An AI approach to analyze the context of publication citations

David Lickorish,[1] Tomas Rees,[2] Hollie Rawlings,[2] Eileen Burkart-Hartman,[1] Shailesh Desai[1]

[1]Takeda Pharmaceuticals U.S.A., Inc., Lexington, MA, USA; [2]Oxford PharmaGenesis, Oxford, UK

Scan the QR code to access the poster and any additional materials on your mobile device as well as forward yourself a link to it via email. If you do not have a QR reader, please enter https://tiny.one/uS01422Py2 in your browser.

Background

- Takeda US has been tracking the impact metrics of articles in publication portfolios for 2 years to provide insight into the variance of metrics across the portfolios; however, these quantitative metrics are limited.
- Artificial intelligence (AI)-powered research platforms that employ large language models (LLMs) are available to delve into the contents of engagement, which may provide additional insights into publication impact.

Objective

- To deepen our understanding of citation context and publication impact by combining conventional article-level impact metrics with AI-powered content analysis.
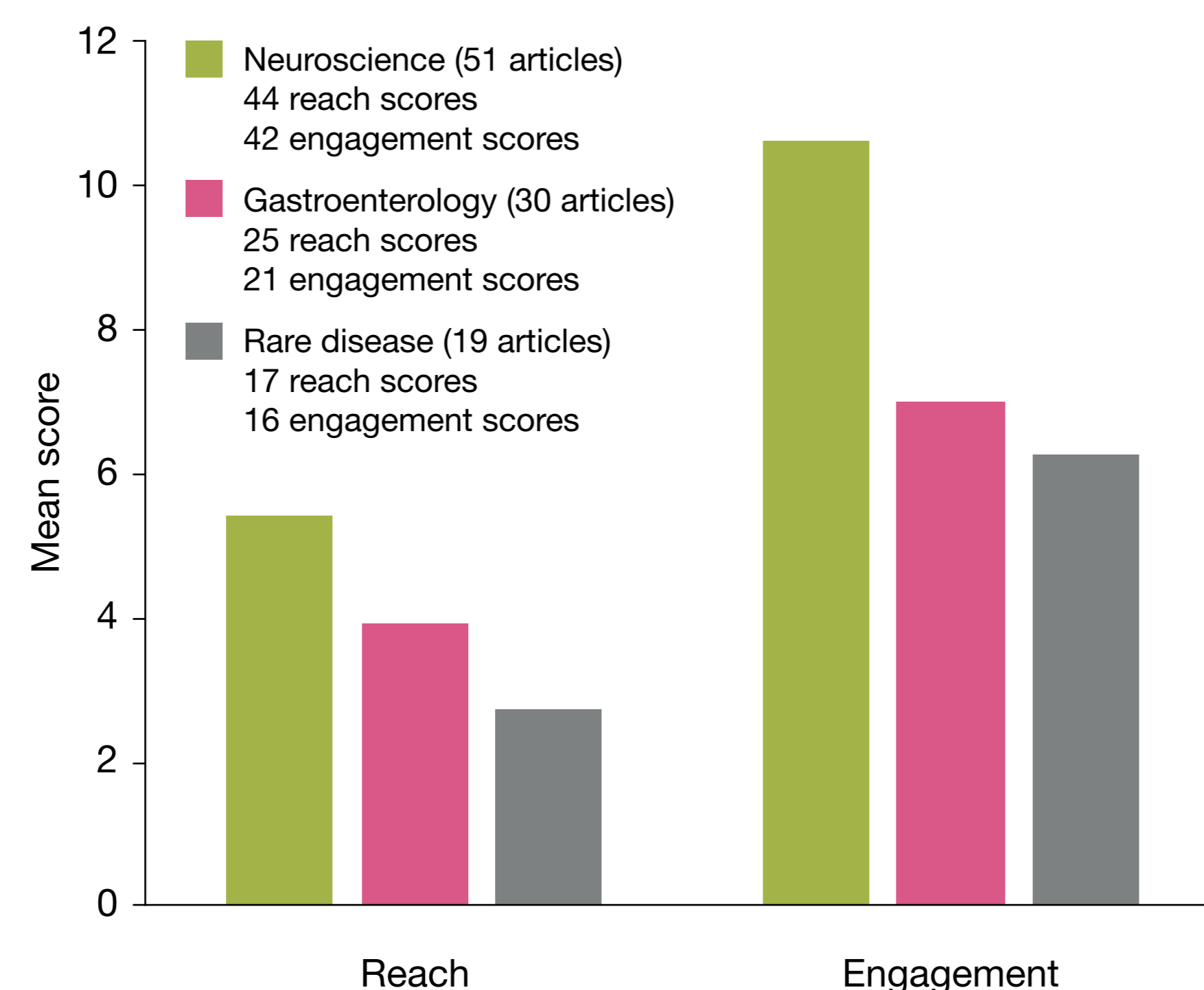
Methods

