

# Assessing PubMed metatag usage for plain language summary discoverability

Adeline Rosenberg<sup>a</sup>, Slávka Baróniková<sup>b</sup>, William Gattrell<sup>c</sup>, Namit Ghildyal<sup>d</sup>, Tim J Koder<sup>a</sup>, Taija Koskenkorva<sup>e</sup>, Andrew Liew<sup>f</sup>, Radha Narayan<sup>g</sup>, Joana Osório<sup>a</sup>, Valérie Philippon<sup>h</sup>, Melissa Shane<sup>a</sup>, Catherine Skobe<sup>i</sup>, Kim Wager<sup>a</sup>

<sup>a</sup>Oxford PharmaGenesis Ltd, Oxford, UK; <sup>b</sup>Galápagos NV, Mechelen, Belgium; <sup>c</sup>Ipsen, Abingdon, UK; <sup>d</sup>Janssen Global Services LLC, Raritan, NJ, USA; <sup>e</sup>Novartis Pharma AG, Basel, Switzerland; <sup>f</sup>Oxford PharmaGenesis Pty Ltd, Melbourne, Australia; <sup>g</sup>Alexion, AstraZeneca Rare Diseases, Boston, MA, USA; <sup>h</sup>UCB, Cambridge, MA, USA; <sup>i</sup>Pfizer Inc., New York, NY, USA

\*Affiliation at the time of original abstract development



## PLAIN LANGUAGE SUMMARY OF THIS POSTER

PubMed is a website that files scientific research articles and shows their abstracts. PubMed can also show plain language summaries (PLS) of these articles when publishers tag the PLS correctly. PLS and open access, or free-to-read, publishing can help readers to find and use published research.

In this study, we downloaded all of the available articles on PubMed and searched for those with a PLS tag. We found that there were 3217 articles with PLS tags on PubMed. To check the accuracy of the PLS tags, we used a computer program to count how many of the PLS tags were correct. We confirmed the results manually. We found that 14.6% (470) of the tags were not tagging PLS but instead tagged:

- non-English language abstracts
- copies of the scientific abstract
- empty content
- other types of content, such as website links or article summary bullet points.

Next, we looked at how many of the journals with correctly tagged PLS were open access. All 105 journals were fully open access or had open access options. We also looked at how many of the individual articles were open access. Of the 2747 articles with correctly tagged PLS, 77.7% (2135) were open access.

Overall, our results suggest that publishers need more guidance on how to correctly use the PLS tag on PubMed so that the PLS can be found. This is important because tagging PLS correctly can help publishers to increase the impact of an article, so that it can be found and used by readers.

## BACKGROUND

- PubMed is one of the most widely used platforms for accessing biomedical research.<sup>1</sup>
- A broader audience can discover text-based, concise plain language summaries (PLS) that are hosted on PubMed when they are tagged correctly.<sup>1</sup>
  - This function was introduced in 2019 and allows retrospective tagging of pre-2019 records.<sup>2</sup>
- Open access (OA) publishing can also enhance discoverability, which increases publication accessibility and usage.<sup>3</sup>

## OBJECTIVES

We aimed to:

- determine the proportion of PubMed records correctly using the PLS tag and the reasons for incorrect usage
- establish the journal-level and article-level OA status of records with PLS on PubMed.

## DESIGN

### PLS TAGGING

- The entire PubMed database was downloaded (up to **9 February 2022**) and searched for PLS indexed with an Extensible Markup Language (XML) <plain-language-summary> tag in the 'Other Abstract' field.
- Records were deduplicated, and incorrectly tagged PLS were programmatically excluded for improper tag usage (i.e. non-PLS content) and confirmed with manual spot checks.

### OA STATUS

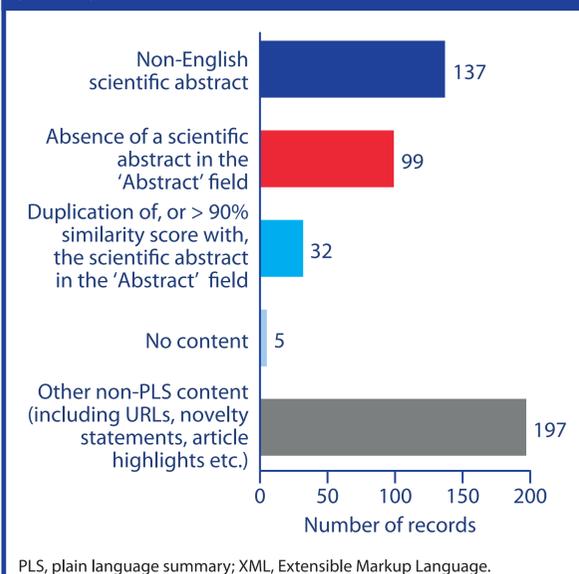
- Correctly tagged PLS were categorized by journal and assessed for overall OA status using Journal Selector (Sylogent LLC, Bristol, PA, USA) or using information on journal websites for those not indexed on Journal Selector.
- Article-level OA status was assessed using the Simple Query Tool from Unpaywall (OurResearch, Sanford, NC, USA).

