OTOLARYNGOLOGY–HEAD AND NECK SURGERY FOUNDATION

Original Research—Otology and Neurotology

Otolaryngology–Head and Neck Surgery
145(2) 282–287

DOI: 10.1177/0194599811403882
http://otojournal.org

SAGE

Tinnitus Handicap Inventory for Evaluating Treatment Effects: Which Changes Are Clinically Relevant?

Florian Zeman, MathD[1], Michael Koller, PhD[1], Ricardo Figueiredo, MD[2,3], Andreia Aazevedo, MD[2,3], Marcello Rates, MD[3,4], Claudia Coelho, MD, PhD[3,5], Tobias Kleinjung, MD[3,6], Dirk de Ridder, MD[3,7], Berthold Langguth, MD[3,8], and Michael Landgrebe, MD[3,8]

Sponsorships or competing interests that may be relevant to content are disclosed at the end of this article.

Abstract

Objective. To determine the minimum change of the Tinnitus Handicap Inventory (THI) score that could be considered clinically relevant, the authors compared the absolute change of the THI with the Clinical Global Impression–Improvement (CGI-I) score.

Study Design. International studies register with standardized data collection.

Setting. Tinnitus Research Initiative (TRI).

Subjects and Methods. Two hundred ten patients of the TRI database were eligible for this study. In the first analysis, the THI score change and CGI-I ratings were compared with equipercentile linking. In a second analysis, the authors categorized the CGI-I into the 4 groups *much better or better*, *minimally better*, *no change*, and *worse* and calculated the corresponding differences of the THI score and the effect sizes. An effect size separating the minimally better and the no-change groups was chosen, and the referring THI mean score difference was calculated.

Results. According to the linking method, a CGI-I value of 3 (minimally better) corresponded to a THI score reduction of 6 to 16, whereas the CGI-I value of 4 (no change) corresponded to the range between improvement by 5 points and worsening by 4 points. For separating the no-change and minimally better groups, an effect size $d = 0.5$ was determined, resulting in a minimal clinically relevant difference of $\Delta THI = 7$.

Conclusion. Two different methods yielded comparable results in identifying a reduction in the THI score of 6 and 7 points, respectively, as the minimal clinically relevant change. This study provides a first orientation for sample size calculations and for planning the design of future studies.

Keywords

tinnitus, Tinnitus Handicap Inventory

Received December 23, 2010; revised February 15, 2011; accepted February 23, 2011.

Prospective placebo-controlled randomized trials are obligatory for proving the efficacy of new treatments. In the past few years, a large variety of new treatment options for tinnitus have been proposed, which has increased the need for clinical studies in this field. The most critical part in the design of clinical studies is the choice of outcome measurements. In tinnitus research, the quantification of severity is challenging for many reasons. First, tinnitus is a purely subjective sensation lacking objectively measurable variables. Second, tinnitus has many aspects and dimensions that vary from patient to patient. Some patients are most bothered by the loudness of their tinnitus, whereas others mainly suffer from tinnitus-related insomnia or concentration difficulties. Several approaches have been developed for the quantification of tinnitus. Besides psychometric measures of tinnitus

[1]Center for Clinical Studies, University Hospital Regensburg, Germany
[2]Otosul, Otorrinolaringologia Sul-Fluminense, Volta Redonda, Brazil
[3]Tinnitus Research Initiative Tinnitus Clinics Network
[4]Centro de Tratamento e Pesquisa em Zumbido, Belo Horizonte, Brazil
[5]Otolaryngology Department, University of São Paulo Medical School, São Paulo, Brazil
[6]Department of Otolaryngology, University of Regensburg, Germany
[7]Brain & TRI Tinnitus Clinic, Department of Neurosurgery, University of Antwerp, Belgium
[8]Department of Psychiatry, Psychosomatics and Psychotherapy, University Regensburg, Germany

Corresponding Author:
Florian Zeman, MathD, Center for Clinical Studies, University Hospital Regensburg, Franz-Josef-Strauss-Allee 11, 93053 Regensburg, Germany
Email: florian.zeman@klinik.uni-regensburg.de

loudness or minimal masking levels, visual analog scales and numeric rating scales of tinnitus loudness or annoyance have been used. Moreover, different standardized questionnaires for the assessment of tinnitus-related handicaps have been developed, validated, and translated into different languages. The situation is further complicated by the relatively low correlation between the different assessment instruments.[1] The best test-retest reliability and the best validity with quality of life have been shown for standardized tinnitus questionnaires.[2] However, none of these questionnaires has been developed for the assessment of treatment-related changes; furthermore, only limited data are available about what changes in the score can be considered clinically relevant.[3]

