# Evaluating the quality of systematic reviews and meta-analysis abstracts generated by ChatGPT versus original abstracts

**Pravin Bolshete*, Madhulika Malode, Snehal Khanolkar and Sneha Badgujar**

Sqarona Medical Communications LLP, Pune, India. *Presenting author

32

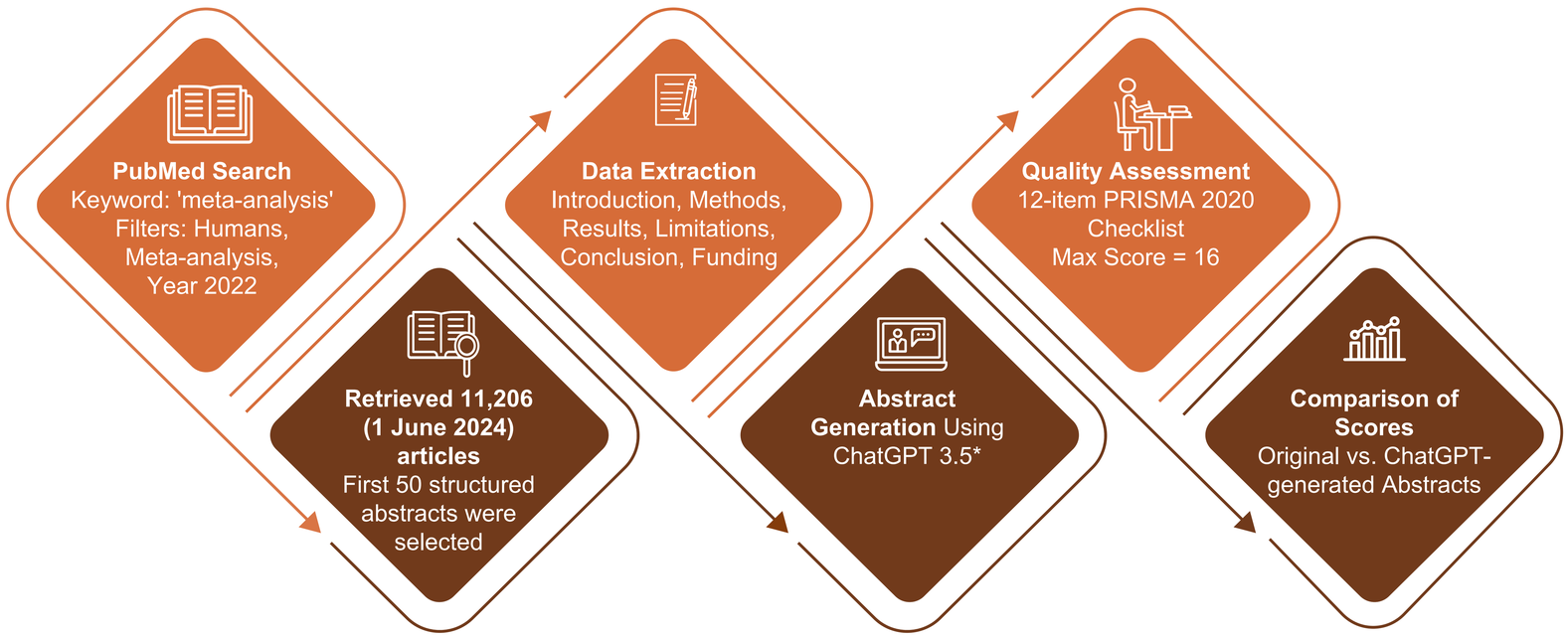
## INTRODUCTION

- Systematic reviews and meta-analyses are vital for evidence-based clinical decision-making.
- Abstracts serve as a key entry point for readers to assess the relevance and quality of a study.
- High-quality abstracts are critical, as many readers make decisions based solely on abstracts without accessing full articles.
- The PRISMA 2020 for Abstracts checklist provides a structured framework to ensure transparent and complete reporting.[1]
- With the increasing use of AI tools like ChatGPT in scientific writing, it is essential to assess if AI-generated abstracts meet established reporting standards.

## OBJECTIVE

- To evaluate the quality of abstracts generated by ChatGPT versus original abstracts, using the PRISMA 2020 checklist as a benchmark.

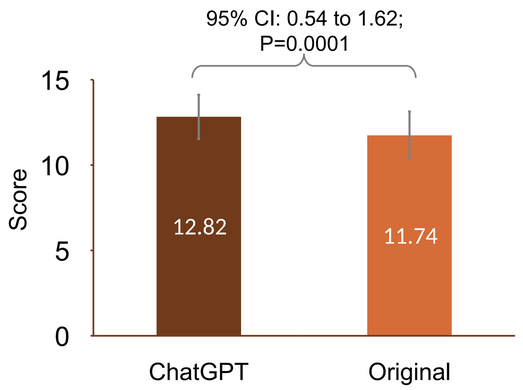
## METHODS



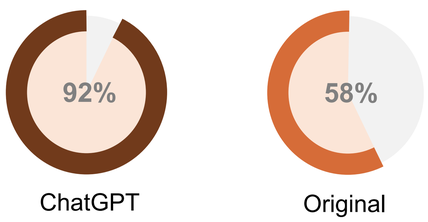

## RESULTS

- The mean PRISMA checklist score was significantly higher for ChatGPT abstracts (12.82) compared to original abstracts (11.74).
- The scores ranged from 7 to 14 for ChatGPT abstracts and 9 to 15 for original abstracts.

