each study are displayed in table x below. A detailed explanation for the reasoning behind each RoB subscore assignment (L = low, S = some concerns, H=high) can be found in section 2.3 and in supplementary material for each study specifically.

Table 4: RoB overall and section scores (organized 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 (2016)	L	L	L	L	L	L
Maharjan (2018)	L	L	L	L	L	L
Addorisio (2019)	L	S	L	L	S	S
Aihua (2014)	S	H	L	S	S	S
Andreas (2019)	L	S	L	L	L	S
Antonino (2017)	L	L	S	L	S	S
Badran (2018)	L	L	L	L	S	S
Bretherton (2019)	L	L	L	L	S	S
Burger (2019)	L	L	L	L	S	S
Busch (2013)	L	S	L	L	S	S
Cakmak (2017)	L	H	L	L	L	S
Clancy (2014)	L	S	L	L	S	S
De Couck (2017)	L	L	L	L	S	S
Hein (2013)	L	L	L	S	S	S
Huang (2014)	S	L	L	L	S	S
Janner (2018)	L	L	L	L	S	S
Juel (2017)	L	S	L	L	S	S
Kovacic (2017)	L	S	L	L	L	S
Rong (2014)	L	S	L	L	S	S
Stavrakis (2015)	L	S	L	L	S	S
Stavrakis (2020)	L	L	L	L	S	S
Tran (2019)	L	S	L	L	S	S
Tutar (2020)	L	S	L	S	S	S
Afanasiev (2016)	H	S	H	L	S	H
Borges (2019)	L	S	L	L	S	H
Fisher (2018)	S	H	H	S	H	H
Frøkjaer (2016)	S	S	L	L	S	H
Hasan (2015)	L	H	L	L	S	H
Johnson (1991)	S	H	L	S	S	H
Kutlu (2020)	S	H	L	L	L	H
Laqua (2014)	L	S	L	L	H	H
Napadow (2012)	S	H	S	L	H	H
Salama (2020)	L	H	L	S	S	H
Stowell (2019)	L	S	L	L	S	H
Straube (2015)	L	H	L	L	L	H
Tobaldini (2019)	S	H	L	L	S	H
Yu (2017)	L	H	L	L	S	H
Zamotrinsky (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.

Only two studies (Bauer et al., 2016 and Maharjan et al., 2018) were assigned an overall low risk of bias. This is unsurprising, as the risk of bias assessment is rigorous and it is difficult to design and implement RCTs with low risk of bias. Subsection and overall score percentages are illustrated in Fig. 4 below.

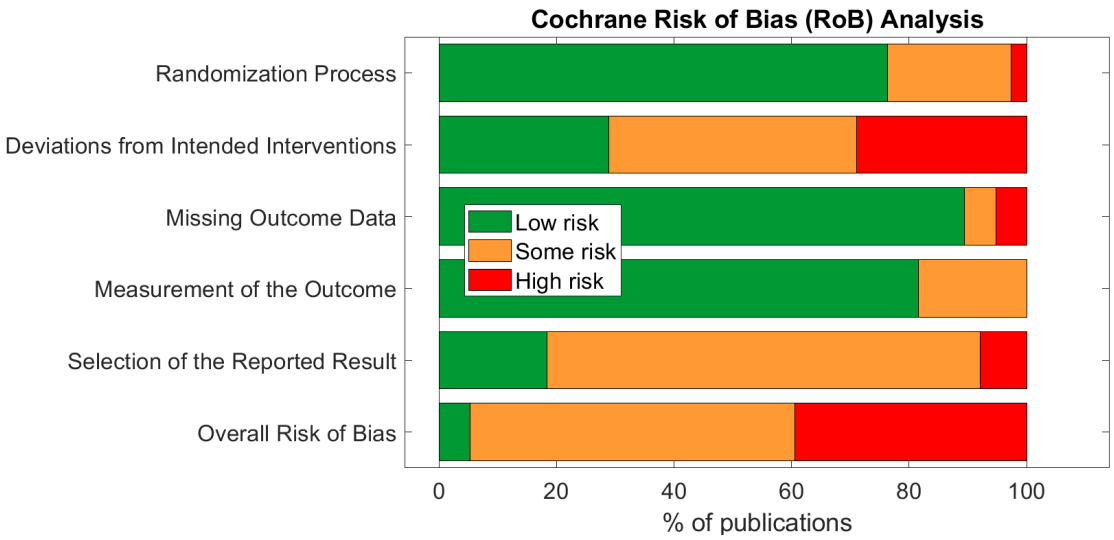

Figure 4: Summary of Cochrane Risk of Bias subsections and overall scores for 38 publications reviewed.

