Journal of Health Communication
International Perspectives

ISSN: 1081-0730 (Print) 1087-0415 (Online) Journal homepage: https://www.tandfonline.com/loi/uhcm20

A "fitness" Theme May Mitigate Regional Prevalence of Overweight and Obesity: Evidence from Google Search and Tweets

Bo Liang, Ye WangYE WANG & Ming-Hsiang TsouMING-HSIANG TSOU

To cite this article: Bo Liang, Ye WangYE WANG & Ming-Hsiang TsouMING-HSIANG TSOU (2019): A "fitness" Theme May Mitigate Regional Prevalence of Overweight and Obesity: Evidence from Google Search and Tweets, Journal of Health Communication, DOI: 10.1080/10810730.2019.1657526

To link to this article: https://doi.org/10.1080/10810730.2019.1657526

Published online: 30 Aug 2019.

Journal of Health Communication, 0: 1–10, 2019

ISSN: 1081-0730 print/1087-0415 online
DOI: https://doi.org/10.1080/10810730.2019.1657526

ORIGINAL ARTICLE

A "fitness" Theme May Mitigate Regional Prevalence of Overweight and Obesity: Evidence from Google Search and Tweets

BO LIANG[1], YE WANG[2], and MING-HSIANG TSOU[3]

[1]*Department of Business, Saint Vincent College, Latrobe, PA, USA*
[2]*Department of Communication Studies, University of Missouri-Kansas City, Kansas City, MO, USA*
[3]*Department of Geography, San Diego State University, San Diego, CA, USA*

Taking ecological perspectives to overweight and obesity, the current study applies data mining approach to examine the association between information and social environments and regional prevalence of overweight and obesity. In particular, we focus on online search and social media data since the increasing popularity of location-based geo-targeting could be an influential source of regional differences in health information and social environment. In Study 1, we calculated the correlation between regional overweight and obesity rates with regional Google searches for a time period of 12 years (2004 to 2016). The findings showed that in regions with high overweight and obesity rates, people were looking for and obtaining information on weight-loss and diet;, but in regions with low overweight and obesity rates, people were looking for and obtaining information on fitness services and facilities. In Study 2, we analyzed and compared 4010 tweets from Houston, a city with a high overweight and obesity rate, and 3281 tweets from San Diego, a city with a low overweight and obesity rate. The tweets were collected from August 2015 to August of 2016. We analyzed the textual content of tweets by word frequency analysis and topic modeling. The findings suggest that San Diego has a social environment that focuses on fitness and combining exercising with dieting. In contrast, Houston's social environment emphasizes dieting. The implication of these findings is that health practitioners should push a paradigm shift to a stronger focus on "healthy life" (combining exercising and dieting) in regions with high overweight and obesity rates.

One of the goals of Healthy People 2020 is to "create social and physical environments that promote good health for all" (US Department of Health and Human Services, 2011). However, a challenge facing "good health for all" in the next decade is to combat higher overweight and obesity rates in certain regions in the United States. Specifically, prevalence of overweight and obesity is much higher in the South than in other regions (Bennett, Wolin, & Duncan, 2008; Qvortrup, 2010). Regional prevalence of overweight and obesity requires a comprehensive examination of health behaviors in larger contexts, such as cultural, social, and geographical settings.

From an ecological perspective, weight control behaviors occur in a setting made up of physical as well as information, social, and cultural environments (Ackerson & Viswanath, 2009; Golden & Earp, 2012; Sallis, Owen, & Fisher, 2015). A contemporary approach to public health suggests where people live relates to their attitudes towards diet and exercise – two most effective weight management methods (Miller, 1999; Miller, Koceja, & Hamilton, 1997; Wing, 1999), and subsequently their body weight (Bennett et al., 2008; Qvortrup, 2010).

Address correspondence to Bo Liang, Saint Vincent College, Latrobe, PA. E-mail: bo.liang@stvincent.edu

Following this line of inquiry, empirical studies have looked at the density of food outlets (e.g., fast-food outlets) (De Vogli, Kouvonen, & Gimeno, 2014; Janssen, Davies, Richardson, & Stevenson, 2018), access to physical activity facilities, and walkability of neighborhoods (Matsaganis & Wilkin, 2015; Saelens, Sallis, & Frank, 2003). These studies revealed that easier access to physical activities and higher walkability was associated with more physical activities (Saelens et al., 2003). Meanwhile, easier access to food outlets, especially fast-food outlets, was associated with higher BMI or obesity rate (De Vogli et al., 2014; Janssen et al., 2018). These findings suggest that variance in regional overweight and obesity prevalence is related to the location-specific physical environment.

In addition to physical environment, the contemporary approach to public health also emphasizes information, social, and cultural environment, i.e., health disparities across populations (Marmot, 2005; Reidpath, Burns, Garrard, Mahoney, & Townsend, 2002). Information environment is made up of health information sources, including news, advertisements, interpersonal discussions, health-care professionals, and the Internet (Sallis et al., 2015). Studies on health information suggest variability across populations due to targeting practices of media and advertisers. For example, one study found that there were more local health news stories in newspapers serving Hispanic communities or the national audience than in Black newspapers (Wang & Rodgers, 2013).

With the increasing popularity of geo-targeting techniques, variability across populations has rapidly evolved into regional differences in online information, thus making online data, like Google search and Twitter, a rich source to compare informational environments across regions. IP address-based geographic targeting, GPS-tracking, and geo-tagging methods in smart devices enable businesses to deliver messages to any geographical community that has shown an interest in certain search terms (Poese, Uhlig, Kaafar, Donnet, & Gueye, 2011; Tsou & Lusher, 2015). For example, Google Ads allows messages to be delivered to specific cities, counties, and states based upon users' search interests. If a user is residing in a designated geographic region, the relevance between a search keyword and online content determines what local search results and advertisements this user will see on search engine result pages. Thus, users in regions with greater access to exercise facilities may see more search results and advertisements of local fitness businesses (e.g., a local Planet Fitness club). In contrast, users in regions with greater access to food outlets and particularly fast-food outlets may see more search results and advertisements about food and fast food due to local food businesses' usage of geo-targeting advertising. Accordingly, neighborhood attributes that vary geographically may be reflected as differences in regional information environment.

