

## Article

# Deciphering Latent Health Information in Social Media Using a Mixed-Methods Design

George Shaw, Jr. <sup>1,\*</sup>, Margaret Zimmerman <sup>2</sup>, Ligia Vasquez-Huot <sup>3</sup> and Amir Karami <sup>4</sup>

<sup>1</sup> Department of Public Health Sciences, School of Data Science, University of North Carolina at Charlotte, Charlotte, NC 28223, USA

<sup>2</sup> School of Information Science, Florida State University, Tallahassee, FL 32306, USA

<sup>3</sup> Department of Public Health Sciences, University of North Carolina at Charlotte, Charlotte, NC 28223, USA

<sup>4</sup> Collat School of Business, University of Alabama Birmingham, Birmingham, AL 35233, USA

\* Correspondence: gshaw11@uncc.edu; Tel.: +1-704-562-0021

**Abstract:** Natural language processing techniques have increased the volume and variety of text data that can be analyzed. The aim of this study was to identify the positive and negative topical sentiments among diet, diabetes, exercise, and obesity tweets. Using a sequential explanatory mixed-method design for our analytical framework, we analyzed a data corpus of 1.7 million diet, diabetes, exercise, and obesity (DDEO)-related tweets collected over 12 months. Sentiment analysis and topic modeling were used to analyze the data. The results show that overall, 29% of the tweets were positive, and 17% were negative. Using sentiment analysis and latent Dirichlet allocation (LDA) topic modeling, we analyzed 800 positive and negative DDEO topics. From the 800 LDA topics—after the qualitative and computational removal of incoherent topics—473 topics were characterized as coherent. Obesity was the only query health topic with a higher percentage of negative tweets. The use of social media by public health practitioners should focus not only on the dissemination of health information based on the topics discovered but also consider what they can do for the health consumer as a result of the interaction in digital spaces such as social media. Future studies will benefit from using multiclass sentiment analysis methods associated with other novel topic modeling approaches.

**Keywords:** sentiment analysis; text-mining; obesity; social media; topic modeling; health communication



**Citation:** Shaw, G., Jr.; Zimmerman, M.; Vasquez-Huot, L.; Karami, A. Deciphering Latent Health Information in Social Media Using a Mixed-Methods Design. *Healthcare* **2022**, *10*, 2320. <https://doi.org/10.3390/healthcare10112320>

Academic Editors: Andrea Tittarelli and Daniele Giansanti

Received: 30 August 2022

Accepted: 13 November 2022

Published: 19 November 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Obesity is a complex health problem and continues to be a major health concern in the United States (U.S.). To encourage physicians to pay more attention to the condition and address the way health insurance companies pay for various treatments, the American Medical Association recently recognized obesity as a disease [1]. There is a need to identify health concerns related to obesity, chronic conditions associated with the disease, and modifiable behavior factors such as proper dieting and increasing physical activity [2,3]. Interviews and surveys are traditional data collection methods for federal and state public health agencies to collect behavioral health data concerning obesity [4–6]. While these are well-developed data collection methods [7,8], social media (SM) provides an additional data source to collect behavioral health data, and computational social science provides additional data collection methods [9,10]. Through SM, researchers can effectively and economically collect data about health behaviors and health risk factors.

People are using SM platforms to disseminate their health experiences and communicate with public health professionals or people with similar health experiences [5,11,12]. This adds a dynamic layer to health information-seeking behavior (HISB) in which such information seeking online is no longer strictly dependent upon static platforms. Within the context of SM, HISB is a layered, complex mechanism across a spectrum of actions and users that can include public health agencies disseminating quality information to fat-shaming

conversations on Twitter. While there is value across the spectrum of SM data, many public health agencies are not harnessing the knowledge that resides in these unstructured data and using SM platforms to create meaningful interactions with health consumers [13]. The information shared by users on SM platforms has been harnessed to analyze influenza, E. coli outbreaks, conjunctivitis, and heart disease [14–18]. When looking specifically at Twitter data, initial data collection focused on communicable diseases and began to include noncommunicable diseases as computational methods improved [15,16,19,20]. The improvement of computational science methods is changing how we conduct content analyses aimed at behaviors associated with noncommunicable diseases.

According to Lacy et al. (2015), a content analysis—from its original conceptual understanding—involves the process of categorizing data based on human input to answer a more significant research question surrounding the data [21]. While insightful, traditional content analysis is labor intensive and unfeasible with big data sets, computational approaches expedite this process [22]. Computational content analysis has been used on topics concerning social justice, business, and health [21,23,24]. From a health perspective, the content analysis of user-generated SM data has provided insights into spatial physical activity presence, the prediction of heart disease, and communication of shared user health behaviors [18,25,26].

Prior studies have used social media and various computational approaches to analyze diet, diabetes, exercise, and obesity (DDEO). The authors of [27] sought to identify the influence of social media on public health related to communicated health information using networking modeling. Another study conducted geospatial analysis of tweets to measure happiness, diet, and physical activity [28]. Ref. [2] studied the temporal trends in weight-loss-related posts. These and several additional studies used variations of sentiment analysis, topic modeling, or content analysis to analyze the data. However, these studies did not analyze DDEO topics collectively using SM data. Additionally, there has been limited work using a mixed-methods design to analyze and evaluate DDEO topics [24,29].

This research study adds to the breadth of knowledge that uses SM data to analyze health topics but focuses on topic coherence, qualitatively analyzing the relationship among four health topics (diet, diabetes, exercise, and obesity) and distinguishing SM association from HISB. While some public health departments are performing well with disseminating health information, there are opportunities for public health agencies to move beyond basic information dissemination [30]. Many public health agencies lack the support necessary for thoughtful SM engagement. SM has the potential to enhance the communication between individuals and public health agencies [31]. Moreover, understanding the topic discourse that is represented within SM allows public health agencies to be more strategic with information dissemination through this channel of communication [13]. Computational approaches can improve public health department response times to the volume and velocity of data that are generated by SM; refining how quickly we derive knowledge from these data is also harnessed through computational approaches.

With this study, we attempted to answer the question: What are the positive and negative topical sentiments among diet, diabetes, exercise, and obesity tweets? We attempted to provide a framework for analyzing DDEO health concerns hidden within SM data. The computational experiment is the leading focus of this work; however, secondary to the computational experiment is understanding the topics that are represented with DDEO. This study was designed to be hypothesis generating. Through this research experiment, the two aims of our research question were to:

- (1). Characterize DDEO topics through sentiment analysis and computational topic modeling;
- (2). Qualitatively identify the relationships among DDEO topics using the results from the two text-mining procedures.

### 1.1. Background

Obesity prevalence has increased over the past several years with 42.4% of the U.S. population suffering with obesity [7]. Obesity is a well-known risk factor for chronic conditions such as diabetes. People with obesity also experience higher medical costs [1]. Proper dieting and exercising are modifiable lifestyle behaviors that can help with reducing obesity and some of the various chronic conditions associated with it, in particular diabetes [3,31]. While conventional research methodologies have been utilized to gain insight into and characterize behaviors associated with obesity, DDEO data collected from SM require emerging computational methods for their analysis [32].

SM has become a fascinating lens through which we can surveil HISB. Never before has there been such a constant stream of residual data to offer insight into the HISB that can be striated so conveniently by population, topic, and time period. In seeking and exchanging health information through SM profiles, it is possible to group users by other public identifiers with some reliability. In this section, the current uses of SM to seek and disseminate health information will be explored, with special attention given to the platform Twitter, as it is the subject of this research.

According to the Pew Research Center, 72% of Americans use at least one SM platform [33]. While uptake is higher among people under 30 than under 50 (90% and 82%, respectively), users between the ages of 50 and 64 are the fastest-growing demographic with 69% using SM as of June 2019 [33]. SM usage is high, above 65%, in all groups when looking at each of the demographics of race, gender, income, education, and community type, such as urban or rural [33]. With such a large proportion of the population using SM, health information has the potential to reach a larger audience as 93 million Americans report that they look for health information online.