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

Notably, the 'deviations from intended interventions' subsection tended to have the highest risk of bias. This was mainly due to issues with possible subject unblinding as a result of easily perceptible differences between active and control groups. For example if in a crossover design, a placebo control with an active intervention group whose intervention involved clearly perceptible paresthesia was used and the same subject felt both the paresthesia inducing active and the no stimulation placebo control.

'Missing outcome data' was also a low risk of bias section. Studies scoring high or some concerns had unreported outcome data without explicit justification, with a non-trivial difference in the proportion of missing data between interventions that was not robust to this discrepancy.

'Measurement of the outcome' was a similarly low risk of bias section. In order to score some concerns, outcome assessor blinding to subject intervention had to be compromised. For a high RoB score in this section, studies were measuring endpoints that could be influenced by investigator unblinding, such as in the case of 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 4.3. Far fewer studies had issues with the randomization process, missing outcome data, or measurement of the outcome.

3.4 Summary of outcome measures across indications

Here, a summary of the effects of aVNS on common indications is presented. Keeping the RoB assessment in mind, the common measurement outcomes are qualitatively synthesized across studies.

The RoB assessment is used to point out instances where trial design or reporting has the potential to influence interpretation of the trial outcomes.

Cardiac related effects of aVNS

The primary cardiac effects assessed were changes to heart rate (HR) and sympathovagal balance. To measure sympathovagal balance, heart rate variability (HRV) was used. Results were conflicted across studies for heart rate changes and sympathovagal balance but suggest aVNS may have an effect on both. There are concerns that results may be attributed to trial design and inconsistent measurement methods.

Studies reporting a change in HR, report a modest mean drop of 2-3 BPM in the active group. However, almost half of the eleven trials reporting HR effects reported no significant difference in effect between or within sham 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 till a decrease in HR was measured. They report a mean stimulation threshold, to elicit a HR decrease, above the mean threshold for discomfort. In addition, Frojaker et al. (2016) and Juel et al. (2017) report a significant decrease in HR during sham at the earlobe, but not during active stimulation at the conchae and tragus. This suggests that decrease in HR may not be vagally mediated but perhaps mediated by the trigeminal or cervical nerve branches in the auricle (see Fig. 1a). Taken together, there is evidence for the effects of auricular stimulation on decreasing HR at high stimulation amplitudes, but it may not be 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 5 are different ways to quantify HRV, creating multiple ways to analyze ECG data for HRV, allowing for multiple comparisons in search of statistical significance, which may not be appropriately corrected for. 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. With more variance in HR during breathing indicating more vagal control and hence a shift in cardiac sympathovagal balance to parasympathetic (Goldberger, 1999). Contradictory results on the parasympathetic effects of aVNS indicate that the effects of aVNS on HRV are inconsistent or that HRV is an unreliable measure of cardiac sympathovagal balance (Bootsma et al., 2003). Overall, the effects of aVNS on sympathovagal balance are conflicted but suggest that aVNS may improve sympathovagal balance towards parasympathetic activity.

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, as well as regression to the mean in the reported results. For example, there was some evidence showing a small decrease in blood pressure during aVNS. However, the effect is small and likely not clinically relevant for most patients. Since severity of disease state was an inclusion criterion, the small decrease could also be due to regression to the mean. At the same time, a clinical trial visit could increase stress levels and BP of the subjects and

mask the therapeutic effects of aVNS on blood pressure. Another example of subject familiarization is related to HRV. Borges et al. (2019) tried to account for subject familiarization but baseline HRV measurements were taken immediately after these familiarization stimulation sessions. Hence, there is no true baseline measurement of HRV in Borges et al. (2019) and casts doubt on the findings of the study which claims 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.

