Check for updates

Original Article

CHRONIC STRESS

Chronic Stress
Volume 2: 1–6
© The Author(s) 2018
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/2470547018779066
journals.sagepub.com/home/css
SAGE

Risk of Bias in Randomized Clinical Trials on Psychological Therapies for Post-Traumatic Stress Disorder in Adults

Juliana Martins Scalabrin[1], Marcelo F. Mello[1], Walter Swardfager[2,3], and Hugo Cogo-Moreira[1]

Abstract

Objective: To evaluate the factorial validity and internal consistency of a measurement model underlying risk of bias as endorsed by Cochrane for use in systematic reviews; more specifically, how the risk of bias tool behaves in the context of studies on psychological therapies used for treatment of post-traumatic stress disorder in adults.

Methods: We applied confirmatory factor analysis to a systematic review containing 70 clinical trials entitled "Psychological Therapies for Chronic Post-Traumatic Stress Disorder in Adults" under a Bayesian estimator. Seven observed categorical risk of bias items (answered categorically as low, unclear, or high risk of bias) were collected from the systematic review.

Results: A unidimensional model for the Cochrane risk of bias tool items returned poor fit indices and low factor loadings, indicating questionable validity and internal consistency.

Conclusion: Although the present evidence is restricted to psychological interventions for post-traumatic stress disorder, it demonstrates that the way risk of bias has been measured in this context may not be adequate. More broadly, the results suggest the importance of testing the risk of bias tool, and the possibility of rethinking the methods used to assess risk of bias in systematic reviews and meta-analyses.

Keywords

psychotherapy, post-traumatic stress disorder, psychometrics, scale evaluation

Received 23 February 2018; Accepted 4 May 2018

Introduction

According to the Diagnostic and Statistical Manual of Mental Disorders, fifth Edition,[1] post-traumatic stress disorder (PTSD) is a condition characterized by the development of specific symptoms after exposure to one or more traumatic events, including violence. Violence is not an unusual experience worldwide; over two-thirds of individuals report an experience in their lifetime.[2] Due to the heterogeneity of the world's population, it is important to note that exposure to traumatic events is not entirely random, and it depends in part on country of residence, sociodemographic characteristics, and history of prior exposure.[2]

The symptoms of PTSD usually appear in the first three months after the trauma; their duration varies widely, but generally, this is a chronic disease. The clinical picture includes a reliving of the traumatic situation in dreams; flashbacks; strong and persistent negative expectations; aggressive, imprudent, and self-destructive behavior; and difficulty concentrating and sleeping.[1] Many therapeutic interventions are available for PTSD in adults. The first-line treatments recommended by most guidelines are cognitive behavioral therapy (CBT) and selective serotonin reuptake inhibitors.[3] Other interventions recommended for PTSD are eye movement

[1]Department of Psychiatry, Universidade Federal de Sao Paulo, Sao Paulo, Brazil
[2]Department of Pharmacology and Toxicology, University of Toronto, Toronto, Ontario, Canada
[3]Hurvitz Brain Sciences Program, Sunnybrook Research Institute, Toronto, Ontario, Canada

Corresponding author:
Hugo Cogo-Moreira, Department of Psychiatry, Universidade Federal de Sao Paulo, Rua Borges Lagoa, 570, First Floor, Sao Paulo, 04038-000, Brazil.
Email: hugocogobr@gmail.com

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (http://www.creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).

desensitization and reprocessing (EMDR), group therapy, and psychodynamic therapy.[4] There are systematic reviews of clinical trials for some of these non-pharmacological interventions.[5,6]

Systematic reviews of clinical trials for PTSD have used several criteria to assess risk of bias (RoB). RoB refers to the possibility of under- or overestimating the actual effect of the intervention, leading to an unsupported conclusion. In chapter 8 of the *Cochrane Handbook for Systematic Reviews of Interventions*,[7] eight items are listed that aim to assess RoB associated with studies included in systematic reviews.