### 1.2. Health Information and SM

The behaviors related to health information seeking and SM are multifarious. SM is often used as a source of social support [33–35]. The combination of the private, insular nature of communicating from behind a device and the large community of users with diverse and potentially relevant experience is compelling, particularly with stigmatized issues such as obesity and diabetes [36–38]. There is, however, legitimate concern regarding the quality of user-generated health content as SM—including Twitter—has been used by groups and individuals who seek to dissuade others against advice from the medical community [39,40].

Quality assessments of SM information in academic literature are limited, with varying results reported. One study found that half of the health-related tweets analyzed contained false information. In addition, the tweets that did not contain false information were likely to originate from a medical institution [41]. An assessment of user-perceived quality of diabetes-related information on Twitter and Facebook was rated 62 out of a possible 100 [42]. Another study found that while there was high-quality information being disseminated on Twitter, users would need higher literacy skills than the average population's literacy skills to understand it [43]. An assessment of the usefulness of hashtags for organizing cancer information on Twitter assessed the information to be of high quality but did find that privacy was a great concern regarding sharing medical information in the public domain [44].

Another vein of SM research characterizes the types of conversations that users are having on a specific topic [45]. One article explored how humor was used to characterize obesity on Twitter [46]; derogatory jokes were retweeted more than positive ones, and significant attention was given to individual-level instead of societal-level causes for obesity. Mejorva found that fat shaming, or the practice of criticizing a person based on the size of their body, was present in a large share of the discourse happening in the 1.5 million tweets analyzed in their research [38]. Karami and his colleagues explored the various topics present in 4.5 million tweets that discussed diet, diabetes, exercise, and obesity [47]. To demonstrate the relationships between each of the primary topics, subtopics were used to

analyze the relationships, and strong correlations were found between exercise and obesity, as well as diabetes and obesity.

### 1.3. Credible SM Information Sources

It is also difficult to differentiate user-generated content from that produced by health professionals. Mejordva's work incorporating obesity and diabetes discovered that approximately half of the tweets were not affiliated with verifiable, reputable sources [38]. Moreover, tweets from nonreputable sources had a higher likelihood of being retweeted. Another study agreed with this; when assessing retweeting as a metric of reputation on Twitter, it was demonstrated that celebrities and news organizations are more likely to receive a high score than health organizations [48]. A newer study developed a predictive model that assesses the expertise of the user with some success, though vetting for accuracy on SM is an area that warrants considerable concern [49].

Regardless of these issues, there is a legitimate, though not prolific, argument made in scholarship that public health campaigns launched over SM can positively impact users [27,49,50]. SM has been found to be a valuable tool by which to engage the public in order to spread health information [51]. Twitter, in particular, has been utilized to successfully deliver behavioral weight loss interventions and vital diabetes information [28,52–54].

## 2. Materials and Methods

### 2.1. Study Design

In order to best address the research aims of this study, we used a sequential explanatory mixed-methods design. Mixed-method approaches in social media research have increased recently. Social media, as a data source, generates data that benefit from the data analysis strengths associated with quantitative and qualitative research. To characterize the topics, we placed more emphasis on the qualitative data [55]. There are an estimated forty mixed-methods research designs [56]. The sequential explanatory mixed-methods design incorporates the quantitative and qualitative findings in order to create more robust results and provide greater depth than either singular analysis would [55,56]. The sequential explanatory design used for this study consists of a quantitative phase that includes data collection and computational analysis, followed by a qualitative phase that incorporates qualitative data analysis to analyze the results from the topic model for evaluation purposes. The quantitative phase for this study incorporates computational steps to collect tweets, clean the data, conduct natural language processing to identify sentiment polarity, and conduct topic modeling. This type of research has been found to be particularly useful in the spectrum of health research [57]. These two phases inform each other, with the qualitative analysis based on results from the quantitative data; the qualitative phase is used for agreement and the evaluation of the quantitative phase topic model results [56]. Once both the quantitative and qualitative phases have occurred in sequence, the final analysis integrates the findings to enhance the value of the mixed-methods research [57,58]. The following sections outline the analytical framework used for this mixed-methods study.

### 2.2. Data Collection and Cleaning

Data used in this study were collected over a three-month period (June 2016–August 2016). These data were extracted from a larger data set that collected data over a 12-month period in 2016 and 2017 and demonstrated that diet (one of the DDEO topics) is important in relation to diet preferences and the political orientation of the state [59]. Using Java programming (Twitter4j) software, the Twitter API was used to amass the data set. Tweets collected were based on their meta-description of English-language, U.S.-based tweets. This method of data collection from Twitter allows you to collect data in real-time; however, this method has several drawbacks: (1) The Twitter API only allows you to stream roughly 10% of the publicly available tweets, (2) specific geo-location information is not always available for every tweet, and (3) there is an absence of observational context to inform the data captured. Therefore, this work did not attempt to analyze the topics according to

geographic location. Prior studies have demonstrated dieting behaviors and engagement in physical activity according to geographic location [25]. Health data pertaining to chronic conditions (“diabetes” and “obesity”) and modifiable behaviors associated with chronic health conditions (“diet” and “exercise”) were chosen as query terms. The hashtag and non-hashtag versions of each word in DDEO were used as query terms to search the Twitter API and generate the respective data set for each word. For the query terms, the two versions were used independently of each other during the search process within the Twitter API. The hashtag results and non-hashtag results were merged into one data set, reiterating the need to clean the data.

The data collection method used for this study involved passive monitoring. Passive monitoring is a low-cost and easy approach to data collection [10,60] (p. 24). Researchers are able to gain insight into the sentiments of users without actively engaging them. Passive monitoring has been used in politics, business, and other health topics [60–64]. Processing of the data collected required cleaning by removing stop words—such as and, of, the—based on a standard list of stop words. Additionally, leading whitespace, numbers, and special characters were removed from the data. This allowed the topic modeling toolkit, used to discover topics, to efficiently identify the topics for analytics purposes.

### 2.3. Sentiment Analysis

Sentiment analysis is a text mining method used to find the polarity (positive, negative, or neutral) in a data corpus. With success, previous studies have used sentiment analysis to detect opinion polarity concerning health topics [65]. This study used the lexicon-based approach to identify the sentiments; the linguistic inquiry and word count (LIWC) tool was used to perform this step of the study. Sentiment analysis was performed on each query term to identify the positive and negative sentiments [66]. The neutral sentiments were not included as part of the analysis. The study focused on sentiment expression for the health topics based on a positive or negative polarity. This approach is often used when capturing positive and negative sentiments using natural language processing techniques [67,68]. Based on this approach, we acquired a total of eight data sets representing the positive and negative polarity for DDEO.

### 2.4. Topic Modeling

A myriad of health information is communicated in SM spaces. As noted, reputable health care organizations struggle with reaching some intended audiences due to the volume of information disseminated by less credible sources [39,45]. To discover the latent semantic structure and knowledge represented in the data corpora, we conducted text analysis using an unsupervised topic modeling approach. Unsupervised topic modeling is used to discover patterns and describe the knowledge that is represented in unstructured data [68,69]. Using the machine learning for language toolkit (MALLET), the latent Dirichlet allocation (LDA) topic model was used [44,70]. LDA is a common topic-modelling approach, and its performance has been well-documented in other health-related studies involving Twitter data [62,71]. When examining the LDA model the LDA results are two matrices with  $m$  words and  $t$  topics for a given  $n$  of documents. LDA distributes topics over the words  $P(W_i | T_k)$  or is expressed as the probability of each word in each topic and the probability of each topic within each document (in this case, tweets)  $P(T_k | D_j)$ . This allows for a semantically coherent word set [72].

While there is no gold standard for determining the number of topics, several methods have been used to provide objective measures for the optimal number of topics to be analyzed [73]. For this study, we selected 100 topics for each sentiment. Computationally and qualitatively, we determined that this topic number would provide a sufficient representation of the data corpora to successfully perform the analysis for this study [74]. To evaluate the topics identified by the LDA model, we used a qualitative approach. This method does not consider objective analysis with regard to the performance of the model; however, the approach allows for a more in-depth analysis of performance based on topic

coherence. Topics were evaluated through the statistical measure of agreement (inter-rater reliability) [72].

