Plagiarism, quality, and correctness of ChatGPT-generated vs human-written abstract for research paper

Tien-Wei Hsu[a], Shih-Jen Tsai[b], Chih-Hung Ko[c,d,e], Trevor Thompson[f], Chih-Wei Hsu[g], Fu-Chi Yang[h], Chia-Kuang Tsai[h], Yu-Kang Tu[i,j], Szu-Nian Yang[k,l,m], Ping-Tao Tseng[n,o,p,q,*], Chih-Sung Liang[k,r,*], Kuan-Pin Su[s,t,u]

a. Department of Psychiatry, Kaohsiung Veterans General Hospital, Kaohsiung, Taiwan
b. Department of Psychiatry, Taipei Veterans General Hospital, Taipei, Taiwan
c. Department of Psychiatry, Faculty of Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan.
d. Department of Psychiatry, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan.
e. Department of Psychiatry, Kaohsiung Municipal Siaogang Hospital, Kaohsiung Medical University, Kaohsiung, Taiwan.
f. Centre for Chronic Illness and Ageing, University of Greenwich, London, UK
g. Department of Psychiatry, Kaohsiung Chang Gung Memorial Hospital
h. Department of Neurology, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan
i. Institute of Epidemiology & Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan
j. Department of Dentistry, National Taiwan University Hospital, Taipei, Taiwan
k. Department of Psychiatry, Beitou Branch, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan
l. Department of Psychiatry, Armed Forces Taoyuan General Hospital, Taoyuan, Taiwan
m. Graduate Institute of Health and Welfare Policy, National Yang Ming Chiao Tung University, Taipei, Taiwan
n. Institute of Biomedical Sciences, National Sun Yat-sen University, Kaohsiung, Taiwan
o. Department of Psychology, College of Medical and Health Science, Asia University, Taichung, Taiwan
p. Prospect Clinic for Otorhinolaryngology & Neurology, Kaohsiung, Taiwan
q. Institute of Precision Medicine, National Sun Yat-sen University, Kaohsiung, Taiwan
r. Department of Psychiatry, National Defense Medical Center, Taipei, Taiwan
s. College of Medicine, China Medical University, Taichung, Taiwan
t. Mind-Body Interface Laboratory (MBI-Lab), China Medical University and Hospital, Taichung 404, Taiwan
u. An-Nan Hospital, China Medical University, Tainan 709, Taiwan

* contributed equally to this article as corresponding author

Running title: Human-written vs ChatGPT-generated abstracts

Corresponding authors:
Chih-Sung Liang, MD
Department of Psychiatry, Beitou Branch, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan
No. 60, Xinmin Road, Beitou District, Taipei City 11243, Taiwan

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4429014

Fax: +886-2-8978919
Tel: +886-2-28959808
Email: lcsyfw@gmail.com
ORCID: 0000-0003-1138-5586

And

Ping-Tao Tseng
Prospect Clinic for Otorhinolaryngology & Neurology, Kaohsiung City, Taiwan
No. 252, Nanzixin Road, Nanzi District, Kaohsiung City 81166, Taiwan
Tel: +886-7-3524100
Email: ducktseng@gmail.com
ORCID: 0000-0001-5761-7800

Author emails:
twhsu@vghks.gov.tw
sjtsai@vghtpe.gov.tw
chihhungko@gmail.com
T.Thompson@greenwich.ac.uk
harwicacademia@gmail.com
fuji-yang@yahoo.com.tw
jiakuang@yahoo.com.tw
yukangtu@ntu.edu.tw
ysn56725@ms4.hinet.net
ducktseng@gmail.com
lcsyfw@gmail.com
cobolsu@gmail.com

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4429014

Research in context

Evidence prior to this study

The chat generative pre-trained transformer (ChatGPT), a state-of-the-art artificial intelligence chatbot built by OpenAI, was launched in November 2022 and has rippled widespread discussions in the fields of medical research and healthcare. Several studies have shown that new-generation chatbots can produce high-quality coherent texts. In medicine, AI chatbots are believed to assist with documentation, consultation, differential diagnosis, medical education, and healthcare communication. A pilot study evaluated the readability, factual correctness, and humanness of ChatGPT-generated clinical letters to patients, thereby revealing that ChatGPT-generated clinical letters had high overall readability, accuracy, and humanness. Another study used only “title” and “journal style” as prompts and requested ChatGPT to generate an abstract for a medical research paper. This study found that although the generated abstracts had clarity, they contained fabricated data and had high percentages of plagiarism and AI-generated content.

Our research question for this study is: whether ChatGPT can help write a high-quality and accurate abstract given that researchers have already written the full text of a medical research paper. The MEDLINE, Cochrane Central Register of Controlled Trials (CENTRAL), EMBASE, ClinicalTrials.gov, PsycINFO, and Google Scholar databases were systematically searched without language restrictions, starting from database inception to 12 April, 2023. Moreover, to the best of our knowledge, there are no studies that have answered this research question thus far.

