RESEARCH ARTICLE



Analyzing user-generated content using natural language processing: a case study of public satisfaction with healthcare systems

Anna Ruelens¹

Received: 6 July 2020 / Accepted: 6 October 2021 / Published online: 29 October 2021 © The Author(s), under exclusive licence to Springer Nature Singapore Pte Ltd. 2021

Abstract

While user-generated online content (UGC) is increasingly available, public opinion studies are yet to fully exploit the abundance and richness of online data. This study contributes to the practical knowledge of user-generated online content and machine learning techniques that can be used for the analysis of UGC. For this purpose, we explore the potential of user-generated content and present an application of natural language pre-processing, text mining and sentiment analysis to the question of public satisfaction with healthcare systems. Concretely, we analyze 634 online comments reflecting attitudes towards healthcare services in different countries. Our analysis identifies the frequency of topics related to healthcare services in textual content of the comments and attempts to classify and rank national healthcare systems based on the respondents' sentiment scores. In this paper, we describe our approach, summarize our main findings, and compare them with the results from cross-national surveys. Finally, we outline the typical limitations inherent in the analysis of user-generated online content and suggest avenues for future research.

Keywords User-generated online content \cdot Sentiment analysis \cdot Text mining \cdot Healthcare systems \cdot Public attitudes \cdot Computational social science

Introduction

In recent years, user-generated online content (UGC)—including, but not limited to, social media—has accumulated large amount of data on individual attitudes, behaviors, and experiences. While the actual contribution of these data to the study of public opinion is still under discussion, the potential insights derived from the UGC are expected to be significant [4]. In addition, the costs associated with collecting

Research Institute for Work and Society, University of Leuven, Parkstraat 47, Box 5300, 3000 Leuven, Belgium



Anna Ruelens
anna.ruelens@kuleuven.be

and processing of such information using automated technologies are generally low, especially compared to the more conventional means of gathering data on public opinion.

While user-generated content is increasingly available, its practical use in addressing relevant research questions and the knowledge of analytical techniques to examine such data among social science researchers remain limited. As a consequence, many current public opinion studies do not take a full advantage of the available information. Hence, the main purpose of this study is to explore the possibilities of UGC for the study of public opinion. We do so, by analyzing 634 online reader comments reflecting attitudes towards healthcare systems across different countries. Until now, comparative studies on public satisfaction with healthcare services have mostly utilized cross-national survey data (e.g. [23, 28, 43]. To the best of our knowledge, this is the first study that assesses public attitudes towards healthcare system in comparative perspective using unsolicited user-generated online data. In addition to providing an example of how UGC can be used for the analysis of public attitudes, we demonstrate the practical implementation of analytical tools to explore the user-generated content.

The paper begins by providing a brief discussion of user-generated online content. Consequently, we introduce an example of how UGC can be used to study public attitudes towards healthcare systems. In the following section, we outline the proposed methodology and the analytical procedure. The paper concludes with a discussion of the results and limitations of the present study, and suggestions for future research.

User-generated online data

User-generated online content is considered to be one particular type of big data [22]. As such, UGC represents a new form of data that has not previously been available for public opinion research. User-generated content is defined as information in the form of text, media or metadata that are posted by users online, often on social networking sites (SNSs) [26]. User posts on Facebook or tweets generated by users are typical examples of UGC. Although until now most social research using UGC has focused on analyzing Twitter and Facebook data (e.g. [1, 32, 41, 42]), user generate content can be found across different types of outlets, including posts on online forums, customer reviews, newspaper comments, and interaction on social media.

One particular source of UGC and the one used in the present study are online readers' comments. Such comments often appear following the (news) articles published in online versions of newspapers and are often encouraged by the newspapers to foster reader engagement, online deliberation, and "citizen journalism" [27]. While representing a relatively new phenomenon, online reader comments have already been heralded as "new opinion pipeline" [36]. Indeed, the amount of information shared by readers in their comments makes them an invaluable source of insights into public opinion on a wide variety of topics.

The distinguishing feature of UGC, in general, and online user comments, in particular, is that these data are not solicited by researchers [34]. In contrast to the



conventional means of obtaining information on public opinion such as surveys or focus groups, which are specifically designed for research purposes, the unsolicited online data are considered to be "naturally occurring" [14], "organic" [16], or "found" data [18]. The use of UGC for research purposes presents both opportunities and challenges for public opinion research. Many of these, in particular in comparison with survey methods, have been well documented in the recent literature so we will limit ourselves to a brief discussion.

Regarding the main challenges associated with the use of UGC, one of the oftencited issues is the limited generalizability of the findings. Considering that such data are not sampled in the conventional way, they may not be representative of the entire population. Connected with the issue of generalizability is the issue of selection bias, given that the users who post online are often self-selected. Moreover, not every topic is equally likely to be discussed by the online community. As aptly captured by Couper [9]: "if we are to make use of the vast amount of public information on the Internet, we need more work to understand how those who willingly share information with the broader public differ from those who do not, and what kinds of topics are more or less susceptible to selection biases" (p. 903). At the same time, it is important to remember that the generalization to the entire population is not a requirement for every study. For some research purposes, having the data on only one specific group can be of value, especially in the context of studying opinions of sub-groups or "hard-to-reach" populations (see, for example, [10].

