Am J Otolaryngol 40 (2019) 61–66

Contents lists available at ScienceDirect

Am J Otolaryngol

journal homepage: www.elsevier.com/locate/amjoto

The fragility of statistically significant results in otolaryngology randomized trials

Mason Skinner[a], Daniel Tritz[b], Clayton Farahani[a], Andrew Ross[b,*], Tom Hamilton[a], Matt Vassar[b]

[a] Oklahoma State University Medical Center, 744 W 9th St, Tulsa, OK 74127, United States of America
[b] Oklahoma State University Center for Health Sciences, 1111 W 17th St, Tulsa, OK 74107, United States of America

A B S T R A C T

Objectives: The American Academy of Otolaryngology–Head and Neck Surgery regards randomized controlled trials as class A evidence. A novel method to determine the robustness of outcomes in trials is the fragility index. This index represents the number of patients whose status would have to change from a non-event to an event to make a statistically significant result non-significant.

Methods: Investigators included otolaryngology journals listed in the top 10 of one or both of Google Scholar Metrics and Clarivate Analytics' Journal rankings. For inclusion, a randomized controlled trial needed to report a one-to-one random assignment of participants to condition, contain two parallel arms or have used a two-by-two factorial design, and report at least one statistically significant dichotomous outcome.

Results: Sixty-nine trials met inclusion criteria. The median fragility index was three events (interquartile range 1–7.5). Median sample size was 72 (interquartile range 50–102.5). Modest correlations were observed between fragility index and total sample size ($r = 0.27$) and fragility index and event rate ($r = 0.46$). Investigators found no correlation between fragility index and impact factor or Science Citation Index. In 39% (27/69) of trials, the number lost to follow-up was equal to or greater than the fragility index.

Conclusion: A median fragility index of 3 indicates that three people, on average, are needed to alter the outcomes in otolaryngology trials. This indicates that the results of two-group randomized controlled trials reporting binary endpoints published in otolaryngology journals may frequently be fragile.

1. Introduction

Randomized trials are regarded as the gold standard in clinical research and have played an important role in developing recommendations for clinical practice guidelines. Of the American Academy of Otolaryngology–Head and Neck Surgery's 16 current practice guidelines, 9 state that randomized trials are considered class A evidence [1–9], whereas 7 regard them as class B evidence [10–16]. Recommendations such as the prescription of oral steroids to patients with Bell's palsy [4], diagnosis of idiopathic sudden sensorineural hearing loss based on audiometric results [8], prescription of antimicrobials to patients with acute otitis externa [3], placement of tympanostomy tubes for otitis media with effusion [5], and indications for tonsillectomy [9] are all underpinned by randomized trials, attesting to their importance. Since these trials form the basis for such recommendations, it is imperative that their results be robust. Here, the authors propose the fragility index (FI) as a mechanism to determine the robustness of results from randomized trials that form the foundation of guideline recommendations.

The FI is a tool used to determine the robustness of statistically significant dichotomous outcomes [17]. This index represents the number of patients whose status would have to change from a non-event to an event to make a statistically significant result non-significant. Conversely, the FI can be applied to nonsignificant trials by calculating the number of patients who do not develop the outcome of interest required for the trial outcome to become statistically significant. A small FI indicates that a statistically significant outcome relies on relatively few individuals. Whereas, a large FI relies on a greater number of individuals, making it a more robust outcome. For example, in one randomized trial, patients undergoing total laryngectomy received either proton pump inhibitors or placebo with the intent to decrease the incidence of pharyngocutaneous fistula. One of the 21 patients receiving proton pump inhibitors developed a pharyngocutaneous fistula, while 6 of the 19 patients receiving placebo developed a pharyngocutaneous fistula. The authors reported a statistically significant outcome with a *P* value of 0.04 [18]. Therefore, the study's results would have been deemed non-significant had only one additional patient in the proton pump inhibitor group developed a pharyngocutaneous fistula. Thus, for this outcome, the fragility index is one event.

