Crowdsourcing adverse events associated with monoclonal antibodies targeting calcitonin gene-related peptide (CGRP) signaling for migraine prevention: Natural language processing analysis of social media

Pengfei Zhang, Brad K. Kamitaki, Thien Phu Do

Submitted to: JMIR Formative Research
on: March 08, 2024

Table of Contents

Crowdsourcing adverse events associated with monoclonal antibodies targeting calcitonin gene-related peptide (CGRP) signaling for migraine prevention: Natural language processing analysis of social media

Pengfei Zhang[1] MD; Brad K. Kamitaki[1] MD; Thien Phu Do[2] MD

[1]Department of Neurology Rutgers-Robert Wood Johnson Medical School New Brunswick US
[2]Danish Headache Center, Department of Neurology, Rigshospitalet Glostrup Faculty of Health and Medical Sciences University of Copenhagen Copenhagen DK

Corresponding Author:
Pengfei Zhang MD
Department of Neurology
Rutgers-Robert Wood Johnson Medical School
125 Paterson Street
Suite 6200
New Brunswick
US

Abstract

Background: Real-world data, such as those contained within social media platforms, can summarize diverse patient experiences to detect treatment-related adverse events.

Objective: To characterize adverse events related to novel calcitonin gene-related peptide (CGRP) therapeutics on Reddit, a large online social media forum.

Methods: We examined differences in word frequencies from medication-related posts extracted from the Reddit subforum r/Migraine over a ten-year period (2010-2020) using computational linguistics. In the validation phase, we compared the medications propranolol versus topiramate, as well as propranolol and topiramate each against randomly selected posts. In the application phase, we examined posts discussing the CGRP therapeutics erenumab and fremanezumab to determine frequently discussed adverse events.

Results: From 22,467 Reddit r/Migraine posts, we extracted 402 propranolol posts, 1423 topiramate posts, 468 erenumab posts, and 73 fremanezumab posts. Comparing topiramate against propranolol identified several expected adverse events. Erenumab compared against a random selection of terms identified “constipation” as a recurring key word. Erenumab against fremanezumab identified “constipation,” “depression,” “vomiting,” and “muscle.” No adverse events were identified for fremanezumab.

Conclusions: Computational linguistics, when applied to social media, can identify potential adverse events of interest for novel therapeutics. Social media data represents a promising avenue for pharmacovigilance, but further work is needed to improve reliability and usability.

(JMIR Preprints 08/03/2024:58176)
DOI: https://doi.org/10.2196/preprints.58176

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Crowdsourcing adverse events associated with monoclonal antibodies targeting calcitonin gene-related peptide (CGRP) signaling for migraine prevention: Natural language processing analysis of social media

Authors:

Pengfei Zhang, MD[1] (Corresponding Author); Brad K. Kamitaki, MD[1]; Thien Phu Do, MD[2,3,4,5]

Corresponding Author's Information:

125 Paterson St., Suite 6200

New Brunswick, New Jersey 08901

United States of America

Phone: 732-235-7729, E-mail: pz124@rwjms.rutgers.edu

[1]Department of Neurology, Rutgers University-Robert Wood Johnson Medical School, New Brunswick, New Jersey, United States.

[2]Department of Neurology, Danish Headache Center, Copenhagen University Hospital - Rigshospitalet, Copenhagen, Denmark.

[3]Department of Neurology, Copenhagen University Hospital - Herlev and Gentofte, Herlev, Denmark.

[4]Danish Knowledge Center on Headache Disorders, Glostrup, Denmark.

[5]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark

Competing interests:

TPD declares no competing interests. BK has received research grant support from the American Epilepsy Society, the New Jersey Health Foundation, and the Resource Center for Alzheimer's and Dementia Research in Asian and Pacific Americans under NIH/NIA Grant P30-AG083257. The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. PZ has received honorariums from Lundbeck Biopharmaceuticals, Board Vitals, and Fieve Clinical Research. He collaborates with Headache Science Incorporated without

receiving financial support. He has ownership interest in Cymbeline LLC.

Funding: This research received no specific grant from any funding agency.

Acknowledgement: None

Data Availability Statement: The data sets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Abstract:

BACKGROUND: Clinical trials demonstrated efficacy and tolerability of medications targeting calcitonin gene-related peptide (CGRP) signaling for migraine prevention. However, these trials may not accurately reflect the real-world experiences of more diverse and heterogeneous patient populations, who often have a higher disease burden and more comorbidities. Therefore, post-marketing safety surveillance is warranted. Regulatory organizations encourage marketing authorization holders to screen digital media for suspected adverse reactions, applying the same requirements as for spontaneous reports. Real-world data from social media platforms constitute a potential venue to capture diverse patient experiences and help detect treatment-related adverse events. However, while social media holds promise for this purpose, its use in pharmacovigilance is still in its early stages. Computational linguistics, which involves the automatic manipulation and quantitative analysis of oral or written language, offers a potential method for exploring this content.

OBJECTIVE: To characterize adverse events related to monoclonal antibodies targeting CGRP signaling on Reddit, a large online social media forum, by using computational linguistics.

