The Evidence Review Accelerator Toolkit

PRESENTER: SARAH BATEUP, SENIOR RESEARCH ASSISTANT

WORKSHOP DESIGN TEAM: JUSTIN CLARK AND SARAH BATEUP

Agenda

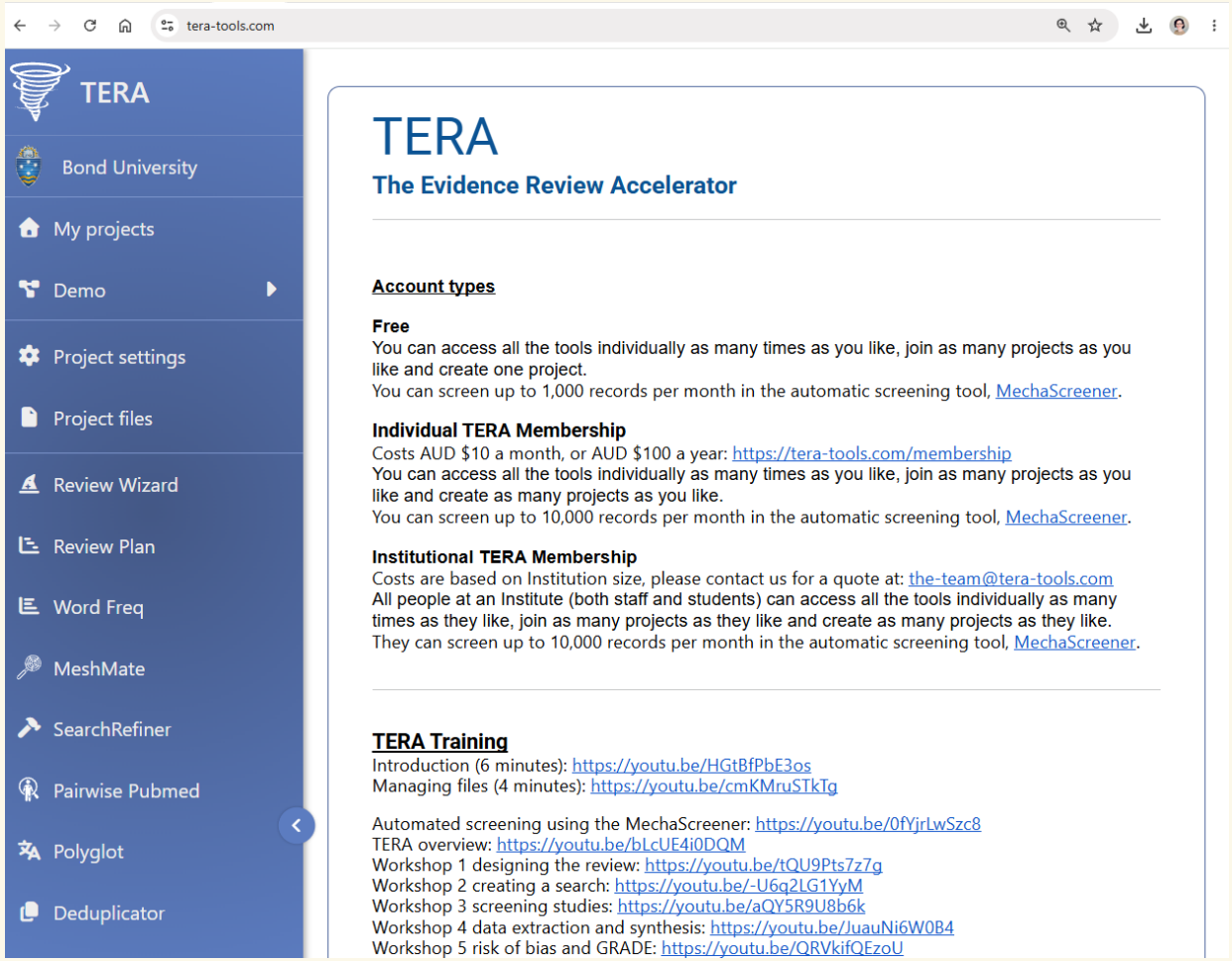
1. INTRODUCTION
2. DESIGNING AND WRITING A REVIEW USING THE REVIEW WIZARD
3. CREATING FOCUSED SEARCHES USING THE WORD FREQ, SEARCHREFINER, POLYGLOT AND THE DEDUPLICATOR
4. DEDUPLICATION STUDY*
5. REDUCING SCREENING LOAD WITH MECHASCREENER
6. CITATION SEARCHING WITH SPIDERCITE
7. SUMM-AI-RISER STUDY*