While traditional surveys can boast extensive metadata on respondents, user-generated data often lack information on basic socio-demographic characteristics of users. Nevertheless, while lacking some socio-demographic information, user generate data often contain a lot of other relevant metadata that can be used in the analysis, including geographical location, frequency of posting, among others. Another challenge associated with UGC is that such data are often complex, messy, and unstructured. Therefore, the quality of UGC cannot always be guaranteed and it depends on the amount of noise in the data. This latter characteristic makes user-generated data similar to the information obtained by qualitative research methods. At the same time, as with the qualitative data, UGC often provides richer information on specific user attitudes than do survey responses. Moreover, researchers analyzing UGC can often derive additional information on the emotional state of the individual, while such information is generally unavailable in survey data.

It is also important to note that, although it is not always a requirement, to exploit UGC in the manner that is both accurate and reliable, public opinion researchers may need to develop certain specialized computational skills when working with such data. Recently more traditional machine learning (ML) techniques applied to natural language processing (NLP), such as keyword extraction, topic recognition, and sentiment analysis have been enhanced by the application of deep learning methods. For example, in their recent paper Souma and colleagues [39] apply recurrent neural network (RNN) with long short-term memory (LSTM) units to forecast financial news sentiments. In another study, Naeem and colleagues [30] use a novel

¹ For a comprehensive comparison between social media and surveys, see Schober et al. [37].



deep learning framework for clickbait detection on social area network. The application of deep learning methods is more technically demanding and may present an additional barrier to the wide use of UGC in public opinion research among social scientists.

Another relevant aspect that should be considered when using UGC for research purposes is that the use of such data may have certain legal and/or ethical barriers. In contrast to the clear procedures for ensuring consent of survey respondents, individuals who leave their 'digital traces' online often do not give their explicit consent for their data to be used for research purposes. When such data are in public domain this may present less of an issue for the researcher interested in analyzing such information. Nevertheless, considerable privacy concerns, especially when users can be identified, need to be acknowledged and addressed.

In what follows, we provide an illustration of how online user comments can be used to gain insights into public opinion about different healthcare systems. This example is only one possible path among many lines of investigation available to the researcher interested in analyzing UGC.

Illustration: examining public satisfaction with healthcare systems

Background

As its fundamental institution and the largest consumer of its resources, health-care is central to the welfare state [43]. The last several decades have been characterized by increasing pressure to transform healthcare systems with a number of reforms implemented in many European countries, as well as in the United States. In the European context, the reforms have been almost continuous [5]. These policy changes and their effects on the individual citizens have been reflected in public evaluation of healthcare system performance.

In addition to representing a key indicator of healthcare quality, public attitudes on healthcare services act as a reflection of popular legitimacy of healthcare systems [28] and of welfare institutions, in general [43]. Furthermore, recent studies found that high levels of public satisfaction with the performance of a healthcare system are strongly associated with trust in government [6]. This makes the study of public attitudes towards healthcare systems an important line of inquiry both for public opinion researchers and policy experts.

Previous research has highlighted the important role of the individual and institutional characteristics for the satisfaction with healthcare systems [25, 28, 43]. At the individual level, the determinants of user satisfaction with healthcare services include the usual socio-demographic characteristics, such as age, gender, marital status, and educational level, as well as health status and income level [28, 33]. A number of studies have also emphasized the importance of ideology for healthcare satisfaction [12, 13]. In addition, personal experience with the healthcare services has been shown to affect the level of user satisfaction [2], but cf. [3].

At the institutional level, healthcare regimes and their characteristics, including type and level of financing (public, private, mixed), as well as the key quality



indicators of healthcare provision (such as the ratio of doctors and nurses to patients, the number of hospital facilities, the number of hospital beds, among others) are often cited as important factors for determining the level of support for the healthcare system [43]. In general, scholars consider healthcare satisfaction to be a multidimensional attitude, with most factors related to its evaluation to belong to one of the three domains: access, quality, and affordability.

Traditionally, most of the comparative data on public satisfaction with healthcare systems came from the large cross-national surveys. In the European context, such surveys include the Eurobarometer, the European Social Survey (ESS), and the European Quality of Life Survey (EQLS). However, implementing such surveys is costly and, similar to other international surveys, they are increasingly characterized by high non-response rate, which impacts the validity of the conclusions. In addition, the attitudes towards healthcare systems across different world regions are usually not reflected within the same survey, making it difficult to compare healthcare systems. While user-generated content containing information on healthcare evaluation is increasingly available, these online resources are yet to be utilized to their full potential. The COVID-19 pandemic has served as an impetus for a number of studies using online data. For example, Havey [19] analyzes six misinformation topics related to the COVID-19 pandemic using sentiment analysis of Twitter data. In another study, Shahsavari and colleagues [38] examine online forum discussions and news reports to detect the emergence of conspiracy theories related to the pandemic. Yet in another paper, Uyheng and Carley [41] analyze online conversation and hate speech around the COVID-19 crisis in the United States and the Philippines. These studies demonstrate the richness of insights that can be derived from the analysis of online data. In the present study, we use a particular type of UGC, namely newspaper readers' comments, to answer the following research questions about general attitudes towards healthcare systems:

- 1. What healthcare related topics do readers perceive as important when comparing healthcare systems?
- 2. Are these topics reflective of one of the three dimensions of healthcare system evaluation, namely accessibility, affordability, and quality?
- 3. Can we identify in the body of comments certain specific individual or institutional factors that are deemed important for the evaluation of healthcare systems?
- 4. Is it possible to generate a ranking of healthcare systems based on the user comments?
- 5. How do results of the analysis of readers' comparative evaluation of healthcare systems correspond to the results obtained from the cross-national surveys?

