AMERICAN MEDICAL WRITERS ASSOCIATION AMWA JOURNAL V38 N3 / 2023

THEME ARTICLE

Science Without Borders – Can Translation Tools Bridge the Language Gap?

Claire Beeby[1]; Eleanor J. Raynsford, PhD[2]; and Charles Pollitt, PhD[2] / [1]Oxford PharmaGenesis Ltd, London, United Kingdom; [2]Ipsen, Slough, United Kingdom

ABSTRACT

Although access to scientific information has improved for the general public since the introduction of plain language summaries (PLSs) and the open-access publishing movement, language barriers still impede the widespread dissemination of information. Most scientific articles are published in English language only, despite English speakers comprising just 17% of the world's population. Here we present a pilot analysis that aimed to compare the translation quality of PLSs and abstracts translated by a selected browser-based translation software. We translated abstracts and PLSs from 5 medical journal publications into French, German, Mandarin, and Slovenian using Google Translate. Four bilingual reviewers with a scientific background assessed the translation quality using pre-defined survey questions that covered the appropriateness of word/phrase selection, grammar, and clarity. We assessed the number of errors of each type and used a 5-point Likert scale to measure the impact of these errors on the meaning of the text. Translations of both PLSs and abstracts were considered accurate and readable, although PLSs scored higher across most measures. For overall accuracy, translated PLSs scored higher on the Likert scale than translated abstracts (mean, 4.60 vs 4.30, respectively), with 60% of PLS translations considered to be "very accurate" compared with 45% of abstract translations. PLSs were also considered less likely to be misinterpreted (mean, 4.55 vs 4.25, respectively), with 60% of PLS translations compared with 45% of abstract translations reported as "definitely not" likely to be misinterpreted. Based on our findings, Google Translate potentially offers a quick and easy approach to translating scientific/medical information summaries for non-English speakers. However, before these articles can be translated, they must be discoverable by non-English speakers. Engagement and collaboration with medical publishers are needed to improve access for non-English speakers, including provision and PubMed indexing of PLSs that can be translated easily.

PLAIN LANGUAGE SUMMARY

Articles that report results from scientific studies are often written in technical language that can be difficult to understand. Scientific articles usually begin with a short summary, called an abstract. Sometimes, plain language summaries (PLSs) are also available which are written using straightforward language. The aim of including a PLS is to make sure that the scientific information can be understood easily by the general public. However, there is another language barrier that can make it difficult for people to read scientific articles: most are only written in English. Specialist services can be used to translate articles into other languages, but this can be expensive and time-consuming and so is not done often. In our study, we measured how well a free, online translation tool (Google Translate) could translate PLSs and abstracts from 5 English-language articles into French, German, Mandarin, and Slovenian. Four people who spoke English and one of the 4 languages read the translations and answered a survey about the translation quality. Overall, translations of both PLSs and abstracts were accurate and easy to read, but PLS translations were slightly better than abstract translations across all the measures. The results of our study show that Google Translate offers a free, quick, and easy way to accurately translate summaries of scientific information which could help people who do not speak English to understand the information. Importantly, before articles can be translated, non-English speakers need to be able to find them. To improve access to scientific articles, we suggest that scientists work with publishers to increase the number of articles that have PLSs, and to make sure that these can be found easily by people who do not speak English.

BACKGROUND

Accessibility of scientific information is an ongoing topic of discussion. Most scientific articles are written in technical language, which is not easily understood by all readers and is a barrier to the widespread accessibility of scientific information. In fact, evidence suggests that scientific literature

ISSN 2163-5315

is becoming less easy to understand, with long words, long sentences, and jargon preventing easy comprehension.[1] This is counterintuitive given the current focus on making science accessible to all. The proportion of freely available scientific literature has continued to grow since open-access publishing was proposed 20 years ago,[2,3] but if the average reader cannot easily understand the information, these efforts seem hollow. Plain language summaries (PLSs) offer a solution to this problem, and are increasingly popular with the aim of supporting nonspecialists (as well as time-poor readers) to understand the content of research articles easily, thereby further enhancing research accessibility.[1]

