Inter-rater reliability of the extended Composite Quality Score (CQS-2) - Protocol

Steffen Mickenautsch (neem@global.co.za)
University of the Western Cape
Stefan Rupf
Saarland University
Ivana Miletić
University of Zagreb
Veerasamy Yengopal
University of the Western Cape

Method Article

Keywords: Protocol, CQS, inter-rater reliability

Posted Date: June 16th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1763870/v1

Inter-rater reliability of the extended Composite Quality Score (CQS-2) - Protocol

Steffen Mickenautsch[1,2,3*], Stefan Rupf[4], Ivana Miletić[5], Veerasamy Yengopal[1]

[1]Faculty of Dentistry, University of the Western Cape, Francie van Zijl Avenue, Tygerberg/Cape Town, 7505, South Africa.

[2] Honorary/Department of Community Dentistry, School of Oral Health Sciences, Faculty of Health Sciences, University of the Witwatersrand, 7 York Rd., Parktown/Johannesburg 2193, South Africa.

[3]Review Center For Health Science Research, 84 Concorde Road East, Bedfordview/Johannesburg, 2008, South Africa.

[4]Chair of Synoptic Dentistry, Saarland University, Building 73, 66421, Homburg, Germany.

[5]Department of Endodontics and Restorative Dentistry, School of Dental Medicine, University of Zagreb, Gunduliceva 5, 10 000, Zagreb, Croatia..

*Corresponding author

Email: neem@global.co.za (SM)

Abstract

Objectives: The aim of this study is to establish the inter-rater reliability of the CQS-2 and to test the null-hypotheses that it does not differ significantly from that of the first CQS version (CQS-1).

Methods: Four independent raters will be selected to rate 45 clinical trial reports using CQS-1 and CQS-2. The raters will remain unaware regarding each other's participation in this study until all rating is completed. Each rater will receive only one rating-template at a time in a random sequence for CQS-1 and CQS-2 rating. Raters will complete each template and send these back to the principal investigator via email. Each rater will receive her/his next template two weeks after submission of the completed previous template. The inter-rater reliabilities for the overall appraisal score of the CQS-1 and the CQS-2 will be established by use of the Brennan-Prediger coefficient. The coefficients of both CQS versions will be compared by use of the two-sample z-test. All data analysis will be carried out using SAS statistical software. A 5% significance level will be used. During secondary analysis, we will establish the inter-rater reliability for each single criterion and each corroboration level for both CQS versions.

Reporting: The results of this study will be reported in line with the Guidelines for Reporting Reliability and Agreement Studies (GRAAS). The final report will be made available online as preprint in one of the major preprint repositories and submitted to a peer-reviewed journal.

Introduction

The Composite Quality Score (CQS) is a novel appraisal tool for prospective, controlled, clinical therapy trials, based on the deductive falsification approach [1]. Trial appraisal that follows such an approach, assumes that any trial design characteristic (or the lack thereof) which lies outside a particular set of applied trial appraisal criteria, such as that of the Jadad scale [2] or Cochane's Risk of Bias (RoB) tool [3,4], may completely falsify the truthfulness of trial results and thus rejects any confidence in 'low bias risk' or high 'Quality of Evidence' (QoE). Consequently the approach accepts that in principle, it is impossible to establish, i.e.: inductively verify, 'low bias risk' or high QoE for any trial. Instead, the CQS follows the concept that, although high QoE cannot be proven, it is possible to establish with high certainty, whether the quality of evidence is low. Low QoE is recognized when characteristics that are essential for a therapy trial, in order to reflect the true effect estimate, are absent [5].

The first version of the CQS (CQS-1) was developed as a composite of trial appraisal categories for, both, systematic and random error. The CQS-1 appeared to have been sufficient for trial appraisal in the field of restorative dentistry, where only two from the total of 683 trial reports where not rated with high confidence as of low QoE [6]. In addition, Mickenautsch et al. (2021) investigated the CQS-1 inter-rater reliability. The results showed an 'almost perfect' inter-rater reliability, according to the Landis/Koch Kappa's Benchmark Scale (Brennan-Prediger coefficient 0.95; 95% CI: 0.87 – 1.00) that was statistically significantly higher that that of the first version of Cochrane's RoB tool [7].

However, while the current CQS-1 appeared to have been sufficient for clinical trial appraisal in the field of restorative dentistry, other fields of clinical therapy may contain a higher number of trials that would pass its three simple, non-restrictive criteria. For that reason a new CQS version (CQS-2) was developed based on meta-epidemiological study evidence. Subsequently, one new criterion concerning double blinding was added and criterion II and III of the original CQS version

amended [8]. These changes raise the questions whether the CQS-2 is associated with high inter-rater reliability, too, and whether such reliability differs from that of the CQS-1?