Given the widespread acceptance of randomized trials for clinical decision making, a method should be adopted to determine the robustness of trials. This study's primary goal is to estimate the robustness of statistically significant otolaryngology trials using the FI. The

* Corresponding author.
E-mail address: aeross@okstate.edu (A. Ross).

https://doi.org/10.1016/j.amjoto.2018.10.011
Received 18 October 2018

secondary outcomes are to investigate the relationship between the FI and the Science Citation Index, impact factor, total sample size, and event rate.

2. Methods

This study was not subject to institutional review board oversight since it did not meet the regulatory definition of human subject research as defined in 45 CFR 46.102(d) and (f) of the Department of Health and Human Services' Code of Federal Regulations [19]. The investigators applied relevant Statistical Analyses and Methods in the Published Literature (SAMPL) guidelines for reporting descriptive statistics [20].

2.1. Journal selection

The investigators retrieved journal rankings from both the Clarivate Analytics' Science Citation Index (previously a resource of Thomson Reuters) and Google Scholar Metrics: Otolaryngology subcategory. To qualify for inclusion, a journal must have been listed in the top 10 of at least one of these ranking systems. Investigators gave preference to general otolaryngology journals above sub-specialty journals during selection. Table 1 shows a breakdown of the journals used for this study and the number of articles from each.

2.2. Eligibility criteria

The investigators included randomized trials published between 2011 and 2016. To qualify for inclusion in this study, the trial had to report a 1:1 random assignment of participant to condition, contain either two arms or have used a two-by-two factorial design, and report at least one statistically significant dichotomous outcome ($P < 0.05$). The investigators included both primary and secondary outcomes. Crossover designs, trials with more than two arms, and cluster trials were excluded.

2.3. Search strategy

Two investigators (MV and MS) conducted a search of PubMed, which includes Medline, to identify randomized trials. To conduct their search query, they applied the Cochrane Collaboration's Highly Sensitive Search Strategy for Identifying Randomized Trials in Medline: Sensitivity- And Precision-Maximizing Version (2008 revision); PubMed format [21]. The final search query is in the online supplement. The search was conducted on March 2, 2017.

Table 1
Characteristics of randomized controlled trials meeting full text criteria ($n = 69$).

Characteristic	Studies	
	(*n*)	(%)
Year of publication		
2011	11	16
2012	10	14
2013	15	22
2014	7	10
2015	16	23
2016	10	14
Journal		
Laryngoscope	12	17
Clinical Otolaryngology	1	1
Otolaryngology: Head and Neck Surgery	14	20
Ear and Hearing	2	3
European Archives of Oto-Rhino-Laryngology	16	23
International Journal of Pediatric Otorhinolaryngology	11	16
JAMA Otolaryngology: Head and Neck Surgery	5	7
Head and Neck	7	10
Hearing Research	1	1
Journal of the Association for research in Otolaryngology	0	0

2.4. Screening by title and abstract

Two investigators (MV and MS) reviewed the inclusion criteria to help improve the consistency of the screening process. The initial 10 abstracts were reviewed jointly as a training exercise. Each abstract was discussed, and a decision was made regarding its suitability for inclusion in the study. The investigators were in agreement regarding the screening of these studies, so MS screened the remaining abstracts based on several criteria. Trials were included based on the previously mentioned inclusion criteria.

2.5. Full-text screening and data extraction

Two investigators (MS and DT) performed simultaneous full-text screening and data extraction. Before beginning this process, they met to extract data from five trials as a means of pilot testing a Google form developed specifically for this study. The remainder of records was divided evenly between them.

Records retained from the title and abstract screening were reviewed by these investigators using the full-text published report. Trials without dichotomous outcomes, or cases in which the dichotomous outcome was not statistically significant, were excluded. If primary and secondary dichotomous outcomes were found, the investigators extracted the primary outcome. In the case of multiple dichotomous and statistically significant primary outcomes, secondary outcomes, or unspecified outcomes, we followed the methodology of previous studies [22,23] by using the GRADE Network's approach [24] to identify the most patient-important outcome. To apply this method, a board-certified otolaryngologist (TH) independently reviewed the outcomes in question, rating them from 1 (low importance for decision making) to 10 (critical for decision making). The outcome with the highest rating was used. If the study included multiple significant outcomes of the same variable, the variable that occurred first temporally was used. In studies that compared two widely accepted modalities, the modality that was developed first was considered the control, and the newer modality was considered the intervention.