METHODS: We examined differences in word frequencies from medication-related posts on the Reddit subforum r/Migraine over a ten-year period (2010-2020) using computational linguistics. The study had two phases: a validation phase and an application phase. In the validation phase, we compared posts about propranolol and topiramate, as well as posts about each medication against randomly selected posts, to identify known and expected adverse events. In the application phase, we analyzed posts discussing two monoclonal antibodies targeting CGRP signaling—erenumab and fremanezumab— to identify potential adverse events for these medications.

RESULTS: From 22,467 Reddit r/Migraine posts, we extracted 402 propranolol posts, 1423 topiramate posts, 468 erenumab posts, and 73 fremanezumab posts. Comparing topiramate against propranolol identified several expected adverse events, e.g., “appetite”, “weight”, “taste”, “foggy”, “forgetful”, and “dizziness”. Erenumab compared against a random selection of terms identified

"constipation" as a recurring key word. Erenumab against fremanezumab identified "constipation", "depression", "vomiting", and "muscle". No adverse events were identified for fremanezumab.

CONCLUSIONS: The validation phase of our study accurately identified common adverse events for oral migraine preventive medications. For example, typical adverse events such as "appetite" and "dizziness" were mentioned in posts about topiramate. When we applied this methodology to monoclonal antibodies targeting CGRP or its receptor—fremanezumab and erenumab, respectively—we found no definite adverse events for fremanezumab. However, notable flagged words for erenumab included "constipation," "depression," and "vomiting." In conclusion, computational linguistics applied to social media may help identify potential adverse events for novel therapeutics. While social media data shows promise for pharmacovigilance, further work is needed to improve its reliability and usability.

Key Words: Internet, Patient Reported Outcome, headache, health information, Reddit, registry, monoclonal antibody

Introduction:

Preventive medications for migraine include both non-disease specific drugs (e.g., anti-hypertensives and anti-convulsants) and, more recently, disease-specific anti-calcitonin gene-related peptide (CGRP) therapies[1]. Randomized clinical trials (RCTs) demonstrate the clinically relevant efficacy and tolerability of anti-CGRP monoclonal antibodies (mAbs)[2]. However, they may not accurately reflect real-world experiences of more heterogeneous and diverse patient populations who often have a higher disease burden and more comorbidities[3-6]. In addition to direct reports of potential adverse events from healthcare providers and patients, regulatory organizations encourage marketing authorization holders to screen digital media for suspected adverse reactions, which should be handled similarly to spontaneous reports with the same requirements applied[7]. While social media represents an obvious venue in this context, its usage for pharmacovigilance is still in its infancy[8]. Complementary methods of post-marketing surveillance for adverse events with social media are needed.

The last fifteen years were characterized by a vast proliferation of social media platforms, allowing for both the rapid dissemination of medical information, as well as direct insight into patient perspectives. Indeed, migraine is most prevalent in individuals between 15-49 years of age, who are among the most active users of the Internet[9]. This is also reflected by a growth of migraine-related content on the Internet[10-12]. The anonymous social media discussion board Reddit has over 430 million monthly active users and serves as a popular forum for health-related discussions[13]. Unlike other digital media platforms, Reddit provides easily accessible and legal ways of accessing historic forum data. Anonymous forums like Reddit allow users to share personal medical experiences and advice without social stigma. Collectively, the migraine-related subforums (or “subreddits”) such as r/Migraine have a userbase exceeding 54,100. While these platforms allow for a democratic exploration of patient experiences and perspectives, they are relatively understudied.

Computational linguistics, a discipline encompassing automatic manipulation and quantitative analysis of oral or written language, represents a potential way to explore this content. Computational linguistics can be applied to identify clusters of words in a dataset with the goal of deriving dominant thematic elements. Identifying word distribution frequencies in specific groups of texts is a common technique used in computational linguistics with multiple wide-ranging applications[14]. In this study, we applied computational linguistics, specifically word frequency analysis, to characterize and identify self-reported adverse events associated with the new therapeutic anti-CGRP mAbs reported by users on Reddit. To ascertain the reliability and validity of our methodology, we first conducted a preliminary study with older generation migraine drugs to see we could identify known adverse reactions. After validation, we applied the methodology to the newer generation of CGRP therapies for migraine prevention.

Methods:

Study Design

This was a cross-sectional survey of posts on the r/Migraine subforum on Reddit. Data was analyzed using custom back-end software. The study consisted of four phases: (1) Data Extraction, (2) Automated Data Organization, (3) Validation of Computational Linguistics Approach, and (4) Application of Computational Linguistics Approach.

Data Extraction

We extracted posts in JavaScript Object Notation (JSON) format from the Reddit r/Migraine subforum published from January 1, 2010 through January 1, 2020 using the pushshift.io, a publicly available Reddit repository, in order to extract data for analysis[15]. The timestamp, title, and the text of each post were obtained and combined into a comma-separated values (CSV) file for further evaluation. Of note, our search did not include any comments responding to individual posts. Furthermore, usernames were excluded to ensure protection of personal data. Once the above data was downloaded, we then converted the entire document to lower case characters for ease of data

search and extraction.