Web search activities reflect the public's real-time exposure, interests, concerns, and intentions at regional levels (Askitas & Zimmermann, 2009; Battelle, 2005; Ettredge, Gerdes, & Karuga, 2005; Liang & Scammon, 2013; Scheitle, 2011). Geographic mapping of search data has been conducted in the area of public health. For example, Ginsberg et al. (2009) found that regional variations in Google search for flu-related symptoms were highly correlated with regional variations in CDC ILI (influenza-like illness) data.

In addition to search data, another major source of information for the public is social media. Being a platform for sharing health information (Wang, Willis, & Rodgers, 2014), social media also create a sense of "community" that fosters a set of social norms and creates social climate (Preece, Nonnecke, & Andrews, 2004). By interacting with each other, members of online weight-loss groups form a collective understanding of the social meaning of weight-loss behaviors, weight-loss products/brands, and services (Willis & Wang, 2016). Thus, social media reflects not only differences in health information but also variability in social environments across regions.

Language use reflects individuals' social and psychological processes (Gergen, 1995; Pennebaker & King, 1999). Recent research investigated regional patterns of language use by examining geotagged tweets. For example, Ghosh and Guha (2013) identified spatial patterns of obesity-related themes on Twitter. Schwartz et al. (2013) found that the language used in tweets from over 1000 US counties predicted the subjective well-being of people living in those counties.

Accordingly, the current study examines Twitter content in two cities, to understand possible differences in social environment regarding weight loss. This understanding of social environment surrounding weight loss is important to health promotion and education, particularly when recent studies suggest a shift of paradigm from being weight-centered to health-centered (Miller & Jacob, 2001). The weight-centered approach, which often dominates societal discourse, focuses on losing weight as the only way to health (Miller & Jacob, 2001). This paradigm, as argued by Salas (2015), has an unintended consequence of excessive weight preoccupation, which may lead to dieting, even eating disorder or suicide. Under this weight-centered paradigm, weight loss often has short-term effects, and intervention programs tend to have low retention rates (Bacon & Aphramor, 2011). In contrast, the health-centered approach creates a social environment in which health is the main theme of health promotion, without making weight-loss a necessary mediating path to it (Bacon & Aphramor, 2011). While there are doubts about the effectiveness of the health-centered approach, a recent randomized-controlled trial found that the health-centered education program had equally beneficial health outcomes as the weight-centered intervention (Mensinger, Calogero, & Tylka, 2016). This shift of paradigm in health education will lead to fundamental changes in the social setting of weight management. As social media archives public discourse, it allows us to examine the social environment of weight loss in regions of different obesity rates, and thus can shed light on the future directions of health education.

To summarize, the research question of this study is whether there is an association between regional prevalence of obesity and overweight and regional information and social environments. Accordingly, we conducted two studies to achieve our objective. In Study 1, we investigated the relationship between state-wise obesity and overweight rates and Google search volumes. We used Google search data at state-level since state-level obesity and overweight rates and search volumes are more inclusive and accurate than metropolitan or city-level data. CDC provides annual BMI (Body Mass Index) data of all states but not all metropolitan statistical areas (MMSAs) in the United States. Google provides search volume data by states, metros, and cities, but its metro categorization is different from CDC's MMSAs.

In Study 2, we compared tweets from two cities instead of using state-level data. There is a large amount of tweets about weight loss and fitness even at the city level. Due to our limited data processing capacity, we were not able to process statewide or across-state tweets data. Considering similarities in demographics (e.g., income, education, population) and ranking data of fattest and fittest cities by Men's Fitness Magazine, we selected San Diego and Houston in Study 2.

Study 1: Google Search and Obesity and Overweight Rate

We used Google Trends as our major source for online search data. Google is the most dominant search engine, serving the largest percentage of queries 65% in the US (Comscore, 2015). In Study

1, we collected state-level data of obesity and overweight rates and Google search interest in weight loss-related topics.

Data Collection

We acquired data of 2004–2016 obesity and overweight rates by state from the Behavioral Risk Factor Surveillance System (BRFSS), a state-based health surveillance system that tracks behavioral risk factors. The prevalence of obesity and overweight varied across the states, with West Virginia having the highest prevalence (71.1%) and District of Columbia the lowest prevalence (54.4%) in 2016.

Following the methods used in previous studies (Baram-Tsabari, 2011; Breyer & Eisenberg, 2010; Vosen & Schmidt, 2011), we used Google Trends to collect regional search data. Google Trends provides longitudinal search interest indexes that quantify search volumes as values ranging from 0 to 100 in relation to the peak volume for the given region and time (see trends.google.com). The highest volume during the reporting period has a value of 100, and a value of 50 means the search volume drops by half (see trends.google.com). Zero means that there was not enough data to calculate the search interest index (see trends.google.com).

Google Trends provides search volume data for search terms and search topics. A search topic is a group of terms that share the same concept. For example, all search terms relating to “exercise” constitute the search topic of “exercise”. To have a comprehensive list of search terms, we employed a broad-match search-topic strategy. Based upon our research question, we chose the following search topics: weight loss, diet, physical exercise, gym, and physical fitness. We acquired search data on the five search topics for 51 regions (50 states + Columbia district) in the United States from 1/1 2004 to 12/31/2016.

Analysis & Results

We used Pearson’s Correlation Coefficient to find the relationship between average obesity and overweight rate and relative search volume for weight loss-related topics by state. We present results for Pearson’s Correlation analysis in Table 1.

Regional obesity and overweight rate was strongly and positively correlated with search volume for the topics of weight loss ($r = 0.729$, $p = .000$) and diet ($r = 0.603$, $p = .000$). In other words, in high obesity and overweight regions, people had more searches for weight loss and diet. Regional obesity and overweight rate was moderately and negatively correlated with search volume for the topics of gym ($r = - 0.556$, $p = .000$) and physical fitness ($r = - 0.403$, $p = .002$). In lower obesity and overweight regions, people had more searches for gym and physical fitness. Regional obesity and overweight rate was not significantly correlated with search volume for the topic of physical exercise. To sum up, Study 1 showed that in regions with high overweight and obesity rates, people were looking for and obtaining information on weight-loss and diet;, but in regions with low overweight and obesity rates, people were looking for and obtaining information on fitness services and facilities.

