

Twitter analysis of the orthodontic patient experience with braces vs Invisalign

Daniel Noll^a; Brendan Mahon^b; Bhavna Shroff^c; Caroline Carrico^d; Steven J. Lindauer^e

ABSTRACT

Objective: To examine the orthodontic patient experience having braces compared with Invisalign by means of a large-scale Twitter sentiment analysis.

Materials and Methods: A custom data collection program was created that collected tweets containing the words “braces” or “Invisalign” for a period of 5 months. A hierarchical Naïve Bayes sentiment analysis classifier was developed to sort the tweets into five categories: positive, negative, neutral, advertisement, or not applicable. Each category was then analyzed for specific content.

Results: A total of 419,363 tweets applicable to orthodontics were collected. Users posted significantly more positive tweets (61%) than they did negative tweets (39%; $P \leq .0001$). There was no significant difference in the distribution of positive and negative sentiment between braces and Invisalign tweets ($P = .4189$). Positive orthodontics-related tweets often highlighted gratitude for a great smile accompanied with selfies. Negative orthodontic tweets frequently focused on pain.

Conclusion: Twitter users expressed more positive than negative sentiment about orthodontic treatment with no significant difference in sentiment between braces and Invisalign tweets. (*Angle Orthod.* 2017;87:377–383)

KEY WORDS: Twitter; Invisalign; Braces; Aligners; Clear; Experience

INTRODUCTION

Communication plays a critical role in health care. Providers seek to improve patient care by connecting with patients and understanding their experiences. Traditionally, health care providers gathered this information through surveys, reviews, and word of mouth. In the past decade, communication methods have rapidly changed with the growth of social media.

Founded in 2006, Twitter is an online, fast-paced microblog wherein users share posts in 140 characters or less. Traditional blogs allow for longer, more static content, while microblogs like Twitter focus on shorter, more frequent posts. With 320 million active monthly users, Twitter has grown exponentially and become a primary method of multipurpose communication throughout the world.¹

People use Twitter every day for communication, information, and entertainment. However, people primarily utilize Twitter to express their current thoughts and feelings. Kelly² categorized the content of Twitter posts and found that 41% of Twitter posts are “pointless babble” and another 38% of tweets are conversations between users. News, information, spam, and self-promotion made up the remaining 21% of the posts. Eighty percent of users now access Twitter through their mobile device, allowing people to tweet in the moment.¹

These written thoughts and feelings posted as tweets are unsolicited and publicly available. As a result, Twitter is a unique source of data. Traditional surveys often introduce recall bias and are difficult to conduct on a large scale. Twitter data are collected in

^a Private Practice, Cincinnati, Ohio.

^b Program Engineer, Boston, Mass.

^c Professor and Program Director, Department of Orthodontics, School of Dentistry, Virginia Commonwealth University, Richmond, Va.

^d Assistant Professor, Department of Research Administration, School of Dentistry, Virginia Commonwealth University, Richmond, Va.

^e Professor and Chair, Department of Orthodontics, School of Dentistry, Virginia Commonwealth University, Richmond, Va.

Corresponding author: Dr Bhavna Shroff, Professor and Program Director, Department of Orthodontics, VCU School of Dentistry, Box 980566, Richmond, VA 23298-0566

Accepted: November 2016. Submitted: June 2016.

Published Online: January 06, 2016.

© 2017 by The EH Angle Education and Research Foundation, Inc.

real time, free from recall bias.³ With millions of tweets posted per day, the data source is vast.

Twitter data are best analyzed on a large scale with sentiment analysis.^{4,5} Sentiment analysis, often referred to as opinion mining, is a method to extract and characterize subjective information. Twitter sentiment analysis has been employed to study many fields, from stock market indicators to political election predictions.⁶⁻⁸ Companies seek ways to mine Twitter for consumer feedback about their goods and services.⁹ Twitter's immense information source is largely untapped in orthodontics.

