Insulin pricing and other major diabetes-related concerns in the USA: A study of 46,400 Tweets between 2017 and 2019

Short title: Concerns around diabetes in the USA

Adrian Ahne, MSc[1,2]; Francisco Orchard[2]; Xavier Tannier, PhD[3,4]; Camille Perchoux, PhD[5]; Beverley Balkau, PhD[1]; Sherry Pagoto, PhD[6]; Jessica L Harding, PhD[7]; Thomas Czernichow, PhD[2]; Guy Fagherazzi, PhD[1,8]

Affiliations
1: Inserm U1018, Center for Research in Epidemiology and Population Health (CESP), Paris Saclay University, Gustave Roussy Institute, 114 rue Edouard Vaillant, 94800 Villejuif, France
2: Epiconcept, 47 rue de Charenton, 75012 Paris, France
3: Inserm UMRS 1142, LIMICS, 15 rue de l'école de médecine, 75006 Paris, France
4: Sorbonne Université, 4 Place Jussieu, 75005 Paris, France
5: Luxembourg Institute of Socio-Economic Research (LISER), Esch/Alzette, 11 Porte des Sciences, 4366 Esch-sur-Alzette, Luxembourg
6: UConn Center for mHealth & Social Media, University of Connecticut, 2006 Hillside Road, Unit 1248, Storrs, CT 06269-1248, Connecticut, United States of America
7: Division of Transplantation, Department of Surgery, Emory University School of Medicine, Emory University Hospital, Room B206, 1364 Clifton Road, NE Atlanta, GA 30322, United States of America
8: Luxembourg Institute of Health; Department of Population Health, 1A-B, rue Thomas Edison, L-1445 Strassen, Luxembourg

Correspondence to:
Guy Fagherazzi
Luxembourg Institute of Health; Department of Population Health;
1A-B, rue Thomas Edison, L-1445 Strassen, Luxembourg
Telephone: +352 621589554
Email: Guy.Fagherazzi@lih.lu
Twitter: @GFaghe

Word count (Abstract): 247
Word count (Manuscript): 3173
Word count (Significance of study): 202
References: 37
Number tables: 0
Number figures: 4
Online Supplementary material word count: 1052

Online Supplementary number tables: 3
Online Supplementary number figures: 1

Keywords: Emotion, Epidemiology, Psychological Stress, Methodology

ABSTRACT

Objective

Little is known about the concerns of people living with diabetes. This study aims to identify key diabetes-related concerns in the USA and primary emotions associated with those concerns using information shared on Twitter.

Research Design and Methods

A total of 11.7 million diabetes-related tweets in English were collected between April 2017 and July 2019. Machine learning methods were used to filter tweets with personal content, to geolocate (to the US) and to identify clusters of tweets with emotional elements. A sentiment analysis was then applied to each cluster.

Results

We identified 46,400 tweets with emotional elements in the USA from which 30 clusters were identified; five clusters (18% of tweets) were related to insulin pricing with both positive emotions (*joy, love)* referring to advocacy for affordable insulin and *sadness* emotions related to the frustration of insulin prices, five clusters (12% of tweets) to solidarity and support with a majority of *joy* and *love* emotions expressed. The most negative topics (10% of tweets) were related to diabetes distress (24% *sadness,* 27% *anger*, 21% *fear* elements), to diabetic and insulin shocks (45% *anger*, 46% *fear*) and comorbidities (40% *sadness*).

Conclusions

Our work has enabled us to describe the main topics of interest about diabetes and their associated emotions. Insulin pricing is a central diabetes-related concern and is associated with feelings of sadness. Social media data are complementary to traditional diabetes

epidemiology and can contribute to public decision making around key topics of interest around diabetes.

Significance of Study

What is already known about this subject?

- It is very challenging to collect data a population level to truly understand what are the key concerns of people with diabetes in the real life
- Social media such as Twitter can serve as a relevant source of information
- There are key challenges for a global and fair access to insulin

What are the new findings?

- This work represents an unprecedented overview of the large diabetes online community in the USA
- With 18% of the tweets related to insulin pricing, we have observed that it is a major concern in the diabetes community in the USA. People regularly express fear, anger and sadness about potential diabetes-related complications and comorbidities. However, there is a great amount of support and solidarity among the diabetes online community with numerous posts related to positive emotions

How might these results change the focus of research or clinical practice?

- Our work presents a reproducible approach to easily capture relevant information about the main concerns of people with diabetes that is usually not available in typical clinical or epidemiological settings. This could help to shape more relevant public health actions by targeting what really matters to people with diabetes.

INTRODUCTION

Stress, fears and negative emotions related to the management of diabetes are considered to be the most important psychosocial health factor in the management of diabetes.[1] It has been shown that diabetes distress is associated with decisional conflict and therefore has an impact on day-to-day disease management and the long term risk of diabetes-related complications.[2] Reducing diabetes-related distress may improve HbA1c and reduce the burden of disease among people with diabetes.[3,4] However, current diabetes management guidelines dedicate little attention to the prevention and management of diabetes-related distress. This is due, in part, to the limited knowledge of the sources of stress, anxiety and concerns among people with diabetes and to the difficulty to capture[5,6] them with existing evaluation scales such as the Problem Areas in Diabetes Scale (PAID or PAID-T for adolescents)[7] or the Diabetes Distress Scale (DDS).[5,8,9] By neglecting the impact of psychological factors, future interventions may not adequately address the needs of people with diabetes. Moreover, there is no extensive study in the literature about the impact of socioeconomic environment on diabetes-related concerns.

Social media data offer a unique opportunity to supplement current measures of diabetes-related distress and assess the sentiments of people with diabetes given the very active online diabetes community,[10,11] in particular on Twitter. Twitter is a microblogging and social networking service with 321 million monthly active users.[12]

The aims of the present study were to identify diabetes-related concerns in the USA and to identify the primary emotions and sentiments associated with them.

