Dysphagia
https://doi.org/10.1007/s00455-022-10428-2

ORIGINAL ARTICLE

Statistical Power and Swallowing Rehabilitation Research: Current Landscape and Next Steps

James C. Borders[1] · **Alessandro A. Grande**[2] · **Michelle S. Troche**[1]

Received: 2 November 2021 / Accepted: 14 February 2022

Abstract

Despite rapid growth in the number of treatments to rehabilitate dysphagia, studies often demonstrate mixed results with non-significant changes to functional outcomes. Given that power analyses are infrequently reported in dysphagia research, it remains unclear whether studies are adequately powered to detect a range of treatment effects. Therefore, this review sought to examine the current landscape of statistical power in swallowing rehabilitation research. Databases were searched for swallowing treatments using instrumental evaluations of swallowing and the penetration–aspiration scale as an outcome. Sensitivity power analyses based on each study's statistical test and sample size were performed to determine the minimum effect size detectable with 80% power. Eighty-nine studies with 94 treatment comparisons were included. Sixty-seven percent of treatment comparisons were unable to detect effects smaller than $d = 0.80$. The smallest detectable effect size was $d = 0.29$ for electrical stimulation, $d = 0.49$ for postural maneuvers, $d = 0.52$ for non-invasive brain stimulation, $d = 0.61$ for combined treatments, $d = 0.63$ for respiratory-based interventions, $d = 0.70$ for lingual strengthening, and $d = 0.79$ for oral sensory stimulation. Dysphagia treatments examining changes in penetration–aspiration scale scores were generally powered to reliably detect larger effect sizes and not smaller (but potentially clinically meaningful) effects. These findings suggest that non-significant results may be related to low statistical power, highlighting the need for collaborative, well-powered intervention studies that can detect smaller, clinically meaningful changes in swallowing function. To facilitate implementation, a tutorial on simulation-based power analyses for ordinal outcomes is provided (https://osf.io/e6usd/).

Keywords Swallowing rehabilitation · Meta-science · Statistical power · Dysphagia · Deglutition disorders

Introduction

The field of dysphagia has experienced rapid growth in the number and types of treatments to rehabilitate swallowing dysfunction. Despite these scientific advances, studies examining the effectiveness of these treatments often yield mixed results with non-significant changes to functional outcomes. These null findings are often associated with a lack of evidence for an intervention, prompting some to question their efficacy [1, 2]. However, clinically meaningful findings do not always align with statistical significance [3]. Non-significant results may be attributed to inadequate statistical power to detect smaller, but potentially clinically meaningful, treatment effects. Statistical power is defined as the probability of detecting a "true" effect (when the effect exists) and involves four parameters in its analysis: power, alpha level, effect size, and sample size.

In the context of dysphagia rehabilitation, there are several swallowing-specific factors that should motivate researchers to design studies that can detect smaller treatment effects. First, dysphagia can be impacted by multiple, complex mechanisms of dysfunction, which may also vary within and between patient populations; therefore, it is unlikely that one treatment alone will result in a large effect. Secondly, bolus, task, and disease characteristics may increase swallowing variability, which can substantially reduce statistical power [4–6]. Finally, effect sizes become increasingly smaller as the number of factors that influence a behavior increases [7]; thus, dysphagia interventions seeking to improve functional outcomes in patients with multiple

✉ James C. Borders
jcb2271@tc.columbia.edu

[1] Laboratory for the Study of Upper Airway Dysfunction, Department of Biobehavioral Sciences, Teachers College, Columbia University, New York, NY, USA

[2] Department of Statistics, Columbia University, New York, NY, USA

Published online: 28 February 2022

underlying mechanisms of dysfunction will require study designs, analyses, and sample sizes that have a high likelihood of detecting smaller effects. To confidently evaluate the ability of interventions to improve swallowing function, studies will require sufficient statistical power to detect a range of clinically meaningful effect sizes.

Though statistical power is often recommended to be 80%, this threshold is arbitrary and results in missing a "true" treatment effect 1 in 5 times [8]. Power is not a binary classification (e.g., "well-powered" versus "underpowered"); instead, it exists on a curve, affording varying degrees of power depending on the effect size of interest [9, 10]. For example, a study may have 90% power to detect a 'large' effect (e.g., $d = 1.20$) but only 40% power to detect a smaller magnitude effect (e.g., $d = 0.30$). Additionally, it is important to understand that power extends beyond merely the number of participants collected and is specific to a study's design and statistical analysis, such that certain designs (e.g., within- versus between-subject) and analyses (e.g., parametric versus non-parametric) afford higher statistical power [11].

There has been an increased awareness of the prevalence and impact of low-powered studies across many disciplines because of the importance of reproducibility and minimizing error [12–14]. Statistical power affects one's ability to accurately detect and estimate the direction and magnitude of an effect, which impacts the reliability of research findings [15]. Studies with low power are not only less likely to detect an effect, but also have a higher false positive rate when a statistically significant result is reported [12, 16, 17]. This means that studies with low power may mistakenly make a 'false discovery', indicating that a treatment effect is present when there is no true treatment effect. The effect size estimate can also be inflated in low-powered studies, overestimating its true magnitude [18]. This overestimation is most notable in studies with less than 50% power to detect a true effect [15]. These errors contribute to publication bias and affect reproducibility, often resulting in different conclusions across studies [19].

It remains unclear whether swallowing rehabilitation research demonstrates adequate statistical power to detect a range of treatment effects. Given recent findings that only 9% of studies using the penetration–aspiration scale reported a power analysis, studies may not be appropriately powered to detect treatment effects with this outcome [20]. Therefore, this review aimed to examine the current landscape of statistical power in swallowing rehabilitation research. Since statistical power is unique to a given research question and analysis, we chose to investigate studies examining changes to the penetration–aspiration scale—an outcome measure with widespread clinical and research use in the field of dysphagia [21]. The minimum effect size detectable with 80% power was then calculated for each study. Across all studies, we used a common effect size metric, namely Cohen's d, to describe the relative sensitivity of swallowing rehabilitation research to detect a range of effects. Notably, these effect sizes do not reflect each study's results; instead, they indicate the minimum effect size that was detectable with 80% power given the study design, sample size, and analysis. In this sense, studies with higher statistical power have a greater likelihood to detect smaller effect sizes.

Methods

Search Strategy

The search strategy was conducted in September 2021 according to PRISMA guidelines [22]. Two databases (Web of Science and PubMed) were queried for peer-reviewed publications citing "A Penetration–Aspiration Scale" [21] in order to identify studies using this outcome. Relevant systematic reviews and meta-analyses were also searched. For inclusion in the review, studies needed to have been interventions on adult populations (≥ 18 years of age) using the penetration–aspiration scale as an outcome measure during instrumental assessments of swallowing (flexible endoscopic evaluations of swallowing or videofluoroscopic swallowing studies). Exclusion criteria included studies descriptively reporting penetration–aspiration scale results without statistical analysis, non-English articles, pediatric populations, surgical treatments, and compensatory strategies (e.g., chin tuck, bolus modifications). Case series with less than 4 participants were also excluded since analyses with these sample sizes are typically descriptive in nature. Studies that did not provide sufficient information to calculate the minimum effect size detectable were excluded.

Study Selection and Data Abstraction

After removal of duplicates, titles and abstracts were screened for inclusion. Full-text articles were then assessed for final inclusion. The following variables were extracted from each article: treatment type, sample size, patient population, study design, whether a power analysis was reported, type of statistical analysis and comparison (i.e., between versus within-subject), comparison p value, and alpha level. A conservative approach to power estimation was used, such that the statistical test and sample size from the comparison that afforded the highest power was chosen. For example, if a study performed both between (i.e., experimental vs control group)- and within-subject (i.e., pre- to post-intervention for the experimental group) comparisons with the penetration–aspiration scale then the statistical test and sample size for the comparison that provided the highest power was used. Sensitivity analyses did not include additional covariates (e.g., bolus consistency, age).

Statistical Analysis

Sensitivity power analyses were performed in R version 4.0 for parametric statistical tests [23] and G*Power version 3.1 for non-parametric tests [24]. Despite strict statistical assumptions imposed in G*Power (i.e., normal distribution of difference scores for the Wilcoxon signed-rank test), we decided to use this software given its prevalence in clinical research. Sensitivity power analyses were performed based on the statistical test, sample size, and alpha level to determine the minimum effect size detectable with 80% power. Effect sizes were calculated based on the statistical test performed, then converted to Cohen's *d* to provide a standardized measure of effect size across studies. Though Cohen's *d* is an effect size measure for continuous outcomes and is not recommended for ordinal outcomes (e.g., the penetration–aspiration scale), we used this effect size since most studies reported Cohen's *d*. Thus, this reduced the number of effect size conversions and provided a common metric for comparisons across highly heterogeneous studies. Given that studies did not consistently report correlations between pre- and post-treatment outcomes for within-subject comparisons, a "moderate" correlation was assumed when converting from Cohen's *dz* to Cohen's *d*. The following formula was used for this conversion, where $\rho = 0.50$ [25].

$$d = dz \times \sqrt{2 \times (1 - \rho)}$$

Cohen's *d* represents a standardized mean difference, which is calculated by dividing the difference in means by sources of variation. These values can then be interpreted as a percentage of the standard deviation; for example, a Cohen's *d* value of 0.50 means the difference between two groups equals half a standard deviation [26]. Though conventional guidelines for "small" ($d = 0.20$), "medium" ($d = 0.50$), and "large" ($d = 0.80$) effect sizes were used to provide a general framework for the magnitude of effects that studies were adequately powered to detect [11], raw effect size values were also examined for more precise interpretation. In this review, these effect size values are presented in the context of each study's sensitivity (i.e., power) to detect a range of effects. Importantly, these values do not represent actual effect size results from these studies. Power-determination analyses were also performed across a range of effect sizes ($d = 0.1$–1.0) for each study.

