

Elena Dang, John Watson, Stephen Bublitz, Suzanne Lipman, Susan Croce, and Jeffrey Stumpf

MedVal/PharmaWrite, Princeton, NJ

8

Improved Readability of Plain Language Summaries Using ChatGPT



WHY ARE PLAIN LANGUAGE SUMMARIES (PLS) IMPORTANT?

- PLS are an important tool to increase engagement in healthcare, by making research more understandable to nonexperts¹
- Assessing the reading level of PLS gives insight into the accessibility for the target audience
- However, authors may not consider the appropriate reading level for the target audience when writing PLS²



WHAT READING GRADE LEVEL SHOULD AUTHORS TARGET?

Considering Accessibility and Precision in Targeting Various Reading Levels

READING GRADE LEVEL	6-8 ^a	9-12	13-16
TARGET AUDIENCE	Patients and caregivers including individuals with lower literacy levels, language barriers, or diminished cognition	Nonexpert but scientifically engaged patients, caregivers, policy makers, and so on	General healthcare professionals and experts in that therapeutic area or field
EXAMPLE MATERIALS	Patient handouts and administration guides	Clinical trial summaries and PLS	Review articles and meta-analyses

MORE ACCESSIBLE ← → MORE PRECISE

^aReading Grade Level recommended by NIH for patient education materials and average Reading Grade Level of U.S. adults^{3,5}



HOW COULD AUTHORS IMPROVE COMMUNICATIONS FOR BROADER AUDIENCES?

- Generative artificial intelligence (AI) tools such as ChatGPT (OpenAI, San Francisco, CA) could be prompted to create and modify PLS at different reading levels
 - However, determining the best practices for using ChatGPT and understanding the accuracy of relevant information in the generated PLS is necessary



WHAT DID WE EVALUATE?

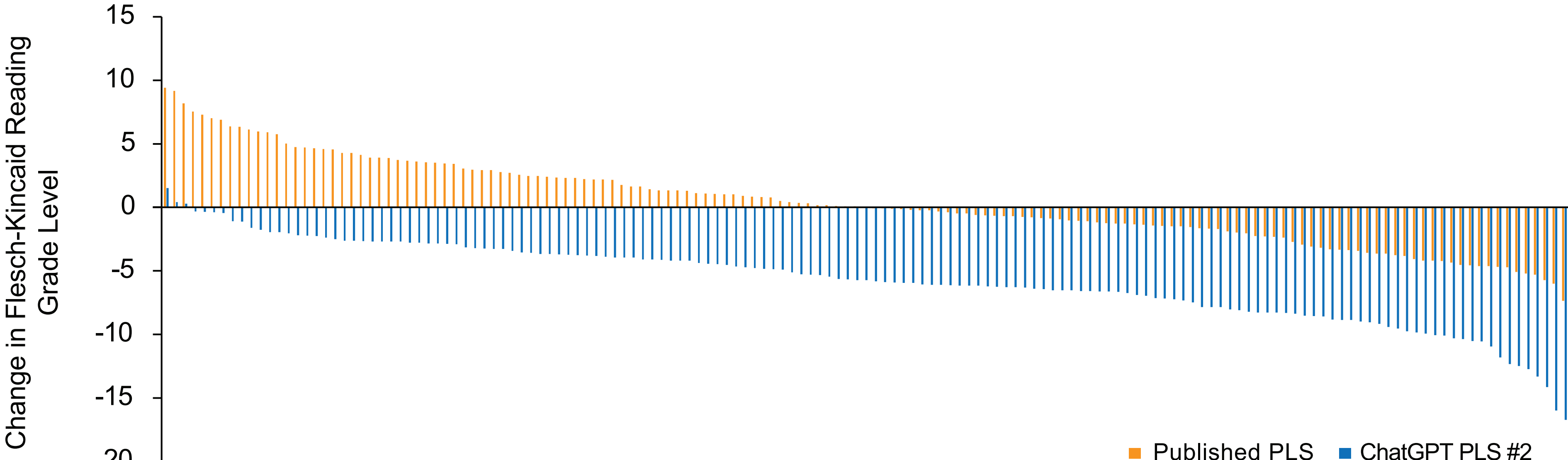
190  articles were identified

153  abstracts with PLS available on PubMed were evaluated for the Flesch-Kincaid Grade Level^b