Although the volume of freely available scientific literature is increasing, as well as the number of articles that include PLSs, the proportion of articles published in languages other than English is decreasing. In the early 1900s, around one-third of scientific articles were written in English.[4] This had risen to around three-quarters of scientific articles published in English by 2013.[4] However, around 83% of the world's population is non-English speaking, leaving a huge accessibility gap for both lay people and the scientific community.[4]

To close this gap, there is a need to improve accessibility of medical information for non-English speaking physicians, researchers, policymakers, patients, and caregivers. Physicians who do not speak English may be at a disadvantage if they do not have timely access to important scientific information in their own language, for example results of clinical trials. Physicians who speak English as a second language may also find it harder to understand[5] or remember[6] scientific information that they have read in English than information supplied in their native language. Even high-profile research funded by non-English speaking governments is likely to be published in English, limiting accessibility within its country of origin if no translation is provided.[7]

Although specialist translation services offer high-quality translations of scientific text, time and cost may be barriers for most individuals and organizations to get articles they want to read translated regularly. Browser-based translation tools offer the potential for quick, easy, and free-of-charge translation of scientific articles. Free translation tools are largely trained on nontechnical language rather than scientific literature, and so may not translate scientific articles as clearly or accurately as plain language text.[8] Although abstracts provide a convenient condensed summary of a scientific article's content, most are written in highly-technical language. The rise in popularity of PLSs may therefore allow for enhanced access to scientific information for non-English speakers, provided that browser-based tools can accurately translate these summaries.

In August 2022, we performed a pilot analysis comparing the quality of translation of PLSs and scientific abstracts when carried out using a selected browser-based translation software. This analysis was presented as a poster at both the European and annual meetings of the International Society for Medical Publication Professionals (ISMPP) in 2023.[9,10]

METHODS

We selected 5 Ipsen-sponsored articles with accompanying PLSs for translation (Figure 1).[11-15] Our decision to use articles and PLSs from a single source was designed to limit any impact of differences in the quality of written English in