Automated Data Organization

For the validation phase of the study, we algorithmically obtained, through the Haskell filter function, a list of documents containing posts with either the brand or trade name for propranolol and topiramate. These medications were selected as evidence-based preventive treatments for migraine in accordance with American Headache Society guidelines and because their side effect profiles are well known[16]. Next, we algorithmically compiled all posts containing each drug in separate documents. To use computational linguistics terminology, each document represents a *corpus* for a medication; each post within that corpus further represents a *text*.

We subsequently identified posts not including any of the above medications, nor including any CGRP therapeutic medication, to serve as controls for comparison. To conduct multiple tests, we randomly divided these controls into four groups: Group 1, Group 2, Group 3, and Group 4. To accomplish this, we assigned each post in the control a random number between 1 to 4. All posts with randomly assigned digits 1 were grouped into Group 1. Those with a digit of 2 were grouped into Group 2, and so on.

Validation of Computational Linguistics Approach

Comparing the relative occurrence of words between two groups of texts is a well-defined technique in computational linguistics. For drugs with different distinct side effect profiles (e.g., the anticonvulsant topiramate versus the beta-blocker propranolol), we hypothesized there would be a difference in the frequency of word distributions between corpora for these different medications.

Although the chi square or log-likelihood ratio tests are commonly used, Lijiffijit and colleagues recently demonstrated that the Wilcoxon rank-sum test, Welch’s t-test, or bootstrap test are better choices to utilize to avoid overconfidence bias[17]. Consequently, we employed and validated Welch’s t-test for each of these comparisons given ease of implementation and intuitiveness.

In accordance with the abovementioned methodology, given a corpus for drug S and a corpus for drug T, we denoted each individual word by q, the total number of words in each corpus by n_1 and n_2, and x_1 and x_2 as the mean frequency of q over corpus S and corpus T, respectively. We then applied Welch's t-test[17]. To simplify our calculations, words with fewer than three characters in length, symbols, and numbers were filtered out. We identified any words showing statistical significance at $p < 0.05$ using a one-tailed test, since to reject the null hypothesis, it is sufficient to show that a specific word occurrence is greater – not less – in one corpus against another, or a t-statistic > 1.73. Because this process involved the repeated application of Welch's t-test $n_1 + n_2$ times to the same dataset, the Benjamini-Hochberg procedure was used to decrease multi-sampling bias. Specifically, once statistically significant words were obtained after a given comparison between two corpora, the list of significant words was then ranked by p-value. We then applied the Benjamini-Hochberg procedure to this ranked ordered list, only retaining those words passing this procedure.

To validate our method and the reliability of the Reddit data, we applied this approach first comparing propranolol and topiramate. Of note, if a text existed in both the propranolol and the topiramate corpora, then it was taken out of both. Subsequently, we compared propranolol vs. Group1 and topiramate vs. Group 2.

Application of Computational Linguistics Approach

The above approach was then applied to the corpus between erenumab vs. Group 3 in order to establish a baseline side effect profile for erenumab. Similarly, fremanezumab was compared to Group 4 to establish a baseline side effect profile for fremanezumab. We then compared the erenumab corpus against the fremanezumab corpus to ascertain any difference between the two medication side effect profiles. This is clinically relevant, as erenumab targets the CGRP receptor, while fremanezumab targets the CGRP ligand. We were interested to see whether there were fewer reports of specific adverse events with erenumab than with fremanezumab. Again, when comparing erenumab and fremanezumab, shared posts were removed.

For evaluation of our data, all contributing authors reviewed words identified through the above processes, where a word is counted as describing a potential adverse event by consensus. The above algorithms, including data download, parsing, filtering, and data analysis were implemented through custom back-end software using a combination of the Clojure and Haskell programming languages.

Ethical Considerations

Our data source was obtained from Pushshift.io, a social media data collection site, with the explicit purpose of collecting Reddit data for the purpose of research[15]. Our project proposal was approved by Rutgers University Institutional Review Board as non-human subjects research.

To ensure anonymity of users in our data processing, we removed all usernames from our dataset. Furthermore, since our methodology studies posts in aggregate, individual identifiable data are not published nor evaluated as part of this study.

Results:

Dataset

We extracted a total of 22,467 Reddit posts from r/Migraine, spanning from January 1, 2010, to December 31, 2019. In the validation phase of the study, we identified 402 posts (1.79%) that mentioned propranolol and 1,423 posts (6.33%) that mentioned topiramate. During the application phase, focusing on mAbs targeting CGRP signaling, we found 468 posts (2.08%) that mentioned erenumab and 73 posts (0.32%) that mentioned fremanezumab [Figure 1].