Heavilin et al.¹⁰ found that the public uses Twitter to broadcast experiences and thoughts about dental pain in real time. Their Twitter results were similar to traditional surveys about dental pain, potentially validating Twitter as a data source in the dental field. Henzell et al.¹¹ analyzed 131 orthodontics-related tweets and found that orthodontic patients use social media sites such as Twitter to convey positive and negative feelings about their treatment.

The current literature regarding the patient experience with braces compared with Invisalign is sparse and conflicting. Miller et al.¹² compared the two treatment methods and found that Invisalign patients experienced less discomfort, pain, and analgesic use during their first week of orthodontic treatment than patients with traditional appliances. However, Shalish et al.¹³ found no statistically significant differences in pain levels, analgesic use, or speech dysfunctions between patients treated with Invisalign and traditional appliances. Patients wearing traditional braces reported more oral sores and food accumulation but similar levels of sleep and daily life disturbances. Given the increasing popularity of clear aligners, further research is needed to investigate other aspects of the patient experience such as esthetics and treatment satisfaction. Twitter provides a new and exciting medium in which to examine the impact of orthodontic treatment on everyday life.

The aim of this study was to examine orthodontic patient experience with braces compared with that of Invisalign by means of a large-scale Twitter analysis. The null hypothesis was that there is no difference in sentiment between tweets about braces and tweets about Invisalign.

MATERIALS AND METHODS

The Virginia Commonwealth University Institutional Review Board granted an exemption for this because no individual human subject was enrolled and no identifiable information was collected. Tweets were collected over a 5-month period from April through September 2015. All tweets were publicly accessible

from Twitter's database. Inclusion criteria consisted of any tweet that contained either of the two keywords, "braces" or "Invisalign." Each tweet was classified into one of five categories: positive, negative, neutral, advertisement, or not applicable. Applicable tweets were defined as pertaining to orthodontics and written in the English language.

The software programs for this project consisted of two sections: data collection and data interpretation. The data collection program was written to interact with Twitter's servers and continuously collect all tweets that met the inclusion criteria. A second program was written to interpret the entire collected database. Each tweet was classified into one of the five previously listed categories by machine-learning sentiment analysis. The program was constructed using a Hierarchical Naïve Bayes classifier, the preferred method for Twitter sentiment analysis.⁴ Naïve Bayes classifiers are probabilistic classifiers that break down a block of text into a group of independent words and classify the text into a category based on the text's similarity to precategorized texts.¹⁴ Thus, the context-aware program "learns" from the precategorized texts.

Naïve Bayes classifiers require manual classification of a number of tweets to act as reference material to "train" the program. In this study, an independent reviewer manually sorted 3784 tweets into one of the five categories. These preclassified tweets, referred to as a corpus, were used to achieve two objectives: to train the program and to test it. From the corpus, 71% (2706 tweets) were used to train the classifier on which words and features were most representative of each category. The other 1078 tweets in the corpus were used to test agreement between the independent reviewer and the program.

Text classifiers are most effective when classifying text into one of two categories. The program sorted tweets into the five categories in a specific sequence (Figure 1). This method is known as hierarchical classification.¹⁵ The first classifier determined whether the text was applicable to orthodontics. Examples of nonapplicable tweets included posts such as "Britain braces for election gridlock" or tweets about knee braces. If the tweet was classified as applicable, the text advanced to the second classifier, which determined whether the tweet was an advertisement. If it was not an advertisement, the text was sent to the third classifier, which determined whether the tweet was neutral or expressed a strong sentiment. If the tweet was not neutral, it advanced to the fourth and final classifier, which determined whether the tweet expressed a positive or a negative sentiment.