Data and method

Data

In this study, we analyze a unique dataset of user-generated online comments from *The New York Times* (https://www.nytimes.com). The comments represent responses



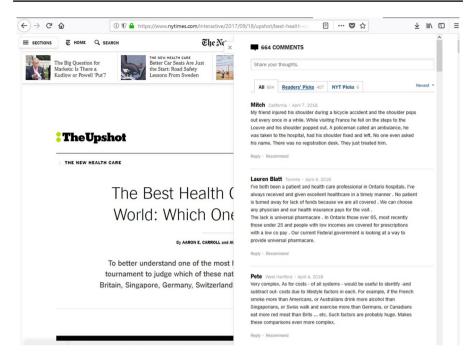


Fig. 1 Screenshot of readers' comments on nytimes.com. Retrieved from https://www.nytimes.com/interactive/2017/09/18/upshot/best-health-care-system-country-bracket.html

of the readers to the original article by Carroll² and Frakt³ titled "The best health care system in the world: which one would you pick?" that was written as a contribution on the ongoing debate on healthcare policy in the United States (US) and published online on September 18, 2018. The article contains opinions of an expert panel consisting of five distinguished health researchers who discuss and compare healthcare systems across eight countries: Canada, Britain, Singapore, Germany, Switzerland, France, Australia, and the United States. The amount of online comments at the beginning of our analysis was 636, and after deleting duplicate comments, our completed dataset contained 634 unique reader comments (Fig. 1). The following section explains the method and analytical steps and procedures used to analyze these comments. The online corpus of comments was freely available on the New York Times website at https://www.nytimes.com/interactive/2017/09/18/upshot/best-health-care-system-country-bracket.html.

³ Austin Frakt is a director of the Partnered Evidence-Based Policy Resource Center at the V. A. Boston Healthcare System, associate professor with Boston University's School of Public Health; and adjunct associate professor with the Harvard T. H. Chan School of Public Health.



 $^{^2}$ Aaron Carol is a health services researcher and professor of pediatrics at Indiana University School of Medicine.

Method

Previous research of online readers' comments has often relied on manual analysis, which is a limited factor in the ability to analyze larger corpora of comments. In this study, we use automated natural language processing (NLP) tools that facilitate this type of textual analysis. The two main text mining methods applied in this study are word frequency analysis and sentiment analysis. Both methods are extensively used in the studies based on computational approach (e.g. [24, 39, 40]).

The first method, word frequency analysis, as its name suggests, involves identifying most frequently used terms in the body of user comments. This technique is based on the idea that words which are most frequently used by the commenter indicate issues of higher importance to the user [35]. This method is inductive in identifying the topics of relevance, compared, for example, to the predefined structure of the survey instrument. Given the focus of this study, we were particularly interested in the most common terms related to healthcare and medical services.

In the second part of our analysis, we employed text sentiment analysis also known as emotional polarity computation. The main aim of sentiment analysis is to determine the subjective opinions of online users with respect to a specific topic (for an overview of sentiment analysis see [31]. These opinions can be of an evaluative or of a judgmental nature. Moreover, they can also reflect the emotional state of the user, revealed either intentionally or unintentionally. For the purposes of this study, higher positive sentiment score reflects a more positive attitude towards health-care systems. Although sentiment analysis is a mainstream tool in text mining, this method continues to evolve and achieves greater accuracy and validity.

Analytical strategy and tools

In our analytical approach to examining textual content we build on the typical steps employed in text analytics and we develop these steps further to make them appropriate for our analyses. Concretely, in Step 1 (extraction phase) we extracted the readers' comments by adapting a publicly available script written by Caren [7]. In this step, we used web-scraping techniques to retrieve the comments and to convert them into suitable format for further analysis.

In Step 2, we conducted data cleaning using natural language pre-processing techniques. The goal of this step was to transform the raw data into a usable format for textual analysis [20]. Although frequently regarded as time-consuming and tedious, this stage of data pre-processing is essential to ensure quality results.⁴ At this stage, we removed spaces and special symbols from the text corpus to be analyzed.

In Step 3, we used the Natural Language Toolkit (NLTK) toolbox implemented in Python to conduct Named Entity Recognition (NER) to extract the names of the countries mentioned in the text corpora. One of the main challenges of unstructured text data is that the input is not standardized. In this particular instance, the location

⁴ A common notion in a computing: 'garbage in garbage out' highlights the importance of the quality of input data.



names included not only the names of countries but also the names of cities, US states, Canadian provinces, and various municipalities. Some locations were also abbreviated in multiple different ways (e.g. US, US, USA, America), misspelled, or referred to by their unofficial name (e.g. "Holland" instead of "the Netherlands"). To ensure the consistency of country names we implemented the algorithm using a "Geocoder" module in Python.