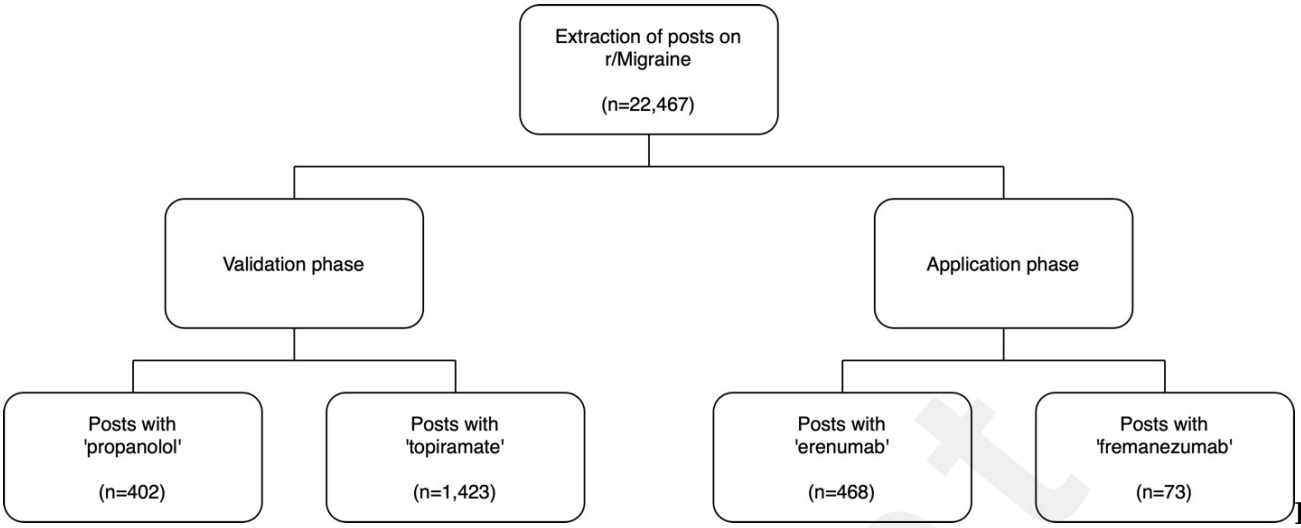

Figure 1. Study flow diagram of the validation and application phases of a computer linguistics analysis of r/Migraine social media posts on Reddit.

Since these posts did not exclusively reference the target medications, additional steps were taken to exclude co-occurring texts when comparing medications, as detailed in the Methods section. For the control corpora, Group 1 contained 5,058 posts, Group 2 had 4,996 posts, Group 3 included 5,130 posts, and Group 4 comprised 5,009 posts.

Validation of Computational Linguistics Approach

When topiramate posts were compared against propranolol posts using the above methodology, side effect-related words including “appetite,” “cognitive,” “tingling,” “remember,” “taste,” “dopamax” (a frequently used slang term describing the cognitive adverse events of topiramate), “tastes,” and “kidney” were statistically more frequent [Table 1].

Table 1. Key words identified when comparing topiramate with propranolol from the Reddit r/Migraine subforum (2010-2020)

topamax, topiramate, **appetite***, effects, posts, ready, realized, serious, throw, **cognitive***, loss, side, treating, level, flat, **tingling***, situation, blocks, opinion, improved, ntopamax, options, woman, possibility, issues, reddit, glasses, awesome, cymbalta, grateful, **remember***, version, depressants, decent, dumb, believe, future, **taste***, angry, evidence, **dopamax***, tapering, someone, practice, trust, shortly, honest, sign, **kidney***, steroids, refused, tons, disorders, cambia, learned, nerve, feet, ahead, essentially, posted, **tastes***, allergic, emotional, harder, ride, limited, white, spoke, parts, comment, optimistic, hurt, count

***Bolded and starred words were identified as potential side-effect related words of interest**

Table 2. Key words identified when comparing propranolol with Group1 from the Reddit r/Migraine subforum (2010-2020)

propranolol, taking, been, months, topamax, effects, prescribed, tried, with, take, medication,

doctor, side, dose, have, neurologist, preventative, amitriptyline, daily, since, also, inderal, currently, anyone, started, month, years, weeks, from, beta, which, migraines, medications, some, about, life, week, meds, made, drug, really, time, year, days, will, other, first, nortriptyline, after, because, dosage, effect, again, just, didn, stopped, they, haven, topiramate, help, well, terrible, went, starting, this, past, blockers, still, doesn, almost, maxalt, nothing, feel, sumatriptan, that, back, like, seemed, what, stop, preventatives, appointment, long, good, advice, weight, blood, npropranolol, give, neuro, blocker, relief, imitrex, worked, results, going, headaches, only

Table 3. Key words identified when comparing topiramate with Group2 from the Reddit r/Migraine subforum (2010-2020)