Therefore, the aim of this study is to establish the inter-rater reliability of the CQS-2 and to test the null-hypotheses that it does not differ significantly from that of the CQS-1.

Methodology

Rater selection

One independent rater will be selected from a cohort of academics known to each of the investigators: SM, SR, IM and VY based on the following criteria (to the best of the investigators' knowledge):

(i) Knowledge of research methodology;

(ii) Potential and/or demonstrated past interest in conducting systematic reviews of clinical trials;

(iii) Independent from each other and from the investigators: SM, SR, IM and VY (e.g. no joint publication listed in PubMed or other known prior academic collaboration);

(iv) Positive response to the written invitation for participation as rater.

Hence a total of four independent investigators will be selected. From the potential number of raters contacted, the first four will be selected who agree to participate. The number of raters (N = 4) is determined in accordance with a similar study to assess the inter-rater reliability of the CQS-1, published elsewhere [7]. Rater selection will be quasi-random, i.e. although no selection according to a random sequence will be conducted, acceptance to participate by each rater will be left to chance. Raters will be free to accept or decline a once-off written invitation without any further afford by the investigators to secure study participation. In order to assure rater independence, no rater-interaction will take place during the rating process, thus avoiding any interaction effect on the results.

Rater blinding

The raters will remain unaware regarding each other's participation in this study until all rating is completed. However, in order to investigate the use of the CQS-2 under conditions as close as possible to the practical routine of trial appraisal, the raters will not be blinded to the trial reports' references, author names and affiliations, as well as acknowledgments and funding sources. In addition, in order to obtain raters' informed consent regarding their participation in this study they will receive information about the full content of the study protocol. Hence, each rater will be aware that her/his judgment will be compared with those of other raters.

Sample size calculation

The number of required trial reports is calculated based on a minimum expected agreement between raters of 70%, and a 95% confidence interval of 15%, using the appropriate formula for sample size calculation: $N = 1/E^2$ (with N = number of required articles and E = confidence interval) [9]. In line with the applied sample size calculation method, a minimum number of 44 (rounded to 45) required trial reports is determined.

Trial report selection

The 45 trial reports will be selected from PubMed. The database will be searched using the search term 'prospective AND clinical AND controlled AND trial' with the set limits: 'Abstract', 'Free full text' [Text availability], 'Clinical trial' [Article type], 'From 2022/1/1 to 2022/05/31' [Publication date], 'Best match' [Display options]. Citation abstracts will be checked whether they describe a prospective, clinical, controlled trial, published in English language. The first 45 relevant citations will be selected (trial protocols or trials in other publication language than English will not be included).

Trial rating process

Each rater will receive from the principal investigator (SM) via email a download link for the 45 trial reports and a MS Excel assessment template for both CQS versions that are prepared in line with published specifications for each appraisal method [7,8]. Each rater will receive only one template at a time in a random sequence for CQS-1 and CQS-2 rating. The random sequence will be generated using block randomization (Block size = 2) out of a total of 8 rating events. Raters will enter their rating results into the template and send these back to the principal investigator via email. Each rater will receive her/his next template two weeks after submission of the completed previous template.

The Composite Quality Score (CQS)

The CQS includes: (i) binary trial report rating per appraisal criterion (Scores: 0 = invalid/falsified, 1 = corroborated); (ii) multiplication of individual rating scores to an overall appraisal score and (iii) identification of invalid/falsified trial reports based on a zero overall appraisal score.

(i) ***CQS-1***

The CQS-1 was originally developed as a composite of two trial appraisal categories for systematic and random error [7]. For each category the following criteria were set:

(a) Systematic error (Randomisation)

Criterion I: 'Randomisation' for allocation to treatment groups is in some form reported in the text (Yes = 1 / No = 0);

Criterion II: Concealing of the random allocation is in some form reported in the text (Yes = 1 / No = 0).

(b) Random error (Sample size)

Criterion III: The sample size of any particular treatment group reported in the trial report is not less than N = 200 (Yes = 1 / No = 0).

The minimum sample size limit (N) was calculated using the formula: $N = \{([P1 \times (100 - P1)] + [P2 \times (100 - P2)])/(P2 - P1)^2\} \times f(\alpha,\beta)$ [10] and was based on the assumption that the difference in intervention effect between study groups (P1 - P2) is not less than 10%, with $\alpha = 5\%$ and $\beta = 20\%$, i.e.: $f(\alpha,\beta) = 7.9$ [11].

