

Comparative analysis of plain language summary abstracts: generative artificial intelligence versus human medical writers



Poster 9

Laura Watts,^a David McMinn,^b
Walter Bender,^b Wilson Joe,^a
Tim Becker^a

^aLumantia Communications Inc., Yardley, PA, USA; ^bSorcerer, Washington, DC, USA.

Introduction



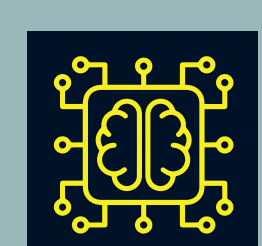
Scientific abstracts, especially those in clinical medicine, are routinely deemed “very difficult” to comprehend and are often characterized by reading levels that are consistent with a college education or beyond^{1,2}



Plain language summary abstracts (PLSAs) for scientific publications serve as an important mechanism to relay research results in a more accessible way to audiences who may not have a scientific background, including patients and caregivers^{3,4}



Regulatory agencies have recognized the importance of making clinical research more understandable and transparent to the general public by mandating lay summaries for risk management plans and clinical trial results^{5,6}



With a clear need and increasing demand for PLSAs, generative artificial intelligence (GenAI) has emerged as a powerful tool with the capacity to accelerate PLSA development and promote the readability and understanding of scientific communications for lay audiences⁷⁻¹⁰

Objective

- To assess whether GenAI could produce first-draft PLSAs of comparable quality to those developed by human medical writers (HMWs)

Methods

- Two open-access articles that were published in 2021 and 2022 from distinct therapeutic areas (oncology and infectious disease), with accompanying PLSAs, were identified via PubMed^{11,12}
 - The PLSAs for these articles served as the HMW PLSA comparators
 - GenAI PLSAs were developed using a bespoke AI pipeline that incorporated task-specific coding, with the prompt specifying a target audience of “middle school students”; the primary scientific abstract from each article served as input
 - The primary scientific abstracts, HMW PLSAs, and GenAI PLSAs are available at the QR code above
- The primary scientific abstracts, HMW PLSAs, and GenAI PLSAs were assessed using 5 standard readability metrics (automated readability index [ARI], Gunning Fog Index, Flesch-Kincaid grade level, Simple Measure of Gobbledygook [SMOG] Index, and Linsear Write Readability Formula)¹³
 - All of these readability metrics output an approximate US grade level, with lower scores/grade levels indicating that the content is easier to read
- The PLSAs were also scored by 2 blinded reviewer groups (medical writing professionals and patients/patient advocates) using a custom comprehensive readability assessment scale that evaluated 11 items (readability, comprehensibility, jargon use, clarity and flow, inclusion of key data, accuracy of content, summarization of overall message, clear description of the findings’ relevance, proper grammar, actionability [ie, was the information provided in a manner that readers could use it for informed decision making?], and empathy/understanding) on a 5-point Likert scale (1=very poor; 2=poor; 3=fair; 4=good; 5=excellent)
 - For this scale, higher scores indicate better quality
- Content accuracy was evaluated by an unblinded professional medical writer trained in hematology and infectious diseases

Limitations

- Interpretation of results is limited by the small sample size of survey participants, particularly the number of patients/patient advocates, and the limited number of PLSAs evaluated
- In addition, the novel comprehensive readability assessment scale was developed for this study and has not been validated

Results

Standard Readability Metrics

- The primary scientific abstracts had an overall mean (standard deviation [SD]) readability score/US grade approximation of 14.7 (1.7; college sophomore) across all scales, with the lowest mean result from the SMOG Index at 12.8 (0.12; high school senior) and the highest from the Gunning Fog Index at 17.1 (0.21; college graduate; **Figure 1**)
- The HMW PLSAs had an overall mean (SD) readability score/US grade approximation of 13.7 (3.5; college freshman) across all scales, ranging from 11.8 (3.0; high school junior; SMOG Index) to 15.7 (4.1; college junior; Gunning Fog Index)
- The GenAI PLSAs had an overall mean (SD) readability score/US grade approximation of 8.9 (1.2; eighth grade) across all scales, ranging from 7.4 (0.9; seventh grade; SMOG Index) to 10.1 (0.8; high school sophomore; Gunning Fog Index)