METHODS

Data collection

A tweet extraction engine was developed in April 2017 and since then has collected more than 11.7 million (status July 2019) diabetes-related tweets in English via Twitter's Streaming Application Programming Interface (API), based on a list of diabetes-related keywords, such as *diabetes, hypoglycemia, hyperglycemia* and *insulin*, from all over the world (see Supplement S1 for the full list of keywords used). All data collected in this study were publicly posted on Twitter. Therefore, according to the privacy policy of Twitter,[13] users agree to have this information available to the general public. In this study, we restricted the analysis to 46,400 US-based tweets with personal and emotional content (see Figure 1 for an overview over the workflow).

For further details concerning data representation, data preprocessing, training of machine learning classifiers to filter tweets with personal content, from institutional tweets (organisations, advertisement, research news, etc.) related to diabetes; detect and exclude jokes; predict the gender and type of diabetes of users as well as the geolocation algorithm to determine tweets in the USA, please refer to Supplementary 2. For the present work, the statistical unit considered is a tweet and not a user profile. This implicates that a user may have several tweets.

Identifying emotions

To identify emotions among diabetes-related tweets, we combined detailed Parrot's classification of emotions[14] with a dictionary of emotional keywords present in the two most

common tools to assess psychological health in people with diabetes, namely the Problem Areas in Diabetes questionnaire and the Diabetes Distress Scale. Parrot identified over 100 emotions and conceptualized them as a tree structured list (full list Supplement S3) with the six primary emotions being *joy, love, surprise, sadness, anger* and *fear*. To capture as many tweets as possible containing emotions, we hypothesized that the synonyms of the words from Parrot's classification should be included as well. Synonyms were identified using the WordNet database.[15]

Moreover, due to the nature of Twitter as a microblogging service (short messages), people use emoticons and emojis that express special meanings. Emoticons constitute a metacommunicative pictorial representation of a facial expression using punctuation marks and letters, such as ':-)' or ':D'.[16] Emojis are incorporated into sets of characters available in mobile phones, such as '😀' or '😭'. We used Wolny's categorisation of emojis and emoticons into emotional categories[17] similar to the six primary emotions defined by Parrot. This allowed us to determine a subset of *emotional* tweets containing either emotional words or emojis/emoticons.

Sentiment Analysis

Sentiment analysis is commonly used to investigate the positive or negative opinion within a tweet. We used a widely accepted[18–20] human-validated sentiment analysis tool Valence Aware Dictionary for Sentiment Reasoning (VADER).[21] VADER was specifically developed for social media data and combines a lexicon and the processing of the sentence characteristics to determine sentence polarity. VADER computes sentiment and valence for each word level and provides positive, negative and neutral scores at the sentence level. We

used the compound score, which is unidimensional and a normalized measure of sentiment between -1 (most negative) and +1 (most positive).

Topic extraction

We applied the unsupervised machine learning algorithm K-means[22] using cosine similarity, to group the emotional tweets into topics/clusters. In order to define the right number of clusters *k*, the Silhouette score,[23] measuring how on average each data point is closer to its cluster's center than to any other cluster, was used. We obtained the highest score for $k=30$ clusters. Each topic/cluster was then labelled according to the most contributing tweets (those closest to the topic center) and the most frequent words (top words).

Assessment of the mean income

We studied the associations between the topics of interest for people with or talking about diabetes and the mean income of the city, based on data from the 2017 American Community Survey.[24] The tertiles for the mean household income were calculated: low income - [$24,609, $67,224], medium income - [$67,224, $86,758], high income - [$86,758, $394,259]. Each geolocated tweet was associated with the mean income for its corresponding city and then assigned to its respective tertile: low income, medium (med) income, high income. Associations between topics and mean income tertile groups were assessed with Chi-squared tests. For each topic the p-value was calculated using a Chi2 test between the binary variable if a tweet belongs to the corresponding topic and the tertile categories of the city-level mean income from the tweet's origin.

Python (version > 3.6) and the packages scikit-learn[25] and gensim[26] were used to implement the methods described above. Algorithms related to the present study are open source under the following address: https://github.com/WDDS/Tweet-Diabetes-Classification.

RESULTS

Spatial distribution of diabetes-related tweets

This analysis was based on 167,743 geolocated, diabetes-related tweets from the USA, see Figure 2. The highest number of tweets were seen in California (N: 18,551) and Texas (N: 14,237), whereas Vermont (N: 197) and Wyoming (N: 131) had a low number of tweets. At the city level, New York City (N: 9,663), Los Angeles (N: 5,301) and Chicago (N: 4,884) had the highest number of tweets.

Topics of interest

Among all US diabetes-related tweets, 46,407 were identified as *emotional* tweets (28% of tweets) of which 14,485 (31%) were written by men, 20,228 (44%) by women and 11,694 (25%) from unknown sex; 20,285 (44%) were predicted as from people with type 1 diabetes, 4,375 (9%) from type 2 diabetes and 21,747 (47%) diabetes type unknown. Figure 3 shows the detailed description of the 30 topics of interest about diabetes. In most topics, an over-representation of women could be observed, in particular in topics 10, 23 and 26 related to the importance of affordable insulin and the oral glucose tolerance test (OGTT). In contrast, men more frequently discussed topics around diabetes-related stories (topic n°6) and 'diabetic/insulin shock' (topic n°29). Discussions about 'advocacy for affordable insulin' (topic n°10) and enjoying the exchange in the diabetes online community (topic°5) were

almost exclusively dominated by people with type 1 diabetes. The only topics where people with type 2 diabetes had significantly more tweets than on average is topic 19 related to the confusion between type 1 and type 2 diabetes or in topic 17 when they tweeted about the anniversary of their diagnosis.

