

Why did it go viral? An informatics-based case study of exaggerated language in news and social media

Hollie Rawlings,¹ Tomas Rees,¹ Lubaina Koti,² Avishek Pal,³ Andrew Liew¹

¹Oxford PharmaGenesis, Oxford, UK; ²Dasman Diabetes Institute, Kuwait City, Kuwait;

³Novartis, Basel, Switzerland



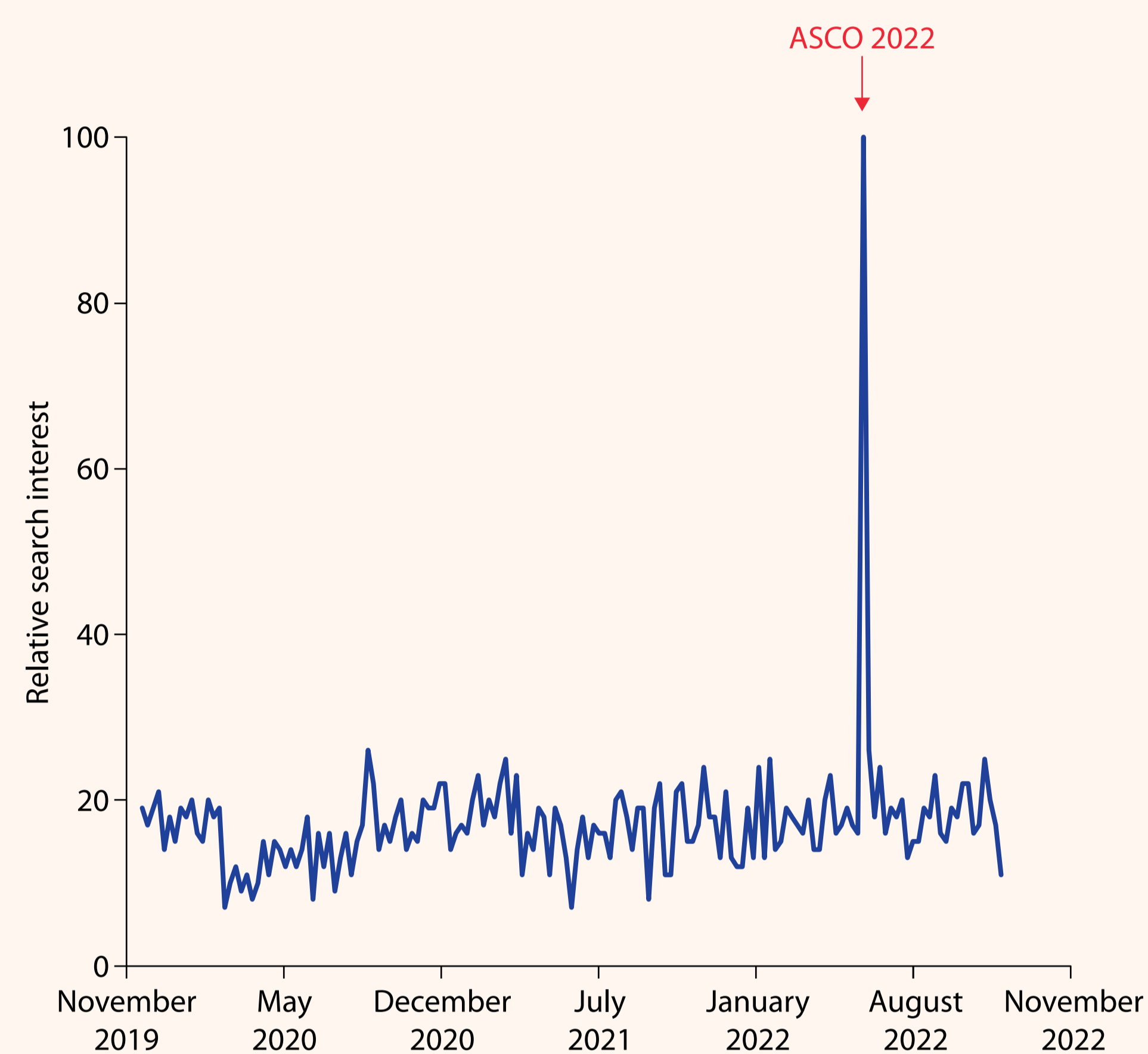
1. Introduction

- News stories and social media coverage about medical innovations can be exaggerated^{1,2} leading to 'viral' stories that may erode public trust in science and medicine.
- We selected a case study to examine this phenomenon: a study of a novel rectal cancer treatment in 12 patients presented at the 2022 American Society of Clinical Oncology (ASCO 2022) Annual Meeting, and simultaneously published in *NEJM*.³
- This publication generated extremely high news and social media interest, and so provides a rich source of media content.

