

Original Paper

Attitudes of Crohn's Disease Patients: Infodemiology Case Study and Sentiment Analysis of Facebook and Twitter Posts

Marco Roccetti^{1*}, Dott Ing; Gustavo Marfia^{2*}, PhD; Paola Salomoni^{1*}, Dott; Catia Prandi^{3*}, PhD; Rocco Maurizio Zagari^{4*}, MD; Faustine Linda Gningaye Kengni^{4*}, MD; Franco Bazzoli^{4*}, MD; Marco Montagnani^{4*}, MD

¹Department of Computer Science and Engineering, University of Bologna, Bologna, Italy

²Department for Life Quality Studies, University of Bologna, Rimini, Italy

³Madeira Interactive Technologies Institute, Funchal, Portugal

⁴Gastroenterology Unit, Department of Medical and Surgical Sciences, University of Bologna, Bologna, Italy

* all authors contributed equally

Corresponding Author:

Gustavo Marfia, PhD

Department for Life Quality Studies

University of Bologna

Corso D'Augusto 237

Rimini,

Italy

Phone: 39 0541 434 044

Fax: 39 0541 434 044

Email: gustavo.marfia@unibo.it

Abstract

Background: Data concerning patients originates from a variety of sources on social media.

Objective: The aim of this study was to show how methodologies borrowed from different areas including computer science, econometrics, statistics, data mining, and sociology may be used to analyze Facebook data to investigate the patients' perspectives on a given medical prescription.

Methods: To shed light on patients' behavior and concerns, we focused on Crohn's disease, a chronic inflammatory bowel disease, and the specific therapy with the biological drug Infliximab. To gain information from the basin of big data, we analyzed Facebook posts in the time frame from October 2011 to August 2015. We selected posts from patients affected by Crohn's disease who were experiencing or had previously been treated with the monoclonal antibody drug Infliximab. The selected posts underwent further characterization and sentiment analysis. Finally, an ethnographic review was carried out by experts from different scientific research fields (eg, computer science vs gastroenterology) and by a software system running a sentiment analysis tool. The patient feeling toward the Infliximab treatment was classified as positive, neutral, or negative, and the results from computer science, gastroenterologist, and software tool were compared using the square weighted Cohen's kappa coefficient method.

Results: The first automatic selection process returned 56,000 Facebook posts, 261 of which exhibited a patient opinion concerning Infliximab. The ethnographic analysis of these 261 selected posts gave similar results, with an interrater agreement between the computer science and gastroenterology experts amounting to 87.3% (228/261), a substantial agreement according to the square weighted Cohen's kappa coefficient method ($w^2K=0.6470$). A positive, neutral, and negative feeling was attributed to 36%, 27%, and 37% of posts by the computer science expert and 38%, 30%, and 32% by the gastroenterologist, respectively. Only a slight agreement was found between the experts' opinion and the software tool.

Conclusions: We show how data posted on Facebook by Crohn's disease patients are a useful dataset to understand the patient's perspective on the specific treatment with Infliximab. The genuine, nonmedically influenced patients' opinion obtained from Facebook pages can be easily reviewed by experts from different research backgrounds, with a substantial agreement on the classification of patients' sentiment. The described method allows a fast collection of big amounts of data, which can be easily analyzed to gain insight into the patients' perspective on a specific medical therapy.

(*JMIR Public Health Surveill* 2017;3(3):e51) doi: [10.2196/publichealth.7004](https://doi.org/10.2196/publichealth.7004)

KEYWORDS

health information systems; public health informatics; consumer health information; social networking

Introduction

Patient opinions are highly valued in many medical studies for the assessment of their well-being. However, it is not always easy to collect patients' feedbacks for clinical studies. Interestingly, the advent of means of one-to-many communication, including the Web and social media, support peer-to-peer and one-to-many exchanges and comparisons of patients' experiences and feelings. Such Web-based tools have also radically changed the scenario in front of caregivers; patients are set in front of many more stimuli and sources of information than before (one-third of adult American citizens consider the Web a diagnostic tool), although no guarantee is granted on the quality of the retrieved information [1-3].

Nonetheless, Web-based anonymity may boost frankness and sincerity, as its privacy is often perceived as absolute, also when compared with the direct patient-doctor interactions. Sharing their experiences on the Web, patients provide a very useful knowledge base of insights to both rookies and medical researchers [4]: the former could learn how to handle given situations, and the latter could gather more sincere and unbiased feedback or even acquire further knowledge in their field of clinical study.

Although the reasons for understanding what is shared on the Web in relation to a given disease are clear, no well-established method exists today. Challenges, in fact, may be found and are not limited to (1) data gathering, (2) filtering of any unwanted or unnecessary information, (3) key topics individuation and interpretation, and (4) comparison to any related state-of-the-art in medical research.