Table 1. Pearson’s correlation of obesity and overweight rate and google search volume by state

	1	2	3	4	5	6
1. Obesity & Overweight Rate	1					
2. Search: Weight Loss	.729*	1				
3. Search: Diet	.603*	.648*	1			
4. Search: Physical Exercise	0.263	.446*	0.455*	1		
5. Search: Gym	−.556*	−.384*	−.373*	−0.12	1	
6. Search: Physical Fitness	−.403*	−.416*	−.382*	−0.11	.361*	1

* Correlation is significant at the 0.01 level (2-tailed).

Study 2: Tweets and Obesity and Overweight Rate

In Study 2, we conducted a comparative case study of two cities: Houston and San Diego. We intended to select two American cities with significantly different obesity and overweight rates but comparable population and median household income. Our selection was based upon the ranking system of fattest and fittest cities by Men’s Fitness Magazine, which measures BMI, behavioral (e.g., physical activities), and environmental (e.g., hours spent in front of the television, levels of fruit and vegetable consumption) factors (Koch, 2014, 2015). This ranking system has been used by researchers in the fields of community health, urban environment, and geography (e.g., Evans, Crookes, & Coaffee, 2012; Roberts, Won, Vasudevan, Ko, & Guileyardo, 2016). This ranking system is published every year since 2002. Houston was ranked among the top 10 fattest cities in most years and was the fattest city in the years of 2002, 2003, 2005, and 2012. In contrast, San Diego was ranked among the top 10 fittest cities in most years and was the third fittest city in 2002.

According to CDC, the year of 2012 is the most recent year for which BMI data is available for both Houston and San Diego MMSAs (BRFSS, 2012). In this year, obesity and overweight rate for Houston MMSA was 64.3%, and San Diego MMSA, 59% (BRFSS, 2012). Based on z-test, the difference of the two proportions is significant ($z = -3.4167$, $p = .000634$). The CDC data were line with the ranking data by Men’s Fitness Magazine. Houston was the fattest city, and San Diego was the eighth fittest city in 2012 (Millado, 2012).

In terms of population and median household income, both cities are among the 10 most populous metropolitan areas, with similar median household income ranging from 60,000 to 69,000 (Census, 2017). According to US Census Bureau (2016), Houston had a population of 2.3 million, and San Diego had a population of 1.4 million.

More than 23% of online adults use Twitter (Duggan, 2015). This information-rich media has been examined as a source for health information in various studies (Ghosh & Guha, 2013; Scanfeld, Scanfeld, & Larson, 2010). Based on previous studies on tweets analysis (Ghosh & Guha, 2013; Graham & Ackland, 2015; Prier, Smith, Giraud-Carrier, &

Hanson, 2011), we took two major steps to explore tweets in this study: (1) collecting tweets, and (2) text mining.

Data Collection

We collected geocoded tweet data with the help of The Center of Human Dynamics In the Mobile Age (HDMA), a research institution at San Diego State University. The HDMA developed a SMART dashboard (Social Media Analytics and Research Testbed) to collect Twitter data with user-defined keywords and geolocations using Twitter Search APIs. SMART dashboard is an online geo-targeted search and analytics tool, including an automatic data processing procedure to help researchers to 1) search tweets in different cities; 2) filter noise; 3) analyze social media data from a spatiotemporal perspective, and 4) visualize social media data (Tsou et al., 2015).

We started with popular search terms discovered in Study 1, but these terms generated a large amount of tweets. Due to our limited capacity for data collection, we had to narrow down the scope of our research by using AND Boolean search strategies. Specifically, we used two and three-word search queries, including: "lost pound(s)" and "losing weight" under the topic of weight loss, and "dieting" and "on a diet" under the topic of "diet." All these queries are status-related terms indicating individuals' engagement in weight loss or dieting. As the volume of tweets on the topic of gym/fitness was extremely large, we had to use nine AND Boolean broad-match search queries that combine concepts of gym/fitness and exercises. Table 2 shows the search topics (named as "Theme Categories" in the table) and queries for Tweet collection. Based on the selected queries, we collected a total of 14,571 tweets by Houston users, and 27,207 tweets by San Diego users from August of 2015 to August of 2016.

Table 2. Tweet data: Terms and matching options

Theme Categories	Queries	Matching Option
Weight loss	losing weight	broad match
	lost pound(s)	broad match
Diet	dieting	broad match
	on a diet	exact match
Gym/fitness	gym fit	broad match
	gym fitness	broad match
	gym workout	broad match
	gym exercise	broad match
	fit workout	broad match
	fit exercise	broad match
	fitness workout	broad match
	Fitness exercise	broad match

1. broad match: tweets containing all the words in a term in any order; any word is allowed to be added to the middle of the term.
2. exact match: tweets containing all the words in a term in the same order; no word is allowed to be added to the middle of the term.

Text Mining

We analyzed the textual content of tweets using tm package in the R environment (Feinerer 2012; Meyer, Hornik, & Feinerer, 2008). Text mining is conceptually similar to thematic analysis, which summarizes the most important themes in content. To prepare the dataset for topic modeling, we pre-processed the text by two steps. First, we removed retweets and duplicated tweets based on pure text match. This process resulted in a dataset of 3281 tweets from users in San Diego, and 4010 tweets from users in Houston. Then, we cleaned the text by removing URLs (words starting with 'http://'), HTML entities (e.g., """), punctuation characters, number characters, stop words (i.e., common words such as "the" "I"), and converting words to lowercase. We then examined the frequency of occurrences of terms used in tweets. We made a list of terms with a frequency of occurrences more than 15.

Our second step was to generate topic models with the latent dirichlet allocation (LDA) approach to discover hidden topics for each of the three theme categories: weight loss, diet, and gym and fit/fitness. Topic modeling is a powerful tool to algorithmically identify latent text patterns in documents (Blei & La□erty, 2009). LDA is an unsupervised machine learning technique used to identify latent topic information in large document collections. LDA uses a "bag of words" model, with each document representing as a bag of its words. Each topic is represented as a probability distribution over a number of words. Topic modeling with the LDA approach has been applied widely in studies on Twitters (e.g., Ghosh & Guha, 2013; Ramage, Dumais, & Liebling, 2010). For the current study, we used the "topicmodels" package in R (Hornik & Grün , 2011). We specified the number of topics desired as 20, and the number of words assigned to each topic as eight.