2. Objective

- We sought to understand the triggers and independent channels of the dissemination of this study and how the viral language used by these independent channels differed from the scientific language used by the investigators/authors in the originals.

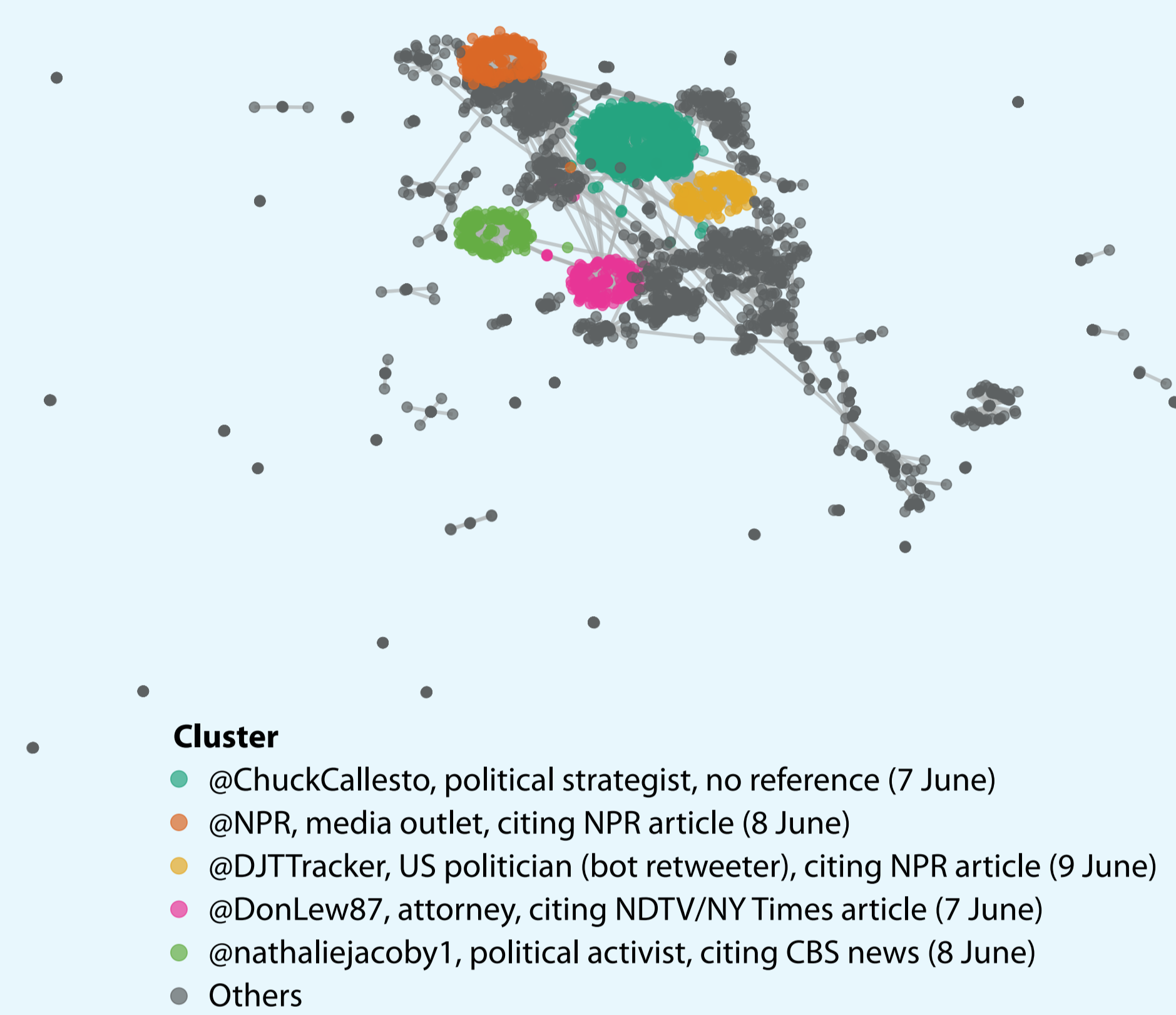
3. A presentation at ASCO 2022 generated very high interest in rectal cancer



How this was discovered

- Google trends data for 'rectal cancer' in the period 30 November 2019–30 November 2022.