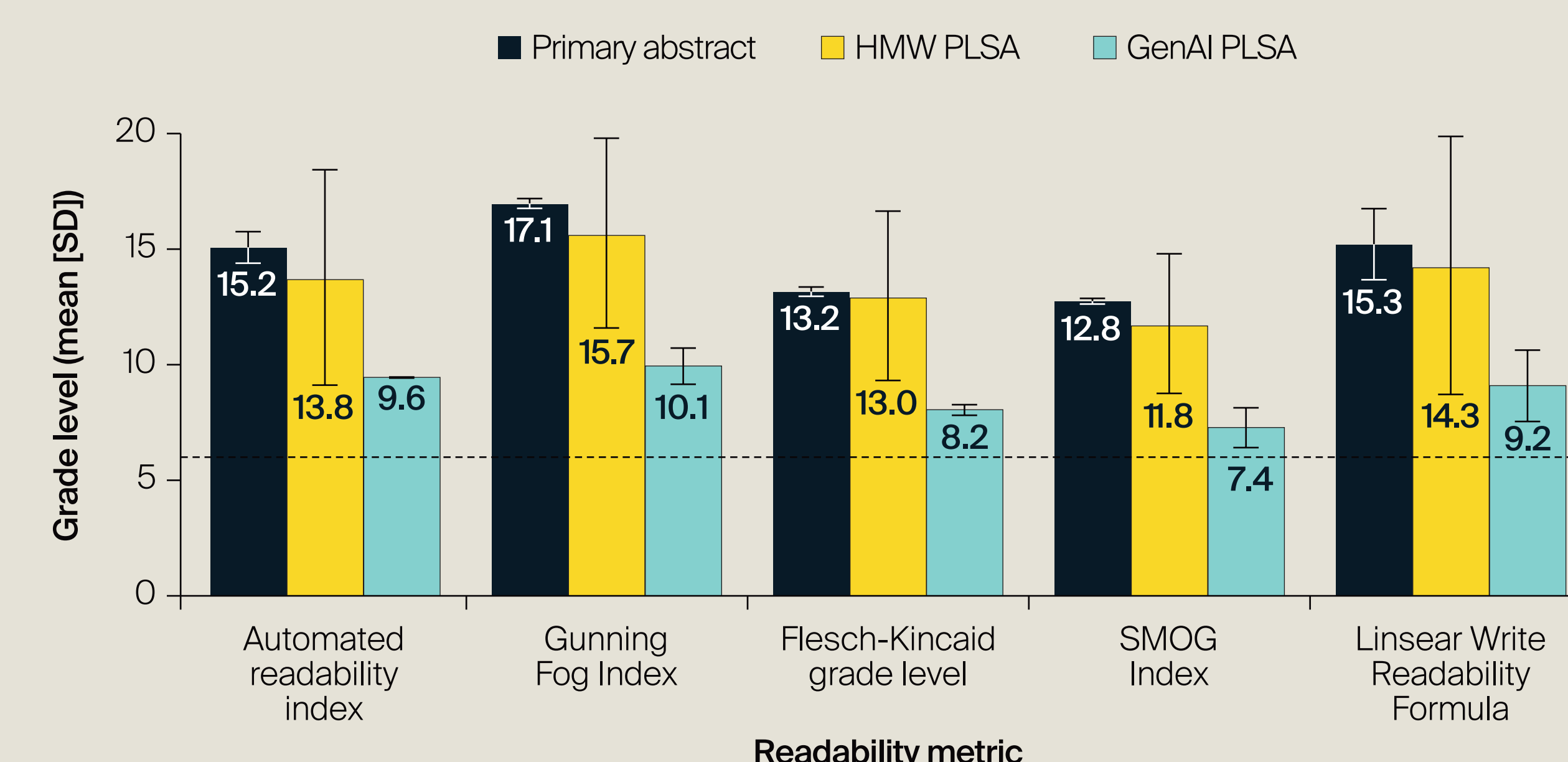
Comprehensive Readability Assessment Scale