### 2.5. Topic Evaluation

To evaluate the topics that were identified from the LDA model, Cohen's kappa was calculated. As previously noted, LDA is an unsupervised topic modeling approach to discovering patterns within a data corpus. Essentially, the model can be trained to cluster together words into topics, which then allows documents with similar topics to be clustered [10]. In this study, we used LDA for the exploratory discovery of topics. Human involvement is necessary for determining themes (topics) and discovering relevant study topics that are difficult to identify when using a topic-modeling method that does not require annotated data [75]. Inter-rater reliability was used to ensure homogeneity in identifying the topics and the stratified relationships among them. If the word in the topic cluster contained a high probability as identified by the model and could be semantically related to another topic, it was identified as being related to another topic. Cohen's kappa seeks to determine the level of agreement over and above the agreement that is expected through chance [76]. Using this measure, we were able to analyze the topic model results by incorporating a qualitative approach. That is, the topics were evaluated qualitatively with the intent to contextualize the topics. The topic evaluation process involved five steps:

Step 1: The LIWC tool was used to computationally identify health-related topics and polarity (positive or negative) of the four query terms [47].

Step 2: LDA topic modeling was performed on the positive and negative health-related topics as identified through the use of the LIWC tool. Analyzing over one million tweets would have required a substantial amount of human effort. Computationally, LDA performs the process exponentially faster while addressing issues of sparsity related to text mining [77].

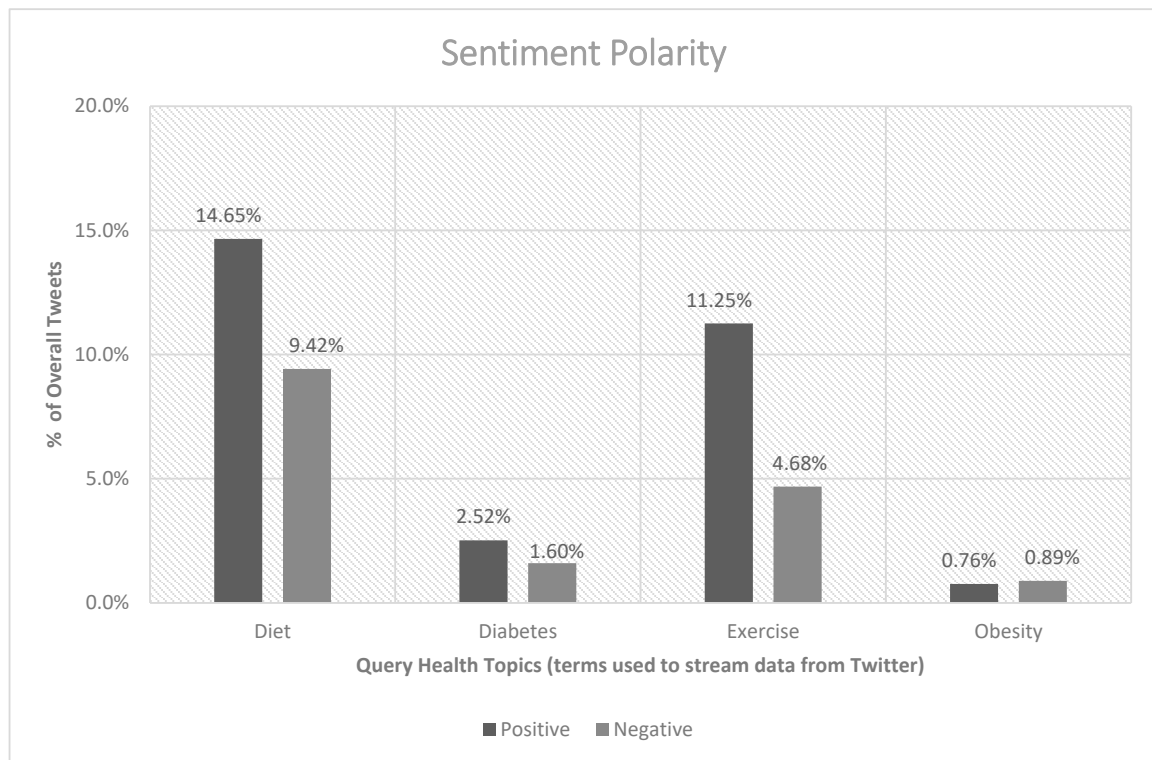
Step 3: The topic model results were then reviewed by two coders. They identified the topics as being related or unrelated to a DDEO health topic. If they were unrelated to a DDEO health topic, topics were removed, and no additional analyses were conducted on those topics.

Step 4: After all the non-DDEO-related health topics were removed, the coders were tasked with confirming topic coherence according to their characterization (labeling) as being DDEO related [10,14]. However, unlike the labeling performed in predictive computational studies, the labeling performed in this study was based on analyzing the representative word cluster for each topic.

Step 5: After the coders characterize the topics independently, they met to discuss disagreements. Once completed, Cohen's kappa was calculated to measure the agreement after the meeting.

## 3. Results

A total of 15 million tweets represented the data set used in this study. After removing retweets as part of the data cleaning process, the final data corpus consisted of 1.7 million tweets. Our first aim of this research involved characterizing the DDEO health topics using the aforementioned computational approaches. The following sections detail the descriptive statistics of the DDEO topics. When examining the overall positive and negative sentiment compositions of the tweets, 29% were positive and 17% were negative (see Figure 1); the remaining 54% of the tweets were neutral. Among the DDEO topics, the diet data corpus contained the highest number of positive and negative tweets. Positive and negative obesity-related tweets were the least among the DDEO topics.



**Figure 1.** Total sentiment polarity percentages for DDEO.

Eight hundred topics (100 for each DDEO sentiment) were chosen for the topic analysis. Using the LIWC dimension setting of health on the 800 topics [47], a total of 78 topics were unrelated to their respective health topic (Table 1). Through the qualitative approach, we identified an additional 250 topics that were not DDEO related (Table 1). This approach involved two researchers analyzing the topics according to word clusters. Overall, 59% (473) of the topics were coherent. Obesity was the most-identified topic based on the applied approach; exercise was the least-identified topic (Table 2).

**Table 1.** Total number of topics removed using LIWC and inter-rater agreement.

Health Topic	Topics Removed by LIWC (Positive and Negative)	Topics Removed by Coders (Positive and Negative)	Total Topics Removed (Positive and Negative)
Diet	13	54	67
Diabetes	26	65	91
Exercise	11	84	95
Obesity	28	46	73
Total	78	250	328

**Table 2.** Count and frequency distribution of topics after step 2 was completed.

	Positive	Negative	Total
Diet	59 (49%)	62 (51%)	121 (26%)
Diabetes	52 (43%)	70 (57%)	122 (26%)
Exercise	57 (54%)	48 (46%)	105 (22%)
Obesity	58 (46%)	66 (54%)	124 (26%)
Total	226 (48%)	246 (52%)	472

We also examined the prevalence of the remaining topics after step 1 (subsequently removing the 328 unrelated DDEO topics). Diet, diabetes, and obesity showed similar total frequency distribution, with exercise showing the least among the topics. In comparison, negative topics showed a higher prevalence across the topics; exercise was the exception, with a higher distribution across positive topics (Table 2). Our second aim of this research consisted of qualitatively identifying the relationships in DDEO using the results from the sentiment analysis and subsequent LDA model. When stratifying the DDEO topics to evaluate associations based on the topics, obesity had the highest association with the other topics (Table 3). While previous work has utilized statistical approaches to analyze correlations with other topics [47], the qualitative approach allowed for a more nuanced analysis of these topic associations. Although diabetes topics represented 26% of the total number of topics, diabetes had the fewest associations across the other topics based on the content analysis approach used.

