SWAR 38: Training and Experience in Study Selection (TESS): A pilot randomised trial within a systematic review

Objective of this SWAR
The aim of this Study Within a Review (SWAR) [1] is to answer the research question: Does training and level of experience within the screening pair affect the reliability of decisions made by novice screeners during study selection for a systematic review?

There are three objectives:
(a) To examine whether experience level of the screening partner and the provision of training affects the reliability of screening decisions made by novice, student screeners, compared to the consensus-based decisions made by screeners with content and methodological expertise.
(b) To examine whether pairing a novice, student screener with an experienced screener (versus another novice screener), alongside the provision of task-specific training (versus minimal guidance), improves reliability estimates within the screening pair during the study selection process.
(c) To explore the feasibility and acceptability of training materials for study selection by novice, student screeners.

Study area: Study Identification
Sample type: Review Authors, Novice screeners
Estimated funding level needed: Low

Background
Systematic reviews require time and significant human resources [2]. Specifically, study selection is a resource-intensive step requiring two independent screeners, as per best practice guidelines. By accurately removing records at the earliest screening stages, study selection improves efficiency in the review processes and is integral to identifying studies to include in the systematic review [3]. Researchers with content and methodological expertise may not be available to complete vast amounts of screening within defined project periods, which likely serves as a barrier to timely completion of the review. Thus, the composition of review teams may include colleagues with skills and experience in systematic reviews [4], as well as novice screeners such as student supervisees (for example: [5]).

Study selection methods are relatively under investigated compared to other steps in the systematic review process, such as data extraction and study appraisal [6]. They are typically oriented to guidelines on the screening procedure itself (e.g., single screening versus double screening; screening titles first versus concurrent title and abstract screening). Within this, practical guidelines on the composition of the team that will undertake screening, such as the experience level within the screening pair, have received little attention [7, 8]. Further, we are not aware of any studies on the role of training to improve study selection outcomes among novice screeners, although training has been assessed for other steps such as study appraisal (for example: [9, 10]).

Design, Participants, and Materials
This SWAR will use a 2x2 factorial randomised design to examine the role of training and level of experience within the screening pair on the reliability of decisions made by novice, student screeners at title/abstract screening. Participants will remain blinded to allocation and will be asked not to discuss the training they receive with others to minimise the risk of contamination.

Eligibility criteria for participation include: (a) 18+ years; (b) student in higher education; (c) no prior training or experience in the conduct of evidence synthesis; and (d) access to a laptop or computer with reliable internet.

Study selection will use Covidence systematic review screening software. Each participant will have a separate review page, set to 'dual screener' mode, meaning that each study record requires a decision (yes, no, maybe) by two independent screeners. A convenience sample of records at title/abstract will be loaded on to the respective review page and will be consistent

across all SWAR groups. The participant will be ‘Screener 2’, while ‘Screener 1’ will be a member of the review team, either a novice screener from the research team (research assistant) or an experienced screener with moderate methodological expertise in systematic review, in accordance with the randomisation to a SWAR group. In advance, members of the research team will independently complete screening for the sample of records as 'Screener 1'. Their screening decisions will be loaded to the respective review page on Covidence, before the participant commences (as 'Screener 2'). This ensures standardisation within the SWAR groups in terms of potential conflict identification and post-conflict adjustment, in response to the decisions made by 'Screener 1'. For each record screened, the participant will be notified by Covidence as to whether they are in agreement or in conflict with 'Screener 1'. The participant will be blinded to the characteristics of 'Screener 1' (i.e. they will not be aware of their experience level).

Based on calculations using the kappaSize package and PowerBinary function on r, 219 records will need to be screened by each participant at the title/abstract level to achieve 80% power to detect a significant difference in reliability (relative to the expert standard) between the respective SWAR groups. The expert standard is determined by the consensus-based decisions made by two reviewers on the team with content and methodological expertise. The sample of records will be taken from an on-going systematic review and network meta-analysis (PROSPERO registration: CRD42022324367).

Interventions and Comparators
Intervention 1: (a) Training intervention: standardised outline of eligibility criteria for study selection and a comprehensive, recorded online training session, which is task-specific for the host review. The content of the training will be piloted by the review team before finalisation. The recorded online training session will provide, but is not limited to, an overview of the topic area of the host review, a detailed discussion on eligibility criteria for study selection, an introduction to the systematic review software screening interface, a practice and demonstration of screening for a sample of ~10 records at title/abstract, with clear rationale provided for the screening decisions made and, finally, best practice guidelines for study selection, as relevant to this review topic. Similar online training approaches have been implemented for novice screeners ('crowd', non-specialists) on systematic review projects (for example: [11]). The estimated completion time for the training intervention is 45 minutes; and (b) Moderately experienced screening partner [12].

Intervention 2: (a) As per the training intervention (task-specific training) described above in Intervention 1; and (b) Novice screening partner with minimal experience [12].

Intervention 3: (a) Training control: standardised outline of eligibility criteria for study selection and a minimal guidance, generic, recorded online session. The purpose of the recorded online session is to provide an overview of the topic area of the host review, a general introduction to study selection and Covidence systematic review screening software, alongside best practice guidelines for study selection. Covidence Support provides a bank of online training videos which will serve as a point of reference in developing this session. The duration of training will be matched to the task-specific training for an estimated completion time of 45 minutes; and (b) Moderately experienced screening partner [12].