CRICOS 00017B TEQSA PRV12072

Introduction

SET UP

- Create a free personal account at **tera-tool.com**
- Create one project to access the tools.
  Optional: To invite others to the project, they need to have a TERA account.
- Open **bit.ly/teraworkshop**
  We will use the PMIDs and the database export files (RIS and NBIB) later in the session.



Creating focused searches

EXAMPLE SYSTEMATIC REVIEW

Balneotherapy for chronic venous insufficiency - 2023
https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD013085.pub3/

INCLUDED STUDIES

- 7/12 indexed in PubMed
- 3/12 trials
- 1/12 from 1992 (journal indexed in PubMed, only from 2007 onwards)



Comparing eight deduplication automation tools for speed and accuracy

PRESENTER: SARAH BATEUP

AUTHOR TEAM: SARAH BATEUP, HELEN FULBRIGHT, KLAS MOBERG, KAITLYN HAIR, EMMY PETERSON, CLAIRE STANSFIELD, RIAZ QURESHI, JUSTIN CLARK

PUBLICATION: WWW.DOI.ORG/10.1017/RSM.2026.10100

Outline

1. AIMS
2. METHODS
3. OUTCOMES
4. STRENGTHS AND LIMITATIONS

1. Aims

OUR STUDY

We compared eight deduplication tools.

We wanted to use a large dataset – we tested the tools using 26,778 records from 5 reviews.

We measured the tools' performance across the following three outcomes:

1. Singulars (unique studies) removed
2. Duplicate records retained
3. Time spent using the tool for deduplication

We wanted to help research teams choose deduplication tools that would suit their experience, timeframe and budget.

2. Methods

Selected five Cochrane reviews

Searched PubMed for recent Cochrane Revies: **(("Cochrane Database Syst Rev"[jour]) AND (2020/01/01:2022/12/31[Date - Publication]))**

Used Google random number generator to select reviews. **1469** reviews were found.

Worked through selection criteria until 5 samples were obtained.

Obtained the sample to be deduplicated

The selection criteria:

- Between 2-8 databases searched.
- Between 1000-20,000 results found at the time of the review.
- Search strings for all databases published.
- Searches were recreated and rerun using the same databases as the original review.

Created the gold standard sets

Each set was deduplicated by two authors independently, using EndNote, and sorting the records by multiple fields, e.g. title, then author etc.

Disputes were resolved using the SRA Disputatron.

We defined a singular record as an article published in a unique place, i.e. unique vol, issue, and page range within a journal.

2. Methods

Tested each of the eight tools against the gold standard sets

Each of the 5 sets of records was deduplicated using the automation tools, then manually deduplicated if required by the tool. The combined time for these two steps was recorded.

The set of deduplicated records was loaded into an Endnote Library, along with the gold standard set, and checked for errors independently by two members of the study team.

Finally, the two authors compared their results and resolved disputes through discussion. If an agreement couldn't be reached, Justin was consulted for the final decision.

Compared each tool's performance against the three stated outcomes

Logistic mixed models were fitted to compare the proportion of errors (separate models for missed duplicates and removed singulars) by the eight tools used for deduplication.

A linear mixed model was fitted to compare time taken. Time taken was log transformed prior to analysis to remove positive skew.