Value addition provided by this study

We provided the ChatGPT model (using ChatPDF) with the full text of a published medical research paper excluding the abstract, and requested it to generate a structured or unstructured abstract with the same style as the original paper. We invited five experts to evaluate the quality of the original and generated abstracts in a blind manner using a Likert scale ranging from 0 to 10. Both ChatGPT-generated structured and unstructured abstracts had low percentages of similarity and plagiarism, whereas ChatGPT-generated unstructured abstracts contained a high percentage of AI content. For structured abstracts, the blinded reviewers rated the original and generated abstracts similarly. However, for the unstructured abstracts, the generated abstracts received lower scores than the original abstracts. The proportion of words in each subheading (importance, objective, intervention, design, interventions, outcomes, results, and conclusions) differed between the generated and original structured abstracts. Human authors wrote more words in the results

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4429014

and design subheadings, whereas ChatGPT generated more words for the remaining five subheadings. Interestingly, the accuracy of differentiating between the original and generated abstracts was only 40% for the structured abstracts and 73% for the unstructured abstracts. However, 30% of the conclusions in the ChatGPT-generated abstracts were incorrect.

Implications of the available evidence

The expert assigned a quality rating of low to the ChatGPT-generated unstructured abstract, which further improved upon structuring. Although AI algorithms demonstrated good capacity to replicate some aspects of the human process of constructing an abstract, the validity of the conclusions drawn from AI-generated abstracts may be uncertain, and the proportion of words in each subheading may differ from that in human-written abstracts.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4429014

Summary

Background: There has been growing interest in the application of ChatGPT, an artificial intelligence (AI)-based chatbot, in both medical practice and publishing. We compared the quality, plagiarism, and accuracy of the ChatGPT-generated abstracts with those of the original human-written abstracts for the same full-text research papers.

Methods:

We provided 20 full-text published papers, excluding abstracts, to the ChatGPT model (using ChatPDF), and prompted it to generate a structured or unstructured abstract with the same style as that in the original paper. We invited five experts to evaluate the quality of the original and generated abstracts in a blind approach (Likert scale: 0 to 10). The generated abstracts were evaluated on similarity, plagiarism, and AI content. Finally, five blinded reviewers identified the abstracts written by the original authors and validity of the conclusions.

Findings: We analysed 10 unstructured and structured abstracts each. The similarity between the generated and original abstracts was 16·35% (in terms of standard deviation: 9·08%). Plagiarism accounted for 7·55% of the generated abstracts. The AI content was 31·48% (20·75%) for the generated structured abstracts and 75·58% (26 ·76%) for the generated unstructured abstracts. For quality evaluation, the generated and original structured abstracts were rated similarly (-0·16; -0·89 to 0·57; p=0·67), whereas the generated unstructured abstracts had lower scores than the human-written abstracts (-2·30; -3·00 to -1·60; p<0·01). Moreover, more words were written by the original authors in the results and design subheadings, whereas in the remaining five subheadings, ChatGPT generated more words. The identification accuracy was 40% for the structured abstracts and 73% for the unstructured abstracts. However, 30% of the conclusions in the ChatGPT-generated abstracts were incorrect.

Interpretation: ChatGPT generated better-structured abstracts; however, the interpretation of the study findings may not be valid.

Funding: None

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4429014

Introduction

The title and abstract are the most important parts of a research paper, which provide readers with the first impressions of the paper [1]. Most readers only read the title and abstract, with only a small fraction of "interest one" reading the entire article[2]. The abstracts of research papers can be classified as structured or unstructured. The subheadings of structured abstracts typically include background/objective, methods, results, and discussion/conclusion, which are more elaborate, informative, and easier to read than unstructured abstracts[3]. However, unstructured abstracts provide more flexibility, which allows authors to focus on the most important aspects of their research without being constrained by a particular structure, and offer increased engagement for readers through enhanced narrative and personalised summary of the research. Regardless of the type, the abstract must be concise, precise, functional, unbiased, comprehensive, and self-sufficient to summarise the article[4].

An AI-based conversational large language model (LLM), ChatGPT, was released in November 2022. ChatGPT (Open AI, San Francisco, CA, USA) can respond to multiple languages and generate refined and highly sophisticated responses based on advanced modeling.[5] It was trained on an almost infinite amount of text data (including human language) and can interact with people in a conversational manner. Several studies have shown that new-generation chatbots can produce high-quality coherent texts. In medicine, ChatGPT is believed to assist with documentation, consultation, differential diagnosis, medical education, and healthcare communication.[6] A pilot study evaluated the readability, factual correctness, and humanness of ChatGPT-generated clinical letters to patients, thereby revealing that these letters had high overall readability, accuracy, and humanness.[7] Another study provided only "title" and "journal style," and requested the ChatGPT to generate an abstract for a medical research paper.[8] This study found that the abstracts were clearly written but contained fabricated data and had high percentages of plagiarism and AI-generated content.

