Annals of Epidemiology 28 (2018) 625–628

Contents lists available at ScienceDirect

Annals of Epidemiology

journal homepage: www.annalsofepidemiology.org

Original article

Risk of bias in randomized controlled trials of psychological treatments for bulimia nervosa and binge eating

Juliana Pagotto Trevizo [a], Phillipa Hay, PhD [b], Walter Swardfager, PhD [c,d], Hugo Cogo-Moreira, PhD [a,*]

[a] Department of Psychiatry, Universidade Federal de São Paulo, São Paulo, Brazil
[b] School of Medicine and Translatinal Institute for Health Research, Western Sydney University, Sydney, Australia
[c] Department of Pharmacology & Toxicology, University of Toronto, Toronto, Canada
[d] Hurvitz Brain Sciences Program, Sunnybrook Research Institute, Toronto, Canada

ARTICLE INFO

Article history:
Received 6 September 2017
Accepted 16 June 2018
Available online 3 July 2018

Keywords:
Risk of bias
Randomized clinical trial
Eating disorder
Psychological intervention
Confirmatory factor analysis

ABSTRACT

Purpose: In the context of Cochrane systematic reviews/meta-analyses of randomized clinical trials, risk of bias (RoB) is assessed using categorical indicators (low, unclear, or high RoB). This study sought to evaluate the indicators of the Cochrane RoB tool available for construct validity as applied to randomized clinical trials of psychological treatments for bulimia nervosa and binge eating.
Methods: Bayesian confirmatory factor analysis was used to test the construct validity of the measurement model underlying the set of five categorical items, and the reliability of these indicators to measure RoB.
Results: In 48 primary randomized clinical trials, the model showed good fit indices and factor loadings higher than 0.4.
Conclusions: The results support the construct validity of the Cochrane RoB tool and the reliability of three of five items in this health intervention context.

Introduction

Eating disorders are common and severe mental health problems associated with impaired health-related quality of life [1]. They include individuals both overeating and the act of not eating enough, and weight and body shape preoccupations. The major disorders delineated in the *Diagnostic and Statistical Manual of Mental Disorders* (fifth ed.; *DSM–5*; American Psychiatric Association, 2013) are bulimia nervosa, anorexia nervosa, and binge-eating disorder (BED). One of the leading psychological therapies for eating disorders is cognitive-behavioral therapy (CBT) [2]. The evidence base for CBT rests on systematic reviews, which utilize transparent and reproducible methods in assessing studies [3,4]. Beyond the CBT, other intervention include "third wave" therapies such as dialectical behavior therapy but their evidence base is less strong [5].

In the context of Cochrane systematic reviews/meta-analyses of randomized clinical trials (RCTs), risk of bias (RoB) is assessed using ordinal scale indicators (low, unclear, or high RoB). Version 5.0 of the Cochrane Handbook (2008), specifies six RoB criteria: 1) Sequence generation 2) Allocation Concealment 3) Blinding of participants, personnel, and outcome assessors 4) Incomplete outcome data 5) Selective outcome reporting and 6) Other sources of bias. Version 5.1 (2011), suggests seven criteria: 1) Random sequence generation, 2) Allocation concealment, 3) Blinding of participants and personnel, 4) Blinding of outcome assessment, 5) Incomplete outcome data, 6) Selective reporting, and 7) Other sources of bias. Some differences might be pointed out regarding the two versions: 1) most of the criteria are intended to capture the same underlying concepts; 2) "Blinding of participants, personnel, and outcome assessors" were divided into two separated items "Blinding of participants and personnel" and "Blinding of outcome assessment", and 3) the scale was changed: previously, there were "Yes" or "No" questions, whereas in Version 5.1, RoB were judged to be low, high, or unclear.

Earlier work has examined empirical evidence supporting the relationship between various bias categories and outcomes in RCTs across bodies of evidence [6,7]. However, only recently, was the Cochrane RoB as tool evaluated in terms of its convergent validity and factorial validity. Studies across different areas of psychiatry

Availability of data and materials: Data were collected from RoB's information contented in *Psychological treatments for bulimia nervosa and binging* systematic review conducted by Phillipa Hay and collaborators.

Conflict of interest: P.H. is lead author of the systematic review that was the subject of this study. The other coauthors declared no potential conflict of interests.

* Corresponding author. Universidade Federal de São Paulo, Department of Psychiatry, 04039-032 São Paulo, Brazil.
E-mail address: hugocogobr@gmail.com (H. Cogo-Moreira).

https://doi.org/10.1016/j.annepidem.2018.06.006
1047-2797/

have shown that the tool may poorly inform RoB as a latent construct. For example, in clinical trials of methylphenidate for attention deficit hyperactivity disorder [8], the Cochrane RoB tool exhibited good fit indices, but the majority of items were not reliable (i.e., evaluated via the magnitude of the factor loadings) in capturing RoB. Those results prompted replication in other areas of research, prompting similar results to emerge where the five RoB indicators showed poor reliability in assessing interventions for autism spectrum disorders ($\omega = 0.687$; 95% confidence interval ranging from 0.613 to 0.761) [9] and psychological therapies for adult post-traumatic stress disorder [10].

