Reliability and validity of MINCIR scale for methodological quality in dental therapy research

Ricardo Andrés CARTES-VELÁSQUEZ[a]
Carlos MANTEROLA[a]
Pedro ARAVENA[a]
Javier MORAGA[b]

[a]Programa de Doctorado en Ciencias Médicas, Facultad de Medicina, Universidad de La Frontera, Temuco, Temuco, Chile.

[b]Centro de Investigación Biomédica, Facultad de Ciencias de la Salud, Universidad Autónoma de Chile, Santiago, Santiago, Chile.

Corresponding Author:
Ricardo Andrés Cartes-Velásquez
E-mail: cartesvelasquez@gmail.com

Abstract: The clinical paradigm of evidence-based medicine requires a foundation of good quality research upon which clinical and epidemiological decisions can be based. Several instruments have been designed to assess research quality and validated, though most have limitations. The MINCIR scale was designed to determine the methodological quality (MQ) of clinical research, its psychometric properties for dental research involving large-scale evaluations have not yet been determined. The aim of this study was to determine the validity and reliability of the MINCIR scale for assessment of the MQ of dental therapy studies published in journals indexed in Institute for Scientific Information (ISI) databases. A validation study was performed on a sample of 99 articles from four representative ISI dental journals. Criterion validity was determined in relation to level of evidence (LoE) classification as described by the Oxford Center for Evidence-Based Medicine (OCEBM) ranking system, reliability was determined by calculation of intra-class correlation coefficient (ICC) values, and internal consistency was determined by calculation of Cronbach's alpha. Very good inter-observer reliability (ICC = 0.93), excellent temporal stability (ICC = 0.97), good internal consistency (Cronbach's alpha = 0.77), and a strong (inverse) correlation with OCEBM LoEs (-0.807; $p < .0001$) were obtained. These results indicate that the MINCIR scale has adequate psychometric properties and therefore is a valid option for use in the assessment of MQ in dental therapy research articles.

Keywords: Reproducibility of Results; Research Design; Dental Research.

Introduction

The paradigm shift in medical practice from experience to evidence has increased demand for research in the health sciences.[1] Rigorous studies are needed to provide scientific support and inform clinical practices from the bedside level up to the level of global public health policies. However, quality and utility are variable among studies. Indeed, quality decreases as a function of the number and severity of biases in planning, implementation, and presentation.[1,2,3]

Several grading systems or level of evidence (LoE) ranking rubrics have been developed to differentiate among studies of differing quality, such as the Sackett and the Oxford Center for Evidence-Based Medicine (OCEBM) system.[1] In addition, reading/writing checklists that identify particular items and pieces of information that are necessary for reports to be con-

sidered of good quality, such as CONSORT, STROBE, and PRISMA, are available for many study designs.[2]

Research can also be evaluated with the use of instruments that employ methodological quality (MQ) scales, such as the Jadad, PEDro, and MINCIR, which consider internal and external validity. Such systems enable enhanced assessment of the research's usefulness.[3] Unfortunately, evaluation instruments for dentistry research are scarce and most that are available are limited to specific specialties, geographic areas, or research designs.[4,5] Evaluation methods that cover diverse dental therapies have the potential to provide relevant information to clinicians and improve dental practice.

Originally, the MINCIR scale was designed and validated for surgical research.[6] It assesses the methodological quality of a study based on research design, sample size, and methodology. MQ assessments cannot replace LoE, or other, evaluations but can complement traditional quality evaluations of clinical research.[6] Furthermore, instruments that show psychometric properties are needed to perform any evaluation. The aim of this study was to determine the validity and reliability of the MINCIR scale for determining MQ assessments of dental studies published in 2012 in journals indexed in an Institute for Scientific Information (ISI) database.

Methodology
Study design and instrumentation

This study was designed to assess the validity of the MINCIR scale. The MINCIR scale[6] includes three sections of items, namely study design, population (amount and justification of sample size), and methodology (objective, design, selection criteria, and sample size calculation), as shown in Table 1. The MINCIR scale total score ranges from a minimum of 6 points for articles of the lowest MQ (mainly poor quality case reports), to a maximum of 36 points for well-performed and reported multi-center, double-blinded randomized clinical trials that include data from at least 200 patients (with that subject number being justified).