A random intercept was specified to account for within review correlation and least squares means estimated and compared across the eight tools.

3. Outcomes

PERFORMANCE OF INDIVIDUAL TOOLS FOR DEDUPLICATION ACROSS ALL OUTCOMES

TOTAL RECORDS WITH DUPLICATES = 26,778

Tool	Removed Singulars	Missed Duplicates	Total time (HH:MM)	Degree of automation	Cost
ASySD	22	**34**	00:53	Partial - manual checking required	**Free**
Covidence	8	276	00:32	**Full automation**	Subscription
Deduklick	11	81	00:14	**Full automation**	Subscription
EPPI-Reviewer	20	74	01:03	Partial - manual checking required	Subscription
PICO Portal	16	136	05:07	Partial - manual checking required	Subscription
Rayyan	**2**	**40**	20:34	Partial - manual checking required	Subscription
SRA (TERA) Deduplicator: Focused	9	81	02:58	Partial - manual checking required	**Free**
SRA (TERA) Deduplicator: Relaxed	11	280	**00:01**	**Full automation**	**Free**

4. Strengths and Limitations

STRENGTHS

- Randomised design.
- Large dataset.
- Same gold standard sets used for each tool comparison.
- Two team members per tool.

LIMITATIONS

- Small number of errors in the gold standard later in the study.
- Different pairs of authors for each tool.
- Reviews had well-defined study types.
- Our records came from health sciences databases only.



Exploring a new AI summarisation tool: the Summ-AI-Riser

PRESENTER: SARAH BATEUP

AUTHOR TEAM: SARAH BATEUP, PAUL GLASZIOU, TAMMY HOFFMANN AND JUSTIN CLARK

Outline

1. BACKGROUND
2. AIMS
3. METHODS
4. STRENGTHS AND LIMITATIONS

1. Background

Currently, access to SR findings is limited to audiences that can read and interpret the complex information within reviews, therefore access to the information in SRs is limited, for example:

- **In clinical care:** Due to the challenges in reading and interpreting SRs, clinicians misinterpret the direction or strength of effect in around 20% of the time.
- **In the media:** Review findings can also be difficult to interpret for people without medical training. Journalist may report on SR findings; however they increasingly rely on press releases to base their reports. Press releases can vary in quality, with lower quality releases failing to include key contextual information and study limitations.

A key strength of Summ AI Riser is its ability to generate tailored summaries for four distinct audiences: clinicians, researchers, journalists, and members of the public, each with different information needs.