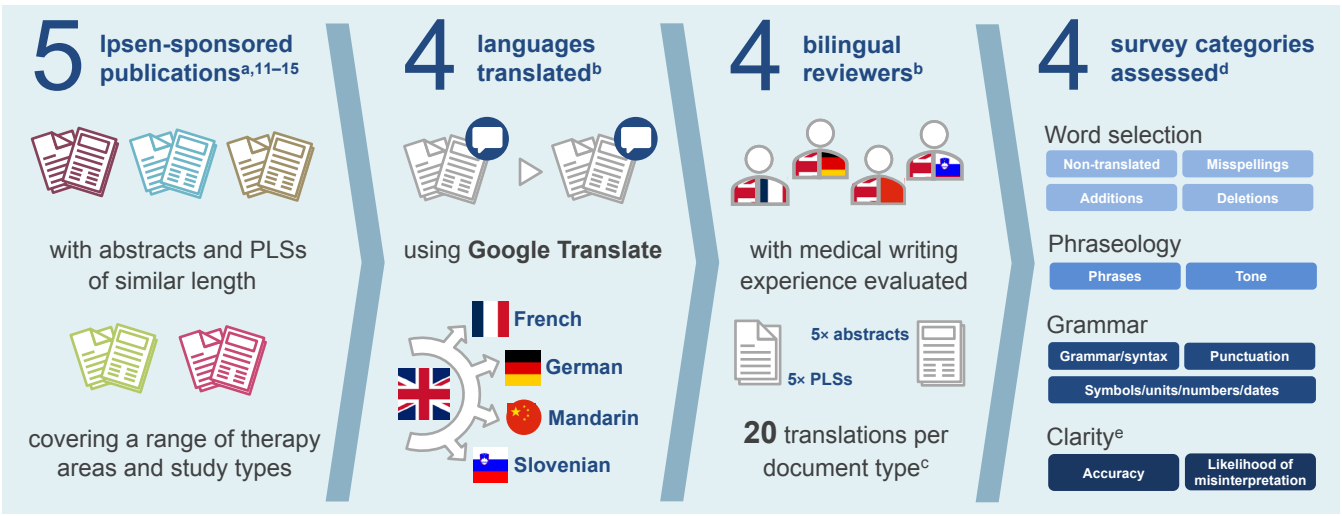

Figure 1. Study design. PLS, plain language summary.
[a]Ipsen-sponsored publications were selected because they were readily accessible and known to have both abstracts and PLSs.
[b]Languages were selected as representative of some widely spoken language families, and because these were the native languages of 4 in-house bilingual employees.
[c]One reviewer per translated language.
[d]Using standardized assessments, reviewers judged the impact of each category on the meaning, understanding, and readability of translated text.
[e]"Accuracy" and "likelihood of misinterpretation" were assessed using a 5-point Likert scale; 5 = highest accuracy and lowest likelihood of misinterpretation.

original documents on the translations. All of the selected publications had utilized medical writing assistance, ensuring that they were written in high-quality English.

Google Translate was used to translate the PLSs and abstracts into 4 languages that are representative of some widely-spoken language families (French, German, Mandarin, and Slovenian). Although other browser-based translation tools are available, we selected Google Translate because it is well known, free to use, and is incorporated in Google Chrome, the most widely-used Internet browser (as of July 2023).[16]

For each language, 1 bilingual reviewer with a scientific background assessed the translation quality using predefined survey questions that covered the appropriateness of word/phrase selection, grammar, and clarity. Reviewers assessed the number of errors of each type, and used a 5-point Likert scale[17] to measure the impact of these errors on the meaning of the text.

RESULTS

Word Selection

When assessed at the level of individual words, translations of PLSs and abstracts performed similarly (Figure 2). Inappropriate word insertion (0.2% vs 0.1% words), omission (0.2% vs 0.2%), and misspelling (0.05% vs 0.02%) rates were low in both translated PLSs and abstracts respectively, and most errors had little or no impact on the meaning of the text. However, translated PLSs had lower proportions of untranslated and mistranslated words than abstracts (both comparisons 0.6% vs 1.0%).

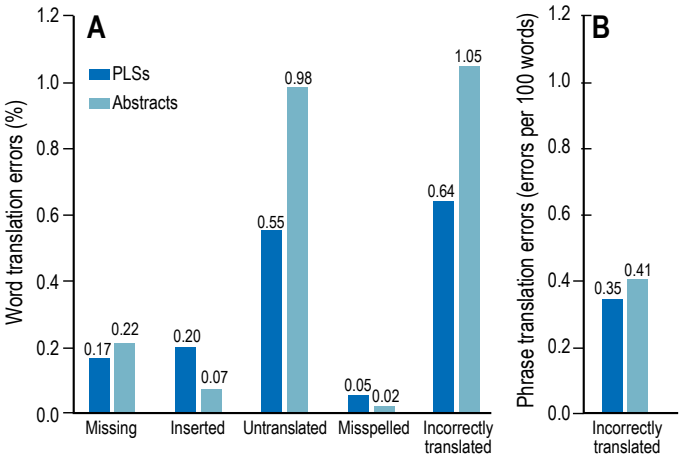

Figure 2. Translation errors (**A**. word selection and **B**. phraseology). PLS, plain language summary.

Phraseology

There were fewer incorrect phrase translations in translated PLSs than in translated abstracts (0.35 vs 0.41 errors/100 words).

Grammar

Translated PLSs had similar rates of grammatical/syntax errors to abstracts, but a lower proportion of these errors was identified as having the potential to lead to misinterpretation (17.2% vs 40%, respectively).

Clarity

On a 5-point Likert scale, translated PLS scored favorably compared with abstracts for overall translation accuracy (mean, 4.60 vs 4.30, respectively) and likelihood of misinterpretation (mean, 4.55 vs 4.25, respectively). For overall translation accuracy, the proportions of translations considered to be "very accurate" were 60% for PLSs and 45% for abstracts (Figure 3). Regarding how likely it was that translation errors would lead to misinterpretation of the information, 60% of translated PLSs and 45% of translated abstracts were scored as "definitely not" likely to be misinterpreted (Figure 4).

There were no notable differences in results between different languages, although the sample size was not large enough to detect inter-language differences.