**Table 3.** Stratified distribution of DDEO topic relationships.

	Diet	Diabetes	Exercise	Obesity
Diet		14	28	32
Diabetes	2		3	22
Exercise	27	6		15
Obesity	7	22	5	

Each topic is represented by T and the numeric value of its positioning among the topics. As noted in Table 4, T1 for positive diet topics represents the first topic (T) from the list of topics (1). Diet-related topics were the most inferable health topic. Diabetes, second to exercise, contained a significant portion of incoherent subtopics. Fifty-eight percent of the topics identified were related to negative sentiments. When analyzing the subtopics, a reoccurring theme we identified was chronic diseases (as noted by T4). The authors of [78] identified chronic disease with a large frequency distribution across negative topics regarding diabetes. When analyzing the subtopics for exercise, many of the positive and negative topics discussed user engagement in physical activity (positive—T4; negative—T36). Additionally, obesity was the only DDEO topic with slightly more negative sentiments than positive sentiments.

#### *Inter-Rater Reliability and DDEO Relationship*

The qualitative content analysis performed on the LDA topic results was also used to establish the reliability of the topics and the relationships among them. Inter-rater reliability demonstrated high reliability with regard to topic coherence of using the LDA topic results for topic analysis regarding DDEO. Additionally, all of the DDEO relationships coded revealed almost perfect agreement between the raters (Table 5). These results indicate the potential of this mixed-methods analytical approach for analyzing topics using unsupervised machine learning. A random sample of coders from a diversified population should be investigated to extend the evidence for and reliability of the analytical approaches we used.



**Table 4.** Sample of LDA topics representing each DDEO element (topics were conveniently selected).

		Positive	Negative	
Diet	T1	diet—meat—healthy—fruit—fruits—veggies—vegetables	T17	bad—craving—whataburger—train—break—habit—crossfit
	T4	based—plant—vegan—health benefits—healthy—vegetarian	T18	coke—mcdonald—large—bottle—fridge—hangover—addicted
	T10	diet—diabetes—exercise - blood food—nutrition - sugar	T28	poor—health—problems—obesity emotional—physical—activity
Diabetes	T8	diabetes—care—supplies—insulin—medical—insurance—money	T4	fibrosis—cystic celiac—causing—mellitus—epidemic—disease—endocrine
	T11	diabetes—healing—cancer—god pray—hypertension—energy	T7	loss—weight—diet—exercise—surgery—prediabetes—patient
	T19	Ice—cream—chocolate—love—sugar—coffee—donuts	T20	meat—cancer—antibiotics—hormones—dairy—vegan—diseases
Exercise	T4	exercise—body—stress—yoga—soul—meditation—breathing	T6	weight—lose—diet—fat—eating—food—pills
	T14	fitness—workout—gym—health fitfam—training—cardio	T25	stress—depression—anxiety—helps endorphins—brain—mood
	T41	exercise—pokemon—playing pokemongo—people—walk—game	T36	hate—running—gym—worst—working—kind—stupid
Obesity	T16	obesity—activity—physical—social—reduce—active—fitness	T7	poor—diabetes—dental—warning soda—consumption—health
	T27	obesity—pokemon—childhood epidemic—america—pokemongo walking	T16	poka—obesity—bmi—time—game proportional—complication
	T70	diabetes—obesity—cancer disease—cholesterol—hypertension—insulin	T44	syrup—corn—obesity—promoted fructose—markets—household

**Table 5.** Cohen’s kappa agreement for each topic based on DDEO coherence agreement and identification of topic relationships.

Topic	Positive		Negative	
	Inter-Rater Reliability		Inter-Rater Reliability	
	Topic Coherence	DDEO Topic Relation	Topic Coherence	DDEO Topic Relation
Diet	1.000	1.000	0.955	1.000
Diabetes	0.958	1.000	0.969	1.000
Exercise	0.979	0.921	0.978	1.000
Obesity	0.885	1.000	1.000	0.969

**4. Discussion**

It is difficult to infer the three dominated messages normally found on Twitter—commentaries and opinions, highly personal moment-to moment sentiments and emotions, and informational—through topic model results alone [79]. However, these topics provide insight for health care practitioners who are interested in quickly analyzing large unstructured SM data sets to understand the information being communicated regarding a particular health topic. More importantly, this method uncovers hidden patterns of data (information) that would normally be discarded due to the topics that have a higher frequency distribution within the data set. The following discussion section utilizes the results from the qualitative analytical process and represents the hypothesis generating discussion that would be replicated by health care practitioner’s or public health agencies. Pseudocode was used to increase the anonymity of the tweets analyzed in this study while retaining the original sentiments of the users. However, this process removes the semantical structure of their original communication.

#### 4.1. Analyzing the Health Topics Diet

When analyzing positive and negative subtopics for diet, many of the topics appear to reference food or specific diets. As seen in the positive diet topic T4 (Table 3), we infer that the topic is referencing a vegan or vegetarian diet. Several studies have indicated the benefits of a plant-based diet; particularly with reducing people's risk to chronic conditions like diabetes, cardiovascular disease, and high cholesterol [80–82].

Contrary to the health benefits from a plant-based diet, the negative topics associated with diet indicate the consumption of processed food, in-addition to exercising. One twitter reader tweeted “So my dad's supposed to be on this 30-day diet challenge thing, right? Why did I find a stash of KitKats a few moments ago . . . .” This sentiment is supported by T17. Moreover, T28 also illuminates the emotions that are involved with proper dieting behavior. When we are dealing with negative emotions, impulsive behavior is a mechanism that we use to cope with stress. In some cases, this can lead to overeating and consuming excess calories in a dissociative manner [83].

##### 4.1.1. Diabetes

The positive topics for diabetes covered an array of subtopics like food, spiritual healing, diabetes management, and emotions. As noted in T19, this topic serves as an oxymoron with regard to the diabetes health topic and our interpretation of this topic (sweets). The word cluster for this topic contains foods that are high in sugar with no nutritional value [84]. One user tweeted “my midnight snacks consist of sugar and bagels. Diabetes is what I may have if I continue to eat this way.” Another user says, “Sweat tea from McDonalds is that diabetes in a cup.” Absent from the analysis was the tracking of users over time and the geolocation information. Therefore, we are not able to make inferences about particular geographic regions. However, Nguyen et al. have demonstrated the relationship between healthy food references and economically disadvantaged census tract locations [25].

A latent negative topic inferred from the analytical approach was family history and the relationship with diabetes. One twitter user mentioned the connection between diabetes in their family and current diabetic symptoms. While research does support that people have genetic disposition to the disease, family culture and behavioral factors regarding food consumption plays a role in diabetes prevalence [25,85,86].

##### 4.1.2. Exercise

The sentiment complexity of the exercise topic is captured in the following tweet: “Freedom, exercise, and me time is what my bike has meant to me . . . more than I can express in words.” Another user tweeted, “On this journey, dieting is so much easier than exercise. I need a personal trainer to get my fitness motivation back suggestions.” For health care practitioners, the latter tweet provides opportunities for user engagement, particularly with improving active participation and two-way communication between SM users and public health agencies. Currently, there is a lack of engagement from public health agencies and health care professionals. Health care practitioners will benefit from creating engaged communities through SM interactions [87,88]. Increased SM engagement also allows health care practitioners to disseminate credible information in spaces that can be dominated by misinformation [89].

Within our positive topics, we also noticed that individuals use Twitter as a digital space to disseminate mobile gaming behavior. Gaming applications are changing how people and researchers view the activities that reflect physical activity [90]. The augmented reality (AR) game—Pokémon Go is an example of mobile gaming behavior that was identified through the topic evaluation (T41). However, this AR application can also lead to unattended accidents due to mobile vehicle distraction and pedestrians lack of awareness in their surroundings [91]. Again, situations like these present opportunities for health care practitioners—public health in particular—to not only disseminate but create engagement with users regarding the drawback of this physical activity behavior.

For this research study, the textual analysis processing task used on the content was completed using n-gram analysis. As a result, this creates an added layer of complexity in the topic analysis process by using the qualitative method. A user tweeted, “I would like to say that the Olympics has inspired me, but it is really due to the fat shaming I expect in California next month.” Based on their tweet, it appears to be some behavioral motivation expressed for exercising, but the remaining portion of their sentiment expresses an alternative motivation factor. The use of another text analysis processing method may have represented these distinguished sentiments better and improved step three of the analytical framework.