An open question amounts to understand what the medical community could learn from the information that is shared on the Web [5-8]. Such new interesting area of research is part of the novel infoveillance and infodemiology fields. A few studies have considered such a problem in relation to different chronic diseases [9-16]. However, to the best of our knowledge, a general approach to this class of problems, based on the use of a combination of different technologies, is missing. This requires expertise that cannot stop to the medical or statistical fields but must also include techniques developed in computer science in addition to others from econometrics, ethnographic research, and psychometrics areas of study.

We borrowed the techniques from the aforementioned scientific areas to investigate a well-defined community of chronic illness patients affected by Crohn's disease. The choice of such a community is motivated by the following important fact: Crohn's disease is a chronic illness with increasing incidence, especially in western countries where it is often diagnosed in young people (in the age range of 15-30 years) who typically spend a lot of time on the Web [17]. Crohn's disease is therefore a good study model for our purposes.

The method that we present builds upon steps that we have previously developed [18,19]. In an initial analysis [18], computer science and econometrics techniques led us to find that (1) Crohn's disease patients share more frequently information on Facebook pages rather than in Twitter streams, and (2) the pharmaceutical treatment that is most often cited, in both positive and negative terms for Crohn's disease, is Infliximab. Further contributions have been made [19], where we put our findings in relation with small and large scale medical trials.

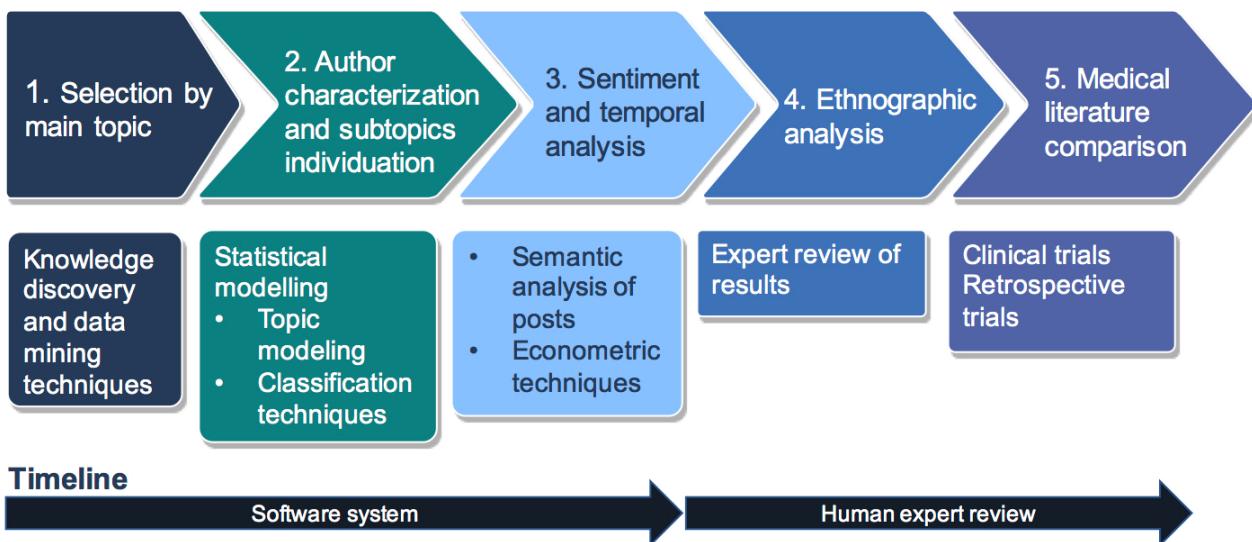
Now, the logic and contribution of this paper is to present a method on how Web-based patient information could be obtained and evaluated. To this aim, the following research questions (RQs) are considered:

1. Between Twitter and Facebook, which social media platform do people post on most frequently?
2. Which topics trigger the most patient reactions (eg, medical therapy satisfaction or dissatisfaction)?
3. What kind of attitude do patients have toward the most debated topic (eg, positive, neutral, or negative)?

The results of this study should be integrated with traditional research approaches to help clinicians understand patients' perspectives.

Methods

Answers to RQ1, RQ2, and RQ3 were obtained following the methods delineated in [Figure 1](#), where the problem of finding and analyzing Web-based data involves two steps. The first one (leftmost part of the timeline) relies completely on software components, whereas the second includes the intervention of human operators. Why this architectural choice has been made will become clear in the following subsections.

Figure 1. Web-based patient feedback analysis.

Topic Selection

To understand where patients share their experiences, we implemented a selection procedure (selection by main topic in Figure 1), a well-known operation in data mining and knowledge discovery [20,21]. In fact, no a priori knowledge may be available regarding where patients prefer sharing their experiences. Often the burden of such discovery process is very limited, as many forums and social media pages are often entirely dedicated to the discussion of given diseases. Hence, it is often simple to carry out this step accessing a great quantity of relevant data.