topamax, side, effects, taking, topiramate, been, have, with, neurologist, just, dose, about, that, take, started, tried, week, this, since, which, months, doctor, because, month, migraines, weeks, from, also, some, well, daily, other, prescribed, like, years, going, back, what, effect, really, preventative, only, work, working, neuro, days, made, know, much, long, loss, starting, need, even, after, currently, first, medication, meds, still, worked, didn, time, then, propranolol, again, good, stopped, taken, more, nothing, point, **appetite***, dosage, drug, imitrex, would, helped, when, control, works, medications, getting, night, will, think, preventatives, already, were, there, every, little, **weight***, went, amitriptyline, gave, said, issues, appointment, maxalt, increased, want, last, should, feel, anything, anyone, trying, seems, being, before, pretty, very, having, wants, here, finally, twice, until, able, give, helping, **remember***, while, chronic, where, experiences, haven, experience, next, over, three, year, triptans, stop, used, options, down, botox, could, severe, into, better, keep, they, beta, away, help, read, worse, blockers, birth, through, great, them, feet, **memory***, success, depakote, though, another, almost, increase, past, doses, under, female, anymore, these, frequency, decided, seeing, same, many, effective, switch, came, symptoms, actually, gabapentin, magnesium, different, without, needed, aimovig, else, seem, depression, lost, course, things, recently, thing, half, trokendi, lose, advice, desperate, wanted, problems, increasing, horrible, switched, diagnosed, couldn, took, normal, life, **tingling***, history, start, suggested, mood, drugs, several, sure, doesn, maybe, concerned, awful, **cognitive***, constant, might, find, **anxiety***, insurance, wasn, changed, called, experiencing, break, everything, feeling, doing, agreed, medicine, reading, function, never, makes, drive, neurologists, abortive, nthanks, talk, times, sent, headaches, ntopamax, myself, anti, nortriptyline, pill, treatment, verapamil, prescription, instead, results, than, told, once, withdrawal, prevention, nervous, supposed, august, doctors, seen, relief, probably, including, less, completely, worth, hands, felt, switching, nausea, finding, primary, care, reaction, pills, either, hoping, tired, change, least, antidepressants, continue, prescribe, april, something, plus, fioricet, background, post, period, current, basically, those, hell, deal, frustrated, hello, difference, question, sleepy, afraid, however, story, **taste***, around, cannot, list, terrible, says, cymbalta, wondering, stopping, suggestions, idea, medical, asked, **zombie***, fine, issue, such, serious, hard, combo, realize, right, changes, july, blood, body, pharmacist, initially, brain, ready, thinks, treating, happy, tonight, gone, suffering, people, done, developed, talked, **tingly***, abortives, supplements, couple, risk, small, frequent, stories, flat, became, eating, looking, zomig, stuff, free, gained, problem, stayed, **dizzy***, **fatigue***, opinion, wean, except, given, hair, gain, situation, possible, lasted, various, pounds, zero, words, causing, minimal, metoprolol, **forget***, reduced, taper, steroids, physician, relpax, definitely, single, upped, **kidney***, ntopiramate, possibly, eventually, weaning, steroid, elavil, tests, making, nafter, whole, exercise, insomnia, january, ever, zonegran, september, pain, thinking, second, generic, slowly, depressants, follow, major, **foggy***, **dopamax***, please, specialist, nsaids, level, functional, plan, monday, hospital, shot, drink, kept, reduction, toradol, none, surprised, gotten, cycle, rescue, positive, seemed, intractable, stay, make, failed, chance, dropped,

ncurrently, nightly, crazy, live, acupuncture, average, family, seizure, **forgetful***, levels, quit, stomach, calcium, ability, early, practice, selftext, others, thanks, discuss, occasionally, june, propanolol, high, most, decrease, schedule, losing, gradually, believe, guess, both, stones, coming, bring, march, everyone, look, biggest, previous, option, higher, helps, found, longer, tolerance, october, onto, slow, morning, previously, constantly, wits, fairly, negative, especially, tapering, proven, posts, quitting, upping, common, kind, reason, possibility, number, waiting, online, bedtime, goes, name, pharmacy, best, super, improved, handle, treat, nbut, **dizziness***, using, thank, complete, main, vitamins, fast, stick, anyway, bumped, willing, miracle, hasn, shocked, clinic, decent, mris, noticed, weaned, refused, talking, candidate, together, quick, nanyway, occasional, round, efficacy, tolerate, improvement, **carbonated***, nany, suggest, chest, **tastes***, psychiatrist

***Bolded and starred words were identified as potential side-effect related words of interest**

When propranolol was compared to the control corpus Group 1, 98 words appeared more frequently, none of which we identified as potential adverse events [Table 2]. When topiramate was compared against Group 2, there were 563 words that appeared more often [Table 3]. The following words were determined to be potential adverse effects: “appetite,” “weight,” “remember,” “memory,” “cognitive,” “anxiety,” “taste,” “zombie,” “tingly,” “dizzy,” “fatigue,” “forget,” “kidney,” “foggy,” “dopamax,” “forgetful,” “dizziness,” “carbonate,” and “tastes.”

Application of Computational Linguistics Approach

We next identified frequent words of interest by comparing erenumab to Group 3 [Table 4] and fremanezumab to Group 4. For erenumab, 36 words appeared more frequently, but “constipation” was the only potential adverse effect. The only word appearing more often for fremanezumab versus the control corpus was “ajovy,” a brand name for fremanezumab.

Table 4. Key words identified when comparing erenumab with Group 3 from the Reddit r/Migraine subforum (2010-2020)

aimovig, dose, month, injection, botox, neurologist, effects, months, drug, first, insurance, emgality, injections, tried, topamax, side, free, trial, neuro, next, shot, approved, preventatives, since, hoping, ajovy, cgrp, **constipation***, effect, plan, working, here, received, shots, doses, propranolol

***Bolded and starred words were identified as potential side-effect related words of interest**

When comparing erenumab to fremanezumab, there were 180 words that were significantly associated with erenumab, with the following words representing potential adverse events: “depression”, “constipation”, and “vomiting” [Table 5]. Notably, the word “muscle” was also

associated with erenumab, but it was unclear whether this described an adverse event [Table 5]. The only word considered significant for fremanezumab was “ajovy” with no adverse effect-related words detected.