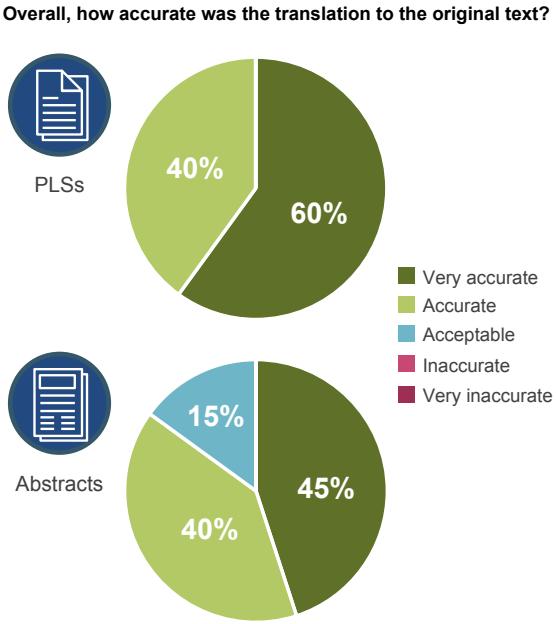

Figure 3. Translation accuracy. None of the reviewers found the translated text of PLSs or abstracts in any language to be "inaccurate" or "very inaccurate." PLS, plain language summary.

DISCUSSION

The language barrier is a big hurdle for information accessibility in scientific publishing. One solution could be for English-language journals to provide alternative languages for abstracts. Nevertheless, there are many difficulties with this approach, not least the cost, workforce capacity, and expertise needed to review translations.

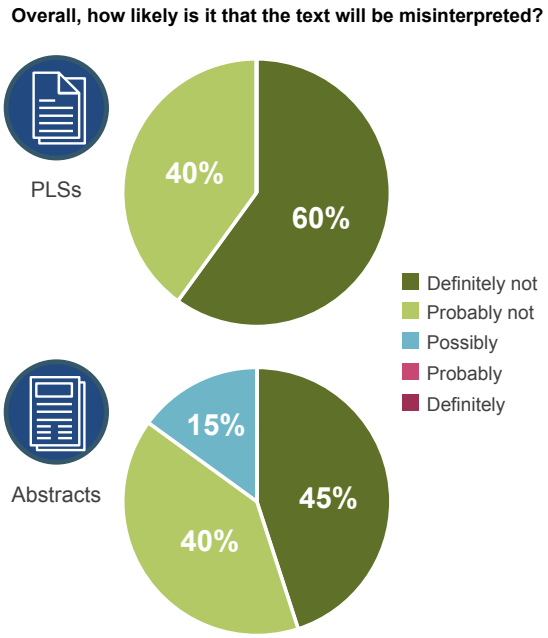

Figure 4. Likelihood of misinterpretation of translated text. None of the reviewers found the translated text of PLSs or abstracts in any language to be "probably" or "definitely" misinterpreted. PLS, plain language summary.

If freely available tools can translate scientific information clearly and accurately, non-English speakers could use these tools to read scientific abstracts in their own language. Our pilot study showed that, although not perfectly, both plain language and technical language scientific information were considered accurate and readable when translated using Google Translate. PLSs scored higher across most measures, possibly owing to the inclusion of complex sentences, abbreviations, and scientific terminology within the abstracts. However, the differences were generally small, and no statistical testing was conducted to establish whether the differences were statistically significant.

Based on comments provided by the bilingual reviewers in our study, there are potential considerations when writing PLSs and abstracts that may help to make the text more easily understood when translated via browser-based software. Translations of text that used the active voice were more readable and natural than those that used the passive voice. Acronyms were not translated consistently; although Google Translate could often recognize an acronym when it was first defined, acronyms were often lost in translation when used subsequently, or when an "s" was added to create the plural form. Practical guidance has previously been given for creating PLSs that are accessible for laypersons,[18,19] and a similar set of recommendations for creating translation-friendly summaries would go some way to helping non-English speakers have easier access to scientific information. Learnings from this study inform some initial recommendations, but more work will be needed to refine this list in the future.