- Overall, 22 medical writing professionals and 5 patients/patient advocates rated the PLSAs using the comprehensive readability assessment scale
- When rated by medical writing professionals on a scale of 1 to 5, the overall mean (SD) score across all items was 3.3 (0.5) for HMW PLSAs and 3.9 (0.7) for GenAI PLSAs (higher scores=better quality)
 - When comparing HMW PLSAs with GenAI PLSAs, the largest improvements using GenAI were observed for jargon usage (3.0 [0.2] vs 3.9 [0.5]), inclusion of key data (2.9 [0.5] vs 3.9 [0.7]), and accuracy (3.0 [0.8] vs 4.0 [0.7]; **Figure 2A**)
- When rated by patients/patient advocates, mean (SD) scores across all items were 3.5 (0.3) for HMW PLSAs and 4.1 (0.2) for GenAI PLSAs
 - When comparing HMW PLSAs with GenAI PLSAs, the largest improvements using GenAI were observed for jargon use (3.1 [0.4] vs 4.0 [0]) and empathy/understanding (2.9 [0.1] vs 4.0 [0.3]; **Figure 2B**)

Unblinded Accuracy Assessment

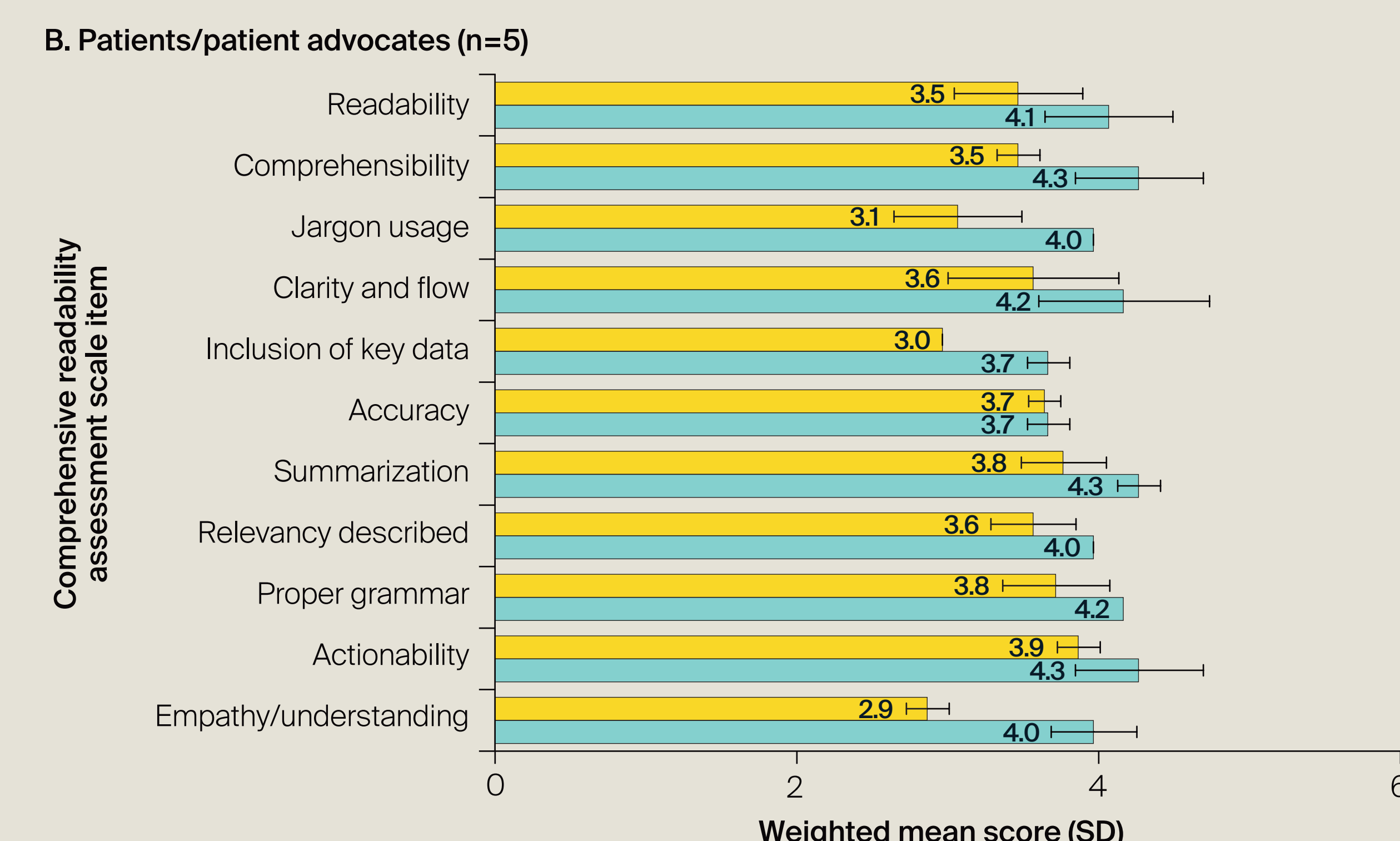
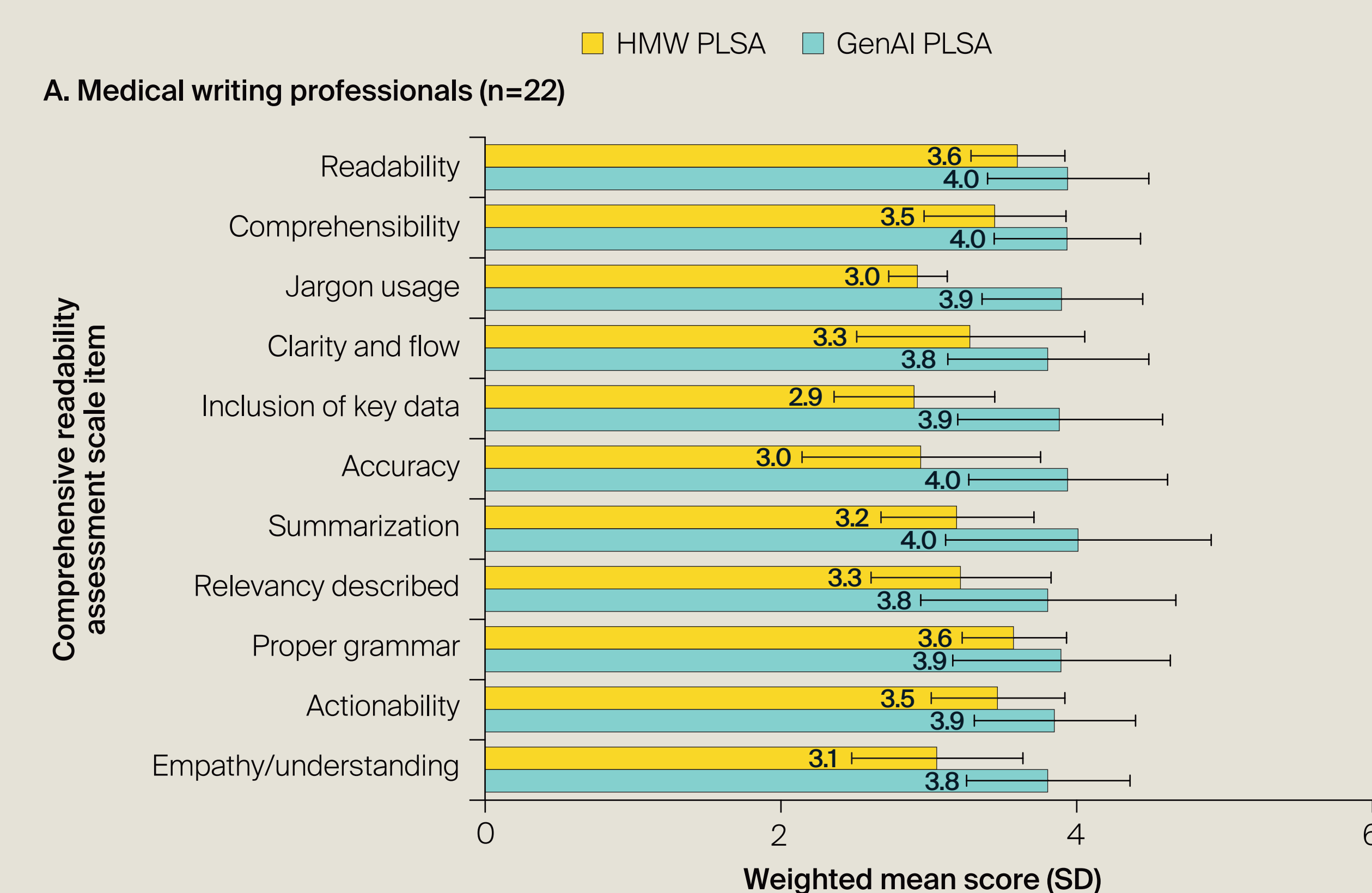
- A content error was identified in both the HMW PLSA and GenAI PLSA related to COVID-19
 - In the HMW PLSA, the phrase “COVID-19 virus” was used despite the fact that the phrase would have more accurately been written as “COVID-19, which is caused by the SARS-CoV-2 virus”
 - Similarly, in the GenAI PLSA, COVID-19 was described as “another type of virus” in a manner that suggested it was distinct from SARS-CoV-2
 - The accuracy item scores for the HMW PLSA and GenAI PLSA related to COVID-19 were similar when rated by both medical writing professionals (3.6 vs 3.5, respectively) and patients/patient advocates (3.6 vs 3.6, respectively)