#### 4.1.3. Obesity

Positive topic 27 for obesity indicates the potential impact Pokémon Go (Exercise: T41) and other AR gaming can have with addressing childhood obesity. However, there is bleak optimism on AR gaming applications like Pokémon Go and impacting childhood obesity. There are questions regarding the lack of sustainability by these game applications. Physical activity returns to baseline performance after a few weeks [92]. A positive twitter comment supports the link that scientists have made between obesity and 13 types of diseases. These types of comments are identified through topics like T70. Diabetes in men, hypertension, and cholesterol are all chronic conditions that have been associated with obesity [93,94].

A twitter user expressed negative sentiments concerning obesity related conditions: “there is something when you know your life is slowly slipping away because of obesity-related health problems.” For public health departments that focus on oral health, T7 indicates the opportunity to disseminate and engage individuals regarding their oral health. According to the CDC, tooth decay is one of the most prevalent chronic diseases in the United States. Health risk behaviors that consist of drinking and eating foods that are high in sugar, are significant contributors to this problem [95]. State health departments communicating dental health can benefit from the information gathered through SM and the content users disseminate through these platforms [5]. Early SM research involving state health departments and health communication showed low user engagement [13]. However, the use of SM by local or state health departments should focus not only on the dissemination of health information but also consider what the agency can do for the health consumer through those SM interactions.

#### 4.2. Implications

This study adds to the breadth of knowledge regarding mixed methods approaches for computational topic discovery. This study also used open-source and low-cost text mining methods to analyze the data. For many public health agencies with limited resources or lack of staff with analytical expertise, these methods can be deployed within their health care setting without significant disruption to current workflows. Additionally, public health practitioners can apply this method to qualitative survey data. Analyzing qualitative survey data using this method may elicit topics that can be important for addressing process measures impacting quality of care for public health care organizations. When considering possible use cases specific to public health practitioners in large cities, this method can be used to possibly identify health concerns through geocoded tweets. This method provides practitioners with a data-driven approach to understanding the needs of the community they serve by using big data to inform decision making [96]. This work also has implications for clinical settings that rely on patient feedback to improve their processes.

From a research perspective, this work adds to the breadth of methodological approaches that seek to discover and interpret the knowledge provided by these data sources regarding DDEO. While this data-driven research is grounded in data science computational methods [10], this work generated a hypothesis that allowed for the application of information-seeking theoretical frameworks. With an effective strategy, this analytical

method can be used for other unstructured data sets that are collected by health care practitioners and public health agencies.

#### 4.3. Study Limitations

One limitation of this study is that we did not seek to analyze agreement prior to the coders meeting. There were distinct domain differences between the coders related to DDEO, and we expected a weak disagreement between the coders. The lack of context is another drawback of research involving topic modeling. Understanding the relational dynamics of DDEO topic communication on Twitter can be improved by the use and evaluation of other topic model approaches such as the correlated topic model (CTM). CTM allocates the relationships across topics and extends the topical functions of the LDA model [97]. Analyzing the quantity and interaction of DDEO information dissemination among credible sources is an opportunity for additional research. We also did not consider the temporal and spatial data of the tweets. The data used for our study were collected during the summer, and this might have impacted the volume of diet- and exercise-related tweets. Lastly, the sentiment analysis tool utilized in this study calculated sentiment polarity based on the overall sentiment expressed by the tweets. Future studies will benefit from using multiclass sentiment analysis methods associated with machine learning techniques like BERT in conjunction with novel topic modeling approaches.

#### 5. Conclusions

People use Twitter and other SM platforms to communicate their health sentiments. These sentiments include health experiences that contain complex semantical structures. Sentiment analysis and topic modeling are effective text mining approaches for topically inferring information from these voluminous data sets. Using these two approaches, we were able to demonstrate the analysis process based on the analytical framework outlined.

When examining the entire composition of the final data corpus (1.7 million tweets), 29% were positive and 17% were negative. Using the computational and qualitative methods, we removed 328 topics that were not DDEO related. However, during the qualitative phase of the topic removal process, we were able to identify three times the number of unrelated DDEO topics. Except for exercise, most of the topics representing DDEO were negative. Diet was the most inferable topic; based on our sample subtopic analysis, food and diets were the most specific topics represented with regard to diet.

Unlike computational approaches that are largely rule-based when classifying topics, the qualitative approach creates challenges when classifying a tweet as DDEO related. Coders infuse their positionality into the process. However, the use of an agreement measure adds an additional method of identifying and evaluating the varying degree coders may have despite a clear coding protocol or equal category proportions [98]. The framework used in this study provides an additional opportunity for transdisciplinary work to be conducted as it relates to DDEO topics. While this framework can be generalized to other social media topics, the nuances involved with examining the word clusters could create concerns regarding the quality of the results. Despite these concerns, additional research with a strong interdisciplinary team is warranted for understanding the potential concerns related to the quality of the results from this analytic framework.

As a digital space, Twitter is a popular SM platform for health communication [99], but many public health practitioners and agencies are using the platform for the one-way dissemination of information. Limited resources and training are needed to conduct this methodology. SM information dissemination should be an initial step in the interaction process to engage SM users and create a relationship beyond the digital space.