Research has focused on the possibility of using artificial intelligence (AI) chatbots in medical practice and publishing. ChatGPT has shown impressive results in generating scholarly content (e.g., university essays),[9] though debate still exists regarding its use in academic writing. One potential research direction is to determine whether ChatGPT can help generate high-quality and accurate abstracts given that researchers have already written the full text of a research paper. The aim of the current study is to examine the quality of ChatGPT-generated versus human-written

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4429014

abstracts for the same full text of a research paper. We also examined the similarity, plagiarism, and correctness of the generated abstracts. We hypothesised that the generated abstracts would differ from the original author-written abstracts in terms of quality and several aspects of academic writing.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4429014

Methods

We aimed to evaluate the quality and accuracy of abstracts generated by ChatGPT. Medical abstracts can be classified as structured or unstructured; therefore, we selected 10 papers of each type. For homogeneity, we selected only randomised controlled trials in the field of psychiatry. Papers with unstructured abstracts were selected from the journals: *Molecular Psychiatry*, *Neuropsychobiology*, and *Translational Psychiatry*, whereas those with structured abstracts were selected from *JAMA Psychiatry*. The abstract structure of JAMA Psychiatry includes the following seven subheadings: (i) importance; (ii) objective; (iii) design, setting, and participants; (iv) interventions; (v) main outcomes and measures; (vi) results; and (vii) conclusions and relevance.

Generation of abstract

We used ChatPDF, (https://www.chatpdf.com/, Mathis Lichtenberger und Moritz Lage GbR, Berlin, Germany) which is based on ChatGPT 3·5, to analyse the PDF file content efficiently. ChatPDF offers users a summary and answers questions regarding PDF files without non-existent or self-created content, which is one of the major concerns of ChatGPT.[5] We provided ChatPDF with full texts of the selected papers after excluding abstracts. The prompt provided to the ChatPDF was “Please write a 250-word summary” for papers with unstructured abstracts, and “Please write a 350-word summary including 7 paragraphs, namely "importance," "objective," "design, setting, and participants," "interventions," "main outcomes and measures," "results," and "conclusions and relevance"” for papers with structured abstracts.

Similarity, plagiarism, AI-content, subheading

We evaluated similarity using Plagiarism Comparison (an online tool; www.prepostseo.com/plagiarism-comparison-search) which compares the generated and original abstracts for duplicate content. We evaluated plagiarism using PlagScan (an online plagiarism detection tool; www.plagscan.com/en/) which checks for plagiarism from various sources, including the Internet, publishers, and user-submitted documents. ChatGPT may generate abstracts using the sources of the original papers; therefore, we calculated an updated plagiarism index after excluding the self-plagiarised part (using sentences in the original papers). Finally, we examined the AI output detectors for both generated and original abstracts using a GPT-2 Output Detector (openai-openai-detector. hf.space), Copyleaks (copyleaks.com/ai-content-detector), Sapling (sapling.ai/ai-content-detector), and GPTZero (gptzero.me). The average percentage of the AI content in the first three

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4429014

detectors was calculated. GPTZero determined the content either as "entirely by AI," "parts by AI," or "entirely by a human" in each original and generated abstracts. For structured abstracts, the proportion of words in each subheading was compared between the generated and original abstracts.

Experts and quality evaluation

Five experts in psychiatry were asked to evaluate the quality of the abstracts after reading full texts of the research papers. The quality score was assigned using a Likert scale ranging from 0 to 10 (worst=0, not bad=5, extremely good=10), while considering the concise, precise, functional, unbiased, comprehensive, and self-sufficient aspects of the abstracts. Additionally, the experts were asked to identify the abstracts written by human authors. Finally, the experts determined the validity of the conclusions generated by ChatPDF during the unblinding phase.

Statistical analyses

Data were analysed using Microsoft Excel 365. Statistical analyses were performed using Statistical Package for Social Sciences V.26 (SPSS Inc., Chicago, Illinois, USA). Violin plots, bar graphs, and x-y plots were generated using GraphPad Prism V.8 (GraphPad Software Inc., San Diego, CA, USA). The Bland–Altman plot (difference plot) was constructed using MedCalc V.20 (MedCalc Software, Ostend, Belgium). Pie charts and matrix plots were generated using Microsoft Excel 365.

Group differences were analysed using two-sample t-tests (for continuous variables) and chi-square tests (for categorical variables). If more than 20% of the cells had expected cell counts of less than 5, Fisher's exact test was conducted for categorical variables. Correlations between two variables were examined using Pearson's correlation coefficients. Bland–Altman plots were used to assess the agreement between the scores of the generated and original abstracts. Linear regression analysis was performed to determine predictors of the scores. All analyses were two-tailed, and $p < 0{\cdot}05$ was considered significant.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4429014

Preprint not peer reviewed

Result

Characteristics of the five experts and included articles

The H-indices of the five experts ranged from 26 to 54, and their research experience from 16 to 29 years (eTable 1). The selected 20 articles were published between 2021 and 2023 (eTable 2). The full texts of the generated structured and unstructured abstracts are presented in eTable 3. The average number of words in the structured and unstructured abstracts were 397·40 (43·02) and 249·80 (25·36), respectively.

Similarity and plagiarism

Similarity (duplicate content) between the original and the generated abstracts was 16·35% (Figure 1a; structured: 17·30%; unstructured: 15·4%). The percentage of plagiarism was 18·75%, of which 11·20% were self-plagiarised (Figure 1b). After excluding the self-plagiarised section of the original paper, the percentage of adjusted plagiarism was 7·55% (structured: 6·60%; unstructured: 8·85%).