- Takeda US articles published between January 1, 2022 and November 30, 2023 in the neuroscience, gastroenterology and rare disease portfolios were monitored.
- Each article was assessed 30 days and 6 months after publication to provide reach and engagement scores, respectively.[1]
  - The reach score is derived from the number of mentions of the article by news outlets and social media, whereas the engagement score depends on the number of article citations and Mendeley saves.
- Citation statements for each article were retrieved using scite (https://scite.ai/) and analyzed using a custom AI-powered tool that employed an LLM in a secure environment.
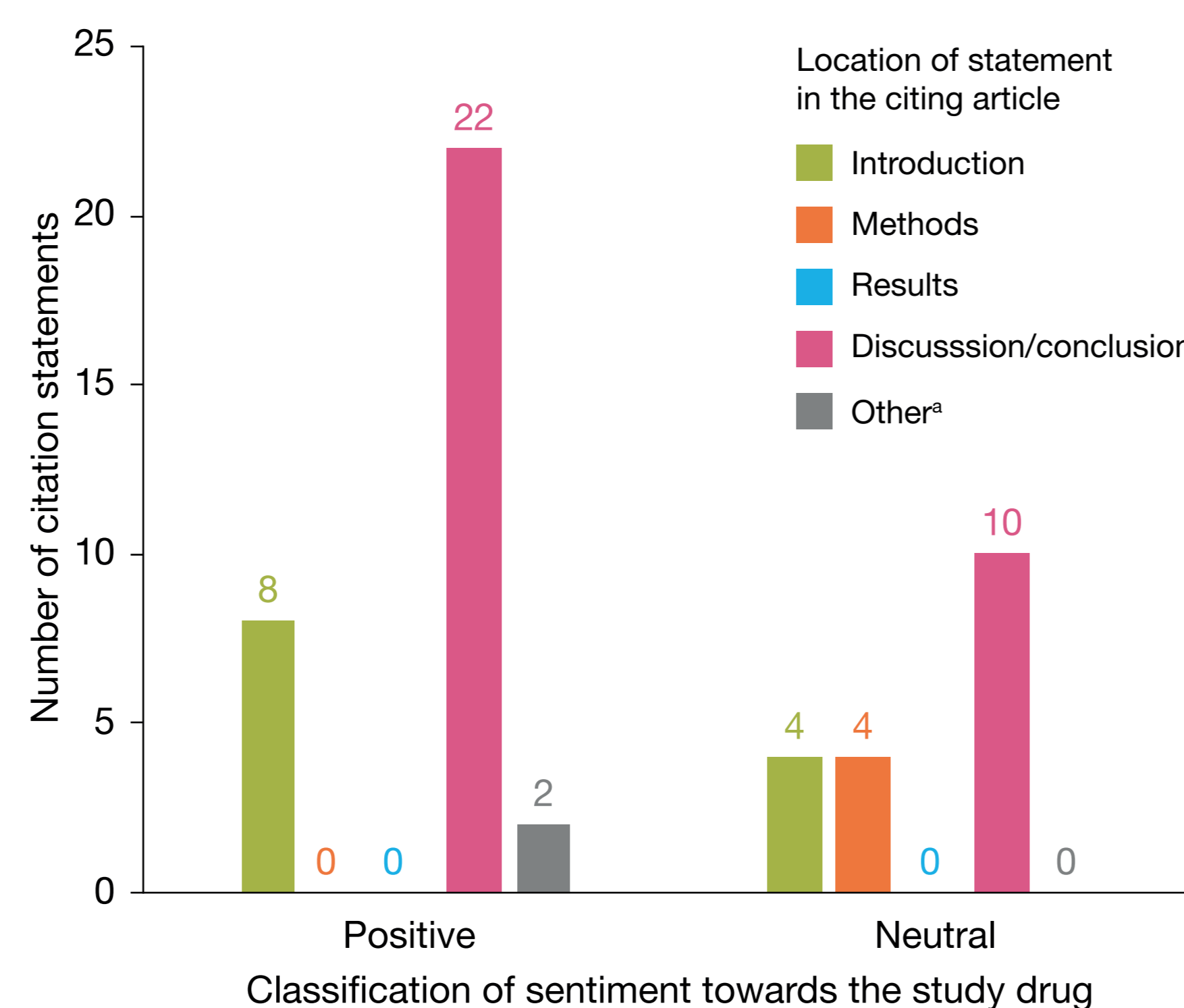
Results

- Overall, 100 published articles from the portfolios of neuroscience (51), gastroenterology (30) and rare disease (19) were analyzed.
- During the period of monitoring, there were 29 news mentions, 250 article citations and 1084 Mendeley saves related to the articles across the portfolios.
- At the cutoff of November 30, 2023, there were 86 reach scores available (44, 25 and 17 scores for the neuroscience, gastroenterology and rare disease portfolios, respectively) and 79 engagement scores available (42, 21 and 16 scores for the neuroscience, gastroenterology and rare disease portfolios, respectively).
- Mean reach and engagement scores were highest for articles in the neuroscience portfolio (**Figure 1**).