Table 5. Key words identified when comparing erenumab with fremanzeumab from the Reddit r/Migraine subforum (2010-2020)

aimovig, where, felt, medications, doesn, effect, much, starting, makes, enough, stopped, such, come, everything, home, triptan, next, possible, half, wait, worth, being, live, high, talk, preventatives, trial, question, nhas, medicine, diagnosed, thinking, hour, means, received, know, definitely, best, caused, into, happy, sumatriptan, given, lost, issue, short, here, **depression***, frustrated, early, should, weird, aren, knows, problem, especially, leave, advance, mind, ones, painful, story, **constipation***, real, cost, school, list, thank, related, note, support, told, nothing, this, follow, nbut, nthis, stories, wouldn, another, available, https, thread, difficult, info, okay, regular, bring, vision, wrong, migraine, specialist, often, made, abortive, considering, history, intense, notice, sometimes, sort, **vomiting***, break, causing, expensive, experiencing, learned, middle, asking, physical, seemed, vent, muscle, coming, concerned, monday, water, current, program, work, changes, medical, make, didn, blocks, happen, recommended, along, case, horrible, extra, specific, chest, fact, fall, letting, nalso, reason, types, type, serious, reading, barely, situation, therapy, treatments, erenumab, nand, began, clinic, dealing, minute, trials, freaking, ideas, tuesday, incredibly, move, nfor, paying, strong, benefits, continue, friends, mostly, naimovig, ndoes, news, nnow, rest, unfortunately, wonder, hold, reddit, update, toradol, prescribe, disease, switched, rebound

***Bolded and starred words were identified as potential side-effect related words of interest**

Discussion:

In this study, we applied computational linguistics methods to unfiltered discussions about migraine treatments on Reddit to distinguish patient-reported adverse effects associated with these medications. The validation phase of our study accurately identified common adverse events for standard migraine preventive treatments. For example, typical adverse events such as “taste” and “tingling” appeared in posts for topiramate. After applying this methodology to monoclonal antibodies targeting CGRP or its receptor, fremanezumab and erenumab, respectively, no definite adverse events were flagged for fremanezumab. Notable flagged words for erenumab included “constipation”, depression,” and “vomiting.”

In our proof of concept analysis, we confirmed that key words discussed in relation to topiramate from the Reddit r/Migraine subforum are consistent with the most common adverse events observed in randomized clinical trials[18]. Interestingly, dizziness, a common side effect, did

not appear when comparing propranolol to either the control corpus or topiramate[19]. A possible explanation could be that many different words are used to describe this side effect, e.g., “lightheadedness,” “dizzy,” which were detected individually but did not meet statistical significance after correcting for multiple comparisons. Because patients may use a variety of different words to describe a single symptom, pooling these terms may be useful in future studies when analyzing social media data.

While constipation was not a major concern during randomized clinical trials with erenumab, emerging real-world data report that it is among the most frequent reasons for treatment discontinuation[20,21]. Constipation was subsequently added to the United States Food and Drug Administration (US FDA) label after regulatory approval[22]. During randomized clinical trials, approximately 2% of participants reported constipation compared to approximately 10-20% of real-world users[20,23,24]. Our findings further support this, since Reddit users frequently discuss constipation and vomiting in relation to erenumab. The discrepancy between data from randomized clinical trials and real-world experiences may be due to a more heterogenous patient population receiving erenumab in the post-marketing phase. This population likely includes patients with other medical comorbidities or factors that could have excluded them from consideration in clinical trials. Increasing evidence suggests that individuals with migraine may have a higher prevalence of comorbid gastrointestinal disorders when compared with the general population[25,26].

Interestingly, the word “muscle” was associated with posts discussing erenumab when compared with fremanezumab. Although we did not include this word as a potential adverse event due to its ambiguity, it could be related to muscle cramps, which was added to the US FDA warning label for erenumab after approval[22]. Muscle spasms occurred in approximately 2% of patients in randomized clinical trials[20]. Whether “muscle”-related adverse are occurring at a more significant rate in real-world usage is unclear. Additional investigation into the extent and severity of muscle disease due to erenumab could be considered in the future or when examining other post-marketing

surveillance data sources for verification.

Another notable post-marketing safety update was the addition of hypertension to the US FDA label for erenumab; however, this was not flagged in our analysis. In RCTs, less than 1% of participants reported hypertension, but participants with severe cardiovascular risk factors were excluded from these trials, and hypertension has not been commonly noted in other studies[23,24]. It is possible that cardiovascular risk factors are screened to a higher degree than gastrointestinal disorders. Patients at risk for hypertensive complications might therefore be more likely to decline treatment with this drug. Finally, worsening of hypertension is typically asymptomatic, and patients may not recognize this as a potential adverse event to raise in social medial discussions.

There were no key words flagged in the fremanezumab dataset. That constipation was flagged for erenumab but not for fremanezumab could be related to pharmacodynamic differences in the mechanism of action. Erenumab targets the CGRP receptor, while fremanezuamb targets its ligand. One study offers a possible mechanistic model that strengthens this hypothesis, specifically, via direct effects of CGRP on intestinal motility modulated by other receptors in addition to the canonical CGRP receptor[27]. However, galcanezumab, another monoclonal antibody targeting the CGRP ligand, also had constipation listed as a common adverse reaction in clinical studies and post-marketing reports[28]. Further data is needed to clarify any potential differences in mechanism.