AI-content

The mean percentages of AI-content for the generated structured (Figure 2a) and unstructured abstracts (Figure 2b) were 31·48% and 75·58%, respectively. Group differences between the original and generated abstracts were significant for both the structured (p=0·001) and unstructured (p<0·0001) abstracts. For ChatGPT-generated abstracts, eight unstructured abstracts were determined to be entirely written using AI, whereas only four structured abstracts were determined to be entirely written using AI (group difference: p=0·01; Figure 2c). For ChatGPT-generated abstracts, three unstructured abstracts were determined to be entirely written by AI, and none of structured abstracts were determined to be entirely written using AI (Figure 2d). Details of the percentages of each AI detector are listed in eTable 4.

Quality, correlation, and agreement

For structured abstracts, there was no difference in the quality of the generated [7·04 (1·48)] and the original [7·20 (0·30)] abstracts (difference: -0·16; -0·89 to 0·57; p=0·67) (Figure 3, eTable 5). However, for the unstructured format, the quality of the generated abstracts (5·24 [1·86]) was lower than that of the original abstracts (7·54 [1·68]) (difference: -2·30; -3·00 to -1·60; p<0·001). Combining all abstracts, the quality of the generated [6·14 (1·90)] abstracts was lower than the original [7·37 (1·93)] abstracts (difference: -1·23; -1·76 to -0·70; p<0·001).

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4429014

There is a negative correlation between the quality scores of ChatPDF-generated structured abstracts and those of original structured abstracts (eFigure 1a and eTable 6; r=-0·43; -0·63 to -0·17; p=0·002); however, the correlation is not significant for the unstructured abstracts (p=0·91). The consistency between the scores of the generated and original abstracts mostly fell within a 95% agreement (eFigure 1b).

Predictors of quality scores

A structured format was associated with higher quality scores in ChatGPT-generated abstracts (coefficient: 2·067, p<0·001) (eTable 7), whereas a higher H-index was associated with higher quality scores in the original abstracts (coefficient: 0·079, p<0·001).

Proportion of words in each subheading

In the structured abstracts (Figure 4 and eTable 8), the proportion of words in each subheading differed between generated and original. The original human authors wrote more words in the subheadings: "results" (34.0% vs 16.4%, p<0.001) and "design, setting, and participants" (21·2% vs 13·8, p=0·004), whereas ChatGPT generated more words in the remaining five subheadings. The total word counts and word proportions of each subheading are listed in eTable 9.

Accuracy of identification and validity of conclusions

Three experts were asked to identify the abstracts written by the original authors. The accuracy of identification was only 40% for structured abstracts (Figure 5a), whereas it was 73% for unstructured abstracts (Figure 5b). The accuracy differed significantly between the structured and unstructured abstracts (p=0·009). Importantly, 30% (6/20; i.e., 2 unstructured and 4 structured abstracts) of the abstracts generated by ChatGPT revealed incorrect conclusions (eTable 3 and Figure 5c). For example, in 3rd structured abstract, ChatGPT concluded that "The TUNED study found no evidence to support the use of home-based tDCS for reducing inattention symptoms in adults with ADHD." However, the actual conclusion of this article was "The TUNED study supported that home-based tDCS would improve attention in adult patients with ADHD. Home-based tDCS could be a non-pharmacological alternative for patients with ADHD." (eTable 3)

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4429014

Preprint not peer reviewed

Discussion

The main findings of this study are as follows: First, the abstracts generated by ChatGPT do not reveal a high percentage of similarity or plagiarism when using full texts as prompts. Second, the quality of the generated structured abstracts was similar to that of the original abstracts, whereas the quality of the generated unstructured abstracts was lower than that of the original ones. Third, among the generated abstracts, eight unstructured, in contrast to only four structured abstracts were determined to have been written entirely by AI. Fourth, in structured abstracts, humans tended to focus on the results and study design subheadings, whereas ChatGPT generated a higher proportion of words in the remaining five subheadings. Fifth, the scores of the generated and original abstracts mostly fell within 95% agreement. Sixth, the identification accuracy was only 40% for structured abstracts and 73% for unstructured abstracts. Finally, six of the twenty abstracts generated by ChatGPT drafted incorrect conclusions.

Concerns have been raised regarding potential plagiarism issues in articles generated by ChatGPT.[10,11] We examined the similarity (duplicate content) and plagiarism of ChatGPT-generated abstracts using the full text of a research paper as a prompt. Because both humans and ChatGPT may utilise some sentences from the full texts while writing the abstracts, we found that the generated and original abstracts contained 16·35% duplicate content. Furthermore, because we used the full texts of a research paper as a prompt, 11·20% of plagiarism was self-plagiarised (i.e., the source was the sentence of the original published paper), which will not occur when we submit a new paper for consideration for publication. The adjusted plagiarism rate was 7·55% for the generated abstracts, and we found that electronic news about original articles still accounted for a certain proportion of plagiarism. A previous study by Khalil et al. requested ChatGPT to generate essays on various topics and evaluated the plagiarism.[8] They found that 40 of 50 essays showed high originality, with a similarity score of less than 20 %. In summary, ChatGPT-generated abstracts did not reveal serious similarities or plagiarism when full texts were used as prompts.

In the current study, ChatGPT generated higher-quality structured abstracts than unstructured abstracts. The subheading instructions helped ChatGPT generate more concise, precise, and self-sufficient content. However, the proportion of words in each subheading was significantly different between the original author-written and generated abstracts. We found that ChatGPT missed statistical values in the results (e.g., p-values and confidence intervals). Moreover, ChatPDF tended to generate

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4429014

more words in the subheadings: "importance (i.e., background)" and "conclusions and relevance." However, the original authors focused on the subheadings: "results" and "study design." We believe that if a prompt includes a specific proportion of subheadings, this difference may reduce.