In Step 4, using genderize.io API, we determined the gender of a commenter based on his/her first name. While on some online platforms identification of gender presents a challenge with many users providing only a username or commenting as "anonymous", the users of the NYT comments tended to provide their first name.

In Step 5, we conducted a word frequency count analysis using "tm" package in R [11] to identify most frequently used words associated with healthcare in the body of comments. Given that no separate online dictionary exists that would allow automatized allocation of healthcare terms to healthcare attitudes domains, we accomplished this part manually.

In the final step, Step 6, we conducted a sentiment analysis using Vader (Valence Aware Dictionary for sEntiment Reasoning) rule-based module from the NLTK toolbox [21]. Vader outputs scores for a positive, negative, and neutral sentiment. In addition, it provides a compound sentiment score that reflects the overall sentiment associated with the given piece of text. The compound score ranges from -1 (most negative) to 1 (most positive). First, we applied Vader to determine sentiment of the entire comment. Consequently, we parsed each comment into separate sentences and conducted sentiment analysis on the sentences containing country names. This technique allowed us to analyze the association between a particular country and a corresponding sentiment. When several countries were used in one given sentence the same compound sentiment score was assigned to them. Afterwards, we averaged a compound score for each country to generate a country ranking.

Results

Descriptive information

After removing the duplicates, we identified 634 unique comments, as of April 16th, 2018. After aggregating and standardizing the names and abbreviations, sixty-one unique geographical entities were identified at the country level. Regarding the gender composition, 167 comments (26.3%) were assigned to female commenters, 289 comments (45.6%) were assigned to male commenters and 178 comments (28.1%) were non-identified based on the username.

Word frequency analysis

Word frequency analysis was used to identify the most frequently occurring terms in the body of comments. Table 1 presents fifteen high-frequency words



Table 1 Top 15 words with high frequencies related to healthcare							
1. System(s) [650]	2. Care [621]	3. Health [521]	4. Cost(s) [384]	5. Insurance [326]			
6. Pay (id, ment) [295]	7. Doctor(s) [252]	8. Hospital(s) [223]	9. Healthcare [213]	10. Medical [194]			
11. People [156]	12. Private [151]	13 Access [111]	14. Coverage [102]	15. Free [98]			

Table 2 High-frequency words per healthcare domain

Domain	Words
Affordability	Cost(s) [384], insurance [326], pay(id, ment) [295], private [151], coverage [102], free [98], need [94], universal [78], public [73]
Accessibility	Access [111], time(s) [175], wait [42], hour(s) [32]
Quality	Doctor(s) [252], hospital(s) [223], quality [60], specialist(s) [55], physicians [51]

in the text corpus, related to healthcare. Given the topic of the newspaper article, it is not surprising that the top three words include "system(s)", "care", and "health".

The high-frequency terms were associated with the three healthcare domains of accessibility, affordability, and quality, outlined in the literature on patient satisfaction. Most of the identified terms belong to the dimension of affordability, followed by the dimension of quality. The words associated with access to healthcare services, while relevant were less frequently used by the commenters (Table 2).

A comparison of high-frequency terms used by female and male commenters indicates that most words are common to both genders, although female readers appear to focus more on the affordability of healthcare services frequently using the terms such as "free", "coverage", "taxes" and "need", while male commenters appear to be more concerned with the organizational and structural aspects of healthcare systems, frequently using such words as "government" and "private". In addition, the word frequency analysis demonstrates the relevance of personal experiences with healthcare systems for their evaluations. This is indicated by the frequency of such terms as "experience" [60], "experiences" [10], "experienced" [36], and "personal" [14].

Sentiment analysis

In the first part of sentiment analysis, we determined the emotional polarity at the level of a comment. Table 3 provides an example of two comments, one was assigned a positive sentiment score and another was assigned a negative sentiment score.

Based on the sentiment score of the comment, we compared the most frequently mentioned healthcare topics for the 10% most satisfied and 10% least satisfied commenters. The findings indicate that most frequently used terms related to healthcare are common to both groups. This is the case, for example, for the topics of "insurance", "doctor(s)", and "cost(s)". Some differences are worth noting, however. For



Table 3	Example of a positive
and a ne	gative comment

Sentiment	Comment
Positive	'Anne Jolly' I am an aged pensioner, living in Australia. I have completely free doctor and hospital visits, including x rays. Pathology etc. My dental care is also free, as is optical care. I have a heart condition requiring medication, one of which is [very] expensive but all my prescribed medications are covered by a subsidy which means I pay only \$6.30 and all prescriptions are free once I reach a yearly ceiling. People who are working have a medicare levy as part of their talevy. That covers basic health care, and then they are expected to have private health insurance for subsidiaries like dental, optical etc. Those who can pay something for their health care are expected to, but people like
Negative	me are very well looked after 'Chris' What isn't mentioned here is that Briton rations drugs and/or does not make them available at all through a system bearing the horrendous acronym NICE. Hospitals are hopelessly understaffed, badly managed an hotbeds for infectious disease (immune-suppressed patients sharing bathrooms—in one of the most modern facilities in London). Access is shambolic, waiting times are endless, especially in the big cities. Ther is NO choice of primary care doctors, you have to go where the state tells you to go. There is hardly any dental care. Nobody in Europe considers the UK to have health care on the level of an advanced industrialized

Capitalization, punctuation and syntax are preserved The spelling in the brackets is corrected by the authors

example, most positive commenters mention "access", "free", and "patients" while the least positive commenters frequently discuss "private", "wait times", "coverage", and "poor".