Our final step was to analyze the results (words in each topic) of topic modeling for each theme category. Within each theme category, we grouped topics into themes based on the words contained in each topic. We explored commonality and variation in themes between the two cities by analyzing the relative frequency of each word contained in each theme. We calculated the relative frequency of each word as the ratio of the frequency of each word and the total frequency of all words within each theme category. We conducted z-test to determine differences in the relative frequency of words between San Diego and Houston Twitter users. The themes and related statistics were reported below.

The Theme Category of Weight Loss

In the theme category of weight loss, there were 1,082 tweets from San Diego, and 1401 tweets from Houston. We grouped 20 topics identified by topic modeling into seven themes: progress update, dieting, promoted information, positive thoughts/feelings, fat, difficulty, and exercise. Table 3 presents these themes, the words with high frequency (at least 15 occurrences for one of the two cities), and z-test statistics on comparison of relative word frequencies between the two cities.

Table 3. The theme category of weight loss: Words with high frequency

	San Diego			Houston		
Themes	Words	Absolute Frequency	Relative Frequency (%)	Words	Absolute Frequency	Relative Frequency (%)
Progress update	week	72	10.93	week	99	14.89
		Comparison (San Diego vs. Houston	*Z test stat*	−2.1482	*p value*	0.03*
	day	62	9.40	day	64	9.62
		Comparison (San Diego vs. Houston	*Z test stat*	−0.124	*p value*	0.90
	month	35	5.15	month	61	9.17
		Comparison (San Diego vs. Houston	*Z test stat*	−2.7852	*p value*	0.01*
	year	16	2.35	year	19	2.86
		Comparison (San Diego vs. Houston	*Z test stat*	−0.5217	*p value*	0.60
	Total	185	27.83	*Total*	243	36.54
		Comparison (San Diego vs. Houston	*Z test stat*	−3.388	*p value*	0.001*
Eat	eat	84	12.75	eat	86	12.93
	diet	41	6.22	diet	29	4.36
	food	36	5.46	food	20	3.01
	meal plans	19	2.88	Meal plans	3	0.45
	Total	180	27.31	*Total*	138	20.75
		Comparison (San Diego vs. Houston	*Z test stat*	2.79376	*p value*	0.005*
Positive feelings	good	30	4.55	better	29	4.36
	best	28	4.25	good	52	7.82
	great	16	2.43	great	15	2.26
	better	16	2.43	best	18	2.71
	happy	15	2.28	happy	18	2.71
	Total	105	15.93	*Total*	132	19.85
		Comparison (San Diego vs. Houston	*Z test stat*	−1.860	*p value*	0.063
Promoted information	reasons	50	7.59	ways	19	2.86
	tips	34	5.16	help	19	2.86
	ways	16	2.43	tips	17	2.56
	things	22	3.34	reasons	10	1.50
	help	16	2.43	things	8	1.20
	Total	138	20.94	Total	73	10.98
		Comparison (San Diego vs. Houston	*Z test stat*	4.931	*p value*	0.000*
Fat	fat	20	3.03	fat	30	4.51
		Comparison (San Diego vs. Houston	*Z test stat*	−1.432	*p value*	0.152
Difficulty	hard	9	1.37	hard	38	5.71
		Comparison (San Diego vs. Houston	*Z test stat*	−4.228	*p value*	0.000*
Exercise	exercise	22	3.34	exercise	11	1.65
		Comparison (San Diego vs. Houston	*Z test stat*	1.926	*p value*	0.040*

Progress update was the most popular theme for both cities. Users had a short-term frame of mind for weight loss. The word "week" had the highest frequency for this theme across the two cities. Users preferred to report their weight-loss progress weekly. The word "day" was the second most frequently used word across the two cities. Eating and expressing positive feelings were also popular themes. Users shared their positive feelings about their weight loss experience. Frequently used words on this theme included: good, great, better, best, and happy.

Twitter is a platform used not only by regular users to share their daily life, but also by bloggers, professionals, and businesses to disseminate information. We found tweets suggesting easy weight loss tips. We named this theme as promoted information.

Three themes emerge but are not as popular as the themes mentioned above, including: fat, difficulty, and exercise. Users were aware of two facts about fat: (1) fat is part of their weight, (2) fat is different from muscle. Users felt that losing weight was hard, especially when they adopted dieting. Users often mentioned exercise and weight loss simultaneously.

Based on z-test, themes of eat, promoted information, and exercises were more popular in San Diego, whereas themes of progress update and positive feelings were more popular in Houston ($p < .05$). Regarding progress update, Houston users mentioned “week” and “month” more frequently than did San Diego users ($p < .05$).

The Theme Category of Diet

In the theme category of diet, there are 653 tweets from San Diego, and 1164 tweets from Houston. Houston had a significantly higher proportion of diet-related tweets among all the tweets collected (30% of total tweets) than San Diego (22% of total tweets) had ($z = 8.961$, $p < .05$). We grouped 20 topics identified by topic modeling into six themes: burn fat quickly without dieting or exercise, weight loss, food, short term, flexible dieting, and difficulty. Table 4 presents these themes and the words with high frequency (at least 15 occurrences for one of the two cities), and z-test statistics on comparison of relative word frequencies between the two cities.

The most popular theme over the time period while data were collected for this study was: burn fat quick without dieting or exercise. This was a promoted message distributed widely via Twitter on September 29, 2015. We could not find the original article about this weight loss strategy because the links on Twitter expired at the time when data were collected. The next two most popular themes were weight loss and food. Regarding food, users talked about their feelings, especially negative feelings for dieting. Users also talked about the food that they loved but had to avoid such as cake and pizza.

Three themes emerged but were not as popular as the themes mentioned above, including: short term, flexible dieting, and difficulty. Users had a short-term frame of mind for dieting. They used words showing short periods of time in their Twitter updates, including: “day” and “week”. Flexible dieting was a currently popular dieting strategy shared by users. This dieting strategy focuses on the counting and tracking of macronutrients (protein, carbohydrates, and fat) to achieve a body composition goal. Like weight loss, most users felt that dieting was hard work.

Based on z-test, themes of weight loss, short term, and flexible dieting were more popular in San Diego, whereas the promoted message, burn fat quickly without dieting or exercise, was a more popular theme in Houston ($p < .01$). Regarding a short-term time frame, San Diego users mentioned “week” more frequently than did Houston users ($p < .01$).