Results

The database search resulted in 1298 studies from Web of Science, 630 studies from PubMed, and 9 from a manual search. Once duplicates were removed, 1376 unique studies remained (Fig. 1). Five studies using multilevel models were excluded since the minimum effect size detectable with 80% power could not be calculated [27–31]. Eighty-nine studies met inclusion criteria, including 39 surface or pharyngeal electrical stimulation [32–70], 14 non-invasive brain stimulation [33, 68, 71–82], 14 respiratory [32, 83–95], nine postural [96–104], six oral sensory stimulation [51, 66, 105–108], five lingual strengthening [109–113], and seven interventions with a combination of treatments [114–120]. Five studies included two treatments [32, 33, 68, 97, 102]; thus, the final number of treatment studies was 94. Fifty-nine studies were randomized controlled trials. The penetration–aspiration scale was the primary outcome of interest in most studies (56%), whereas 21% of studies indicated that it was a secondary outcome. The remaining 23% of studies did not explicitly state whether the penetration–aspiration scale was a primary or secondary outcome. Most (87%) treatment comparisons selected for sensitivity power analyses were within-subject statistical analyses. Eighty-six (91%) treatment comparisons used statistical analyses that provided Cohen's *d* as a measure of effect size, whereas only 3 comparisons used odds ratios (OR) and 5 used an effect size for chi-squared tests (φ). Fifty-nine (63%) treatment comparisons reported a statistically significant result (Table 1). Among studies without a power analysis, 8 studies qualitatively cited low power as a potential reason for a null finding.

Power analyses were reported in 21 studies and thresholds for power ranged from 60 to 90% (Table 1). Two treatment comparisons were powered to detect effect sizes smaller than $d = 0.50$ (Fig. 2). The minimum detectable effect size across studies using a between-subject analysis was $d = 0.58$ for electrical stimulation, $d = 0.74$ for respiratory interventions, $d = 0.74$ for postural maneuvers, $d = 0.93$ for combined treatments, $d = 1.11$ for non-invasive brain stimulation, and $d = 1.15$ for oral sensory stimulation. For studies using a within-subject analysis, the minimum detectable effect size was $d = 0.29$ for electrical stimulation, $d = 0.49$ for postural maneuvers, $d = 0.52$ for non-invasive brain stimulation, $d = 0.61$ for combined treatments, $d = 0.63$ for respiratory interventions, $d = 0.70$ for lingual strengthening, and $d = 0.79$ for oral sensory stimulation. Sixty-seven percent of treatment comparisons were unable to detect effects smaller than $d = 0.80$ with adequate statistical power.

Discussion

Though a variety of treatments to rehabilitate swallowing dysfunction are available to clinicians, inconsistent conclusions across studies obfuscate clinical best practice. This literature is defined by mixed results which may be attributed to inadequate statistical power, affecting a researcher's ability to accurately detect and estimate treatment effects. The present review suggests that swallowing rehabilitation research is generally powered to detect conventionally large effect sizes and not smaller (potentially clinically

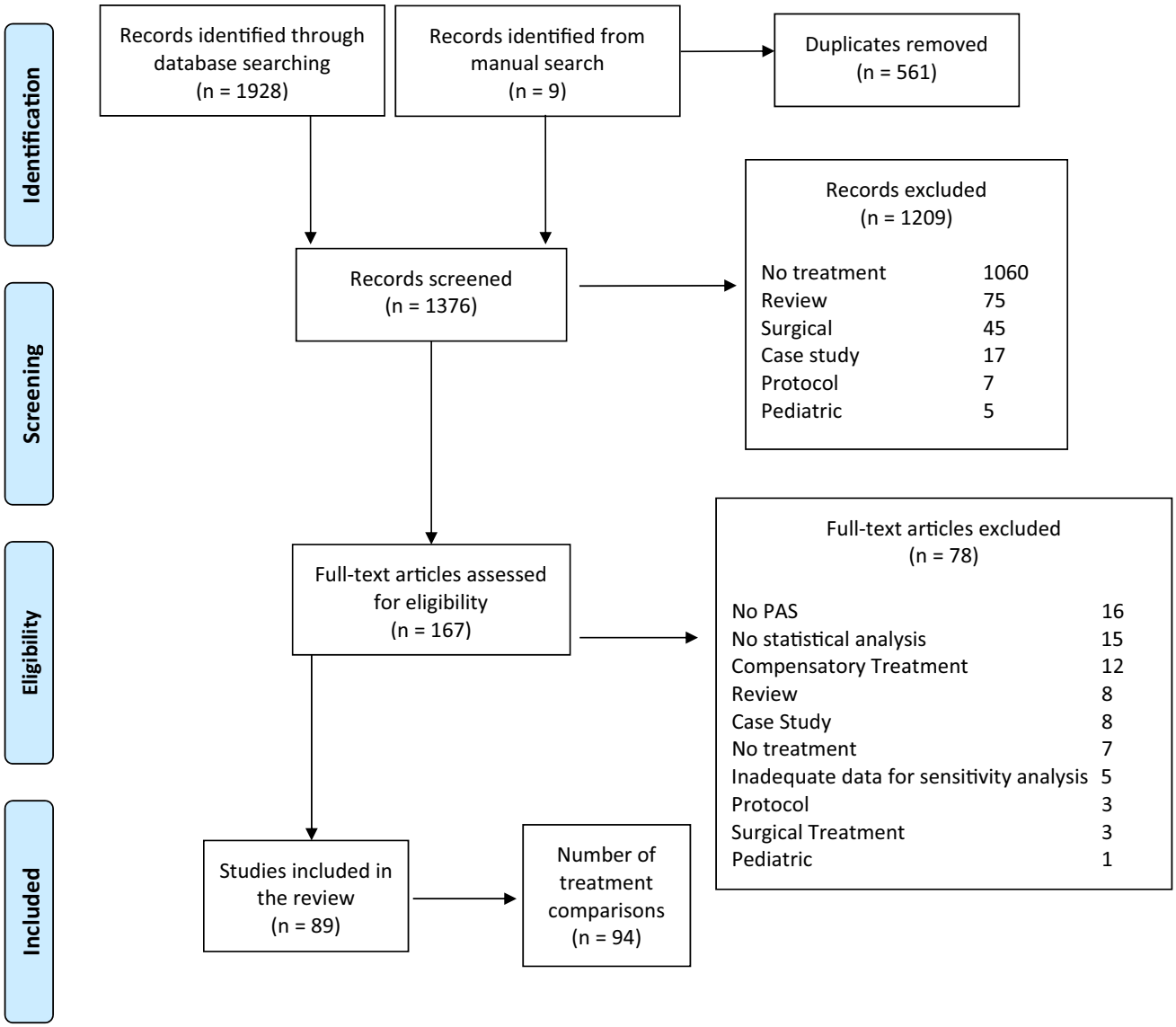

Fig. 1 PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram

meaningful) effects, which may help to explain mixed findings commonly seen in the literature.

Treatments included in this review spanned various domains, including postural maneuvers, non-invasive brain stimulation, and respiratory-based interventions. Across all treatments, adequate sensitivity to detect effects less than $d = 0.50$ was extremely rare. Furthermore, most (67%) treatment comparisons only had sufficient power to detect conventionally 'large' effects (i.e., $d > 0.80$), suggesting that non-significant results may be related to inadequate statistical power to detect smaller, but potentially clinically meaningful effects (Fig. 2). For example, as revealed in this systematic review, non-invasive brain stimulation studies seeking to detect a treatment effect of $d = 0.70$ would have an average of 49% power, meaning that these studies would detect a true treatment effect less than half of the time. In addition to this low sensitivity to detect treatment effects, studies with low statistical power are also more likely to result in inaccurate effect size estimates [15].