Population and sampling

This study looked at dental studies in humans, indexed as "Dentistry, Oral Surgery, and Medicine" in the Web of Science (Thomson-Reuters, formerly ISI) for the year 2012. Adopting the parameters of a 95% confidence level, 1-point accuracy, and a standard deviation of 4.95 points,[6] we determined that the required sample size was 94, which we rounded to 100. Given that MQ would be expected to correlate with impact factor, as in surgical studies,[6] a stratified sampling method was used. We selected one journal per quartile (Q1 to Q4) of the 2012 Journal Citation Report. The number of articles selected for each journal was proportional to the number of articles published in all the journals of that journal's quartile for 2012. In journals that had more articles than we required, we used a random sampling method to select which articles to include.

Selection criteria and study procedures

This study included dental studies in humans published in journals indexed in an ISI database and inclusion was not restricted based on gender or age. We excluded articles published in dental-specialty journals (*e.g.* those only covering maxillofacial surgery or periodontology, *etc.*), review articles, *in vitro* research, and bibliometric studies. Scale application was standardized between two evaluators (RC-V and PA) and applied independently, according to MINCIR scale instructions.[7] Each study was scored by conducting a critical item-by-item analysis (see Table 1). Differences among evaluators were solved by consensus with a third evaluator (JM).

Variables

Validity was determined in accordance with OCEBM LoEs. Briefly, each study was classified as belonging to one of four levels. Good-quality clinical trials, good-quality cohort studies, and good quality case-control studies were classified as belonging to level 1b, 2b, and 3b, respectively. Case reports and low-quality cohort/case-control studies were classified as level 4. Expert opinion articles or reviews were considered to be level-5 studies and were not included in this study.

Internal consistency was defined as the degree of consistency of scale assignments. Inter-observer

Table 1. MINCIR scale item scoring (adapted from Manterola et al.[6])

Item	Points
Study design	
Multicenter randomized clinical trial (double blind)*	12
Randomized clinical trial (double blind)*	9
Clinical trial (simple blind or nonblinded, nonrandomized)†	6
Concurrent cohort studies	4
Case-control studies and historic cohort studies	3
Cross-sectional studies	3
Case series	1
Sample size	
201	6 or 12
151–200	5 or 10
101–150	4 or 8
61–100	3 or 6
31–60	2 or 4
30	1 or 2
Methodology	
Objective	
Precise and definite objectives are presented	3
Imprecise objectives are presented	2
No objectives are presented	1
Design	
Design used is mentioned and justified	3
Design used is only mentioned	2
Design used is neither mentioned nor justified	1
Selection criteria	
Inclusion and exclusion criteria are described	3
Inclusion or exclusion criteria are described	2
No selection criteria are described	1
Sample size	
Justified sample size	3
Nonjustified sample size	1

*Includes clinical trials with restricted randomization and quasi-experimental studies.
†Includes experimental studies (pre-post).

reliability values reflected agreement between two independent evaluators. Temporal stability described how well scores across two assessment trials with an inter-assessment interval of 4 weeks agreed with each other.

Statistical analyses

Inter-observer reliability and temporal stability were indexed by calculating intra-class correlation (ICC) coefficients. Internal consistency was determined using Cronbach's alpha and criterion validity was tested with the Wilcoxon test and Pearson correlation. Data were tabulated using an Excel 2003 (Microsoft Corporation, Redmond, USA) spreadsheet, and analyzed on STATA 10/SE (STATA Corporation, College Station, USA).

Ethical considerations

The identities of researchers and institutions associated with the articles were protected.

Results

Of 100 articles selected, 1 was excluded from the analysis making the final sample number 99. The articles were distributed among four journals as follows: *Clinical Oral Investigations* (Q1, n = 30), *Oral Surgery Oral Medicine Oral Pathology Oral Radiology and Endodontics* (Q2, n = 35), *Medicina Oral Patología Oral Cirugía Bucal* (Q3, n = 19), and the *Journal of Dental Sciences* (Q4, n = 15).

The LoE distribution for the analyzed studies was as follows: level 4, 59.6%; level 3b, 3.0%; level 2b, 21.2%; and level 1b, 16.2%. Overall, the data had very good inter-observer reliability, excellent temporal stability, and good internal consistency. The section and total ICC coefficient values for MINCIR inter-observer reliability and temporal stability are reported in Table 2, and the section and total Cronbach's alpha values for internal consistency of MINCIR scale scores are reported in Table 3.