Knowledge about the minimal change in the score of a specific questionnaire that can be considered clinically relevant is important for both the interpretation and the design of clinical trials. One example is the estimation of the sample size, which determines the number of patients needed for a study to reject the null hypothesis and to confirm the alternative hypothesis (ie, to show that the new treatment has an effect) under a predefined alpha and beta level. For the calculation of the sample size, the most important information is the expected effect of the new treatment compared with placebo. The expected effect has to be both statistically different from placebo and clinically relevant. Information about clinically relevant changes is also needed for the interpretation of study results. A statistically significant score change may not necessarily be also clinically relevant.

The Tinnitus Handicap Inventory (THI) is probably the most widespread validated questionnaire for quantifying tinnitus severity.[4] In recent prospective randomized studies, the THI has also been used as primary or secondary outcome measurements.[5-8] Because of its widespread application, the THI has been recommended in a consensus document to be used as an outcome measurement in clinical trials to allow comparability across studies.[2] With regard to the clinical relevance of changes in the THI score, the only available orientation is the 95% confidence interval of 20 points, suggesting that a change of 20 points can be considered significant.[9]

To estimate clinically relevant changes, we analyzed data from the Tinnitus Research Initiative (TRI) database that contains changes in different tinnitus questionnaires from patients undergoing different treatment trials.[10] Here, we compared changes in the THI with patients' subjective impressions of treatment-related changes of tinnitus. The patents' subjective impressions were assessed with the Clinical Global Impression–Improvement (CGI-I).[11]

Materials and Methods

Database

Data from the TRI database were analyzed. The TRI database contains longitudinal data that are collected in a standardized way from patients undergoing different types of treatment interventions in different study centers and different countries.[10] Data collection within the TRI database has been approved by the local ethics committee of the University of Regensburg, Germany.

Patient Inclusion and Exclusion Criteria

At the time of analysis, 320 patients from 6 different centers who had received 21 different forms of treatment, including behavioral therapy, pharmacologic treatment, and brain stimulation, were included in the database. All data sets, including the THI at baseline and the THI and the CGI-I at the end of treatment, were included in the analysis. No exclusion criteria were defined.

A total of 210 patients met the inclusion criteria. The time interval between baseline and the end of treatment was 2 weeks in 2 patients, 4 weeks in 109 patients, and 12 weeks in 99 patients.

Assessments

The THI is a validated and widely used questionnaire for assessing the impact of tinnitus in daily life.[4,12] The THI is also frequently used for documenting the treatment outcomes of tinnitus, even though the THI has not been developed and validated for this purpose. Besides the original English version, translations into Danish, Spanish, Korean, Portuguese, German, Italian, and Chinese have been validated and published.[13-15] The THI consists of 25 items, each with the 3 response options—yes (4 points), sometimes (2 points), and no (0 points)—resulting in a total score range from 0 to 100. A higher score denotes a higher tinnitus-related handicap.

To assess a patient's subjective perception about the change of tinnitus over time, we applied a modified version of the CGI-I[11] for use in tinnitus complaints. In this version, the patients themselves are asked to "rate the total improvement of their tinnitus complaints compared to before the beginning of treatment." Patients had to mark 1 of the 7 answers: 1, very much better; 2, much better; 3, minimally better; 4, no change; 5, minimally worse; 6, much worse; and 7, very much worse.

The THI and further questionnaires were filled in at baseline as well as the CGI-I at all following visits (which had been 1 to 4 visits over 2 to 16 weeks, depending on the type of treatment), including the end of therapy and follow-up. An example about the collected data for a specific treatment can be found in Landgrebe et al.[10]

Statistical Analysis

Patient characteristics are summarized by means of median values and interquartile ranges (first to third quartiles) for continuous variables, as well as frequency counts and percentages for categorical data. For comparison of 2 continuous variables, the Student *t* test for independent and dependent variables was applied; for 3 or more groups, we used the 1-way analysis of variance (ANOVA) test.

