Artificial intelligence (AI)-augmented analysis of the readability of source abstracts, published plain language summaries (PLSs), and AI-generated alternatives (AI-PLS): Is 'the literature' becoming more accessible?

Emma Leah[a] and Bhartendu Srivastava[b]

[a]Syneos Health Medical Communications, London, United Kingdom

[b]Syneos Health Medical Communications, Gurugram, India

Conclusion

Published PLSs are generally short and difficult to read, often with worse readability than their associated abstract.

As unverified AI-PLS alternatives are so easily accessible by readers, it is of increasing importance to improve readability of peer-reviewed PLS output, perhaps through implementing quality standards, to increase accessibility of trusted literature.

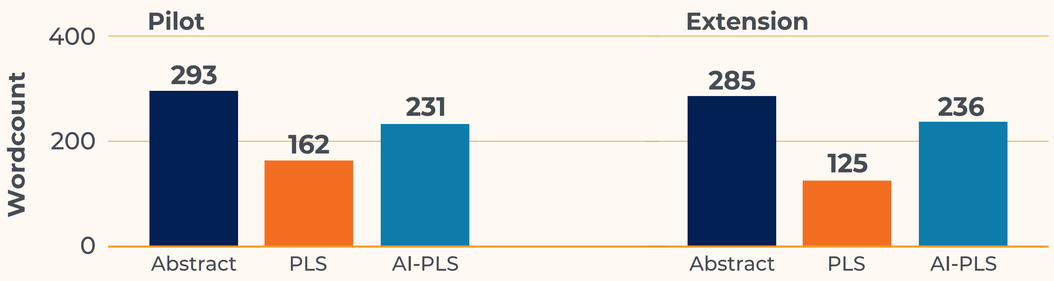
Results

In a pilot assessment by therapy area (TA), there was no apparent trend to differences between TAs. In the extension assessment, we therefore chose to focus on one TA (oncology). Details of the pilot and extension are available in the **supplementary information** (accessible via the QR code).

Wordcounts

Published PLSs of >50 words that are available alongside the abstract on PubMed are, on average, almost half the length of the published abstract. Findings were similar across the pilot and extension.

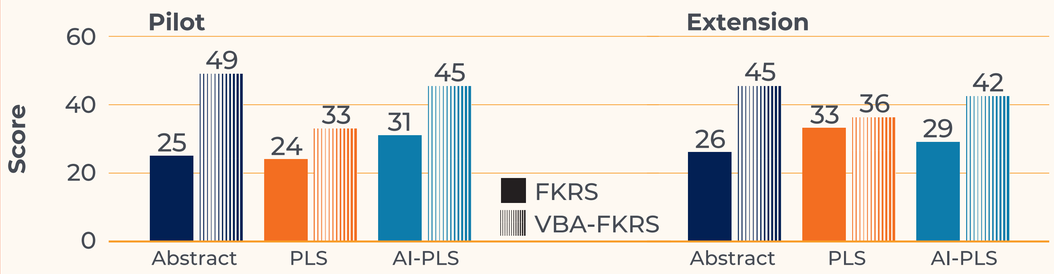
Mean wordcount



The pilot assessment included 25 selected papers across five TAs and the extension included 30 selected papers in oncology.

Readability

Mean readability



All content types assessed — abstracts, published PLSs, and AI-PLSs — were at least difficult to read, according to mean Flesch-Kincaid reading scores (FKRS; score of 30–50). Some were very difficult to read (score of 10–30).

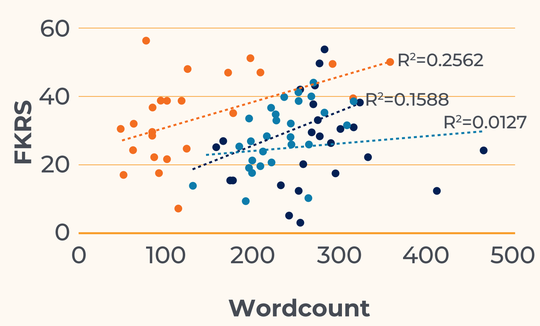
In the pilot, published PLSs were, on average, slightly more difficult to read than their associated abstract (FKRS of 24 and 25, respectively). Conversely in the extension, the PLSs scored higher than the abstracts (FKRS of 33 and 26, respectively). AI-PLSs scored as marginally more readable than the abstracts in both the pilot and extension assessments. Compared to the published PLSs, AI-PLSs were scored slightly less difficult to read in the pilot, and slightly more difficult in the extension.

The Visual Basic FKRS (VBA-FKRS) was developed with the assistance of generative AI to potentially help scale-up readability analysis. VBA-FKRS gave higher scores to all content in comparison with FKRS, although, if applying the same thresholds as for FKRS, all were still rated as difficult to read on average. The least difficult to read content type according to VBA-FKRS was the abstracts, with the published PLSs being the most difficult. Results were similar in the pilot and extension.