Figure 1. PLSAs developed by GenAI were more readable than primary scientific abstracts or PLSAs developed by HMWs (lower=easier to read).



PLSA, plain language summary abstract; GenAI, generative artificial intelligence; HMW, human medical writer; SD, standard deviation. The dashed line represents the recommended reading level for health information.¹⁴

Figure 2. GenAI PLSAs achieved better comprehensive readability assessment scale scores compared with HMW PLSAs when assessed by (A) medical writing professionals and (B) patients/patient advocates (higher=better quality).



GenAI, generative artificial intelligence; PLSA, plain language summary abstract; HMW, human medical writer; SD, standard deviation.

Conclusions

- In line with previous studies, GenAI PLSAs were consistently more readable across standard readability metrics compared with HMW PLSAs⁸⁻¹⁰
- On all but 1 of the comprehensive readability assessment scale items, medical writing professionals and patients/patient advocates rated GenAI PLSAs higher (ie, with better quality outcomes) compared with HMW PLSAs
 - When assessed by patients/patient advocates, the largest overall difference was observed in scores for the empathy/understanding scale item, with GenAI PLSAs outperforming HMW PLSAs
- GenAI has the potential to increase the accessibility of scientific information through the rapid development of high-quality first-draft PLSAs; secondary review by a medical writing professional remains a requirement due to known limitations associated with GenAI, including the potential for the introduction of content errors¹⁵

References

- Gizzi A. *Journal of Information Sciences*. 2011;37(5):273-281.
- Powers-Sigley P, et al. *PLoS One*. 2019;14(7):e0217725.
- Brochard K, Sorcerer SM. *PLoS One*. 2019;14(11):e0224927.
- Kawer M, et al. *Cochrane Database of Systematic Reviews*. 2021;7(1):18986.
- European Medicines Agency. *Guidance on the format of the risk management plan (RMP) in the EU*. In: Integrated format. Accessed March 8, 2024. https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/guidance-format-risk-management-plan-emp-eu-integrated-format-rev-201_en.pdf.

- McMinn D, et al. Presented at: European Meeting of ISMP; January 24-25, 2023; London, UK. Poster #21.
- McMinn D, et al. Presented at: ISPOR Europe; November 12-15, 2023; Copenhagen, Denmark. Poster #44.
- McMinn D, et al. Presented at: European Meeting of ISMP; January 23-24, 2024; London, UK. Poster #32.
- Bender W, et al. Presented at: 20th Annual Meeting of ISMP; April 29-May 1, 2024; Washington, DC, USA. Poster #3.
- Parker CC, et al. *PLoS Med*. 2022;19(6):e1003988.
- Myint PK, et al. *Ther Adv Drug Saf*. 2021;12(2):204209820985690.

Acknowledgments

This study was funded by Lumantia Communications Inc.

LW, WJ, and TB are employees of Lumantia Communications Inc. DM and WB are employees of Sorcerer.

Disclosures

LW, WJ, and TB are employees of Lumantia Communications Inc. DM and WB are employees of Sorcerer.