To analyze changes in the THI score, we favored the absolute score change over the percentage score change for the following reasons: our analyses showed that the absolute score change is less dependent on baseline than the percentage score change and is therefore more appropriate for future covariance analyses.[16] Also, for planning clinical trials with the THI score as the primary outcome, the expected absolute score change is easier to estimate than the percentage score

change because it is less dependent on the baseline values of the study population, which can range from 0 to 100.

Spearman rank correlation was calculated for a comparison of the patients' CGI-I scores at the end of treatment with the changes of the THI scores.[17] Furthermore, 2 different methods were used to get a reliable answer to which change in the THI score patients perceive as an at least minimal improvement.

Equipercentile Linking

To examine the link between CGI-I and the absolute change of the THI score, we used equipercentile linking, a method identifying those scores on both measures with the same percentile rank.[18] This method has the advantage that, unlike in linear regression models, no linearity needs to be assumed. Linear regression would not be an appropriate method because both variables are measured with random error, and the task is not to predict one variable using the other but to concord both variables. Equipercentile linking has been successfully applied in different fields of psychiatry[19-21] and psychology.[22] For our calculations, we applied the SAS macro EQUIPERCENTILE,[23] a realization of the algorithms described by Kolen and Brennan.[18]

Cohen Effect Size d

We used Cohen effect size *d* for dependent variables to estimate the effects of the THI score change within different groups according to the CGI-I scores and to recommend a patient-relevant THI score change.[24] For calculation of the effect size, we used the following formula:

$$d = \frac{\mu_x - \mu_y}{\sigma^2_{\text{pooled}}}$$

where μ_x and μ_y are the mean values of baseline and the end of treatment, and σ^2_{pooled} is the pooled standard deviation.

For a better interpretation of size and direction of the calculated effect sizes, the absolute value in the numerator was left out.

The corresponding 95% confidence intervals for each effect size were calculated according to Smithson.[25]

For analysis, different groups with regard to the CGI-I scores were formed:

Much better: CGI-I < 3
Minimally better: CGI-I = 3
No change: CGI-I = 4
Worse: CGI-I > 4.

To find a minimal clinical and therefore patient-relevant difference in the change of the THI score, the 2 important groups were minimally better and no change, which present the smallest difference between no treatment effect and small treatment effect. The other 2 groups only served as control groups because they should show a more extreme effect in the change of the THI score.

Table 1. Patient Characteristics at Baseline (N = 210)

Age, y, median (IQR)	54.6 (44.6-63.0)
Tinnitus duration, y, median (IQR)	5.2 (1.9-11.1)
Sex, No. (%)[a]	
Male	134 (63.8)
Female	76 (36.2)
Laterality, No. (%)[a]	
Right	55 (26.2)
Left	77 (36.7)
Both sides/inner head	75 (35.7)
Tinnitus severity at baseline based on the THI, No. (%)[a]	
Slight	17 (8.1)
Mild	70 (33.3)
Moderate	65 (31.0)
Severe	41 (19.5)
Catastrophic	17 (8.1)
Etiology, No. (%)[a]	
Blast trauma	24 (11.4)
Injury of cervical spine	3 (1.4)
Change of hearing	17 (8.1)
Stress	56 (26.7)
Head injury	6 (2.9)
Other	100 (47.6)
Type of treatment, No. (%)[a]	
Pharmaceutical drugs	145 (59.5)
Transcranial direct current stimulation	22 (10.5)
Repetitive transcranial magnetic stimulation	43 (30.0)

Abbreviations: IQR, interquartile range; THI, Tinnitus Handicap Inventory.
[a]Percentages do not add up to 100% because of occasional missing values.

The effect sizes and the appropriate confidence intervals were calculated for all subgroups. Afterward, the effect sizes of the minimally better and no-change groups were compared, and an effect size, which separates them, was chosen. To estimate an appropriate standard deviation of the THI difference in clinical studies, we calculated the mean overall standard deviations of all studies with more than 4 patients. Together with the chosen effect size, we calculated the corresponding change of the THI score. As a rough rule of thumb, the effect size of Cohen can be categorized into small effect (around 0.2-0.3), medium effect (around 0.5), and large effect (around 0.8 to infinity).

Statistical analyses were done with PASW 18.0 (SPSS, an IBM Company, Chicago, Illinois) and SAS 9.2 (SAS Institute, Cary, North Carolina).