Statistical criterion validity testing revealed that MQ differed significantly among the four LoE groups of articles ($p < .0001$). A strong correlation (-0.807; $p < .0001$) was observed between LoE (lower level is better) and MQ (higher score is better). The mean MQ scores for each LoE were as follows: level 4, 9.3 ± 2.3; level 3b, 12.3 ± 3.2; level 2b, 15.3 ± 3.2; and level 1b, 22.8 ± 5.9.

Table 2. Inter-observer reliability and temporal stability.

Item	Inter-observer reliability ICC	Temporal stability ICC
Design	0.78	0.98
Sample size	0.97	0.97
Methodology	0.87	0.81
MINCIR scale total	0.93	0.97

Table 3. Internal consistency of the MINCIR scale.

Item	Item-rest correlation	Alpha for deleted-item
Design	0.63	0.70
Sample size	0.54	0.76
Methodology	0.75	0.63
MINCIR scale	–	0.77

Discussion

For clinical application, it is important that research meet the highest LoE criteria possible. Unfortunately, in dentistry[4,5,8] and other biomedical sciences,[6] many published research articles would be considered to have a low LoE. However, LoEs provide an incomplete picture since studies that have the same LoE nominally, may differ in essential methodological aspects (*e.g.* sample size, justification, or selection criteria). Hence, the concept of MQ may be used to complement LoE classification.[3]

There has been a recent increase in the number of articles considering MQ, especially in the realm of systemic reviews' consideration of primary sources.[9] Palys *et al.*[10] advocate multiplying research quality scores rather than adding them such that there may be greater separation of research (articles) by quality and a greater assurance that articles classified as high quality fulfill a high number of, if not all, criteria. There are various scales that may be used to measure this concept.[3] Unfortunately, thus far, they have been focused on specific study designs. An integrative method for evaluating an entire research scenario is needed. For example, the Jadad scale was developed to assess the quality of clinical trials in pain research, and it was extended to assess additional types of clinical trials. But it cannot be used to assess observational studies. A similar circumstance exists for the PEDro scale.[3]

The MINCIR scale overcomes this limitation, allowing assessment of observational and experimental studies. This characteristic is an important advantage considering that clinical trials represent a minor portion of research articles in dentistry.[4,5,8] An additional advantage of the MINCIR scale is that it enables one to perform systematic reviews

of clinical trials, cohort studies, case-control, and case series studies.[11]

The MINCIR scale has been used to assess surgical studies,[6] including studies in the areas of oral and maxillofacial surgery,[8] and has demonstrated psychometric properties.[12] However, these may vary depending on the study population; so it is necessary to have validation procedures to assure adequacy. Our analysis indicates that the MINCIR scale has adequate psychometric properties, consistent with previously reported data for this scale.[9] In particular, scores exceeding 0.9 for reliability and 0.7 for internal consistency indicate excellent performance. These values exceed those reported for the Jadad *et al.*,[13] the PEDro,[14] and other similar scales.[3]

Reliability results for particular items of the MINCIR are promising. However, the design items are hard to assess, perhaps due to unclear descriptions of the studies. For example, some articles may be presented as cohort or case-control studies, but if they were poorly designed and reported, it is difficult to classify them appropriately.

Our internal consistency analysis revealed that all of the MINCIR sections had good item-rest correlation, with the weakest item-rest correlation being observed for the sample size section, which may be due, at least in part, to the fact that the sample size required for particular studies varies depending on the type of research being conducted. For example, a multi-center randomized double-blind clinical trial comparing two interventions may require 100 patients (6 points) or less to obtain valid results, whereas a case series study may include 500 patients (12 points) but still yield poor-quality results. One solution to this issue may be to assign more importance to the justification of sample size than to the number itself. More research is needed to clarify this aspect of scoring.

MINCIR scale scores had a strong correlation with LoE in the present study using the criteria of the OCEBM[1] as an index of validity. This strong correlation (-0.8) provides robust evidence of the MINCIR scale's validity because the OCEBM's LoE rubric is inclusive of observational and experimental studies and is a widely recognized instrument for assessment of research quality in the biomedical sciences.