However, often posts are not written by patients (ie, many report scientific news or drug advertisements). Such a problem requires the implementation of mechanisms capable of identifying sites where patients publish their experiences. In our analysis of Crohn's disease patients, this has been done resorting to two different techniques known for the uncovering of social spammers [22,23].

The first technique simply amounts to identify nonhuman Web-based posts from the number of duplicate ones that may be associated to a single user account. In fact, duplicates are frequently associated to those accounts which are dedicated to post news or advertisements [22]. The second amounts to analyze the behaviors of single writers [23]. To this aim, we performed an additional test, assessing the role of the most prolific users on both social media (please note that this test could be performed automatically by a computer program) to determine whether they were patients or not.

Subtopic Individuation

The second step of interest is that of shaping the corpus of acquired data, characterizing and modeling it in terms of subtopics of interest. Four different subtopics have been individuated: lifestyle, symptoms, treatments, and side-effects, used to define four corresponding dictionaries. Such approach is consistent with previous works on medical data mining [24-26]. Within the lifestyle subtopic, we included all those terms that are related to the behavior of a patient (eg, food

consumption habits and smoker or nonsmoker). Symptoms, treatments, and side-effects contain, instead, the words representing the distinctive signs of a disease (eg, fever and high pressure), the names of the medications utilized to contrast it (eg, tylenol and paracetamol), along with any related side-effect (eg, dizziness), respectively.

For the sake of completeness, we note that the number of subtopics, in general, may be any. The area of topic individuation and modeling is an active area of research whose developments may prove to be very useful in such context, to reveal the topics treated in a corpus of posts [27]. In text data mining, the creation of dictionaries is called feature selection. A wide variety of feature selection methods exist. One of the most common methods for quantifying the discrimination level of a feature is the use of a measure known as the Gini-index [28]. In essence, let $p_i(w)$ be the conditional probability that a document belongs to class i , given the fact that it contains the word w . The Gini-index for word w , denoted by $G(w)$, is defined as $G(w) = \sum p_i(w)^2$ where k amounts to the number of classes. The value of $G(w)$ always lies in $(1/k, 1)$, with higher values of $G(w)$ associated to a higher discriminative power of the word w . Such an approach is very general, however. For the very specific situations, say a situation where we are interested at selecting those posts where users mention a specific medication, setting $w = \text{medication name}$ results a reliable indicator of an ongoing exchange regarding this topic.

Sentiment Analysis

After a topic has been identified and posts containing words pertaining to that topic selected, an additional step is performed to determine the relationship of a patient with the given topic. To this aim, sentiment analysis techniques have been exploited, as their performance is progressively becoming more accurate and reliable [29-33]. In this work we used University of Pittsburgh's OpinionFinder, but additional resources are freely available for the assessment of sentiment values in posts [34]. For example, the Apache OpenNLP framework could be utilized to classify text into predefined categories resorting to the maximum entropy algorithm [35]. Standford's StanfordNLP,

in addition, is a tool trained with 215,154 phrases with fine grained sentiment labeling [36].

Subsequently, in order to verify the correlation between given topics and given sentiment values, econometric approaches (eg, Granger causality) have been employed. Notably, we borrowed such an approach from social media data mining applied to stock exchange analysis [37]. Logistic regression approaches also appear viable for such a domain [38]. Simpler approaches could also be employed to verify the co-occurrence of negative or positive expressions with given key terms. In essence, various statistical analysis methodologies can be utilized to evaluate the importance of a given topic within post sentiment values.

Ethnographic Analysis

The use of software components in the chart shown in Figure 1 ends with the sentiment analysis step. After individuating the topic of greatest interest for patients, we analyzed, by ethnographic approach, the qualitative feeling of the patients on the specific issue. Since the use of the Infliximab therapy was the most discussed topic (see below), we adopted a 3-valued Likert scale to assess the sentiment value of a patient toward Infliximab [39]. A value of 1 was attributed to positive, 0 to neutral, and -1 to negative feelings. Because we wanted to investigate the reliability of such manual assessment, we compared the ethnographic analysis performed by a computer science researcher and a senior gastroenterologist. We then analyzed the concordance of such assessments using the square weighted Cohen's kappa coefficient method. Additionally, we also assessed the patients' feelings according to the 3-point Likert scale using our software system, which relied on OpinionFinder.