For qualifying trials, the investigators extracted the following information using a Google form, developed and pilot tested specifically for this study: journal name, year of publication, funding source, sample size for each group, number lost to follow up for each group, the dichotomous outcome, the type of outcome (primary, secondary, tertiary, unspecified), number of events per group, *P* value, statistical test used for group comparison, journal impact factor, and Science Citation Index for the trial.

2.6. Risk of bias evaluation

Two investigators used the Cochrane Risk of Bias tool 2.0 (RoB 2.0) to evaluate the likelihood and sources of bias in the included trials. Details describing the new risk of bias tool can be found at the Cochrane Training website [25].

To perform the RoB evaluations, MV, DT, and CF read the test manual for RoB 2.0 and viewed the Cochrane Collaboration training videos. They held a series of training sessions to evaluate a subset of trials and discuss each bias judgement in depth. This training was structured to be consistent with da Costa et al.'s intensive risk of bias training [26,27], which improved interrater reliability over other training methods. Following these training sessions, DT and CF independently evaluated three additional trials. Results were compared and discrepancies discussed until resolution was achieved. All trials were independently evaluated by DT and CF, after which the two investigators met to resolve discrepancies. MV was available for third-

party adjudication but was not needed.

2.7. Data analysis

Descriptive statistics, including medians; interquartile ranges; and correlations with variables such as total sample size, event rate, five-year impact factor, and Science Citation Index, were calculated using Microsoft Excel. The FI for each study was calculated using the fragility calculator located at www.fragilityindex.com. The FI formula converts one patient from a non-event to an event in either the control or experimental group. It selects the group with the smaller number of events. Then it recalculates a Fisher's exact test until the *P* value is greater than or equal to 0.05. The fragility quotient (FQ) was calculated by dividing the FI by total sample size. A one-way ANOVA was conducted to compare the effect of overall risk of bias on FI in low, medium, and high risk of bias groups.

3. Results

3.1. Study selection

The PubMed search returned 1090 records. After abstract screenings and full-text reviews, investigators retained 69 studies (Fig. 1). Table 1 details the included trials by journal and year.

3.2. Characteristics of trials and risk of bias

The median sample size of randomized controlled trials (RCTs) was 72 (interquartile range, IQR, 50–102.5), and the median lost to follow-up was 0 (IQR 0–2). The median Web of Science Citation Index for included trials was 4 (IQR 1–9), and the median journal impact factor was 2.089 (IQR 1.63–2.45). Of the 69 outcomes, 55 (80%) were primary and 14 (20%) were secondary. Table 2 details risk of bias across the trials. Overall, 30 of 69 (43.5%) trial outcomes received low risk of bias judgements. The randomization process had the greatest likelihood for bias, owing to a lack of concealing the allocation sequence until the patients were recruited and assigned to interventions. (See full results in Table 2.)

3.3. Fragility index and fragility quotient

The median FI for the included 69 trials was 3 events (IQR 1–7.5; Fig. 2). The median FI for primary outcomes was 3 (IQR 1–8), and the median FI for secondary outcomes was 1.5 (IQR 0.75–3.75). Applying the Fisher's exact test nullified statistical significance in 16 trials, which resulted in a FI of 0. FI is calculated using Fisher's exact test. Chi-square and other methods are commonly used to compare dichotomous outcomes. The *P* value can be discrepant when calculating outcomes with Fisher's exact test versus other tests. In cases where Fisher's exact test would result in a non-significant *P* value the FI would be 0. The Science