**Author Contributions:** Conceptualization, G.S.J.; Methodology G.S.J., M.Z., L.V.-H., and A.K. Software, G.S.J. and A.K.; Computational Analysis, G.S.J. and A.K.; Validation, M.Z. and L.V.-H.; Data Curation, A.K.; Writing-original Draft Preparation, G.S.J., M.Z., and L.V.-H.; Writing-review and Editing, G.S.J., M.Z., and L.V.-H.; Visualization, G.S.J.; Supervision, A.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data available upon request.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Kim, D.D.; Basu, A. Estimating the Medical Care Costs of Obesity in the United States: Systematic Review, Meta-Analysis, and Empirical Analysis. *Value Health* **2016**, *19*, 602–613. [CrossRef]
2. Turner-McGrievy, G.M.; Beets, M.W. Tweet for health: Using an online social network to examine temporal trends in weight loss-related posts. *Transl. Behav. Med.* **2015**, *5*, 160–166. [CrossRef]
3. Wing, R.R.; Goldstein, M.G.; Acton, K.J.; Birch, L.L.; Jakicic, J.M.; Sallis, J.F.; Smith-West, D.; Jeffery, R.W.; Surwit, R.S. Behavioral Science Research in Diabetes: Lifestyle changes related to obesity, eating behavior, and physical activity. *Diabetes Care* **2001**, *24*, 117–123. [CrossRef] [PubMed]
4. Creswell, J.W. Quantitative Methods. In *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*; Sage Publications: Thousand Oaks, CA, USA, 2014.
5. Jha, A.; Lin, L.; Savoia, E. The use of social media by state health departments in the US: Analyzing health communication through Facebook. *J. Community Health* **2016**, *41*, 174–179. [CrossRef] [PubMed]
6. Pierannunzi, C.; Hu, S.S.; Balluz, L. A systematic review of publications assessing reliability and validity of the Behavioral Risk Factor Surveillance System (BRFSS), 2004–2011. *BMC Med. Res. Methodol.* **2013**, *13*, 49. [CrossRef]
7. CDC. Adult Obesity Facts. 2021. Available online: <https://www.cdc.gov/obesity/data/adult.html> (accessed on 1 December 2021).
8. Forrest, K.Y.Z.; Lin, Y. *Comparison of Health-Related Factors between Rural and Urban Pennsylvania Residents Using Behavioral Risk Factor Surveillance System (brfss) Data*; The Center for Rural Pennsylvania: Harrisburg, PA, USA, 2010.
9. Oboler, A.; Welsh, K.; Cruz, L. The danger of big data: Social media as computational social science. *First Monday* **2012**, *17*, 7. [CrossRef]
10. Paul, M.J.; Dredze, M. Social Monitoring for Public Health. *Synth. Lect. Inf. Concepts Retr. Serv.* **2017**, *9*, 1–183. [CrossRef]
11. Chretien, K.C.; Kind, T. Social media and clinical care: Ethical, professional, and social implications. *Circulation* **2013**, *127*, 1413–1421. [CrossRef]
12. Zhou, L.; Zhang, D.; Yang, C.C.; Wang, Y. Harnessing social media for health information management. *Electron. Commer. Res. Appl.* **2017**, *27*, 139–151. [CrossRef]
13. Thackeray, R.; Neiger, B.L.; Smith, A.K.; Van Wagenen, S.B. Adoption and use of social media among public health departments. *BMC Public Health* **2012**, *12*, 242. [CrossRef]
14. Aramaki, E.; Maskawa, S.; Morita, M. Twitter catches the flu: Detecting influenza epidemics using Twitter. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Scotland, UK, 27–31 July 2011; pp. 1568–1576.
15. Culotta, A. *Towards Detecting Influenza Epidemics by Analyzing Twitter Messages*; ACM: Washington, DC, USA, 2010; pp. 1–8.
16. Diaz-Aviles, E.; Stewart, A. Tracking Twitter for epidemic intelligence: Case study. In Proceedings of the Web Science Conference, Boston, MA, USA, 30 June–3 July 2019; pp. 82–85.
17. Deiner, M.S.; Lietman, T.M.; McLeod, S.D.; Chodosh, J.; Porco, T.C. Surveillance Tools Emerging From Search Engines and Social Media Data for Determining Eye Disease Patterns. *JAMA Ophthalmol.* **2016**, *134*, 1024–1030. [CrossRef]
18. Eichstaedt, J.C.; Schwartz, H.A.; Kern, M.L.; Park, G.; Labarthe, D.R.; Merchant, R.M.; Jha, S.; Agrawal, M.; Dziurzynski, L.A.; Sap, M.; et al. Psychological Language on Twitter Predicts County-Level Heart Disease Mortality. *Psychol. Sci.* **2015**, *26*, 159–169. [CrossRef]
19. Chew, C.; Eysenbach, G. Pandemics in the age of Twitter: Content analysis of Tweets during the 2009 H1N1 outbreak. *PLoS ONE* **2010**, *5*, e14118. [CrossRef]
20. Eschler, J.; Dehlawi, Z.; Pratt, W. Self-characterized illness phase and information needs of participants in an online cancer forum. *Proc. Int. AAAI Conf. Web Soc. Media* **2015**, *9*, 101–109. [CrossRef]
21. Lacy, S.; Watson, B.R.; Riffe, D.; Lovejoy, J. Issues and Best Practices in Content Analysis. *Journal. Mass Commun. Q.* **2015**, *92*, 791–811. [CrossRef]
22. Harris, J.K.; Mart, A.; Moreland-Russell, S.; Caburnay, C.A. Diabetes topics associated with engagement on twitter. *Prev. Chronic Dis.* **2015**, *12*, E62. [CrossRef]
23. Bollen, J.; Mao, H.; Zeng, X. Twitter mood predicts the stock market. *J. Comput. Sci.* **2011**, *2*, 1–8. [CrossRef]
24. Liu, Y.; Mei, Q.; Hanauer, D.A.; Zheng, K.; Lee, J.M. Use of Social Media in the Diabetes Community: An Exploratory Analysis of Diabetes-Related Tweets. *JMIR Diabetes* **2016**, *1*, e4. [CrossRef]
25. Nguyen, Q.C.; Kath, S.; Meng, H.-W.; Li, D.; Smith, K.R.; VanDerslice, J.A.; Wen, M.; Li, F. Leveraging geotagged Twitter data to examine neighborhood happiness, diet, and physical activity. *Appl. Geogr.* **2016**, *73*, 77–88. [CrossRef]

26. Abbar, S.; Mejova, Y.; Weber, I. You Tweet what you eat: Studying food consumption through Twitter. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, Seoul, Republic of Korea, 18–23 April 2015; pp. 3197–3206.
27. Harris, J.K.; Moreland-Russell, S.; Tabak, R.G.; Ruhr, L.R.; Maier, R.C. Communication About Childhood Obesity on Twitter. *Am. J. Public Health* **2014**, *104*, e62–e69. [CrossRef]
28. Gore, R.J.; Diallo, S.Y.; Padilla, J.J. You Are What You Tweet: Connecting the Geographic Variation in America’s Obesity Rate to Twitter Content. *PLoS ONE* **2015**, *10*, e0133505. [CrossRef]
29. Salas-Zárate, M.D.P.; Medina-Moreira, J.; Lagos-Ortiz, K.; Luna-Aveiga, H.; Rodríguez-García, M.; Valencia-García, R. Sentiment Analysis on Tweets about Diabetes: An Aspect-Level Approach. *Comput. Math. Methods Med.* **2017**, *2017*, 5140631. [CrossRef]
30. Harris, J.K.; Mueller, N.L.; Snider, D. Social Media Adoption in Local Health Departments Nationwide. *Am. J. Public Health* **2013**, *103*, 1700–1707. [CrossRef]
31. Park, A.; Bowling, J.; Shaw, G.; Li, C.; Chen, S. Adopting social media for improving health: Opportunities and challenges. *North Carol. Med. J.* **2019**, *80*, 240–243. [CrossRef]
32. Flegal, K.M.; Carroll, M.D.; Kit, B.K.; Ogden, C.L. Prevalence of obesity and trends in the distribution of body mass index among US adults, 1999–2010. *J. Am. Med. Assoc.* **2012**, *307*, 491–497. [CrossRef]
33. Pew Research Center, Internet and Technology. Social Media Fact Sheet. 2019. Available online: <https://www.pewresearch.org/internet/fact-sheet/social-media/> (accessed on 1 July 2019).
34. Meng, J.; Martinez, L.; Holmstrom, A.; Chung, M.; Cox, J. Research on Social Networking Sites and Social Support from 2004 to 2015: A Narrative Review and Directions for Future Research. *Cyberpsychology Behav. Soc. Netw.* **2017**, *20*, 44–51. [CrossRef]
35. Naslund, J.A.; Aschbrenner, K.A.; Marsch, L.A.; Bartels, S.J. The future of mental health care: Peer-to-peer support and social media. *Epidemiol. Psychiatr. Sci.* **2016**, *25*, 113–122. [CrossRef]
36. Shepherd, A.; Sanders, C.; Doyle, M.; Shaw, J. Using social media for support and feedback by mental health service users: Thematic analysis of a twitter conversation. *BMC Psychiatry* **2015**, *15*, 29. [CrossRef]
37. Lydecker, J.A.; Galbraith, K.; Ivezaj, V.; White, M.A.; Barnes, R.D.; Roberto, C.A.; Grilo, C.M. Words will never hurt me? Preferred terms for describing obesity and binge eating. *Int. J. Clin. Pr.* **2016**, *70*, 682–690. [CrossRef]
38. Mejova, Y. Information Sources and Needs in the Obesity and Diabetes Twitter Discourse. In Proceedings of the 2018 International Conference on Digital Health, Lyon, France, 23–26 April 2018; pp. 21–29.
39. Schabert, J.; Browne, J.L.; Mosely, K.; Speight, J. Social stigma in diabetes. *Patient-Patient-Cent. Outcomes Res.* **2013**, *6*, 1–10. [CrossRef]
40. Del Vicario, M.; Bessi, A.; Zollo, F.; Petroni, F.; Scala, A.; Caldarelli, G.; Stanley, H.E.; Quattrociocchi, W. The spreading of misinformation online. *Proc. Natl. Acad. Sci.* **2016**, *113*, 554–559. [CrossRef]
41. Dunn, A.G.; Leask, J.; Zhou, X.; Mandl, K.D.; Coiera, E. Associations Between Exposure to and Expression of Negative Opinions About Human Papillomavirus Vaccines on Social Media: An Observational Study. *J. Med. Internet Res.* **2015**, *17*, e144. [CrossRef]
42. Alnemer, K.A.; Alhuzaim, W.M.; Alnemer, A.A.; Alharbi, B.B.; Bawazir, A.S.; Barayyan, O.R.; Balaraj, F.K. Are Health-Related Tweets Evidence Based? Review and Analysis of Health-Related Tweets on Twitter. *J. Med. Internet Res.* **2015**, *17*, e246. [CrossRef]
43. Gabarron, E.; Årsand, E.; Wynn, R. Social Media Use in Interventions for Diabetes: Rapid Evidence-Based Review. *J. Med. Internet Res.* **2018**, *20*, e10303. [CrossRef]
44. Trethewey, S.P. Strategies to combat medical misinformation on social media. *Postgrad. Med. J.* **2019**, *96*, 4–6. [CrossRef]
45. Katz, M.S.; Anderson, P.F.; Thompson, M.A.; Salmi, L.; Freeman-Daily, J.; Utengen, A.; Dizon, D.S.; Blotner, C.; Cooke, D.T.; Sparacio, D.; et al. Organizing Online Health Content: Developing Hashtag Collections for Healthier Internet-Based People and Communities. *JCO Clin. Cancer Inform.* **2019**, *3*, 1–10. [CrossRef]
46. So, J.; Prestin, A.; Lee, L.; Wang, Y.; Yen, J.; Chou, W.-Y.S. What Do People Like to “Share” About Obesity? A Content Analysis of Frequent Retweets About Obesity on Twitter. *Health Commun.* **2015**, *31*, 193–206. [CrossRef]
47. Karami, A.; Dahl, A.A.; Turner-McGrievy, G.; Kharrazi, H.; Shaw, G., Jr. Characterizing diabetes, diet, exercise, and obesity comments on Twitter. *Int. J. Inf. Manag.* **2018**, *38*, 1–6. [CrossRef]
48. Weitzel, L.; de Oliveira, J.P.M.; Quaresma, P. Measuring the Reputation in User-generated-content Systems Based on Health Information. *Procedia Comput. Sci.* **2014**, *29*, 364–378. [CrossRef]
49. Horne, B.D.; Nevo, D.; Adalı, S. Recognizing experts on social media: A heuristics-based approach. *ACM SIGMIS Database DATABASE Adv. Inf. Syst.* **2019**, *50*, 66–84. [CrossRef]
50. Krueger, E.A.; Chiu, C.J.; Menacho, L.A.; Young, S.D. HIV testing among social media-using Peruvian men who have sex with men: Correlates and social context. *AIDS Care* **2016**, *28*, 1301–1305. [CrossRef] [PubMed]
51. Novillo-Ortiz, D.; Hernández-Pérez, T. Social media in public health: An analysis of national health authorities and leading causes of death in Spanish-speaking Latin American and Caribbean countries. *BMC Med. Inform. Decis. Mak.* **2017**, *17*, 16. [CrossRef] [PubMed]
52. Kass-Hout, T.A.; Alhinnawi, H. Social media in public health. *Br. Med. Bull.* **2013**, *108*, 5–24. Available online: [https://www.researchgate.net/profile/Taha-Kass-Hout/publication/257533135\\_Social\\_media\\_in\\_public\\_health/links/596e5d0ca6fdcc2416901343/Social-media-in-public-health.pdf](https://www.researchgate.net/profile/Taha-Kass-Hout/publication/257533135_Social_media_in_public_health/links/596e5d0ca6fdcc2416901343/Social-media-in-public-health.pdf) (accessed on 15 February 2018). [CrossRef] [PubMed]
53. Park, H.; Reber, B.H.; Chon, M.-G. Tweeting as health communication: Health organizations’ use of Twitter for health promotion and public engagement. *J. Health Commun.* **2015**, *21*, 188–198. [CrossRef] [PubMed]