Next, the corpus of 3784 tweets was analyzed and evaluated for specific content. Frequently used words were incorporated into tables of indicator words and

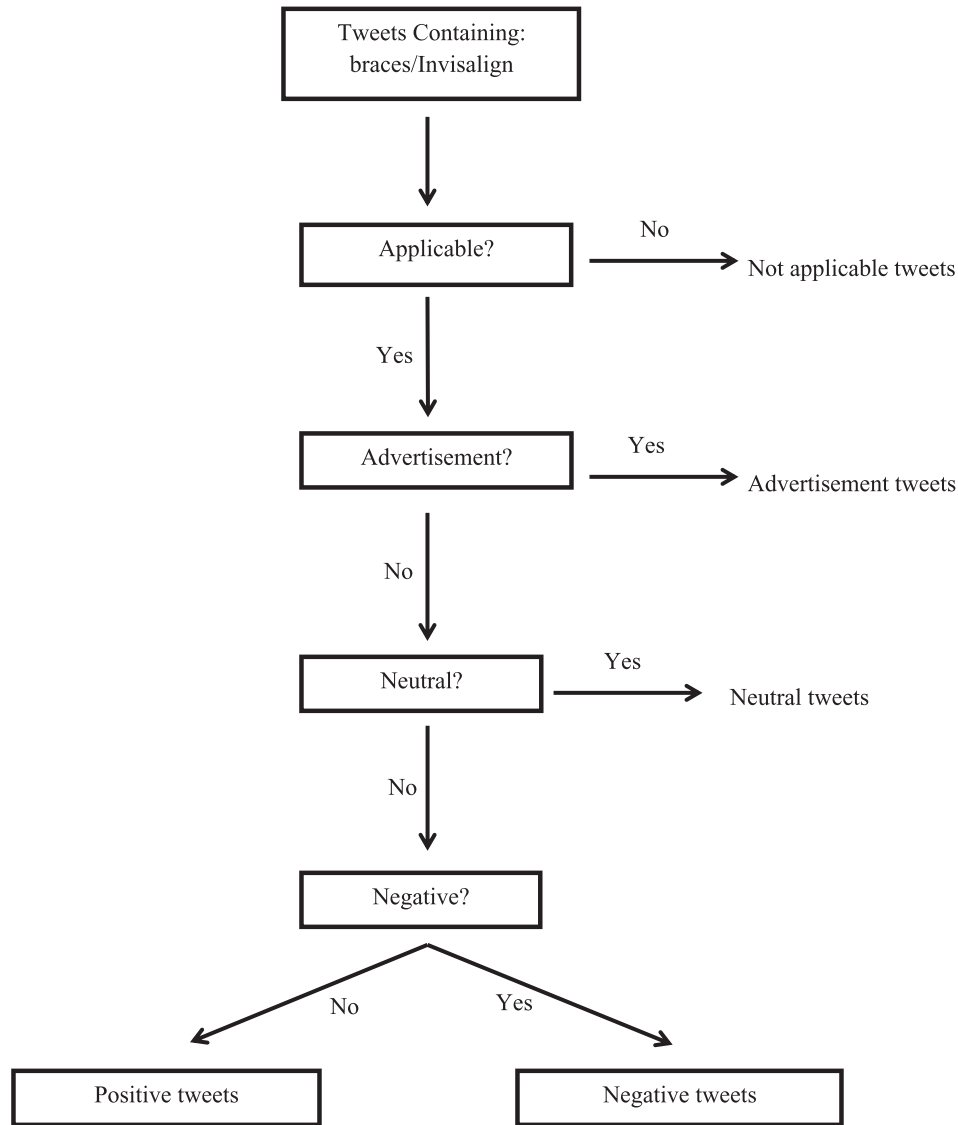


Figure 1. Classifying sequence.

ratios. These indicator ratios showed how likely a specific word was to cause a tweet to be sorted into a certain category, offering insight into the content of each category.¹⁴

Chi-square tests were employed to identify significant differences in the distribution of each category. All analyses were performed in the Statistical Analysis System (SAS) EG v6.1.^{16,17} The kappa (κ) statistic was used to measure agreement between the independent reviewer and the program.