Then, using the published abstracts and 2 successive ChatGPT prompts, PLS were generated and evaluated^c

N=153	PUBLISHED ABSTRACTS	PUBLISHED PLS	ChatGPT-GENERATED PLS	
			PROMPT 1	PROMPT 2
Flesch-Kincaid Grade Level ^b , Mean±SD	15.2±2.7	16.0±4.7	10.9±2.7	9.5±2.4
Reading Grade Level ≤10, %	2.6%	1.9%	38.6%	62.8%
Reading Grade Level Reduction of ≥2 Compared to Published Abstract, %	—	22.9%	73.2%	90.2%

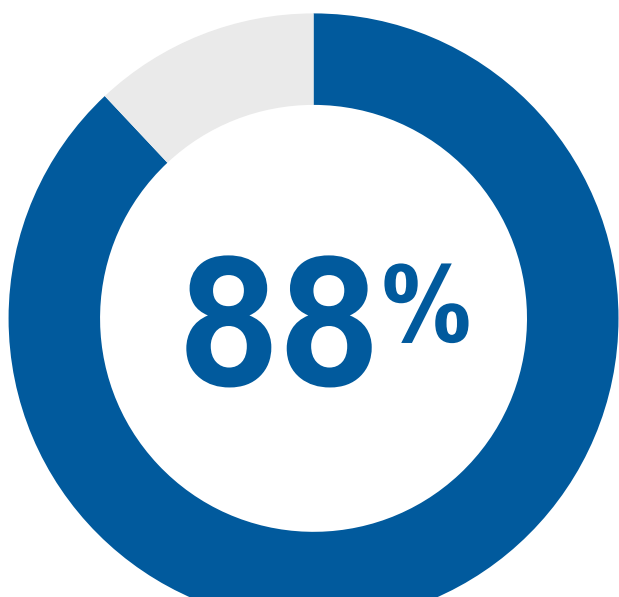
Reading levels of ChatGPT-generated PLS were more consistently reduced than those of published PLS



^dTwo outliers were removed with reading grade levels of 40 and 55, possibly due to limitations in the readability assessment tool.

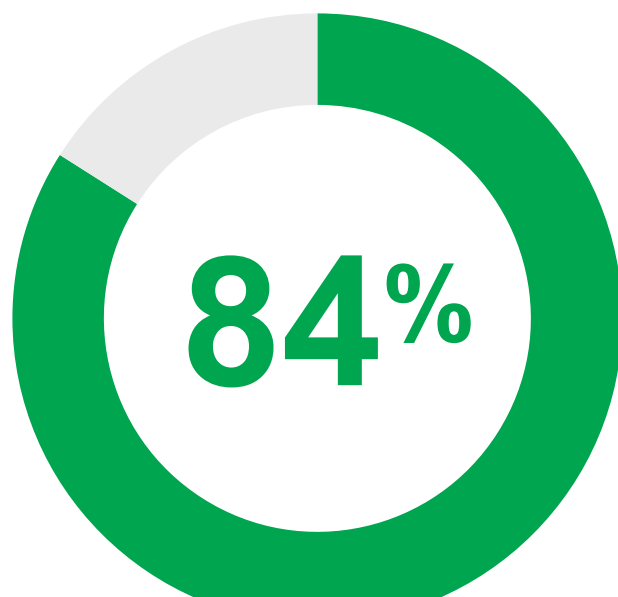
For 25 randomly selected ChatGPT-generated PLS:

- 20% of the randomly selected ChatGPT-generated PLS did not include an acceptable amount of key points from the methods section of the abstract, compared with 12% for the background and 8% for the results