Intervention 4: (a) As per the training control (minimal guidance) described above in Intervention 3; and (b) Novice screening partner with minimal experience [12].

Index Type: Study selection

Method for Allocating to Intervention or Comparator:
Randomisation

Outcome Measures
Primary: Performance outcome data are recorded in Covidence. This allows for the calculation of the primary outcome: between-group reliability (% agreement, kappa statistic), which is defined as the reliability between the decisions of novice screeners with task-specific training and an

experienced screening partner (versus the expert standard), compared to novice screeners across the other conditions (e.g., minimal guidance training and a novice screening partner, versus expert standard).

Following screening completion, participants will be asked to complete an online questionnaire to collect data on two feasibility primary outcomes: (a) efficiency (participant's self-report of the time taken to complete allocated screening); and (b) acceptability (usefulness of the training and decisions of the screening partner in the completion of the study selection task).

Secondary: Performance data which are recorded in Covidence and relevant to secondary outcomes include: (a) within-group reliability (% agreement, kappa statistic), which is defined as the reliability between novice screeners and their allocated screening partner; and (b) validity (false positives, false negatives, sensitivity, specificity).

Secondary feasibility measures include motivation/interest to participate in the research and insights on what could improve the experience of being involved in study selection. Further, following participation, participants will be asked to indicate what training condition they believe they were allocated to in order to determine the success of blinding.

Analysis Plans

Descriptive statistics will be used to summarise the data, with figures such as forest plots used to show differences in reliability between SWAR groups. Inferential tests (e.g., between-group ANOVA) will be conducted to explore if differences in outcomes are statistically significant between SWAR groups, using SPSS/r statistical software.

Possible Problems in Implementing This SWAR

As the outcome data for this SWAR relies on participants who will undertake study selection as part of the host systematic review, the following barriers or problems might be encountered: (a) slow recruitment; (b) poor engagement with training materials and the screening task; (c) participant drop-out; and (d) contamination (e.g., participants allocated to the minimal guidance training control access resources other than those provided as part of the SWAR).

References

1. Devane D, Burke NN, Treweek S, et al. Study within a review (SWAR). Journal of Evidence Based Medicine 2022;15(4):328-32. doi: 10.1111/jebm.12505

2. Borah R, Brown AW, Capers PL, Kaiser KA. Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. BMJ Open 2017;7(2):e012545. doi: 10.1136/bmjopen-2016-012545

3. Polanin JR, Pigott TD, Espelage DL, Grotpeter JK. (2019). Best practice guidelines for abstract screening large-evidence systematic reviews and meta-analyses. Research Synthesis Methods 2019;10(3):330-442. doi: 10.1002/jrsm.1354

4. Lasserson TJ, Thomas J, Higgins JPT. Chapter 1: Starting a review. In Higgins JPT, Thomas J, Chandler J, et al (editors), Cochrane handbook for systematic reviews of interventions version 6.4 (updated August 2023). Cochrane. Available from www.training.cochrane.org/handbook.

5. Tendal B, Higgins JPT, Jüni P, et al. Disagreements in meta-analyses using outcomes measured on continuous or rating scales: Observer agreement study. BMJ 2009;339:b3128. doi: 10.1136/bmj.b3128

6. Robson R, Pham B, Hwee J, et al. Few studies exist examining methods for selecting studies, abstracting data, and appraising quality in a systematic review. Journal of Clinical Epidemiology 2019;106:121-35. doi: 10.1016/j.jclinepi.2018.10.003

7. Cooper M, Ungar WJ, Zlotkin S. An assessment of inter-rater agreement of the literature filtering process in the development of evidence-based dietary guidelines. Public Health Nutrition 2006;9(4):494-500. doi: 10.1079/phn2005877

8. Ng LC, Pitt VJ, Huckvale K, et al. Title and Abstract Screening and Evaluation in Systematic Reviews (TASER): A pilot randomised controlled trial of title and abstract screening by medical students. Systematic Reviews 2014;3(1):121 doi: 10.1186/2046-4053-3-121

9. Da Costa BR, Beckett B, Diaz A, et al. Effect of standardized training on the reliability of the Cochrane risk of bias assessment tool: A prospective study. Systematic Reviews 2017;6(1):44. doi: 10.1186/s13643-017-0441-7

10. Oremus M, Oremus C, Hall GB, McKinnon MC. Inter-rater and test–retest reliability of quality assessments by novice student raters using the Jadad and Newcastle–Ottawa Scales. BMJ Open 2012;2(4):e001368. doi: 10.1136/bmjopen-2012-001368

11. Noel-Storr AH, Redmond P, Lamé G, et al. Crowdsourcing citation-screening in a mixed-studies systematic review: a feasibility study. BMC Medical Research Methodology 2021;21(1):88. doi: 10.1186/s12874-021-01271-4

12. Horton J, Vandermeer B, Hartling L, et al. Systematic review data extraction: cross-sectional study showed that experience did not increase accuracy. Journal of Clinical Epidemiology 2009;63(3):289-98. doi: 10.1016/j.jclinepi.2009.04.007

Publications or presentations of this SWAR design

Examples of the implementation of this SWAR

People to show as the source of this idea: Elayne Ahern, Aoife Whiston, Sarah Dillon, Fiona Lynn
Contact email address: elayne.ahern@ul.ie
Date of idea: 27/09/2024
Revisions made by:
Date of revisions: