



Insulin pricing and other major diabetes-related concerns in the USA: a study of 46 407 tweets between 2017 and 2019

Adrian Ahne, Francisco Orchard, Xavier Tannier, Camille Perchoux, Beverley Balkau, Sherry Pagoto, Jessica Lee Harding, Thomas Czernichow, Guy Fagherazzi

► To cite this version:

Adrian Ahne, Francisco Orchard, Xavier Tannier, Camille Perchoux, Beverley Balkau, et al.. Insulin pricing and other major diabetes-related concerns in the USA: a study of 46 407 tweets between 2017 and 2019. *BMJ Open Diabetes Research & Care*, 2020, 8 (1), pp.e001190. 10.1136/bmjdrc-2020-001190 . hal-02803719

HAL Id: hal-02803719

<https://hal.science/hal-02803719>

Submitted on 7 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Insulin pricing and other major diabetes-related concerns in the USA: a study of 46 407 tweets between 2017 and 2019

Adrian Ahne,^{1,2} Francisco Orchard,² Xavier Tannier,³ Camille Perchoux,⁴ Beverley Balkau,¹ Sherry Pagoto,⁵ Jessica Lee Harding,⁶ Thomas Czernichow,² Guy Fagherazzi ⁷

To cite: Ahne A, Orchard F, Tannier X, *et al.* Insulin pricing and other major diabetes-related concerns in the USA: a study of 46 407 tweets between 2017 and 2019. *BMJ Open Diab Res Care* 2020;**8**:e001190. doi:10.1136/bmjdr-2020-001190

► Additional material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjdr-2020-001190>).

Received 15 January 2020
Revised 24 March 2020
Accepted 4 April 2020



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to
Dr Guy Fagherazzi;
Guy.Fagherazzi@lih.lu

ABSTRACT

Introduction Little research has been done to systematically evaluate concerns of people living with diabetes through social media, which has been a powerful tool for social change and to better understand perceptions around health-related issues. 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 USA) and to identify clusters of tweets with emotional elements. A sentiment analysis was then applied to each cluster.

Results We identified 46 407 tweets with emotional elements in the USA from which 30 clusters were identified; 5 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, 5 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 shock (45% *anger*, 46% *fear*) and comorbidities (40% *sadness*).

Conclusions Using social media data, we have been able to describe key diabetes-related concerns and their associated emotions. More specifically, we were able to highlight the real-world concerns of insulin pricing and its negative impact on mood. Using such data can be a useful addition to current measures that inform public decision making around topics of concern and burden among people with diabetes.

INTRODUCTION

Stress, fears and negative emotions are considered to be the most important psychosocial health factor in the management of diabetes.¹ Diabetes distress is associated with decisional conflict and therefore has an impact on day-to-day disease management

Significance of this study

What is already known about this subject?

- It is very challenging to collect representative data at a population level to understand what are the key concerns of people with diabetes in real life.
- Social media platforms, such as Twitter, may serve as a relevant source of information to supplement traditional population health studies.
- There are worldwide inequalities in access to insulin.

What are the new findings?

- With 18% of the tweets related to insulin pricing, this 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 lot 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 information about key diabetes-related concerns, which is usually not available in typical clinical or epidemiological studies.
- This information can supplement data from clinical or epidemiological studies to inform public health strategies to deal with diabetes-related prevention, management and treatment.

and the long-term risk of diabetes-related complications.² Reducing diabetes-related distress may improve hemoglobin A1c and reduce the burden of disease among people with diabetes.^{3 4} However, there is still limited knowledge about the sources of stress, anxiety and concerns among people with diabetes and it is difficult to capture them with existing evaluation scales such as the Problem Areas in Diabetes Scale (PAID

or PAID-T for adolescents) or the Diabetes Distress Scale (DDS).^{5–9}

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 in particular on Twitter.^{10–11} Twitter is a microblogging and social networking service with 321 million monthly active users.¹²

Using social media to explore psychological information related to diabetes and associations between socioeconomic factors and diabetes-related concerns based in real-life is currently uncharted territory. However, using this resource may provide important insights and allow future interventions regarding prevention, management and treatment of diabetes to be more appropriately tailored.

Using Twitter data, the aims of the current study are 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 >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 online supplementary materials 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, users agree to have this information available to the general public.¹³ In this study, we restricted the analysis to 46 407 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 online supplementary material S2. For the present work, the statistical unit considered is a tweet and not a user profile. Thus a user may have several tweets that are included.