Cochrane's eight indicators fall into five categories. The first is *selection bias,* which includes indicators related to the process used for selection of study participants (*random sequence generation* and *allocation concealment*). The second category, *performance bias*, measures unequal exposure of participants to factors other than the intervention. The items included in this category are *blinding of participants and personnel*, and *other potential threats to validity*. The third category, *detection bias*, evaluates care taken when determining outcomes (*blinding of outcome assessment* and *other potential threats to validity)*. The fourth category is *attrition bias*, and it refers to differences between groups in the number of withdrawals from a study; its single item is *incomplete outcome data*. The fifth and final category, *reporting bias*, refers to differences between reported and unreported findings, or *selective outcome reporting*. Since the review used as a sample in the present study was conducted before the latest update to the Handbook, the aforementioned items differed slightly in some cases.

Other tools have been proposed to evaluate the risk of bias, such the Jadad Scale,[8] which assesses randomization, double blinding, withdrawals, and dropout. Another instrument is the Physiotherapy Evidence Database (PEDro) scale, an 11-item scale created for rating randomized clinical trials in PEDro.[9]

Cochrane's items are observable indicators, and RoB can be conceptualized as a latent variable underlying them, that is, reflective indicator scale (for more details about reflective and formative models, see Bollen[10]). The research areas of psychology and psychiatry inevitably work with latent phenomena (depression, intelligence, mental health, etc.). This concept of latent variables, and approaches used to measure them, can be extended to phenomena not necessarily underlying human behavior, for example, in the representation of RoB developed by Cochrane.

In the same way that a set of symptoms is used to describe a given disease or pathology, and that those symptoms are used as observable indicators to evaluate something not directly observable (a latent trait), we can suppose that a latent RoB construct underlies the items on an RoB instrument. Up to now, only two studies have been conducted to evaluate the fit and reliability of RoB tools. One relates to measuring RoB in studies of attention deficit hyperactivity disorder (ADHD[11]) and the other in studies of autism spectrum disorders (ASD[12]), which found a good fit for the model with the underlying measurement theory, but poor reliability of the individual items.

In the present study, the focus was on evaluating the Cochrane RoB tool measurement model through analysis of a systematic review containing 70 studies on psychological therapies used for treatment of PTSD in adults. The aim was to provide evidence of the construct validity of this tool, which is widely adopted by Cochrane. This tool is not only used in relation to the studies on PTSD interventions but also in other diseases across different medical disciplines, and the investigation of its measurement features across specific contexts is therefore of fundamental interest.

Methods

Sample

The sample consisted of 70 studies included in Cochrane's systematic review "Psychological Therapies for Chronic Post-Traumatic Stress Disorder (PTSD) in Adults,"[5] which was first published in 2005, then updated in 2007, and later in 2013. This review compiles controlled randomized studies of psychological therapies for adults (age 18 years or older) with PTSD performed between 1989 and 2013, with a total of 4761 participants. Relevant randomized controlled trials were selected from The Cochrane Library (all years), MEDLINE (1950 to date), EMBASE (1974 to date), and PsycINFO (1967 to date).

The studies included in the review covered various interventions, such as trauma-focused CBT (TFCBT), EMDR, non-trauma-focused CBT (non-TFCBT), other therapies (supportive therapy, non-directive counseling, psychodynamic therapy, and present-centered therapy), group TFCBT, and group non-TFCBT. The authors of the review also hand searched the *Journal of Traumatic Stress*, contacted experts in the field, searched the bibliographies of included studies, and performed citation searches of identified articles. The heterogeneity of the included studies of these interventions is fundamental, providing variability in the RoB across the different natures of the interventions.

Selection Criteria

To perform the systematic review that provided the sample for this study, a previous version of Cochrane Handbook published in 2011[13] was used. Thus, Bisson

et al.[5] rated each study on the following seven indicators of RoB:

a. random sequence generation (item 1)
b. allocation concealment (item 2)
c. incomplete outcome data (item 3)
d. selective reporting (item 4)
e. other bias (item 5)
f. blinding of participants and personnel (item 6)
g. blinding of outcome assessment (item 7)

To better access all the available information from the included studies, Bisson et al.[5] contacted the authors of the clinical trials to obtain missing data. Later, two review authors independently performed RoB assessments, tabulating the results on a Likert-type scale (choosing one of the three response categories for each item: low, unclear, and high risk of bias).