RESULTS

Agreement between the independent reviewer and the program was examined by testing 1078 common tweets. The agreement was found to be almost perfect ($\kappa = 0.97$).¹⁸

Over a 5-month period, a total of 477,054 tweets were collected that contained the word braces or Invisalign, of which 419,363 were applicable to orthodontics. Many more tweets contained the word braces (96%) than Invisalign (4%).

Figure 2 is a flow chart of all collected tweets and their classification. Tweets not applicable to orthodontics made up 12% (57,691) of all collected tweets and were excluded. Among the applicable tweets, advertisements made up 8% (34,819). The remaining 92% of the tweets applicable to orthodontics were assumed to be from orthodontic patients or people interested in orthodontics. Next, 53,677 tweets were classified as neutral and filtered out. The remaining subset contained 330,867 positive and negative tweets about the orthodontic experience.

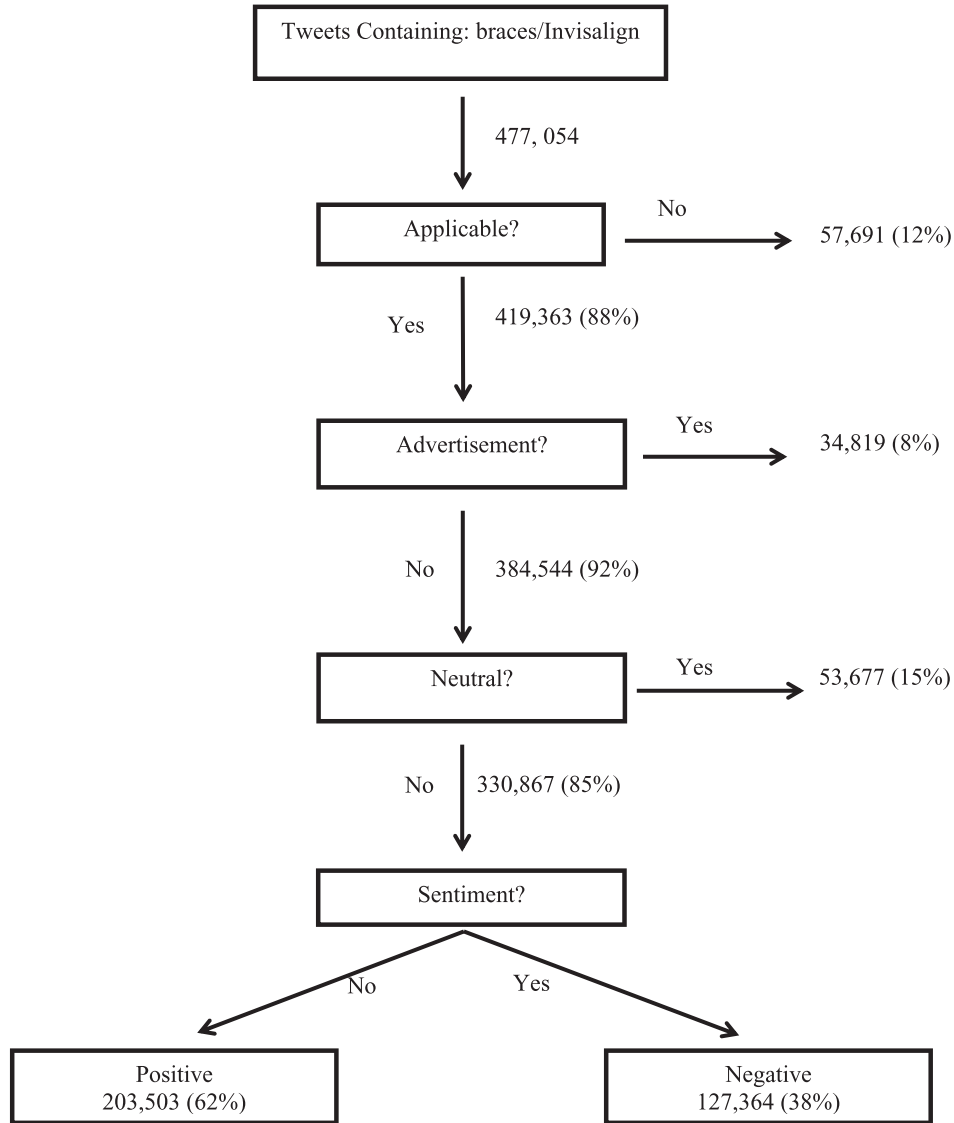


Figure 2. Overall flow diagram results.