Table 4. The theme category of diet: Words with high frequency

	San Diego			Houston		
Themes	Words	Absolute Frequency	Relative Frequency (%)	Words	Absolute Frequency	Relative Frequency (%)
Burn Fat Quick Without Dieting or Exercise	fat	145	12.56	fat	336	16.48
	exercise	85	7.37	exercise	303	14.86
	without	85	7.37	without	298	14.62
	quick	69	5.98	quick	289	14.17
	burn	67	5.81	burn	287	14.08
	Total	451	39.08	*Total*	1513	74.20
		Comparison (San Diego vs. Houston	*Z test stat*	−24.102	*p value*	0.000*
Weight loss	weight	181	15.68	weight	61	2.99
	loss	114	9.88	lose	37	1.81
	lose	96	8.32	loss	13	0.64
	Total	391	33.88	*Total*	111	5.44
		Comparison (San Diego vs. Houston	*Z test stat*	31.900	*p value*	0.000*
Food	food	60	5.20	food	99	4.86
	hate	16	1.39	hate	34	1.67
	cake	27	2.34	cake	34	1.67
	pizza	24	2.08	pizza	41	2.01
	Total	127	11.01	*Total*	208	10.20
		Comparison (San Diego vs. Houston	*Z test stat*	1.079	*p value*	0.281
Short-term	day	51	4.42	day	84	4.12
		Comparison (San Diego vs. Houston	*Z test stat*	0.405	*p value*	0.686
	week	45	3.90	week	43	2.11
		Comparison (San Diego vs. Houston	Z test stat	−28.521	p value	0.000*
	Total	96	8.32	*Total*	127	6.91
		Comparison (San Diego vs. Houston	*Z test stat*	3.400	*p value*	0.001*
Flexible dieting	flexible	58	5.03	flexible	34	1.67
		Comparison (San Diego vs. Houston	*Z test stat*	8.364	*p value*	0.000*
Difficulty	hard	31	2.69	hard	46	2.26
		Comparison (San Diego vs. Houston	*Z test stat*	1.179	*p value*	0.239

The Theme Category of Gym/Fitness

For the theme category of gym/fitness, there were 1310 tweets from San Diego, and 1287 from Houston. San Diego had a much higher proportion of gym/fitness-related tweets (43% of total tweets) than Houston (34% of total tweets) ($z = 6.948$, $p < .001$). We grouped 20 topics identified by topic modeling into eight themes: time, new experience, positive feeling, types of exercises, dieting, calorie burned, working hard, and health. Table 5 presents these themes and the words with high frequency (at least 15 occurrences for one of the two regions), and z-test statistics on comparison of relative word frequencies between the two cities.

Time was the most popular theme across the two cities. Frequently used words included: time, minute, and day. Users had a fixed time for exercise at gym. They liked programs with shorter duration and easiness. Users often reported about their daily gym activities.

Themes of new experience, positive feelings, and types of exercises were also popular. Users were motivated to go to gym by new experiences, such as new exercises and new fitness programs. Users demonstrated positive feelings for going to gyms. The frequently used words included good, fun, and love. Tweets on types of exercises were mainly used for commercial promotion. The three frequently promoted exercises were: cardio, crossfit, and dance.

Four themes emerge but are not as popular as the themes mentioned above are: dieting, calorie burned, working hard, and health. Gym goers cared about food and the number of calories burned in a certain amount of minutes while doing exercises. They recognized that hard work was paid off. Users mentioned the word "health" in their gym-related tweets, but not frequently.

According to z-test, new experience and types of exercises were more popular themes in San Diego, whereas calorie burned was a more popular theme in Houston ($p < .001$). Regarding the theme of time, San Diego users mentioned "time" and "morning" more frequently, whereas Houston users mentioned "day" and "minute" more frequently ($p < .001$).

Table 5. The theme category of gym/fitness: Words with high frequency

	San Diego			Houston		
Themes	Term	Absolute Frequency	Relative Frequency %	Term	Absolute Frequency	Relative Frequency %
Time	time	75	13.02	time	22	4.77
		Comparison (San Diego vs. Houston	*Z test stat*	4.561	*p value*	0.000*
	day	56	9.72	day	72	15.62
		Comparison (San Diego vs. Houston	*Z test stat*	−2.880	*p value*	0.000*
	minute	46	7.99	minute	64	13.88
		Comparison (San Diego vs. Houston	*Z test stat*	−3.061	*p value*	0.000*
	morning	45	7.81	morning	4	0.87
		Comparison (San Diego vs. Houston	*Z test stat*	−4.767	*p value*	0.000*
	Total	222	38.54	*Total*	162	35.14
		Comparison (San Diego vs. Houston	*Z test stat*	1.127	*p value*	0.260
New experience	new	62	10.76	new	23	4.99
		Comparison (San Diego vs. Houston	*Z test stat*	3.366	*p value*	0.000*
Positive feelings	good	44	7.64	love	43	9.33
	fun	28	4.86	good	32	6.94
	love	35	6.08	fun	11	2.39
	Total	107	18.58	*Total*	86	18.66
		Comparison (San Diego vs. Houston	*Z test stat*	−0.033	*p value*	0.974
Types of exercises	cardio	42	7.29	cardio	13	2.82
	dance	33	5.73	dance	35	7.59
	crossfit	15	2.60	crossfit	4	0.87
	Total	90	15.63	*Total*	52	11.28
		Comparison (San Diego vs. Houston	*Z test stat*	2.025	*p value*	0.042*
Dieting	eat	36	6.25	eat	34	7.38
		Comparison (San Diego vs. Houston	*Z test stat*	−0.721	*p value*	0.47
Calorie burned	burn	11	1.91	burn	34	7.38
	calorie	13	2.26	calorie	42	9.11
	Total	24	4.17	*Total*	76	16.49
		Comparison (San Diego vs. Houston	*Z test stat*	−6.679	*p value*	0.000*
Working hard	hard	19	3.30	hard	13	2.82
		Comparison (San Diego vs. Houston	*Z test stat*	0.444	*p value*	0.056
Health	healthy	16	2.78	healthy	15	3.25
		Comparison (San Diego vs. Houston	*Z test stat*	−0.442	*p value*	0.059

Discussion

The two studies examine regional differences in online information and social environments. The Google Search study showed that regions with low overweight and obesity rates had more searches for gym and physical fitness but fewer searches for weight loss and diet, than regions with high obesity and overweight rates. The Twitter study found that San Diego users had a higher proportion of gym/fitness-related tweets and lower proportion of diet-related tweets than did Houston users. Thus, results from Google Search and Twitter combined seem to suggest a fitness-focused social environment in regions with low overweight and obesity rates.

