

Original Paper

# Social Media Mining of Long-COVID Self-Medication Reported by Reddit Users: Feasibility Study to Support Drug Repurposing

Jonathan Koss, MSc; Sabine Bohnet-Joschko, Prof Dr

Department of Management and Entrepreneurship, Faculty of Management, Economics and Society, Witten/Herdecke University, Witten, Germany

**Corresponding Author:**

Jonathan Koss, MSc

Department of Management and Entrepreneurship

Faculty of Management, Economics and Society

Witten/Herdecke University

Alfred-Herrhausen-Str. 50

Witten, 58455

Germany

Phone: 49 2302926475

Email: [jonathan.koss@uni-wh.de](mailto:jonathan.koss@uni-wh.de)

## Abstract

**Background:** Since the beginning of the COVID-19 pandemic, over 480 million people have been infected and more than 6 million people have died from COVID-19 worldwide. In some patients with acute COVID-19, symptoms manifest over a longer period, which is also called “long-COVID.” Unmet medical needs related to long-COVID are high, since there are no treatments approved. Patients experiment with various medications and supplements hoping to alleviate their suffering. They often share their experiences on social media.

**Objective:** The aim of this study was to explore the feasibility of social media mining methods to extract important compounds from the perspective of patients. The goal is to provide an overview of different medication strategies and important agents mentioned in Reddit users’ self-reports to support hypothesis generation for drug repurposing, by incorporating patients’ experiences.

**Methods:** We used named-entity recognition to extract substances representing medications or supplements used to treat long-COVID from almost 70,000 posts on the “/r/covidlonghaulers” subreddit. We analyzed substances by frequency, co-occurrences, and network analysis to identify important substances and substance clusters.

**Results:** The named-entity recognition algorithm achieved an F1 score of 0.67. A total of 28,447 substance entities and 5789 word co-occurrence pairs were extracted. “Histamine antagonists,” “famotidine,” “magnesium,” “vitamins,” and “steroids” were the most frequently mentioned substances. Network analysis revealed three clusters of substances, indicating certain medication patterns.

**Conclusions:** This feasibility study indicates that network analysis can be used to characterize the medication strategies discussed in social media. Comparison with existing literature shows that this approach identifies substances that are promising candidates for drug repurposing, such as antihistamines, steroids, or antidepressants. In the context of a pandemic, the proposed method could be used to support drug repurposing hypothesis development by prioritizing substances that are important to users.

(*JMIR Form Res* 2022;6(10):e39582) doi: [10.2196/39582](https://doi.org/10.2196/39582)

**KEYWORDS**

social media mining; drug repurposing; long-COVID; crowdsourcing; COVID-19; Reddit; social media; content analysis; network analysis; recognition algorithm; treatment

## Introduction

**Background**

Since the beginning of the COVID-19 pandemic, over 480 million people have been infected and more than 6 million

people have died from COVID-19 worldwide [1]. In some patients with acute COVID-19, symptoms manifest over a longer period of time [2]. Owing to this phenomenon, the term “long-COVID” (LC) has emerged [3]. LC refers to both postacute (lasting longer than 4 weeks) and chronic (lasting longer than 12 weeks) symptoms [3,4]. At least one symptom

persists in 32%-87% of previously hospitalized patients [4]. Furthermore, the incidence of LC is estimated to range between 10% and 35% in individuals who have not been hospitalized [5]. The economic costs associated with LC symptomatology could be significant. Reynolds et al [6] stated that chronic fatigue syndrome, which has similar characteristics to LC [7], leads to a 37% reduction in household productivity and a 54% reduction in labor force productivity in the United States. Unmet medical needs have motivated immense research activities. ClinicalTrials.gov lists more than 7000 studies in the field of COVID-19, including more than 600 LC-specific studies [8].

### Retrospective Clinical Analysis

The large number of ongoing studies highlights a key challenge in drug development. There are numerous substances that are potentially effective. It is therefore essential to identify promising substances and narrow down the number of potential drug candidates to those showing the most promise. Given the urgency, scarcity of financial resources, and the high risk of failure in pharmaceutical research [9], drug repurposing (DR) appears to be a promising strategy for LC drug development. The exploitation of existing drugs for new therapeutic purposes usually leads to shorter development cycles with lower costs [10]. For example, existing drugs have proven to be safe for use in humans. Accordingly, phase I clinical trials are not required [10]. From a historical perspective, DR has often been serendipitous [10,11], but systematic approaches also exist to identify promising target leads [12]. One of these approaches is retrospective clinical analysis [12], which has already been used in the context of the COVID-19 pandemic [12]. Retrospective clinical analysis involves learning from real-world experience (eg, evaluating clinical case reports) to hypothesize applications of existing drugs for new indications [13].

### Mining Patients' Experiences From Social Media

Traditionally, retrospective clinical analysis is based on information that is collected and stored in databases, explicitly dedicated for health care system-related applications. Signals for potential DR are subsequently generated by professionals analyzing the data. Nowadays, there is a growing awareness in medical research that the collective intelligence of the affected patients, jointly searching for a solution to improve a medical condition, can be a driver of innovation [14,15] that is leveraged in the innovation process known as "crowdsourcing" [14,16]. While traditional crowdsourcing aims at active collaboration (eg, between a pharmaceutical company and an external patient group), online forums enable a passive approach to collecting real-world data by offering relevant content for analysis, which is also called "passive crowdsourcing" [17,18]. For instance, researchers have analyzed data from disease-specific social media platforms to identify medications that are used outside the approved indication (off-label use) [10,19]. Off-label use provides information to support hypotheses regarding DR [10,19]. These approaches save time and costs associated with data collection, while incorporating patients' real-world experiences. However, social media mining (SMM) [19], a term

that refers to the collection of methods used for conducting passive crowdsourcing, poses significant risks in terms of bias [10,19]. For instance, owing to the age-related user behavior on social media platforms, the data might not be generalizable to the whole population [10].

### Generating Hypotheses on DR From Discussions Among Long-Haulers

In this study, we aimed to capture substances such as medications and supplements that are relevant to the coping strategies of patients with LC. Accordingly, we applied the principle of retrospective clinical analysis using passive crowdsourcing by applying SMM. Since there is no approved drug for LC, promising LC candidates based on off-label properties could not be determined. Instead, we used an exploratory method, mainly consisting of the application of named-entity recognition (NER) and network analysis, aiming to provide an overview of different treatment strategies and important compounds from the patient's perspective for DR hypothesis generation. Methodologically similar approaches have previously been used to identify substances used in self-medication regarding opioid withdrawal [20] or to monitor potential drug interactions and reactions [21]. Furthermore, network analysis has been used to explore discussions related to certain diseases [22] or to explore the public perspective on vaccines [23,24]. For example, Lewis et al [22] used network analysis to analyze reasons for older adults to join a diabetes online community. Luo et al [24] used network analysis to explore public perceptions of the COVID-19 vaccine.

To our best knowledge, this study is the first to explore treatment strategies and important medications to support DR hypothesis generation by applying network analysis.

### Research Objectives

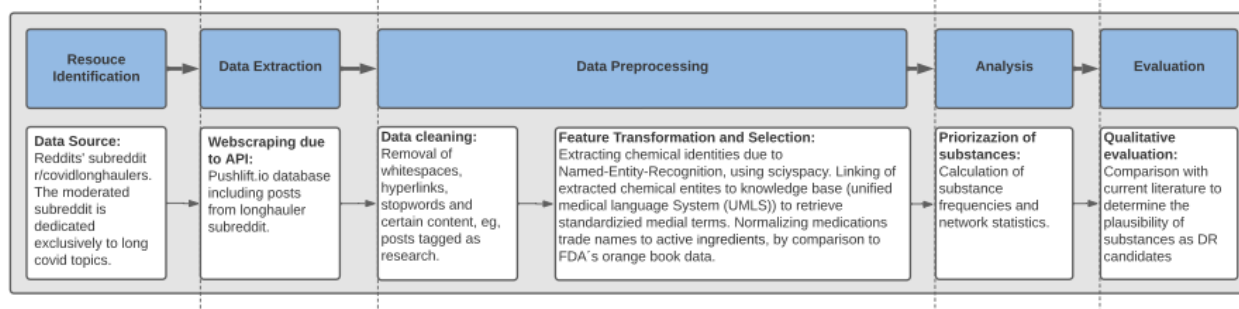
The aim of this feasibility study was to evaluate whether the proposed method can be employed to support DR hypothesis generation based on the experiences of affected individuals shared on Reddit. To this end, we first explored which substances are mentioned in LC online discussions regarding self-medication. Second, we investigated whether there are clusters of substances often discussed together, indicating treatment strategies. Third, we attempted to identify the most important substances in these clusters to indicate respective treatment strategies.