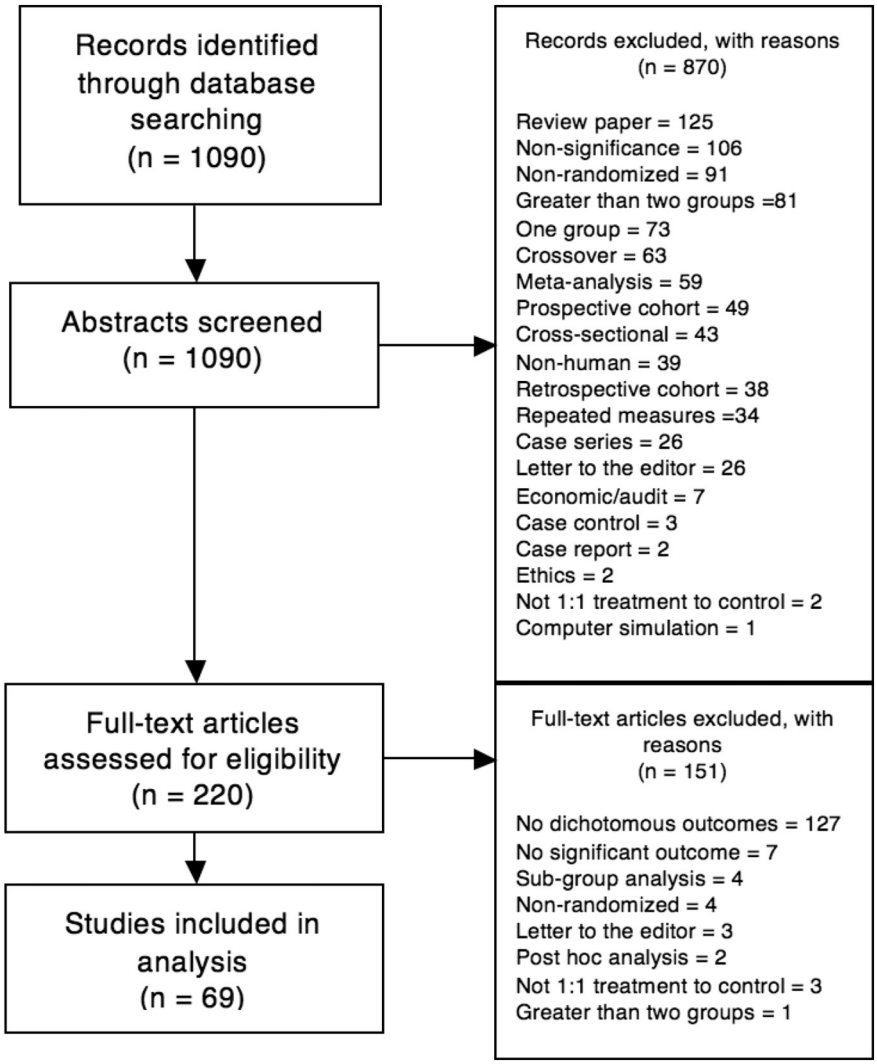

Fig. 1. PRISMA diagram showing exclusions with reasoning during abstract and full text review.

Table 2
Risk of bias evaluation.

Risk of bias category	Low risk *n* (%)	Some concern *n* (%)	High risk *n* (%)
Arising from the randomization process (n = 69)	38 (55.1%)	23 (33.3%)	8 (11.6%)
Due to deviations from intended interventions (n = 69)	60 (87.0%)	1 (1.4%)	8 (11.6%)
Due to missing outcome data (n = 69)	63 (91.3%)	2 (2.9%)	4 (5.8%)
Measurement of the outcome (n = 69)	65 (94.2%)	0 (0%)	4 (5.8%)
Selection of the reported result (n = 69)	52 (75.4%)	2 (2.9%)	15 (21.7%)
Overall bias of trial (n = 69)	30 (43.5%)	14 (20.3%)	25 (36.2%)

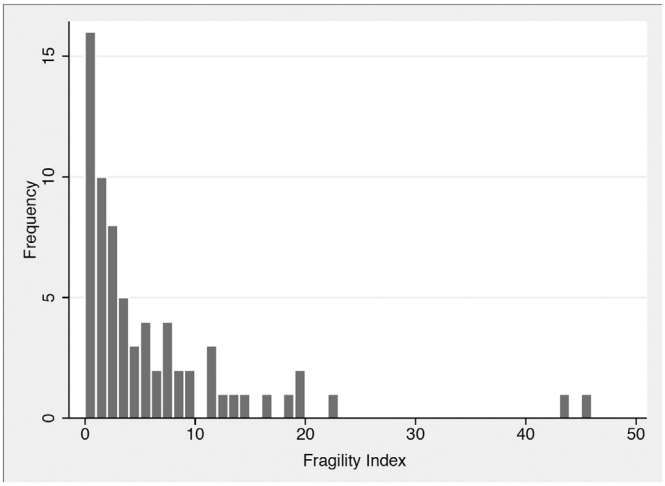

Fig. 2. Distribution of Fragility Index (FI) across 69 trials (median FI = 3; interquartile range 1–7.5).