Table 5: Summary of cardiac studies

Author (y) Indication	**HR Results**	**HRV Results**	**Active level**	**Active waveform**	**Active location**	**Subjects analyzed, Disease**
Antonino (2017) Spontaneous cardiac baroreflex sensitivity (cBRS)	Active: 2-3BPM decrease Sham: NS	Active: LF/HF parasympathetic Sham at earlobe: LF/HF parasympathetic Placebo: LF/HF sympathetic	First sensory	30 Hz, 200 µs PW	Tragus (bilateral)	13, Healthy
Badran (2018) Heart rate	Active: ~2.40 BPM decrease Sham: NS	NA	2x sensory threshold	10 Hz, 500 µs PW	Tragus (left)	35, Healthy
Bretherton (2019) HRV and baroreflex sensitivity	NA	Active: LF/HF ratio parasympathetic Sham: NS	Sensory threshold	30 Hz, 200 µs PW	Inner & outer tragus (side not reported)	14, Healthy
Burger (2019) Negative Thought Occurrence	NA	Active: NS (RMSSD) Sham: NS	Strong sensory	25 Hz, 250 µs PW	Left cymba concha	97, High worriers
Clancy (2014) HRV & Sympathetic Activity	Active: HR decreased (data not reported) Sham: NS	Active: LF/HF ratio parasympathetic Sham: NS	First sensory	30 Hz, 200 µs PW	Inner & outer tragus (side not reported)	48, Healthy
De Couck (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	25Hz, 250 µs PW	Bilateral cymba concha	30, Healthy
Janner (2018) Pain	No significant effects of aVNS on HR (data not reported)	NA	Strong sensory	100 Hz, 200 µs PW	Cymba concha (bilateral)	49, Healthy
Juel (2017) Pain & Gastrointestinal	Active: NS	Active: RMSSD parasympathetic relative to sham, sympathetic relative to baseline	Discomfort	30 Hz, 250 µs PW	Concha (left)	20, chronic pancreatitis patients

Motility	Sham: ~2.8BPM decrease, at earlobe	Sham at earlobe: RMSSD sympathetic				
Tran (2019) Left ventricular strain and autonomic tone (HRV)	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	Tragus (right)	24, diagnosed with diastolic dysfunction by echocardiogram
Yu (2017) Myocardial Ischemia Reperfusion Injury	HR drop stim threshold was 5.5 ± 2.9 mA; absolute HR drop values not reported	NA	Sinus rate slowing threshold	20 Hz, 1 ms PW	Tragus (right)	95, MIRI patients
Borges (2019) cardiac vagal activity	NA	Active: RMSSD sympathetic between resting and second half stim Sham: NS	Strong sensory	25 Hz, PW of 200–300 µs	Left cymba concha	60, Healthy
Frøkjær (2016) Gastroduodenal motility and pain threshold	Active: NS Sham: ~2.2BPM 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	Concha (left)	18, Healthy
Johnson (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	Concha (right)	24, Healthy
Laqua (2014) Pain	No significant effects of aVNS on HR (data not reported)	NA	Strong sensory	Alternating between 2 Hz and 10 Hz, 200 µs PW	Concha (anode) and mastoid (cathode) (bilateral)	21, Healthy
Napadow (2012) Pain	No significant effects of aVNS on HR (data not reported)	NA	Strong sensory	30 Hz, 450 µs PW	Cymba concha and antihelix/cavum concha slope (left)	15, Chronic Pelvic Pain due to endometriosis patients

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 (R-R) intervals. NN50 is the number of times successive heartbeat intervals exceed 50ms (time-domain measurement)
SDNN	Standard deviation of the NN (R-R) intervals (time-domain measurement)

Inflammatory related effects of aVNS.

Several studies measured cytokine levels to investigate the anti-inflammatory effects of aVNS. Cytokine levels are either measured directly in drawn blood or after an *in vitro* endotoxin induced challenge on drawn blood. In the four studies measuring circulating cytokine levels results were somewhat conflicting. Stavrakis et al. (2020) reports a significant decrease in 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. In subjects being treated for myocardial infarction, Yu et al. (2017) reports that the active group is 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) shows an increase in TNF-α, along with a drop in CRP and IL-6. 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 are discussed in (Stoddard et al., 2010), (Yang et al., 2011), and (Thurm et al., 2005). Additionally, (Broekman et al., 2015) provides 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 1h of stimulation, there was a significant decrease in TNF-α and c-reactive protein (CRP) levels in femoral vein draws. CRP is also an acute phase protein whose release from the liver is stimulated by increased levels of IL-6 (Giudice and Gangestad, 2018). In Addorisio et al. (2019), vibrotactile stimulation was applied for only 2 minutes, showing a statistically significant decrease in endotoxin induced cytokine levels of TNF-α, IL-6, and IL-1βin blood drawn one hour after stimulation.