## Methods

### Overview

The methodology used in this study consists of the following steps: (1) extraction of appropriate data, (2) detection of substance entities mentioned in users' posts using NER, and (3) analysis of substance frequencies and co-occurrence networks of substance entities. Figure 1 outlines the end-to-end workflow in detail.

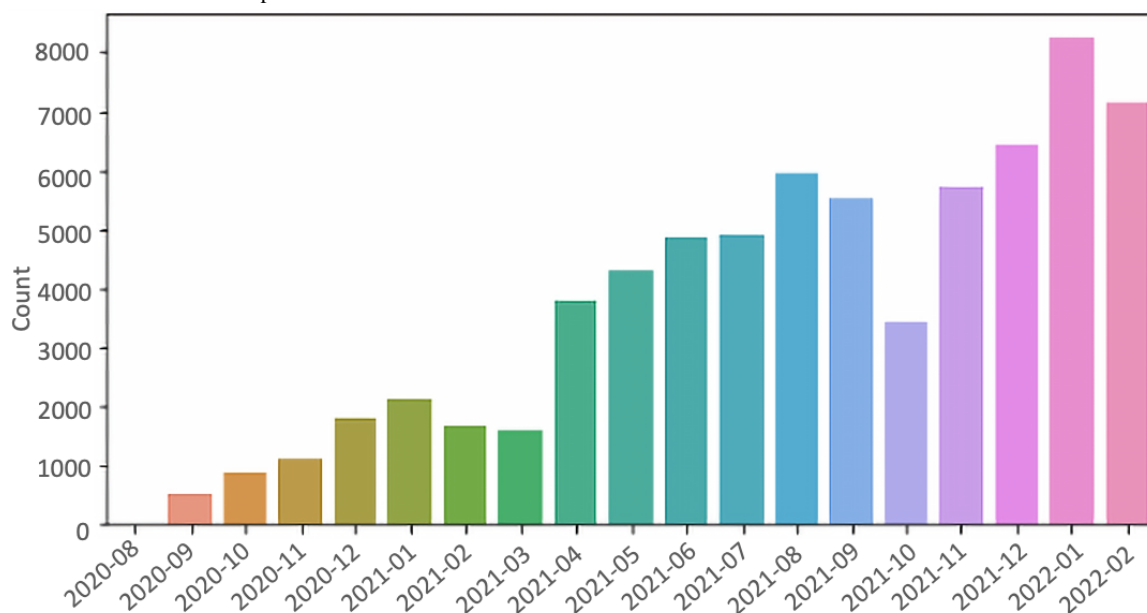
**Figure 1.** End-to-end detailed study workflow. The workflow can be divided into the following steps: resource identification, data extraction, data preprocessing, analysis, and evaluation. API: application programming interface; DR: drug repositioning; FDA: Food and Drug Administration.



## Data Source and Extraction

Reddit is a social media platform that is organized in theme-specific forums called “subreddits” [20]. The data extraction process was performed using Pushshift [25], which is a platform that collects Reddit data and has been available to researchers since 2015 [25]. The extracted data consist of posts and metadata from the subreddit “/r/covidlonghaulers,” which has already been used to explore LC symptoms [26,27]. This subreddit is actively moderated by specific users and provides a medium for LC-related discussions. The content is subject to strict rules prohibiting the promotion of alternative treatment, misinformation, and conspiracy theories. As of January 3, 2022, the subreddit had over 24,000 subscribers and 20,000 threads.

**Figure 2.** Overview of the number of posts at different dates.



## Substance Entity Extraction

First, the text was preprocessed to improve the data quality. For example, hyperlinks, tabs, and blank lines were removed. The substances of interest mentioned by patients in posts needed to be extracted and structured for subsequent analysis [19] using NER [28]. We defined substances of interest as explicitly mentioned substances or groups of substances that can be considered as treatments. For instance, we captured conventional supplements (eg, vitamin supplements) or prescription medications (eg, antidepressants) that were discussed by users.

Users self-report their LC experiences such as discussing symptoms [26] and medications.

Beyond the extraction of posts, metadata such as the username, date of the post, or the so-called “link flair text” can be extracted. Link flair text represent thematic tags that are used to associate posts (initial and subsequent posts) with specific categories. This provides researchers with the ability to exclude data unrelated to the analysis. For example, posts tagged as articles, research articles, or humor posts were excluded from the analysis (see [Multimedia Appendix 1](#)). Additionally, posts without tags were excluded. The analyzed data included 68,268 posts written by 8717 users between August 31, 2020, and March 1, 2022 ([Figure 2](#)).

ScispaCy provides several NER models related to medical issues [29] and was used to extract symptoms from LC Reddit posts [27]. In principle, two ScispaCy models are applicable for this purpose. The first model is the en\_ner\_bc5cdr\_md model [30], which can detect chemical substances and diseases. Since the module covers chemical substances in general and not specific medications or dietary supplements, we defined stop words such as “ethanol” to narrow the focus of the analysis to substances of interest. The second model is the Med7 model [31], which specifically focuses on drug extraction. Running both models together yielded the best results [27]. Negated substances were

excluded from the extraction by considering negotiation terms. For this purpose, we added the Negex algorithm [32] to the NER ScispaCy pipeline. Negex identifies different forms of negation patterns and was initially developed for application to clinical texts [32]. Subsequently, the named entities were normalized and filtered to improve data quality for subsequent analyses. Therefore, the extracted entities were matched against an external knowledge base and were either replaced with standard medical vocabulary or discarded if no match was found. For this purpose, we used the ScispaCy entity linker, which matches entities with the unified medical language system (UMLS) knowledge base [33]. The EntityLinker pipeline performs a string nearest-neighbor search for entities to match them with the UMLS concepts [29]. We considered 0.85 as a threshold value for the overlap with UMLS concepts [29]. To evaluate the entity extraction performance, 500 randomly selected posts from the entire corpus were manually annotated. For data annotation, two annotators were involved. The intercoder reliability was 0.94. The F1 score was used as the evaluation metric [34]. By applying inexact string matching between data from the Food and Drug Administration's Orange Book [35] and the extracted entities, brand names were normalized by their active ingredient [36]. For instance, "zyrtec" was replaced with "ceterizine hydrochloride."

## Network Analysis

The network analysis method offers the possibility of visualizing and evaluating relationships in text. In this study, we used network analysis to obtain an overview of the spectrum of substances and identify potential substance clusters. Similar approaches have been used to identify substances and their effects, including self-medication in opioid withdrawal [20] or pharmacovigilance settings [21].

Features for network analysis consist of the nodes (represented by the extracted entities) and the "edges," which represent the relationship between the nodes as weight based on the co-occurrence of entities. A co-occurrence was defined as the mention of two or more substances in one post [37,38]. Duplicates of entities within one post were removed to avoid assigning more weight to longer posts that mention specific entities more frequently. Subsequently, the information on the co-occurrence of substances was converted into a pointwise mutual information (PMI) matrix [20,39,40]. We only considered associations between entities that co-occurred more frequently than expected based on their overall frequency, also called positive PMI, which was proven to be beneficial for extracting semantic representations [20,41]. To improve the quality of the visualization and analysis, substances occurring less than 10 times and node pairs below the average PMI weight were excluded [22,42]. False-positive nodes were manually removed. Using the PMI matrix, an undirected graph was created and analyzed using Gephi software. Gephi is an open-source software for network analysis, which allows spatialization, filtering, navigation, manipulation, and clustering of entities [43].

## Community Detection

We used clustering, also referred to as "community detection," to identify potential drug and/or supplement strategies for LC.

Community detection describes the clustering of nodes (in our case, substances) that are strongly associated with each other according to their edges. Hence, a cluster consists of substances that are strongly associated and discussed with one another. Communities can be determined using various clustering algorithms. For this purpose, the relatively new Leiden algorithm was used [44]. In contrast to the Louvain algorithm, which has been widely used in network analysis in the past, the Leiden algorithm has several advantages [45] such as having more meaningful partitioning [44]. The modularity value (Q) was used as the quality function [44]. A Q value of at least 0.3 implies meaningful clustering [44]. The result of clustering, and consequently Q, was significantly determined by the preselected resolution [46]. Following an iterative process, we aimed to find a proper balance between the number and relevance of the discovered communities and the resulting modularity by applying different resolution values [46]. To analyze the most important substances in the network, we calculated the degree centrality (number of linkages of a node) [47]; the higher the centrality, the more important the substance is in the network [47,48].