Results

Patient Characteristics

Two hundred ten patients from Argentina, Belgium, Brazil, and Germany, aged between 16 and 88 years (median 54.6 years, interquartile range [IQR], 44.6-63.0 years), were included in this study; the tinnitus duration was between 3 months and 44 years (median 5.2 years; IQR, 1.9-11.1 years). Baseline characteristics and other tinnitus-related information are presented in **Table 1**.

Table 2. Clinical Global Impression–Improvement (CGI-I) Characteristics

CGI-I	No. (%)	ΔTHI[a], Mean (SD)	Duration of Tinnitus, y, Mean (SD)
Very much better	11 (5.2)	30.4 (21.4)	6.8 (7.1)
Much better	37 (17.6)	16.6 (18.1)	7.9 (8.8)
Minimally better	49 (23.3)	9.5 (12.9)	8.4 (9.9)
No change	78 (37.1)	3.2 (12.7)	9.1 (8.9)
Minimally worse	27 (12.9)	−2.3 (10.4)	9.1 (9.3)
Much worse	7 (3.3)	−3.2 (12.6)	11.8 (14.5)
Very much worse	1 (0.5)	-2 (—)	2.3 (—)

[a]ΔTHI: Tinnitus Handicap Inventory (THI) at baseline – THI at last day of therapy.

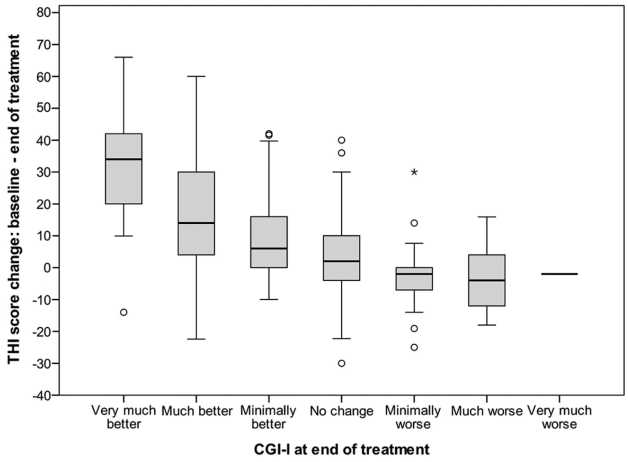

Figure 1. Boxplots of Tinnitus Handicap Inventory (THI) score change categorized by the Clinical Global Impression–Improvement (CGI-I) score at the end of treatment.

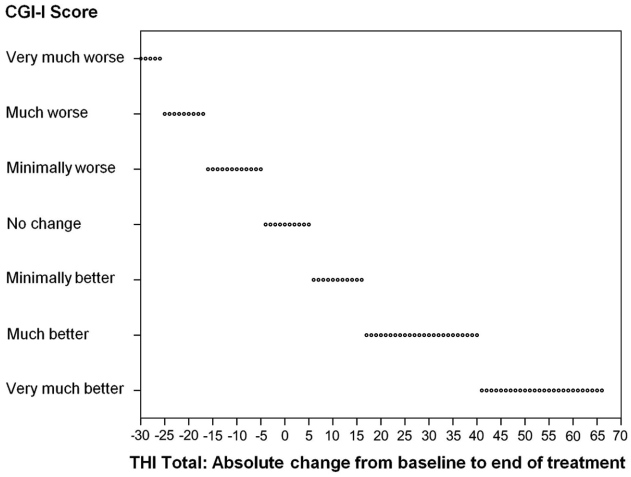

Figure 2. Linking the Clinical Global Impression–Improvement (CGI-I) score with Tinnitus Handicap Inventory (THI) total score change.

Analysis of THI and CGI-I Scores

The THI mean (SD) score at baseline was 45.0 (21.3) and 37.4 (23.6) at the end of therapy, resulting in a significant mean change of 7.6 points ($P < .001$, 95% confidence interval [CI], 5.3-9.7). The frequency distribution of the CGI-I at the end of therapy is shown in **Table 2** together with the corresponding THI mean change and the duration of tinnitus since initial onset. No significant difference exists in the duration of tinnitus for each CGI-I value, but the differences in the THI mean change are highly significant ($P < .001$). The distributions of the change of the THI score, categorized by the CGI at the last day of therapy, are shown in **Figure 1**. The correlation coefficient between the THI mean change and the CGI-I according to Spearman is $r = 0.45$ ($P < .001$).