(ii) ***CQS-2***

The CQS-2 is an update of the CQS-1 and based on a systematic review with meta-analysis of meta-epidemiological study evidence, concerning the lack of trial design characteristics associated with over- or under estimation of the correct effect estimate due to systematic error alone [8]. In contrast to the CQS-1, the CQS-2 does not include a category for random error. The following criteria were set:

Criterion I: 'Randomisation' for allocation to treatment groups is in some form reported in the text (Yes = 1 / No = 0);

Criterion II:

(i) Keeping of the random allocation sequence in a locked computer file; **and**

(ii) Translation of the sequence into identical, coded, serially administered containers and/or sealed, opaque envelopes; **and**

(iii) Reassurance that the person who generated the sequence did not administer it.

are in some form reported in the text (Yes = 1 / No = 0)

Criterion III: Double-blinding or the blinding of at least two out of the three groups: trial participants; trial personnel and trial outcome assessors in some form reported in the text (Yes = 1 / No = 0)

Criterion IV: The sample size of any particular treatment group reported in the trial is not less than N = 100 (Yes = 1 / No = 0)

Statistical analysis

The inter-rater reliabilities for the overall appraisal score of the CQS-1 and the CQS-2 will be established by use of the Brennan-Prediger coefficient [9]. The coefficients of both CQS versions will be compared by use of the two-sample z-test. All data analysis will be carried out using SAS statistical software [12]. A 5% significance level will be used.

During secondary analysis, we will establish the inter-rater reliability for each single criterion and each corroboration level [5] (for CQS-1: C1-3; for CQS-2: C1-4) for both CQS versions, as described above.

Reporting

The results of this study will be reported in line with the Guidelines for Reporting Reliability and Agreement Studies (GRAAS) [13]. The final report will be made available online as preprint in one of the major preprint repositories and submitted to a peer-reviewed journal.

Financial Disclosure

The authors received no specific funding for this work.

Data availability

All data will be made fully available without restriction as part of the preprint and within the manuscript and its Supporting Information files.

Competing Interests

The authors SM, SR and IM have actively contributed to the initial conceptualization, development and dissemination of the CQS-1, prior start of this study. All authors declare to have no financial interest in the study's outcome.

References

[1] Mickenautsch S. Is the Deductive Falsification Approach a Better Basis for Clinical Trial Appraisal? Rev Recent Clin Trials. 2019;14(3):224-228.

[2] Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Controlled Clin. Trials 1996; 17: 1-12.

[3] Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Savovic J, Schulz KF, Weeks L, Sterne JA; Cochrane Bias Methods Group; Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011 Oct 18;343:d5928.

[4] Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, Cates CJ, Cheng HY, Corbett MS, Eldridge SM, Emberson JR, Hernán MA, Hopewell S, Hróbjartsson A, Junqueira DR, Jüni P, Kirkham JJ, Lasserson T, Li T, McAleenan A, Reeves BC, Shepperd S, Shrier I, Stewart LA, Tilling K, White IR, Whiting PF, Higgins JPT. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ. 2019 Aug 28;366:l4898.

[5] Mickenautsch S, Rupf S, Miletić I, Yengopal V. The Composite Quality Score (CQS) as appraisal tool for prospective, controlled, clinical therapy trials: rationale and current evidence (Preprint). Authorea: 561091 [Preprint]. 2022 [cited 2022 June 05]. Available from: https://doi.org/10.22541/au.164848881.12588115/v1.

[6] Mickenautsch S. Are Most of the Published Clinical Trial Results in Restorative Dentistry Invalid? - An Empirical Investigation. Rev Recent Clin Trials. 2020; 15: 122-130.

[7] Mickenautsch S, Miletić I, Rupf S, Renteria J, Göstemeyer G. The Composite Quality Score (CQS) as a trial appraisal tool: inter-rater reliability and rating time. Clin Oral Investig. 2021; 25: 6015-6023.

[8] Mickenautsch S, Rupf S, Miletić I, Yengopal V. Extension of the Composite Quality Score (CQS) as an appraisal tool for prospective, controlled clinical therapy trials (Preprint). Research Square: 1704764/v2 [Preprint]. 2022 [cited 2022 June 09]. Available from: https://doi.org/10.21203/rs.3.rs-1704764/v2.

[9] Gwet KL. Handbook of inter-rater reliability. (2nd ed.) Advanced Analytics, LLC: Gainsburg, MD, USA 2010; pp. 54, 159-160.

[10] Pocock SJ. Clinical trials. A practical approach. John Wiley & Sons, Ltd: Chichester 1988; pp. 126-34.

[11] Geigy. Scientific tables, 7th ed. Geigy: Basle 1970; p. 28.

[12] SAS Institute Inc., SAS Software, version 9.4 for Windows, Cary, NC, USA: SAS Institute Inc. (2002-2012).

[13] Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, Roberts C, Shoukri M, Streiner DL. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. J Clin Epidemiol. 2011 Jan;64(1):96-106.

Annexure

Annex 1 – CQSER-Data collection sheet-CQS-1

Annex 2 - CQSER-Data collection sheet-CQS-2

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Annex1CQSERDatacollectionsheetCQS1.xls
- Annex2CQSERDatacollectionsheetCQS2.xls