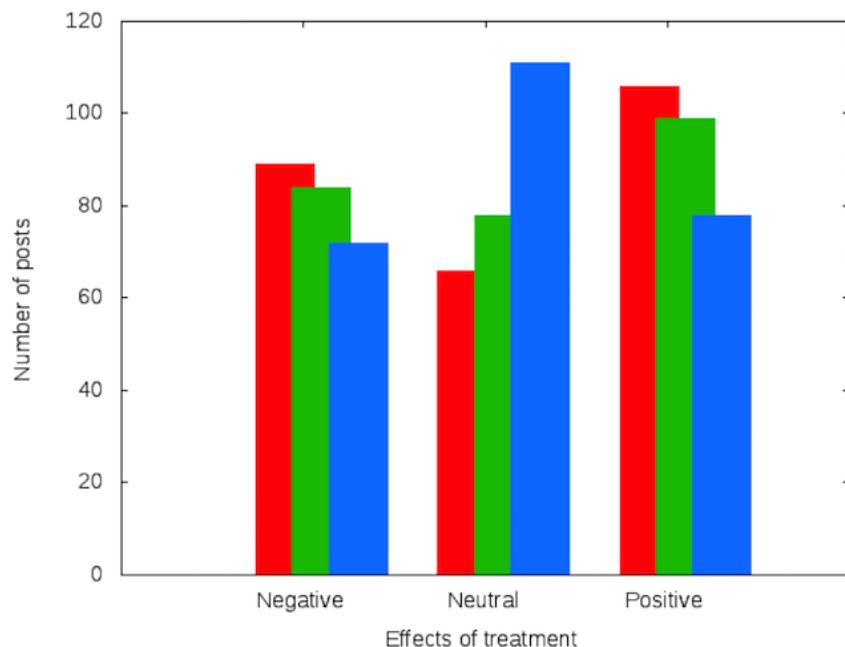
Results

Topics, Subtopics, and Sentiment Analysis

In 2014, 71% and 23% of adults on the Web used Facebook and Twitter, respectively [40]. Because of this fact, our attention focused on the posts that could be found on these two social networks. In fact, such two social networks have the potential of providing spontaneous and uncontrolled patients' opinions differently from thematic and moderated Web-based platforms specifically designed for patients.

Table 1. Subtopic dictionaries.

Subtopics	Dictionary
Lifestyle	Alcohol, bacteria, butter, cake, cell, chocolate, coffee, drink, eggs, food, gene, honey, lactose, map, meat, milk, pasta, smoke, sugar, tnf, virus, vitamin, and wine.
Symptoms and body parts	Abdomen, abscess, agony, anal, anxiety, appetite, arthritis, attack, belly, bladder, bleed, blood, bone, bowel, butt, colitis, constipation, cramp, damage, deficiency, depression, diabetes, diarrhea, digestion, disorder, exhausted, fever, fistula, flare, flu, gastro, grow, hurt, infection, inflamed, intestine, liver, mouth, muscle, nausea, pain, psoriasis, rectum, scar, severe, sleep, stress, suffer, symptom, tired, toilet, ulcer, and vomit.
Treatments	Adalimumab, aloe, antibiotic, asacol, azathioprine, budesonide, calcium, cannabis, capsule, certolizumab, cimzia, colectomy, colonoscopy, colostomy, diagnosis, diet, doctor, dose, drain, entocort, enzyme, fda, ferment, ginger, gp, health care, hospital, humira, ileostomy, imuran, infliximab, infusion, injection, kefir, marijuana, medication, medicine, mercaptopurine, methotrexate, morphine, mri, natural, nutrition, operation, oral, organic, paleo, pentasa, powder, prednisolone, prescribed, prescription, probiotic, rafton, remedy, remicade, resection, reversal, scd, solution, specialist, steroid, surgeon, surgery, test, therapy, transplant, treat, and visit.
Side-effects	Complications, effect, lupus, reaction allergy, and skin.

Figure 2. Computer scientist (red bars), senior gastroenterologist (green bars), and software classification (blue bars).**Table 2.** Square weighted Cohen's kappa coefficient (w2K) for the interrater agreement. The number of patients corresponding to the attributed score ($-1, 0, 1$) is indicated for each different observer (senior gastroenterologist vs computer science expert). The interobserver agreement was substantial: 87.36% (w2K: 0.6470).

		Computer science expert score			Total
		-1	0	1	
Senior gastroenterologist score	-1	62	17	5	84
	0	22	40	16	78
1		11	13	75	99
Total		95	70	96	261

Both expert reviews point to the same conclusions, as confirmed by interrater agreement statistical analysis (data reported in **Table 2**). The interrater agreement was performed using a square weighted Cohen's kappa coefficient (w2K). A substantial agreement (w2K=0.6470, corresponding to 87.36%) was found comparing the computer scientist versus the senior gastroenterologist evaluation of patients' global sentiment. This result indicates that the evaluation of the feeling that was communicated by a post was independent of the scientific background of the reader, although the senior gastroenterologist tended to classify as neutral a slightly larger share of posts, as not deemed relevant from a clinical point of view.