4. Influential tweets came from politically affiliated accounts several days after initial publication

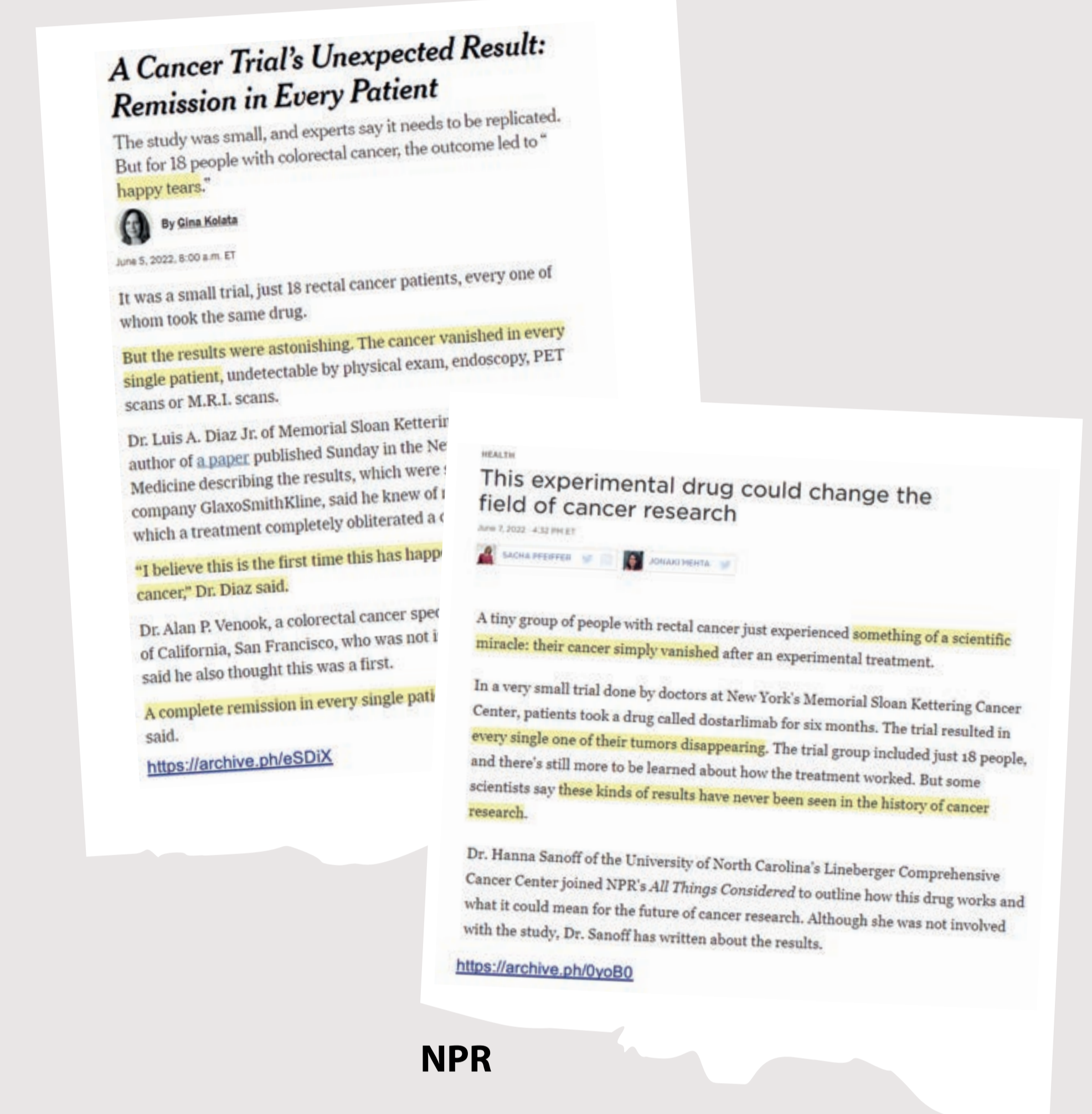


How this was discovered

- Twitter search for 'dostarlimab' and 'rectal cancer' (date range 5 June 2022–7 July 2022). Network cluster analysis based on retweets.