This focus on fitness resonates with the recent shift of paradigm from weight to health that has caught the attention of health-care practitioners and researchers. Previous studies showed that people with a focus on weight tend to set up a cultural "ideal" body weight, and diet to lose weight (Bacon & Aphramor, 2011; Kruger, Galuska, Serdula, & Jones, 2004; O'Hara & Gregg, 2006). In contrast, people with a focus on health and well-being tend to live a healthy life and are not concerned about their body weight (Bacon & Aphramor, 2011; Kruger et al., 2004; O'Hara & Gregg, 2006). Twitter data from our study illustrate regional differences between social environments that emphasize "fitness" versus "weight loss," and suggest a paradigm shift from weight to health in regions with higher overweight and obesity rates.

First, the most popular topic among San Diego users regarding the gym/fitness theme was "time." San Diego users talked about their capability of managing time. They had regular gym time (especially in the morning). Second, new fitness programs or exercises excited San Diego users and motivated them to be in shape. Third, San Diego users attached more positive emotions regarding gym/fitness than did Houston users while they recognized hard work of physical exercises. In contrast, Houston users cared more about the number of calories burned. Fourth, people in San Diego have a different view of the relationship between weight loss, exercising, and dieting from people in Houston. Twitter data showed an association between mentioning of weight loss, dieting, and exercising in San Diego, suggesting that in the local social environment these three concepts are often linked together as an integrated whole. The "fitness" theme encourages people to take a holistic approach that supplements exercise with dieting to effectively control weight in a sustainable way. Combing exercise with dieting to lose weight is recommended as a more effective way to lose weight than focusing on a single strategy (Wing & Phelan, 2005). When the focus is on health, the experience with weight control is often emotionally positive. In contrast, such an association was not found in tweets in Houston, suggesting that in the local social environment dieting and exercising are viewed as two separate strategies to manage weight. Meanwhile, people tended to link weight-loss behaviors with burning calories and dieting, a single-strategy approach. As the frustration and difficulty of dieting take hold, people in regions of high overweight and obesity rates are likely to portrait weight loss as a drudgery journey or even daunting tasks.

These psychological and health benefits of a fitness-focused social environment call for a paradigm shift, and health educator should be leading this effort, to promote holistic approaches that combine exercising with dieting; encourage more exercise time than calories burnt; and, subsequently create positive sentiments to sustain long-term weight control and health outcomes in high obesity regions.

Additionally, our tweet analysis revealed themes on promoted information such as "burn fat quick without dieting or exercise". Misinformation like this needs to be addressed by public health communication in a timely manner. And, recent research on Twitter started to focus on Twitter bots and misinformation (Martinez, Tsou, & Spitzberg, 2019). Health practitioners should employ social media mining for public health surveillance and provide timely responses to the public, to combat misinformation.

Conclusion

The ultimate goal of combating overweight and obesity is to "create social and physical environments that promote good health for all." (US Department of Health and Human Services, 2011). Accordingly, the society needs to create an ecological system with physical, informational, and social environments that facilitate health. Large amounts of geocoded search and social media data allow us to have a fuller picture of information and social environments of geographical regions, which consist of not only Internet activities of individuals with various health conditions and motivations, but also information, promotions, and advocacies by businesses, health-care organizations, and government agencies. Our findings suggest when a region nurtures information and social environments focusing on being fit and healthy rather than calories burned or consumed, people in this region tend to take a holistic approach to weight management, and are often emotionally positive. A fitness/health-centered informational and social environment is beneficial for the closing and reducing of overweight and obesity rates across populations and regions. Ultimately, weight management should be serving the goal of "good health for all."

Limitations and Future Studies

There are limitations to the current approach. First, there was an age bias in our sampling. The age demographics of Twitter are geared toward the youth, nearly 30% of Twitter users aged 18–29 averaging and 25% of users aged 30–49 years old (Aslam, 2018). Second, machine learning allows us to analyze nearly 10,000 tweets, a much larger sample size than what manual analysis can handle. However, limited computation capacity restricted our examination to only two cities. Third, the keyword approach using Google Trend data might be affected by the "Google search suggestions", which was a major problem in the Google Flu Trend prediction case study (Lazer, Kennedy, King, & Vespignani, 2014).

Fourth, in the Pearson correlation analysis, each state has the same "weight" in the correlation, but in real world, California has much more people comparing to other states. However, Google Trends data are relative measures, and population size is absolute data. Without knowing the actual number of searches, we could not apply more advanced modeling to assign different weights to states based upon population size. Future studies may adopt more advanced spatial statistics, such as GWR (geographically weighted regression) if absolute search data are available. Last, the 12-year search data from 1/1 2004 to 12/30/2016 in Study 1 showed that the search volume for all weight-loss related search topics peaked during the first week of January across all states. While time patterns were not the focus of this study, these seasonable peeks could be explained by the new-year resolution effect. Future studies could explore the time-related analysis of search and social media data.

References

Ackerson, L. K., & Viswanath, K. (2009). The social context of interpersonal communication and health. *Journal of Health Communication*, *14* (S1), 5–17. doi:10.1080/10810730902806836

Askitas, N., & Zimmermann, K. F. (2009). Google econometrics and unemployment forecasting. *Applied Economics Quarterly*, *55*(2), 107–120. doi:10.3790/aeq.55.2.107

Aslam, S. (2018). *Twitter by the numbers: Stats, demographics & fun facts*. Retrieved from https://www.omnicoreagency.com/twitter-statistics

Bacon, L., & Aphramor, L. (2011). Weight science: evaluating the evidence for a paradigm shift. *Nutrition Journal*, *10*(1), 9. doi:10.1186/1475-2891-10-9

Baram-Tsabari, A. (2011). Exploring new web-based tools to identify public interest in science. *Understanding of Science*, *20*(1), 130–143. doi:10.1177/0963662509346496

Battelle, J. (2005). The birth of Google. *Wired-san Francisco-*, *13*(8), 102.

