Evaluation of A New Approach to Deduplication of Trial Registry Records in Systematic Reviews

Zabzuni E, *Anzola D, Almuallem L, Cruz F, Cooper C*
STRATENYM Inc, Toronto, Ontario, Canada

Aim

To evaluate the efficiency of the Cooper & Premji method[1] compared to the following established methods of deduplication:

1. Bramer method[2];
2. Bond deduplication tool (automated deduplication)[3];
3. A manual review of duplicates by title screening

We compared each method in terms of duplicates removed and processing time.

"Determining and implementing the most efficient deduplication method for clinical trial registries could save time, resources, and costs"

Background

Trials registries are a core resource to search in systematic literature reviews (SLRs) of intervention effect.[1,4] Cochrane, in their Handbook and Methodological Expectations of Cochrane Intervention Reviews (MECIR) guidance,[5] consider it mandatory for Cochrane reviews and this guidance maps into Health Technology Assessment (HTA) where searches of registry resources are considered best practice.

The issue here is that data in registry resources differ from format and structure to bibliographic data in databases such as MEDLINE. This makes the process of identifying and removing duplicate study records time-intensive.

Cooper & Premji have proposed a new method to deduplicate trial registry records.[1] Their method focuses on the trial registry number as the unit for deduplication, in contrast with other methods that are based on title, abstract and author fields.

Methods

The Cooper & Premji method extracts unique trial registry numbers and consolidates them into a single searchable field for deduplication. Two independent screeners applied the Cooper & Premji method, the Bramer method, the Bond deduplication tool, and manual deduplication to datasets from the International Clinical Trials Registry Platform (ICTRP) and ClinicalTrials.gov (CTG). Manual deduplication was completed by screening titles. The number of duplicates identified and processing time were used to compare performance.

Findings

Figure 1 shows the results of the deduplication procedure from the four tools evaluated.

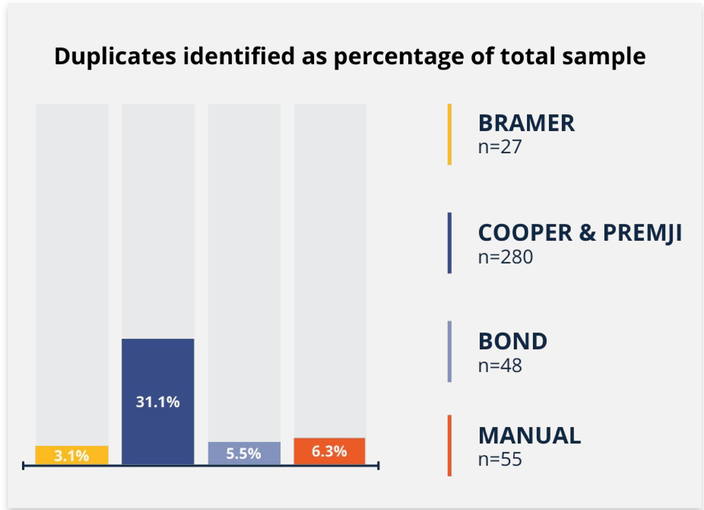

Figure 1*. Proportion of records identified as duplicates, as a percentage of the total sample (N=872) by deduplication tool*

Out of 872 records, the Cooper & Premji method identified 280 duplicates (32.1% of total records), which is 253 more than the Bramer method, 232 more than the Bond method, and 225 more than manual deduplication by title screening.

Findings

Figure 2 shows the processing time for each deduplication method.

The Cooper & Premji method was completed in 5 minutes, 15 times faster than the Bramer method (75 minutes) but 5 times slower than the automated method (1 minute). Manual deduplication was the most time-consuming method, completed in 106 minutes.

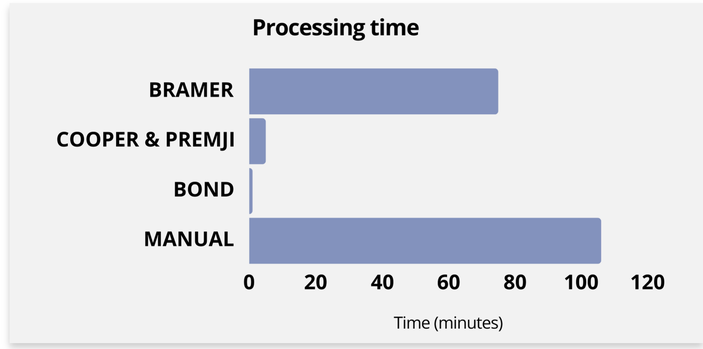

Figure 2. *Processing time (minutes) to complete each deduplication method*

Discussion

Conclusion: The Cooper & Premji method identified the greatest number of duplicates and offered substantial time savings compared with manual deduplication and the Bramer method. The Cooper & Premji method identified duplicate records that had not been flagged by methods that rely heavily on bibliographic fields (Bramer, manual deduplication), such as title, authors, and journal page numbers.

Implications for practice: Faster deduplication of registry records can meaningfully support SLRs in HTA by reducing manual effort, improving study identification accuracy, and helping streamline evidence generation—particularly in early-phase or rare disease contexts where registry data may play a key role.

Overall, The Cooper & Premji method efficiently identifies more duplicate trial registry records than other commonly used methods.

STRATENYM

References

1. Premji Z, Cooper C. *Res Synth Methods*. 2025:1-14. doi.org/10.1017/rsm.2025.20
2. Bramer W M, et al. *JMLA*, 2016;104(3):240–243
3. Forbes C, et al. *Syst Rev,* 2024;13(1):206
4. Kim J S M, et al. *Syst Rev,* 2022;11(1):206
5. Higgins J, et al. *Cochrane*; 2023.

Ersi Zabzuni
Market Access Consultant
STRATENYM Inc.
Ersi.zabzuni@stratenym.com
www.Stratenym.com
linkedin.com/company/stratenym

Scan the QR code to download the poster

Acknowledgements & Sources of Funding
This research received no specific grant from any funding agency. EZ, DA, LA, FC, and CC are funded by STRATENYM.