54. Reuter, K.; Jones, K.; Dejonckheere, M.; Stevens, R.C.; Brawner, B.M.; Kranzler, E.; Giorgi, S.; Lazarus, E.; Abera, M.; Huang, S.; et al. Exploring Substance Use Tweets of Youth in the United States: Mixed Methods Study. *JMIR Public Health Surveill.* **2020**, *6*, e16191. [[CrossRef](#)]
55. Tashakkori, A.; Teddlie, C. Issues and dilemmas in teaching research methods courses in social and behavioural sciences: US perspective. *Int. J. Soc. Res. Methodol.* **2003**, *6*, 61–77. [[CrossRef](#)]
56. Fetters, M.D.; Curry, L.A.; Creswell, J.W. Achieving Integration in Mixed Methods Designs-Principles and Practices. *Health Serv. Res.* **2013**, *48*, 2134–2156. [[CrossRef](#)]
57. Ivankova, N.V.; Creswell, J.W.; Stick, S.L. Using Mixed-Methods Sequential Explanatory Design: From Theory to Practice. *Field Methods* **2006**, *18*, 3–20. [[CrossRef](#)]
58. Dickson, V.V.; Page, S.D. Using mixed methods in cardiovascular nursing research: Answering the why, the how, and the what's next. *Eur. J. Cardiovasc. Nurs.* **2021**, *20*, 82–89. [[CrossRef](#)]
59. Karami, A.; Dahl, A.A.; Shaw, G.; Valappil, S.P.; Turner-McGrievy, G.; Kharrazi, H.; Bozorgi, P. Analysis of Social Media Discussions on (#) Diet by Blue, Red, and Swing States in the US. *Multidiscip. Digit. Publ. Inst. Healthc.* **2021**, *9*, 518.
60. Komito, L. Social media and migration: Virtual community 2.0. *J. Am. Soc. Inf. Sci. Technol.* **2011**, *62*, 1075–1086. [[CrossRef](#)]
61. Tumasjan, A.; Sprenger, T.O.; Sandner, P.G.; Welpe, I.M. Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. In Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media, Washington, DC, USA, 23–26 May 2010; pp. 178–185. [[CrossRef](#)]
62. Gallagher, J.; Ransbotham, S. Social media and customer dialog management at Starbucks. *MIS Quarterly Executive* **2010**, *9*.
63. Comito, C.; Pizzuti, C.; Procopio, N. How people talk about health? Detecting health topics from Twitter streams. In Proceedings of the BDIOT, Beijing, China, 24–26 October 2018; pp. 1–6. [[CrossRef](#)]
64. Ghosh, D.; Guha, R. What are we 'tweeting' about obesity? Mapping tweets with topic modeling and Geographic Information System. *Cartogr. Geogr. Inf. Sci.* **2013**, *40*, 90–102. [[CrossRef](#)] [[PubMed](#)]
65. On, J.; Park, H.-A.; Song, T.-M.; Erdley, W.; Brixey, J.; Bartlett, R. Sentiment Analysis of Social Media on Childhood Vaccination: Development of an Ontology. *J. Med. Int. Res.* **2019**, *21*, e13456. [[CrossRef](#)]
66. Medhat, W.; Hassan, A.; Korashy, H. Sentiment analysis algorithms and applications: A survey. *Ain Shams Eng. J.* **2014**, *5*, 1093–1113. [[CrossRef](#)]
67. Nasukawa, T.; Yi, J. Sentiment analysis: Capturing favorability using natural language processing. In Proceedings of the 2nd International Conference on Knowledge Capture, Sanibel Island, FL, USA, 23–25 October 2003; pp. 70–77.
68. Yang, F.-C.; Lee, A.J.; Kuo, S.-C. Mining Health Social Media with Sentiment Analysis. *J. Med. Syst.* **2016**, *40*, 236. [[CrossRef](#)]
69. Paul, M.J.; Dredze, M. A Model for Mining Public Health Topics from Twitter. *Health* **2012**, *11*, 1.
70. Fong, S.; Zhuang, Y.; Li, J.; Khoury, R. Sentiment analysis of online news using Mallet. In Proceedings of the 2013 International Symposium on Computational and Business Intelligence, New Delhi, India, 24–26 August 2013; pp. 301–304.
71. Wallace, B.C.; Paul, M.J.; Sarkar, U.; Trikalinos, T.A.; Dredze, M. A large-scale quantitative analysis of latent factors and sentiment in online doctor reviews. *J. Am. Med. Inform. Assoc.* **2014**, *21*, 1098–1103. [[CrossRef](#)]
72. Chang, J.; Gerrish, S.; Wang, C.; Boyd-Graber, J.L.; Blei, D.M. Reading tea leaves: How humans interpret topic models. *Adv. Neural Inf. Process. Syst.* **2009**, *22*, 288–296.
73. Graham, S.; Weingart, S.; Milligan, I. Getting Started with Topic Modeling and MALLET. The Editorial Board of the Programming Historian. 2012. Available online: <https://programminghistorian.org/en/lessons/topic-modeling-and-mallet> (accessed on 1 February 2018).
74. Shaw, G., Jr.; Karami, A. Computational content analysis of negative tweets for obesity, diet, diabetes, and exercise. *Proc. Assoc. Inf. Sci. Technol.* **2017**, *54*, 357–365. [[CrossRef](#)]
75. Prier, K.W.; Smith, M.S.; Giraud-Carrier, C.; Hanson, C.L. Identifying health-related topics on twitter. In *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 18–25.
76. Di Eugenio, B.; Glass, M. The Kappa Statistic: A Second Look. *Comput. Linguist.* **2004**, *30*, 95–101. [[CrossRef](#)]
77. Lu, Y.; Mei, Q.; Zhai, C. Investigating task performance of probabilistic topic models: An empirical study of PLSA and LDA. *Inf. Retr. J.* **2010**, *14*, 178–203. [[CrossRef](#)]
78. Shaw, G., Jr.; Sharma, T.; Ramakrishnan, S. Exploring Diabetes and Users' lifestyle choices in Twitter to improve health outcomes. In Proceedings of the Southern Association for Information Systems Conference, Cancun, Mexico, 15–17 August 2019.
79. Finfgeld-Connett, D. Twitter and Health Science Research. *West. J. Nurs. Res.* **2014**, *37*, 1269–1283. [[CrossRef](#)] [[PubMed](#)]
80. Barnard, N.D.; Cohen, J.; Jenkins, D.J.; Turner-McGrievy, G.; Gloede, L.; Jaster, B.; Seidl, K.; Green, A.A.; Talpers, S. A Low-Fat Vegan Diet Improves Glycemic Control and Cardiovascular Risk Factors in a Randomized Clinical Trial in Individuals With Type 2 Diabetes. *Diabetes Care* **2006**, *29*, 1777–1783. [[CrossRef](#)] [[PubMed](#)]
81. Jenkins, D.J.A.; Wong, J.M.W.; Kendall, C.W.C.; Esfahani, A.; Ng, V.W.Y.; Leong, T.C.K.; Faulkner, D.A.; Vidgen, E.; Greaves, K.A.; Paul, G.; et al. The Effect of a Plant-Based Low-Carbohydrate ("Eco-Atkins") Diet on Body Weight and Blood Lipid Concentrations in Hyperlipidemic Subjects. *Arch. Intern. Med.* **2009**, *169*, 1046–1054. [[CrossRef](#)] [[PubMed](#)]
82. Tusso, P.; Stoll, S.R.; Li, W.W. A Plant-Based Diet, Atherogenesis, and Coronary Artery Disease Prevention. *Perm. J.* **2015**, *19*, 62–67. [[CrossRef](#)] [[PubMed](#)]