Citation Index of the 5 most cited studies was 31 (FI = 4), 27 (FI = 7), 26 (FI = 6), 23 (FI = 0), and 20 (FI = 14), with a FI median of 6. Only three trials from the two highest five-year impact factor journals were included in the sample, whereas 27 of the trials in the sample were published in the lowest two impact factor journals. The median FQ for included trials was 0.04 (IQR 0.004–0.11). The number lost to follow-up was greater than or equal to the FI in 27 of 69 (39%) trials.

3.4. Correlation between fragility index and other variables

There was a modest correlation between FI and total sample size (r = 0.27; Fig. 3) and FI and event rate (r = 0.46). Investigators found no correlation between FI and five-year impact factor (r = 0.02) or FI and Science Citation Index (r = 0.08). No significant effect of overall risk of bias on the FI existed [F(66,2) = 1.14, P = 0.32].

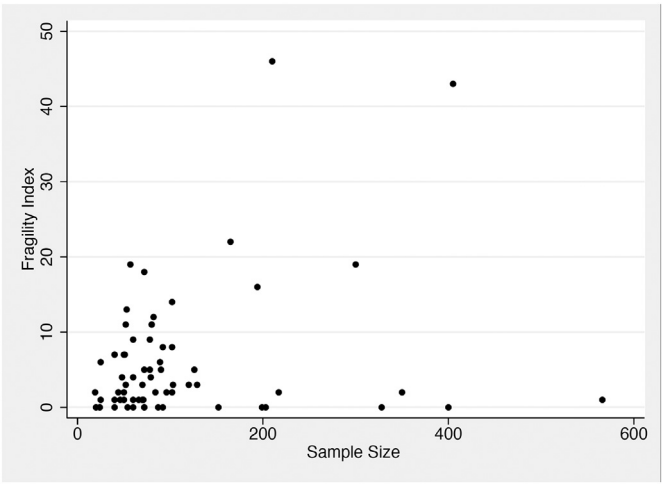

Fig. 3. Fragility index versus total sample size.

4. Discussion

Investigators found that on average only three non-events, modified to events, would render a finding nonsignificant. One example of a fragile trial investigated whether pillar suturing improved adenotonsillectomy outcomes for pediatric obstructive sleep apnea. The study reported that the primary outcome was improvement in the apnea hypopnea index of > 50%; however, this outcome's FI was 0 and was based on 11 intervention events and 6 control events [28]. In this study, investigators noted 16 occasions where trial outcomes were no longer statistically significant when applying the Fisher's exact test instead of chi-square. It has been recommended that chi-square not be used for smaller sample sizes. Fisher's exact test, while more conservative, may be used for both large and small sample sizes [29]. This applies to this study because of the varying sample sizes, the majority of which are small and deal with small numbers of events.

One example of a robust trial looked at the efficacy of topical suspension of bacterial antigens for management of chronic suppurative otitis media. The primary outcome was absence of otorrhea with a FI of 46 based on 67 events in the intervention group and 6 events in the control group [30]. Sample size may be a contributing factor to the robustness of trial results, as demonstrated by the moderate correlation found in this study. While FI is a useful and easily interpreted metric, it has been criticized for not accounting for sample size [31]. The FQ, in which the FI is divided by the trial's sample size, has been proposed as a supplementary measure [31]. The investigators found no current use in the published literature, and no guidance regarding acceptable ranges for this measure exists. To demonstrate the need for reporting both metrics, consider one trial with a total sample size of 566 patients and a FI of 1 [32] compared to a second trial with a total sample size of 25 and a FI of 1 [33]. Initially, the FI indicates that both trials are fragile; however, the FQ for these trials is 0.001 for the first trial and 0.04 for the second. Hence, the second trial seems more robust even though the fragility indices are equal.