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

Table 6: Summary of inflammatory studies

Author (y) Indication	aVNS Cytokine results (compared to control)	Measurement method	Subjects, disease
Addorisio (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 (2019) postop AF	IL-6 NS	Endotoxin induced	40, patients undergoing cardiac surgery
	CRP NS		
Stavrakis (2015) Atrial Fibrillation	TNF-α decrease (p<0.05)	Endotoxin induced	40, paroxysmal AF ablation patients
	IL-6 NS		
	IL-10 NS		
	CRP decrease (p<0.05)		
Stavrakis (2020) Atrial Fibrillation	TNF-α decrease (p=0.0093)	Circulating	53, diagnosed with Paroxysmal AF
	IL-6 NS		
	IL-1β NS		
	IL-17 NS		
Yu (2017) Myocardial Ischemia Reperfusion Injury	TNF-α lower increase than control (p <0.05)	Circulating	95, MIRI 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)		
Afanasiev (2016) Coronary Insufficiency and left ventricular 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 NS; HSP70 increase (p<0.05)		
Salama (2020) Acute inflammatory response after lung lobectomy	TNF-α NS	Circulating	100, lobectomy via thoracotomy in patients with non-small cell lung cancer
	IL-6 lower increase than control (p =0.02)		
	IL-10 NS		
	IL-1β NS		
	IL-18 NS		
	CRP lower increase than control (p=0.01)		

Epilepsy related effects of aVNS

The three studies investigating the antiepileptic effects of aVNS were all tested chronically. All used stimulation frequencies between 20-30 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, showing significance from baseline after a few weeks to months of daily prescribed stimulation.

Based on these three chronic studies, there is some evidence to suggest the anti-epileptic effects of aVNS. The primary outcomes of these non-invasive interventions are comparable to that of implantable cervical 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 versus within group analysis, it is possible that the effects may be attributed to placebo.

Table 7: Summary of epilepsy studies

Author (y) Indication	Results	Active waveform	Active Location	Study Duration
Auricular VNS				
Bauer (2016) Epilepsy	**Between group (NS):** 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	Concha (left)	20 weeks
Aihua (2014) Epilepsy	**Between group (S)**: NS 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	concha and external ear canal (bilateral)	12 months
	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 (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	Cymba concha and cavum concha (side NR)	24 weeks
Cervical implanted VNS				
Author (y)	**Results**			**Study Duration**
Ben-Menachem (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 (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

Pain related effects of aVNS

Eight studies investigated the effects of aVNS on the sensation of pain. Four studies investigated the effects of aVNS on pain threshold levels in healthy patients but using varying pain-assessment methods. In contrast, the other four studies examined the effects of aVNS 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 were highly varied. In studies that investigated pain thresholds for healthy patients, results showed negligible changes in pain threshold levels due to aVNS therapy. For chronic migraines (Straube et al., 2015), aVNS had a therapeutic effect with both 1 Hz and 25 Hz stimulation. Unexpectedly, the 1 Hz stimulation, considered 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 lesser effect on the reduction of headaches (-7 vs -3.3 over 28 days). For GI Pain, fibromyalgia, and chronic pelvic pain, pain was also significantly ameliorated by aVNS therapy. Overall, these studies provide evidence that aVNS may be therapeutic for pain conditions, but more studies are needed to further explore the effectiveness for specific medical conditions and rule out significant contributions by placebo effect.

Table 8: Summary of pain studies

Author (y) Indication	**Pain Results**	**Duration of study**	**Subjects analyzed, disease**
Busch (2013) Pain	Active: NS Sham: NS	2 sessions (48 hrs apart)	48, Healthy
Janner (2018) Pain	Active: NS Sham: NS	4 sessions (48 hrs apart)	49, Healthy
Kovacic (2017) GI pain	Active: Pain decrease Sham: NS	4 weeks	104, Children with GI pain
Johnson (1991) Pain threshold and autonomic function	Active: NS Sham: NS	1 session	24, Healthy
Kutlu (2020) Fibromyalgia	Active: Pain decrease Sham: Pain decrease	4 weeks	52, Fibromyalgia patients
Laqua (2014) Pain	Active: NS Sham: NS	2 sessions (1 week apart)	21, Healthy
Napadow (2012) Pain	Active: Larger decrease in pain ($p<0.05$) Sham: Decrease in pain ($p<0.05$)	2 sessions (1 week apart)	15, Endometriosis patients with chronic pelvic pain
Straube (2015) Chronic Migraine	Active (25Hz): Headache occurence -3.3 per 28 days (17.4 % decrease) Sham (1Hz) : Headache occurrence -7.0 per 28 days (36.4 % decrease)	12 weeks	46, Chronic migraine patients

Other effects of aVNS.