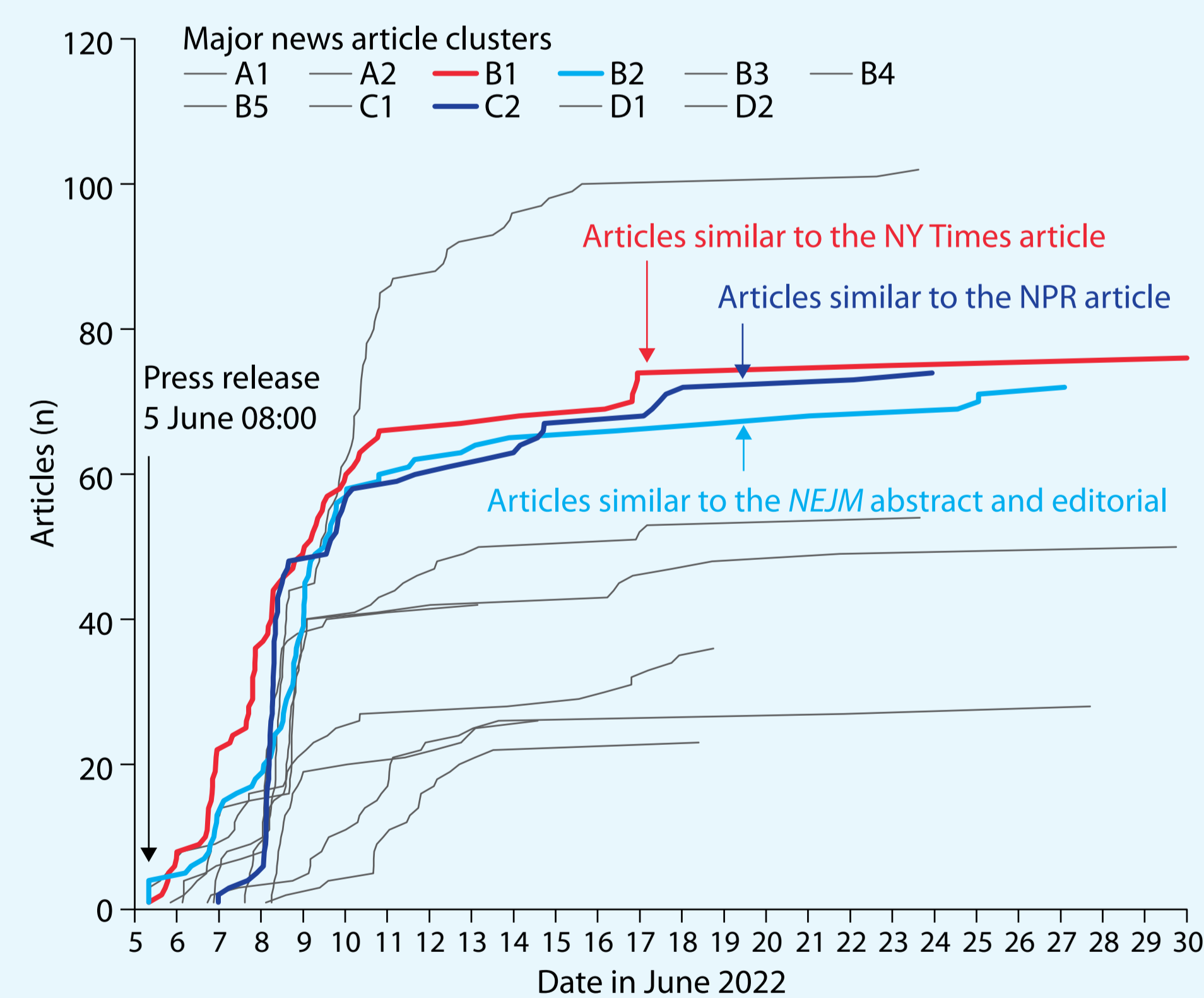
5. Media sources cited by key tweets used exaggerated language

The New York Times



NPR

6. Clustering of news/web articles found large numbers that shared the language of the media sources, and also ones that were more similar to the language of the *NEJM* abstract and editorial