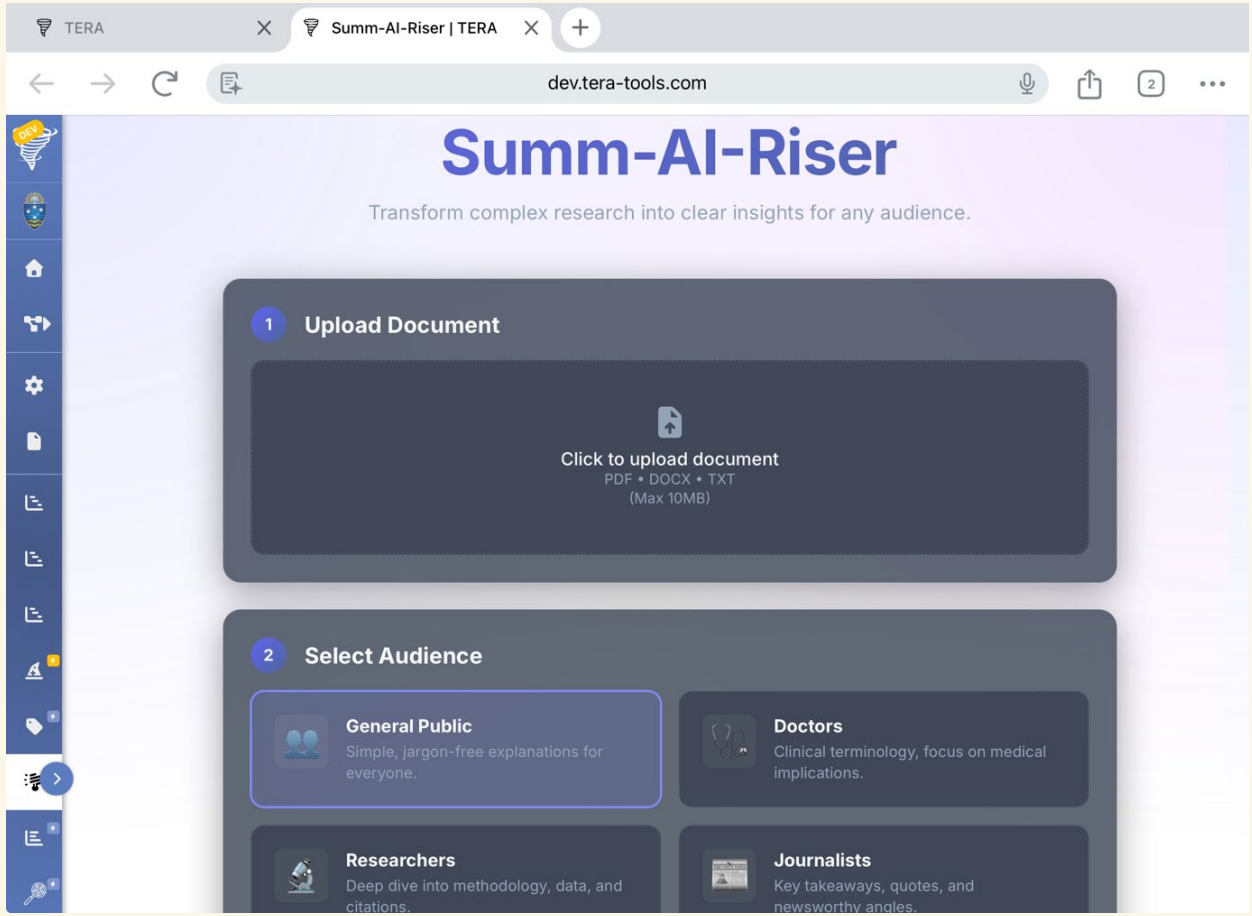
2. Aims

PRIMARY AIM

The overall aim of the two parallel trials is to assess whether the Summ-AI-Riser tool improves understanding of the results of the systematic review and accurately summarises its findings.

SECONDARY AIM

The secondary aim is to evaluate user satisfaction and readability. The findings of this study are important in determining whether the tool delivers information clearly and is enjoyable to read.