Primary emotions related to the topics of interest

Figure 4 illustrates the sentiment and emotions distribution over the 30 topics. In Supplemental S4 the most common emotional words and emojis/emoticons have been listed for each topic. Overall, we found that the most positive topics were related to support and solidarity within the diabetes online community (topic n°1) with 72% of emotions related to *joy* and about inspiring relatives living with diabetes (topic n°2) with *joy* (36%) and *love* (31%) elements. In contrast, the most negative topics were related to 'diabetes-related comorbidities' (topic °30) of which 40% of the emotions were related to *sadness*, 19% to *anger* and 22 % to *fear*; 'diabetic/insulin shock' (topic n°29) with an over-representation of *anger* (45% of all emotions) and *fear* (46%); and "diabetes distress" (topic n°28) with *sadness* (24%), *anger* (27%) and *fear* (21%) emotions expressed. Topics 2 ("Inspiring relatives living with diabetes") and 4 ("Diabetes awareness / Support / Donation") showed a higher proportion of *surprise* emotions. The topics in which people with or talking about diabetes exchange, support each other and show solidarity (topics 1, 2, 3, 4, 5) contained *joy* and *love* emotions most frequently. Topic 9 was related to beauty products with diabetes-related names and was considered irrelevant for our analyses and disregarded.

Insulin pricing

Insulin pricing is a major topic of interest mentioned in 5 of the 30 topics (18% of tweets). A key topic associated with insulin pricing was related to the frustration of not being able to afford insulin (topic n°25), in which many tweets expressed sadness (24%), anger (21%) and fear (20%). Topic 13 regrouped tweets in which people are looking for "glucose guardians" who would be able to support them financially in obtaining insulin. Topic 10 described the tweets of people who are fighting for affordable insulin, which are represented by the hashtag '#insulin4all'. The positive emotions in this group showed mutual support.

Associations between topics of interest and mean income

We observed that topics such as n°10 'advocacy for affordable insulin', n°22: 'insulin' and n°25: 'frustration with insulin prices' were positively associated with mean household city income such that cities with higher incomes were more likely to post tweets about these topics ($P<0.001$ for all) (see Supplementary 5). Similarly, we found that topics 24: 'misunderstandings of diabetes' ($P<0.001$), 19: 'confusion between type 1 and type 2 diabetes' ($P<0.05$), 12: 'life with type 1 diabetes' ($P<0.01$) and 21: 'glucose variability' ($P<0.05$) were positively associated with mean city income. Likewise, positive associations were found between topics 24 and 19 which are related to language use and the way people talk about type 1 and type 2 diabetes with cities with higher incomes being more likely to talk about these subjects.

In contrast, we observed that topics 5: 'diabetes social media advocacy' (#DSMA) group enjoying online support ($P<0.001$), 18: 'day-to-day stories about diabetes' ($P<0.001$) and 26: 'oral glucose tolerance test (OGTT)' ($P<0.001$) were negatively associated with mean city income such that cities with lower income were more likely to post these tweets.

Furthermore, cities with medium mean income tweet about topic 15: 'Insulin pump complications' more frequently ($P<0.001$).

DISCUSSION

Our findings suggest that Twitter is useful to capture diabetes-related topics of interest and the emotions associated with those topics. Our key findings suggest that there is a great amount of support and solidarity among the diabetes online community with numerous tweets related to *joy* and *love* emotions observed. On the other hand, there also is considerable frustration (annoyance) about the ability of people to distinguish between type 1 and type 2 diabetes and fear, anger and sadness related to diabetes-related complications and comorbidities as well as around insulin pricing.

We found that insulin pricing is a major concern in tweets shared in the US (18% of the emotional tweets related to this topic) and is associated with both positive (*joy*, *love*) and negative (*sadness, anger, fear)* emotions. People frequently share their frustration with insulin prices,[27] access to insulin,[28] and search for ways to help them finance it by resorting to a "glucose guardian" or from donations which are often accompanied with negative emotions. Positive emotions are present when it comes to solidarity in the fight for affordable insulin and the exchange over insulin prices in the community in general. We observed associations of topics addressing insulin pricing to be more frequent in cities with high mean incomes. This does not indicate that people living in cities with a high mean household income feel more concerned about the insulin prices, but that they probably have a greater ability to organize and gather around this issue. A large number of tweets geolocated in cities with a high mean household income include the hashtag '#insulin4all', a campaign that unites the diabetes community around the access to treatment for everyone. The Diabetes Social

Media Advocacy (#DSMA) group on Twitter is a community of people exchanging about diabetes related topics, in which the price of insulin is a central concern.

Comparison with the literature

Hughes et al. suggested the importance of understanding emotions and self-control (regulation of thoughts, emotions, and behavior) for health outcomes in adolescents with diabetes.[29] Hagger et al. reviewed diabetes distress among adolescents with type 1 diabetes and found that a substantial proportion experienced elevated diabetes distress and that this is often associated with suboptimal glycemic control.[30] Richman et al. suggested that positive emotions may play a protective role in the development of disease.[31] Ogbera et al. showed that emotional distress is associated with poor self-care in type 2 diabetes.[32] A study conducted by Iturralde et al. showed that anxiety is highly comorbid with depression among individuals with type 2 diabetes.[4] Our study aligns with these results showing that emotions and diabetes distress topics are frequent and concern people with diabetes or people talking about diabetes on social media.

Similar to Nguyen et al.,[33] who showed that individuals living in zip codes with high percentages of happy and physically-active tweets had lower obesity prevalence based on geolocated Twitter data, we have shown gradients between topics related to diabetes on Twitter and the household income level of their city.

It is known that there are key challenges for a global and fair access to insulin.[34] With respect to insulin pricing, we are the first to exhibit and quantify, on a large sample of people with or talking about diabetes, the extent of the crisis in the USA based on social media data. In

addition, we have also been able to highlight the different emotions and fears associated with the crisis around insulin pricing and therefore the potential impacts on the health of people with diabetes.