The classification performed by our software system, instead, provides a different outcome than those given by the computer science expert and by the senior gastroenterologist. In fact, the number of posts classified as neutral increase, as the sentiment analysis algorithm was evidently unable to determine with a precision similar to a human being the underlying meaning of a piece of text. Nonetheless, the proportion between positive and negative posts remains comparable, showing that the algorithmic tool could be useful to determine the existence of situations where positive and negative remarks concerning Infliximab were made.

Discussion

The availability of big data from social networks may be seen as an important source of information in medical research, alternative to the traditional sources of information [43,44]. Obviously, there are limitations, as patient characteristics (eg, age and sex) are often unknown.

We used social networks to analyze the perception of therapies by Crohn's disease patients. Crohn's disease has been chosen because of its well-defined features of chronic and sometimes disabling disease, with a strong impact on the quality of life of patients. Additionally, Crohn's disease is typically diagnosed in young patients (in the age range of 15-30 years), an age group of frequent social network users.

This work expands our previous studies, to propose a method to analyze the information posted on the Web. An important point of this work is that we use data derived from external observation of patients' spontaneous opinions during their daily lives. From this perspective, this study is a meticulous observation of the big data that a social network like Facebook may supply.

Our previous analyses revealed that Facebook (RQ1), with respect to Twitter, is the social network in which it is easier to find Crohn's disease information [18]. Our further studies individuated Infliximab as the most debated drug (RQ2), with both positive and negative sentiments among Crohn's disease patients [19]. This result was justifiable considering that Infliximab has been the first biological treatment (ie, monoclonal antibody) capable of strongly improving Crohn's disease management, with a rapid diffusion in the clinical setting. In addition, social networks usage started a few years after the 1998 approval of the Infliximab therapy for Crohn's disease patients, and this chronological coincidence possibly boosted the discussion on sites such as Facebook. Notably, a good match was found between the sentiment assessments in relation to

Infliximab obtained, with the ethnographic analyses performed by either computer science or gastroenterology experts (RQ3). This indicates that a data mining approach provided material of simple interpretation, regardless of the analysts' scientific and professional background. This represents a good starting point to provide a completely automated approach for the analysis of such data, in substitution of the final ethnographic step performed in this work. Another important finding is that our ethnographic results are in substantial agreement with the medical literature. In fact, medical trials involving large numbers of patients (large-scale retrospective trials) exhibit a percentage of those who experienced a negative reaction to Infliximab falling between 20-40% [45,46].

Acknowledgments

This research has been conducted using data available from public pages on Facebook. No sensitive medical data has been utilized, requiring any permission. The funding has been provided by the Alma Mater University of Bologna. Data, under the form of analyzed Facebook posts, are available upon request emailing to the authors.

Authors' Contributions

The authors declare an equal contribution to this manuscript in all of its phases. These include the design of the research activities, experimental studies, and the writing of the manuscript. In particular, MR, GM, PS, and CP have collaborated on ideating, designing, and developing the computer application that has been utilized to gather and classify data. Additionally, RMZ, FLGK, FB, and MM have contributed to the analysis of the supplied data from a medical and statistical viewpoint.

Conflicts of Interest

None declared.