Multiple mechanisms of dysfunction, including disordered laryngeal vestibule closure, tongue base retraction, or pharyngeal constriction, often contribute to impairments in functional swallowing outcomes (i.e., aspiration or pharyngeal residue). Regardless of whether a given treatment is designed to target one or many mechanisms of swallowing dysfunction, the multifactorial nature of dysphagia makes it such that a single treatment is unlikely to result in large functional improvements to swallowing. Therefore, power analyses that explicitly specify the smallest treatment effect size of interest (i.e., the minimum amount of change

Table 1 Descriptive statistics and sensitivity power analyses

Author, year	Patient population	Study design (total sample size)	Statistical approach	Comparison	Comparison sample size	PAS treatment	Power analysis reported and threshold	Minimum Cohen's *d* detectable at 80% power
Electrical stimulation								
Arreola, 2021	Stroke	RCT (89)	Wilcoxon signed-rank test	Within-subjects	30	Ordinal	Yes (80%)	0.54
Bath, 2016	Stroke	RCT (129)	Repeated-measures ANOVA	Between-subjects	126	Interval	Yes (90%)	0.50
Bath, 2020	Neurogenic	Observational (236)	Paired *t*-test	Within-subjects	98	Interval	Yes (80%)	0.29
Bhatt, 2015	Head and neck cancer	Observational Retrospective (95)	Independent samples *t*-test	Between-subjects	54 (experimental), 41 (control)	Interval	No	0.59
Bogaardt, 2009	Multiple sclerosis	Observational (25)	Wilcoxon signed-rank test	Within-subjects	25	Ordinal	No	0.60
Everton, 2021	Stroke	RCT (72)	Independent samples *t*-test	Between-subjects	38 (experimental), 34 (control)	Interval	No	0.67
Gallas, 2010	Stroke	Observational (11)	Repeated-measures ANOVA	Within- subjects	11	Interval	No	1.86
Guillen-Sola, 2017	Stroke	RCT (62)	Chi-square test	Between-subjects	17 (experimental), 17 (control)	Categorical (1–5, 6–8)	No	1.25
Hagglund, 2020	Stroke	RCT (32)	Wilcoxon signed-rank test	Within-subjects	18	Ordinal	Yes (80%)	0.72
Huang, 2014	Stroke	RCT (29)	Repeated-measures ANOVA	Within-subjects	10	Ordinal	No	1.99
Jayasekeran, 2010	Stroke	RCT (50)	Mann–Whitney U test	Between-subjects	22 (experimental), 28 (control)	Ordinal	Yes (80%)	0.83
Jeon, 2020	Stroke	RCT (34)	Repeated-measures ANOVA	Within-subjects	17	Interval	Yes (80%)	0.99
Ko, 2016	Stroke and traumatic brain injury	Observational (28)	Repeated-measures ANOVA	Within-subjects	12	Interval	No	1.94
Langmore, 2015	Head and neck cancer	RCT (116)	Repeated-measures ANCOVA	Within- subjects	54	Interval	No	0.50
Lee, 2015	Heterogenous	Observational (15)	Wilcoxon signed-rank test	Within-subjects	15	Ordinal	No	0.80
Lee, 2019	Stroke	RCT (40)	Wilcoxon signed-rank test	Within- subjects	20	Ordinal	No	0.68
Lee, 2021	Stroke, brain tumor, encephalitis	RCT (49)	Paired *t*-test	Within-subjects	26	Interval	Yes (80%)	0.57
Lim, 2009	Stroke	RCT (28)	Wilcoxon signed-rank test	Within-subjects	16	Ordinal	No	0.77

Table 1 (continued)

Author, year	Patient population	Study design (total sample size)	Statistical approach	Comparison	Comparison sample size	PAS treatment	Power analysis reported and threshold	Minimum Cohen's *d* detectable at 80% power
Lim, 2014	Stroke	RCT (47)	Mann–Whitney U test	Between-subjects	18 (experimental), 15 (control)	Ordinal	No	1.04
Lin, 2011	Head and neck cancer	RCT (20)	Paired *t*-test	Within-subjects	10	Interval	No	1.00
Ludlow, 2007	Brain injury, cardiovascular disease, brain tumor, Parkinson's disease	Crossover Design (11)	Paired *t*-test	Within-subjects	10	Interval	No	0.85
Martindale, 2019	Stroke and non-stroke	Observational (43)	Repeated-measures ANOVA	Within- subjects	43	Interval	No	0.88
Michou, 2014	Stroke	RCT (18)	Wilcoxon signed-rank test	Within-subjects	6	Ordinal	No	1.49
Miller, 2021	Stroke	RCT (12)	Wilcoxon signed-rank test	Within-subjects	12	Ordinal	No	0.91
Mituuti, 2018	Stroke	Observational (10)	Friedman's ANOVA	Within-subjects	10	Ordinal	No	1.99
Oh, 2019	Stroke	RCT (26)	Paired *t*-test	Within- subjects	14	Interval	No	0.81
Ortega, 2016	Older adults	RCT (38)	Chi-square test	Between-subjects	19 (experimental), 19 (comparison)	Categorical	No	1.15
Park, 2012	Stroke	RCT (18)	Wilcoxon signed-rank test	Within-subjects	9	Ordinal	No	1.10
Park, 2016	Stroke	RCT (50)	Paired *t*-test	Within-subjects	25	Interval	Yes (80%)	0.58
Park, 2018	Parkinson's disease	RCT (18)	Wilcoxon signed-rank test	Within- subjects	9	Ordinal	No	1.10
Park, 2019	Stroke	Observational (10)	Wilcoxon signed-rank test	Within-subjects	10	Ordinal	Yes (80%)	1.03
Restivo, 2013	Multiple sclerosis	RCT (20)	Wilcoxon signed-rank test	Within-subjects	10	Ordinal	No	1.03
Rofes, 2013	Stroke	RCT (20)	Wilcoxon signed-rank test	Within- subjects	10	Ordinal	No	1.03
Seo, 2021	Stroke	RCT (23)	Wilcoxon signed-rank test	Within-subjects	12	Ordinal	No	0.91
Simonelli, 2019	Stroke	RCT (31)	Mann–Whitney U test	Between-subjects	16 (experimental), 15 (control)	Ordinal	No	1.07
Sun, 2013	Stroke	Observational (29)	Wilcoxon signed-rank test	Within-subjects	29	Ordinal	Yes (80%)	0.55

Table 1 (continued)

Author, year	Patient population	Study design (total sample size)	Statistical approach	Comparison	Comparison sample size	PAS treatment	Power analysis reported and threshold	Minimum Cohen's *d* detectable at 80% power
Terre, 2015	Traumatic brain injury	RCT (20)	Wilcoxon signed-rank test	Within- subjects	10	Ordinal	No	1.03
Vasant, 2016	Stroke	RCT (35)	Logistic regression	Between-subjects	35	Categorical (1–2, 3–8)	Yes (80%)	1.45
Verin, 2011	Stroke, multiple sclerosis, Parkinson's disease, progressive supranuclear palsy	Crossover Design (11)	Wilcoxon signed-rank test	Within-subjects	13	Ordinal	No	0.87
Non-invasive brain stimulation								
Khedr, 2019	Parkinson's disease	RCT (30)	Paired *t*-test	Within-subjects	19	Interval	Yes (80%)	0.68
Kim, 2011	Traumatic brain injury	RCT (30)	Wilcoxon signed-rank test	Within-subjects	10	Ordinal	No	1.03
Lee, 2015	Stroke	RCT (24)	Repeated-measures ANOVA	Within-subjects	12	Interval	No	1.20
Lim, 2014	Stroke	RCT (47)	Mann–Whitney U test	Between-subjects	14 (experimental), 15 (control)	Ordinal	No	1.11
Lin, 2018	Stroke	RCT (28)	Wilcoxon signed-rank test	Within-subjects	13	Ordinal	Yes (80%)	0.87
Michou, 2012	Stroke	Observational (6)	Wilcoxon signed-rank test	Within-subjects	6	Ordinal	No	1.49
Michou, 2014	Stroke	RCT (18)	Wilcoxon signed-rank test	Within-subjects	6	Ordinal	No	1.49
Park, 2013	Stroke	RCT (18)	Wilcoxon signed-rank test	Within-subjects	9	Ordinal	No	1.10
Park, 2017	Stroke	RCT (33)	Paired *t*-test	Within-subjects	11	Interval	No	0.94
Park, 2019	Geriatric	Observational (8)	Wilcoxon signed-rank test	Within-subjects	8	Ordinal	No	1.19
Restivo, 2019	Multiple sclerosis	RCT (18)	Wilcoxon signed-rank test	Within-subjects	9	Ordinal	No	1.10
Unluer, 2019	Stroke	RCT (28)	Friedman's ANOVA	Within-subjects	15	Ordinal	Yes (80%)	0.78
Verin, 2008	Stroke	Observational (7)	Repeated-measures ANOVA	Within-subjects	7	Interval	No	2.55
Zhong, 2021	Stroke	RCT (147)	Repeated-measures ANOVA	Within-subjects	36	Interval	No	0.51

Table 1 (continued)

Author, year	Patient population	Study design (total sample size)	Statistical approach	Comparison	Comparison sample size	PAS treatment	Power analysis reported and threshold	Minimum Cohen's *d* detectable at 80% power
Respiratory interventions								
Arnold, 2020	Stroke	Observational (20)	Paired *t*-test	Within-subjects	10	Interval	No	1.00
Eom, 2017	Stroke	RCT (26)	Wilcoxon signed-rank test	Within-subjects	13	Ordinal	Yes (60%)	0.87
Guillen-Sola, 2017	Stroke	RCT (62)	Chi-square test	Between-subjects	16 (experimental), 17 (control)	Categorical (1–4; 5–8)	No	1.28
Hegland, 2016	Stroke	Observational (12)	Repeated-measures ANOVA	Within-subjects	12	Interval	No	1.78
Hutcheson, 2018	Head and neck cancer	Observational (64)	Wilcoxon signed-rank test	Within-subjects	23	Ordinal	Yes (90%)	0.63
Jang, 2019	Stroke	RCT (32)	Wilcoxon signed-rank test	Within-subjects	18	Ordinal	No	0.72
Martin-Harris, 2015	Head and neck cancer	Observational (30)	Test of Proportions	Within-subjects	30	Categorical	Yes (80%)	0.93
Mohannak, 2020	Inclusion Body Myositis	Observational (12)	Paired *t*-test	Within-subjects	12	Interval	No	0.89
Moon, 2017	Stroke	RCT (18)	Wilcoxon signed-rank test	Within-subjects	9	Ordinal	No	1.10
Park, 2016	Stroke	RCT (27)	Wilcoxon signed-rank test	Within-subjects	14	Ordinal	No	0.83
Pitts, 2009	Parkinson's disease	Observational (10)	Wilcoxon signed-rank test	Within-subjects	10	Ordinal	No	1.03
Plowman, 2016	ALS	Observational (15)	Repeated-measures ANOVA	Within-subjects	15	Interval	No	1.69
Plowman, 2019	ALS	RCT (46)	Chi-square test	Between-subjects	23 (experimental), 23 (control)	Categorical (1–2, 3–8)	No	1.00
Troche, 2010	Parkinson's disease	RCT (60)	Repeated-measures ANCOVA	Between-subjects	30 (experimental), 30 (control)	Interval	Yes (80%)	0.74
Combined treatments								
Balou, 2019	Older adults	Observational (9)	Wilcoxon signed-rank test	Within-subjects	9	Ordinal	No	1.1
Furuie, 2019	Head and neck cancer	Observational (30)	Independent samples *t*-test	Between-subjects	30 (experimental), 30 (control)	Interval	No	1.06
Hsiang, 2019	Head and neck cancer	RCT (40)	Mann–WhitneyU test	Between-subjects	20 (experimental), 20 (control	Ordinal	Yes (80%)	0.93
Kraaijenga, 2017	Head and neck cancer	Observational (17)	Paired *t*-test	Within-subjects	17	Interval	Yes (80%)	0.72