Strengths and Limitations

This study has numerous strengths. First, a major advantage of using social media data is that information is expressed spontaneously, on a large scale, in a real-time manner, in what can be considered as an open digital space with flat role hierarchy for information sharing and development of online communities. This potentially minimises biases such as responder bias that you would observe in traditional and observational studies. We evaluated tweets related to diabetes from a large number of people with a great variability in their profiles. The methodologies developed in this study present an innovative way to concentrate on relevant (personal, emotional) geolocated tweets (US), to identify topics of interest and emotions shared within topics. This approach is able to capture trends in the diabetes online community and socioeconomic factors can be associated with social media data at the ecological level. Moreover, topics are very less medically oriented which allows us to identify alternative topics which would not have been detected in a traditional clinical setting.

There are, however, several limitations to consider. First, diabetes-related concerns expressed on Twitter may not be representative of all people with diabetes. However, it has been previously suggested that it can be partially offset by the large variability in the profiles of people sharing data in the first place, which is key in digital epidemiology.[35] Indeed, we have observed a large variability in their profiles, allowing us to identify various clusters of interest. Besides, the precision of our filter classifiers (personal content, jokes), gender and type of diabetes classifiers is not perfectly accurate which means that we cannot guarantee

that 100% of tweets are posted by actual people with diabetes and it was sometimes impossible to define the sex or type of diabetes. Besides, our results should be interpreted in the context of the Twitter population solely. For instance, we found an over-representation of people with type 1 diabetes and women in our study when compared to what is known from the worldwide diabetes epidemiology literature[36] which may be explained by the younger demographic of Twitter users.[37] We were unable to account for several clinical and environmental factors that may help us tease out these associations. The label provided by the researchers for each topic is not exclusive. By refining the tweets in each topic, more subtopics could be defined. This could be a future direction to investigate. Finally, the geolocation of tweets was partially based on locations the users provided, which might not be their true location. No causal inference can be made regarding the link between the mean household income per city and the topics of interest of people residing in the corresponding city; and would be subject to the ecological fallacy. A person tweeting from a city of low mean income does not have necessarily a low income. A natural perspective of our work is to extend our analyses to include more countries and languages.

Conclusion

In this study we investigated diabetes-related topics and their associated emotions. We showed that insulin pricing is a central concern and comes with feelings of sadness. We have also shown that feelings, emotions and stress, which are not usually well taken into account in traditional epidemiology, are essential elements to capture the full picture of the real-life of people with diabetes. Our work should encourage future studies to take them into account. Future intervention studies should focus on reducing the level of stress and fears in people with diabetes.

Social media provides a useful observatory for diabetes issues as it is a direct source to capture information about people with diabetes feelings, emotions, beliefs and fears related to diabetes, diabetes treatment and complications among the large and active diabetes online community. The use of Twitter analysis on diabetes could inform the public debate about diabetes issues and help to contribute directly to public and clinical decision making.

Acknowledgment

Guy Fagherazzi takes full responsibility for the work as a whole, and for the decision to submit and publish the manuscript.

The authors' contributions were as follows: G.F. designed the research; A.A. and G.F. conducted the research; F.O. and A.A collected the data; A.A., F.O. and G.F. analyzed data; A.A., G.F., F.O., T.C, X.T interpreted the data; A.A. and G.F. drafted the article; G.F., F.O., T.C., X.T., B.B., J.L.H. , C.P. and S.P. revised the manuscript critically; GF had primary responsibility for the final content of the manuscript. All authors read and approved the final manuscript.

Role of the funding source

This work has been supported by the MSDAvenir Foundation, the French Speaking Diabetes Society and the Luxembourg Institute of Health. These study sponsors had no role in the design nor the interpretation of the results of the present study. Adrian Ahne, Francisco Orchard and Thomas Czernichow are supported by Epiconcept Company. Epiconcept was involved in the data collection and writing of the report. No study sponsor influenced the decision to submit the paper for publication.

Declaration of Interest

The authors declare that they have no competing interests. The study sponsors had no role in the study design, analyses and interpretation of the results.

References

1. Finer S, Robb P, Cowan K, et al. Top ten research priorities for type 2 diabetes: results from the Diabetes UK–James Lind Alliance Priority Setting Partnership. *The Lancet Diabetes & Endocrinology*.

2. Bruno BA, Choi D, Thorpe KE, et al. Relationship Among Diabetes Distress, Decisional Conflict, Quality of Life, and Patient Perception of Chronic Illness Care in a Cohort of Patients With Type 2 Diabetes and Other Comorbidities. *Diabetes Care*. Epub ahead of print 2 May 2019. DOI: 10.2337/dc18-1256.

3. Cummings DM, Lutes LD, Littlewood K, et al. Randomized Trial of a Tailored Cognitive Behavioral Intervention in Type 2 Diabetes With Comorbid Depressive and/or Regimen-Related Distress Symptoms: 12-Month Outcomes From COMRADE. *Diabetes Care* 2019; 42: 841–848.

4. Iturralde E, Chi FW, Grant RW, et al. Association of Anxiety With High-Cost Health Care Use Among Individuals With Type 2 Diabetes. *Diabetes Care*. Epub ahead of print 18 June 2019. DOI: 10.2337/dc18-1553.

5. Fenwick EK, Rees G, Holmes-Truscott E, et al. What is the best measure for assessing diabetes distress? A comparison of the Problem Areas in Diabetes and Diabetes Distress Scale: results from Diabetes MILES-Australia. *J Health Psychol*. Epub ahead of print 2016. DOI: 10.1177/1359105316642006.

6. Fisher L, Hessler DM, Polonsky WH, et al. When Is Diabetes Distress Clinically Meaningful?: Establishing cut points for the Diabetes Distress Scale. *Diabetes Care* 2012; 35: 259–264.

7. Polonsky WH, Anderson BJ, Lohrer PA, et al. Assessment of diabetes-related distress. *Diabetes Care* 1995; 18: 754–760.