References

1. Fox S, Duggan M. Pew Research Center. 2013. URL: http://www.pewinternet.org/files/old-media/Files/Reports/PIP_HealthOnline.pdf [accessed 2017-07-18] [WebCite Cache ID 6s3KK1rUX]
2. Scanfeld D, Scanfeld V, Larson E. Dissemination of health information through social networks: twitter and antibiotics. *Am J Infect Control*. Apr 2010;38(3):182-188. [doi: [10.1016/j.ajic.2009.11.004](https://doi.org/10.1016/j.ajic.2009.11.004)]
3. Cline R, Haynes KM. Consumer health information seeking on the Internet: the state of the art. *Health Educ Res*. Dec 2001;16(6):671-692. [Medline: [11780707](https://pubmed.ncbi.nlm.nih.gov/11780707/)]
4. Stellefson M, Chaney B, Barry AE, Chavarria E, Tennant B, Walsh-Childers K, et al. Web 2.0 chronic disease self-management for older adults: a systematic review. *J Med Internet Res*. Feb 2013;15(2):e35. [FREE Full text] [doi: [10.2196/jmir.2439](https://doi.org/10.2196/jmir.2439)] [Medline: [23410671](https://pubmed.ncbi.nlm.nih.gov/23410671/)]
5. Blog.google. URL: <https://blog.google/products/search/im-feeling-yucky-searching-for-symptoms/> [accessed 2017-04-18] [WebCite Cache ID 6poMo7QTV]
6. White RW, Tatonetti NP, Shah NH, Altman RB, Horvitz E. Web-scale pharmacovigilance: listening to signals from the crowd. *J Am Med Inform Assoc*. May 01, 2013;20(3):404-408. [FREE Full text] [doi: [10.1136/amiajnl-2012-001482](https://doi.org/10.1136/amiajnl-2012-001482)] [Medline: [23467469](https://pubmed.ncbi.nlm.nih.gov/23467469/)]
7. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. *J Med Internet Res*. 2009;11(1):e11. [FREE Full text] [doi: [10.2196/jmir.1157](https://doi.org/10.2196/jmir.1157)] [Medline: [19329408](https://pubmed.ncbi.nlm.nih.gov/19329408/)]
8. Eysenbach G. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. *AMIA Annu Symp Proc*. 2006:244-248. [FREE Full text] [Medline: [17238340](https://pubmed.ncbi.nlm.nih.gov/17238340/)]
9. Chew C, Eysenbach G. Pandemics in the age of twitter: content analysis of tweets during the 2009 H1N1 outbreak. *PLoS One*. Nov 29, 2010;5(11):e14118. [FREE Full text] [doi: [10.1371/journal.pone.0014118](https://doi.org/10.1371/journal.pone.0014118)] [Medline: [21124761](https://pubmed.ncbi.nlm.nih.gov/21124761/)]
10. Gittelman S, Lange V, Gotway CC, Okoro CA, Lieb E, Dhingra SS, et al. A new source of data for public health surveillance: facebook likes. *J Med Internet Res*. 2015;17(4):e98. [FREE Full text] [doi: [10.2196/jmir.3970](https://doi.org/10.2196/jmir.3970)] [Medline: [25895907](https://pubmed.ncbi.nlm.nih.gov/25895907/)]
11. Matsuda S, Aoki K, Tomizawa S, Sone M, Tanaka R, Kuriki H, et al. Analysis of patient narratives in disease blogs on the internet: an exploratory study of social pharmacovigilance. *JMIR Public Health Surveill*. Feb 24, 2017;3(1):e10. [FREE Full text] [doi: [10.2196/publichealth.6872](https://doi.org/10.2196/publichealth.6872)] [Medline: [28235749](https://pubmed.ncbi.nlm.nih.gov/28235749/)]