Other clinically investigated effects of aVNS are on motor symptoms of Parkinson's, depression, schizophrenia, obesity, impaired glucose tolerance, gastroduodenal motility, and tinnitus. Most of these effects are 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. More studies are needed to further explore the effects of aVNS for these indications.

4. DISCUSSION

This is the first systematic review applying the Cochrane Risk of Bias framework to auricular vagus nerve stimulation clinical trials. Our systematic review of 38 publications, totaling 41 RCTs, across indications including epilepsy, cardiac, inflammatory, and pain show 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 conflicted between studies and were hindered by trial designs including inappropriate washout periods and multiple methods used to quantify HRV. There is early-stage evidence to suggest aVNS may reduce circulating levels or 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. aVNS showed preliminary evidence of ameliorating pathological pain but not induced pain.

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, inconsistent reporting of results - raise the level of concern for bias in the outcomes and therefore the quality of evidence. 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.

While the lack of consistency in outcomes can be partially explained by differences in trial design, it is important to stress that fundamentally there exists no way to directly confirm engagement of on-target neural fibers during aVNS. This inability to measure the engagement of on- and off-target neural fibers critically limits the development of aVNS therapies.

Here, we discuss gaps and improvements to aid the development of aVNS therapies. 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 need and associated challenges in careful design of controls and

maintenance of the double blind. Lastly, in the short term, we make suggestions to improve the reporting of clinical trials results to allow meta-analysis of results across aVNS studies.

4.1 Target Engagement

The lack of direct measures of neural target engagement from nerve trunks innervating the ear in preclinical and early clinical studies is hindering the development of aVNS therapies. Primary measures of local target engagement will enable identification of fiber type activated, serve as biomarkers to identify on- and off-target nerve neural activation, and provide more comprehensive information when drawing lessons from animal models. 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. Data from direct measures of on- and off-target engagement can be used to titrate the therapy by adjusting electrode design, placement, and stimulation waveform parameters to improve local neural target engagement.

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) shows the cymba conchae is innervated only by the auricular vagus in 7 of 7 cadavers. In reality, 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. Variation in peripheral nerve innervation is well studies for other regions of the body, such as the hand (Guru et al., 2015; Bas and Kleinert, 1999). The reliance on the Peuker and Filler study is concerning due to the small sample size and poor representation, presumably drawn only from the German population. Given the importance of the claims in Peuker and Filler, a further dissection study with a more representative and larger sample size is called for. Secondly, functional magnetic resonance imaging (fMRI) evidence is also used to suggest vagal innervation of the concha (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) show that results of MRI studies are 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), activating the same region in the brain when in fact the auricular vagus might not be recruited during stimulation. Still more trigeminal pathways from the auricle lead to the trigeminal spinal nuclei, whose proximity to NTS might lead to apparent activation of the NTS. fMRI of the brainstem is further complicated (Napadow et al., 2019) as distance from the measurement coils is increased, decreasing the effective resolution (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, 2017) 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 may be contaminated and misinterpreted due to EMG leakage (Usami et al., 2013). The common measures of target engagement used are secondary surrogates and are 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 have largely relied 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.

Implantable VNS is believed to recruit A and B fibers (Krahl, 2012) for its anti-epileptic effects and parasympathetic efferent B fibers innervating the heart for its cardiac effects (Sabbah et al., 2011). Despite the similar epineurial cuff electrode designs used in implantable VNS for epilepsy and heart failure, different stimulation parameters (Anand et al., 2020) are used to recruit different fiber pathways. Meanwhile, aVNS is hypothesized to achieve therapeutic potency by activating myelinated A-beta fibers on the auricular branch of the vagus nerve (Kaniusas, 2019), which in the auricle exits as a web of axons. Unlike stimulation of the 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 electrode. Orientation of fibers relative to the stimulation electrode has a large effect on fiber recruitment (Grill, 1999) and could potentially lead to preferential activation of nerve pathways oriented parallel 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 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 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 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 (Rattay, 1999; McNeal, 1976). 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 brings to light firstly, our lack of confidence on which of several nerve trunks innervating the auricle are being activated during aVNS that may be generating the on- and off-target effects. Secondly, given the fundamentals of neural stimulation do not support directly porting stimulation parameters from implantable VNS it is important to understand whether the auricular vagus

is even being activated and if so, what fiber types. This knowledge requires direct measures of local target engagement from the nerves innervating the auricle. Data from local target engagement measurements can be used to optimise the therapy by adjusting electrode design, placement, and stimulation parameters.