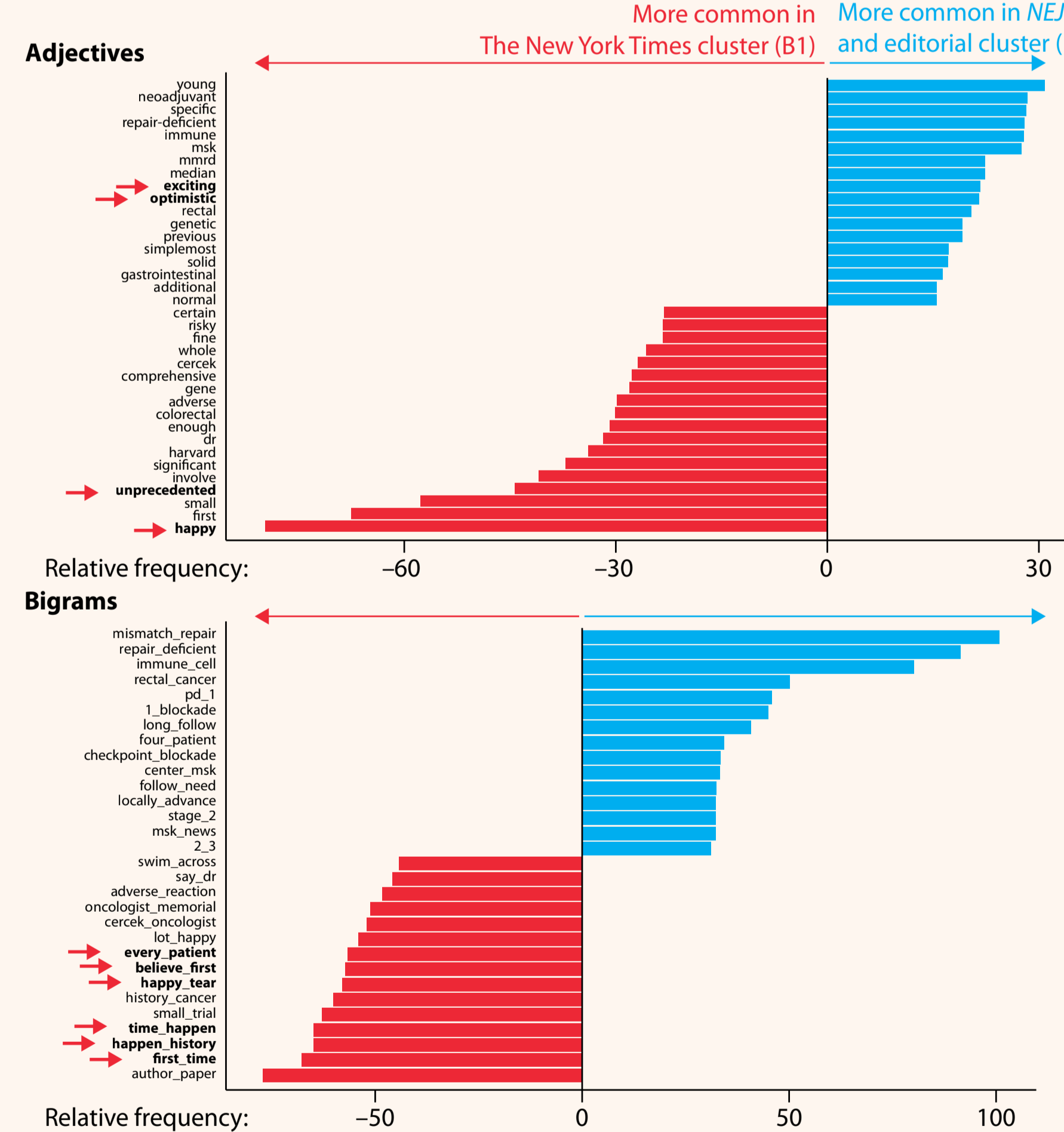


How this was discovered

- Feedly search for news/web articles on 'dostarlimab' and 'rectal cancer'.
- Key articles that were not found in the search were manually downloaded.
- Topic clustering was performed using BERTopic.

7. Comparison of word use between clusters of articles related to the media sources showed that they shared exaggerated language that is less common in more scientific sources