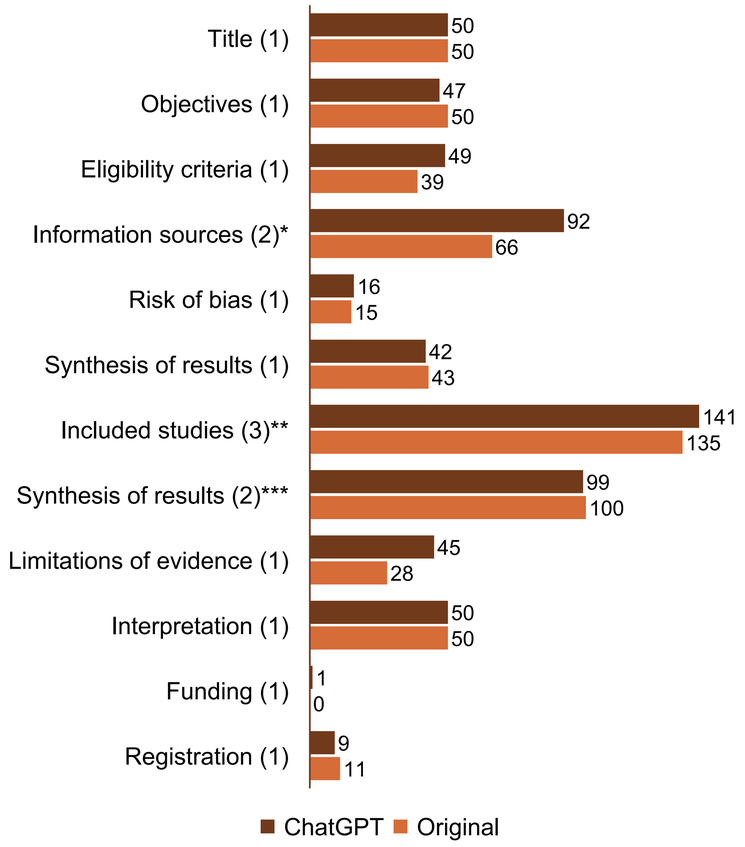
**Mean for ChatGPT vs. Original abstracts**




**Proportion of abstracts achieving ≥12 score**




**Total score for each PRISMA item**




PRISMA, preferred reporting items for systematic reviews and meta-analyses.
Numbers mentioned in the brackets represent the maximum score for the parameter. *Source and date of access; **total number of included studies, participants and summary of study characteristics; ***no. of studies and participants.

- At least one element of the “included studies” and “synthesis of results” was reported in all abstracts of both groups; however, all necessary elements of “included studies” were reported by 84% (ChatGPT) vs. 74% (original) of abstracts, elements of “synthesis of results” by 98% (ChatGPT) vs. 100% (original) of abstracts.
- “Eligibility criteria” was reported by 98% (ChatGPT) vs. 78% (original) of abstracts.
- At least one element of “information sources” was reported by 98% (ChatGPT) vs. 80% (original) of abstracts.
- “Risk of bias” was reported by 32% (ChatGPT) vs. 30% (original) of abstracts and “limitations” by 90% (ChatGPT) vs. 56% (original) of abstracts.

## LIMITATIONS

- The evaluation focused solely on the presence of PRISMA checklist items without assessing the factual accuracy or scientific validity of the ChatGPT-generated abstracts.
- Inclusion of PRISMA elements alone may not equate to the overall quality or reliability of information.
- Abstract selection was consecutive and not randomized, which may introduce selection bias.
- The sample size of 50 abstracts per group was relatively small, limiting the generalizability of the findings.
- The study used only one version of ChatGPT (3.5), and results may differ with newer models.

## CONCLUSION

- ChatGPT-generated abstracts demonstrated higher compliance with the PRISMA 2020 checklist compared to original author-generated abstracts, based on an unweighted scoring system where each checklist item contributed equally to the total score.
- However, the assessment focused on checklist adherence and did not evaluate the factual accuracy or scientific rigor of the content.

**REFERENCES**: [1] Page MJ, et al. BMJ. 2021;372:n71.
**FUNDING**: Sqarona Medical Communications LLP
**CONFLICT OF INTEREST**: All authors are associated with Sqarona Medical Communications LLP, which provides medical writing services.

Poster presented at the 21st Annual Meeting of ISMPP, May 12–14, 2025, Washington, DC, USA.