

Scalability of Meta-Research on Patient-Authored Publications: You Can Do a Lot With a Bot

Bella Yang^a, Dawn Lobban^{b,*}, Karen Woolley^{c,d,e}

^aEnvision Pharma Group, Philadelphia, PA, USA; ^bEnvision Pharma Group, Wilmslow, UK; ^cAdjunct Professor, University of Queensland, Brisbane, Australia; ^dAdjunct Professor, University of the Sunshine Coast, Sippy Downs, Australia; ^eCaregiver author
*At the time of this research

Background & Objective

- As patient authorship increases, publication professionals may face increasing responsibilities in identifying, guiding, and measuring the output of patient authors¹
- Meta-research² on patient-authored publications could help publication professionals meet these responsibilities. This research is feasible but challenging to scale³
- Our objective was to develop a bot to enable scalability of meta-research on patient-authored publications

Methods

- To enable the bot to rapidly extract relevant text from PubMed’s data fields, we developed Natural Language Processing (NLP) algorithms, which leveraged the pretrained Name Entity Recognition (NER) models from ScispaCy and GPT-3.5 model from OpenAI
- We tested the bot on a dataset of patient-authored publications, identified via PubMed’s advanced affiliation search function. The Boolean search query was “patient author[Affiliation] OR patient partner[Affiliation] OR carer[Affiliation] OR caregiver[Affiliation]”. The summary of the analysis below is based on publications that were published from 1994–2022 (N=603)
- To identify bibliographic metrics of patient authors, other databases and tools (eg, Gender API, Worldbank, SpaCy, and Pycountry) were used in the algorithm of the bot
- Analysis was done using Microsoft Power BI to automate data preprocessing

Results

1. Was the Scalability of Meta-Research on Patient-Authored Publications Possible?

- Scalability of research on patient authorship is now possible, given the automated features of the bot and the reduction in manual processing. The bot will be able to analyze increasingly larger datasets as patient authorship increases
- Results from the bot-enabled rapid visualization are shown through an interactive dashboard



References

1. Arnstein L, Wadsworth AC, Yamamoto BA, et al. *Res Involv Engagem.* 2020;6:34. 2. Ioannidis JPA. *PLoS Biol.* 2018;16:e2005468. 3. Oliver J, Lobban D, Dormer L, et al. *Res Involv Engagem.* 2022;8:12.

Disclosures and Acknowledgments

All authors participated in the research, were actively involved in preparing the abstract, and provided approval for submission. BY is an employee and shareholder of Envision Pharma Group, a healthcare communications company that is funded by pharmaceutical companies. KW is a consultant to, and shareholder in, Envision Pharma Group. DL was an employee of, and owns shares in, Envision Pharma Group. DL serves on the ISMPP Patient Engagement Task Force. No payments were made for authorship, but KW was compensated for medical writing tasks. This encore abstract was previously presented at the 19th Annual Meeting of ISMPP (24-26 April 2023; Washington, DC, USA).

Presented at the International Society for Medical Publication Professionals (ISMPP) 2024 European Meeting (23-24 January 2024; London, UK)

Results (continued)

2. How Efficient and Accurate Was the Bot?

- The data extraction using the bot was 66-fold faster than existing methods, demonstrating the overall value of the bot
- The accuracy of the bot in identifying patient authors was greater than 95% based on a manually extracted random sample of 100 publications



3. What Characteristics of Patient-Authored Publications Could the Bot Identify?

- The bot was able to identify all the key characteristics of patient-authored publications and patient authors that have been identified via a manual (human) process. However, data extraction with the bot was more rapid, scalable, and without the risk of human error