Data Analysis

Confirmatory factor analysis (CFA) was used to evaluate the construct validity of Cochrane's RoB tools. To perform the statistical analysis, we used Mplus version 8 software[14] under a Bayesian estimator. The default priors on each loading and threshold (i.e. RoB indicators were considered as ordered-categorical variables (low risk, unclear, and high risk of bias)) is a normal distribution, with 0 mean and variance 5. To evaluate model adjustment, the criteria used to indicate a satisfactory fit were (a) a posterior predictive p-value (PPP) which ranges, as the regular p value, from 0 to 1, but the closest to 0.5, the better and (b) its related Bayesian Posterior Predictive Checking (PPC) 95% confidence interval (CI) for the difference between the observed and the replicated χ^2 values where the lower limit of the band is negative, and zero falls close to the middle of the interval.[15]

We also used McDonald's omega, a parameter that carries less risk of overestimation or underestimation of reliability and has also been shown by many researchers to be a more sensible index of internal consistency when compared, for example, to alpha.[16] Coefficient alpha, also called Cronbach's alpha, is the most common means of assessing internal consistency in the social sciences, but as Cronbach himself concluded, it is not appropriate for scales where questions are designed to target different areas or processes.[17]

Results

The CFA model with seven indicators returned poor fit indices, PPC 95% CI for $\chi^2 = (-8.451, 30.876)$, PPP = 0.272. Regarding the magnitude of the factor loadings, four of the seven items had factor loadings below or close to 0.4 (items 1, 2, 4, and 6); therefore, the majority of the factor loadings indicated that the items were not reliable (i.e. low R^2) indicators, offering poor convergent validity (see Table 1). Figure 1 shows the unidimensional model with the seven indicators and an underlying RoB latent variable.

When the a priori model does not fit the data, this method allows modification of the model and retesting with the same data.[18] Aiming to obtain a model that would make theoretical sense, be reasonably

Table 1. Cochrane's items and corresponding values of R^2 and its confidence interval under Bayesian estimator.

			95% credibility interval	
Items	Estimate	SD	Lower 2.5%	Upper 2.5%
Item 1	0.128	0.062	0.043	0.278
Item 2	0.165	0.175	0.001	0.638
Item 3	0.228	0.164	0.006	0.603
Item 4	0.058	0.120	0.000	0.433
Item 5	0.367	0.196	0.011	0.736
Item 6	0.071	0.162	0.000	0.589
Item 7	0.582	0.192	0.123	0.861

SD: posterior standard deviation.

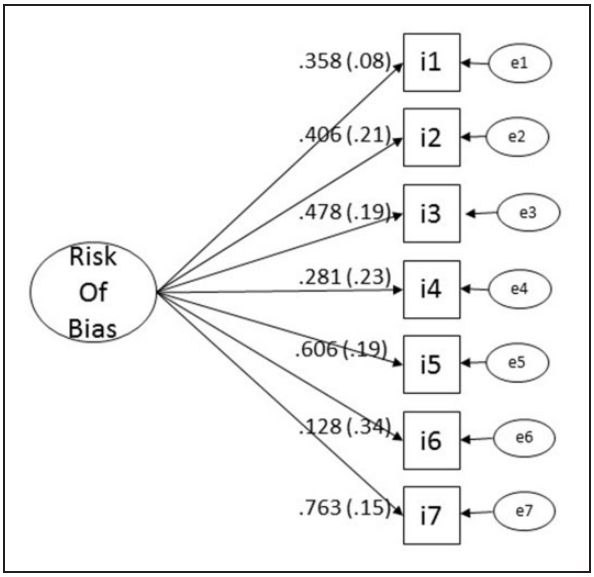

Figure 1. A conceptual model for the associations between risk of bias and Cochrane's items, with factor loadings and posterior standard deviations in parentheses. Note that risk of bias is the latent factor (represented by an oval) underlying the seven observed indictors (represented by squares). Residual variances are represented by circles with labels "e".