Linking the CGI-I Score and the Absolute Change of the THI from Baseline to the End of Treatment

The THI absolute score change from baseline to the end of treatment and the CGI-I score at the end of treatment were linked and are presented in **Figure 2**. The scores were linked as follows: feeling "very much worse" on the CGI (CGI-I score 7) corresponded to a THI score change of –30 to –26, "much worse" (CGI-I score 6) corresponded to a THI score change of –25 to –17, "minimally worse" (CGI-I score 5) corresponded to a THI score change of –16 to –5, "no change" (CGI-I score 4) corresponded to a THI score change of –4 to 5, "minimally better" (CGI-I score 3) corresponded to a THI score change of 6 to 16, "much better" (CGI-I score 2) corresponded to a THI score change of 17 to 40, and "very much better" (CGI-I score 1) corresponded to a THI score change of 41 to 66.

Combination of CGI-I Scores and Changes in THI Scores

According to the 4 groups—much better (CGI-I < 3), minimally better (CGI-I = 3), no change (CGI-I = 4), and worse (CGI-I > 4)—the direction, magnitude, and variation of the changes in the THI scores between baseline and the end of therapy are summarized by effect sizes according to Cohen *d*. For calculation of an effect size, mean and pooled standard deviation of the difference between THI at baseline and THI at the last visit are needed for each group. The effect sizes of all groups are summarized in **Table 3**. The effect size of the minimally better group (CGI = 3) is 0.74 (95% CI, 0.42-1.05) and that of patients with no change (CGI = 4) is 0.26 (0.03-0.48). Although the confidence intervals are wide because of the relative small sample sizes and large standard deviations for each group, a notable difference can be seen between the 2 groups with only a small range (0.42-0.48) of the 95% confidence intervals overlapping.

The cutoff point, which separates the no-change group (CGI = 4) from the minimally better group (CGI = 3), has to be outside the 95% confidence interval of the CGI = 4 group and should also represent a minimal acceptable effect size to favor the CGI = 3 group. Therefore, an effect size of $d = 0.5$ (>0.48) is an acceptable choice, which is also the approximate

Table 3. Effect Sizes of CGI-I Groups

CGI-I Group	No.	ΔTHI,[a] Mean (SD)	*d*[b]	95% CI of *d*
Much better group (<3)	48	19.79 (19.58)	1.01	0.66-1.36
Minimally better group (= 3)	49	9.50 (12.92)	0.74	0.42-1.05
No change group (= 4)	78	3.25 (12.65)	0.26	0.03-0.48
Worse group (>4)	35	−2.44 (10.51)	−0.23	−0.57-0.11

Abbreviations: CGI-I, Clinical Global Impression–Improvement; CI, confidence interval.
[a]ΔTHI: Tinnitus Handicap Inventory (THI) at baseline – THI at last day of therapy.
[b]Cohen's effect size between THI at baseline and last day of therapy.

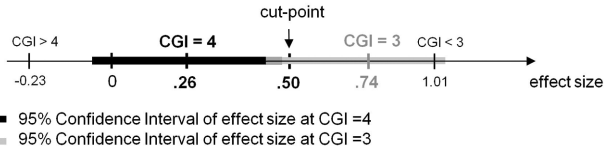

Figure 3. Graphical illustration of effect sizes with confidence intervals. CGI, Clinical Global Impression.

difference of the estimated effect sizes of the 2 groups (**Figure 3**). An effect size of $d = 0.5$ is also in line with the "universality of half a standard deviation" in the interpretation of changes of quality-of-life data.[26] In addition to the effect size, an estimated standard deviation of the before-after difference of the THI score is needed. The mean pooled standard deviation of all used studies with $n \geq 5$ is $SD_{diff} = 13.8 \approx 14$. With an effect size of $d = 0.5$ and an estimated standard deviation of the before-after difference of the THI score of $SD_{diff} = 14$, the change of the THI score can be calculated to $\Delta THI = 7$.

Discussion

The attempt to determine minimal clinically important differences has been made in various assessment areas, such as quality of life,[27,28] depression,[29] schizophrenia,[21] and pain severity.[30] The basic approach for most of these studies is to reference the change of the instrument scores to a categorical rating of changes in patients' health status. According to this approach, by using the CGI-I as the self-assessment of the current health status, we investigated the size of a meaningful change in the THI score.