Patient-authored publication characteristics

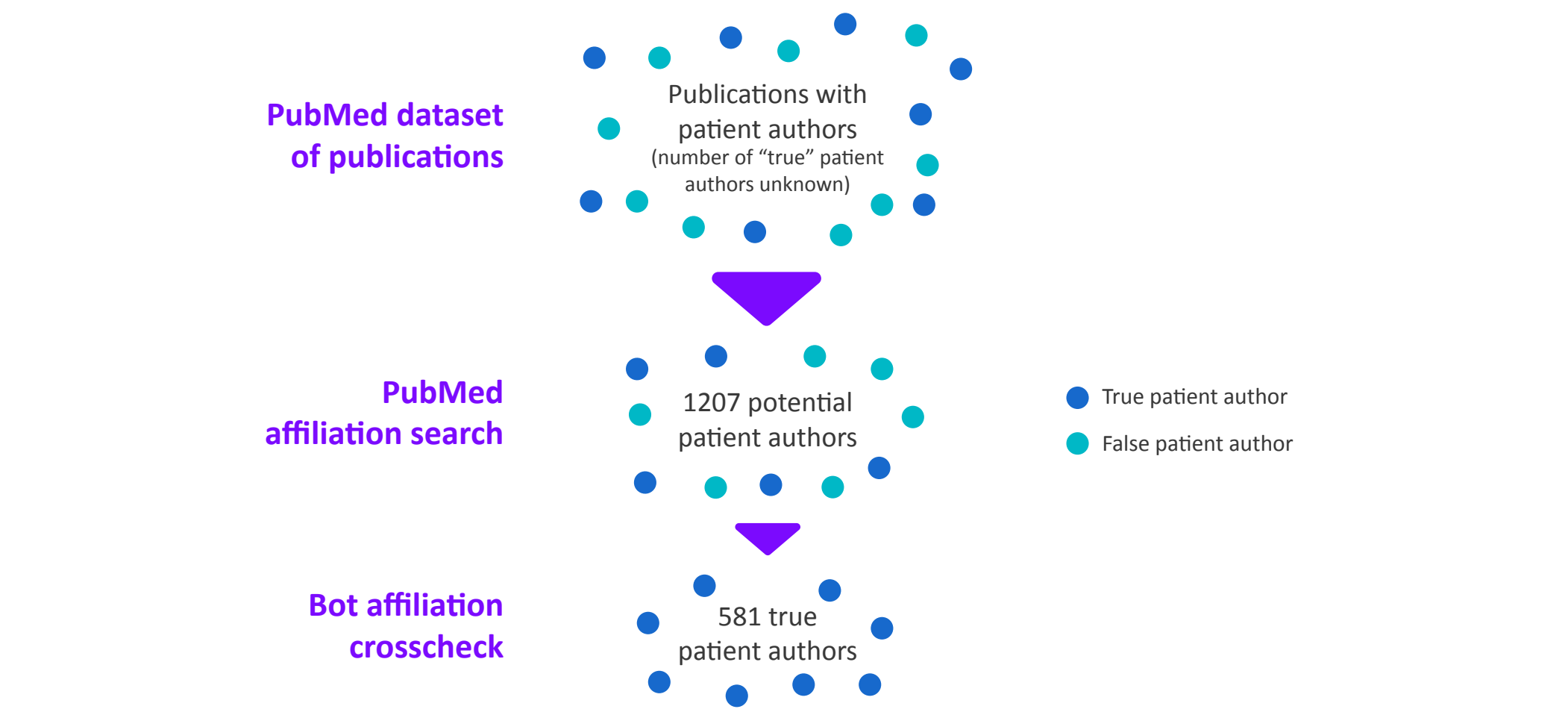
Characteristic	Rapid and reliable extraction/analysis via bot
Number of patient-authored publications	✓
Publication type	✓
Therapeutic area	✓
Language	✓
Open access	✓
Plain language summary	✓
Retraction	✓

Patient author characteristics

Characteristic	Rapid and reliable extraction/analysis via bot
Name	✓
Affiliation	✓
Author order	✓
Gender estimation	✓
Country	✓
Country income classification	✓

4. How Did the Bot Clean the Data?

- PubMed’s advanced search function for affiliation can be used to retrieve patient-authored publications³
- However, PubMed’s output can include false positives. A false positive is a publication that includes an author with a patient author affiliation term (eg, patient author, patient partner, carer, caregiver), but the author is not a patient author
- Examples of false positives include when an author’s affiliation is:
 - Included as part of a government organization (eg, Utah Caregiver Support Program, Utah Division of Aging & Adult Services, Salt Lake City)
 - An acronym for an academic organization (eg, The Center of Advanced Research and Education in Reproduction [CARER], Ben-Gurion University of the Negev, Beer Sheva, Israel)
- The bot was trained to rapidly clean the data by automatically checking for and removing false positives. Completing this task manually is time-intensive, has a risk of human error, and is not scalable
- The affiliation term that yielded the lowest number of false positives was “patient author”. Reaching consensus with diverse stakeholders on a set of standard terms for using PubMed’s affiliation search function would facilitate research on patient authorship. For example, “caregiver author” vs “caregiver” would reduce the number of false positives