## Results

### Substances in Self-Reports

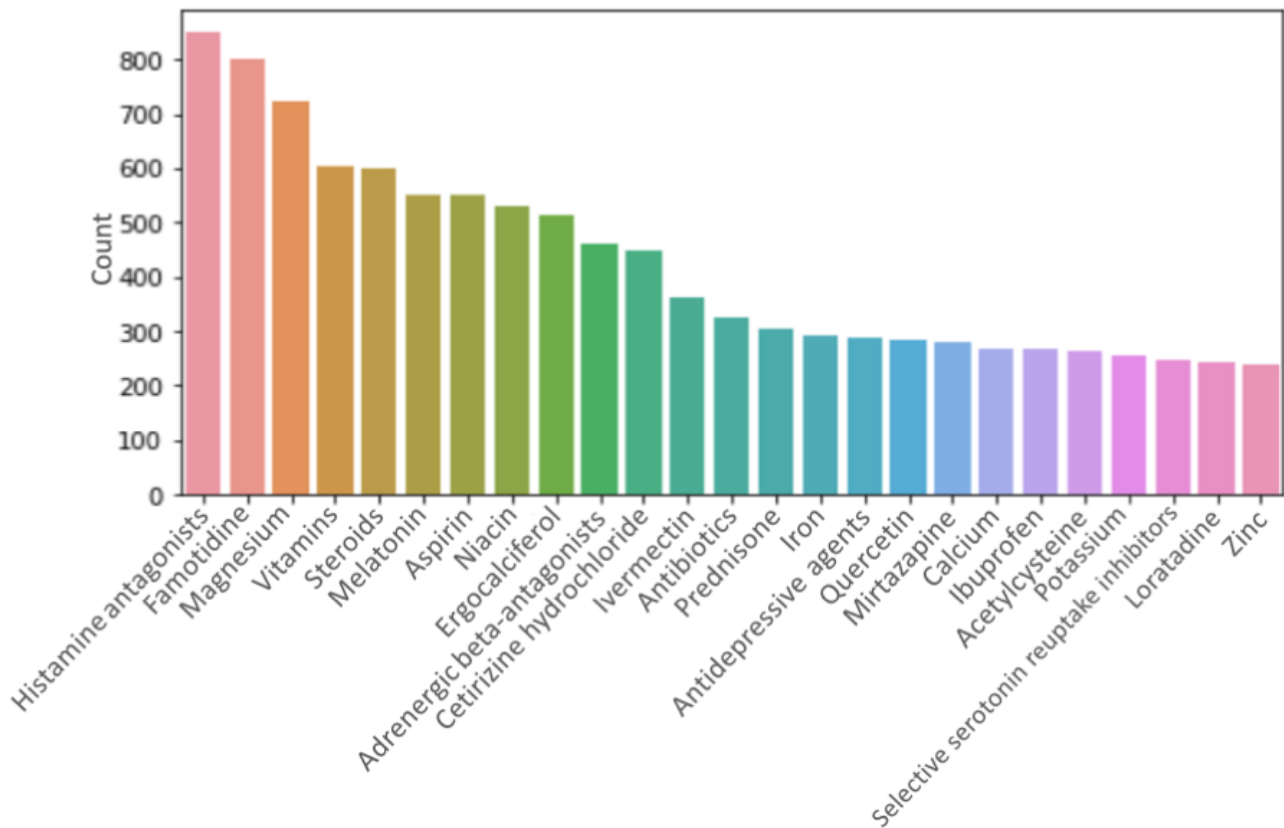
The NER algorithm achieved an F1 score of 0.67 (precision=0.69, recall=0.66). Error analysis was performed on the incorrectly labeled entities. Errors were classified into lexical and dictionary errors [49]. A lexical error (38.5%) refers to the case in which users employ a variety of terms when referring to the substances they use. For example, our model failed to detect and extract the term "benzos," since it is a slang abbreviation for the drug "benzodiazepine" and was not indexed in the model. Another example of expressions that our model did not recognize are compound terms of more than a single word (eg, "anti histamines") with the algorithm only extracting "histamines" and was thus missing the preceding word "anti," giving the extracted entity a different meaning.

A dictionary error (61.5%) refers to certain terms that are not specifying a concrete substance but rather a substance group; for example, "electrolytes" were not captured, whereas explicitly mentioned substances representing electrolytes such as "magnesium" were reliably detected. Furthermore, the algorithm extracted substances that are not considered as treatments, such as "chlorine."

A total of 28,447 substance entities and 5789 word-co-occurrence pairs were extracted. "Histamine antagonists," "famotidine," "magnesium," "vitamins," and "steroids" were the most frequently mentioned substances that appeared at least once in a post (duplicate mentions of a substance in a post were disregarded) (Figure 3). A list of all substances can be found in [Multimedia Appendix 1](#).

The most frequent word pairs are listed in [Table 1](#). For example, the pairing that occurred most frequently with 218 mentions was cetirizine hydrochloride–famotidine.

**Figure 3.** The 25 most frequently mentioned substances that appeared at least once in a post. For example, “histamine antagonists” were discussed in more than 800 different posts.



**Table 1.** The most frequent co-occurrences.

Rank	Substance–Substance pair	Frequency (number of mentions)
1	Cetirizine Hydrochloride–Famotidine	218
2	Famotidine–Histamine Antagonists	135
3	Potassium–Magnesium	106
4	Famotidine–Loratadine	98
5	Ergocalciferol–Magnesium	96
6	Cetirizine Hydrochloride–Histamine Antagonists	95
7	Aspirin–Famotidine	88
8	Loratadine–Histamine Antagonists	82
9	Zinc–Ascorbic Acid	78
10	Famotidine–Melatonin	78

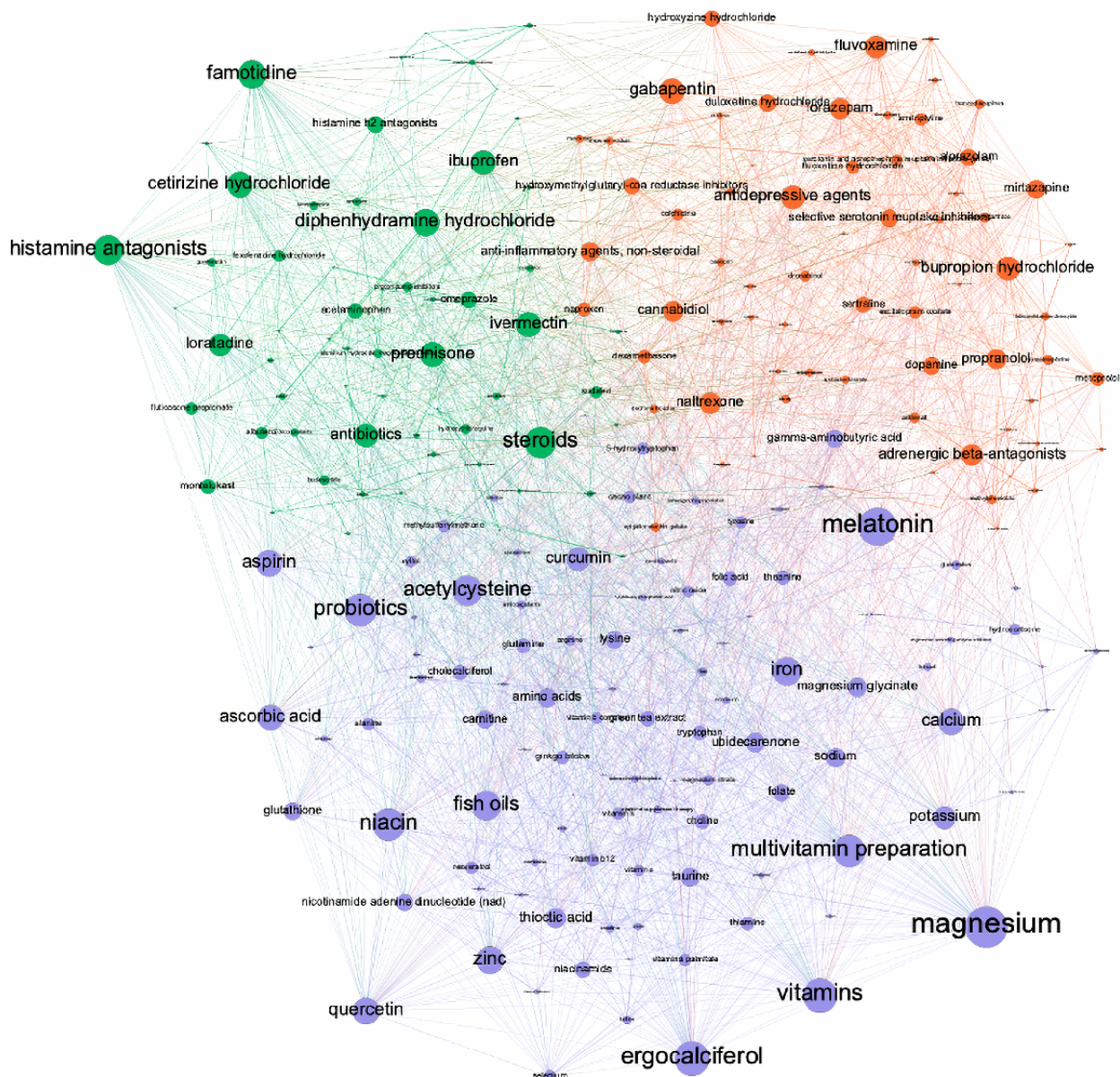
## Substance Clusters

### Overview

Using a resolution of 0.6, three clusters were found. They consisted of 244 nodes and 3570 edges. The modularity value

was 0.48, indicating a reasonable partitioning of communities [50,51]. The average clustering coefficient was 0.414. Overall, these scores indicated that the network (Figure 4) had no random structure [47]. Coloring in the network indicates the community and node size degree centrality.