The incorrect conclusions drafted by ChatGPT reflect a major concern for AI chatbots in academic writing: inaccuracies and misinformation.[5,12] In particular, most incorrect conclusions were derived from the generated structured abstracts (4/10). When we requested ChatGPT with instructions for subheadings, it showed a good ability to generate concise and clear descriptions. However, while synthesising complex content in articles, it would lead to an incorrect judgment of the conclusion. This was not the case for the generated unstructured abstracts, wherein only two (2/10) conclusions were incorrect.

Our study has certain limitations, and its findings require a careful interpretation. First, the small sample size of the articles limits the statistical power. Second, we included only RCT to increase the homogeneity of the study design. Similarly, we selected articles from a limited number of journals and only one type of structured abstract. However, the application of AI chatbots to other study designs or formats of structured abstracts is unknown. Third, the used ChatPDF, is based on ChatGPT 3·5. Therefore, our findings may not be applicable to other AI-language models. Furthermore, the version, function, and logic of the language would get updated over time. Our findings are tentative under the current version of ChatPDF.

In conclusion, we used the full texts of a research paper as prompts for the ChatGPT. We found that generated abstracts may show lower quality and higher accuracy in an unstructured format, but higher quality and lower correctness in a structured format. Our study suggests that although AI algorithms demonstrate a good capacity to replicate some aspects of the human process of constructing an abstract, they produce critical errors in the conclusions. Further comprehensive studies are warranted to replicate our findings and extend them to different study designs and journal styles.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4429014

Acknowledgment

We acknowledge that the entire content of the paper including abstract is written by human authors and edited by all the human coauthors.

Contributors

CSL and PTT conceived and designed the study. TWH and CSL wrote the first draught of the manuscript. SJT, CHK, PTT, CSL, and KPS rated the abstracts. KPS, TT, CWH, FCY, CKT, YKT, and SZK interpreted the data and contributed to the writing of the final version of the manuscript. TWH have accessed and verified the data. CSL and PTT were responsible for the decision to submit the manuscript. All authors confirmed that they had full access to all the data in the study and accept responsibility to submit for publication.

Declaration of interests

All authors declared no conflict of interest.

Data sharing

The data that support the findings of this study are available from the corresponding author (CSL) upon reasonable request.

Figure legend

Figure 1. Similarity (duplicate content) and plagiarism of the ChatGPT-generated abstracts[a]

[a] Number in the parentheses indicates the total percentage of plagiarism (adjusted plagiarism plus self-plagiarized part)

Figure 2. AI-content of the Chat-GPT generated and the original author-written abstracts

Figure 3. Comparisons of the quality of the ChatGPT-generated and the original author-written abstracts[a]

[a] ns: not statistically significant, *: <0·5, **:<0·01, ***:<0·001, ****:<0·0001

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4429014

Figure 4. Comparisons of word proportion of each subheading in the structured abstracts

[a] The details of the values can be found in eTable 8 and eTable 9.

Figure 5. Accuracy of identification and proportion of correct conclusions

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4429014

Reference

1. Tullu MS. Writing the title and abstract for a research paper: Being concise, precise, and meticulous is the key. *Saudi journal of anaesthesia* 2019; **13**(Suppl 1): S12.

2. Tullu M, Karande S. Writing a model research paper: A roadmap. *Journal of postgraduate medicine* 2017; **63**(3): 143.

3. Bavdekar SB, Gogtay NJ. Writing an abstract for a research manuscript: Providing an honest, succinct and complete summary. *The Journal of the Association of Physicians of India* 2015; **63**(12): 64-7.

4. Anstey A. Writing style: abstract thoughts. Blackwell Publishing Ltd Oxford, UK; 2014. p. 205-6.

5. Sallam M. ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. *Healthcare (Basel)* 2023; **11**(6).

6. Lee P, Bubeck S, Petro J. Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine. *N Engl J Med* 2023; **388**(13): 1233-9.

7. Ali SR, Dobbs TD, Hutchings HA, Whitaker IS. Using ChatGPT to write patient clinic letters. *Lancet Digit Health* 2023; **5**(4): e179-e81.

8. Khalil M, Er EJapa. Will ChatGPT get you caught? Rethinking of plagiarism detection. 2023.

9. Liebrenz M, Schleifer R, Buadze A, Bhugra D, Smith A. Generating scholarly content with ChatGPT: ethical challenges for medical publishing. *Lancet Digit Health* 2023; **5**(3): e105-e6.