3. Methods

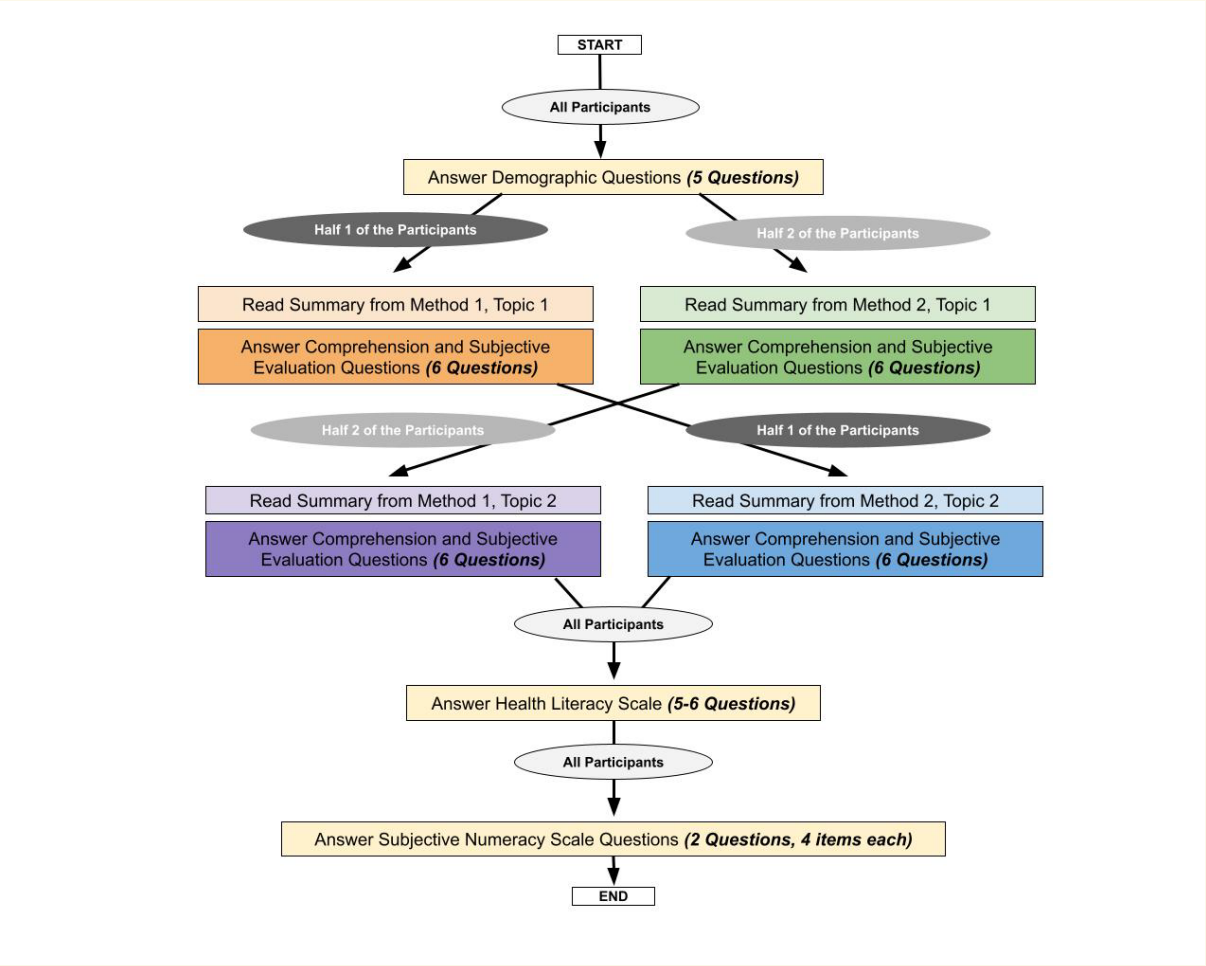
INTERVENTION

The general format for both the General Population and Journalist studies were the same.

Participants were randomly assigned 1:1 to one of two groups. Each participant read two summaries, one produced by Summ-AI-Riser and one from a tradisional source (press release or Plain Language Summary), covering two different systematic reviews (SR-A and SR-B).

- **AB group:** Summ-AI-Riser summary of SR-A, then the traditional summary for SR-B.
- **BA group:** Summ-AI-Riser summary of SR-B, then the traditional for SR-A.

After reading each summary, participants answered questions testing comprehension and readability.



3. Methods

RECRUITMENT AND SAMPLE SIZE

For the **General Population study**, sample size is calculated to be **218** participants. Recruitment took place **online** through a first-party research platform.

We had 232 participants complete the survey in September 2026.

Recruitment for the **Journalists** study took place **in person** at the *Australian Science Communicators Conference* at QUT, Kelvin Grove in September 2026. Sample size was calculated to be **138**.

We had 78 participants complete the survey, and we are currently running an additional session to recruit additional participants.



Summ-AI-Riser

tera-tools.com/summ-ai-riser



Thank you

ANY QUESTIONS?

Sarah Bateup, Institute for Evidence-Based Healthcare,
sbateup@bond.edu.au