To further the development of aVNS, it will be essential to understand local target engagement of the nerves 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 in both preclinical animal models and clinical patients. 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 understand the relation 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 versus 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.

4.2 Problems with control design

The design of an indistinguishable yet nontherapeutic control is central to maintaining the double blind 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, making 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 the double blind in aVNS RCTs.

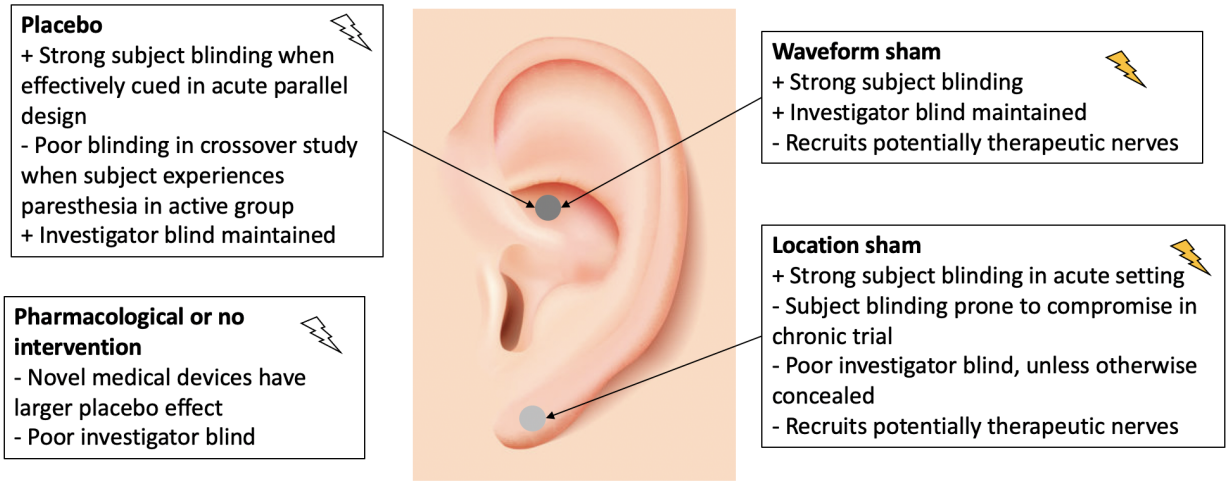

Figure 5: Types of controls used in aVNS clinical trials. Ideal control is indistinguishable from active intervention, to both subject and investigator, yet therapeutically inert.

Common control designs used in aVNS studies are summarized in Fig. 5. Namely, a placebo, which entails similar placement of the electrode and device but no stimulation delivered. In a waveform sham, a different nontherapeutic waveform is delivered at the same location as active intervention. In a location sham, the most common control used in 16 of 41 RCTs reviewed, the same active waveform is delivered at a different location on the auricle that is presumably not engaging a therapeutic nerve. Lastly, no intervention or a pharmacological control may be used. These different control designs are evaluated at length in supplementary material along with recommendations on appropriate control types depending on trial design.

Inappropriate implementation of the control group resulted in compromised blinding in many studies when subjects were able to feel a paresthesia in the active intervention but not in the control and when investigators were able to see differences in electrode placement or stimulator operation. Unblinding due to inappropriate control design was the main contributor leading to risk of bias in the Cochrane analysis deviation from intended intervention section. The design of appropriate controls is difficult for trials testing non-pharmacological interventions - especially so for paresthesia-inducing neuromodulation trials (Robbins and Lipton 2017; Nature Biotechnology, 2019) but is essential to establish a double blind.

Measurement of local target engagement via microneurography of the major nerve trunks innervating the ear will enable understanding of the neural recruitment occurring during active and sham stimulation and guide appropriate design of controls. 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. Over time, the consistent use of control designs and evaluation of blinding will grow our understanding of the concealability and therapeutic potency of various control methods.

4.3 Short Term Solutions – Guide to Reporting

While measurement of local target engagement and improvement of control design using microneurography and consistent evaluation of blinding remain fundamental gaps to improve the quality of evidence for aVNS in the long-term, several steps can be implemented in the short term 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 clinical trial 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. 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.