parsimonious, and show an acceptably close correspondence to the data,[19] we re-ran the model with modifications. The first modified model was created by excluding the item with the lowest factor loading (item 6 with 0.128); the fit indices did not improve, PPC 95% CI for $\chi^2 = (-18.451, 30.876)$, PPP = 0.294. Then, we excluded the item with the next lowest factor loading (item 4 with 0.281), and once again, the fit indices showed poor adjustment, PPC 95% CI for $\chi^2 = (-10.104, 28.191)$, PPP = 0.203. After this, we excluded item 1 (factor loading = 0.358) and the fit indices, although better than the previous figures, still indicated poor adjustment, PPC 95% CI for $\chi^2 = (-14.426, 17.745)$, PPP = 0.322. We did not continue further in the exclusion of items because at least four items are necessary in a unidimensional model to produce an over-identified model (i.e. a testable model).[10]

As shown in Figure 1, the item with the strongest relation to the latent factor (RoB) was item 7 (blinding of outcome assessment), with the highest factor loading, 0.763. Items 1 (random sequence generation), 4 (selective reporting), and 6 (blinding of participants and personnel) showed the weakest associations with the latent factor, with unsatisfactory factor loading values equal to 0.358, 0.281, and 0.128, respectively. Finally, the omega total value was 0.618.

Alternatively, we considered a restructuring of the model, preserving the seven original items. We tested a specification including residual covariances (i.e. items 1 and 2 both related to selection bias and also items 3 and 4 both related to performance bias). Figure 2 shows the unidimensional model with two additional residual covariances.

Although PPP increased (PPC 95% CI for $\chi^2 = (-21.241, 30.370)$, PPP = 0.396), a model cannot be retained based solely on values of fit statistics; the residuals, such as standardized, normalized, correlation, or covariance residuals, must also be considered. Figure 2 shows that the factor loadings for almost half of the items are lower than 0.4. Table 2a and b shows the correlation and covariances between the items, respectively. The former showed the majority of correlations are less than moderate.

Discussion

The RoB tool applied in our context of psychological interventions for PTSD in adults did not return either good fit indices or reliable measures in relation to Cochrane's items; consequently, the way that risk of bias has been measured in this context may not be reliable. Although fit indices (i.e. PPP) improved after the addition of residual correlations, it is a poor practice to decide on whether to retain a model based solely on values of fit statistics because poor model fit at the level of the residuals is not always detected by global fit statistics.[19,20]

The validity of a model is not measured only by the reliability of its indicators considering the factor loadings associated with them. Another way to evaluate reliability is by analyzing the indicators' residual terms, which represent the variance unexplained by the factor that the

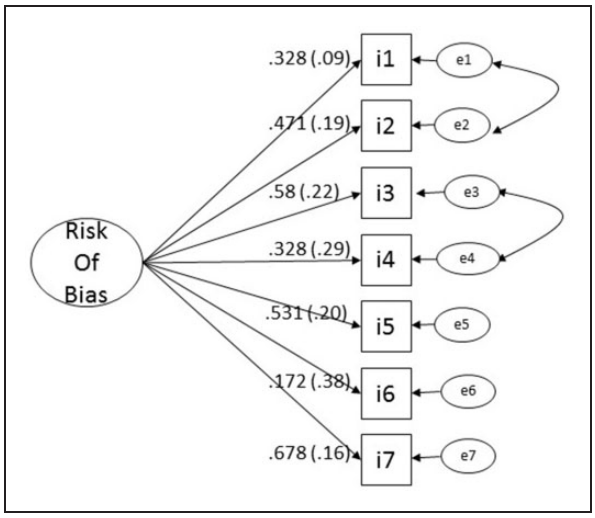

Figure 2. Alternative specification of the risk of bias model with two residual covariances added. Note that risk of bias is the latent factor (represented by an oval) underlying the seven observed indictors (represented by squares). Residual variances are represented by circles with labels "e." Residual variances are the circles with "e". Double-headed arrows indicate residual covariances.