Bennett, G. G., Wolin, K. Y., & Duncan, D. T. (2008). Social determinants of obesity. In F. Hu (Ed.), *Obesity Epidemiology* (pp. 342–376). New York, NY: Oxford University Press.

Blei, D. M., & Lafferty, J. D. (2009). Topic models. In A. Srivastava & M. Sahami (Eds.), *Text mining: Theory and applications* (pp. 101–124). Text Mining: Classification, Clustering, and Applications. Boca Raton, Florida: Chapman & Hall/CRC Press.

Breyer, B. N., & Eisenberg, M. L. (2010). Use of Google in study of noninfectious medical conditions. *Epidemiology*, *21*(4), 584. doi:10.1097/EDE.0b013e3181e9edc4

Census. (2017). Media household income of 25 most populous metropolitan areas. Retrieved from https://www.census.gov/content/dam/Census/library/publications/2017/acs/acsbr16-02.pdf

Comscore. (2015). Comscore releases January 2015 U.S. desktop search engine rankings. Retrieved from https://www.comscore.com/Insights/Market-Rankings/comScore-Releases-January-2015-US-Desktop-Search-Engine-Rankings

De Vogli, R., Kouvonen, A., & Gimeno, D. (2014). The influence of market deregulation on fast food consumption and body mass index: A cross-national time series analysis. *Bulletin of the World Health Organization*, *92*, 99–107A. doi:10.2471/BLT.13.120287

Duggan, M. (2015). Mobile messaging and social media 2015. *Pew Research Center*, *19*, 2015.

Ettredge, M., Gerdes, J., & Karuga, G. (2005). Using web-based search data to predict macroeconomic statistics. *Communications of the ACM*, *48*(11), 87–92. doi:10.1145/1096000

Evans, B., Crookes, L., & Coaffee, J. (2012). Obesity/fatness and the city: Critical urban geographies. *Geography Compass*, *6*(2), 100–110. doi:10.1111/geco.2012.6.issue-2

Feinerer I. (2012). Text Mining Package. R package version 0.5-7.1. 2012. Retrieved from http://CRAN. R-project, org/package= tm.

Gergen, K. J. (1995). Social construction and the educational process. In L. P. Steffe & J. Gale (Eds.), *Constructivism in education* (pp. 17–39). Hillsdale, New Jersey: Lawrence Erlbaum.

Ghosh, D., & Guha, R. (2013). What are we 'tweeting'about obesity? Mapping tweets with topic modeling and geographic information system. *Cartography and Geographic Information Science*, *40*(2), 90–102. doi:10.1080/15230406.2013.776210

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, *457*(7232), 1012. doi:10.1038/nature07634

Golden, S. D., & Earp, J. A. L. (2012). Social ecological approaches to individuals and their contexts: Twenty years of health education & behavior health promotion interventions. *Health Education & Behavior*, *39*(3), 364–372. doi:10.1177/1090198111418634

Graham, T., & Ackland, R. (2015). Topic modeling of tweets in R: A tutorial and methodology.

Hornik, K., & Grün, B. (2011). Topicmodels: An R package for fitting topic models. *Journal of Statistical Software*, *40*(13), 1–30.

Janssen, H. G., Davies, I. G., Richardson, L. D., & Stevenson, L. (2018). Determinants of takeaway and fast food consumption: A narrative review. *Nutrition Research Reviews*, *31*(1), 16–34. doi:10.1017/S0954422417000178

Koch, P. (2014). The fittest & fattest cities in America: We gave America a physical – and looked at much more than just the vitals – then made a house call to the fittest city to learn its secret. *Men's Fitness*, Apr. 82.

Koch, P. (2015). The city that lost a million pounds: The fittest and fattest cities in America. *Men's Fitness*, Apr, 80.

Kruger, J., Galuska, D. A., Serdula, M. K., & Jones, D. A. (2004). Attempting to lose weight: Specific practices among US adults. *American Journal of Preventive Medicine*, *26*(5), 402–406. doi:10.1016/j.amepre.2004.02.001

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: Traps in big data analysis. *Science*, *343*(6176), 1203–1205. doi:10.1126/science.1248506

Liang, B., & Scammon, D. L. (2013). Incidence of online health information search: A useful proxy for public health risk perception. *Journal of Medical Internet Research*, *15*(6), e114. doi:10.2196/jmir.2401

Marmot, M. (2005). Social determinants of health inequalities. *The Lancet*, *365*(9464), 1099–1104. doi:10.1016/S0140-6736(05)74234-3

Martinez, L. S., Tsou, M., & Spitzberg, B. H. (2019). Belief surveillance, sentiment analysis, and identification of informational targets for E-cigarettes interventions. *Proceedings of the 2019 International Conference on Social Media & Society*, *Forthcoming, Toronto, Canada*.

Matsaganis, M. D., & Wilkin, H. A. (2015). Communicative social capital and collective efficacy as determinants of access to health-enhancing resources in residential communities. *Journal of Health Communication*, *20*(4), 377–386. doi:10.1080/10810730.2014.927037

Mensinger, J. L., Calogero, R. M., & Tylka, T. L. (2016). Internalized weight stigma moderates eating behavior outcomes in women with high BMI participating in a healthy living program. *Appetite*, *102*, 32–43. doi:10.1016/j.appet.2016.01.033

Meyer, D., Hornik, K., & Feinerer, I. (2008). Text mining infrastructure in R. *Journal of Statistical Software*, *25*(5), 1–54.

Millado, N. (2012). Fittest & fattest cities in America. *Men's Fitness*, Mar 2012, 28-3, 32–57.

Miller, W. C. (1999). How effective are traditional dietary and exercise interventions for weight loss? *Medicine and Science in Sports and Exercise*, *31*(8), 1129–1134.

Miller, W. C., & Jacob, A. V. (2001). The health at any size paradigm for obesity treatment: The scientific evidence. *Obesity Reviews*, *2*(1), 37–45.