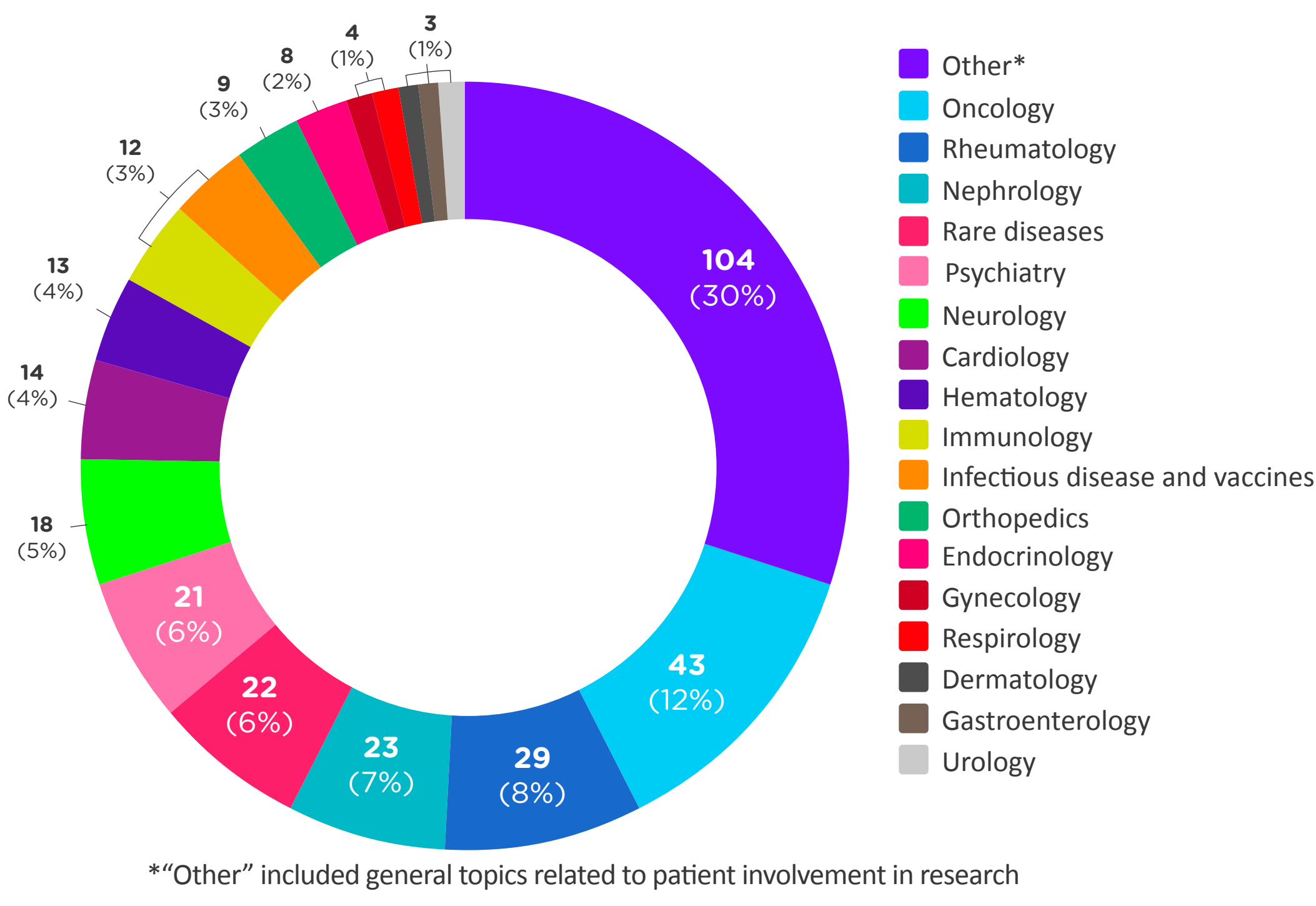


- Of the 603 publications identified in PubMed, 581 patient authors, patient partners, caregivers, or carers were identified by the bot. These 581 patients were involved in 327 publications

Results (continued)