10. Aydın Ö, Karaarslan EJAaS. OpenAI ChatGPT generated literature review: Digital twin in healthcare. 2022.

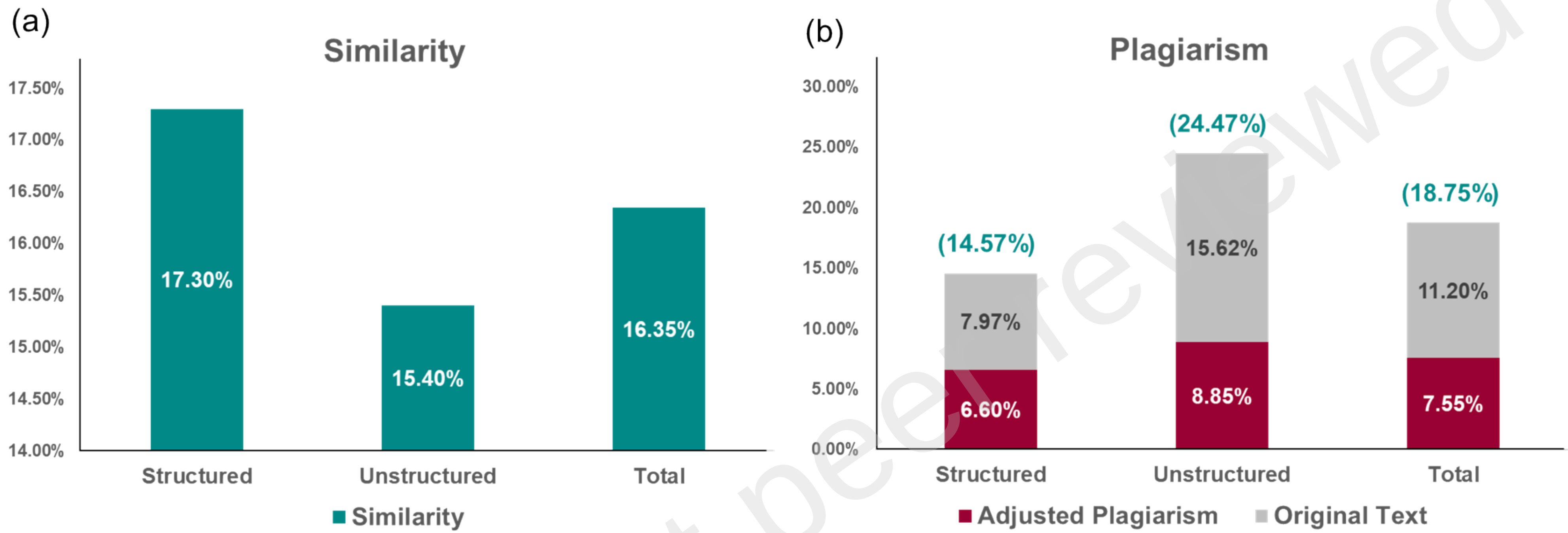
11. Dahmen J, Kayaalp ME, Ollivier M, et al. Artificial intelligence bot ChatGPT in medical research: the potential game changer as a double-edged sword. *Knee Surg Sports Traumatol Arthrosc* 2023; **31**(4): 1187-9.

12. Stokel-Walker C, Van Noorden R. What ChatGPT and generative AI mean for science. *Nature* 2023; **614**(7947): 214-6.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4429014

Figure 1. Similarity (duplicate content) and plagiarism of the ChatGPT-generated abstracts

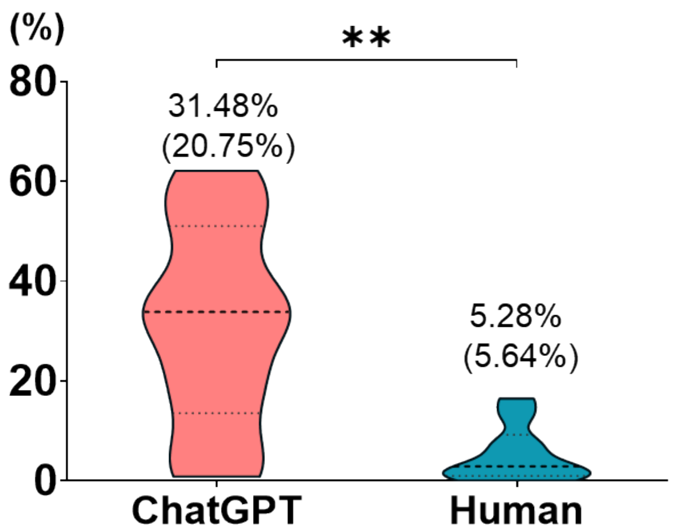


Preprint not peer reviewed

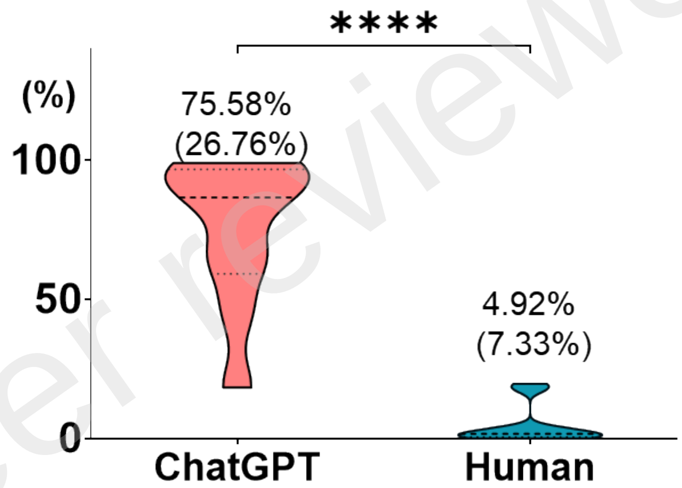
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4429014

Figure 2. Mean AI-content of the Chat-GPT generated and the original author-written abstracts