Table 1 (continued)

Author, year	Patient population	Study design (total sample size)	Statistical approach	Comparison	Comparison sample size	PAS treatment	Power analysis reported and threshold	Minimum Cohen's *d* detectable at 80% power
Tarameshlu, 2019	Multiple sclerosis	RCT (20)	Independent samples *t*-test	Between-subjects	10 (experimental), 10 (control	Interval	No	1.32
van der Molen, 2011	Head and neck cancer	RCT (49)	Wilcoxon signed-rank test	Within-subjects	24	Ordinal	No	0.61
van der Molen, 2014	Head and neck cancer	RCT (49)	McNemar test	Within-subjects	29	Categorical	No	0.93
Lingual strengthening								
Kim, 2017	Stroke	RCT (35)	Paired *t*-test	Within-subjects	18	Interval	No	0.70
Namiki, 2019	Geriatric	Observational (18)	Wilcoxon signed-rank test	Within-subjects	18	Ordinal	Yes (80%)	0.72
Robbins, 2005	Geriatric	Observational (10)	Repeated-measures ANCOVA	Within-subjects	10	Interval	No	1.99
Robbins, 2007	Stroke	Observational (10)	Paired *t*-test	Within-subjects	10	Interval	No	1.00
Steele, 2016	Stroke	RCT (11)	Friedman's ANOVA	Within-subjects	6	Ordinal	Yes (NR)	1.43
Postural maneuvers								
Choi, 2017	Stroke	RCT (32)	Paired *t*-test	Within-subjects	16	Interval	Yes (60%)	0.75
Gao, 2017	Stroke	RCT (90)	Repeated-measures ANOVA	Between-subjects	30 (experimental), 30 (control)	Interval	No	0.67
Kim, 2019	Stroke	RCT (25)	Wilcoxon signed-rank test	Within-subjects	12	Ordinal	Yes (60%)	0.91
Mano, 2015	Spinal and bulbar muscular atrophy	Observational (6)	Paired *t*-test	Within-subjects	6	Interval	No	1.43
Park, 2017	Stroke	RCT (37)	Paired *t*-test	Within-subjects	19	Interval	Yes (80%)	0.68
Park, 2018	Stroke	RCT (22)	Wilcoxon signed-rank test	Within-subjects	11	Ordinal	No	0.97
Park, 2019	Stroke	RCT (37)	Paired *t*-test	Within-subjects	18	Interval	Yes (80%)	0.70
Park, 2020	Stroke	RCT (20)	Wilcoxon signed-rank test	Within-subjects	15	Ordinal	Yes (60%)	0.80
Ploumis, 2018	Stroke	RCT (70)	Wilcoxon signed-rank test	Within-subjects	37	Ordinal	No	0.49
Oral sensory stimulation								
Jakobsen, 2019	Brain injury	RCT (10)	Wilcoxon signed-rank test	Within-subjects	5	Ordinal	No	1.76

Table 1 (continued)

Author, year	Patient population	Study design (total sample size)	Statistical approach	Comparison	Comparison sample size	PAS treatment	Power analysis reported and threshold	Minimum Cohen's *d* detectable at 80% power
Ortega, 2016	Older adults	RCT (38)	Chi-square test	Between-subjects	19 (experimental), 19 (comparison)	Categorical	No	1.15
Power, 2006	Stroke	RCT (16)	Repeated-measures ANOVA	Within-subjects	8	Interval	No	1.51
Rosenbek, 1998	Stroke	RCT (45)	Paired *t*-test	Within-subjects	13	Interval	No	0.85
Tomsen, 2019	Older adults	RCT (28)	Paired *t*-test	Within-subjects	7	Interval	No	1.27

RCT randomized controlled trial, *NR* not reported, *ANOVA* analysis of variance, *ANCOVA* analysis of covariance

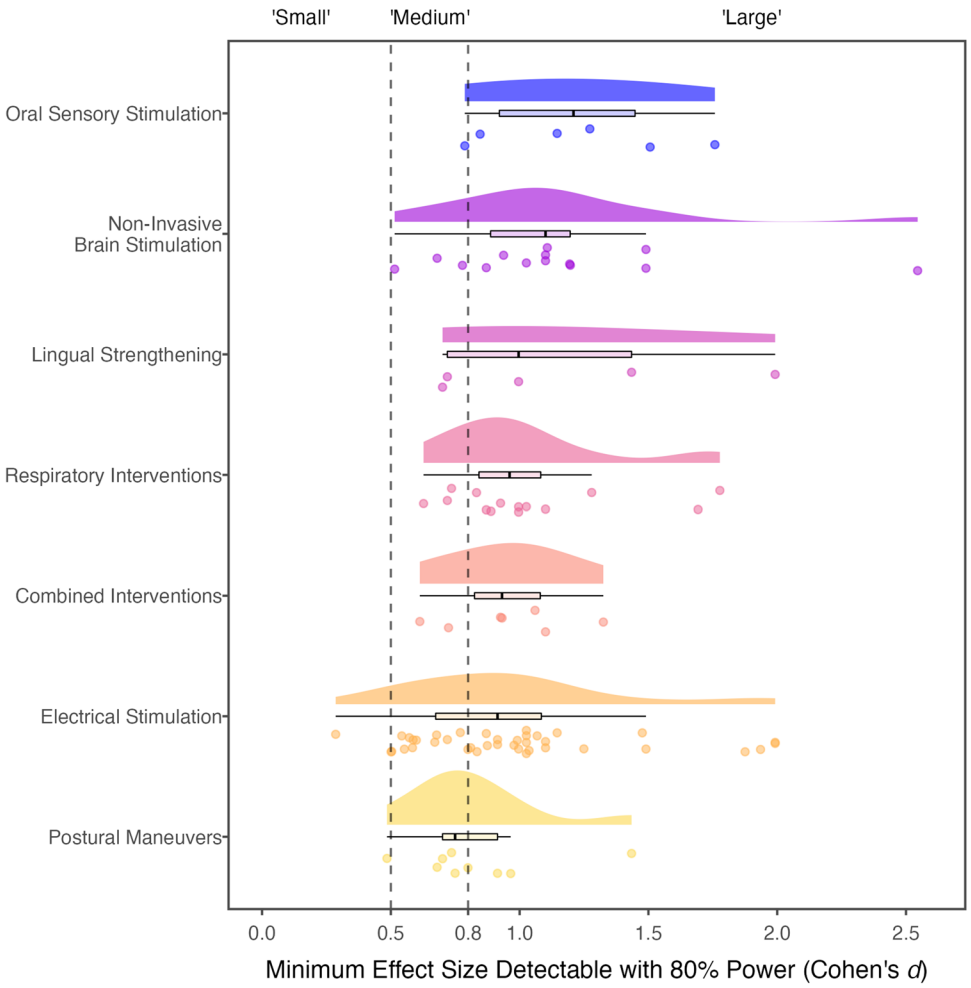

Fig. 2 Minimum Effect Size Detectable with 80% Power Across Treatments. [1]The ability of a study to detect smaller effect sizes is desired. [2]Cohen's *d* conventional benchmarks (i.e., "small," "medium," and "large) are provided for general interpretation. However, these guidelines are relative concepts and depend on clinical significance in the context of a given research question

in an outcome that is meaningful for a study to detect) are imperative to ensure that a study is not only informative, but also falsifiable. This central component of study design and power analyses requires careful consideration to ensure clinically meaningful effects have a high likelihood of detection and accurate estimation given the complex nature of dysphagia.

Rehabilitation research poses significant challenges to one of the most conventional methods of increasing statistical power in treatment studies—the recruitment of large patient samples. Barriers that prohibit merely increasing the sample size include, but are not limited to, the financial and ethical burden of large-scale clinical trials, the rarity of many diseases which result in dysphagia, and heightened variability between and within patient populations [121]. In order to reduce the impact of these barriers, non-conventional analyses and study designs, such as one-tailed statistical tests, multilevel models, and sequential designs, have been proposed as alternative approaches to increase power [122, 123].