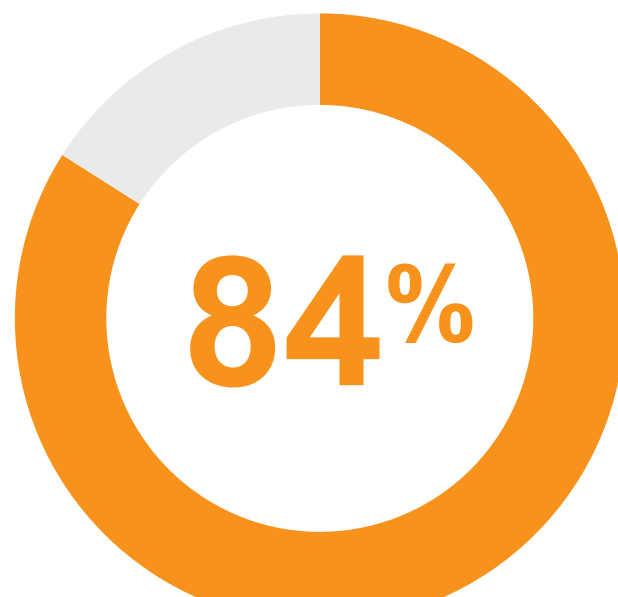
88%

included nearly all key points from the abstract



84%

accurately described the key points



84%

were sufficient to understand the abstract

^bArticles were removed if they were duplications or the PLS were not publicly accessible in PubMed. ^cWith ChatGPT 3.5, two prompts were used to generate new PLS from the published abstracts. Prompt 1: I am a medical writer at a scientific and medical communications company working on plain language summaries for patients and caregivers. The current text is written at too high of a reading level. Please take this text and make it easier to understand, while keeping important terminology, with a limit of 300 words. (Followed by the abstract). Prompt 2: The readability score is still too high. Could you please create another summary less than 300 words that could be read by an eighth grader?

WHAT DID WE LEARN?

- Recently published PLS are not written at a reading level to accommodate wider audiences
- ChatGPT efficiently reduced reading levels of abstracts in generating PLS while sufficiently and accurately maintaining key points
- This research demonstrates the potential of AI to facilitate clear communication of medical information to wider audiences, thereby promoting scientific and medical literacy
 - However, human expertise and oversight is generally recommended to review AI-generated content due to the potential for misinterpretation or hallucinations

REFERENCES

- Edgell C, Rosenberg A. Putting plain language summaries into perspective. *Curr Med Res Opin.* 2022;38(6):871-874.
- Gainey KM, et al. What author instructions do health journals provide for writing plain language summaries? A scoping review. *Patient.* 2023;16(1):31-42.
- Center for Plain Language. What is readability? [Internet]. Center for Plain Language; Accessed March 4, 2024. Available from: <https://centerforplainlanguage.org/what-is-readability/#:~:text=U.S.%20illiteracy%20statistics%20from%20the,12%20to%2014%20years%20old>.
- Kutner M, et al. The health literacy of America's adults results from the 2003 National Assessment of Adult Literacy (NCES 2006-483). US Department of Education. 2006. nces.ed.gov/pubs2006/2006483.pdf.
- Weis BD. Health Literacy: A Manual for Clinicians. American Medical Association Foundation and American Medical Association; Chicago, IL, 2003.
- Readability Formulas - Readability Scoring System. Readability formulas. <https://readabilityformulas.com/readability-scoring-system.php>. Published Sept 2023. Accessed March 11, 2024.

ACKNOWLEDGMENTS: We thank Christy King (Data and Analytics of ClinicalMind) for assistance with the initial formulation of this project.

DISCLOSURES: All authors are employees of MedVal/PharmaWrite. JS is a member of the ISMPP meetings committee.

SUPPORTED BY: MedVal/PharmaWrite, Princeton, NJ.

POSTER PRESENTED AT THE 20TH ANNUAL MEETING OF ISMPP, APRIL 29-MAY 1, 2024, WASHINGTON, DC