Figure 1. Mean reach and engagement scores of articles in the neuroscience, gastroenterology and rare disease portfolios.

Neuroscience (51 articles)
44 reach scores
42 engagement scores

Gastroenterology (30 articles)
25 reach scores
21 engagement scores

Rare disease (19 articles)
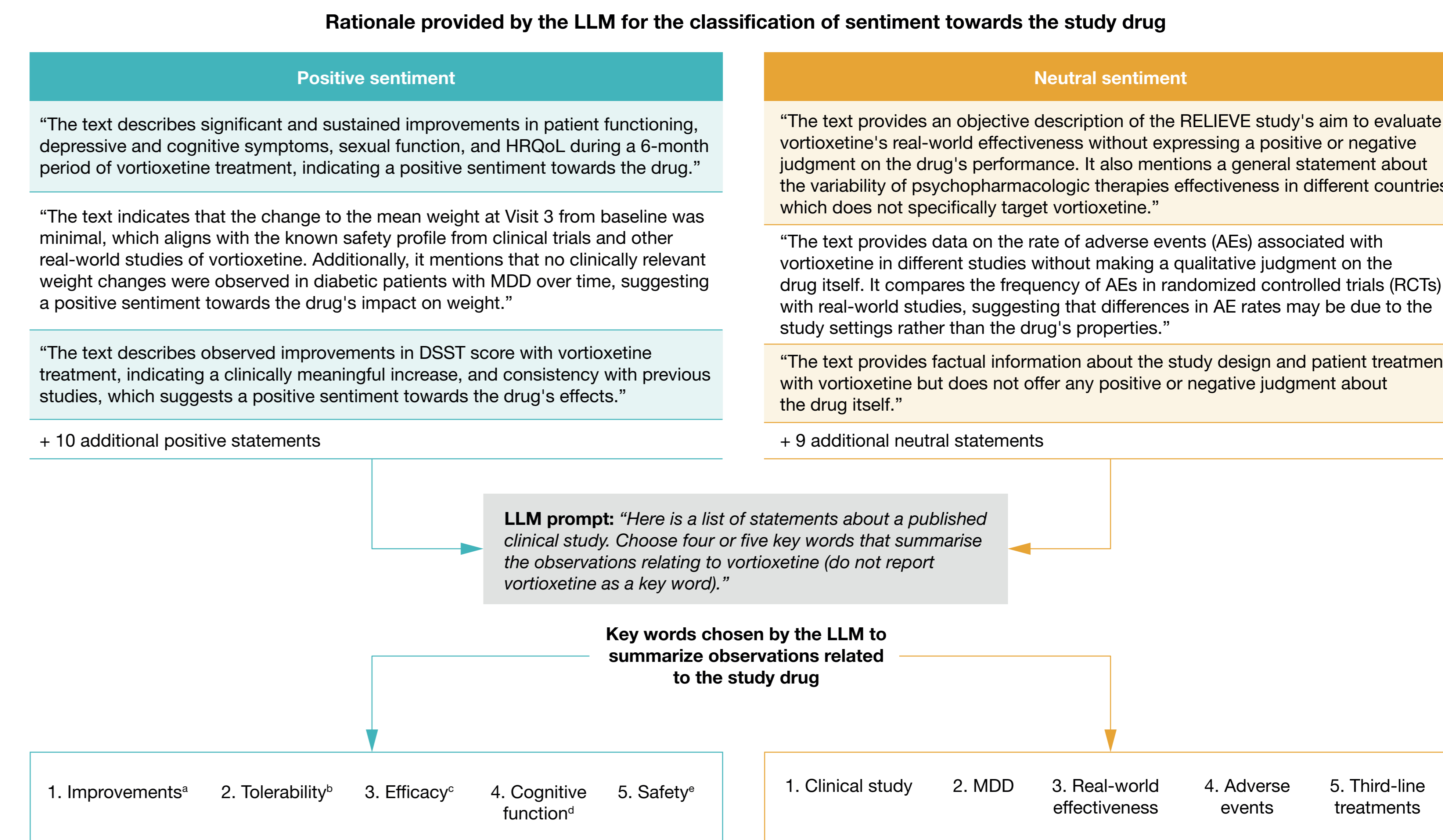17 reach scores
16 engagement scores

- Overall, 105 citation statements were available for 24 articles across the neuroscience, gastroenterology and rare disease portfolios.
- Scite identified 16 supportive citation statements, all of which were for six articles in the therapy area of major depressive disorders (MDDs) in the neuroscience portfolio.
- The LLM identified 50 citation statements mentioning a study drug, mostly located in the discussion or conclusion of the citing articles, and classified the sentiments towards the study drug as positive (n = 32) or neutral (n = 18) (**Figure 2**).