12. Nguyen QC, Li D, Meng H, Kath S, Nsoesie E, Li F, et al. Building a national neighborhood dataset from geotagged twitter data for indicators of happiness, diet, and physical activity. *JMIR Public Health Surveill*. Oct 17, 2016;2(2):e158. [FREE Full text] [doi: [10.2196/publichealth.5869](https://doi.org/10.2196/publichealth.5869)] [Medline: [27751984](https://pubmed.ncbi.nlm.nih.gov/27751984/)]
13. Ling R, Lee J. Disease monitoring and health campaign evaluation using google search activities for HIV and AIDS, stroke, colorectal cancer, and marijuana use in Canada: a retrospective observational study. *JMIR Public Health Surveill*. Oct 12, 2016;2(2):e156. [FREE Full text] [doi: [10.2196/publichealth.6504](https://doi.org/10.2196/publichealth.6504)] [Medline: [27733330](https://pubmed.ncbi.nlm.nih.gov/27733330/)]
14. Delir HP, Kang Y, Buchbinder R, Burstein F, Whittle S. Investigating subjective experience and the influence of weather among individuals with fibromyalgia: a content analysis of twitter. *JMIR Public Health Surveill*. Jan 19, 2017;3(1):e4. [FREE Full text] [doi: [10.2196/publichealth.6344](https://doi.org/10.2196/publichealth.6344)] [Medline: [28104577](https://pubmed.ncbi.nlm.nih.gov/28104577/)]
15. Koschack J, Weibezahl L, Friede T, Himmel W, Makedonski P, Grabowski J. Scientific versus experiential evidence: discourse analysis of the chronic cerebrospinal venous insufficiency debate in a multiple sclerosis forum. *J Med Internet Res*. Jul 01, 2015;17(7):e159. [FREE Full text] [doi: [10.2196/jmir.4103](https://doi.org/10.2196/jmir.4103)] [Medline: [26133525](https://pubmed.ncbi.nlm.nih.gov/26133525/)]
16. Hamad EO, Savundranayagam MY, Holmes JD, Kinsella EA, Johnson AM. Toward a mixed-methods research approach to content analysis in the digital age: the combined content-analysis model and its applications to health care twitter feeds. *J Med Internet Res*. Mar 08, 2016;18(3):e60. [FREE Full text] [doi: [10.2196/jmir.5391](https://doi.org/10.2196/jmir.5391)] [Medline: [26957477](https://pubmed.ncbi.nlm.nih.gov/26957477/)]
17. Thatcher A, Wretschko G, Fridjhon. Online flow experiences, problematic Internet use and internet procrastination. *Comput Human Behav*. Sep 2008;24(5):2236-2254. [doi: [10.1016/j.chb.2007.10.008](https://doi.org/10.1016/j.chb.2007.10.008)]
18. Roccetti M, Casari A, Marfia G, August). Inside chronic autoimmune disease communities: a social networks perspective to Crohn's patient behavior and medical information. 2015. Presented at: Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015; August 25-28, 2015:1089-1096; Paris, France. [doi: [10.1145/2808797.2808813](https://doi.org/10.1145/2808797.2808813)]
19. Roccetti M, Prandi C, Salomoni P, Marfia G. Unleashing the true potential of social networks: confirming infliximab medical trials through facebook posts. *Netw Model Anal Health Inform Bioinforma*. 2016;5:15. [doi: [10.1007/s13721-016-0122-9](https://doi.org/10.1007/s13721-016-0122-9)]
20. Bramer M. Principles of Data Mining. London. Springer; 2016.
21. Fayyad U, Piatetsky-Shapiro G, Smyth P. Knowledge discovery and data mining: towards a unifying framework. In: KDD'96 Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. 1996. Presented at: Second International Conference on Knowledge Discovery and Data Mining; August 02-04, 1996:82-88; Portland, Oregon.
22. Lee K, Caverlee J, Webb S. Uncovering social spammers: social honeypots + machine learning. In: SIGIR '10 Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval. ACM; 2010. Presented at: The 33rd international ACM SIGIR conference on Research development in information retrieval; July 19-23, 2010:435-442; Geneve, Switzerland. [doi: [10.1145/1835449.1835522](https://doi.org/10.1145/1835449.1835522)]
23. Stringhini G, Kruegel C, Vigna G. Detecting spammers on social networks. In: ACSAC '10 Proceedings of the 26th Annual Computer Security Applications Conference. 2010. Presented at: The 26th annual computer security applications conference; December 06-10, 2010:1-9; Austin, TX. [doi: [10.1145/1920261.1920263](https://doi.org/10.1145/1920261.1920263)]
24. Zhou X, Menche J, Barabási A, Sharma A. Human symptoms-disease network. *Nat Commun*. Jun 26, 2014;5:4212. [doi: [10.1038/ncomms5212](https://doi.org/10.1038/ncomms5212)] [Medline: [24967666](https://pubmed.ncbi.nlm.nih.gov/24967666/)]
25. Koh H, Tan G. Data mining applications in healthcare. *J Healthc Inf Manag*. 2005;19(2):64-72. [Medline: [15869215](https://pubmed.ncbi.nlm.nih.gov/15869215/)]
26. Milley A. Healthcare and data mining. *Health Management Technology*. 2000;21(8):44-47.
27. Blei D. Probabilistic topic models. *Commun ACM*. 2012;55(4):77. [doi: [10.1145/2133806.2133826](https://doi.org/10.1145/2133806.2133826)]
28. Aggarwal C, Zhai C. Mining text data. In: Springer Science & Business Media. Berlin/Heidelberg, Germany. Springer Science & Business Media; 2012.
29. Cambria E. Affective computing and sentiment analysis. *IEEE Intell Syst*. Mar 2016;31(2):102-107. [doi: [10.1109/MIS.2016.31](https://doi.org/10.1109/MIS.2016.31)]
30. Smailovic J, Grcar M, Lavrac N, Žnidaršič M. Predictive sentiment analysis of tweets: a stock market application. In: Holzinger A, Pasi G, editors. Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data. Berlin, Heidelberg. Springer; 2013.
31. Duh MS, Cremieux P, Audenrode MV, Vekeman F, Karner P, Zhang H, et al. Can social media data lead to earlier detection of drug-related adverse events? *Pharmacoepidemiol Drug Saf*. Dec 2016;25(12):1425-1433. [FREE Full text] [doi: [10.1002/pds.4090](https://doi.org/10.1002/pds.4090)] [Medline: [27601271](https://pubmed.ncbi.nlm.nih.gov/27601271/)]
32. Fang X, Zhan J. Sentiment analysis using product review data. *J Big Data*. Jun 16, 2015;2(1):1. [doi: [10.1186/s40537-015-0015-2](https://doi.org/10.1186/s40537-015-0015-2)]
33. Medhat W, Hassan A, Korashy H. Sentiment analysis algorithms and applications: a survey. *Ain Shams Eng J*. Dec 2014;5(4):1093-1113. [doi: [10.1016/j.asej.2014.04.011](https://doi.org/10.1016/j.asej.2014.04.011)]
34. Wilson T, Hoffmann P, Somasundaran S, Kessler J, Wiebe J, Choi Y, et al. OpinionFinder: a system for subjectivity analysis. In: HLT-Demo '05 Proceedings of HLT/EMNLP on Interactive Demonstrations. 2005. Presented at: HLT/EMNLP on Interactive Demonstrations; October 07, 2005:34-35; Vancouver, Canada. [doi: [10.3115/1225733.1225751](https://doi.org/10.3115/1225733.1225751)]
35. Baldridge J. Apache. 2017. URL: <http://opennlp.apache.org/index.html> [accessed 2017-04-18] [WebCite Cache ID [6poN28Vpm](https://www.webcitation.org/6poN28Vpm)]