**Figure 4.** Substance network and clusters. Substances are presented by nodes; the larger the size of a node, the higher degree centrality. Coloring refers to detected communities; violet represents cluster 1, orange refers to cluster 2, and green highlights substances of cluster 3.



### Cluster 1

Cluster 1 mainly consisted of supplements and several over-the-counter (OTC) medications, which are often used in the context of flu-like diseases (Figure 5). The top 10 most important substances and the respective substance classes

measured by degree centrality are displayed in Table 2. The retrieved entities belong to the drug classes of electrolyte/mineral replacement, vitamins, respiratory tract agents such as acetylcysteine, and nutritional supplements such as fish oil or probiotics.

Figure 5. Examples of posts containing typical substance co-occurrences of clusters.

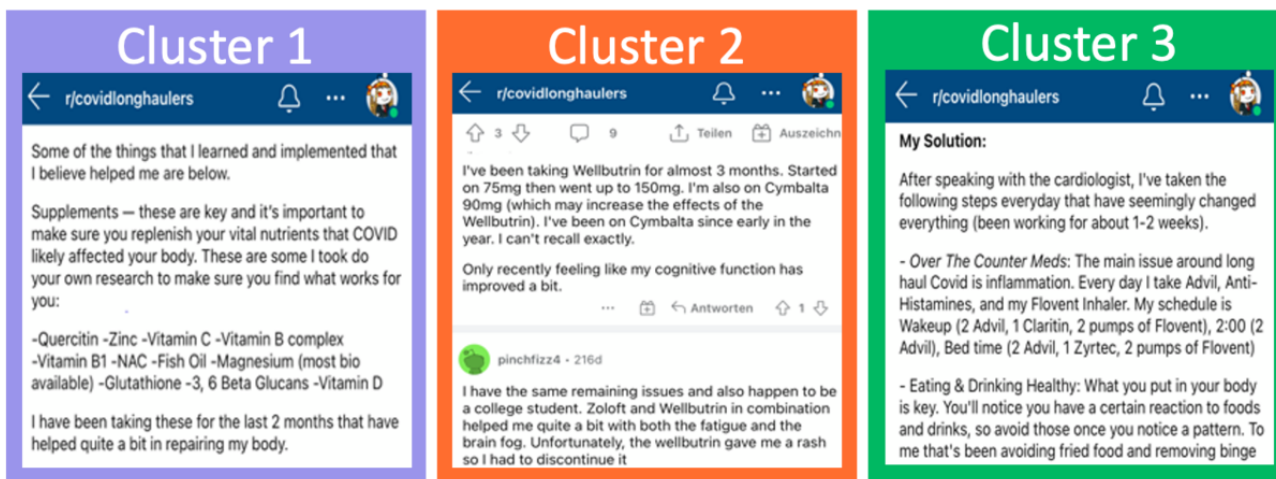


Table 2. Characteristics of clusters.

Cluster	Total share of nodes	Ten most important substances (by degree centrality)
1	42.85%	magnesium, melatonin, ergocalciferol, vitamin, multivitamin preparation, niacin, probiotics, acetylsteine, fish oils, zinc
2	29.51%	gabapentin, bupropion hydrochloride, antidepressive agents, fluvoxamine, adrenergic beta-antagonists, naltrexone, lorazepam, cannabidiol, propranolol, nonsteroidal anti-inflammatory agents
3	26.64%	steroids, histamine antagonists, famotidine, diphenhydramine hydrochloride, cetirizine hydrochloride, prednisone, ibuprofen, antibiotics, loratadine, ivermectin

### Cluster 2

Cluster 2 mostly included prescription medicines such as those used for the treatment of psychological, mental, or neurological disorders (Figure 4). Examples of the 10 most important substances (Table 2) include the anticonvulsant drug gabapentin, antidepressants such as bupropion hydrochloride, and adrenergic β-antagonists such as propranolol. Furthermore, opioid antagonists such as naltrexone, anxiolytics such as lorazepam, and nonsteroidal anti-inflammatory agents such as naproxen incorporated high degree centrality.

### Cluster 3

Cluster 3 mainly included prescription and OTC medicines that are often used to treat allergic reactions and inflammation (Figure 4). The top 10 important substances (Table 2) belonged to the drug classes of steroids such as prednisone, antihistamines such as famotidine, antibiotics, and nonsteroidal anti-inflammatory agents such as ibuprofen.

## Discussion

### Principal Results

Posts were extracted from an LC-specific subreddit to analyze the discussed substances with the aim of evaluating whether the proposed method could be used to support hypothesis development for DR. In the absence of approved medications for LC (which will also be the case in future pandemic situations), all substances can be considered for off-label use, which makes it difficult to evaluate the potential candidates that apply traditional SMM DR approaches. Instead of filtering substances for off-label use, we considered frequencies and

network analysis to facilitate the identification of important substances from the patients’ point of view.

The substances mentioned the most frequently in our feasibility study were antihistamines in general and famotidine, followed by supplements such as magnesium. Moreover, vitamins and steroids were frequently discussed substances. To analyze the strength of the substance-substance combination, the PMI was used to compare the strength of the associations with random associations considering the overall frequency of substances. These substances and their associations formed a nonrandom network consisting of three substance clusters, identified by community detection, implying systematic discussion and usage of substances. For instance, the most frequently mentioned class of substances, histamine antagonists, was found to be highly associated with other inflammatory substances such as steroids.

The clusters, mainly consisting of anti-inflammatory agents and supplements, incorporate the most often mentioned entities. Moreover, medications used for the treatment of psychological, mental, or neurological disorders also formed a cluster. The latter cluster can be assumed to be a less prevalent treatment regime in our sample because its substances occur less frequently. Nevertheless, all clusters reflect the treatment approaches described by the users.

### Supporting DR Hypothesis Generation by Analyzing Patients’ Self-Reports

The results of our feasibility study highlight drugs and self-treatment strategies discussed by long-haulers on Reddit. We were able to successfully identify substance communities (representing different treatment strategies) in the substance network and drugs of high(er) importance to users (based on

degree centrality) within these communities. Comparing the results to the current literature, our findings are supported by the successful identification of promising drug candidates already discussed by the scientific community. For instance, according to Crook et al [52], antihistamines are considered potential DR candidates. Significant improvements in long-term symptoms have been reported in case reports [53] and observational studies [54]. This seems to be backed by discussions from long-haulers: antihistamines are the most frequently discussed substances and are important (measured by degree centrality) to the cluster of anti-inflammatory drugs.

Similarly, Crook et al [52] concluded in their review that antidepressants such as serotonin-norepinephrine reuptake inhibitors and selective serotonin reuptake inhibitors could be repurposed for the treatment of LC, as they have been associated with a reduced risk of death or intubation in acute COVID-19 cases [52,55] and a reduction in peripheral inflammatory markers [52,56]. In April 2021, Sukhatme et al [57] conducted review on the mechanisms of action of fluvoxamine and its role in acute COVID-19 treatment. The authors concluded that it “is also tempting to speculate on a role for fluvoxamine in COVID-19 long-haulers.” In April 2022, Khani et al [58] expressed the hypothesis that the majority of LC symptoms might not be directly due to COVID-19 but likely result from COVID-19-associated inflammation and Epstein-Barr virus (EBV) reactivation. The authors argue that fluvoxamine might have beneficial effects in reducing LC symptoms due to the modulatory effects on central mechanisms (eg, reduction of endoplasmic reticulum stress and inflammation). In medication cluster 2, fluvoxamine was the second most important antidepressant, as measured by degree centrality (Table 2). Interestingly, our data are based on posts up to March 2022, and clearly indicate that users already consider this drug to be important in their LC treatment strategies while the scientific community is still working on hypothesis generation.

The other antidepressant that is most frequently mentioned and most central to long-hauler discussions was bupropion hydrochloride, representing a norepinephrine/dopamine reuptake inhibitor (NDRI) [59]. This drug is currently barely discussed in the research community. There seems to be a significant discrepancy in the perception of importance of NDRIs of long-haulers and the scientific community.