Table 1. Nonapplicable/Applicable Indicator Words

Not Applicable	
Indicator Words	Not Applicable:Applicable
Weather	76.2:1.0
Severe	34.9:1.0
#Suspenders	28.7:1.0
#Menswear	27.1:1.0
#Fashion	11.6:1.0
Applicable	
Indicator Words	Applicable:Not Applicable
Teeth	47.4:1.0
Off	46.4:1.0
Want	13.9:1.0

Table 1 shows a breakdown of the applicable and nonapplicable indicator words among the 3784 braces and Invisalign tweets that were analyzed for specific content. The ratios show how likely a given word was to be sorted into a particular category. For example, Table 1 shows that “weather” had a 76.2:1 ratio of not applicable to applicable. If the word weather appeared in a tweet, it was 76.2 times more likely to be classified as not applicable than applicable. Other tweets that were not applicable to orthodontics contained the words, “severe,” “#suspenders,” “#menswear,” or “#fashion.” In contrast, the word “teeth” was 47.4 times more likely to be classified as applicable to orthodontics. Other tweets applicable to orthodontics included the words “off” and “want.”

There was a significant difference in the proportion of advertisements between Invisalign and braces tweets

Table 2. Distribution of Advertisements for Braces and Invisalign Tweets

	Advertisement	Not an Advertisement
Invisalign	5940 (33%)	12,092 (67%)
Braces	28,879 (7%)	372,452 (93%)

($P < .0001$), with 33% of Invisalign tweets being classified as advertisements and only 7% of braces tweets classified as such. Despite this difference in proportion, a greater number of braces advertisements (28,879) were collected than were Invisalign advertisements (5940). Table 2 displays the total counts and percentages for each category.

Table 3 presents the indicator words for the advertisement tweets broken down between braces and Invisalign tweets. The indicator words listed in the table were much more frequently found in tweets that were classified as advertisements than as nonadvertisements. In contrast to Table 1, the indicator words in Table 3 have a braces column and an Invisalign column to compare and contrast the advertising differences between the two groups. Braces ads often contained the words “smile,” “straight,” and “traditional,” while the Invisalign tweets included the words “offer,” “whitening,” and “alternative.”

Sentiment was then analyzed after separating tweets into the two main categories of braces and Invisalign. There was no significant difference in the distribution of positive and negative tweets for braces compared with Invisalign ($P = .4189$), as 38% of Invisalign tweets were classified as negative and 39% of braces tweets were classified as negative.

Among all braces and Invisalign tweets that expressed polarity, there were significantly more positive tweets than negative ($P < .0001$), as 62% of polarized tweets were positive and 38% were negative. Table 4 displays the total counts and percentages for each category. Table 5 shows a breakdown of indicator words for the positive and negative tweets. Common positive words included “thank,” “#smile,” and

Table 3. Advertisement Indicator Words

Indicator Words	Advertisements			
	Braces		Invisalign	
	Ad:	Not an Ad	Ad:	Not an Ad
Smile!	31.1:	1.0	Offer	8.3:1.0
Straight	26.5:	1.0	Whitening	7.2:1.0
Traditional	21.9:	1.0	Alternative	7.2:1.0
Offer	19.6:	1.0	#Smile	6.2:1.0
#Beauty	17.3:	1.0	Open	6.1:1.0
Free	14.5:	1.0	Http: . . .	5.0:1.0
Clear	14.3:	1.0	Start	3.9:1.0
			Today!	3.9:1.0

Table 4. Distribution of Sentiment for Braces and Invisalign Tweets

	Positive	Negative
Invisalign	5531 (62%)	3402 (38%)
Braces	197,792 (61%)	123,962 (39%)


“#selfie,” while common negative words included “hate,” “pain,” “hurts,” and “food.”