8. Polonsky WH, Fisher L, Earles J, et al. Assessing Psychosocial Distress in Diabetes: Development of the Diabetes Distress Scale. *Diabetes Care* 2005; 28: 626–631.

9. Schmitt A, Reimer A, Kulzer B, et al. How to assess diabetes distress: comparison of the Problem Areas in Diabetes Scale (PAID) and the Diabetes Distress Scale (DDS). *Diabet Med* 2016; 33: 835–843.

10. Bevan A. T1D Exchange: An Online Community for People Touched by Type 1 Diabetes. *AADE in Practice* 2017; 5: 44–46.

11. Balkhi AM, Reid AM, McNamara JPH, et al. The diabetes online community: the importance of forum use in parents of children with type 1 diabetes. *Pediatr Diabetes* 2014; 15: 408–415.

12. Number of monthly active Twitter users worldwide from 1st quarter 2010 to 1st quarter 2019 (in millions). *Statista*,

https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/ (accessed 28 October 2019).

13. Twitter Privacy Policy, Twitter URL: https://twitter.com/en/privacy [accessed 2019-07-23] (2018).

14. Gerrod Parrott W. *Emotions in Social Psychology: Essential Readings*. Psychology Press, 2001.

15. Princeton University. Princeton University 'About WordNet.', https://wordnet.princeton.edu/ (2010, accessed 10 April 2019).

16. Comesaña M, Soares AP, Perea M, et al. ERP correlates of masked affective priming with emoticons. *Computers in Human Behavior* 2013; 29: 588–595.

17. Wolny W. Emotion Analysis of Twitter Data That Use Emoticons and Emoji Ideograms. *25TH INTERNATIONAL CONFERENCE ON INFORMATION SYSTEMS DEVELOPMENT (ISD2016 POLAND)*.

18. Sewalk KC, Tuli G, Hswen Y, et al. Using Twitter to Examine Web-Based Patient Experience Sentiments in the United States: Longitudinal Study. *J Med Internet Res* 2018; 20: e10043.

19. Ribeiro FN, Araújo M, Gonçalves P, et al. SentiBench - a benchmark comparison of state-of-the-practice sentiment analysis methods. *EPJ Data Science*; 5. Epub ahead of print 2016. DOI: 10.1140/epjds/s13688-016-0085-1.

20. Tamersoy A, De Choudhury M, Chau DH. Characterizing Smoking and Drinking Abstinence from Social Media. *HT ACM Conf Hypertext Soc Media* 2015; 2015: 139–148.

21. Hutto and Eric C. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. *Eighth International Conference on Weblogs and Social Media (ICWSM-14), Ann Arbor, MI*.

22. MacQueen J. Some methods for classification and analysis of multivariate observations. In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*. Berkeley, Calif.: University of California Press, 1967, pp. 281–297.

23. Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* 1987; 20: 53–65.

24. U.S. Census Bureau; 2013-2017 American Community Survey 5-Year Estimates using American FactFinder, http://factfinder.census.gov (accessed 12 August 2019).

25. Pedregosa F., Varoquaux G., Gramfort A. et al. Scikit-learn: Machine Learning in Python. *J Mach Learn Res* 2011; 12: 2825–2830.

26. Radim Rehurek PS. Software Framework for Topic Modelling with Large Corpora. In: *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*.

ELRA, 2010, pp. 45–50.

27. Conner F, Pfiester E, Elliott J, et al. Unaffordable insulin: patients pay the price. *Lancet Diabetes Endocrinol* 2019; 7: 748.

28. Fralick M, Kesselheim AS. The U.S. Insulin Crisis - Rationing a Lifesaving Medication Discovered in the 1920s. *N Engl J Med* 2019; 381: 1793–1795.

29. Hughes AE, Berg CA, Wiebe DJ. Emotional processing and self-control in adolescents with type 1 diabetes. *J Pediatr Psychol* 2012; 37: 925–934.

30. Hagger V, Hendrieckx C, Sturt J, et al. Diabetes Distress Among Adolescents with Type 1 Diabetes: a Systematic Review. *Curr Diab Rep* 2016; 16: 9.

31. Richman LS, Kubzansky L, Maselko J, et al. Positive emotion and health: going beyond the negative. *Health Psychol* 2005; 24: 422–429.

32. Ogbera A, Adeyemi-Doro A. Emotional distress is associated with poor self care in type 2 diabetes mellitus. *J Diabetes* 2011; 3: 348–352.

33. Nguyen QC, Brunisholz KD, Yu W, et al. Twitter-derived neighborhood characteristics associated with obesity and diabetes. *Sci Rep* 2017; 7: 16425.