Table 2. The correlation (with "1" in the diagonal) and covariance between the seven risk of bias items.

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7
(a) Correlation							
Item 1	1.000						
Item 2	0.770	1.000					
Item 3	0.197	0.275	1.000				
Item 4	0.105	0.147	0.080	1.000			
Item 5	0.179	0.250	0.329	0.176	1.000		
Item 6	0.060	0.083	0.110	0.058	0.099	1.000	
Item 7	0.225	0.314	0.413	0.220	0.375	0.125	1.000
(b) Covariance							
Item 1	1.120						
Item 2	0.916	1.264					
Item 3	0.261	0.387	1.568				
Item 4	0.118	0.174	0.106	1.115			
Item 5	0.226	0.335	0.492	0.221	1.426		
Item 6	0.064	0.095	0.139	0.063	0.121	1.034	
Item 7	0.327	0.485	0.712	0.320	0.616	0.175	1.892

corresponding indicator is supposed to measure.[19] From their R^2 (factor loadings squared) and residual variance values (Table 1), we can see that the only RoB item with a factor loading higher than the corresponding residual variance was item 7 (58.2% of the variance was related to the latent factor), and that item exhibited the strongest correlation with RoB (Figure 1). Although most of the variance in this item was related to the latent construct, this was still only just over 50%. Furthermore, items 3 and 5, which according to the values shown in Figure 1 could be considered adequate as components of the model, were demonstrated by this analysis not to represent the latent factor adequately (only 22.8% and 36.7% of their variance, respectively, was related to the latent factor).

In the present context, we conclude that the Cochrane RoB model exhibited a limitation related to a lack of convergent validity of its items. Although this conclusion does not preclude the possibility of using this tool in other settings, or its practical utility in this context, the development of alternative tools that could offer this type of validity might also be considered. It is interesting to note that the results of this study agree with recently published findings. Rodrigues-Tartari et al.[11] used CFA to assess RoB in randomized controlled trials of methylphenidate for children and adolescents with ADHD, finding that the majority of the items were not reliable because they exhibited low factor loadings and high values of residual variance. Similarly, Okuda et al.,[12] when evaluating the nine-item Cochrane model as applied to controlled trials for ASD, found that most items were associated with more residual variance than common variance. In both of those analyses, the measurement model returned excellent fit indices as measured by frequentist CFA. Taken together, the evidence identifies a theoretical limitation of the RoB tools and the possibility of rethinking these methods.

Some limitations need to be considered. Here, we use a reflective model to test explicitly whether the items informed the underlying latent construct, which is not the only way to specify a measurement model. Reflective models assume that the latent variable is a common or unique factor[21(p423)]. Alternatively, a formative model specification might have been used, wherein a composite variable is modeled as a weighted sum of the item scores; however, some authors describe formative models as *hard to identify*[22] because indicators are exogenous. This means that their variances and covariances are not explained, which makes it more difficult to assess the validity of a set of indicators. Second, we note imprecision in the credibility interval estimates of the factor loadings (for instance, item 2 showed a credibility interval for R^2 ranging from 0.001 to 0.638); however, if the RoB items had been more closely related to their underlying factors, more precise estimates would have been possible.[19(pp9–10)] The poor factor loadings and their relatively high imprecision have important implications for how authors conducting a systematic review might view the precision of their RoB evaluations. The small number of primary randomized clinical trials included in this study might also have contributed to imprecision[23]; however, large sample sizes are not often available in systematic reviews. Therefore, it might be useful to redefine some RoB indicators with the intention that they should be more strongly related to their intended underlying domains in order to reduce uncertainty in how RoB is being measured.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was funded by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, grant no. 2016/19287-6) and CAPES Thesis award AUXPE: 0374/2016 Process: 23038.009191/2013-76.