A strength of our study is the inherent anonymity of the Reddit forum, which allows us to analyze information that patients might not otherwise feel comfortable discussing or prioritize sharing during a clinical consultation. Querying the Reddit migraine subforum also provides insight into the experience of more than 54,000 unique users, which could include diverse sources ranging from patients and family members to experienced health care providers.

However, the use of Reddit data also confers significant limitations. We cannot correlate our findings with individual patient demographics, previous medical history, or other clinically relevant information. All information is self-reported and subjective, which calls into question the accuracy of

unverifiable data. Although the number of posts differs between erenumab and fremanezumab due to FDA approval time, we do not expect this to significantly change our result since word distribution should be theoretically constant despite sample size. Another limitation is the use of multiple different words for the same symptom, e.g., when lightheadedness, dizzy, and dizziness collectively describe the same entity. Relevant symptoms described in various ways may be missed during this type of analysis. Pooling multiple similar terms could avoid this problem in future studies. Finally, our method did not analyze comments responding to specific posts; incorporating this into future studies utilizing word frequency methodology could increase the accuracy of our conclusions. Future directions for our line of research include the analysis of other digital platforms to compare with our findings.

Post-marketing pharmacovigilance remains critical, especially when randomized clinical trials under detect adverse events. When compared with healthcare professionals, direct patient reporting of adverse drug events identifies more potential signals and in greater detail[29]. Our study provides support that monitoring of patient reports via social media can be incorporated into post-market safety surveillance for preventive migraine medications. While our proposed methodology has its limitations, the data offer a broader post-marketing perspective that complements clinical trial findings and case reports. We hope that consideration by relevant stakeholders will enhance existing tools in the pharmacovigilance armamentarium, enabling quicker and earlier detection of adverse effect signals. However, more work is needed to improve the reliability and validity of social media analysis, which faces the same challenges inherent to big data, including high volume, unstructured capture, and structural regulatory barriers, which all need to be addressed before implementation[30].

References:

1. Ashina M, Buse DC, Ashina H, Pozo-Rosich P, Peres MFP, Lee MJ, Terwindt GM, Halker Singh R, Tassorelli C, Do TP, Mitsikostas DD, Dodick DW. Migraine: integrated approaches to clinical management and emerging treatments. Lancet. 2021 Apr 17;397(10283):1505-1518. doi: 10.1016/S0140-6736(20)32342-4. Epub 2021 Mar 25. PMID: 33773612.
2. Charles A and Pozo-Rosich P. Targeting calcitonin gene-related peptide: a new era in migraine therapy. *Lancet* 2019; 394: 1765-1774. 2019/11/02. DOI: 10.1016/s0140-6736(19)32504-8.
3. Ornello R, Casalena A, Frattale I, et al. Real-life data on the efficacy and safety of erenumab in the Abruzzo region, central Italy. *J Headache Pain*. 2020;21(1):32. Published 2020 Apr 7. doi:10.1186/s10194-020-01102-9
4. Ruff DD, Ford JH, Tockhorn-Heidenreich A, et al. Efficacy of galcanezumab in patients with episodic migraine and a history of preventive treatment failure: results from two global randomized clinical trials. Eur J Neurol. 2019;ene.14114.
5. Ashina M, Tepper S, Brandes JL, et al. Efficacy and safety of erenumab (AMG334) in chronic migraine patients with prior preventive treatment failure: A subgroup analysis of a randomized, double-blind, placebo-controlled study. Cephalalgia. 2018;38:1611–1621.
6. Ferrari MD, Diener HC, Ning X, et al. Fremanezumab versus placebo for migraine prevention in patients with documented failure to up to four migraine preventive medication classes (FOCUS): a randomised, double-blind, placebo-controlled, phase 3b trial. Lancet (London, England). 2019;394:1030–1040.
7. European Medicine Agency. Guideline on good pharmacovigilance practices. https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/guideline-good-pharmacovigilance-practices-gvp-module-vi-collection-management-submission-reports_en.pdf (Accessed October 3, 2021)

8. Pappa, D., Stergioulas, L.K. Harnessing social media data for pharmacovigilance: a review of current state of the art, challenges and future directions. Int J Data Sci Anal 8, 113–135 (2019). https://doi.org/10.1007/s41060-019-00175-3
9. Global, regional, and national burden of migraine and tension-type headache, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol* 2018; 17: 954-976. 2018/10/26. DOI: 10.1016/s1474-4422(18)30322-3.
10. Saffi H, Do TP, Hansen JM, Dodick DW, Ashina M. The migraine landscape on YouTube: A review of YouTube as a source of information on migraine. Cephalalgia. 2020 Oct;40(12):1363-1369. doi: 10.1177/0333102420943891. Epub 2020 Jul 23. PMID: 32703016.
11. Bojazar R, Do TP, Hansen JM, Dodick DW, Ashina M. Googling migraine: A study of Google as an information resource of migraine management. Cephalalgia. 2020 Dec;40(14):1633-1644. doi: 10.1177/0333102420942241. Epub 2020 Jul 23. PMID: 32703015.
12. Viana M, Hougaard A, Do TP. The depiction of migraine visual aura on the internet. Cephalalgia. 2020 Sep;40(10):1113-1118. doi: 10.1177/0333102420927027. Epub 2020 May 27. PMID: 32460539.
13. Reddit's 2019 Year in Review, https://redditblog.com/2019/12/04/reddits-2019-year-in-review/ (accessed September 17, 2020).
14. Lijffijt J, Papapetrou P, Puolamäki K, et al. Analyzing Word Frequencies in Large Text Corpora Using Inter-arrival Times and Bootstrapping. Springer Berlin Heidelberg, 2011, pp.341-357.
15. Baumgartner J, Zannettou S, Keegan B, et al. The Pushshift Reddit Dataset. *Proceedings of the International AAAI Conference on Web and Social Media* 2020; 14.
16. Charles JA and Rapoport AM. The American Headache Society's Position Statement on