Though one-tailed tests are not common practice in the field of dysphagia, when specified a priori they can be a valid approach to maximize statistical power. One-tailed tests are beneficial if an effect is hypothesized to exist in only one direction and the opposite direction is not interesting nor expected. To achieve 80% power, a two-sided test would require a 20% larger sample size compared to a one-sided test. In this sense, one-sided statistical tests maximize data collection efficiency [124]. For example, in one of the studies included in this review, Ludlow and colleagues used a one-tailed t-test with a sample size of 8 participants [61], which afforded a minimum detectable effect size of $d = 0.98$ compared to $d = 1.16$ with a two-sided approach.

Multilevel models, also known as mixed effects or hierarchical models, are another approach to potentially increase statistical power [125]; however, they are rarely utilized in the dysphagia treatment literature (five out of 99 studies in this review). Whereas common statistical tests (e.g., *t*-tests, ANOVA, etc.) require aggregating multiple trials of an outcome to ensure a single data point represents each participant, multilevel models avoid aggregation. This effectively increases the sample size by including repeated trials while also allowing for analyses at the participant level.

Sequential analyses are a common approach in medical trials to optimize data collection efficiency (e.g., [126]). In this design, an a priori power analysis is performed and various data analysis time points (e.g., interim analysis) are prespecified with explicit methods to control the type 1 error rate [123]. A major benefit is that data collection can often be stopped early (i.e., before the sample size specified in the power analysis is reached) given a reasonably high chance of observing a statistically significant finding after collecting less than half of the sample size [123]. Though this type of design is beneficial for investigating whether a treatment effect might exist, effect sizes obtained from interim analyses are subject to the same small sample bias as underpowered studies and may require adjustments or follow-up studies to obtain an accurate effect size estimate [127].

Though power analyses were only reported in 20% of studies in this review, many qualitatively cited "low statistical power" as a reason for obtaining a null finding. However, none of these studies provided a quantitative analysis of the sensitivity of the study design and data to detect a treatment effect. Sensitivity power analyses are one approach to enhance one's understanding of the range of treatment effect sizes that could be reliably detected with an analysis, improving the interpretation of null findings. A sensitivity power analysis is dependent on the statistical analysis approach and provides the minimum detectable effect size given the desired level of power, alpha level, and sample size. For example, if a sensitivity power analysis reveals that a study has 80% power to detect $d = 0.40$ yet finds a non-significant result, then treatment effects larger than $d = 0.40$ are unlikely and treatment effects lower than $d = 0.40$ are possible, but the study design was insufficient to detect them. A major benefit of sensitivity power analyses is that they do not increase researcher burden since they can be performed after data are collected. This type of power analysis implicitly recognizes that resources are limited, and sample size is often based on feasibility constraints. Though sensitivity power analyses can be easily performed for common statistical tests with current software (e.g., [24, 128]), multilevel models require a Monte Carlo simulation approach [129]. A lack of software to perform these simulation-based power analyses, particularly with ordinal outcomes, is a substantial barrier for clinical researchers. Therefore, we have provided a brief supplemental tutorial for simulation-based power analyses with ordinal outcomes for both non-parametric tests (Mann–Whitney *U* and Wilcoxon signed-rank tests) and mixed effects (cumulative link) models (https://osf.io/e6usd/).

A common approach to reconcile multiple treatment studies with mixed findings is to perform a systematic review. These reviews attempt to synthesize available evidence, ultimately providing an assessment of a treatment's efficacy. However, systematic reviews rarely acknowledge statistical power. If underpowered studies predominate, then conclusions based solely on the number of studies that reported a statistically significant result will be biased. An alternate approach is to combine studies in a meta-analysis to provide an overall summary effect. In the field of dysphagia; however, this approach is often untenable due to substantial heterogeneity in study design, patient populations, statistical analyses, assessment types, and swallowing tasks. Furthermore, direct replication studies are exceedingly rare. These barriers prohibit implementing

rigorous meta-analyses to inform patient care. One potential solution which has garnered interest in other fields is open data sets [130]. This not only ensures transparency and reproducibility, but also facilitates meta-analyses. Data sharing provides substantial benefits to the research community, most notably in the presence of mixed results, heterogenous studies, and a growing knowledge base.

There are several limitations to acknowledge in this review. Our results are specific to the penetration–aspiration scale. We acknowledge that interventions may not have been powered or designed to target this outcome. Instead, other outcomes may have been more appropriate given a study's research question. We chose the penetration–aspiration scale as our outcome of interest due to its widespread use in dysphagia management, which permitted inclusion of a large number of studies. Prior studies examining statistical power within a given field have used the summary effect size from meta-analyses as the "true effect" in their power analysis [12, 131]. However, this approach was not feasible in the dysphagia treatment literature due to a low number of meta-analyses. Furthermore, meta-analysis estimates from studies with predominantly low power may not reflect the true population effect. Instead, we used an approach to detect the sensitivity of each study by determining the minimum effect size detectable with 80% power. We used Cohen's *d* as the measure of effect size to summarize sensitivity across studies but acknowledge that conversion between effect sizes may affect their interpretation. Additionally, we assumed a "moderate" correlation for time points for within-subject statistical tests (e.g., Wilcoxon signed-rank test) and acknowledge that different magnitudes of within-subject correlations across studies may have affected our effect size estimates from sensitivity power analyses. However, studies did not commonly report this correlation which prohibited uniformly incorporating it into our analyses. Studies included in this review included diverse methodologies and analyses which may have affected their sensitivity to detect effects, such as the type of statistical test, level of comparison, alpha level, and statistical use of the penetration–aspiration scale (i.e., interval, ordinal, or categorical). Since we used an approach that maximized the sensitivity of each study, this may have overestimated statistical power, most notably in situations where parametric analyses (i.e., Cohen's *d*) were used. However, we were unable to perform re-calculations with appropriate statistical analyses without access to the original data. We used conventional guidelines for "small," "moderate," and "large" Cohen's *d* when interpreting minimum detectable effect sizes, though we recognize that these benchmarks are relative concepts and fully dependent on one's subfield, research context, and the smallest effect size of interest. The use of these effect size benchmarks may result in misrepresentation of the smallest effect size of interest for a given study's primary aim and outcome of interest. However, understanding the smallest effect size of interest for each study is not necessary to evaluate power across swallowing rehabilitation research. Future research will be necessary to better define clinically significant change in swallowing outcomes in order to inform meaningful effect sizes for power analyses.

Conclusions

Though statistical power is a central component of study design, power analyses are infrequently reported in swallowing rehabilitation research. The current review suggests that swallowing interventions examining the penetration–aspiration scale are generally powered to only reliably detect larger effect sizes, whereas smaller (but potentially clinically meaningful) effects have a low likelihood of detection. These findings may help to explain mixed results commonly seen in the dysphagia treatment literature. Non-conventional study designs and statistical analyses may be important considerations to increase power in smaller samples. To promote higher levels of evidence in the context of meta-analysis, open data sets and transparent reporting may also improve the quality of inferences. Moving forward, a comprehensive understanding of clinically meaningful change in swallowing outcomes should be a priority to not only assist in sample size justifications, but also to ensure falsifiable and impactful findings that inform clinical practice.

Funding N/A.

Data Availability and Supplemental Power Analysis Tutorial The data analysis script and a supplemental tutorial on performing simulation-based power analyses with ordinal outcomes are available on the Open Science Framework at the following url: https://osf.io/65atf/.

Declarations

Conflict of interest All authors have no conflict of interest to disclose.

References

1. Mancopes R, Smaoui S, Steele CM. Effects of expiratory muscle strength training on videofluoroscopic measures of swallowing: a systematic review. Am J Speech Lang Pathol. 2020;30:1–22.
2. Langmore SE, Pisegna JM. Efficacy of exercises to rehabilitate dysphagia: a critique of the literature. Int J Speech Lang Pathol. 2015;17(3):222–9.