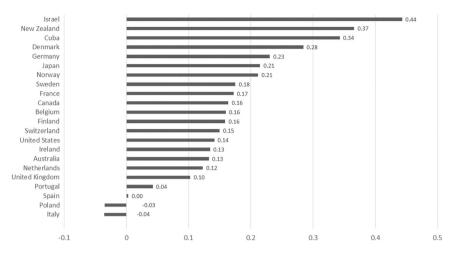


Fig. 2 Continuous sentiment score at the country level

Sentiment analysis: male vs. female commenters

Literature on healthcare attitudes based on survey data indicate that gender is an important factor in the evaluation of healthcare systems. Nevertheless, the results of the empirical studies on the impact of gender on healthcare satisfaction remain mixed [8, 19, 28]. To test whether any differences between female and male commenters can be observed in our set of comments, we calculated the mean compound sentiment score for the male and the female group separately. The results indicate that the mean sentiment score in the female group (0.397) is somewhat lower than that in the male group (0.433). However, the actual difference between the means of the two groups does not appear to be statistically significant in the analyzed body of comments.

Sentiment analysis: country level

In the following step, we identified the sentiment score for each specific country based on the sentiment analysis conducted at the level of a sentence. In the cases in which a particular country was mentioned more than once, we used the average compound sentiment score. The number of times that a specific country appeared in the body of comments differ greatly. For instance, the United States appeared 654 times, which is not surprising given that the comments address the article published in the U.S. newspaper. To provide other examples, France was mentioned 330 times, Germany 67 times, Mexico 40 times, and Japan 22 times. We performed accuracy check which resulted in the set of twenty-two countries with the assigned compound sentiment score. The results of the sentiment analysis indicate that when collapsed into sentiment categories, eighteen countries (82%) were assigned a positive sentiment (compound score \geq 0.05) while four countries (18%) were assigned a neutral sentiment (compound score between - 0.05 and



Fig. 3 Sentiment scores across European countries



0.05). A number of countries were originally assigned a negative sentiment but after averaging the result the compound score turned out to be neutral or positive.

Figure 2 presents the sentiment score for twenty-two countries. A continuous sentiment score shows that Israel, New Zealand, Cuba, and Denmark are associated with the most positive sentiment score, while Portugal Spain, Poland, and Italy with the least positive sentiment score. Additionally, Figure 2 presents an interesting comparison between public approval of national healthcare systems across different world regions. This information is usually unavailable from the survey data, considering that most surveys that gather information on this topic generally focus on one region, such as Europe or North America.

Focusing on the European region, we plotted a sentiment score per European country (Fig. 3). Figure 3 shows that Northern European countries such as Sweden, Norway, and Denmark together with Germany and France have higher positive sentiments compared to Southern European countries, such as Italy, Portugal, Spain. Poland, the only East European country in the set of comments, also has a low sentiment score associated with it.

Comparison of country rankings with survey results

In this section, we address the question of how our results compare to the findings from the commonly used cross-national surveys. The questions around satisfaction with healthcare systems are routinely included in major European surveys, such as



Table 4 Country mean scores of the EQLS, the ESS, and the sentiment score

Country	EQLS	ESS	Sentiment
Belgium	7.6	7.7	5.8
Denmark	7.4	_	6.4
Finland	7.6	7.5	5.8
France	7.4	6.8	5.9
Germany	7.3	6.6	6.2
Ireland	5.9	4.6	5.7
Italy	5.8	5.8	4.8
Netherlands	7.3	6.7	5.6
Norway	_	7.4	6.1
Poland	5.4	4.5	4.8
Portugal	6.3	5.7	5.2
Spain	7.2	5.9	5.0
Sweden	7.3	6.1	5.9
Switzerland	_	7.5	5.8
United Kingdom	6.9	6.2	5.5

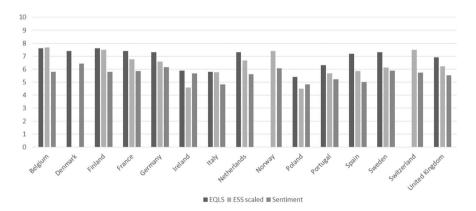


Fig. 4 Ranking of ESS, EQLS and the results of sentiment analysis per country

the European Quality of Life Survey (EQLS), the European Social Survey (ESS), and the Eurobarometer. We have chosen to compare the country-level results of the sentiment analysis to the results of the EQLS 2016 and the ESS 2018, as these two surveys use a similar wording of the question. Additionally, the measurement scale of the two surveys can be more easily compared.

In the EQLS, the question wording is: "In general, how would you rate the quality of each of the following public services in [country]? Please tell me on a scale of 1–10, where 1 means very poor quality and 10 means very high quality. (a) Health services." The ESS asks the question in the following way: "[still] using this card, please say what you think overall about the state of health services in [country] now-adays?, where "0" is "extremely bad" and "10" is "extremely good."