Additional drugs that appear to be important to long-haulers but that are barely discussed in clinical research include naltrexone, adrenergic  $\beta$ -antagonists, and prednisone. For instance, prednisone, as a corticosteroid, emerged as the most central steroid used for self-treatment by long-haulers in our study. Chen et al [60] found a high incidence of EBV coinfection in acute COVID-19 cases and concluded that patients may be advised to use a corticosteroid. Following the hypothesis of Khani et al [58] that EBV is also of central importance in LC, corticosteroid could be useful in the treatment of LC, and thus explain the identified central position in the long-haulers discussion identified in this study. In fact, Goel et al [61] reported that systemic steroids are helpful in hastening the recovery of a selected subset of patients with LC [61]. Another recently published single-center interventional pre-post study demonstrated that low-dose naltrexone is safe in patients with

LC and may improve well-being (eg, reduce symptomatology) [62].

In summary, reports on clinical research indicate that our proposed method might support the early identification of promising DR candidates. When incorporating patients' experiences regarding specific drugs, network analysis has proven to be especially useful to slice the data in a meaningful way. In particular, network analysis enables the identification of different (self-) treatment strategies and corresponding drugs beyond raw frequencies. For example, medications used for the treatment of psychological, mental, or neurological disorders represent a medication cluster; however, these drugs are generally reported at lower frequency. This might result from the fact that most of these drugs are only available on prescription and therefore access is limited in contrast to OTCs. Substance community detection and degree centrality highlight the strategies and drugs that otherwise could be overlooked. Furthermore, network analysis allows to distinguish between systematic and random discussions of substances. This is important information, as a systematic discussion might be an indicator for data quality and the knowledge of the crowd. Our results imply that patients' experiences shared on social media influence others' self-treatment decisions [63]. Positive experiences reported by users will lead to other users adopting the same approaches, leading to the increase of discussion of potentially helpful substances [64].

However, it is beyond the scope of this study to review all of the identified compounds for their potential relevance to LC DR. We encourage professionals to consider the findings as a starting point for hypothesis generation by narrowing down potential DR candidates to drugs that are important for long-haulers. Clearly, we cannot derive any conclusions about effectiveness based on this analysis. Nevertheless, these drugs appear to be frequently used by long-haulers as treatments. Therefore, these drugs should be further evaluated by the scientific community to determine whether they might be effective or even harmful, which would also be of relevance to communicate from a public health perspective.

### Limitations and Future Work

Our study has several limitations in line with comparable studies based on similar methods. As our NER algorithm relies on pretrained models, the error analysis we performed implies that a custom model trained on annotated examples of the posts used in this study would increase the accuracy of the results. To avoid missing entities in the normalization process (eg, due to the use of slang), a customized dictionary could be defined, which links slang terms of substances used by patients to their medical terminology. Further, substances that fail in the normalization process could be revised and normalized manually. However, the performance of the NER algorithm can be considered to be appropriate. Other studies applying pretrained NER algorithms to extract medical entities from Reddit data showed similar F1 scores. Foufi et al [49] used PKDE4J [65] to identify biomedical substances from disease-specific subreddits, and 71.48% of the extracted entities were correctly labeled. Šćepanović et al [66] evaluated six state-of-the-art pretrained language models in a bidirectional long short-term memory-conditional random



field (BiLSTM-CRF) model to identify medical entities in disease-specific subreddits; the F1 scores ranged from 0.64 to 0.73. Approaches combining different pretrained language models can improve performance. For example, Šćepanović et al [66] built a customized NER system by combining a BiLSTM-CRF sequence labeling architecture with contextual embeddings, which scored 0.71 on symptoms and 0.77 on drugs.

### Determining the Outcome of Substances

We analyzed the importance of substances in the long-haulers' discussions, but we did not analyze whether the substances were helpful in terms of the outcomes, which should be considered in future studies. Several approaches can be applied to approximate outcomes. The gold standard is the manual examination of medication-specific posts by medical professionals incorporating domain knowledge. Further programmatic approaches could include analyses of average sentiments regarding drugs to determine whether a treatment is perceived to be useful by users [67]. Another possibility could be to capture correlations between substances and effects mentioned in a sentence by the application of dependency parsing [68] or by implementing an observational study design [69]. However, reliable assessment of causality of underlying treatment effects is impossible because of various limitations [19]; for instance, inaccurate use of medical terminology by users would bias the results, even if machine-learning algorithms perform with perfect accuracy. Moreover, our analysis indicates a trend of polymedication, which would confound the analysis of single substances. Furthermore, the data quality is low compared with data obtained in traditional study designs, and several confounders such as demographic variables are unknown. In general, even if holistic patient information is available, social media data should be interpreted with caution. For instance, ivermectin was considered a potential DR candidate for acute COVID-19. However, this is suspected to be supported by flawed research [70] and it was later found to be ineffective in randomized clinical trials [71]. While we cannot evaluate the efficacy of ivermectin in patients with LC, it can be assumed that users' treatment decisions may also be influenced by potentially flawed information.

### Generalizability of Results and Inference of Indications

Even though important substances were found from the users' perspective, the results cannot be generalized for all patients

with LC since demographics or symptom distributions were not analyzed. Although previous studies [26,27] indicate that the spectrum of symptoms in the subreddit broadly agrees with recently published studies on LC [26,27], it can be assumed that the treatment choice of users depends on their symptoms. Therefore, the implementation of a bimodal network to reveal correlations between symptoms and medications could be useful. Moreover, the demographic distribution of users might not be representative of the entire long-hauler population. In fact, research suggests that the distribution of Reddit is skewed by age and gender [72], which is a major limitation to SMM in general [19], indicating that the data source represents a young male subpopulation [72]. There are two possible improvements in future SMM studies: (1) combining data from multiple platforms could lower user bias depending on the specific platform; and (2) algorithms can be applied to infer demographic variables and analyze diverse user data, including metadata [19].

### Conclusions

In this feasibility study, we tested the application of SMM methods to support the development of hypotheses on LC DR. To this end, we extracted substance entities to analyze frequencies and co-occurrences and subsequently used them to identify substance clusters. Our results highlight certain approaches to DR, such as antihistamines, steroids, or antidepressants, while also indicating that patients experiment with a wide range of substances in a systematic manner. This feasibility study demonstrates that network analysis can be used to characterize the medication strategies discussed. Comparison with existing literature indicates that the approach identifies substances that are plausible candidates for drug repurposing. The comparison also showed that some substances are important from the users' point of view while they are barely discussed in the scientific community. These substances should be reviewed by experts to assess their potential efficacy or harmfulness. The result could either lead to a DR hypothesis or underline the need to communicate the potential risks of the drug in the community of long-haulers. In the context of a pandemic, the proposed method might be used to support DR hypothesis development by prioritizing substances that are important to users in a cost- and time-effective manner.

---

### Acknowledgments

This research was supported as part of the ATLAS project "Innovation and digital transformation in healthcare" by the state of North Rhine-Westphalia, Germany (grant number: ITG-1-1).

---

### Conflicts of Interest

None declared.

---

### Multimedia Appendix 1

Link flair text (LFT) tags included and excluded from the analysis, and total substances extracted from posts. [[DOCX File, 25 KB-Multimedia Appendix 1](#)]