(a) Structured



(b) Unstructured



(c) ChatGPT

ChatGPT	Unstructrued	Structured
Entirely by AI	8	4
Entirely by a human	2	0
Parts by AI	0	6

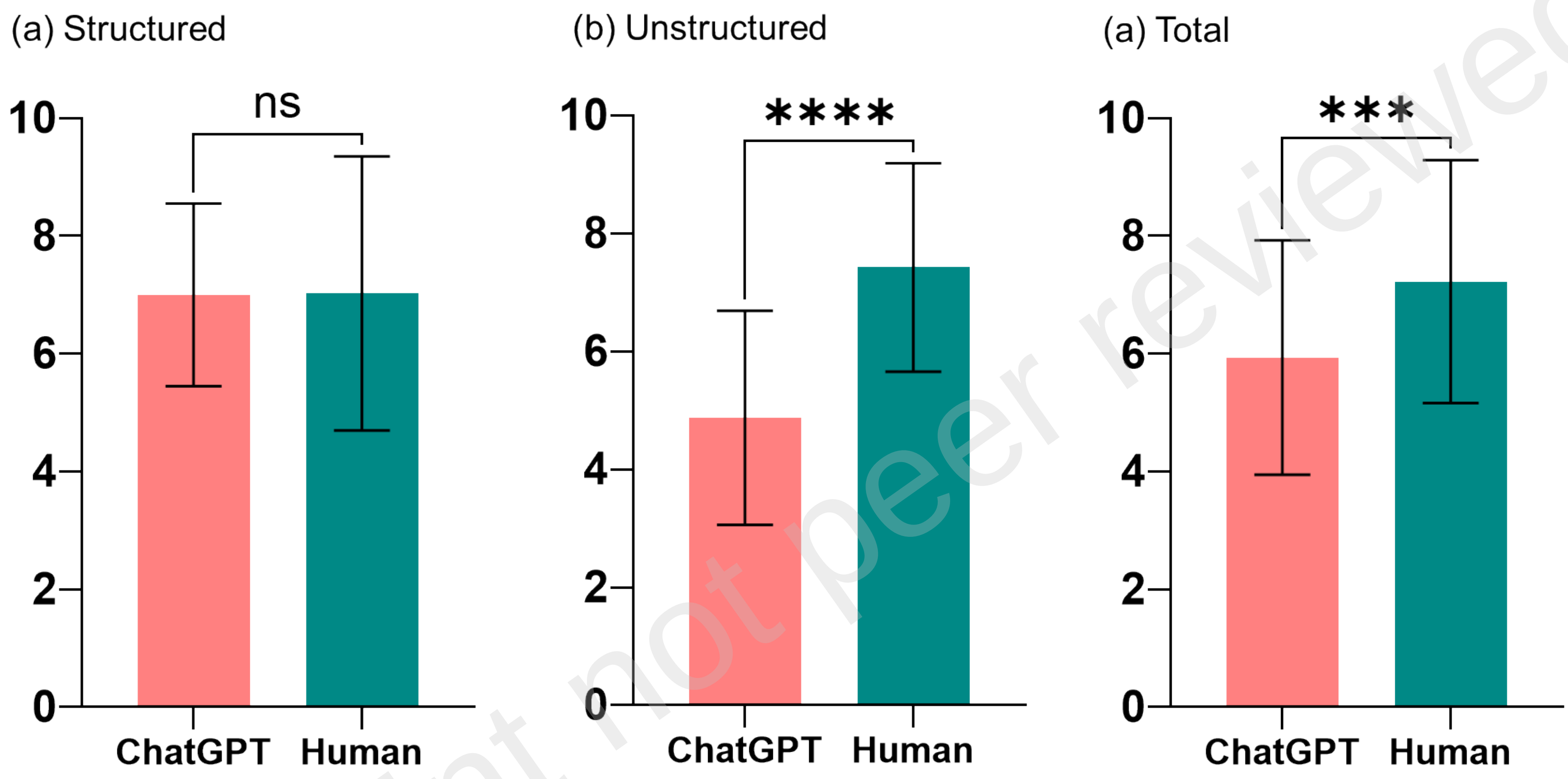
(d) Human

Human	Unstructrued	Structured
Entirely by AI	3	0
Entirely by a human	7	3
Parts by AI	0	7