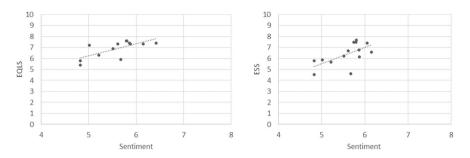


Fig. 5 Scatterplot of country scores from the EQLS and the ESS versus sentiment scores

After rescaling the ESS mean scores and the sentiment scores to match the scale of the EQLS (from 1 to 10), we obtained the results presented in Table 4 and visually depicted in Fig. 4. Although not identical, the ranking across the three sources appears to be similar. To ease the comparison, we correlated the mean scores of the ESS and the EQLS with the country sentiment scores. The results from the ESS show similar patterns to that of the EQLS and the correlation between the two country rankings is high (0.88). While lower than the correlation value between the two analyzed surveys, the correlation between sentiment scores and the EQLS mean scores is also strong and positive (0.70). The correlation between sentiment scores and the ESS mean scores is somewhat less and is considered to be moderate (0.63).

The scatterplots of the relationship between the mean scores from the EQLS, the ESS and the sentiment scores are presented in Fig. 5. As can be observed, the relationship between the EQLS mean scores and the scores from the sentiment analysis is linear and positive, confirming the results of the correlation analysis. The scatterplot of the ESS mean scores versus sentiment scores also indicates positive and linear relationship.

Discussion and conclusion

The present study was motivated by a desire to explore the usefulness of user-generated online content for public opinion research. While the amount of information shared by individuals online has increased dramatically, these data on personal experiences, opinions, and attitudes have been generally underutilized. Our second aim was to determine a level of correspondence between the results obtained from the analysis of UGC compared to those obtained by analyzing conventional survey data. Finally, the third purpose of this study was to provide a practical example of tools and procedures for analyzing UGC. To achieve these three goals, we focused specifically on public attitudes towards healthcare systems. Below we highlight the main findings of this study.

First, the results obtained from the word frequency analysis reflected all three dimensions of healthcare evaluation, namely affordability, accessibility, and quality of health services. The words with the highest frequencies included "cost(s)",



"insurance", "pay(id)", "doctor(s)" and "hospital(s)". This result indicates that the most commenters are greatly concerned with the affordability and quality domains, although the accessibility domain is also important. While it is not clear without further analysis what exactly the users mean by "access", the commenters mention frequently "wait times", indicating that this issue presents a significant barrier to the accessibility of healthcare services.

The analysis of word frequency for female and male commenters separately revealed that although most of the high-frequency words were identical, female users tend to focus more on the affordability of healthcare services, indicated by such words as "free", "coverage", "taxes", and "need", while male users tend to prioritize the institutional structure of healthcare systems using such words as "government" and "private".

A comparison of high-frequency words related to healthcare between 10% of commenters with the most positive sentiment score and 10% with the least positive sentiment score showed that although such factors as insurance, doctors, and costs are relevant to both groups, the most positive group frequently mentioned such aspects as "access", "free", and "patients", while the least positive group were concerned with "private", "wait times", "coverage", and "poor".

Taken together, the results of the word frequency analysis provide support for the importance of the institutional factors in determining public satisfaction with health-care systems. In particular, the manner in which healthcare provision is organized and funded (public or private), as well as certain aspects of quality (number of doctors, number of hospitals) and access (wait times) play an important role in influencing public opinion about healthcare services. These findings provide additional validity to the institutional indicators collected as part of the official healthcare services statistics by the agencies such as the World Health Organization (WHO) or the Eurostat.

Literature on public attitudes towards healthcare systems have shown inconsistent results in regard to the influence played by gender [8, 17, 28]. Based on the analyzed set of readers' comments, we found that female commenters tend to have a lower compound sentiment score than do the male commenters, indicating that women readers are less satisfied with healthcare services. However, it is important to note that the difference between the means of two groups was not found to be statistically significant. At the same time, in line with the literature on the individual-level determinants of healthcare attitudes the results of the word frequency analysis confirmed that personal experiences with the healthcare system play an important role for healthcare system evaluation [2].

Regarding other individual-level characteristics relevant to healthcare satisfaction, while we were able to include gender into our analysis, it was not possible to investigate the effect of age, given that the information on this variable was missing and could not be directly inferred from the user comments. On the other hand, we can assume that the individuals posting on the NYT website have similar sociodemographic profiles in terms of educational level and income. As our analysis shows, despite similar socio-demographic backgrounds, these individuals exhibited divergent opinions about healthcare systems and their comparison. Thus, in contrast



to previous studies, this result suggests that this group of commenters should not be treated as homogeneous.

Using sentence-level sentiment analysis, we obtained a sentiment score for 61 countries, distributed across several continents. To ensure the robustness of our findings, we focused on the set of 22 countries that were assigned the sentiment score more than once in the body of comments. Most of these countries belong to the European region but we also obtained the sentiment score for Australia, Cuba, Israel, and New Zealand. This allowed us to compare countries from different regions across the globe, which has not been possible using cross-national survey data that are often confined to a single region. Interestingly, although the European healthcare systems traditionally enjoy high levels of public support compared to other regions [25, 29], in our analysis the top three countries with the most positive sentiment score were found to be outside the European area (Israel, New Zealand, and Cuba).