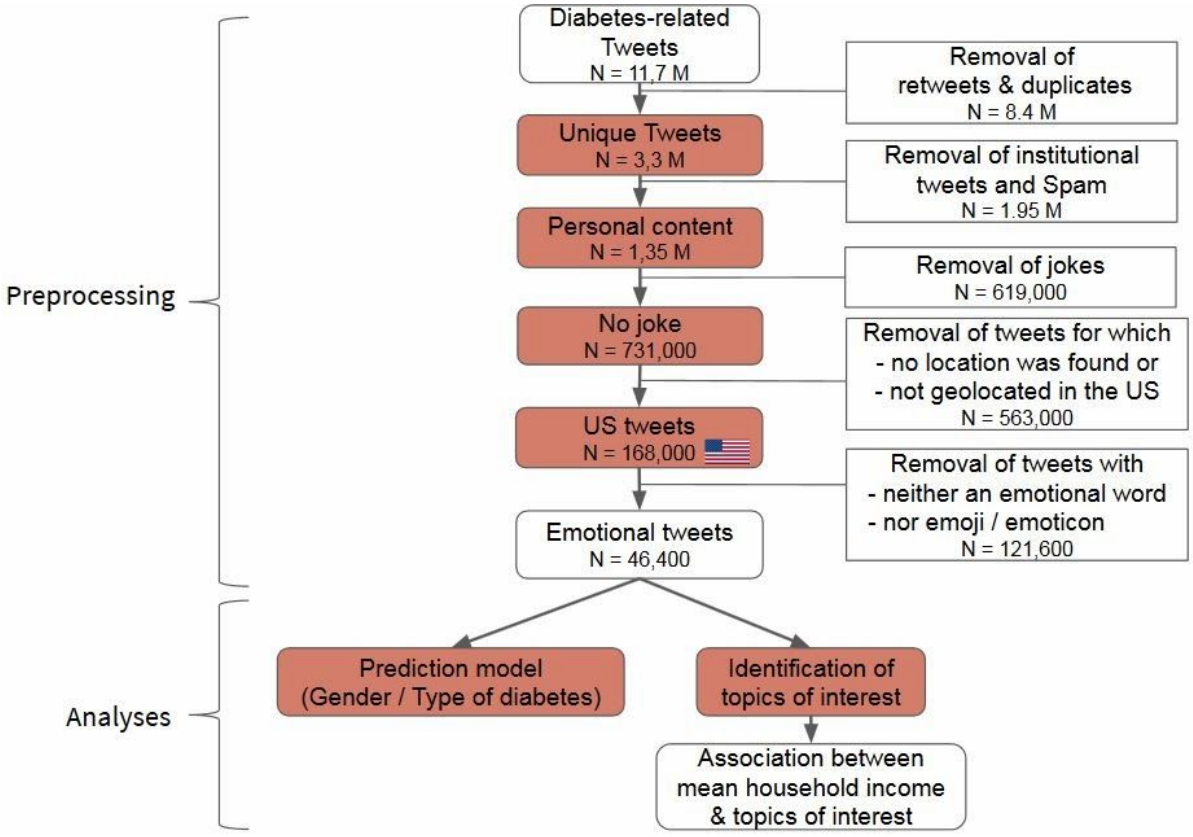
34. Luo J, Gonsalves G, Greene J. Insulin for all: treatment activism and the global diabetes crisis. *Lancet* 2019; 393: 2116–2117.

35. Adrover C, Bodnar T, Huang Z, et al. Identifying Adverse Effects of HIV Drug Treatment and Associated Sentiments Using Twitter. *JMIR Public Health Surveill* 2015; 1: e7.

36. International Diabetes Federation. Brussels, Belgium. IDF Diabetes Atlas, 9th edn, https://www.diabetesatlas.org/en/ (2019).

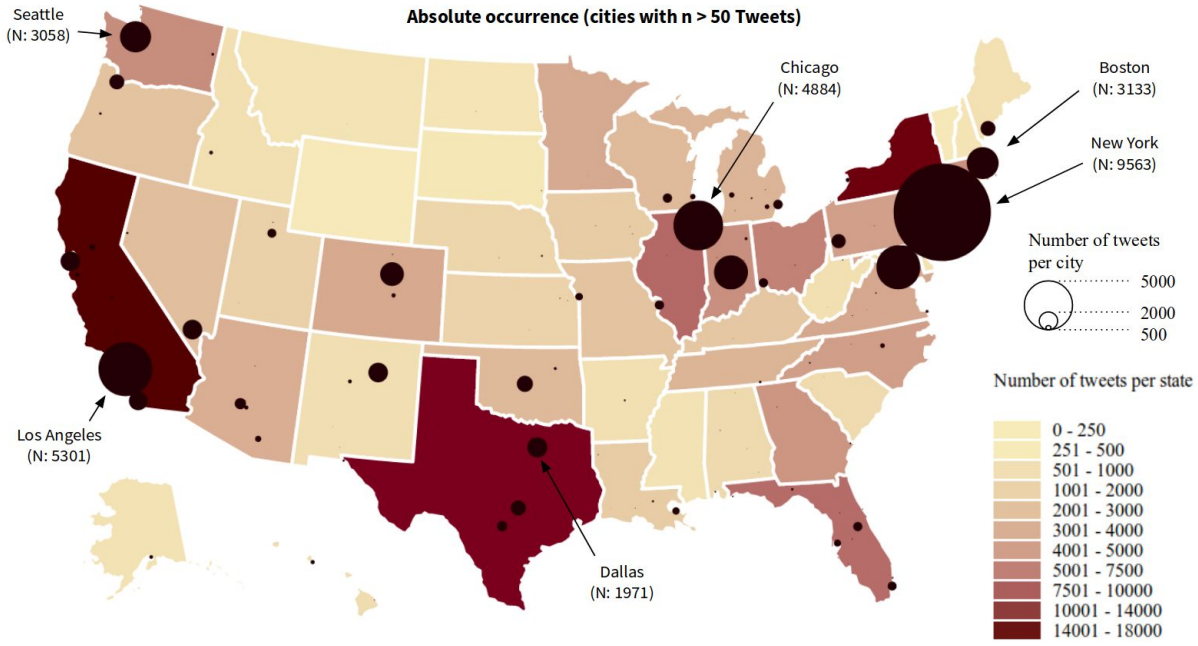
37. Percentage of U.S. adults who use Twitter as of February 2019, by age group. *Statista*, https://www.statista.com/statistics/265647/share-of-us-internet-users-who-use-twitter-by-age-group/ (accessed accessed 28 October 2019).

Figure 1: Workflow: from the data collection to the analysis



Workflow from data collection to data analysis. Red boxes correspond to steps where machine learning methods were applied.

Figure 2: Spatial distribution of the diabetes-related tweets over the USA



The number of tweets in each state is represented by color ranging from few tweets in white-yellow to a lot of tweets in dark red. The size of the black circles is proportional to the number of tweets in a given city.