This study is the second to report on the fragility of RCTs in otolaryngology. Noel et al. [34] evaluated RCTs in head and neck literature where surgery was the primary intervention with a median sample size of 67.5 and a median FI of 1 event (IQR 0–2.5). Other clinical disciplines have provided evidence of trial fragility. For example, Evaniew et al. [23] evaluated RCTs in spine surgery with a median sample size of 132 patients and a median FI of 2 events (IQR 1–3). Walsh et al. [17] evaluated trials from high-impact general medical journals and found a median sample size of 682 patients and a median FI of 8 events (IQR 3–18). In comparison, this study found a median sample size of 72 patients and a median FI of 3 events (IQR 1–7). The similarity in fragility between this study, Noel et al. [34], and Evaniew et al. [23] is not surprising, since they were conducted in surgical specialties. This study was also the first, to the investigators' knowledge, to use the Cochrane RoB 2.0. Investigators found that a large portion of trial outcomes had a high risk of bias. The area most in need of improvement is randomization, which is important in balancing known and unknown prognostic

factors.

The American Academy of Otolaryngology–Head and Neck Surgery regards RCTs as class A evidence, and the specialty has noted an increase in the number of RCTs performed each year. Between 2000 and 2005, 202 RCTs were reported [35], whereas between 2011 and 2013, 189 RCTs were found. [36] Increases in research have generated large amounts of data to be considered when making clinical decisions. Since these decisions often stem from a statistically significant finding regarding therapeutic effectiveness, clinicians and surgeons turn to *P* values. Despite the pervasiveness of the *P* value, it has been criticized for its inability to provide an interpretation regarding the magnitude of treatment effects [37]. Readers often place similar degrees of confidence in *P* values regardless of other trial determinants. In this study, investigators found that many RCTs in otolaryngology were characterized by small sample sizes and outcome events, the results of which should be interpreted with caution, given that as few as three patients were needed to nullify statistical significance in evaluated trials. Since interpretation of clinical research is fundamental to the evidence-based medicine model of patient care, otolaryngologists should be equipped with the necessary tools to interpret outcomes and make well-informed clinical decisions. One tool is the FI. The FI is easily understood and allows for quick determination of statistical robustness [17]. It should be used in conjunction with the *P* value to aid in the interpretation of statistical results. Investigators recommend this combination be augmented with the FQ. In addition to FI, number needed to treat is a way clinicians and patients can understand the clinical relevance of dichotomous outcomes [38]. The investigators included number needed to treat alongside FI/FQ in the supplemental table.

This search strategy considered the top 10 otolaryngology journals with the highest rankings from two independent sources, Clarivate Analytics' Science Citation Index and Google Scholar Metrics. Since investigators only included RCTs published in these journals, the results may have been biased toward studies with larger sample sizes [39]; therefore, since investigators possibly excluded RCTs with smaller sample sizes from lower ranked journals, this sample may comprise trials with higher fragility indices. Furthermore, the investigators note the limitations of the FI, which is limited by its strict application of one-to-one randomization and binary outcomes [17]. This limitation prohibits the inclusion of other clinical trials. FI is limited to binary outcomes since it depends on *P* values, which rely on a single null hypothesis that is either accepted or rejected based on the data. This does not account for other hypotheses that may play a role. Bayesian models can help address this problem by comparing the null hypothesis against different alternative hypotheses. This could better determine the robustness of clinical trial outcomes [40]. Last, approximately 20% of the outcomes included for analysis were not primary outcomes; thus, the conclusions regarding the degree of fragility in the literature and its implication on the interpretation of results may be slightly overstated.

5. Conclusions

Randomized trials in otolaryngology journals publish statistically significant outcomes that are, oftentimes, fragile. This fragility is, in part, due to small sample sizes and low numbers of outcome events. The FI and the FQ are new and intuitive metrics that complement *P* values and provide insight into the interpretation of statistically significant outcomes of RCTs. Outcomes that have a low FI should be interpreted with caution. We recommend qualifying trials report FI and lost to follow up. In the event that FI is smaller than lost to follow up, the outcome should be interpreted carefully because the lost events could render the outcome non-significant.