83. Dalen, J.; Smith, B.W.; Shelley, B.M.; Sloan, A.L.; Leahigh, L.; Begay, D. Pilot study: Mindful Eating and Living (MEAL): Weight, eating behavior, and psychological outcomes associated with a mindfulness-based intervention for people with obesity. *Complement. Ther. Med.* **2010**, *18*, 260–264. [[CrossRef](#)]
84. Lloyd-Jones, D.M.; Hong, Y.; Labarthe, D.; Mozaffarian, D.; Appel, L.J.; Van Horn, L.; Greenlund, K.; Daniels, S.; Nichol, G.; Tomaselli, G.F.; et al. Defining and setting national goals for cardiovascular health promotion and disease reduction: The American heart association's strategic impact goal through 2020 and beyond. *Circulation* **2010**, *121*, 586–613. [[CrossRef](#)]
85. American Diabetes Association 2. Classification and Diagnosis of Diabetes. *Diabetes Care* **2016**, *40*, S11–S24. [[CrossRef](#)]
86. Oomen, J.S.; Owen, L.J.; Suggs, L.S. Culture Counts: Why Current Treatment Models Fail Hispanic Women With Type 2 Diabetes. *Diabetes Educ.* **1999**, *25*, 220–225. [[CrossRef](#)]
87. Neiger, B.L.; Thackeray, R.; Van Wagenen, S.A.; Hanson, C.L.; West, J.H.; Barnes, M.D.; Fagen, M.C. Use of social media in health promotion: Purposes, key performance indicators, and evaluation metrics. *Health Promot. Pract.* **2012**, *13*, 159–164. [[CrossRef](#)] [[PubMed](#)]
88. Neiger, B.L.; Thackeray, R.; Burton, S.H.; Thackeray, C.R.; Reese, J.H. Use of Twitter Among Local Health Departments: An Analysis of Information Sharing, Engagement, and Action. *J. Med. Int. Res.* **2013**, *15*, e177. [[CrossRef](#)] [[PubMed](#)]
89. Chou, W.-Y.S.; Oh, A.; Klein, W.M.P. Addressing Health-Related Misinformation on Social Media. *JAMA* **2018**, *320*, 2417–2418. [[CrossRef](#)] [[PubMed](#)]
90. LeBlanc, A.G.; Chaput, J.-P. Pokémon Go: A game changer for the physical inactivity crisis? *Prev. Med.* **2017**, *101*, 235–237. [[CrossRef](#)] [[PubMed](#)]
91. Joseph, B.; Armstrong, D.G. Potential perils of peri-Pokémon perambulation: The dark reality of augmented reality? *Oxf. Med. Case Rep.* **2016**, *2016*, omw080. [[CrossRef](#)] [[PubMed](#)]
92. Nemet, D. Childhood obesity, physical activity, and exercise. *Pediatr. Exerc. Sci.* **2017**, *29*, 60–62. [[CrossRef](#)]
93. Lukic, L.; Lalic, N.M.; Rajkovic, N.; Jotic, A.; Lalic, K.; Milicic, T.; Seferovic, J.P.; Macesic, M.; Gajovic, J.S. Hypertension in Obese Type 2 Diabetes Patients is Associated with Increases in Insulin Resistance and IL-6 Cytokine Levels: Potential Targets for an Efficient Preventive Intervention. *Int. J. Environ. Res. Public Health* **2014**, *11*, 3586–3598. [[CrossRef](#)]
94. Koh-Banerjee, P.; Wang, Y.; Hu, F.B.; Spiegelman, D.; Willett, W.C.; Rimm, E.B. Changes in Body Weight and Body Fat Distribution as Risk Factors for Clinical Diabetes in US Men. *Am. J. Epidemiol.* **2004**, *159*, 1150–1159. [[CrossRef](#)]
95. CDC. Adult Oral Health. 2020. Available online: <https://www.cdc.gov/oralhealth/basics/adult-oral-health/index.html> (accessed on 1 April 2020).
96. Maramba, I.D.; Davey, A.; Elliott, M.N.; Roberts, M.; Roland, M.; Brown, F.; Burt, J.; Boiko, O.; Campbell, J.; Sokolova, M.; et al. Web-Based Textual Analysis of Free-Text Patient Experience Comments From a Survey in Primary Care. *JMIR Med. Inform.* **2015**, *3*, e20. [[CrossRef](#)]
97. Alghamdi, R.; Alfalqi, K. A Survey of Topic Modeling in Text Mining. *Int. J. Adv. Comput. Sci. Appl.* **2015**, *6*, 147–153. [[CrossRef](#)]
98. van Oest, R. A new coefficient of interrater agreement: The challenge of highly unequal category proportions. *Psychol. Methods* **2019**, *24*, 439. [[CrossRef](#)] [[PubMed](#)]
99. Muralidhara, S.; Paul, M.J. #Healthy Selfies: Exploration of Health Topics on Instagram. *JMIR Public Health Surveill.* **2018**, *4*, e10150. [[CrossRef](#)] [[PubMed](#)]