Figure 3: Overview of the 30 topics of interest for people with or talking about diabetes and their gender and type of diabetes distributions

No.	Topic label	Top words	Gender M/F (%)	Type Diabetes* T1D ‖ T2D(%)	Sample tweet
1	Support/solidarity in diabetes community	happy, birthday, day, #dsma, #t1d, insulin	M 0.27 F 0.44	0.62 0.03	HAPPY BIRTHDAY TO MY FAVORITE DIABETIC I'm glad we're close friends, come back soon I miss u:(I love you. . .
2	Inspiring relatives living with diabetes	love, t1d, amazing, #t1d, awesome	M 0.32 F 0.47	0.8 0.06	What a talented family! You were amazing yesterday. Thank you for being a great example for living beyond T1D diagnosis
3	Sharing hope and encouraging	hope, well, #dsma, get, soon, better	M 0.37 F 0.37	0.42 0.14	as a fellow t1d, your tweet made me smile. hope you get to say you no longer need batteries one day soon
4	Diabetes awareness / Support / Donation	help, love, please, us, let, awareness	M 0.30 F 0.43	0.55 0.06	**ATTENTION** Please come out & support T1D that Lynn Taylor is hosting! Help us find a cure for this disease 💙
5	DSMA enjoying online support	#dsma, good, :), glad, tonight, love	M 0.28 F 0.46	0.92 0.02	Thanks for a good chat everyone :) I'm going to get to bed early tonight. #DSMA
6	DSMA sharing diab. related stories	good, bad, news, really, know, like	M 0.37 F 0.32	0.32 0.12	Well they did say my legs are healing pretty good so that is a good sign now I need to see if I can get a primary doctor for my diabetes
7	T1D hashtags	#t1d, #diabetes, #type1diabetes, love	M 0.25 F 0.53	0.8 0.15	Big Day! I love you Jacob. #thisishome #hokies #HokieNation #T1D #Type1diabetes #diabetes — feeling proud
8	Diabetes care	care, take, health, insulin, one, taking	M 0.33 F 0.42	0.43 0.05	Taking care of your mental health is important. But dont eat for the first time in 3 days then make a thread about ur struggle w diabetes:/
9	Bloodsugar palette (beauty products)	#bloodsugar, palette, excited, love, got	M 0.14 F 0.63	0.06 0.06	@JeffreeStar is coming out with the new #thirsty palette and I'm still trying to get my hands on the #BloodSugar palette 😩😩😩
10	Advocacy for affordable insulin	#insulin4all, insulin, good, thank, love, us	M 0.18 F 0.70	0.95	I am THRILLED to see all of the publicity the insulin crisis is getting– let's keep it coming! #insulin4all
11	Life with and without diabetes	would, love, could, insulin, like, one	M 0.37 F 0.38	0.36 0.13	I dream what my life would be like without Type 1 Diabetes. I wouldn't be depressed. I wouldn't sick all the time or worry.
12	Life with type 1 diabetes	#t1d, #diabetes, love, #type1diabetes	M 0.25 F 0.51	0.72 0.16	All good things must come to an end. My first high blood sugar reading in 3 weeks. Oh well. #T1D #dsma #DOC #livebeyond
13	Glucose guardian	glucose, guardian, need, love, father	M 0.23 F 0.46	0.05 0.01	sometimes i wish i had a glucose guardian 😔 😔
14	Chatting about insulin	love, insulin, good, #dsma, one, glucose	M 0.34 F 0.39	0.35 0.06	Love seeing my family can't afford insulin and Cory Booker could out-freestyle Bernie Sanders on the timeline. Someone please kill me
15	Insulin pump complications	insulin, pump, love, hope, got, new, feel	M 0.31 F 0.47	0.53 0.01	It doesn't take five minutes, but I hate changing my insulin pump site and refilling the insulin reservoir. Maybe b. . .
16	Diabetes in family	family, runs, friends, know, type, history	M 0.29 F 0.42	0.35 0.06	My mom was with one of my sisters and diabetes runs on her side of the family so I'm worried lol just fo. . .
17	Diaversary	years, ago, two, type, one, insulin	M 0.36 F 0.42	0.49 0.25	Happy Diabirthday Kid. 11 years since dx. Remember when 'they' said there would be a cure in 5 years? tick.....tock. . .
18	Day-to-day stories about diabetes	day, today, insulin, glucose, good, last	M 0.29 F 0.47	0.45 0.08	YES! I remembered to test and take my morning insulin! 130 is a good reading fir me. All my diabetic family out...
19	Confusion between T1D and T2D	type, one, two, good, love, people, feel	M 0.32 F 0.40	0.53 0.4	#diabetes I swear. I have type 1, and nothing is more frustrating then when people try to give me some type 2 lecture
20	Diagnosis	diagnosed, one, two, type, since, year, old	M 0.32 F 0.47	0.62 0.17	Good evening, as many of you know, my daughter Kate was diagnosed with Type 1 Diabetes 3 years ago. Every year, we. . .
21	Glucose variability	glucose, blood, high, sugar, insulin, levels	M 0.34 F 0.38	0.19 0.17	Every time my blood sugar is good I think I can eat whatever without giving insulin, an hour later I'm 350 with moderate ketones 😅
22	Insulin	insulin, good, like, need, people, feel	M 0.37 F 0.38	0.35 0.05	Can't stress enough how much I need an insulin pump. It would help greatly!! I hate carrying around my insulin
23	Diabetes pop star Nick Jonas	nick, jonas, cried, remember, found	M 0.16 F 0.61	0.53 0.02	I still remember I cried at a Jonas brothers concert while nick was giving his speech about living with diabetes.
24	Misunderstandings of diabetes	people, know, like, one, understand	M 0.34 F 0.39	0.39 0.11	Idk if living alone with uncontrolled type one diabetes was a good idea. Hell even if it was controlled you just never know.
25	Frustration with insulin prices	insulin, afford, cost, cannot, insurance	M 0.36 F 0.44	0.55	My insulin is $1800. Utterly ridiculous! I can't afford mine either and so.etimes I have to go to an urgent care or ER for insulin
26	OGTT	glucose, test, hour, today, tomorrow	M 0.10 F 0.66	0.08 0.02	Taking my glucose tolerance test Monday and I honestly am dreading it
27	Parents and diabetes	mom, dad, got, died, insulin, one, bad, lost	M 0.33 F 0.43	0.28 0.1	Today, I was diagnosed with diabetes. My mom texted "I'm sorry, babygirl". I cried b/c I've seen her fight with this my whole entire life
28	Diabetes Distress	feel, like, hate, sick, really, bad, feeling	M 0.31 F 0.41	0.39 0.08	I know it's all I talk about but I hate this. I hate being diabetic so much I feel so contricted. Ranting about it makes it hurt less sorry
29	Diabetic/Insulin Shock	shock, insulin, went, going, coma, would	M 0.44 F 0.31	0.13 0.02	Diabetic shock at 23. I went into shock after eating three Klondike bars (I didn't even know I was d..
30	Diabetes-related comorbidities	anxiety, depression, pain, neuropathy, heart, cancer	M 0.34 F 0.35	0.25 0.1	Makes me fucking sick that as a T1 diabetic w a mother who has breast cancer father w heart issues and we lost insurance