Although this study focused on abstracts and PLSs, it highlights the importance of using clear and simple language in general. Ideally these learnings should also be applied to full-length articles. In 2020, Future Science Group was the first publisher to offer full-length PLSs of publications (PLSPs)—standalone summaries of entire articles written in nontechnical language.[19] Several of these summaries have been made available in several languages,[20-23] despite the original scientific article being published in English language only. Additional publishers have now also started to offer similar opportunities to publish in plain language. However, fewer than 100 PLSPs have been published to-date, so they do not yet offer a broad opportunity to make full-length plain-language texts available for translation by non-English speakers.[24]

These study results are promising, but are small adjustments to the way we write enough to enhance access to medical information for non-English speakers?

Some journals already offer translated abstracts, and multilingual journals publish abstracts in multiple languages. Other journals allow the opportunity to upload a translated abstract in the author's native language or in additional languages. Despite this, even when a translation is available, it is not always easy to find. Without knowing in advance which journals offer abstracts in their native language, non-English speakers must search for them, so how can we ensure non-English speakers can find the articles they need? Language filters are available on PubMed for articles written in non-English language, but for English-language publications, PubMed displays abstracts in English by default. When a translation is available, this must be accessed via a link.[25] There are tools available to help non-English speakers to use PubMed to find articles written in their native language. Technical solutions have been proposed to allow non-native English speakers to search for English-language articles on PubMed, for example, a Web-based tool that helps users to build PubMed searches in several languages (multilingual Query Builder).[26] However, to the best of our knowledge, these search tools are neither readily available nor in common use.

Whether they speak English or not, laypersons wishing to access scientific literature may not be experts in searching for scientific information. There may be additional need to ensure that PLSs are easily found. Although some PLSs are indexed and tagged in PubMed, there may be a need for a lay-friendly database or search engine dedicated to PLSs that would ideally be searchable in any language. Additionally, publishers and the scientific community

should work together with providers of translation software to ensure that, when professional translations of technical-language publications have been provided, these are made available for training of machine translation software.

It is likely to be some time, if ever, before Google Translate (and similar software) is sufficient for people to remove the need for specialist translation of scientific articles. This is particularly pertinent to the pharmaceutical industry, in which companies must comply with professional standards, regulations, and laws to ensure clinical trial results are reported accurately. Currently, translations from browser-based software could not be used without professional review of the translation.

Study Limitations

With only 5 publications, 4 languages, and 1 reviewer per language, the sample size in this study was limited. Although efforts were made to include different therapy areas and study types, the publications were developed by a single sponsor and were of uniform quality, and the writing style was similar across the 5 publications. For the most part, there is a lot of Internet content written in the languages we used in the pilot study, and Google Translate is likely to have been trained extensively in French, German, and Mandarin. Browser-based translation tools may not perform as well with languages that have a smaller Internet presence. Although multiple alternative tools are available for automatic translation (DeepL, Microsoft Translate, and ChatGPT), Google Translate was the only software used in this study. Although all these popular translation tools employ machine learning techniques (a subset of AI),[27-29] there is currently a lot of public interest in ChatGPT and future advancements in similar large language models. Recent studies suggest that the current iteration of ChatGPT does not yet consistently outperform Google Translate or Microsoft Translator, and performs worse with less widely spoken languages.[30] Although there is a lot of excitement about the future of ChatGPT in many fields, it is still reliant on the availability of training data.

The results of this study may not easily be extrapolated to other publication types because we included only short text-based summaries. These were easy to handle using Google Translate, but this approach may be less practical for longer texts or articles in which a lot of the information is embedded in figures and tables. Although Google Translate does have capabilities for translating whole documents, the complexity of the document formatting can affect how successful this is.

Finally, a survey-based approach was used to assess translation quality. The quality assessment could be expanded to also include the reverse translation method, in which Google-translated abstracts and PLSs are translated back to English by a translator who has no knowledge of the original text, with the results compared with the original to check for equivalence of the wording.