## RESULTS

### PLS TAGGING

- Out of the entire PubMed database of 31,817,472 records, only 3217 (0.01%) had an XML <plain-language-summary> tag in the 'Other Abstract' field, of which slightly over half (1644 [51.1%]) were published in 2021.
- Of the 3217 records, 470 (14.6%) used the <plain-language-summary> tag incorrectly (**Figure 1**) and 2747 records (85.4%) used the tag correctly.
  - This represents an incidence rate of true PLS for all of 2021 of 929 per 1,000,000 records (n = 1644/1,769,389).
- Within the 2747 correctly tagged records, there were 124 records using the <plain-language-summary> tag to index both non-English scientific abstracts and English PLS, in addition to an English scientific abstract in the 'Abstract' field.

**Figure 1:** Categories of incorrect usage of the XML <plain-language-summary> tag in the 'Other Abstract' field (n = 470).



### OA STATUS

- The 2747 records correctly using the <plain-language-summary> tag for PLS were published in 105 journals.
  - All (100%) of these journals were full/gold OA journals or offered OA options, and none were closed/subscription only.
  - Of the 105 journals, 30 (28.6%) were full/gold OA journals and 75 (71.4%) offered OA options.
- At the article level, 2135 of these records (77.7%) were available under some form of OA licence (**Table 1**).
  - Of these 2135 OA records, 1593 (58.0%) were published in the 30 full/gold OA journals.
  - Of the 1154 articles published in the other 75 journals with OA options, only 542 (47.0%) were in fact published under an OA licence.

**Table 1:** Article-level OA status (n = 2747).

OA status	Number of records, n (%)
Open	2135 (77.7)
<b>Gold</b>	<b>1593 (58.0)</b>
Hybrid	409 (14.9)
Green	69 (2.5)
Bronze	64 (2.3)
Paywalled	607 (22.1)
Unknown	5 (0.2)

OA, open access.

## CONCLUSIONS

- Despite the use of the <plain-language-summary> tag increasing over time,<sup>4</sup> records using this tag represent a very small minority of all PubMed records (0.01%), and the tag is commonly used incorrectly for several reasons.
- **There is an unmet need for explicit guidance on both the processes of indexing and the correct usage of the <plain-language-summary> tag, which could help improve uptake and correct tagging.**
- To date, all PLS available on PubMed are published in journals with OA options, and more than half are published in full/gold OA journals. These journals are likely to benefit from a PLS through increased discoverability and publication accessibility.
- Limitations of this analysis include a lack of PLS quality assessment and small sample size, largely due to low publisher uptake and correct tagging.
- **Ultimately, these findings highlight an opportunity for publishers to increase the impact of their journals' content and reach a broader audience by ensuring correct PubMed tagging as they expand their PLS offerings.**

## REFERENCES

1. Ossom Williamson P, Minter CJ. *J Med Libr Assoc* 2019;107:16–29.
2. Collins M. *NLM Tech Bull* 2019;(426):e4.
3. Li H et al. *J Informetrics* 2021;15:101154.
4. Gattrell W et al. *Curr Med Res Opin* 2022;38(sup1):25–45.

## ACKNOWLEDGEMENTS

This abstract is an encore of the original analysis, which was presented in a poster at the 9th International Congress on Peer Review and Scientific Publication, 8–10 September 2022, Chicago, IL, USA, available online via the QR code above. Editorial support was provided by Charlotte Watson of Oxford PharmaGenesis, Oxford, UK. The authors also thank Sarah Griffiths and Ishmam Nawar of Oxford PharmaGenesis, Oxford, UK for reviewing a version of this poster.

## CONFLICTS OF INTEREST DISCLOSURES

At the time of original abstract development, WG was an employee and shareholder of, or held stock or stock options in, Ipsen, Abingdon, UK and is now an employee of Bristol Myers Squibb, Uxbridge, UK. All other authors are employees of their respective affiliation, and TJK, SB, NG, RN and CS may be shareholders, or hold stock or stock options, for their respective affiliations.

## ORCID iDs

AR, 0000-0003-4599-4291; SB, 0000-0001-8072-5690; WG, 0000-0002-9733-6330; NG, 0000-0002-3908-3507; TJK, 0000-0001-6152-7365; TK, 0000-0002-7349-2435; AL, 0000-0003-4792-9637; RN, not available; JO, 0000-0003-2713-5370; VP, 0000-0002-7444-6027; MS, 0000-0002-6590-8904; CS, 0000-0002-5220-2566; KW, 0000-0001-8107-9600.

## FUNDING/SUPPORT

This work was produced by representatives of Open Pharma. Open Pharma is a multi-sponsor collaboration facilitated by Oxford PharmaGenesis and receives sponsorship funding from AstraZeneca Pharmaceuticals LP, Boehringer Ingelheim International GmbH, ER Squibb & Sons LLC, F. Hoffmann-La Roche AG, Galápagos NV, Gilead Sciences, Inc., GlaxoSmithKline Biologicals SA, Ipsen Biopharm Ltd, Janssen Global Services LLC, John Wiley & Sons Ltd, Novartis Pharma AG, Novo Nordisk A/S, Pfizer Inc., Takeda Development Center Americas, Inc. and UCB Biopharma SRL.

## ROLE OF FUNDER/SPONSOR STATEMENT

This poster represents the work of the individual authors and not of their respective affiliations. The organizations providing sponsorship funding to Open Pharma were not involved in the design and conduct of the study, collection, management, analysis and interpretation of the data, preparation, review, or approval of the abstract or poster, or decision to submit the abstract for presentation.