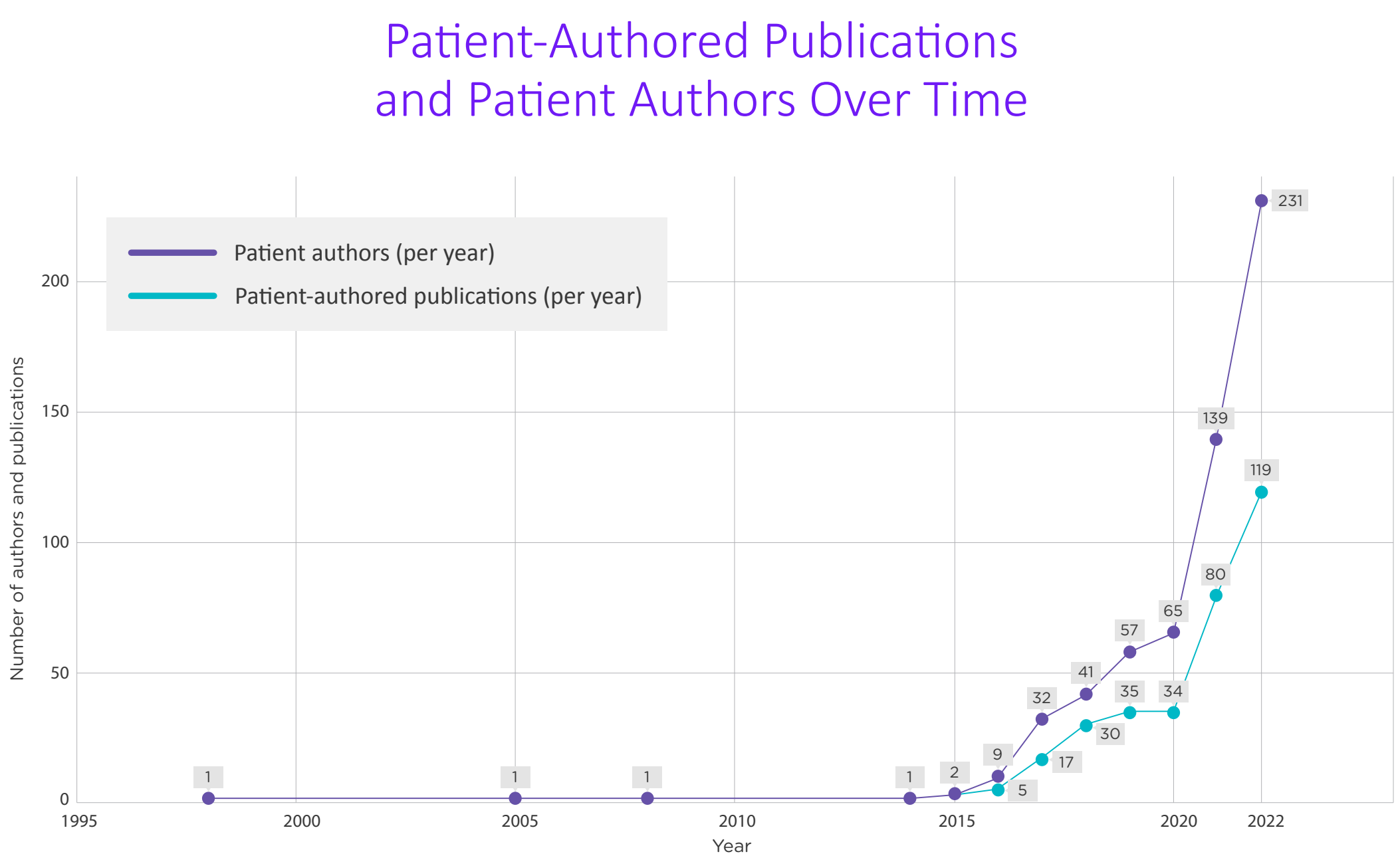
5. Which Therapeutic Area Is Most Common Among Patient-Authored Publications?

- The top three therapeutic areas that patient authors were involved in were oncology, rheumatology, and nephrology



6. How Prevalent and Accessible Are Patient-Authored Publications?

- Patients were more likely to be involved in open access publications (75%) than publications behind the paywall (25%)
- There was a rapid increase in patient authors and patient-authored publications in the past 5 years



Conclusions

Scalability of meta-research on patient authorship is possible with the bot’s ability to analyze a larger dataset with greater efficiency than the original methods

- With combinations of NLP and large language models such as GPT, the efficiency and accuracy of meta-research on patient authorship can be improved
- Publication professionals can use evidence-based insights gained from scalable meta-research to better understand the rapidly emerging field of patient-authored publications and to support patient-author partnerships

Please scan this QR code to view an interactive analysis report of this data

Please scan this QR code to hear an audio summary of the poster

Please scan this QR code to view a digital version of the poster