The present study suggests that a minimum clinically significant change in the THI score can be defined by a before-after difference of 7 points, based on a calculated effect size of $d = 0.5$ and an estimated standard deviation of the THI mean score change in clinical studies of $SD_{Diff} = 14$. This definition is also in agreement with the linking method, calculating that a THI score change of 6 to 16 points corresponds with a CGI-I score of 3 (minimally better), and a THI score change of –4 to 5 corresponds with a CGI-I score of 4 (no change). Thus, both statistical methods show that patients perceive a THI score reduction of at least 7 points as an improvement. Furthermore, the equilinkage method indicates that a reduction of 17 points or more is perceived as a highly relevant improvement (much better). These data suggest a reduction of 7 points in the THI as a meaningful response criterion in clinical trials. A reduction of 17 points could be used as a criterion for a "super response."

We are well aware that this analysis does not account for possible covariables, such as the THI baseline score, age, etiology, duration of tinnitus, or type of treatment. All these factors may influence the subjective perception of how tinnitus is changing over time. For example, elderly patients with a 10-year history of tinnitus will rate their improvement of tinnitus complaints rather differently than younger patients who have had tinnitus just for several months. Unfortunately, the sample size of this study did not allow us to analyze the role of these potential covariables. Therefore, further studies with larger samples are needed to confirm our findings and to evaluate the role of potentially influencing factors.

Conclusion

This study serves as an orientation of what difference in the THI score corresponds to a clinically meaningful improvement in individual patients. Further analyses from larger and independent samples are needed to cross-validate and confirm this result. Furthermore, the role of potential influencing factors, such as the THI baseline score, age, etiology, or the duration of tinnitus, should be investigated.

Author Contributions

Florian Zeman, design, analysis and interpretation, drafting the paper; **Michael Koller**, design and interpretation, reviewing paper, final approval; **Ricardo Figueiredo**, acquisition of the data, reviewing paper, final approval; **Andreia Aazevedo**, acquisition of the data, reviewing paper, final approval; **Marcello Rates**, acquisition of the data, reviewing paper, final approval; **Claudia Coelho**, acquisition of the data, reviewing paper, final approval; **Tobias Kleinjung**, acquisition of the data, reviewing paper, final approval; **Dirk de Ridder**, acquisition of the data, reviewing paper, final approval; **Berthold Langguth**, design, interpretation and acquisition of the data, reviewing paper, final approval; **Michael Landgrebe**, interpretation and acquisition of the data, reviewing paper, final approval.

Disclosures

Competing interests: Florian Zeman received payment of a half-time position as a statistician for the past 24 months by the Tinnitus Research Initiative. Claudia Coelho is a grant recipient of the Tinnitus Research Initiative.

Sponsorships: Tinnitus Research Initiative.

Funding source: Tinnitus Research Initiative.

References

1. Hiller W, Goebel G. Factors influencing tinnitus loudness and annoyance. *Arch Otolaryngol Head Neck Surg*. 2006;132:1323-1330.
2. Langguth B, Goodey R, Azevedo A, et al. Consensus for tinnitus patient assessment and treatment outcome measurement: Tinnitus Research Initiative meeting, Regensburg, July 2006. *Prog Brain Res*. 2007;166:525-536.