Funding and conflicts of interest

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.amjoto.2018.10.011.

References

[1] Seidman MD, Gurgel RK, Lin SY, et al. Clinical practice guideline: Allergic rhinitis. Otolaryngol. Head Neck Surg. 2015;152(1 Suppl):S1–43.
[2] Tunkel DE, Bauer CA, Sun GH, et al. Clinical practice guideline: tinnitus. Otolaryngol. Head Neck Surg. 2014;151(2 Suppl):S1–40.
[3] Rosenfeld RM, Schwartz SR, Cannon CR, et al. Clinical practice guideline: acute otitis externa. Otolaryngol. Head Neck Surg. 2014;150(1 Suppl):S1–24.
[4] Baugh RF, Basura GJ, Ishii LE, et al. Clinical practice guideline: Bell's palsy. Otolaryngol. Head Neck Surg. 2013;149(3 Suppl):S1–27.
[5] Rosenfeld RM, Schwartz SR, Pynnonen MA, et al. Clinical practice guideline: Tympanostomy tubes in children. Otolaryngol. Head Neck Surg. 2013;149(1 Suppl):S1–35.
[6] Chandrasekhar SS, Randolph GW, Seidman MD, et al. Clinical practice guideline: improving voice outcomes after thyroid surgery. Otolaryngol. Head Neck Surg. 2013;148(6 Suppl):S1–37.
[7] Roland PS, Rosenfeld RM, Brooks LJ, et al. Clinical practice guideline: polysomnography for sleep-disordered breathing prior to tonsillectomy in children. Otolaryngol. Head Neck Surg. 2011;145(1 Suppl):S1–15.
[8] Stachler RJ, Chandrasekhar SS, Archer SM, et al. Clinical practice guideline. Otolaryngol. Head Neck Surg. 2012;146(3_suppl):S1–35.
[9] Baugh RF, Archer SM, Mitchell RB, et al. Clinical practice guideline: tonsillectomy in children. Otolaryngol. Head Neck Surg. 2011;144(1 Suppl):S1–30.
[10] Schwartz SR, Cohen SM, Dailey SH, et al. Clinical practice guideline: hoarseness (dysphonia). Otolaryngol. Head Neck Surg. 2009;141(3 Suppl 2):S1–31.
[11] Pynnonen MA, Gillespie MB, Roman B, et al. Clinical practice guideline: evaluation of the neck mass in adults. Otolaryngol. Head Neck Surg. 2017;157(2_suppl):S1–30.
[12] Bhattacharyya N, Gubbels SP, Schwartz SR, et al. Clinical practice guideline: benign paroxysmal positional vertigo (update). Otolaryngol. Head Neck Surg. 2017;156(3_suppl):S1–47.
[13] Ishii LE, Tollefson TT, Basura GJ, et al. Clinical practice guideline: improving nasal form and function after rhinoplasty. Otolaryngol. Head Neck Surg. 2017;156(2_suppl):S1–30.
[14] Schwartz SR, Magit AE, Rosenfeld RM, et al. Clinical practice guideline (update): earwax (cerumen impaction). Otolaryngol. Head Neck Surg. 2017;156(1_suppl):S1–29.
[15] Rosenfeld RM, Shin JJ, Schwartz SR, et al. Clinical practice guideline: otitis media with effusion (update). Otolaryngol. Head Neck Surg. 2016;154(1 Suppl):S1–41.
[16] Rosenfeld RM, Piccirillo JF, Chandrasekhar SS, et al. Clinical practice guideline (update): adult sinusitis. Otolaryngol. Head Neck Surg. 2015;152(2 Suppl):S1–39.
[17] Walsh M, Srinathan SK, McAuley DF, et al. The statistical significance of randomized controlled trial results is frequently fragile: a case for a fragility index. J. Clin. Epidemiol. 2014;67(6):622–8.
[18] Stephenson KA, Fagan JJ. Effect of perioperative proton pump inhibitors on the incidence of pharyngocutaneous fistula after total laryngectomy: a prospective randomized controlled trial. Head Neck 2015;37(2):255–9.
[19] of Health USD, Services H, et al. Protection of human subjects. Title 45 Code of Federal Regulations, part 46. 2009.
[20] Lang TA, Altman DG. Basic statistical reporting for articles published in biomedical journals: the “statistical analyses and methods in the published literature” or the SAMPL guidelines. Int. J. Nurs. Stud. 2015;52(1):5–9.
[21] Higgins JPT, Green S. Cochrane handbook for systematic reviews of interventions. Wiley; 2008.
[22] Khan M, Evaniew N, Gichuru M, et al. The fragility of statistically significant findings from randomized trials in sports surgery a systematic survey. Am. J. Sports Med. 2016;45(9):2164–70. (0363546516674469).
[23] Evaniew N, Files C, Smith C, et al. The fragility of statistically significant findings from randomized trials in spine surgery: a systematic survey. Spine J. 2015;15(10):2188–97.
[24] Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. J. Clin. Epidemiol. 2011;64(4):395–400.
[25] Higgins JPT, Sterne JAC, Savović J, et al. A revised tool for assessing risk of bias in randomized trials. Cochrane Database Syst. Rev. 2016;10(Suppl. 1):29–31.
[26] da Costa BR, Beckett B, Diaz A, et al. Effect of standardized training on the reliability of the Cochrane risk of bias assessment tool: a prospective study. Syst. Rev. 2017;6(1):44.
[27] da Costa BR, Resta NM, Beckett B, et al. Effect of standardized training on the reliability of the Cochrane risk of bias assessment tool: a study protocol. Syst. Rev. 2014;3:144.
[28] Chiu P-H, Ramar K, Chen K-C, et al. Can pillar suturing promote efficacy of adenotonsillectomy for pediatric OSAS? A prospective randomized controlled trial. Laryngoscope 2013;123(10):2573–7.
[29] Lydersen S, Pradhan V, Senchaudhuri P, Laake P. Choice of test for association in small sample unordered r x c tables. Stat. Med. 2007;26(23):4328–43.
[30] Mora R, Salzano FA, Mora E, Guastini L. Efficacy of a topical suspension of bacterial antigens for the management of chronic suppurative otitis media. Eur. Arch. Otorhinolaryngol. 2012;269(6):1593–7.
[31] Ahmed W, Fowler RA, McCredie VA. Does sample size matter when interpreting the fragility index? Crit. Care Med. 2016;44(11):e1142–3.