The findings from the sentiment analysis at the country level for the European region produced results that are quite similar to those obtained on the basis of aggregated cross-national survey data. The results of the correlation analysis indicated a relatively high correspondence of country sentiment scores with the mean scores from the EQLS (0.70) and a moderate correspondence with the mean scores from the ESS (0.63). This is an encouraging result, considering the ease and the low cost associated with obtaining the UGC data, especially in comparison with the comparative survey data.

We should acknowledge certain limitations of this study that offer several opportunities for future work. First, like with much of the research using user-generated content, we should be cautious in generalizing our results. The majority of the individuals who participated in commenting online represent a general NYT readership. As mentioned above, in terms of the socio-demographics, this group of commenters is characterized by medium to high level of education and income. In addition, many of these individuals have either lived or studied in a different country. Thus, the results of this study may not be generalizable to other groups of respondents, particularly those with lower level of education and income, less experience abroad, and those with limited experience with the Internet.

Another limitation common to many studies employing user-generated data, is that the body of comments may grow while the platform for commenting remains open. Hence, it is possible that the findings based on the body of comments at the given time may be different when another time frame is analyzed. To illustrate, when we first accessed the data in March 2018, the NYT article "The Best Health-care System in the World" had 636 comments. As of January 24, 2020, the body of comments has grown to 771 comments, with the last comment made in April 12, 2019, when the administrators closed the comments section for this article. Much of UGC is real-time or streaming data and, as such, it is generated continuously. Implementing streaming APIs and similar tools will assist with gathering, managing, and analyzing such data, as they are generated.

Finally, as mentioned above, user-generated data are often complex and noisy. In this regard, UGC presents similar challenges to those of qualitative data. And much like the analysis of qualitative information, the analysis of UGC does not often lends itself easily to replicability, especially when not all steps in the analysis



are automatized. In this study, as part of the validation step,⁵ we had to conduct a manual check of the results of sentiment analysis and remove a number of cases to ensure consistency. This was possible considering the small set of comments analyzed but would be difficult and time-consuming with a larger corpus of comments. We hope that the development of better algorithms and analytical tools will ensure a greater replicability and reproducibility of UGC analysis in the future.

Despite these challenges, we believe that user-generated online data offer promising opportunities for public opinion research. This study provides only one preliminary examination of how user-generated data can be analyzed and future opportunities are abundant. For example, future studies may consider analyzing a larger corpus of online comments and/or different online platforms for gathering information on healthcare attitudes. While we do not consider the online user-generated content to be a direct competitor to more traditional survey data, we are certain that the richness of such data combined with appropriate analytical tools would be of great benefit to researchers of public opinion.

Acknowledgements The author would like to thank Seppe vanden Broucke for his valuable suggestions in the earlier stages of the paper.

Funding Not applicable.

Availability of data and materials Not applicable.

Code availability Not applicable.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

- Badawy, A., & Ferrara, E. (2018). The rise of jihadist propaganda on social networks. *Journal of Computational Social Science*, 1(2), 453–470.
- 2. Blendon, R. J., Benson, J., Donelan, K., Leitman, R., Taylor, H., Koeck, C., & Gitterman, D. (1995). Who has the best health care system? A second look. *Health Affairs*, 14(4), 220–230.
- Bleich, S. N., Özaltin, E., & Murray, C. J. (2009). How does satisfaction with the health-care system relate to patient experience? *Bulletin of the World Health Organization*, 87, 271–278.
- Bonikowski, B. (2017). Big data: challenges and opportunities for comparative historical sociology. Trajectories Newsletter of the ASA Comparative and Historical Section, 28(2), 29–32.
- 5. Bonoli, G., & Palier, B. (1998). Changing the politics of social programmes: Innovative change in British and French welfare reforms. *Journal of European Social Policy*, 8(4), 317–330.
- Cammett, M., Lynch, J., & Bilev, G. (2015). The influence of private health care financing on citizen trust in government. *Perspectives on Politics*, 13(4), 938–957.
- 7. Caren (2012). https://nealcaren.github.io/.
- Cohen, G. (1996). Age and health status in a patient satisfaction survey. Social Science & Medicine, 42(7), 1085–1093.
- Couper, M. P. (2011). The future of modes of data collection. Public Opinion Quarterly, 75(5), 889–908.

⁵ For an excellent discussion of the importance of validation see Grimmer and Stewart [15].