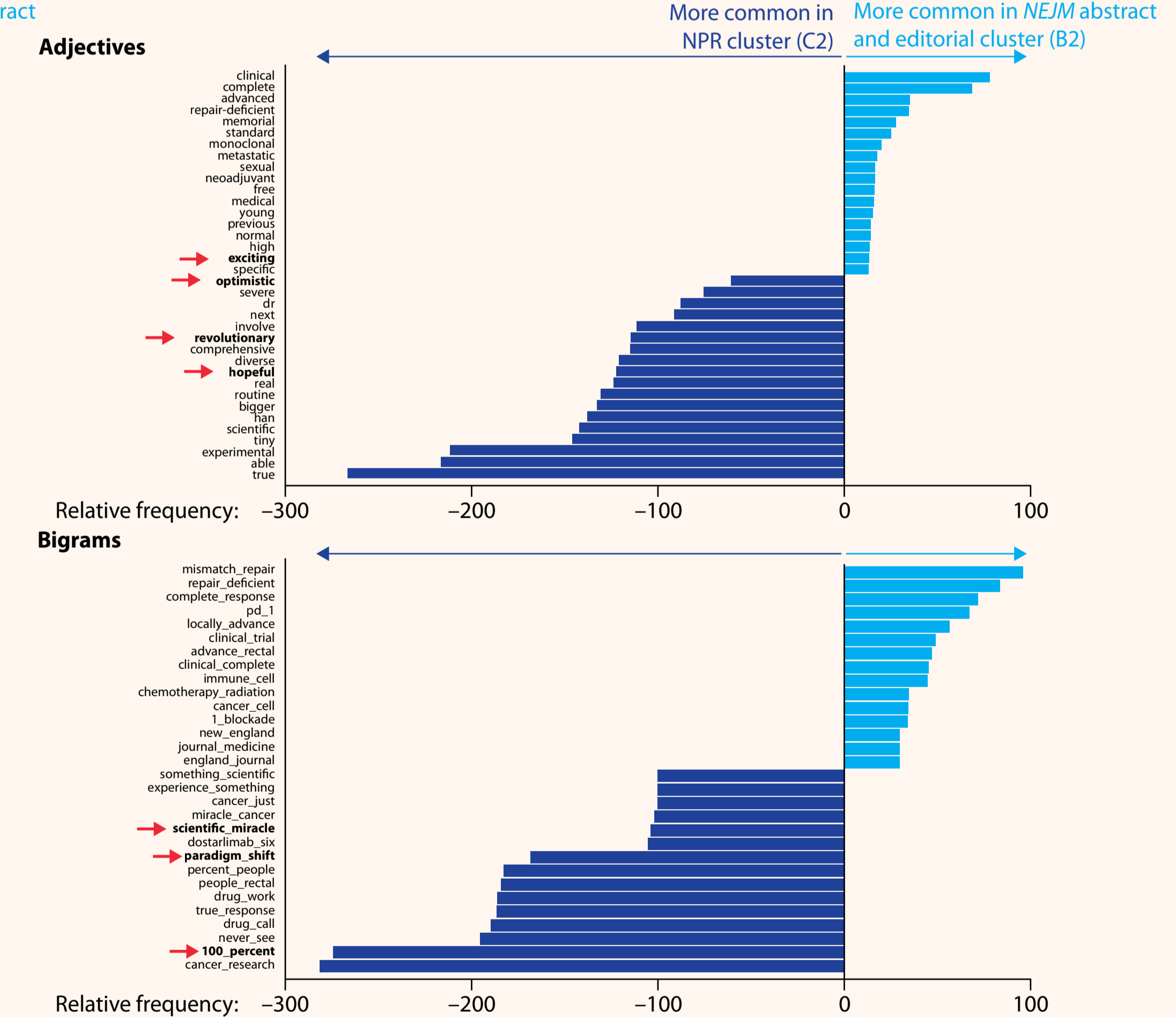
A. The New York Times cluster versus *NEJM* abstract and editorial cluster



How this was discovered

- Text cleaning removed non-English articles and standard stop-words and performed lemmatization.
- Network analysis of influential terms identified several articles with large amounts of irrelevant text – these articles were excluded from further analysis.
- Part-of-speech (pos) tagging was used to identify adjectives.

B. NPR cluster versus *NEJM* abstract and editorial cluster



Conclusions

- We found that news articles from major publishers used exaggerated language that was picked up by other media and by influential non-medical social media accounts.
- Exaggerated language used in some news articles, which substantially overemphasized the narrative of this trial, likely played a key role in the viral spread of information among the general public in news and social media.
- Plain language summaries can enable access to accurate, fair and balanced interpretation of medical research,⁴ which may counter the perils of exaggerating or extrapolated reporting.
- This AI approach combining natural language models with text analytics allows analysis of large volumes of text to gain insights into the spread of messages.

References

- Sumner P et al. *BMJ* 2014;349:g7015. 2. Suarez-Lledo V, Alvarez-Galvez J. *J Med Internet Res* 2021;23:e17187. 3. Cercek A et al. *N Engl J Med* 2022;386:2363–76. 4. Pushparajah DS et al. *Ther Innov Regul Sci* 2018;52:474–81.

Disclosures

HR, TR, AL: Employees of Oxford PharmaGenesis. LK: Nothing to disclose. AP: Employee of Novartis with no direct financial interest related to the assets or topics discussed in this poster.