ORCID iD

Walter Swardfager http://orcid.org/0000-0002-0030-8908

References

1. American Psychological Association. *Manual diagnóstico e Estatístico de Transtornos Mentais: DSM-5*. Porto Alegre, Brazil: Artmed Editora, 2014.
2. Benjet C, Bromet E, Karam E, et al. The epidemiology of traumatic event exposure worldwide: results from the World Mental Health Survey Consortium. *Psychol Med*. 2016; 46(2): 327–343.
3. Bernik M, Laranjeiras M, Corregiari F. Tratamento farmacológico do transtorno de estresse pós-traumático [Pharmacological treatment of posttraumatic stress disorder]. *Rev Bras Psiquiatr*. 2003; 25: 46–50.
4. Foa EB, Keane TM, Friedman MJ, Cohen JA. *Effective Treatments for PTSD: Practice Guidelines From the International Society for Traumatic Stress Studies*, 2nd ed. New York, NY: Guilford Press, 2008.
5. Bisson JI, Roberts NP, Andrew M, Cooper R, Lewis C. Psychological therapies for chronic post-traumatic stress disorder (PTSD) in adults. *Cochrane Database Syst Rev*. 2013; 12: CD003388.
6. Lawrence S, De Silva M, Henley R. Sports and games for post-traumatic stress disorder (PTSD). *Cochrane Database Syst Rev* 2010; 20: CD007171.
7. Higgins JPT, Altman DG. Assessing risk of bias in included studies. In Higgins JPT, Green S, eds. *Cochrane Handbook for Systematic Reviews of Interventions* (Version 5.1.0.). The Cochrane Collaboration. www.training.cochrane.org/handbook. Updated June 2017. Accessed May 11, 2018.

8. Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials. 1996* 17(1): 1–12.
9. Maher CG, Sherrington C, Herbert RD, Moseley AM, Elkins M. Reliability of the PEDro scale for rating quality of randomized controlled trials. *Phys Ther*. 2003; 83(8): 713–721.
10. Bollen KA. *Structural Equations With Latent Variables*. New York, NY: John Wiley & Sons, 1989.
11. Rodrigues-Tartari R, Swardfager W, Salum GA, Rohde LA, Cogo-Moreira H. (2018). Assessing risk of bias in randomized controlled trials of methylphenidate for children and adolescents with attention deficit hyperactivity disorder (ADHD). *Int J Meth Psychiatr Res*. doi: 10.1002/mpr.1586.
12. Okuda PMM, Klaiman C, Bradshaw J, Reid M, Cogo-Moreira H. Assessing risk of bias in randomized controlled trials for Autism Spectrum Disorder (ASD). *Front Psychiatry* 2017; 8: 265.
13. Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. http://handbook.cochrane.org.
14. Mplus Version 8 [Computer Software]. Los Angeles, CA: Muthén & Muthén.
15. Muthén B, Asparouhov T. Bayesian structural equation modeling: a more flexible representation of substantive theory. *Psychol Meth*. 2012; 17(3): 313.
16. Dunn TJ, Baguley T, Brunsden V. From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *Br J Psychol*. 2014; 105(3): 399–412.
17. Cronbach LJ, Shavelson RJ. My current thoughts on coefficient alpha and successor procedures. *Educ Psychol Meas*. 2004; 64(3): 391–418.
18. Jöreskog KG. Testing structural equation models. *Sage Focus Editions* 1993; 154: 294–294.
19. Kline RB. Principles and Practice of Structural Equation Modeling, 3rd ed. New York, NY: The Guilford Press, 2015.
20. Hoyle RH, Isherwood JC. Reporting results from structural equation modeling analyses in Archives of Scientific Psychology. *Arch Sci Psychol*. 2013; 1(1): 14–22. doi:10.1037/arc0000004sion 5.1.0.
21. Bentler PM. Multivariate analysis with latent variables: Causal modeling. *Annu Rev Psychol*. 1980; 31: 419–456.
22. Treiblmaier H, Bentler PM, Mair P. Formative constructs implemented via common factors. *Struct Equ Model Multidiscip J*. 2011; 18(1): 1–17.
23. Marsh HW, Hau K-T, Balla JR, Grayson D. Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavior Res*. 1998; 33(2): 181–220. doi:10.1207/s15327906mbr3302_1.