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4429014

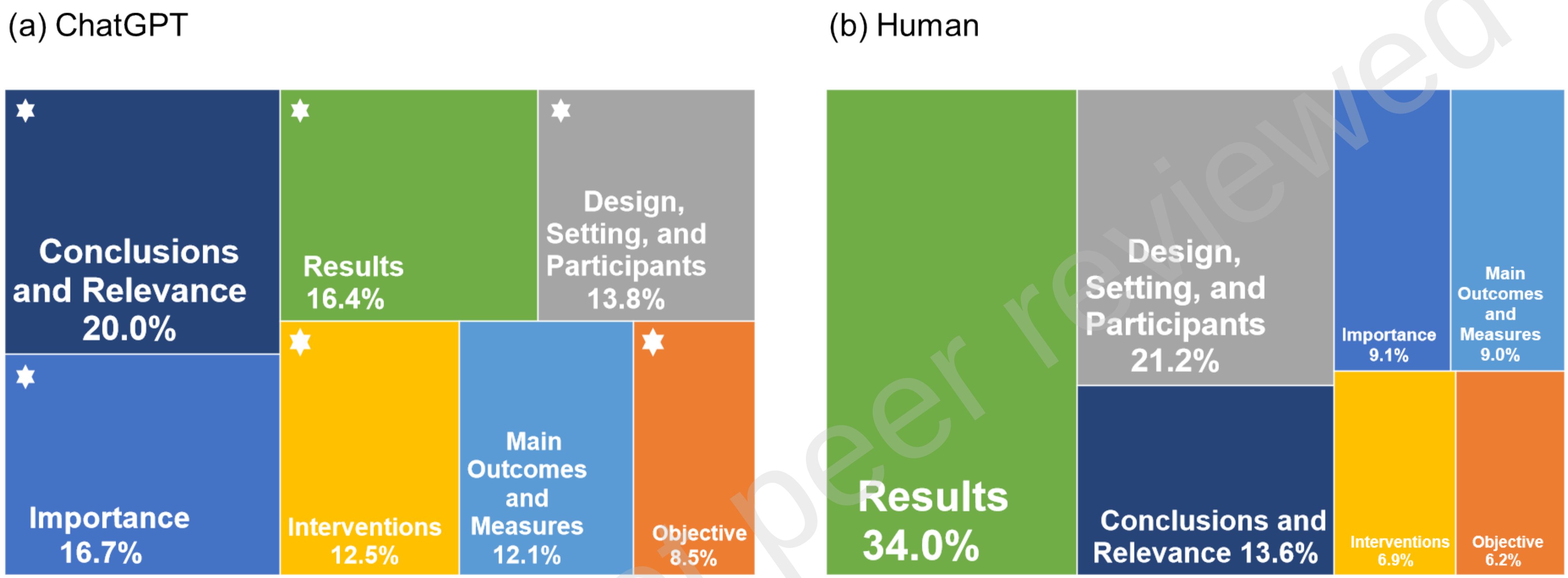
Figure 3. Comparisons of the quality of the ChatGPT-generated and the original author-written abstracts



Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4429014

Figure 4. Comparisons of word proportion of each subheading in the structured abstract

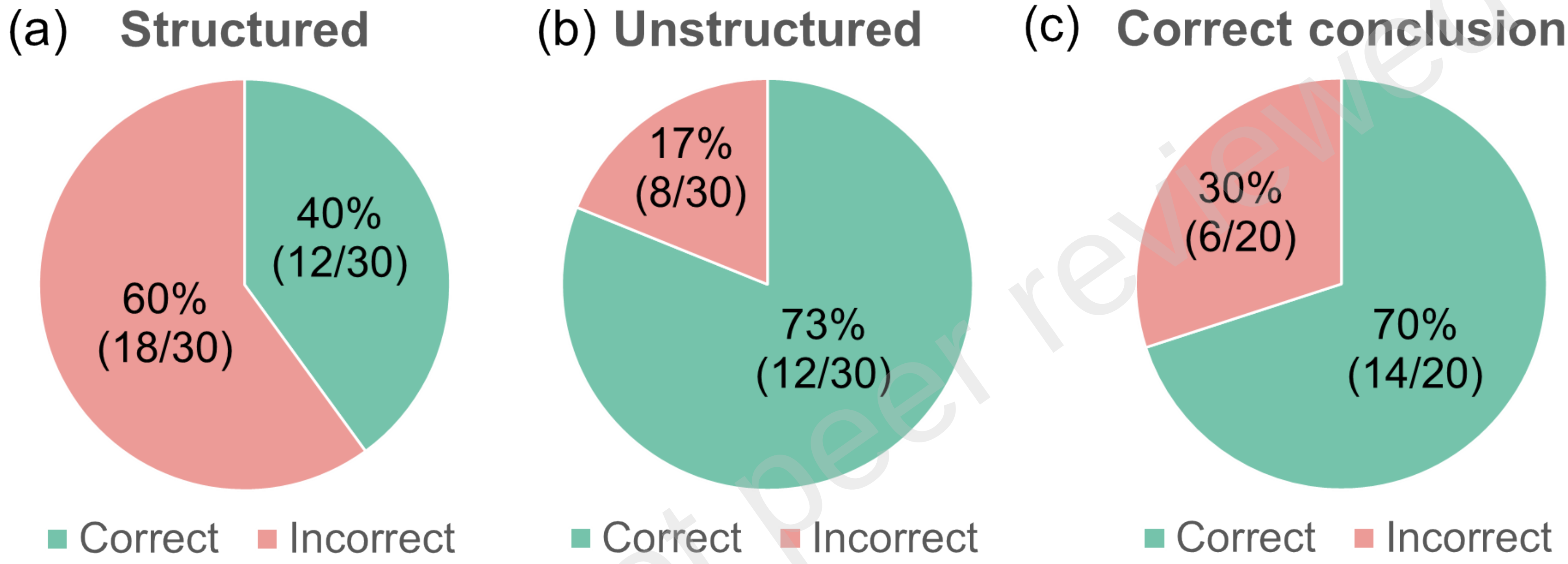


✡: indicate significant group differences

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4429014

Figure 5. Accuracy of identification and proportion of correct conclusions



Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4429014