Miller, W. C., Koceja, D. M., & Hamilton, E. J. (1997). A meta-analysis of the past 25 years of weight loss research using diet, exercise or diet plus exercise intervention. *International Journal of Obesity*, *21*(10), 941. doi:10.1038/sj.ijo.0800499

O'Hara, L., & Gregg, J. (2006). The war on obesity: A social determinant of health. *Health Promotion Journal of Australia*, *17*(3), 260–263.

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology*, *77*(6), 1296. doi:10.1037/0022-3514.77.6.1296

Poese, I., Uhlig, S., Kaafar, M. A., Donnet, B., & Gueye, B. (2011). IP geolocation databases: Unreliable? *ACM SIGCOMM Computer Communication Review*, *41*(2), 53–56. doi:10.1145/1971162

Preece, J., Nonnecke, B., & Andrews, D. (2004). The top five reasons for lurking: Improving community experiences for everyone. *Computers in Human Behavior*, *20*(2), 201–223. doi:10.1016/j.chb.2003.10.015

Prier, K. W., Smith, M. S., Giraud-Carrier, C., & Hanson, C. L. (2011). Identifying health-related topics on twitter. In Salerno, J., Yang, S. J., Nau, D., Chai, S.-K. (Eds.,) *International conference on social computing, behavioral-cultural modeling, and prediction* (pp. 18–25). Berlin, Heidelberg: Springer.

Qvortrup, M. (2010). Social determinants of obesity. In Tsichlia, G., Johnstone, A. (eds.,) *Fat Matters: From Sociology to Science* (pp. 13–21). Cumbria, CA: M & K Publishing.

Ramage, D., Dumais, S. T., & Liebling, D. J. (2010). Characterizing microblogs with topic models. *ICWSM*, *10*(1), 16.

Reidpath, D. D., Burns, C., Garrard, J., Mahoney, M., & Townsend, M. (2002). An ecological study of the relationship between social and environmental determinants of obesity. *Health & Place*, *8*(2), 141–145. doi:10.1016/S1353-8292(01)00028-4

Roberts, W. C., Won, V. S., Vasudevan, A., Ko, J. M., & Guileyardo, J. M. (2016). Frequency of massive cardiac adiposity (floating heart) at necropsy and comparison of clinical and morphologic variables with cases with nonmassive cardiac adiposity at a single Texas hospital, 2013 to 2014. *The American Journal of Cardiology*, *117*(6), 1006–1013. doi:10.1016/j.amjcard.2015.12.039

Saelens, B. E., Sallis, J. F., & Frank, L. D. (2003). Environmental correlates of walking and cycling: findings from the transportation, urban design, and planning literatures. *Annals of Behavioral Medicine*, *25*(2), 80–91. doi:10.1207/S15324796ABM2502_03

Salas, X. R. (2015). The ineffectiveness and unintended consequences of the public health war on obesity. *Canadian Journal of Public Health*, *106*(2), e79–e81. doi:10.17269/cjph.106.4757

Sallis, J. F., Owen, N., & Fisher, E. (2015). Ecological models of health behavior. *Health Behavior: Theory, Research, and Practice*, *5*, 43–64.

Scanfeld, D., Scanfeld, V., & Larson, E. L. (2010). Dissemination of health information through social networks: Twitter and antibiotics. *American Journal of Infection Control*, *38*(3), 182–188. doi:10.1016/j.ajic.2009.11.004

Scheitle, C. P. (2011). Google's insights for search: A note evaluating the use of search engine data in social research. *Social Science Quarterly*, *92*(1), 285–295. doi:10.1111/ssqu.2011.92.issue-1

The Centers for Disease Control and Prevention (2012). *BRFSS Prevalence & Trends Data*. Retrieved from https://nccd.cdc.gov/BRFSSPrevalence/rdPage.aspx?rdReport=DPH_BRFSS.ExploreByTopic&irbLocationType=StatesAndMMSA&islClass=CLASS14&islTopic=TOPIC09&isl-Year=2016&rdRnd=49611

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Lucas, R. E., Agrawal, M., … Ungar, L. (2013, June). Characterizing geographic variation in well-being using tweets. In *Seventh International AAAI Conference on Weblogs and Social Media, Cambridge, Massachusetts*.

Tsou, M. H., Jung, C. T., Allen, C., Yang, J. A., Gawron, J. M., Spitzberg, B. H., & Han, S. (2015, July). Social media analytics and research test-bed (SMART dashboard). In *Proceedings of the 2015 International conference on social media & society* (pp. 2). Bern, Switzerland: International Cartographic Association. Retrieved from http://dl.acm.org/citation.cfm?id=2789196

Tsou, M. H., & Lusher, D. (2015). Mapping web pages by internet protocol (IP) addresses: Analyzing spatial and temporal characteristics of web search engine results. *Proceedings of the International Symposium on Cartography in Internet and Ubiquitous Environments 2015*, 17th–19th March, Tokyo.

US Department of Health and Human Services. (2011). *About Healthy People*. Retrieved from https://www.healthypeople.gov/2020/About-Healthy-People

US Census Bureau. (2016). *American FactFinder*. Retrieved from https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk

Vosen,S. & Schmidt, T. (2011). Forecasting private consumption: Survey based indicators Vs. Google trends. *Journal of Forecasting*, *30*(6), 565–578.

Wang, Y., & Rodgers, S. (2013). Reporting on health to ethnic populations: A content analysis of local health news in ethnic versus mainstream newspapers. *Howard Journal of Communications*, *24*(3), 257–274. doi:10.1080/10646175.2013.805991

Wang, Y., Willis, E., & Rodgers, S. (2014). Consumer insights from peer-to-peer communication in an online health community of weight management. *Journal of Consumer Health on the Internet*, *18*(2), 143–156. doi:10.1080/15398285.2014.902273

Willis, E., & Wang, Y. (2016). Blogging the brand: Meaning transfer and the case of Weight Watchers' online community. *Journal of Brand Management*, *23*(4), 457–471. doi:10.1057/bm.2016.16

Wing, R. R. (1999). Physical activity in the treatment of the adulthood overweight and obesity: current evidence and research issues. *Medicine and Science in Sports and Exercise*, *31*(11 Suppl), S547–52.

Wing, R. R., & Phelan, S. (2005). Long-term weight loss maintenance. *The American Journal of Clinical Nutrition*, *82*(1), 222S–225. doi:10.1093/ajcn/82.1.222S