---

### References

1. WHO Coronavirus (COVID-19) Dashboard. World Health Organization. URL: <https://covid19.who.int> [accessed 2022-08-30]
2. Lopez-Leon S, Wegman-Ostrosky T, Perelman C, Sepulveda R, Rebolledo P, Cuapio A, et al. More than 50 long-term effects of COVID-19: a systematic review and meta-analysis. *Sci Rep* 2021 Aug 09;11(1):16144. [doi: [10.1038/s41598-021-95565-8](https://doi.org/10.1038/s41598-021-95565-8)] [Medline: [34373540](https://pubmed.ncbi.nlm.nih.gov/34373540/)]
3. Michelen M, Manoharan L, Elkheir N, Cheng V, Dagens A, Hastie C, et al. Characterising long COVID: a living systematic review. *BMJ Glob Health* 2021 Sep 27;6(9):e005427 [FREE Full text] [doi: [10.1136/bmjgh-2021-005427](https://doi.org/10.1136/bmjgh-2021-005427)] [Medline: [34580069](https://pubmed.ncbi.nlm.nih.gov/34580069/)]
4. Nalbandian A, Sehgal K, Gupta A, Madhavan MV, McGroder C, Stevens JS, et al. Post-acute COVID-19 syndrome. *Nat Med* 2021 Apr 22;27(4):601-615 [FREE Full text] [doi: [10.1038/s41591-021-01283-z](https://doi.org/10.1038/s41591-021-01283-z)] [Medline: [33753937](https://pubmed.ncbi.nlm.nih.gov/33753937/)]
5. Pavli A, Theodoridou M, Maltezou HC. Post-COVID syndrome: incidence, clinical spectrum, and challenges for primary healthcare professionals. *Arch Med Res* 2021 Aug;52(6):575-581 [FREE Full text] [doi: [10.1016/j.arcmed.2021.03.010](https://doi.org/10.1016/j.arcmed.2021.03.010)] [Medline: [33962805](https://pubmed.ncbi.nlm.nih.gov/33962805/)]
6. Reynolds KJ, Vernon SD, Bouchery E, Reeves WC. The economic impact of chronic fatigue syndrome. *Cost Eff Resour Alloc* 2004 Jul 21;2(1):4 [FREE Full text] [doi: [10.1186/1478-7547-2-4](https://doi.org/10.1186/1478-7547-2-4)] [Medline: [15210053](https://pubmed.ncbi.nlm.nih.gov/15210053/)]
7. Wong TL, Weitzer DJ. Long COVID and myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS)-a systemic review and comparison of clinical presentation and symptomatology. *Medicina* 2021 May 26;57(5):418 [FREE Full text] [doi: [10.3390/medicina57050418](https://doi.org/10.3390/medicina57050418)] [Medline: [33925784](https://pubmed.ncbi.nlm.nih.gov/33925784/)]
8. Clinical studies related to COVID-19. *ClinicalTrials.gov*. URL: <https://clinicaltrials.gov/ct2/results?cond=COVID-19> [accessed 2022-08-30]
9. Pammolli F, Magazzini L, Riccaboni M. *Nat Rev Drug Discov* 2011 Jul 1;10(6):428-438. [doi: [10.1038/nrd3405](https://doi.org/10.1038/nrd3405)] [Medline: [21629293](https://pubmed.ncbi.nlm.nih.gov/21629293/)]
10. Pushpakom S, Iorio F, Eyers PA, Escott KJ, Hopper S, Wells A, et al. Drug repurposing: progress, challenges and recommendations. *Nat Rev Drug Discov* 2019 Jan 12;18(1):41-58. [doi: [10.1038/nrd.2018.168](https://doi.org/10.1038/nrd.2018.168)] [Medline: [30310233](https://pubmed.ncbi.nlm.nih.gov/30310233/)]
11. Cha Y, Erez T, Reynolds IJ, Kumar D, Ross J, Koytiger G, et al. Drug repurposing from the perspective of pharmaceutical companies. *Br J Pharmacol* 2018 Jan 18;175(2):168-180. [doi: [10.1111/bph.13798](https://doi.org/10.1111/bph.13798)] [Medline: [28369768](https://pubmed.ncbi.nlm.nih.gov/28369768/)]
12. Ng YL, Salim CK, Chu JJH. Drug repurposing for COVID-19: approaches, challenges and promising candidates. *Pharmacol Ther* 2021 Dec;228:107930 [FREE Full text] [doi: [10.1016/j.pharmthera.2021.107930](https://doi.org/10.1016/j.pharmthera.2021.107930)] [Medline: [34174275](https://pubmed.ncbi.nlm.nih.gov/34174275/)]
13. Cavalla D. Using human experience to identify drug repurposing opportunities: theory and practice. *Br J Clin Pharmacol* 2019 Apr 03;85(4):680-689. [doi: [10.1111/bcp.13851](https://doi.org/10.1111/bcp.13851)] [Medline: [30648285](https://pubmed.ncbi.nlm.nih.gov/30648285/)]
14. Tucker J, Day S, Tang W, Bayus B. Crowdsourcing in medical research: concepts and applications. *PeerJ* 2019;7:e6762. [doi: [10.7717/peerj.6762](https://doi.org/10.7717/peerj.6762)] [Medline: [30997295](https://pubmed.ncbi.nlm.nih.gov/30997295/)]
15. Sharma N. Patient centric approach for clinical trials: current trend and new opportunities. *Perspect Clin Res* 2015;6(3):134-138 [FREE Full text] [doi: [10.4103/2229-3485.159936](https://doi.org/10.4103/2229-3485.159936)] [Medline: [26229748](https://pubmed.ncbi.nlm.nih.gov/26229748/)]
16. Boudreau K, Lakhani K. Using the crowd as an innovation partner. *Harv Bus Rev* 2013 May;91(4):60-9, 140 [FREE Full text] [Medline: [23593768](https://pubmed.ncbi.nlm.nih.gov/23593768/)]
17. Charalabidis Y, Loukis E, Androutsopoulou A, Karkaletsis V, Triantafyllou A. Passive crowdsourcing in government using social media. *Transform Gov: People, Process Policy* 2014;8(2):283-306. [doi: [10.1108/TG-09-2013-0035](https://doi.org/10.1108/TG-09-2013-0035)]
18. Ahmed S, Rajput AE, Sarirete A, Aljaberi A, Alghanem O, Alsheraigi A. Studying unemployment effects on mental health: social media versus the traditional approach. *Sustainability* 2020 Oct 02;12(19):8130. [doi: [10.3390/su12198130](https://doi.org/10.3390/su12198130)]
19. Koss J, Rheinlaender A, Truebel H, Bohnet-Joschko S. Social media mining in drug development-Fundamentals and use cases. *Drug Discov Today* 2021 Dec;26(12):2871-2880 [FREE Full text] [doi: [10.1016/j.drudis.2021.08.012](https://doi.org/10.1016/j.drudis.2021.08.012)] [Medline: [34481080](https://pubmed.ncbi.nlm.nih.gov/34481080/)]
20. Preiss A, Baumgartner P, Edlund MJ, Bobashev GV. Using named entity recognition to identify substances used in the self-medication of opioid withdrawal: natural language processing study of Reddit data. *JMIR Form Res* 2022 Mar 30;6(3):e33919 [FREE Full text] [doi: [10.2196/33919](https://doi.org/10.2196/33919)] [Medline: [35353047](https://pubmed.ncbi.nlm.nih.gov/35353047/)]
21. Correia R, Li L, Rocha L. Monitoring potential drug interactions and reactions via network analysis of Instagram user timelines. *Pac Symp Biocomput* 2016;21:492-503 [FREE Full text] [Medline: [26776212](https://pubmed.ncbi.nlm.nih.gov/26776212/)]
22. Lewis JA, Gee PM, Ho CL, Miller LMS. Understanding why older adults with type 2 diabetes join diabetes online communities: semantic network analyses. *JMIR Aging* 2018 Jul 28;1(1):e10649 [FREE Full text] [doi: [10.2196/10649](https://doi.org/10.2196/10649)] [Medline: [31518243](https://pubmed.ncbi.nlm.nih.gov/31518243/)]
23. Wang Q, Zhang W, Cai H, Cao Y. Understanding the perceptions of Chinese women of the commercially available domestic and imported HPV vaccine: A semantic network analysis. *Vaccine* 2020 Dec 14;38(52):8334-8342. [doi: [10.1016/j.vaccine.2020.11.016](https://doi.org/10.1016/j.vaccine.2020.11.016)] [Medline: [33190947](https://pubmed.ncbi.nlm.nih.gov/33190947/)]
24. Luo C, Chen A, Cui B, Liao W. Exploring public perceptions of the COVID-19 vaccine online from a cultural perspective: semantic network analysis of two social media platforms in the United States and China. *Telemat Inform* 2021 Dec;65:101712 [FREE Full text] [doi: [10.1016/j.tele.2021.101712](https://doi.org/10.1016/j.tele.2021.101712)] [Medline: [34887618](https://pubmed.ncbi.nlm.nih.gov/34887618/)]
25. Baumgartner J, Zannettou S, Keegan B, Squire M, Blackburn J. The pushshift Reddit dataset. 2020 Presented at: Fourteenth International AAAI Conference on Web and Social Media (ICWSM 2020); June 8-11, 2020; Atlanta, GA.