DISCUSSION

Strengths of this study include accurate classification of a large volume of tweets and quantitative analyses of each category of tweets. Eke¹⁹ raised concern about the use of Twitter for research, worrying that because context is not taken into account when extracting specific words, such a method could result in low predictive values. To reduce this risk, the Naïve Bayes classifying technique utilized “context-aware” machine learning. As a result, sorting by the program had nearly perfect agreement with the manual human sorting. A limitation of Twitter studies is the inability to gather demographic information, as users’ demographics are not linked to their profile.

Align Technology (San Jose, Calif) advertises Invisalign treatment as offering an improved patient experience over traditional braces, emphasizing comfort, convenience, and an improved lifestyle.²⁰ The results of this study did not find a significant difference in the distribution of sentiment between the braces and Invisalign tweets. Therefore, the null hypothesis that there is no difference in sentiment between tweets about braces and tweets about Invisalign was not rejected.

Table 5. Sentiment Indicator Words

Indicator Words	Positive
	Pos:Neg
Thank	6.4:1.0
#Smile	6.0:1.0
#Selfie	4.7:1.0
	3.7:1.0
Indicator Words	Negative
	Neg:Pos
Hate	26.5:1.0
Pain	17.5:1.0
Hurts	13.5:1.0
Food	8.5:1.0
Rubber	7.5:1.0
Lisp	6.5:1.0
Ugly	6.5:1.0
School	5.7:1.0
Retainers	5.5:1.0
Broke	5.5:1.0
Bands	5.5:1.0
Sick	4.5 :1.0

Most of the tweets expressed positive sentiments regarding both major treatment modalities—braces and Invisalign. Many users expressed gratitude for their orthodontic treatment, using the word thank to express appreciation for their new smile. Many of the positive #selfie tweets were accompanied by photographs of patients showing their new smile upon completion of treatment. This finding demonstrates that appliance removal defines a significant day in the life of patients. Social media can be employed to capture these important moments and connect a practice to the community.

The negative tweets gave insight into the experiences considered irritating by patients. The most frequent complaint about treatment was pain. Most of the negative tweets containing the word food were focused on the pain from chewing and not necessarily the food restrictions limiting sticky and hard foods. Other Twitter users bemoaned the challenges of wearing “rubber bands.” “Lisps” developed from Invisalign aligners was another objection, while others said they were “sick” of braces and their “ugly” appearance. Other sources of frustration were “retainers” and broken appliances. Orthodontic providers need to have a thorough understanding of these common negative reactions to treatment in order to improve the orthodontic experience.

Interestingly, the word off was almost always applicable to orthodontics at a ratio of 46.4:1, but off was not found in the list of positive or negative indicator words. While some users excitedly tweeted about getting their braces off, others complained that their orthodontist refused to remove their braces.

Orthodontic advertisements on Twitter emphasized smiles. Ads often contained the word “offer.” Some of these tweets stated the services offered by the office, and others announced special offers to begin orthodontic treatment. Ads for braces sometimes attempted to attract new patients by showcasing “clear” braces. Invisalign was often marketed as an “alternative” to braces, and some offices offered whitening along with Invisalign treatment. Other providers used Twitter to distribute practical information such as hours of operation and practice Web site links.

This study analyzed Twitter posts about orthodontics to gain a better understanding of the patient experience. On average, over 2700 orthodontic tweets were collected each day. Most of the tweets about orthodontics were positive, and there was no significant difference in the sentiment distribution between tweets about braces versus tweets about Invisalign.