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 be reported. Of the 41 aVNS RCTs reviewed, 13 RCTs pre-registered, but only 5 of these 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 appropriate method for a controlled study is a between group comparison of the active arm versus the control arm. Several studies claimed statistically significant findings even if the between group analysis was non-significant, based just on the within group analysis. An example illustrating this difference is found in supplementary material. 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 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, worsening of symptoms (if any) in the non-responders, and the distribution of the results in low sample size studies. In Stavrakis et al. (2020) there is worsening of symptoms in the non-responders (53% of the active group) at the three-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.

Justifying clinical relevance of outcome measures. The outcome measure itself may not be established as clinically relevant. For example, an *in vitro* endotoxin induced cytokine measurement was used to proxy *in vivo* immune response in Addorisio et al. (2019) and Stavrakis et al. (2015). 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 [] to measure blood pressure. Clancy et al. (2014) used both a finger based PPG, Finometer®, and a traditional arm sphygmomanometer to measure BP and concluded that the increase in BP measured using the Finometer®, which persisted into the recovery phase after stimulation, 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.

Contextualizing 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 outcome from intervention to the outcome of the standard of care or another therapy. For example, Cakmak et al. (2017), in an aVNS trial for Parkinson's showed a 5.3 points improvement on the UPDRS part 3 (Goetz et al., 2008) in the active group. This could be contextualized with the 18.4 points improvement in DBS

(Kahn, 2019). A direct comparison may be disputable due to varying disease severity of the subjects, but it does provide contextualization of the UPDRS part 3 scale.

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 homage should be paid to this fact when discussing clinical significance of the findings.

Risk of bias in the selection of reported results found in the RoB analysis can be addressed by pre-registration of trials, use of appropriate statistical analysis, justifying clinical relevance of outcome measures, and contextualizing clinical significance of results. These ideas are summarized in Table 9.

Table 9: Checklist for trial reporting

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

4.4 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 towards 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 (DiMasia et al., 2016; Paul et al., 2010). In Gupta et al. (2011), several factors were identified that hindered the successful translation of drug therapies from early-stage results to market approval that are relevant to aVNS.

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 versus 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 scale, neuroanatomy, and neurophysiology from animal models to humans. A concern already 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, as outlined above, can be deployed across preclinical models and early clinical studies, increasing translatability of findings, and providing data to titrate electrode design, placement, and stimulation waveform parameters to optimize for target engagement in the development and deployment of neuromodulation therapies.

4.5. aVNS compared to other 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 VNS and SCS will inform stimulation parameters, possible mechanisms of action, and electrode design and placement to maximize target nerve recruitment and minimize therapy limiting off-target effects such as muscle recruitment (Nicolai et al., 2020; Yoo et al., 2013). 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.

Paresthesia inducing neuromodulation therapies make the design of an indistinguishable yet nontherapeutic sham challenging. At the same time, an appropriate control is essential to the maintenance of a double blind and design of an RCT. 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., 2020). 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., 2019). A systematic review of dorsal root ganglion (DRG) stimulation for pain also showed serious concern for bias across all studies reviewed due to compromise in subject and investigator blinding (Deer et al., 2013). The suggestions laid forth on measuring local target engagement and consistent post-hoc evaluation of blinding in subjects and investigators will grow our knowledge of effective control design for paresthesia inducing neuromodulation therapies such as aVNS, SCS, and implantable VNS.

4.6 Limitations of this review

Firstly, none of the authors have conducted an aVNS clinical trial. The 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. 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 an additional 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 was 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.

5. Conclusion

Based on our review of 41 aVNS clinical RCTs, we conclude with the overall impression that aVNS shows physiological effects but has not yet shown strong clinically significant effects. Progress in the field is limited by lack of direct measures of target engagement at the site of stimulation. Measures of target engagement will inform therapy design and control design for maintenance of the double blind. Using the risk of bias tool, we found concerns in the design of trials, particularly control and blinding, and incomplete reporting of information. The studies are currently exploratory in nature, which is appropriate given the early-stage of research the aVNS field is in. The non-invasive nature and low side effect profile of auricular stimulation holds potential to make it a first-line therapy in the treatment of a variety of illnesses and disease states. Rigor in trial design and reporting, along with the systematic study of mechanisms of action starting with local target engagement, will accelerate the development and clinical translation of aVNS based therapies.