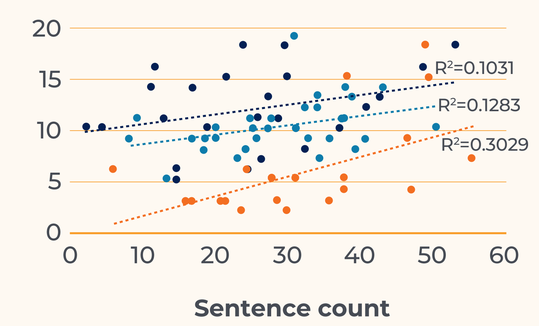
Relationship between readability and word or sentence counts

Readability by FKRS or VBA-FKRS increased with higher wordcount and sentence count in the pilot (see **supplementary information**), and extension assessments. Using FKRS in the extension, these trends were more noticeable for PLSs than for abstracts and AI-PLSs.

Readability (FKRS) by wordcount



Readability (FKRS) by sentence count



Using VBA-FKRS in the extension, the trend to less difficulty in reading with higher wordcount was similar across content types; however, the influence of higher sentence count on readability was more noticeable for abstracts than for PLSs and AI-PLS (see **supplementary information**).

Restricting to published PLSs >100 words

Given the influence of wordcount on readability, we tried limiting the analysis of readability by wordcount to abstracts with published PLSs >100 words (instead of >50 words). Readability as measured by both FKRS (abstract, 25; PLS, 33; AI-PLS, 39) and VBA-FKRS (47, 37, and 42, respectively) was similar to the full sample. Interestingly, the readability of abstracts according to FKRS or VBA-FKRS in this subanalysis was no longer associated with increase in wordcount. The increase in readability associated with increased wordcount was nevertheless similar for PLSs and AI-PLSs as in the whole sample. These preliminary findings warrant further investigation.

Introspection

Using the same approach, the abstract for this poster scored 40 (FKRS) and 59 (VBA-FKRS), higher than an AI-PLS (which scored 34 by FKRS and 53 by VBA-FKRS). Nevertheless, the FKRS level is "Difficult" for both content types. The poster scored 36 (FKRS) and 62 (VBA-FKRS).

Background and objective

PLSs, AI-augmented or not, have potential to increase knowledge equity. When available alongside published abstracts in PubMed, peer-reviewed PLSs may increase accessibility of information and save desk-research time by providing clearly-worded synopses. Nevertheless, implementation has been heterogeneous. Generative AI offers a shortcut to similarly summarised content, but the output should be verified by an expert reviewer to ensure accuracy. We investigated availability and comparative readability of abstracts, their PubMed-adjacent PLSs, and AI-PLSs generated from the same source abstracts via simple prompting.

Methodology

A pilot assessment was conducted across five TAs (more details in the supplementary information), totalling 25 abstracts.

Abstracts were identified via PubMed search for plain-language summary. All abstracts in the extension assessment (n=30) were about cancer, and were published between 1 Oct 2024 and 10 Dec 2024. Stand-alone PLSs and abstracts with PLS ≤50 words were excluded.

AI-PLSs were generated by Perplexity AI (chosen in the pilot assessment), via a simple prompt to produce a PLS from the abstract provided, with no instruction regarding length.

FKRSs were obtained via an online calculator. To reduce the need for manual cut-and-paste, and potentially enhance the ability to analyse at scale, a tool was created in Microsoft Excel using generative AI (Copilot) to help write and refine a VBA based on FKRS (VBA-FKRS).

Readability by wordcount and sentence count was analysed descriptively. Outlier PLSs (those with scores <10) were manually screened to confirm poor readability. Content that could not be rated by the online FKRS calculator (returning a result of "NA") and/or was scored with a negative value in the VBA-FKRS was treated as an error and excluded from the analysis.

FKRS definitions

FKRS	Readability level
0–10	Extremely difficult
10–30	Very difficult
30–50	Difficult
50–60	Fairly difficult
60–70	Plain English
70–80	Fairly easy
80–90	Easy
90–100	Very easy

Limitations

This was an exploratory assessment with no formal statistical analysis. Identified PLSs were only those tagged and readily available via simple PubMed search. Quality of AI-PLS was not systematically screened. The VBA-FKRS was created and refined with the assistance of Copilot, and requires further refinement and validation in a separate dataset to confirm reliability. A larger sample size is needed to further investigate any potential differences between TAs. Wordcount and sentence count are components of the FKRS. Alternative methods of assessing readability and analysing associations with text length and structure should be investigated. See supplementary information for more details on limitations.

Disclosures: Both authors are employees of Syneos Health.

Acknowledgements: The authors are indebted to Zaavan Baildon of Syneos Health, for assistance with the extension assessment and initial drafting of the poster, and to Chiara Lee, PhD and Lee Lemon for providing editorial and typesetting support, respectively (in accordance with fourth GPP [2022] and ICMJE guidelines).

Author contributions: Both authors contributed to the planning, data collection and analysis, and writing/review of the abstract and poster.

Poster presented at the European Meeting of the International Society for Medical Publication Professionals (ISMPP) | London, 27–29 January 2025

Syneos Health Communications