3. Bothe AK, Richardson JD. Statistical, practical, clinical, and personal significance: definitions and applications in speech-language pathology. Am J Speech Lang Pathol. 2011;20(3):233–42.
4. Hedström J, Tuomi L, Andersson M, Dotevall H, Osbeck H, Finizia C. Within-bolus variability of the penetration-aspiration scale across two subsequent swallows in patients with head and neck cancer. Dysphagia. 2017;32(5):683–90.
5. Molfenter SM, Steele CM. Physiological variability in the deglutition literature: hyoid and laryngeal kinematics. Dysphagia. 2011;26(1):67–74.
6. Molfenter SM, Steele CM. Temporal variability in the deglutition literature. Dysphagia. 2012;27(2):162–77.
7. Ahadi S, Diener E. Multiple determinants and effect size. J Pers Soc Psychol. 1989;56(3):398–406.
8. Neyman J, Pearson ES. The testing of statistical hypotheses in relation to probabilities a priori. Math Proc Camb Philos Soc. 1933;29(4):492–510.
9. Bacchetti P. Current sample size conventions: flaws, harms, and alternatives. BMC Med. 2010;8(1):17.
10. Neyman J. Frequentist probability and frequentist statistics. Synthese. 1977;36(1):97–131.
11. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale: L. Erlbaum Associates; 1988.
12. Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, et al. Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci. 2013;14(5):365–76.
13. Szucs D, Ioannidis JPA. Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. PLoS Biol. 2017;15(3):1–18.
14. Yarkoni T. Big correlations in little studies: inflated fMRI correlations reflect low statistical power—commentary on Vul et al. (2009). Perspect Psychol Sci. 2009;4(3):294–8.
15. Gelman A, Carlin J. Beyond power calculations: assessing type S (sign) and type M (magnitude) errors. Perspect Psychol Sci. 2014;9(6):641–51.
16. Ioannidis JPA. Why most published research findings are false. PLoS Med. 2005;2(8):696–701.
17. Ioannidis JPA, Tarone R, McLaughlin JK. The false-positive to false-negative ratio in epidemiologic studies. Epidemiology. 2011;22(4):450–6.
18. Ioannidis JPA. Why most discovered true associations are inflated. Epidemiology. 2008;19(5):640–8.
19. Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan A-W, Cronin E, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. PLoS ONE. 2008;3(8):1–31.
20. Borders JC, Brates D. Use of the penetration-aspiration scale in dysphagia research: a systematic review. Dysphagia. 2019;35(4):583–97.
21. Rosenbek JC, Robbins JA, Roecker EB, Coyle JL, Wood JL. A penetration-aspiration scale. Dysphagia. 1996;11:93–8.
22. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev. 2015;4:1–9.
23. R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2018. https://www.R-project.org/.
24. Faul F, Erdfelder E, Lang A-G, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods. 2007;39(2):175–91.
25. Caldwell AR, Lakens D, Parlett-Pelleriti CM. Power analysis with superpower. 2019. https://aaroncaldwell.us/SuperpowerBook/. Accessed 20 Nov 2020.
26. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front Psychol. 2013. https://doi.org/10.3389/fpsyg.2013.00863.
27. Baijens LWJ, Speyer R, Passos VL, Pilz W, van der Kruis J, Haarmans S, et al. Surface electrical stimulation in dysphagic parkinson patients: a randomized clinical trial: electrical stimulation for dysphagia in Parkinson. Laryngoscope. 2013;123(11):E38-44.
28. Rogus-Pulia N, Rusche N, Hind JA, Zielinski J, Gangnon R, Safdar N, et al. Effects of device-facilitated isometric progressive resistance oropharyngeal therapy on swallowing and health-related outcomes in older adults with dysphagia. J Am Geriatr Soc. 2016;64(2):417–24.
29. Wall LR, Ward EC, Cartmill B, Hill AJ, Isenring E, Byrnes J, et al. Prophylactic swallowing therapy for patients with head and neck cancer: a three-arm randomized parallel-group trial. Head Neck. 2020;42(5):873–85.
30. Messing BP. Prophylactic swallow therapy for patients with head and neck cancer undergoing chemoradiotherapy: a randomized trial. Dysphagia. 2017;32(4):487–500.
31. Van Daele DJ, Langmore SE, Krisciunas GP, Lazarus CL, Pauloski BR, McCulloch TM, et al. The impact of time after radiation treatment on dysphagia in patients with head and neck cancer enrolled in a swallowing therapy program. Head Neck. 2019;41(3):606–14.
32. Guillén-Solà A, Messagi Sartor M, Bofill Soler N, Duarte E, Barrera MC, Marco E. Respiratory muscle strength training and neuromuscular electrical stimulation in subacute dysphagic stroke patients: a randomized controlled trial. Clin Rehabil. 2017;31(6):761–71.
33. Lim KB, Lee HJ, Yoo J, Kwon YG. Effect of low-frequency rTMS and NMES on subacute unilateral hemispheric stroke with dysphagia. Ann Rehabil Med. 2014;38(5):592–602.
34. Terré R, Mearin F. A randomized controlled study of neuromuscular electrical stimulation in oropharyngeal dysphagia secondary to acquired brain injury. Eur J Neurol. 2015;22(4):687-e44.
35. Simonelli M, Ruoppolo G, Iosa M, Morone G, Fusco A, Grasso MG, et al. A stimulus for eating. The use of neuromuscular transcutaneous electrical stimulation in patients affected by severe dysphagia after subacute stroke: a pilot randomized controlled trial. NeuroRehabilitation. 2019;44:103–310.
36. Lee HY, Hong JS, Lee KC, Shin YK, Cho SR. Changes in hyolaryngeal movement and swallowing function after neuromuscular electrical stimulation in patients with dysphagia. Ann Rehabil Med. 2015;39(2):199–209.
37. Sun S-F, Hsu C-W, Lin H-S, Sun H-P, Chang P-H, Hsieh W-L, et al. Combined neuromuscular electrical stimulation (NMES) with fiberoptic endoscopic evaluation of swallowing (FEES) and traditional swallowing rehabilitation in the treatment of stroke-related dysphagia. Dysphagia. 2013;28(4):557–66.
38. Ko KR, Park HJ, Hyun JK, Seo IH, Kim TU. Effect of laryngopharyngeal neuromuscular electrical stimulation on dysphonia accompanied by dysphagia in post-stroke and traumatic brain injury patients: a pilot study. Ann Rehabil Med. 2016;40(4):600–10.
39. Park J-S, Hwang N-K, Kim H-H, Lee G, Jung Y-J. Effect of neuromuscular electrical stimulation combined with effortful swallowing using electromyographic biofeedback on oropharyngeal swallowing function in stroke patients with dysphagia: a pilot study. Medicine. 2019;98(44):1–6.
40. Rofes L, Arreola V, López I, Martin A, Sebastián M, Ciurana A, et al. Effect of surface sensory and motor electrical stimulation on chronic poststroke oropharyngeal dysfunction. Neurogastroenterol Motil. 2013;25(11):888–96.
41. Lin PH, Hsiao TY, Chang YC, Ting LL, Chen WS, Chen SC, et al. Effects of functional electrical stimulation on dysphagia