26. Sarker A, Ge Y. Mining long-COVID symptoms from Reddit: characterizing post-COVID syndrome from patient reports. *JAMIA Open* 2021 Jul;4(3):ooab075 [FREE Full text] [doi: [10.1093/jamiaopen/ooab075](https://doi.org/10.1093/jamiaopen/ooab075)] [Medline: [34485849](https://pubmed.ncbi.nlm.nih.gov/34485849/)]
27. Sarabadani S, Baruah G, Fossat Y, Jeon J. Longitudinal changes of COVID-19 symptoms in social media: observational study. *J Med Internet Res* 2022 Feb 16;24(2):e33959 [FREE Full text] [doi: [10.2196/33959](https://doi.org/10.2196/33959)] [Medline: [35076400](https://pubmed.ncbi.nlm.nih.gov/35076400/)]
28. Mansouri A, Affendey L, Mamat A. Named entity recognition approaches. *Int J Comput Sci Netw Secur* 2008;8(2):339-344.
29. Neumann M, King D, Beltagy I, Ammar W. ScispaCy: fast and robust models for biomedical natural language processing. arXiv. 2019. URL: <https://arxiv.org/abs/1902.07669> [accessed 2022-08-30]
30. Li J, Sun Y, Johnson R, Sciaky D, Wei C, Leaman R, et al. BioCreative V CDR task corpus: a resource for chemical disease relation extraction. *Database* 2016;2016:baw068 [FREE Full text] [doi: [10.1093/database/baw068](https://doi.org/10.1093/database/baw068)] [Medline: [27161011](https://pubmed.ncbi.nlm.nih.gov/27161011/)]
31. Kormilitzin A, Vaci N, Liu Q, Nevado-Holgado A. Med7: a transferable clinical natural language processing model for electronic health records. *Artif Intell Med* 2021 Aug;118:102086. [doi: [10.1016/j.artmed.2021.102086](https://doi.org/10.1016/j.artmed.2021.102086)] [Medline: [34412834](https://pubmed.ncbi.nlm.nih.gov/34412834/)]
32. Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. A simple algorithm for identifying negated findings and diseases in discharge summaries. *J Biomed Inform* 2001 Oct;34(5):301-310 [FREE Full text] [doi: [10.1006/jbin.2001.1029](https://doi.org/10.1006/jbin.2001.1029)] [Medline: [12123149](https://pubmed.ncbi.nlm.nih.gov/12123149/)]
33. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Res* 2004 Jan 01;32(Database issue):D267-D270 [FREE Full text] [doi: [10.1093/nar/gkh061](https://doi.org/10.1093/nar/gkh061)] [Medline: [14681409](https://pubmed.ncbi.nlm.nih.gov/14681409/)]
34. Hripcsak G. Agreement, the F-measure, and reliability in information retrieval. *J Am Med Inform Assoc* 2005 Jan 31;12(3):296-298. [doi: [10.1197/jamia.m1733](https://doi.org/10.1197/jamia.m1733)]
35. Orange book: approved drug products with therapeutic equivalence evaluations. US Food and Drug Administration. URL: <https://www.fda.gov/drugs/drug-approvals-and-databases/approved-drug-products-therapeutic-equivalence-evaluations-orange-book> [accessed 2022-08-30]
36. Ma'ayan A, Jenkins S, Goldfarb J, Iyengar R. Network analysis of FDA approved drugs and their targets. *Mt Sinai J Med* 2007 May;74(1):27-32 [FREE Full text] [doi: [10.1002/msj.20002](https://doi.org/10.1002/msj.20002)] [Medline: [17516560](https://pubmed.ncbi.nlm.nih.gov/17516560/)]
37. Marshall S, Yang C, Ping Q, Zhao M, Avis N, Ip E. Symptom clusters in women with breast cancer: an analysis of data from social media and a research study. *Qual Life Res* 2016 Mar;25(3):547-557 [FREE Full text] [doi: [10.1007/s11136-015-1156-7](https://doi.org/10.1007/s11136-015-1156-7)] [Medline: [26476836](https://pubmed.ncbi.nlm.nih.gov/26476836/)]
38. Vitte J, Gao F, Coppola G, Judkins AR, Giovannini M. Timing of Smarcb1 and Nf2 inactivation determines schwannoma versus rhabdoid tumor development. *Nat Commun* 2017 Aug 21;8(1):300. [doi: [10.1038/s41467-017-00346-5](https://doi.org/10.1038/s41467-017-00346-5)] [Medline: [28824165](https://pubmed.ncbi.nlm.nih.gov/28824165/)]
39. Liu C, Lu X. Analyzing hidden populations online: topic, emotion, and social network of HIV-related users in the largest Chinese online community. *BMC Med Inform Decis Mak* 2018 Jan 05;18(1):2 [FREE Full text] [doi: [10.1186/s12911-017-0579-1](https://doi.org/10.1186/s12911-017-0579-1)] [Medline: [29304788](https://pubmed.ncbi.nlm.nih.gov/29304788/)]
40. Teng C, Lin Y, Adamic L. Recipe recommendation using ingredient networks. 2012 Presented at: 4th annual ACM Web Science Conference (WebSci '12); June 22-24, 2012; Evanston, IL. [doi: [10.1145/2380718.2380757](https://doi.org/10.1145/2380718.2380757)]
41. Bullinaria JA, Levy JP. Extracting semantic representations from word co-occurrence statistics: stop-lists, stemming, and SVD. *Behav Res Methods* 2012 Oct 19;44(3):890-907. [doi: [10.3758/s13428-011-0183-8](https://doi.org/10.3758/s13428-011-0183-8)] [Medline: [22258891](https://pubmed.ncbi.nlm.nih.gov/22258891/)]
42. Calabrese C, Ding J, Millam B, Barnett GA. The uproar over gene-edited babies: a semantic network analysis of CRISPR on Twitter. *Env Commun* 2019 Dec 19;14(7):954-970. [doi: [10.1080/17524032.2019.1699135](https://doi.org/10.1080/17524032.2019.1699135)]
43. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. 2009 Presented at: Third International AAAI Conference on Weblogs and Social Media; May 17-20, 2009; San Jose, CA.
44. Traag VA, Waltman L, van Eck NJ. From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep* 2019 Mar 26;9(1):5233. [doi: [10.1038/s41598-019-41695-z](https://doi.org/10.1038/s41598-019-41695-z)] [Medline: [30914743](https://pubmed.ncbi.nlm.nih.gov/30914743/)]
45. Hairol Anuar SH, Abas ZA, Yunus NM, Mohd Zaki NH, Hashim NA, Mokhtar MF, et al. Comparison between Louvain and Leiden Algorithm for Network Structure: A Review. 2021 Presented at: 1st International Conference on Material Processing and Technology (ICMProTech 2021); July 14-15, 2021; Perlis, Malaysia. [doi: [10.1088/1742-6596/2129/1/012028](https://doi.org/10.1088/1742-6596/2129/1/012028)]
46. Arroyo-Machado W, Torres-Salinas D, Robinson-Garcia N. Identifying and characterizing social media communities: a socio-semantic network approach to altmetrics. *Scientometrics* 2021 Oct 12;126(11):9267-9289 [FREE Full text] [doi: [10.1007/s11192-021-04167-8](https://doi.org/10.1007/s11192-021-04167-8)] [Medline: [34658460](https://pubmed.ncbi.nlm.nih.gov/34658460/)]
47. Dotsika F, Watkins A. Identifying potentially disruptive trends by means of keyword network analysis. *Technol Forecast Soc Change* 2017 Jun;119:114-127. [doi: [10.1016/j.techfore.2017.03.020](https://doi.org/10.1016/j.techfore.2017.03.020)]
48. Hoser B, Hotho A, Jäschke R, Schmitz C, Stumme G. Semantic network analysis of ontologies. In: Sure Y, Domingue J, editors. *The Semantic Web: Research and Applications*. ESWC 2006. Lecture Notes in Computer Science, vol 4011. Berlin, Heidelberg: Springer; 2006.
49. Foufi V, Timakum T, Gaudet-Blavignac C, Lovis C, Song M. Mining of textual health information from Reddit: analysis of chronic diseases with extracted entities and their relations. *J Med Internet Res* 2019 Jun 13;21(6):e12876 [FREE Full text] [doi: [10.2196/12876](https://doi.org/10.2196/12876)] [Medline: [31199327](https://pubmed.ncbi.nlm.nih.gov/31199327/)]
50. Varsha K, Patil K. An overview of community detection algorithms in social networks. 2020 Presented at: International Conference on Inventive Computation Technologies (ICICT), 2020; February 26-28, 2020; Coimbatore, Tamilnadu, India. [doi: [10.1109/iciict48043.2020.9112563](https://doi.org/10.1109/iciict48043.2020.9112563)]