Integrating New Migraine Treatments into Clinical Practice - Comments. *Headache* 2019; 59: 629. 2019/04/12. DOI: 10.1111/head.13496.

17. Lijffijt J, Nevalainen T, Saily T, et al. Significance testing of word frequencies in corpora. 2014. DOI: 10.1093/llc/fqu064.
18. Linde M, Mulleners WM, Chronicle EP, et al. Topiramate for the prophylaxis of episodic migraine in adults. *Cochrane Database Syst Rev* 2013; 2013: Cd010610. 2013/06/26. DOI: 10.1002/14651858.Cd010610.
19. Linde K and Rossnagel K. Propranolol for migraine prophylaxis. *Cochrane Database Syst Rev* 2004: Cd003225. 2004/04/24. DOI: 10.1002/14651858.CD003225.pub2.
20. Lattanzi S, Brigo F, Trinka E, et al. Erenumab for Preventive Treatment of Migraine: A Systematic Review and Meta-Analysis of Efficacy and Safety. *Drugs* 2019; 79: 417-431. 2019/02/23. DOI: 10.1007/s40265-019-01069-1.
21. Do T.P., Younis S., Ashina M. (2021) Erenumab. In: Maassen van den Brink A., Martelletti P. (eds) Monoclonal Antibodies in Headache. Headache. Springer, Cham. https://doi.org/10.1007/978-3-030-69032-8_9
22. Amgen Inc. Aimovig (erenumab-aooe) [package insert]. U.S. Food and Drug Administration website. https://www.accessdata.fda.gov/drugsatfda_docs/label/2018/761077s000lbl.pdf. Revised May 2018. Accessed September 22, 2020.
23. Ornello R, Casalena A, Frattale I, et al. Real-life data on the efficacy and safety of erenumab in the Abruzzo region, central Italy. *J Headache Pain* 2020; 21: 32. 2020/04/09. DOI: 10.1186/s10194-020-01102-9.
24. Lambru G, Hill B, Murphy M, et al. A prospective real-world analysis of erenumab in refractory chronic migraine. *J Headache Pain* 2020; 21: 61. 2020/06/04. DOI: 10.1186/s10194-020-01127-0.
25. Aamodt AH, Stovner LJ, Hagen K, et al. Comorbidity of headache and gastrointestinal

complaints. The Head-HUNT Study. *Cephalalgia* 2008; 28: 144-151. 2008/01/17. DOI: 10.1111/j.1468-2982.2007.01486.x.

26. Martami F, Ghorbani Z, Abolhasani M, et al. Comorbidity of gastrointestinal disorders, migraine, and tension-type headache: a cross-sectional study in Iran. *Neurol Sci* 2018; 39: 63-70. 2017/10/13. DOI: 10.1007/s10072-017-3141-0.
27. Falkenberg K, Bjerg HR and Olesen J. Two-Hour CGRP Infusion Causes Gastrointestinal Hyperactivity: Possible Relevance for CGRP Antibody Treatment. *Headache* 2020; 60: 929-937. 2020/04/01. DOI: 10.1111/head.13795.
28. Eli Lilly and Company. Emgality (Galcanezumab) [Summary of Product Characteristics]. European Medicines Agency. https://www.ema.europa.eu/en/documents/product-information/emgality-epar-product-information_en.pdf. Accessed September 29, 2020
29. Avery, A.J., Anderson, C., Bond, C.M., Fortnum, H., Gifford, A., Hannaford, P.C., Murphy, E.: Evaluation of patient reporting of adverse drug reactions to the UK Yellow Card Scheme: literature review, descriptive and qualitative analyses, and questionnaire surveys. Health Technol. Assess. **15**(20), 1–234 (2011)
30. Bousquet, C., Dahamna, B., Guillemin-Lanne, S., Darmoni, S.J., Faviez, C., Huot, C., Katsahian, S., Leroux, V., Pereira, S., Richard, C., Schck, S., Souvignet, J., Lillo-Le Lout, A., Texier, N.: The adverse drug reactions from patient reports in social media project: five major challenges to overcome to operationalize analysis and efficiently support pharmacovigilance process. JMIR Res. Protoc. **6**(9), e179 (2017)

Supplementary Files

Figures

Study flow diagram of the validation and application phases of a computer linguistics analysis of r/Migraine social media posts on Reddit.