

Technology-assisted Risk of Bias Assessment in RCTs using RobotReviewer: An Evaluation Study

Julian Hirt^{1,2}, Jasmin Meichlinger¹, Petra Schumacher³, Gerhard Mueller³

¹ Institute for Applied Nursing Science, University for Applied Sciences FHS St.Gallen, Switzerland

² Institute for Health and Nursing Science, Medical Faculty, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany

³ Institute for Nursing Science, Department for Nursing Science and Gerontology, UMIT – Health & Life Sciences University, Hall in Tyrol, Austria

Background and aim

Online tools may support researchers in conducting systematic reviews [1]. RobotReviewer enables an automated risk of bias assessment in RCTs using machine learning [2]. An evaluation study [3] including 1180 trials from different health topics, demonstrated a Cohen's kappa agreement between .10 and .48 between RobotReviewer and human assessment in different risk of bias domains. Furthermore, the analysis yielded a sensitivity between .28 and .76 and a specificity between .72 and .90 for detecting a low risk of bias in different domains. Since the sample of trials was not nursing-specific, it is unclear whether risk of bias can be assessed with the RobotReviewer in nursing-related RCTs. Therefore, the aim of this study was to evaluate the reliability of RobotReviewer's risk of bias assessment in nursing-related RCTs.

Method

Research design: Evaluation study

- Index test: Risk of bias assessment of nursing-related RCTs via RobotReviewer
- Reference test: Risk of bias assessment of nursing-related RCTs reported in Cochrane reviews

Literature search process

Cochrane reviews with nurs* in title were identified in MEDLINE via PubMed on August 30th, 2018 (Figure 1)

Inclusion criteria

Electronical availability of full text of RCTs and English language of RCTs

Data extraction and assessment

Two independent research teams

Results

The selection process yielded 190 RCTs published between 1959 and 2016 in 23 Cochrane reviews published between 2000 and 2018 (Figure 1). Missing assessments of risk of bias domains in Cochrane reviews or RobotReviewer yielded varying sample sizes per risk of bias domain. Cohen's Kappa were moderate for randomization (.52), allocation concealment (.60), and for blinding of personal/patients (.43). Blinding of outcome assessors had only slight agreement (.04). Percentage observed agreement was in all domains $\geq .50$. Sensitivity ranged from .44 to .88 and specificity from .48 to .95. Positive predictive value was highest for allocation concealment (.79) and lowest for blinding assessors (.25) (Tables 1 and 2).

Figure 1: Flowchart of literature search process

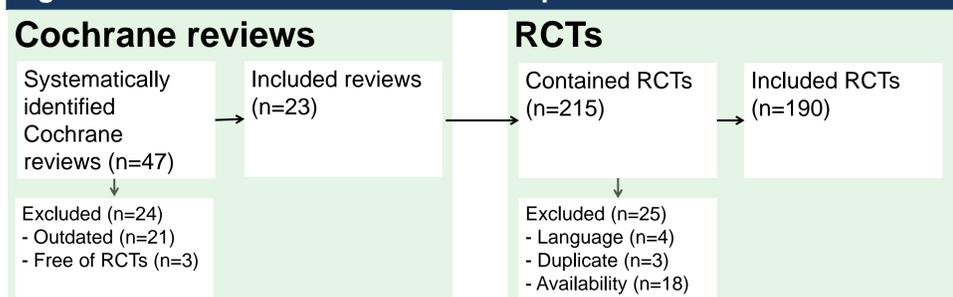


Table 1: Cohens Kappa, percentage observed agreement, and Phi coefficient of RobotReviewer's and human's assessment

Domain (n)	Cohens's Kappa (CI 95 %)	Percentage observed agreement	Phi
Randomisation (129)	.52 (.36;.68)	.78	.53
Allocation concealment (180)	.60 (.48;.72)	.80	.60
Blinding of personal/patients (109)	.43 (.14;.72)	.87	.43
Blinding of outcome assessors (115)	.04 (-1.14;.22)	.50	.05

Table 2: Sensitivity, specificity, and predictive values of RobotReviewer's and human's low risk of bias assessment

Domain (n)	Sn (CI 95 %)	Sp (CI 95 %)	Pos. PV (CI 95 %)	Neg. PV (CI 95 %)
Randomisation (129)	.88 (.81;.95)	.62 (.48;.75)	.77 (.69;.86)	.78 (.69;.87)
Allocation concealment (180)	.77 (.68;.86)	.82 (.75;.90)	.79 (.7;.88)	.81 (.72;.89)
Blinding of personal/patients (109)	.44 (.19;.68)	.95 (.90;.99)	.58 (.3;.86)	.91 (.74;1.07)
Blinding of outcome assessors (115)	.58 (.39;.77)	.48 (.38;.59)	0.25 (1.4;.35)	.80 (.7;.9)

Notes: Neg. PV=Negative predictive value; Pos. PV=Positive predictive value; Sn=Sensitivity; Sp=Specificity.

Discussion

This first evaluation study of RobotReviewer's performance in nursing-related RCTs yielded moderate agreement with human's assessment in Cochrane reviews for randomization and allocation concealment as well as an adequate sensitivity for detecting low risk of selection bias. The results can only be compared with limitations to an earlier evaluation in RCTs from different health topics [3]. Further evaluations should be based on a priori sample size calculations to ensure more precise results. RobotReviewer's performance might be better with good reported RCTs and the use of consistent wording.

Conclusions

The use of RobotReviewer for the risk of bias assessment in nursing-related RCTs can be supportive, but should be supervised by human assessment.

References

- [1] Kohl, C., McIntosh, E. J., Unger, S., Haddaway, N. R., Kecke, S., Schiemann, J., & Wilhelm, R. (2018). Online tools supporting the conduct and reporting of systematic reviews and systematic maps: A case study on CADIMA and review of existing tools. *Environmental Evidence*, 7(1), 2420. <https://doi.org/10.1186/s13750-018-0115-5>.
- [2] Marshall, I. J., Kuiper, J., & Wallace, B. C. (2016). Robotreviewer: evaluation of a system for automatically assessing bias in clinical trials. *Journal of the American Medical Association*, 316(1), 193–201. <https://doi.org/10.1093/jama/ocv044>.
- [3] Gates, A., Vandermeer, B., & Hartling, L. (2018). Technology-assisted risk of bias assessment in systematic reviews: a prospective cross-sectional evaluation of the RobotReviewer machine learning tool. *Journal of Clinical Epidemiology*, 96, 54–62. <https://doi.org/10.1016/j.jclinepi.2017.12.015>.

Poster download

Follow the link <https://bit.ly/2CnaTZI> or scan the QR code to download this poster.