As a field, neuromodulation has ways to go in attaining social normality and gaining widespread adoption as a first-line therapy (Payne and Prudic, 2009; Li et al., 2020; Daniel et al., 2004). 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. As a field, we strive to be known for a high standard of rigor and quality of evidence.

Acknowledgements

Eric H. Chang for notes on endotoxin induced cytokine assays. Carly Frieders and Erik Lovett for reviewing the manuscripts.

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 grants no. N66001-17-2-4010.

References

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. Yu., 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. *Annals of Noninvasive Electrocardiology* 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 & Behavior* 39, 105–110. doi:10.1016/j.yebeh.2014.08.005.

Anand, I. S., Konstam, M. A., Klein, H. U., Mann, D. L., Ardell, J. L., Gregory, D. D., et al. (2020). Comparison of symptomatic and functional responses to vagus nerve stimulation in ANTHEM-HF, INOVATE-HF, and NECTAR-HF. *ESC Heart Failure* 7, 76–84. doi:10.1002/ehf2.12592.

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: Arrhythmia and Electrophysiology* 12. doi:10.1161/CIRCEP.119.007711.

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 Stimulation* 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 Stimulation* 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. *JoVE*, 58984. doi:10.3791/58984.

Bas, H., and Kleinert, J. M. (1999). Anatomic variations in sensory innervation of the hand and digits. *The Journal of Hand Surgery* 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 Stimulation* 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 Research International* 2017, 1–10. doi:10.1155/2017/7830919.

Bootsma, M., Swenne, C. A., Janssen, M. J. A., Cats, V. M., and Schalij, M. J. Heart rate variability and sympathovagal balance: pharmacological validation. 10.

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. Variability in the analysis of a single neuroimaging dataset by many teams. 26.

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.

Brijesh Nair (2019). Clinical Trial Designs. *Indian Dermatol Online J*. doi:10.4103/idoj.IDOJ_475_18.

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., 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. *Biological Psychology* 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 Stimulation* 6, 202–209. doi:10.1016/j.brs.2012.04.006.

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). 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].

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 Stimulation* 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 Neuroscience* 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. *International Journal of Cardiology* 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 Medicine* 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, Behavior, and Immunity* 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. *Journal of Health Economics* 47, 20–33. doi:10.1016/j.jhealeco.2016.01.012.

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: Technology at the Neural Interface* 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.

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 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 Stimulation* 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? *American Journal of Physiology-Heart and Circulatory Physiology* 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. *Anatomy Research International* 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.

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 Journal* 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. https://training.cochrane.org/handbook/current/chapter-08 [accessed November 26, 2020]

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. doi: 10.17226/18494.

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.

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. *Clinical Neurology and Neurosurgery* 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. *Journal of Athletic Training* 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.

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. *The Lancet Gastroenterology & Hepatology* 2, 727–737. doi:10.1016/S2468-1253(17)30253-4.

Krahl, S. (2012). Vagus nerve stimulation for epilepsy: A review of the peripheral mechanisms. *Surg Neurol Int* 3, 47. doi:10.4103/2152-7806.91610.

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 Research International* 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 Neuroscience* 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. 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.

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, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Medicine* 6. doi:10.1371/journal.pmed.1000097.

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.

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 Neuroscience* 199, 48–53. doi:10.1016/j.autneu.2016.06.004.

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 Reports* 4, e745. 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.

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: *Journal of Psychiatric Practice* 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.

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.

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. doi:10.4081/cp.2016.856.

Robbins, M. S., and Lipton, R. B. (2017). Transcutaneous and Percutaneous Neurostimulation for Headache Disorders. *Headache: The Journal of Head and Face Pain* 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. *Clinical Science*, CS20130518. doi:10.1042/CS20130518.

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.

Shen, D. Estimate Carryover Effect in Clinical Trial Crossover Designs. 7.

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. *Journal of the American College of Cardiology* 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: Clinical Electrophysiology* 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*, 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 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, USA: John Wiley & Sons, Inc.), im0718bs66. doi:10.1002/0471142735.im0718bs66.

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. of Cardiovasc. Trans. Res.* 12, 221–230. doi:10.1007/s12265-018-9853-6.

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

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. *American Journal of Otolaryngology* 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 Stimulation* 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 USA* 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 Neurosurgery* 96, 293–301. doi:10.1016/j.wneu.2016.08.102.

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

Yang, B., Pham, T., Goldbach-Mansky, R., 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: Cardiovascular Interventions* 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. 8.