Although there has been recent focus on making scientific information more widely available, there remains a gap for non-English speakers. Ideally, professional translations would be available for all English-language scientific research, but this is not a practical solution. It is possible that, with technological advances, it will get easier for non-English speakers to search for, find, translate, and understand literature originally published in English. In the meantime, we propose that the publishing community should increase its commitment to PLSs, including improving their discoverability (eg, ensuring correct indexing on PubMed). Wider availability and accessibility of these lay-friendly summaries should not only provide an accessibility benefit to English-speaking readers but should also improve access for non-English speakers by enhancing the accuracy of translations using browser-based tools.

Acknowledgements

We thank Nicolas Bertheleme, Claudia Brockmeyer, Shufei Song, and Maša Švent who were the language reviewers involved in the study. Development of this article was jointly funded by Ipsen and Oxford PharmaGenesis Ltd.

Author declarations and disclosures: *All authors made substantial contributions to study conception/design, or acquisition/analysis/interpretation of data; drafted the publication or revised it critically for important intellectual content; and gave their final approval of the publication. Claire Beeby is an employee of Oxford PharmaGenesis Ltd; Eleanor J. Raynsford and Charles Pollitt are employees of Ipsen. Oxford PharmaGenesis Ltd is a HealthScience Consultancy contracted to support Ipsen with medical communication and medical writing activities.*

Author contact: *claire.beeby@pharmagenesis.com*

References

1. Edgell C, Rosenberg A. Putting plain language summaries into perspective. *Curr Med Res Opin*. 2022;38(6):871-874. doi:10.1080/03007995.2022.2058812
2. Frank J, Foster R, Pagliari C. Open access publishing - noble intention, flawed reality. *Soc Sci Med*. 2023;317:115592. doi:10.1016/j.socscimed.2022.115592
3. Piwowar H, Priem J, Larivière V, et al. The state of OA: A large-scale analysis of the prevalence and impact of open access articles. *PeerJ*. 2018;6:e4375. doi:10.7717/peerj.4375
4. Bahji A, Acion L, Laslett AM, Adinoff B. Exclusion of the non-English-speaking world from the scientific literature: recommendations for change for addiction journals and publishers. *Nordisk Alkohol Nark*. 2022;40(1):6-13. doi:10.1177/14550725221102227
5. Rostadmo M, Strømme S, Nylenna M, et al. How well do doctors understand a scientific article in English when it is not