Lacking evidence to support the validity of the bias tool used, confidence in the reliability, and "strength of evidence" provided by systematic reviews and meta-analyses might be called into question. Important criteria for the bias tool include 1) the context in which the tool is being used and 2) the reliability of the individual items. First, it is possible that the tool might behave differently in terms of fit when assessing psychological interventions for eating disorders compared to other areas, such as pharmaceutical trials for schizophrenia. Second, the items may differ in the way they capture the risks of bias, with some items having greater reliability, and therefore more likelihood to indicate a pertinent risk when "unclear" or "probable" risks are identified. For meta-analyses involving diagnostic accuracy studies, the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) is used to assess RoB [11]. Evaluating the revised version, the QUADAS-2, in Alzheimer's disease diagnostic accuracy studies, two dimensions (RoB and applicability concerns) were considered and the authors found poor factor loadings [12], raising the question of whether bias assessment instruments could be redesigned based on modern item response theory to provide a more meaningful measurement model. Using a Monte Carlo simulation, it was concluded that at least 90 primary studies would be needed to properly estimate RoB with 80% power across the QUADAS-2′ seven ordinal items.

The aims of this study were to investigate the construct validity of the indicators used by Cochrane as applied in a systematic review of RCTs of psychological treatments for bulimia nervosa and BED and to assess which of the indicators best captured RoB in the context of eating disorders. In this study, we used confirmatory factor analysis (CFA) to determine construct (factorial) validity. Commonly, CFA is applied when research subjects are humans and the indicators are performance of these individuals in tests that inform a latent psychological/mental, academic skills, or psychiatric construct. Moreover, CFA has been extensively used for other areas of knowledge to test different nature of modeling as in economics [13], business [14], and widely as a way to understand the reliability of observed indicators via latent variable modeling [15].

Method

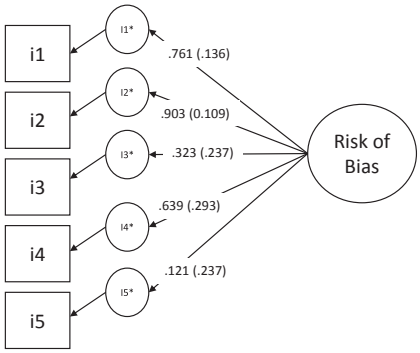
The systematic review "Psychological treatments for bulimia nervosa and binging" included 48 clinical trials ($n = 3054$). The observable units (subjects) considered were the 48 studies. For each included primary study, five ordinal criteria (called here as items) were collected: random sequence generation, concealment allocation, blinding of participants, blinding of assessor, and incomplete outcomes [16]. Each item was classified as having low RoB, unclear, or high RoB (i.e., a categorical ordinal variable) serving as our observed indicators. Data for this manuscript were extracted from Hay, Bacaltchuk [2], presented in the subheading "characteristics of included studies".

Confirmatory Factor Analysis (CFA) was used to model the latent phenomenon underlying the set of Cochrane RoB indicators; CFA is one approach of structural equation modeling. It evaluates the construct validity of different tools (i.e. inventories, batteries, questionnaires, tests, and scales). All these tools intent to measure an underlying feature, not directly observed (i.e., latent) such as RoB. Each primary study will have an amount of RoB, which the tool is intended to identify. As previously defined by Bollen [17], p. 182, "a measurement model specifies a structural model connecting latent variables to one or more measures or observed variables". The observed variables (i.e., the items) are represented by squares in Figure 1. The latent variables, which are not directly observed attributes, are represented by ovals/circles. We used version 7.4 of the Mplus software [18] under a Bayesian estimator (using non-informative prior) due to the reduced sample size, considering the five items as ordinal. The default priors on each loading and threshold are a normal distribution, with 0 mean and variance 5. Because of the Bayesian estimator, the fit indices used to evaluate the goodness of fit as proposed by Muthen and Asparouhov [19] were 1) a posterior predictive p-value (PPP) value within 0.4–0.6 (e.g., the more centered on 0.5, the better), and 2) a 95% confidence interval where a lower band is negative and a zero fit statistic difference falls close to the middle of the interval. To evaluate the correlation between the items of the tool and the underlying RoB factor, we used standardized factor loadings, where values closer to one indicate stronger correlations with the RoB factor. Values close to zero indicate a lack of association between the factor underlying to the item and the RoB. It is possible to transform the factor loadings into common variance (CV) by squaring the factor loading. CV is a percentage measure and offers a measure of reliability [20], ranging from 0 to 1, where values closer to 1 (i.e., equivalent to 100% of CV) would indicate a reliability of 100%.

It is important to state that although our sample size was small ($n = 48$ studies), there was adequate heterogeneity in terms of inclusion criteria, including adults with bulimia nervosa, BED and/or eating disorder not otherwise specified of a bulimic type, and in terms of interventions (CBT and other psychotherapies). Moreover, the sample size was sufficient to conduct the analysis based on the commonly accepted number of around 10 subjects per indicator [21]. Other systematic reviews reporting the RoB table might also be evaluated in this way (we provide Mplus syntax as supplementary material for this purpose).

Results

Table 1 presents a descriptive analysis of the results for each of the items from the 48 studies tabulated according to the author of the Likert scale review (three categories of responses: low risk, unclear, and high risk).



Legend: i1=Random Sequence Generation; i2=Concealment of Allocation; i3=Blinding of Assessor; i4=Blinding of Participants; i5=Incomplete Outcomes

Fig. 1. Shows the measurement model, consisting of the five indicators and their factor loadings onto the RoB latent attribute. * indicates latent response variables for i1–i5.