caused by radiation therapy in patients with nasopharyngeal carcinoma. Support Care Cancer. 2011;19(1):91–9.
42. Jeon YH, Cho KH, Park SJ. Effects of neuromuscular electrical stimulation (NMES) plus upper cervical spine mobilization on forward head posture and swallowing function in stroke patients with dysphagia. Brain Sci. 2020;10(8):1–10.
43. Park JS, Oh DH, Hwang NK, Lee JH. Effects of neuromuscular electrical stimulation combined with effortful swallowing on post-stroke oropharyngeal dysphagia: a randomised controlled trial. J Oral Rehabil. 2016;43(6):426–34.
44. Lee KW, Kim SB, Lee JH, Lee SJ, Park JG, Jang KW. Effects of neuromuscular electrical stimulation for masseter muscle on oral dysfunction after stroke. Ann Rehabil Med. 2019;43(1):11–8.
45. Park JS, Oh DH, Hwang NK, Lee JH. Effects of neuromuscular electrical stimulation in patients with Parkinson's disease and dysphagia: a randomized, single-blind, placebo-controlled trial. NeuroRehabilitation. 2018;42(4):457–63.
46. Mituuti CT, Arone MMAS, Rosa RR, Berretin-Felix G. Effects of sensory neuromuscular electrical stimulation on swallowing in the elderly affected by stroke. Top Geriatr Rehabil. 2018;34(1):71–81.
47. Langmore SE, McCulloch TM, Krisciunas GP, Lazarus CL, Van Daele DJ, Pauloski BR, et al. Efficacy of electrical stimulation and exercise for dysphagia in patients with head and neck cancer: a randomized clinical trial. Head Neck. 2015;38:1221–31.
48. Park JW, Kim Y, Oh JC, Lee HJ. Effortful swallowing training combined with electrical stimulation in post-stroke dysphagia: a randomized controlled study. Dysphagia. 2012;27(4):521–7.
49. Huang KL, Liu TY, Huang YC, Leong CP, Lin WC, Pong YP. Functional outcome in acute stroke patients with oropharyngeal dysphagia after swallowing therapy. J Stroke Cerebrovasc Dis. 2014;23(10):2547–53.
50. Bhatt AD, Goodwin N, Cash E, Bhatt G, Silverman CL, Spanos WJ, et al. Impact of transcutaneous neuromuscular electrical stimulation on dysphagia in patients with head and neck cancer treated with definitive chemoradiation. Head Neck. 2015;37(7):1051–6.
51. Lim KB, Lee HJ, Lim SS, Choi YI. Neuromuscular electrical and thermal-tactile stimulation for dysphagia caused by stroke: a randomized controlled trial. J Rehabil Med. 2009;41(3):174–8.
52. Martindale N, Stephenson J, Pownall S. Neuromuscular electrical stimulation plus rehabilitative exercise as a treatment for dysphagia in stroke and non-stroke patients in an NHS setting: feasibility and outcomes. Geriatrics. 2019;4(4):53.
53. Gallas S, Marie JP, Leroi AM, Verin E. Sensory transcutaneous electrical stimulation improves post-stroke dysphagic patients. Dysphagia. 2010;25(4):291–7.
54. Oh D-H, Park J-S, Kim H-J, Chang M-Y, Hwang N-K. The effect of neuromuscular electrical stimulation with different electrode positions on swallowing in stroke patients with oropharyngeal dysphagia: a randomized trial. J Back Musculoskelet Rehabil. 2019;33(4):637–44.
55. Bogaardt H, van Dam D, Wever NM, Bruggeman CE, Koops J, Fokkens WJ. Use of neuromuscular electrostimulation in the treatment of dysphagia in patients with multiple sclerosis. Ann Otol Rhinol Laryngol. 2009;118(4):241–6.
56. Arreola V, Ortega O, Álvarez-Berdugo D, Rofes L, Tomsen N, Cabib C, et al. Effect of transcutaneous electrical stimulation in chronic poststroke patients with oropharyngeal dysphagia: 1-year results of a randomized controlled trial. Neurorehabil Neural Repair. 2021;17:154596832110231.
57. Bath PM, Scutt P, Love J, Clavé P, Cohen D, Dziewas R, et al. Pharyngeal electrical stimulation for treatment of dysphagia in subacute stroke: a randomized controlled trial. Stroke. 2016;47(6):1562–70.
58. Everton LF, Benfield JK, Michou E, Hamdy S, Bath PM. Effects of pharyngeal electrical stimulation on swallow timings, clearance and safety in post-stroke dysphagia: analysis from the swallowing treatment using electrical pharyngeal stimulation (STEPS) trial. Stroke Res Treat. 2021;2021:1–8.
59. Hägglund P, Hägg M, Levring Jäghagen E, Larsson B, Wester P. Oral neuromuscular training in patients with dysphagia after stroke: a prospective, randomized, open-label study with blinded evaluators. BMC Neurol. 2020;20(1):405.
60. Jayasekeran V, Singh S, Tyrrell P, Michou E, Jefferson S, Mistry S, et al. Adjunctive functional pharyngeal electrical stimulation reverses swallowing disability after brain lesions. Gastroenterology. 2010;138(5):1737–46.
61. Lee SY, Park D, Jang J, Jang EG, Lee JC, Park Y, et al. Compensatory effects of sequential 4-channel neuromuscular electrical stimulation for the treatment of acute, subacute, and chronic dysphagia in a prospective, double-blinded randomized clinical trial. Neurorehabil Neural Repair. 2021. https://doi.org/10.1177/15459683211029891.
62. Ludlow CL, Humbert I, Saxon K, Poletto C, Sonies B, Crujido L. Effects of surface electrical stimulation both at rest and during swallowing in chronic pharyngeal dysphagia. Dysphagia. 2007;22(1):1–10.
63. Restivo DA, Casabona A, Centonze D, Marchese-Ragona R, Maimone D, Pavone A. Pharyngeal electrical stimulation for dysphagia associated with multiple sclerosis: a pilot study. Brain Stimul. 2013;6(3):418–23.
64. Seo K-H, Jang J, Jang EG, Park Y, Lee SY, Kim BR, et al. Clinical effectiveness of the sequential 4-channel NMES compared with that of the conventional 2-channel NMES for the treatment of dysphagia in a prospective double-blind randomized controlled study. J NeuroEng Rehabil. 2021;18(1):90.
65. Miller S, Diers D, Jungheim M, Schnittger C, Stürenburg HJ, Ptok M. Studying effects of neuromuscular electrostimulation therapy in patients with dysphagia: which pitfalls may occur? A translational phase I study. Ger Med Sci. 2021;19:Doc07.
66. Ortega O, Rofes L, Martin A, Arreola V, López I, Clavé P. A comparative study between two sensory stimulation strategies after two weeks treatment on older patients with oropharyngeal dysphagia. Dysphagia. 2016;31(5):706–16.
67. Bath PM, Woodhouse LJ, Suntrup-Krueger S, Likar R, Koestenberger M, Warusevitane A, et al. Pharyngeal electrical stimulation for neurogenic dysphagia following stroke, traumatic brain injury or other causes: main results from the PHADER cohort study. EClinicalMedicine. 2020;28:100608.
68. Michou E, Mistry S, Jefferson S, Tyrrell P, Hamdy S. Characterizing the mechanisms of central and peripheral forms of neurostimulation in chronic dysphagic stroke patients. Brain Stimul. 2014;7(1):66–73.
69. Vasant DH, Michou E, O'Leary N, Vail A, Mistry S, Hamdy S. Pharyngeal electrical stimulation in dysphagia poststroke. Neurorehabil Neural Repair. 2016;30(9):866–75.
70. Verin E, Maltete D, Ouahchi Y, Marie J-P, Hannequin D, Massardier EG, et al. Submental sensitive transcutaneous electrical stimulation (SSTES) at home in neurogenic oropharyngeal dysphagia: a pilot study. Ann Phys Rehabil Med. 2011;54(6):366–75.
71. Park J, Kim H, Park T, Yeo J, Hong H, Oh J. A pilot study of the effects of high-frequency repetitive transcranial magnetic stimulation on dysphagia in the elderly. Neurogastroenterol Motil. 2019;31(5):1–6.
72. Restivo DA, Alfonsi E, Casabona A, Hamdy S, Tassorelli C, Panebianco M, et al. A pilot study on the efficacy of transcranial direct current stimulation applied to the pharyngeal motor cortex for dysphagia associated with brainstem involvement in multiple sclerosis. Clin Neurophysiol. 2019;130(6):1017–24.

73. Lee JH, Kim SB, Lee KW, Lee SJ, Lee JU. Effect of repetitive transcranial magnetic stimulation according to the stimulation site in stroke patients with dysphagia. Ann Rehabil Med. 2015;39(3):432–9.
74. Kim L, Chun MH, Kim BR, Lee SJ. Effect of repetitive transcranial magnetic stimulation on patients with brain injury and dysphagia. Ann Rehabil Med. 2011;35(6):765–71.
75. Park E, Kim MS, Chang WH, Oh SM, Kim YK, Lee A, et al. Effects of bilateral repetitive transcranial magnetic stimulation on post-stroke dysphagia. Brain Stimul. 2017;10(1):75–82.
76. Ünlüer NÖ, Temuçin ÇM, Demir N, Serel Arslan S, Karaduman AA. Effects of low-frequency repetitive transcranial magnetic stimulation on swallowing function and quality of life of post-stroke patients. Dysphagia. 2019;34(3):360–71.
77. Verin E, Leroi AM. Poststroke dysphagia rehabilitation by repetitive transcranial magnetic stimulation: a noncontrolled pilot study. Dysphagia. 2009;24(2):204–10.
78. Michou E, Mistry S, Jefferson S, Singh S, Rothwell J, Hamdy S. Targeting unlesioned pharyngeal motor cortex improves swallowing in healthy individuals and after dysphagic stroke. Gastroenterology. 2012;142(1):29–38.
79. Park JW, Oh JC, Lee JW, Yeo JS, Ryu KH. The effect of 5Hz high-frequency rTMS over contralesional pharyngeal motor cortex in post-stroke oropharyngeal dysphagia: a randomized controlled study. Neurogastroenterol Motil. 2013;25(4):324–31.
80. Khedr EM, Mohamed KO, Soliman RK, Hassan AMM, Rothwell JC. The effect of high-frequency repetitive transcranial magnetic stimulation on advancing parkinson's disease with dysphagia: double blind randomized clinical trial. Neurorehabil Neural Repair. 2019;33(6):442–52.
81. Lin WS, Chou CL, Chang MH, Chung YM, Lin FG, Tsai PY. Vagus nerve magnetic modulation facilitates dysphagia recovery in patients with stroke involving the brainstem—a proof of concept study. Brain Stimul. 2018;11(2):264–70.
82. Zhong L, Rao J, Wang J, Li F, Peng Y, Liu H, et al. Repetitive transcranial magnetic stimulation at different sites for dysphagia after stroke: a randomized, observer-blind clinical trial. Front Neurol. 2021;12:625683.
83. Troche MS, Okun MS, Rosenbek JC, Musson N, Fernandez HH, Rodriguez R, et al. Aspiration and swallowing in Parkinson disease and rehabilitation with EMST: a randomized trial. Neurology. 2010;75(21):1912–9.
84. Park JS, Oh DH, Chang MY, Kim KM. Effects of expiratory muscle strength training on oropharyngeal dysphagia in subacute stroke patients: a randomised controlled trial. J Oral Rehabil. 2016;43(5):364–72.
85. Moon JH, Jung J-H, Won YS, Cho H-Y, Cho K. Effects of expiratory muscle strength training on swallowing function in acute stroke patients with dysphagia. J Phys Ther Sci. 2017;29:609–12.
86. Eom MJ, Chang MY, Oh DH, Kim HD, Han NM, Park JS. Effects of resistance expiratory muscle strength training in elderly patients with dysphagic stroke. NeuroRehabilitation. 2017;41(4):747–52.
87. Hutcheson KA, Barrow MP, Plowman EK, Lai SY, Fuller CD, Barringer DA, et al. Expiratory muscle strength training for radiation-associated aspiration after head and neck cancer: a case series. Laryngoscope. 2017;128(5):1044–51.
88. Pitts T, Bolser D, Rosenbek J, Troche MS, Okun MS, Sapienza C. Impact of expiratory muscle strength training on voluntary cough and swallow function in Parkinson disease. Chest. 2009;135(5):1301–8.
89. Plowman EK, Tabor-Gray L, Rosado KM, Vasilopoulos T, Robison R, Chapin JL, et al. Impact of expiratory strength training in amyotrophic lateral sclerosis: results of a randomized, sham-controlled trial. Muscle Nerve. 2019;59(1):40–6.
90. Plowman EK, Watts SA, Tabor L, Robison R, Gaziano J, Domer AS, et al. Impact of expiratory strength training in amyotrophic lateral sclerosis: expiratory training in ALS. Muscle Nerve. 2016;54(1):48–53.
91. Hegland KW. Rehabilitation of swallowing and cough functions following stroke: an expiratory muscle strength training trial. Arch Phys Med Rehabil. 2016;97(8):1345–51.
92. Mohannak N, Pattison G, Radich B, Hird K, Godecke E, Mastaglia F, et al. Exploring the efficacy of the expiratory muscle strength trainer to improve swallowing in inclusion body myositis: a pilot study. Neuromuscul Disord. 2020;30(4):294–300.
93. Arnold RJ, Bausek N. Effect of respiratory muscle training on dysphagia in stroke patients—a retrospective pilot study. Laryngosc Investig Otolaryngol. 2020;5(6):1050–5.
94. Jang K, Lee S, Kim S, Lee K, Lee J, Park J. Effects of mechanical inspiration and expiration exercise on velopharyngeal incompetence in subacute stroke patients. J Rehabil Med. 2019;51(2):97–102.
95. Martin-Harris B, McFarland D, Hill EG, Strange CB, Focht KL, Wan Z, et al. Respiratory-swallow training in patients with head and neck cancer. Arch Phys Med Rehabil. 2015;96(5):885–93.
96. Choi J-B, Shim S-H, Yang J-E, Kim H-D, Lee D-H, Park J-S. Effects of Shaker exercise in stroke survivors with oropharyngeal dysphagia. NeuroRehabilitation. 2017;41(4):753–7.
97. Gao J, Zhang H-J. Effects of chin tuck against resistance exercise versus Shaker exercise on dysphagia and psychological state after cerebral infarction. Eur J Phys Rehabil Med. 2017;53(3):426–32.
98. Kim H, Park J. Efficacy of modified chin tuck against resistance exercise using hand-free device for dysphagia in stroke survivors: a randomised controlled trial. J Oral Rehabil. 2019;46(11):1042–6.
99. Mano T, Katsuno M, Banno H, Suzuki K, Suga N, Hashizume A, et al. Head lift exercise improves swallowing dysfunction in spinal and bulbar muscular atrophy. Eur Neurol. 2015;74(5–6):251–8.
100. Park JS, Hwang NK, Oh DH, Chang MY. Effect of head lift exercise on kinematic motion of the hyolaryngeal complex and aspiration in patients with dysphagic stroke. J Oral Rehabil. 2017;44(5):385–91.
101. Park JS, An DH, Oh DH, Chang MY. Effect of chin tuck against resistance exercise on patients with dysphagia following stroke: a randomized pilot study. NeuroRehabilitation. 2018;42(2):191–7.
102. Park J, Lee G, Jung Y. Effects of game-based chin tuck against resistance exercise vs head-lift exercise in patients with dysphagia after stroke: an assessor-blind, randomized controlled trial. J Rehabil Med. 2019;51(10):749–54.
103. Park J-S, An D-H, Kam K-Y, Yoon T, Kim T, Chang M-Y. Effects of resistive jaw opening exercise in stroke patients with dysphagia: a double-blind, randomized controlled study. BMR. 2020;33(3):507–13.
104. Ploumis A, Papadopoulou SL, Theodorou SJ, Exarchakos G, Givissis P, Beris A. Cervical isometric exercises improve dysphagia and cervical spine malalignment following stroke with hemiparesis: a randomized controlled trial. Eur J Phys Rehabil Med. 2019;54(6):845–52.
105. Jakobsen D, Poulsen I, Schultheiss C, Riberholt CG, Curtis DJ, Petersen TH, et al. The effect of intensified nonverbal facilitation of swallowing on dysphagia after severe acquired brain injury: a randomised controlled pilot study. NRE. 2019;45(4):525–36.
106. Power ML, Fraser CH, Hobson A, Singh S, Tyrrell P, Nicholson DA, et al. Evaluating oral stimulation as a treatment for dysphagia after stroke. Dysphagia. 2006;21(1):49–55.