their first language? A randomised controlled trial. *BMJ Open*. 2021;11(6):e043444. doi:10.1136/bmjopen-2020-043444
6. Gulbrandsen P, Schroeder TV, Milerad J, Nylenna M. Paper or screen, mother tongue or English: which is better? A randomized trial. *JAMA*. 2002;287(21):2851-2853. doi:10.1001/jama.287.21.2851
7. Fung ICH. Open access for the non-English-speaking world: overcoming the language barrier. *Emerg Themes Epidemiol*. 2008;5(1):1. doi:10.1186/1742-7622-5-1
8. Steigerwald E, Ramírez-Castañeda V, Brandt DYC, et al. Overcoming language barriers in academia: machine translation tools and a vision for a multilingual future. *BioScience*. 2022;72(10):988-998. doi:10.1093/biosci/biac062
9. Beeby C, Thomlinson V, Raynsford E, Pollitt C. P05 - accessibility of scientific information for non-English speakers: using browser-based tools to translate plain language summaries (PLS) and abstracts. Presented at: 2023 European Meeting of ISMPP; January 24-25, 2023; London, United Kingdom.
10. Beeby C, Thomlinson V, Raynsford E, Pollitt C. P18 - accessibility of scientific information for non-English speakers: using browser-based tools to translate plain language summaries (PLS) and abstracts. Paper presented at: 19th Annual Meeting of ISMPP; April 24-26, 2023; Washington, DC.
11. Kelley RK, Mollon P, Blanc JF, et al. Comparative efficacy of cabozantinib and regorafenib for advanced hepatocellular carcinoma. *Adv Ther*. 2020;37(6):2678-2695. doi:10.1007/s12325-020-01378-y
12. Danchenko N, Johnston KM, Whalen J. The cost-effectiveness of abobotulinumtoxinA (Dysport) and onabotulinumtoxina (Botox) for managing spasticity of the upper and lower limbs, and cervical dystonia. *J Med Econ*. 2022;25(1):919-929. doi:10.1080/13696998.2022.2092354
13. Paz-Ares L, Spigel DR, Chen Y, et al. RESILIENT part 1: a phase 2 dose-exploration and dose-expansion study of second-line liposomal irinotecan in adults with small cell lung cancer. *Cancer*. 2022;128(9):1801-1811. doi: 10.1002/cncr.34123
14. Turner-Stokes L, Fheodoroff K, Jacinto J, et al. The spasticity-related quality of life 6-dimensions instrument in upper-limb spasticity: part I development and responsiveness. *J Rehabil Med*. 2022;54:jrm00244. doi:10.2340/jrm.v53.690
15. Adelman DT, Van Genechten D, Megret CM, Truong Thanh XT, Hand P, Martin WA. Co-creation of a lanreotide autogel/depot syringe for the treatment of acromegaly and neuroendocrine tumours through collaborative human factor studies. *Adv Ther*. 2019;36(12):3409-3423. doi: 10.1007/s12325-019-01112-3
16. Desktop browser market share worldwide. StatsCounter. Published 2023. Accessed May 23, 2023. https://gs.statcounter.com/browser-market-share/desktop/worldwide
17. Likert R. A technique for the measurement of attitudes. *Arch Psychol*. 1932;22(140):5-55.
18. Rosenberg A, Baróniková S, Feighery L, et al. Open Pharma recommendations for plain language summaries of peer-reviewed medical journal publications. *Curr Med Res Opin*. 2021;37(11):2015-2016. doi:10.1080/03007995.2021.1971185
19. Dormer L, Walker J. Plain language summary of publication articles: helping disseminate published scientific articles to patients. *Future Oncol*. 2020;16(25):1873-1874. doi:10.2217/fon-2020-0784
20. Modi S. Trastuzumab deruxtecan in previously treated HER2-positive metastatic breast cancer: plain language summary of the DESTINY-Breast01 study. *Future Oncol*. 2021;17(26):3415-3424. doi:10.2217/fon-2021-0427
21. Oaknin A, Tinker AV, Gilbert L, et al. Clinical activity and safety of the anti-PD-1 monoclonal antibody dostarlimab for patients with recurrent or advanced dMMR endometrial cancer. *Future Oncol*. 2021;17(29):3781-3785. doi:10.2217/fon-2021-0598
22. Fizazi K, Blue I, Nowak JT. Darolutamide and survival in nonmetastatic, castration-resistant prostate cancer: a patient perspective of the ARAMIS trial. *Future Oncol*. 2021;17(14):1699-1707. doi:10.2217/fon-2020-1291
23. Savarirayan R, Irving M, Hoover-Fong J, et al. Vosoritide treatment accelerates bone growth in children with achondroplasia. *Future Rare Dis*. 2021;1(3):FRD14. doi:10.2217/frd-2021-0009
24. Plain language summaries. Future Medicine Group. Accessed May 5, 2023. https://www.futuremedicine.com/plainlanguagesummaries
25. PubMed user guide. PubMed. Published 2023. Accessed May 5, 2023. https://pubmed.ncbi.nlm.nih.gov/help/
26. Schuers M, Joulakian M, Kerdelhué G, et al. Lost in translation? A multilingual query builder improves the quality of PubMed queries: a randomised controlled trial. *BMC Med Inform Decis Mak*. 2017;17(1):94. doi:10.1186/s12911-017-0490-9
27. Translate and communicate with ease. DeepL. Accessed May 5, 2023. https://www.deepl.com/en/why-deepl-pro
28. Neural machine translation. Microsoft. Accessed May 5, 2023. https://www.microsoft.com/en-us/research/project/neural-machine-translation/
29. A neural network for machine translation, at production scale. Google Research. Published September 27, 2016. Accessed May 5, 2023. https://ai.googleblog.com/2016/09/a-neural-network-for-machine.html
30. Peng K, Ding L, Zhong Q, et al. Towards making the most of ChatGPT for machine translation. *arXiv*. Preprint posted online March 24, 2023. doi:10.48550/arXiv.2303.13780