- Enghoff, O., & Aldridge, J. (2019). The value of unsolicited online data in drug policy research. *International Journal of Drug Policy*, 73, 210–218.
- 11. Feinerer, I. (2008). An introduction to text mining in R. The Newsletter of the R Project volume 8/2, October 2008 8 (2008):19.
- Gelissen, J. (2000). Popular support for institutionalised solidarity: A comparison between European welfare states. *International Journal of Social Welfare*, 9(4), 285–300.
- Gevers, J., Gelissen, J., Arts, W., & Muffels, R. (2000). Public health care in the balance: Exploring popular support for health care systems in the European Union. *International Journal of Social Welfare*, 9(4), 301–321.
- Golato, A. (2017). Naturally occurring data. The Routledge Handbook of Pragmatics (pp. 21–26). Routledge.
- 15. Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21(3), 267–297.
- 16. Groves, R. M. (2011). Three eras of survey research. Public Opinion Quarterly, 75(5), 861-871.
- Hall, J. A., & Dornan, M. C. (1990). Patient sociodemographic characteristics as predictors of satisfaction with medical care: A meta-analysis. Social Science & Medicine, 30(7), 811–818.
- 18. Harford, T. (2014). Big data: A big mistake? Significance, 11(5), 14–19.
- Havey, N. F. (2020). Partisan public health: How does political ideology influence support for COVID-19 related misinformation? *Journal of Computational Social Science*, 3(2), 319–342.
- He, W., Tian, X., Tao, R., Zhang, W., Yan, G., & Akula, V. (2017). Application of social media analytics: a case of analyzing online hotel reviews. *Online Information Review*, 41, 921–935.
- 21. Hutto, C. J., & Eric, G. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: *Eighth international AAAI conference on weblogs and social media*. 2014.
- Japec, L., Kreuter, F., Berg, M., Biemer, P., Decker, P., Lampe, C., Lane, J., Cathy, O., & Usher, A. (2015). Big data in survey research: AAPOR task force report. *Public Opinion Quarterly*, 79(4), 839–880.
- Jensen, C., & Naumann, E. (2016). Increasing pressures and support for public healthcare in Europe. Health Policy, 120(6), 698–705.
- 24. Kleinberg, B., van der Isabelle, V., & Paul, G. (2021). The temporal evolution of a far-right forum. *Journal of Computational Social Science*, 4(1), 1–23.
- Kohl, J., & Wendt, C. (2004). Satisfaction with health care systems. A comparison of EU countries.
 In W. Glatzer, S. V. Below, & M. Stoffregen (Eds.), Challenges for Quality of Life in the Contemporary World (pp. 311–331). Kluwer Academic Publishers.
- Kurian, J. C. (2015). Facebook use by the open access repository users. Online Information Review., 39, 903–922.
- Manosevitch, E., & Walker, D. (2009, April). Reader comments to online opinion journalism: A space of public deliberation. In *International Symposium on Online Journalism* Vol. 10, pp. 1–30.
- Missinne, S., Meuleman, B., & Bracke, P. (2013). The popular legitimacy of European healthcare systems: A multilevel analysis of 24 countries. *Journal of European Social Policy*, 23(3), 231–247.
- Mossialos, E. (1997). Citizens' views on health care systems in the 15 member states of the European Union. *Health Economics*, 6(2), 109–116.
- Naeem, B., Khan, A., Beg, M. O., & Mujtaba, H. (2020). A deep learning framework for clickbait detection on social area network using natural language cues. *Journal of Computational Social Science*, 3, 1–13.
- 31. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1-2), 1-135.
- Piña-García, C. A., Mario-Siqueiros-García, J., Robles-Belmont, E., Carreón, G., Gershenson, C.,
 & Amador-Díaz-López, J. (2018). From neuroscience to computer science: a topical approach on Twitter. *Journal of Computational Social Science*, 1(1), 187–208.
- 33. Rahmqvist, M., & Bara, A. C. (2010). Patient characteristics and quality dimensions related to patient satisfaction. *International Journal for Quality in Health Care*, 22(2), 86–92.
- Robinson, K. M. (2001). Unsolicited narratives from the Internet: A rich source of qualitative data. Qualitative Health Research, 11(5), 706–714.
- 35. Ryan, G., & Bernard, H. (2003). Techniques to identify themes. Field Methods, 15(1), 85-109.
- 36. Santana, A. D. (2011). Online readers' comments represent new opinion pipeline. *Newspaper Research Journal*, 32(3), 66–81.
- 37. Schober, M. F., Pasek, J., Guggenheim, L., Lampe, C., & Conrad, F. G. (2016). Social media analyses for social measurement. *Public Opinion Quarterly*, 80(1), 180–211.



- 38. Shahsavari, S., Holur, P., Wang, T., Tangherlini, T. R., & Roychowdhury, V. (2020). Conspiracy in the time of corona: Automatic detection of emerging COVID-19 conspiracy theories in social media and the news. *Journal of Computational Social Science*, *3*(2), 279–317.
- Souma, W., Vodenska, I., & Aoyama, H. (2019). Enhanced news sentiment analysis using deep learning methods. *Journal of Computational Social Science*, 2(1), 33–46.
- van der Vegt, I., Maximilian, M., Paul, G., & Bennett, K. (2021). Online influence, offline violence: language use on YouTube surrounding the 'Unite the Right'rally. *Journal of Computational Social Science*, 4(1), 333–354.
- Uyheng, J., & Carley, K. M. (2020). Bots and online hate during the COVID-19 pandemic: Case studies in the United States and the Philippines. *Journal of Computational Social Science*, 3(2), 445–468
- Wang, A.H.-E., Mei-chun, L., Min-Hsuan, W., & Puma, S. (2020). Influencing overseas Chinese by tweets: text-images as the key tactic of Chinese propaganda. *Journal of Computational Social Sci*ence, 3(2), 469–486.
- 43. Wendt, C., Kohl, J., Mischke, M., & Pfeifer, M. (2010). How do Europeans perceive their healthcare system? Patterns of satisfaction and preference for state involvement in the field of healthcare. *European Sociological Review*, 26(2), 177–192.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