107. Rosenbek JC, Robbins J, Willford WO, Kirk G, Schiltz A, Sowell TW, et al. Comparing treatment intensities of tactile-thermal application. Dysphagia. 1998;13(1):1–9.
108. Tomsen N, Ortega O, Rofes L, Arreola V, Martin A, Mundet L, et al. Acute and subacute effects of oropharyngeal sensory stimulation with TRPV1 agonists in older patients with oropharyngeal dysphagia: a biomechanical and neurophysiological randomized pilot study. Ther Adv Gastroenterol. 2019. https://doi.org/10.1177/1756284819842043.
109. Steele CM, Bayley MT, Peladeau-Pigeon M, Nagy A, Namasivayam AM, Stokely SL, et al. A randomized trial comparing two tongue-pressure resistance training protocols for post-stroke dysphagia. Dysphagia. 2016;31:452–61.
110. Robbins J, Kays SA, Gangnon RE, Hind JA, Hewitt AL, Gentry LR, et al. The effects of lingual exercise in stroke patients with dysphagia. Arch Phys Med Rehabil. 2007;88(2):150–8.
111. Robbins J, Gangnon RE, Theis SM, Kays SA, Hewitt AL, Hind JA. The effects of lingual exercise on swallowing in older adults. J Am Geriatr Soc. 2005;53(9):1483–9.
112. Namiki C, Hara K, Tohara H, Kobayashi K, Chantaramanee A, Nakagawa K, et al. Tongue-pressure resistance training improves tongue and suprahyoid muscle functions simultaneously. CIA. 2019;14:601–8.
113. Kim HD, Choi JB, Yoo SJ, Chang MY, Lee SW, Park JS. Tongue-to-palate resistance training improves tongue strength and oropharyngeal swallowing function in subacute stroke survivors with dysphagia. J Oral Rehabil. 2017;44(1):59–64.
114. Balou M, Herzberg EG, Kamelhar D, Molfenter SM. An intensive swallowing exercise protocol for improving swallowing physiology in older adults with radiographically confirmed dysphagia. CIA. 2019;14:283–8.
115. Hsiang C-C, Chen AW-G, Chen C-H, Chen M-K. Early postoperative oral exercise improves swallowing function among patients with oral cavity cancer: a randomized controlled trial. Ear Nose Throat J. 2019;98(6):E73–80.
116. Kraaijenga SAC, van der Molen L, Stuiver MM, Takes RP, Al-Mamgani A, van den Brekel MWM, et al. Efficacy of a novel swallowing exercise program for chronic dysphagia in long-term head and neck cancer survivors. Head Neck. 2017;39(10):1943–61.
117. Tarameshlu M. The effect of traditional dysphagia therapy on the swallowing function in patients with multiple sclerosis: a pilot double-blinded randomized controlled trial. J Bodyw Mov Ther. 2019;23(1):171–6.
118. van der Molen L, van Rossum MA, Burkhead LM, Smeele LE, Rasch CR, Hilgers FJ. A randomized preventive rehabilitation trial in advanced head and neck cancer patients treated with chemoradiotherapy: feasibility, compliance, and short-term effects. Dysphagia. 2011;26(2):155–70.
119. van der Molen L, van Rossum MA, Rasch CRN, Smeele LE, Hilgers FJM. Two-year results of a prospective preventive swallowing rehabilitation trial in patients treated with chemoradiation for advanced head and neck cancer. Eur Arch Otorhinolaryngol. 2014;271(5):1257–70.
120. Furuie H, Hamamoto T, Chikuie N, Kono T, Taruya T, Ishino T, et al. Evaluation of role of prophylactic swallowing rehabilitation in chemoradiotherapy for advanced head and neck cancer using novel software analysis of videofluorography images. Hiroshima J Med Sci. 2019;68(2–3):27–34.
121. Boukrina O, Kucukboyaci NE, Dobryakova E. Considerations of power and sample size in rehabilitation research. Int J Psychophysiol. 2020;154:6–14.
122. Knottnerus JA, Bouter LM. The ethics of sample size: two-sided testing and one-sided thinking. J Clin Epidemiol. 2001;54:109–10.
123. Lakens D. Performing high-powered studies efficiently with sequential analyses: sequential analyses. Eur J Soc Psychol. 2014;44(7):701–10.
124. Lakens D. Will knowledge about more efficient study designs increase the willingness to pre-register? https://osf.io/svzyc. Accessed Mar 2017.
125. Brysbaert M, Stevens M. Power analysis and effect size in mixed effects models: a tutorial. J Cogn. 2018;1(1):9.
126. Polack FP, Thomas SJ, Kitchin N, Absalon J, Gurtman A, Lockhart S, et al. Safety and efficacy of the BNT162b2 mRNA covid-19 vaccine. N Engl J Med. 2020. https://doi.org/10.1056/nejmoa2034577.
127. Zhang JJ, Blumenthal GM, He K, Tang S, Cortazar P, Sridhara R. Overestimation of the effect size in group sequential trials. Clin Cancer Res. 2012;18(18):4872–6.
128. Project J. The jamovi project. 2020. (jamovi). https://www.jamovi.org.
129. Johnson PCD, Barry SJE, Ferguson HM, Mu P. Power analysis for generalized linear mixed models in ecology and evolution. Methods Ecol Evol. 2014;6:133–42.
130. Quintana DS. A synthetic dataset primer for the biobehavioural sciences to promote reproducibility and hypothesis generation. Elife. 2020;9:1–12.
131. Quintana DS. Most oxytocin administration studies are statistically underpowered to reliably detect (or reject) a wide range of effect sizes. Compr Psychoneuroendocrinol. 2020;4:1–4.

James C. Borders MS, CCC-SLP

Alessandro A. Grande BS, MPhil

Michelle S. Troche PhD, CCC-SLP