3. Newman CW, Sandridge SA. Tinnitus questionnaires. In: James BS, ed. *Tinnitus: Theory and Management*. Hamilton, Ontario: BC Decker; 2004:237-254.
4. Newman CW, Jacobson GP, Spitzer JB. Development of the Tinnitus Handicap Inventory. *Arch Otolaryngol Head Neck Surg*. 1996;122:143-148.
5. Anders M, Dvorakova J, Rathova L, et al. Efficacy of repetitive transcranial magnetic stimulation for the treatment of refractory chronic tinnitus: a randomized, placebo controlled study. *Neuro Endocrinol Lett*. 2010;31(2):238-249.
6. Neri G, De Stefano A, Baffa C, et al. Treatment of central and sensorineural tinnitus with orally administered melatonin and sulodexide: personal experience from a randomized controlled study. *Acta Otorhinolaryngol Ital*. 2009;29(2):86-91.
7. Claire G, Preben HS, Torben W, Jan S. Effectiveness of a tinnitus management programme: a 2-year follow-up study. *BMC Ear Nose Throat Disord*. 2009;9:6.
8. Landgrebe M, Binder H, Koller M, et al. Design of a placebo-controlled, randomized study of the efficacy of repetitive transcranial magnetic stimulation for the treatment of chronic tinnitus. *BMC Psychiatry*. 2008;15:8-23.
9. Newman CW, Sandridge SA, Jacobson GP. Psychometric adequacy of the Tinnitus Handicap Inventory (THI) for evaluating treatment outcome. *J Am Acad Audiol*. 1998;9:153-160.
10. Landgrebe M, Zeman F, Koller M, et al. The Tinnitus Research Initiative (TRI) database: a new approach for delineation of tinnitus subtypes and generation of predictors for treatment outcome. *BMC Med Inform Decis Mak*. 2010;10:42.
11. Guy W. *ECDEU Assessment Manual for Psychopharmacology*. Rockville, MD: US Department of Heath, Education, and Welfare Public Health Service Alcohol, Drug Abuse, and Mental Health Administration; 1976.
12. McCombe A, Bagueley D, Coles R, McKenna L, McKinney C, Windle-Taylor P. Guidelines for the grading of tinnitus severity: the results of a working group commissined by the British Association of Otolaryngologists, Head and Neck Surgeons. *Clin Otolaryngol Allied Sci*. 2001;16:388-393.
13. Letícia PS, Vanessa NT, Celso DI, Daniel D, Mariana MS. Brazilian Portuguese Language version of the "Tinnitus Handicap Inventory": validity and reproducibility. *Rev Bras Otorrinolaringol*. 2006;72(6):808-810.
14. Monzani D, Genovese E, Marrara A, et al. Validity of the Italian adaptation of the Tinnitus Handicap Inventory: focus on quality of life and psychological distress in tinnitus-sufferers. *Acta Otorhinolaryngol Ital*. 2008;28(3):126-134.
15. Zachariae R, Mirz F, Johansen L, Andersen S, Bjerring P, Pedersen C. Reliability and validity of a Danish adaptation of the Tinnitus Handicap Inventory. *Scand Audiol*. 2000;29(1):37-43.
16. Kaiser L. Adjusting for baseline: Change or percentage change. *Stat Med*. 1989;8:1183-1190.
17. Spearman C. The proof and measurement of association between two things. *Am J Psychol*. 1904;15:72-101.
18. Kolen MJ, Brennan RL. *Test Equating, Scaling, and Linking: Methods and Practices*. 2nd ed. New York: Springer-Verlag GmbH; 2004.
19. Leucht S, Kane JM, Etschel E, Kissling W, Hamann J, Engel RR. Linking the PANSS, BPRS, and CGI: clinical implications. *Neuropsychopharmacology*. 2006;31:2318-2325.
20. Leucht S, Kane JM, Kissling W, Hamann J, Etschel E, Engel R. Clinical implications of Brief Psychatric Rating Scale scores. *Br J Psychiatry*. 2005;187:366-371.
21. Leucht S, Kane JM, Kissling W, Hamann J, Etschel E, Engel RR. What does the PANSS mean? *Schizophr Res*. 2005;79:231-238.
22. Jensen AR, Saccuzzo DP, Larson GE. Equating the advanced and standard form of the Raven progressive matrices. *Educ Psychol Meas*. 1988;48:1091-1095.
23. Price LR, Lurie A, Wilkins C. EQUIPERCENT: A SAS program for calculating equivalent scores using the equipercentile method. *Appl Psychol Meas*. 2001;25:332.
24. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum; 1988.
25. Smithson M. *Confidence Intervals*. Thousand Oaks, CA: Sage; 2002.
26. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality-of-life: the remarkable universality of half a standard deviation. *Med Care*. 2003;41:582-592.
27. Osoba D, Rodrigues G, Myles J, Zee BJ. Interpreting the significance of changes in health-related quality-of-life scores. *J Clin Oncol*. 1998;16:139-144.
28. Osoba D. Interpreting the meaningfulness of changes in health-related quality of life scores: lessons form studies in adults. *Int J Cancer Suppl*. 1999;12:132-137.
29. Montgomery SA. Clinically relevant effect sizes in depression. *Eur Neuropsychopharmacol*. 1994;4:283-284.
30. Todd KH, Funk KG, Funk JP, Bonacci R. Clinical significance of reported changes in pain severity. *Ann Emerg Med*. 1996;27:485-489.