In this study, stratified sampling (one journal per quartile) was selected because of the presumed correlation between study MQ and journal impact factor.[6] Hence, the stratified sampling method was chosen to ensure greater variability in the studies assessed and to obtain a representative sample of articles. We selected articles published in only one year (*i.e.* 2012) because journals' impact factors vary annually and, therefore, including more than a single year may have undermined the study. We also thought it unlikely that MQ varied considerably between years. Research processes and output quality depend on several factors (*e.g.* financial, educational, public-policy, *etc.*) that evolve over long periods. Moreover, previous examination of MQ using the MINCIR scale, in the field of surgery, showed little variation of MQ across articles published between 2000 and 2004.[15] Finally, it may be argued that the four selected journals may not have been representative of all dental research. However, they do cover most areas of dental research, albeit with a possible over-representation of surgical-pathologic research.

Conclusions

The MINCIR scale has adequate psychometric properties. It is, therefore, a valid option for assessment of MQ in dental therapy research articles.

References

1. Manterola C, Zavando D. Cómo interpretar los "Niveles de Evidencia" en los diferentes escenarios clínicos. Rev Chil Cir. 2009 Dec;61(6):582-95.

2. Richards D. The EQUATOR network and website. Evid Based Dent. 2007;8(4):117.

3. Silva FC, Valdivia Arancibia BA, Iop RR, Gutierres Filho PJB, Silva R. Escalas y listas de evaluación de la calidad de estudios científicos. Rev Cubana Inf Cienc Salud. 2013 Sep;24(3):295-312.

4. Sequeira-Byron P, Fedorowicz Z, Jagannath VA, Sharif MO. An AMSTAR assessment of the methodological quality of systematic reviews of oral healthcare interventions published in the Journal of Applied Oral Science (JAOS). J Appl Oral Sci. 2011 Sep-Oct;19(5):440-7.

5. Tu Y-K, Maddick I, Kellett M, Clerehugh V, Gilthorpe MS. Evaluating the quality of active-control trials in periodontal research. J Clin Periodontol. 2006 Feb;33(2):151-6.

6. Manterola C, Pineda V, Vial M, Losada H. What is the methodologic quality of human therapy studies in ISI surgical publications?. Ann Surg. 2006 Nov;244(5):827-32.

7. Moraga J, Manterola C, Cartes-Velásquez R, Burgos M, Aravena P, Urrutia S, Grupo MINCIR. Instrucciones para la utilización de la escala MINCIR para valorar calidad metodológica de estudios de terapia. Int J Morphol. 2014 Mar; 32(1): 294-8.

8. Aravena P, Cartes-Velásquez R, Manterola C. Productividad y calidad metodológica de artículos clínicos en cirugía oral y máxilofacial en Chile. Período 2001-2012. Rev Chil Cir. 2013 Sep;65(5):382-8.

9. PubMed.gov [homepage]. Bethesda MD: National Center for Biotechnology Information, U.S. National Library of Medicine; 2014 [cited 2014 Jan 23] Available from: http://www.ncbi.nlm.nih.gov/pubmed/?term=%22methodological+quality%22.

10. Palys KE, Berger VW, Alperson S. Trial quality checklists: on the need to multiply (not add) scores. Clin Oral Investig. 2013 Sep;17(7):1789-90.

11. Manterola C, Vial M, Pineda V, Sanhueza A. Systematic review of literature with different types of designs. Int J Morphol. 2009 Dec;27(4):1179-86.

12. Moraga J, Burgos M, Manterola C, Sanhueza A, Cartes-Velásquez R, Urrutia S. Confiabilidad de la escala MINCIR para valorar calidad metodológica de estudios de terapia. Rev Chil Cir. 2013 Jun;65(3):222-7.

13. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJM, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control ClinTrials. 1996 Feb;17(1):1-12.

14. Sherrington C, Herbert RD, Maher CG, Moseley AM. PEDro. A database of randomized trials and systematic reviews in physiotherapy. Man Ther. 2000 Nov;5(4):223-6.

15. Pineda V, Manterola C, Vial M, Losada H. ¿Cuál es la calidad metodológica de los artículos referentes a terapia publicados en la Revista Chilena de Cirugía?. Rev Chil Cir. 2005 Dec;57(6):500-7.