36. Manning C, Surdeanu M, Bauer J, Finkel J, Bethard S, McClosky D. The Stanford CoreNLP natural language processing toolkit. 2014. Presented at: Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations; June 23-24, 2014:55-60; Baltimore, MD.
37. Bollen J, Mao H, Zeng X. Twitter mood predicts the stock market. *Journal of Computational Science*. Mar 2011;2(1):1-8. [doi: [10.1016/j.jocs.2010.12.007](https://doi.org/10.1016/j.jocs.2010.12.007)]
38. Hosmer J, Lemeshow S, Sturdivant RX. Applied logistic regression. Hoboken, NJ. John Wiley & Sons; 2013.
39. Allen I, Seaman C. ASQ. 2007. URL: <http://asq.org/quality-progress/2007/07/statistics/likert-scales-and-data-analyses.html?s=qp> [accessed 2017-07-18] [WebCite Cache ID 6s3b9uZhh]
40. Duggan M, Ellison N, Lampe C, Lenhart A, Madden M. Pew Research Center. 2015. URL: http://www.pewinternet.org/files/2015/01/PI_SocialMediaUpdate20144.pdf [accessed 2017-07-18] [WebCite Cache ID 6s3bJPKYZ]
41. Blomberg J, Burrell M, Guest G. An ethnographic approach to design. In: The human-computer interaction handbook. New Jersey. Lawrence Erlbaum Associates, Inc; 2002:964-986.
42. Roccetti M, Marfia G, Salomoni P, Prandi C, Zagari R, Gningaye F, et al. Unibo. 2017. URL: <http://www.cs.unibo.it/~marfia/Lista-Posts.docx> [accessed 2017-05-16] [WebCite Cache ID 6qVFMOfZJ]
43. Eysenbach G, Wyatt J. Using the internet for surveys and health research. *J Med Internet Res*. Nov 2002;4(2):E13. [FREE Full text] [doi: [10.2196/jmir.4.2.e13](https://doi.org/10.2196/jmir.4.2.e13)] [Medline: [12554560](https://pubmed.ncbi.nlm.nih.gov/12554560/)]
44. Eysenbach G. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. *J Med Internet Res*. 2008;10(3):e22. [FREE Full text] [doi: [10.2196/jmir.1030](https://doi.org/10.2196/jmir.1030)] [Medline: [18725354](https://pubmed.ncbi.nlm.nih.gov/18725354/)]
45. Colombel J, Loftus E, Tremaine W, Egan L, Harmsen WS, Schleck C, et al. The safety profile of infliximab in patients with Crohn's disease: the Mayo clinic experience in 500 patients. *Gastroenterology*. Jan 2004;126(1):19-31. [Medline: [14699483](https://pubmed.ncbi.nlm.nih.gov/14699483/)]
46. Caspersen S, Elkjaer M, Riis L, Pedersen N, Mortensen C, Jess T, et al. Danish Crohn Colitis Database. Infliximab for inflammatory bowel disease in Denmark 1999-2005: clinical outcome and follow-up evaluation of malignancy and mortality. *Clin Gastroenterol Hepatol*. Nov 2008;6(11):1212-7; quiz 1176. [doi: [10.1016/j.cgh.2008.05.010](https://doi.org/10.1016/j.cgh.2008.05.010)] [Medline: [18848503](https://pubmed.ncbi.nlm.nih.gov/18848503/)]

Edited by G Eysenbach; submitted 16.Nov.2016; peer-reviewed by R Altman, C Arnold, D Frohlich; comments to author 03.Feb.2017; revised version received 18.May.2017; accepted 13.Jun.2017; published 09.Aug.2017

Please cite as:

Roccetti M, Marfia G, Salomoni P, Prandi C, Zagari RM, Gningaye Kengni FL, Bazzoli F, Montagnani M. Attitudes of Crohn's Disease Patients: Infodemiology Case Study and Sentiment Analysis of Facebook and Twitter Posts. *JMIR Public Health Surveill* 2017;3(3):e51
URL: <http://publichealth.jmir.org/2017/3/e51/>
doi: [10.2196/publichealth.7004](https://doi.org/10.2196/publichealth.7004)
PMID: [28793981](https://pubmed.ncbi.nlm.nih.gov/28793981/)

©Marco Roccetti, Gustavo Marfia, Paola Salomoni, Catia Prandi, Rocco Maurizio Zagari, Faustine Linda Gningaye Kengni, Franco Bazzoli, Marco Montagnani. Originally published in JMIR Public Health and Surveillance (<http://publichealth.jmir.org>), 09.Aug.2017. 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 Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on <http://publichealth.jmir.org>, as well as this copyright and license information must be included.