51. Newman MEJ. Modularity and community structure in networks. *Proc Natl Acad Sci U S A* 2006 Jul 06;103(23):8577-8582 [FREE Full text] [doi: [10.1073/pnas.0601602103](https://doi.org/10.1073/pnas.0601602103)] [Medline: [16723398](https://pubmed.ncbi.nlm.nih.gov/16723398/)]
52. Crook H, Raza S, Nowell J, Young M, Edison P. Long covid-mechanisms, risk factors, and management. *BMJ* 2021 Jul 26;374:n1648. [doi: [10.1136/bmj.n1648](https://doi.org/10.1136/bmj.n1648)] [Medline: [34312178](https://pubmed.ncbi.nlm.nih.gov/34312178/)]
53. Pinto MD, Lambert N, Downs CA, Abraham H, Hughes TD, Rahmani AM, et al. Antihistamines for postacute sequelae of SARS-CoV-2 infection. *J Nurse Pract* 2022 Mar;18(3):335-338 [FREE Full text] [doi: [10.1016/j.nurpra.2021.12.016](https://doi.org/10.1016/j.nurpra.2021.12.016)] [Medline: [35153633](https://pubmed.ncbi.nlm.nih.gov/35153633/)]
54. Glynne P, Tahmasebi N, Gant V, Gupta R. Long COVID following mild SARS-CoV-2 infection: characteristic T cell alterations and response to antihistamines. *J Investig Med* 2022 Jan 05;70(1):61-67 [FREE Full text] [doi: [10.1136/jim-2021-002051](https://doi.org/10.1136/jim-2021-002051)] [Medline: [34611034](https://pubmed.ncbi.nlm.nih.gov/34611034/)]
55. Hoertel N, Sánchez-Rico M, Vernet R, Beeker N, Jannot A, Neuraz A, AP-HP/Universities/INSERM COVID-19 Research Collaboration AP-HP COVID CDR Initiative. Association between antidepressant use and reduced risk of intubation or death in hospitalized patients with COVID-19: results from an observational study. *Mol Psychiatry* 2021 Sep 04;26(9):5199-5212. [doi: [10.1038/s41380-021-01021-4](https://doi.org/10.1038/s41380-021-01021-4)] [Medline: [33536545](https://pubmed.ncbi.nlm.nih.gov/33536545/)]
56. Köhler CA, Freitas TH, Stubbs B, Maes M, Solmi M, Veronese N, et al. Peripheral alterations in cytokine and chemokine levels after antidepressant drug treatment for major depressive disorder: systematic review and meta-analysis. *Mol Neurobiol* 2018 May 13;55(5):4195-4206. [doi: [10.1007/s12035-017-0632-1](https://doi.org/10.1007/s12035-017-0632-1)] [Medline: [28612257](https://pubmed.ncbi.nlm.nih.gov/28612257/)]
57. Sukhatme VP, Reiersen AM, Vayttaden SJ, Sukhatme VV. Fluvoxamine: a review of its mechanism of action and its role in COVID-19. *Front Pharmacol* 2021 Apr 20;12:652688. [doi: [10.3389/fphar.2021.652688](https://doi.org/10.3389/fphar.2021.652688)] [Medline: [33959018](https://pubmed.ncbi.nlm.nih.gov/33959018/)]
58. Khani E, Entezari-Maleki T. Fluvoxamine and long COVID-19; a new role for sigma-1 receptor (S1R) agonists. *Mol Psychiatry* 2022 May 06:online ahead of print [FREE Full text] [doi: [10.1038/s41380-022-01545-3](https://doi.org/10.1038/s41380-022-01545-3)] [Medline: [35388182](https://pubmed.ncbi.nlm.nih.gov/35388182/)]
59. Krystal AD, Thase ME, Tucker VL, Goodale EP. Bupropion HCL and sleep in patients with depression. *Curr Psych Rev* 2007 May 01;3(2):123-128. [doi: [10.2174/157340007780599096](https://doi.org/10.2174/157340007780599096)]
60. Chen T, Song J, Liu H, Zheng H, Chen C. Positive Epstein-Barr virus detection in coronavirus disease 2019 (COVID-19) patients. *Sci Rep* 2021 May 25;11(1):10902. [doi: [10.1038/s41598-021-90351-y](https://doi.org/10.1038/s41598-021-90351-y)] [Medline: [34035353](https://pubmed.ncbi.nlm.nih.gov/34035353/)]
61. Goel N, Goyal N, Nagaraja R, Kumar R. Systemic corticosteroids for management of 'long-COVID': an evaluation after 3 months of treatment. *Monaldi Arch Chest Dis* 2021 Dec 03;92(2):1981. [doi: [10.4081/monaldi.2021.1981](https://doi.org/10.4081/monaldi.2021.1981)] [Medline: [34730322](https://pubmed.ncbi.nlm.nih.gov/34730322/)]
62. O'Kelly B, Vidal L, McHugh T, Woo J, Avramovic G, Lambert JS. Safety and efficacy of low dose naltrexone in a long covid cohort; an interventional pre-post study. *Brain Behav Immun Health* 2022 Oct;24:100485 [FREE Full text] [doi: [10.1016/j.bbih.2022.100485](https://doi.org/10.1016/j.bbih.2022.100485)] [Medline: [35814187](https://pubmed.ncbi.nlm.nih.gov/35814187/)]
63. Benetoli A, Chen T, Aslani P. How patients' use of social media impacts their interactions with healthcare professionals. *Patient Educ Couns* 2018 Mar;101(3):439-444. [doi: [10.1016/j.pec.2017.08.015](https://doi.org/10.1016/j.pec.2017.08.015)] [Medline: [28882545](https://pubmed.ncbi.nlm.nih.gov/28882545/)]
64. Ogink T, Dong JQ. Stimulating innovation by user feedback on social media: The case of an online user innovation community. *Technol Forecast Soc Change* 2019 Jul;144:295-302. [doi: [10.1016/j.techfore.2017.07.029](https://doi.org/10.1016/j.techfore.2017.07.029)]
65. Song M, Kim WC, Lee D, Heo GE, Kang KY. PKDE4J: entity and relation extraction for public knowledge discovery. *J Biomed Inform* 2015 Oct;57:320-332 [FREE Full text] [doi: [10.1016/j.jbi.2015.08.008](https://doi.org/10.1016/j.jbi.2015.08.008)] [Medline: [26277115](https://pubmed.ncbi.nlm.nih.gov/26277115/)]
66. Scepanovic S, Martin-Lopez E, Quercia D, Baykaner K. Extracting medical entities from social media. 2020 Presented at: CHIL '20: ACM Conference on Health, Inference, and Learning; April 2-4, 2020; Toronto, ON. [doi: [10.1145/3368555.3384467](https://doi.org/10.1145/3368555.3384467)]
67. Gräßer F, Kallumadi S, Malberg H, Zaunseder S. Aspect-based sentiment analysis of drug reviews applying cross-domain and cross-data learning. 2018 Presented at: DH '18: 2018 International Conference on Digital Health; April 23-26, 2018; Lyon, France. [doi: [10.1145/3194658.3194677](https://doi.org/10.1145/3194658.3194677)]
68. Doan S, Yang EW, Tilak SS, Li PW, Zisook DS, Torii M. Extracting health-related causality from twitter messages using natural language processing. *BMC Med Inform Decis Mak* 2019 Apr 04;19(Suppl 3):79 [FREE Full text] [doi: [10.1186/s12911-019-0785-0](https://doi.org/10.1186/s12911-019-0785-0)] [Medline: [30943954](https://pubmed.ncbi.nlm.nih.gov/30943954/)]
69. Saha K, Sugar B, Torous J, Abrahao B, Kıcıman E, De Choudhury M. A social media study on the effects of psychiatric medication use. *Proc Int AAAI Conf Weblogs Soc Media* 2019 Jul 07;13:440-451 [FREE Full text] [Medline: [32280562](https://pubmed.ncbi.nlm.nih.gov/32280562/)]
70. Meyerowitz-Katz G, Wieten S, Medina Arellano MDJ, Yamey G. Unethical studies of ivermectin for covid-19. *BMJ* 2022 Apr 14;377:o917. [doi: [10.1136/bmj.o917](https://doi.org/10.1136/bmj.o917)] [Medline: [35422425](https://pubmed.ncbi.nlm.nih.gov/35422425/)]
71. Roman Y, Burela P, Pasupuleti V, Piscocoya A, Vidal J, Hernandez A. Ivermectin for the treatment of coronavirus disease 2019: a systematic review and meta-analysis of randomized controlled trials. *Clin Infect Dis* 2022 Mar 23;74(6):1022-1029 [FREE Full text] [doi: [10.1093/cid/ciab591](https://doi.org/10.1093/cid/ciab591)] [Medline: [34181716](https://pubmed.ncbi.nlm.nih.gov/34181716/)]
72. Amaya A, Bach R, Keusch F, Kreuter F. New data sources in social science research: things to know before working with Reddit data. *Soc Sci Comput Rev* 2019 Dec 18;39(5):943-960. [doi: [10.1177/0894439319893305](https://doi.org/10.1177/0894439319893305)]

## Abbreviations

**BiLSTM-CRF:** bidirectional long short-term memory–conditional random field

**DR:** drug repurposing  
**EBV:** Epstein-Barr virus  
**LC:** long COVID  
**NDRI:** norepinephrine/dopamine reuptake inhibitor  
**NER:** named-entity recognition  
**OTC:** over the counter  
**PMI:** pointwise mutual information  
**SMM:** social media mining  
**UMLS:** unified medical language system

*Edited by A Mavragani; submitted 16.05.22; peer-reviewed by A Sarker, X Wang; comments to author 06.06.22; revised version received 27.06.22; accepted 09.08.22; published 03.10.22*

*Please cite as:*

*Koss J, Bohnet-Joschko S*

*Social Media Mining of Long-COVID Self-Medication Reported by Reddit Users: Feasibility Study to Support Drug Repurposing*

*JMIR Form Res 2022;6(10):e39582*

*URL: <https://formative.jmir.org/2022/10/e39582>*

*doi: [10.2196/39582](https://doi.org/10.2196/39582)*

*PMID: [36007131](https://pubmed.ncbi.nlm.nih.gov/36007131/)*

©Jonathan Koss, Sabine Bohnet-Joschko. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 03.10.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.