[32] Kopke R, Slade MD, Jackson R, et al. Efficacy and safety of N-acetylcysteine in prevention of noise induced hearing loss: a randomized clinical trial. Hear. Res. 2015;323:40–50.
[33] Fawaz SA, Sabri SM, Sweed AS, Hegazi MA, Riad MA. Use of local mitomycin C in enhancing laryngeal healing after laser cordotomy: a prospective controlled study. Head Neck 2014;36(9):1248–52.
[34] Noel CW, McMullen C, Yao C, et al. The fragility of statistically significant findings from randomized trials in head and neck surgery. Laryngoscope April 2018. https://doi.org/10.1002/lary.27183.
[35] Yao F, Singer M, Rosenfeld RM. Randomized controlled trials in otolaryngology journals. Otolaryngol. Head Neck Surg. 2007;137(4):539–44.
[36] Banglawala SM, Lawrence LA, Franko-Tobin E, Soler ZM, Schlosser RJ, Ioannidis J. Recent randomized controlled trials in otolaryngology. Otolaryngol. Head Neck Surg. 2015;152(3):418–23.
[37] Bhandari M, Montori VM, Schemitsch EH. The undue influence of significant p-values on the perceived importance of study results. Acta Orthop. 2005;76(3):291–5.
[38] Citrome L. Quantifying clinical relevance. Innov. Clin. Neurosci. 2014;11(5–6):26–30.
[39] Bala MM, Akl EA, Sun X, et al. Randomized trials published in higher vs. lower impact journals differ in design, conduct, and analysis. J. Clin. Epidemiol. 2013;66(3):286–95.
[40] Kruschke JK, Liddell TM. The Bayesian new statistics: hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. Psychon. Bull. Rev. 2018;25(1):178–206.