Figure 2. Classification of the sentiments of 50 citation statements towards a study drug as positive or neutral.

Location of statement in the citing article
- Introduction
- Methods
- Results
- Discusssion/conclusion
- Other[a]

[a]'Other' includes one citation statement that was located in an editorial without any headers and another citation statement that was located under a header that followed the conclusion.

Figure 3. Using an LLM to assess and to summarize citations of a clinical trial publication.

Rationale provided by the LLM for the classification of sentiment towards the study drug

Positive sentiment	Neutral sentiment
"The text describes significant and sustained improvements in patient functioning, depressive and cognitive symptoms, sexual function, and HRQoL during a 6-month period of vortioxetine treatment, indicating a positive sentiment towards the drug."	"The text provides an objective description of the RELIEVE study's aim to evaluate vortioxetine's real-world effectiveness without expressing a positive or negative judgment on the drug's performance. It also mentions a general statement about the variability of psychopharmacologic therapies effectiveness in different countries, which does not specifically target vortioxetine."
"The text indicates that the change to the mean weight at Visit 3 from baseline was minimal, which aligns with the known safety profile from clinical trials and other real-world studies of vortioxetine. Additionally, it mentions that no clinically relevant weight changes were observed in diabetic patients with MDD over time, suggesting a positive sentiment towards the drug's impact on weight."	"The text provides data on the rate of adverse events (AEs) associated with vortioxetine in different studies without making a qualitative judgment on the drug itself. It compares the frequency of AEs in randomized controlled trials (RCTs) with real-world studies, suggesting that differences in AE rates may be due to the study settings rather than the drug's properties."
"The text describes observed improvements in DSST score with vortioxetine treatment, indicating a clinically meaningful increase, and consistency with previous studies, which suggests a positive sentiment towards the drug's effects."	"The text provides factual information about the study design and patient treatment with vortioxetine but does not offer any positive or negative judgment about the drug itself."
+ 10 additional positive statements	+ 9 additional neutral statements

LLM prompt: *"Here is a list of statements about a published clinical study. Choose four or five key words that summarise the observations relating to vortioxetine (do not report vortioxetine as a key word)."*

Key words chosen by the LLM to summarize observations related to the study drug

1. Improvements[a]	2. Tolerability[b]	3. Efficacy[c]	4. Cognitive function[d]	5. Safety[e]

1. Clinical study	2. MDD	3. Real-world effectiveness	4. Adverse events	5. Third-line treatments

[a]"Improvements in patient functioning, depressive and cognitive symptoms, sexual function and HRQoL." [b]"Minimal mean weight change, well tolerated in elderly patients with Alzheimer's disease." [c]"Reduction in SDS score, first-line treatment benefits, effective in specific patient subgroups." [d]"Improvements in DSST score, cognitive symptoms and performance." [e]"Consistent safety profile, mild to moderate adverse events, no significant weight changes." DSST, digit symbol substitution test; HRQoL, health-related quality of life; LLM, large language model; MDD, major depressive disorder; RELIEVE, Real-Life Effectiveness of Vortioxetine in Depression; SDS, Sheehan Disability Scale.

- The citation statements were for 12 articles; 10 were from the neuroscience portfolio and 2 were from the gastrointestinal portfolio.
  - Of the 50 citation statements, 44 were for one study drug in the MDDs therapy area.
- Furthermore, the LLM was able to rationalize the classification of citation statements and choose key words after prompting to summarize observations related to the study drug.
  - **Figure 3** shows an example for 25 citation statements (13 positive and 12 neutral) that pertained to the study drug in a clinical trial publication.

Limitations

- The LLM was given citation statements from articles to analyze and may have missed the overall context of the articles when classifying sentiments towards the study drug.
- Although promising, the LLM requires evaluation with a larger data set that has human-provided sentiment classifications.

Conclusions

- We combined conventional impact metrics with an AI-powered LLM to identify and to characterize citation statements that discuss a target drug entity.
- Similar approaches could be used to support insight analysis relating to a wide variety of questions and content sources (e.g. news articles).

References

1. Rees T *et al*. Development of an alternative to journal impact factor using a broader range of article-level metrics [poster]. Presented at the 2023 European Meeting of ISMPP, January 24–25, 2023, London, UK.