CONCLUSIONS

- Twitter users shared more positive than negative sentiment about their orthodontic experiences.

- There was no significant difference in the distribution of positive and negative sentiment between tweets about braces and tweets about Invisalign.
- Negative orthodontic tweets frequently focused on pain.
- Positive orthodontic tweets often highlighted gratitude for a great smile accompanied by selfies.

ACKNOWLEDGMENTS

This study was supported by the Southern Association of Orthodontists and the Virginia Orthodontic Education and Research Foundation. The authors report no conflict of interest with Align Technology.

REFERENCES

1. Twitter Headquarters. Twitter Usage/Company Facts. Available at: <https://about.twitter.com/company>. Accessed November 13, 2015.
2. Kelly R. Twitter Study. 2009 (August). Available from: <https://web.archive.org/web/20110715062407/www.pearanalytics.com/blog/wp-content/uploads/2010/05/Twitter-Study-August-2009.pdf>. Accessed January 12, 2016.
3. Nascimento TD, DosSantos MF, Danciu T, et al. Real-time sharing and expression of migraine headache suffering on Twitter: a cross-sectional infodemiology study. *J Med Internet Res*. 2014;16:e96.
4. Pak A, Paroubek P. Twitter as a corpus for sentiment analysis and opinion mining. *LREc*. 2010;10:1320–1326.
5. Go A, Bhayani R, Huang L. Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*. 2009:1–12.
6. Tomasjan A, Sprenger TO, Sandner PG, Welpel IM. Election forecasts with Twitter: how 140 characters reflect the political landscape. *Soc Sci Comput Rev*. 2010;0894439310386557.
7. Zhang X, Fuehres H, Gloor PA. Predicting stock market indicators through Twitter “I hope it is not as bad as I fear.” *Procedia—Social and Behavioral Sciences*. 2011;26:55–62.
8. Diakopoulos NA, Shamma DA. Characterizing debate performance via aggregated Twitter sentiment. *ACM*; 2010:1195–1198.
9. Kouloumpis E, Wilson T, Moore J. Twitter sentiment analysis: The good the bad and the omg! *lcwsm*. 2011;11:538–541.
10. Heavilin N, Gerbert B, Page JE, Gibbs JL. Public health surveillance of dental pain via Twitter. *J Dent Res*. 2011;90:1047–1051.
11. Henzell RM, Knight MA, Morgaine K, S. Antoun J, Farella M. A qualitative analysis of orthodontic-related posts on Twitter. *Angle Orthod*. 2013;84:203–207.
12. Miller KB, McGorray SP, Womack R, et al. A comparison of treatment impacts between Invisalign aligner and fixed appliance therapy during the first week of treatment. *Am J Orth and Dentofacial Orthop*. 2007;131:302. e1–e9.
13. Shalish M, Cooper-Kazaz R, Ivgi I, et al. Adult patients' adjustability to orthodontic appliances. Part I: a comparison between labial, lingual, and Invisalign. *Eur J Orthod*. 2012;34:724–730.

14. Rish I. *An Empirical Study of the Naïve Bayes Classifier*. New York: IBM; 2001:41–46.
15. Neetu. Hierarchical classification of Web content using Naïve Bayes approach. *Int J Comput Sci Eng*. 2013;5:402.
16. Nordlie RC, Foster JD, Lange AJ. Regulation of glucose production by the liver. *Ann Rev Nutr*. 1999;19:379–406.
17. Wilcox G. Insulin and insulin resistance. *Clin Biochem Rev*. 2005;26:19–39.
18. McHugh ML. Interrater reliability: the kappa statistic. *Biochemia medica*. 2012;22:276–282.
19. Eke PI. Using social media for research and public health surveillance. *J Dent Res*. 2011;90:1045–1046.
20. Align T. The advantage over braces. Available at: <http://www.invisalign.com/why-invisalign/advantage-over-braces>. Accessed 11/21, 2015.