* unknown type diabetes prediction not shown

For each topic the following information is provided: the topic label, the most frequent words, gender and type of diabetes distribution and a sample tweet.

Figure 4: Sentiment and emotion distributions of the 30 topics of interest of people with or talking about diabetes

No.	Topic label	% (N)	SA score	SA distr. neg=[-1,1]=pos	Primary emotions (%) joy‖love‖surprise‖sad‖anger‖fear					
1	Support / solidarity in diabetes community	1.3 (613)	0.68	−1 1	0.72	0.19	0.02	0.04	0.02	0.02
2	Inspiring relatives living with diabetes	3.3 (1552)	0.67	−1 1	0.36	0.31	0.21	0.05	0.03	0.05
3	Sharing hope and encouraging	1.2 (542)	0.54	−1 1	0.45	0.45	0.01	0.05	0.01	0.03
4	Diabetes awareness / Support / Donation	3.2 (1499)	0.48	−1 1	0.35	0.38	0.4	0.12	0.05	0.08
5	DSMA enjoying online support	3.1 (1459)	0.47	−1 1	0.48	0.26	0.06	0.09	0.04	0.07
6	DSMA sharing diabetes related stories	3.3 (1513)	0.36	−1 1	0.3	0.24	0.06	0.17	0.08	0.15
7	T1D hashtags	1.8 (812)	0.33	−1 1	0.37	0.38	0.06	0.08	0.07	0.05
8	Diabetes care	2.9 (1353)	0.28	−1 1	0.05	0.44	0.01	0.05	0.03	0.43
9	Bloodsugar palette (beauty product)	1.8 (818)	0.26	−1 1	0.28	0.49	0.03	0.13	0.02	0.04
10	Advocacy for affordable insulin	1.1 (505)	0.22	−1 1	0.24	0.29	0.09	0.15	0.13	0.1
11	Life with and without diabetes	2.7 (1239)	0.19	−1 1	0.31	0.26	0.06	0.16	0.08	0.13
12	Life with type 1 diabetes	3.3 (1514)	0.18	−1 1	0.3	0.27	0.04	0.16	0.11	0.11
13	Glucose guardian	0.7 (343)	0.16	−1 1	0.37	0.25	0.01	0.25	0.06	0.05
14	Chatting about insulin	5.0 (2302)	0.14	−1 1	0.32	0.23	0.04	0.18	0.12	0.11
15	Insulin pump complications	4.4 (2036)	0.09	−1 1	0.35	0.23	0.03	0.18	0.13	0.09
16	Diabetes in family	0.9 (436)	0.07	−1 1	0.22	0.26	0.03	0.18	0.1	0.22
17	Diaversary	3.6 (1685)	0.07	−1 1	0.28	0.2	0.06	0.21	0.11	0.14
18	Day-to-day stories about diabetes	5.7 (2666)	0.05	−1 1	0.28	0.17	0.04	0.21	0.15	0.15
19	Confusion between T1D and T2D	3.9 (1804)	0.04	−1 1	0.22	0.21	0.05	0.22	0.14	0.17
20	Diagnosis	2.2 (1023)	0.02	−1 1	0.24	0.21	0.06	0.21	0.13	0.16
21	Glucose variability	5.1 (2363)	0.01	−1 1	0.22	0.19	0.04	0.24	0.15	0.17
22	Insulin	8.6 (3970)	0.01	−1 1	0.27	0.18	0.05	0.2	0.14	0.16
23	Diabetes pop star Nick Jonas	0.5 (218)	-0.07	−1 1	0.19	0.19	0.01	0.49	0.04	0.09
24	Misunderstandings of diabetes	7.0 (3244)	-0.08	−1 1	0.16	0.17	0.04	0.29	0.14	0.2
25	Frustration with insulin prices	6.3 (2916)	-0.12	−1 1	0.15	0.17	0.03	0.24	0.21	0.2
26	OGTT	4.3 (2012)	-0.13	−1 1	0.2	0.1	0.02	0.31	0.18	0.19
27	Parents and diabetes	4.0 (1859)	-0.14	−1 1	0.16	0.17	0.04	0.29	0.14	0.21
28	Diabetes Distress	4.5 (2067)	-0.19	−1 1	0.14	0.11	0.04	0.24	0.27	0.21
29	Diabetic/Insulin Shock	0.6 (284)	-0.33	−1 1	0.03	0.02		0.03	0.45	0.46
30	Diabetes-related comorbidites	3.8 (1760)	-0.36	−1 1	0.09	0.08	0.02	0.4	0.19	0.22

For each topic the following information is provided: the topic label, the frequency of the topic, the average sentiment score, the sentiment distribution and the frequency of primary emotions.