Identifying emotions

To identify emotions among diabetes-related tweets, we combined detailed Parrot's classification of emotions with a dictionary of emotional keywords present in the two most common tools to assess psychological health in people with diabetes, namely the PAID questionnaire and the DDS.¹⁴ Parrot identified over 100 emotions and conceptualized them as a tree structured list (full list in online supplementary material 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

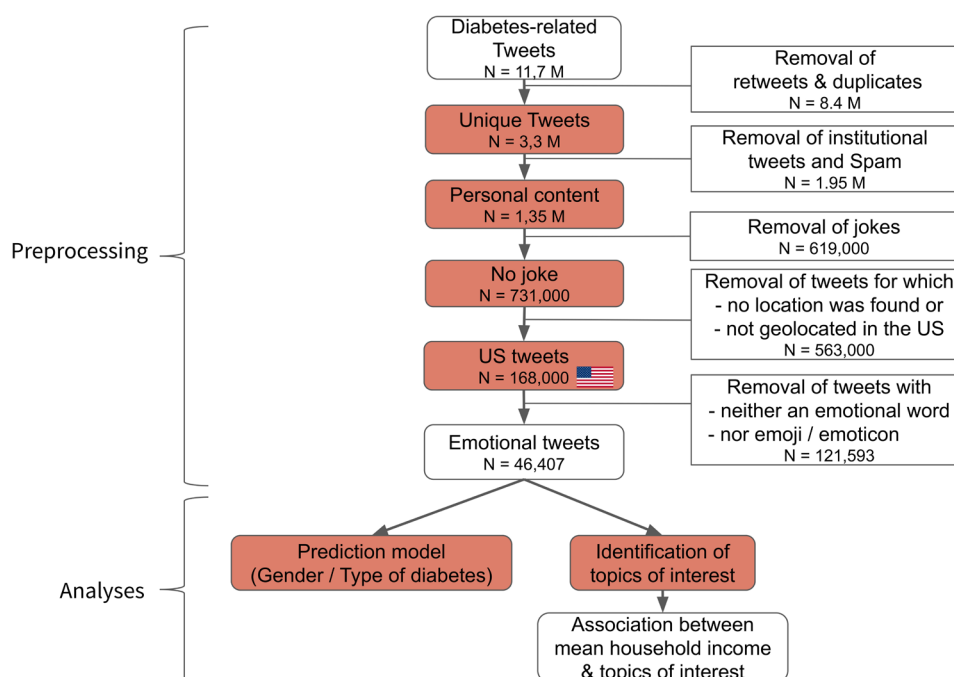


Figure 1 Workflow: from the data collection to the analysis. Red boxes correspond to steps where machine learning methods were applied.

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 similar to the six primary emotions defined by Parrot.¹⁷ This allowed us to determine a subset of *emotional* tweets containing either emotional words or emojis/emoticons. Emojis reproduced with permission from Twitter under CC-BY 4.0: <https://creativecommons.org/licenses/by/4.0/>

Sentiment analysis

Sentiment analysis is commonly used to investigate the positive or negative opinion within a tweet. We used a widely accepted human-validated sentiment analysis tool Valence Aware Dictionary for Sentiment Reasoning (VADER).^{18–21} 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 as our main metric for the sentiment analysis (referred to as the *SA score* later in the text), which is unidimensional and a normalized measure of sentiment between –1 (most negative) and +1 (most positive).

Topic extraction

All tweets are represented via their word vector representations, with the consequence that words similar in semantics are also similar in the word vector space (see online supplementary material S2 for more details). We applied the unsupervised machine learning algorithm K-means, using cosine similarity as distance measure, to group the emotional tweets that are close to each other based on their word vector representations into topics/clusters.²² In order to define the right number of clusters *k*, the Silhouette score, measuring how on average each data point is closer to its cluster’s center than to any other cluster, was used as input parameter for the K-means algorithm.²³ We obtained the highest score for *k*=30 clusters. All tweets were then assigned to one of the 30 clusters/topics. Each topic/cluster was then given a label by two authors (AA, GF) according to the 10 most contributing tweets (those closest to the topic center) and the most frequent words (top words) in the cluster.

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—(US\$24 609, US\$67 224), medium income—(US\$67 225, US\$86 758), high income—(US\$86 759, US\$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 χ^2 tests. For each topic, the p value was calculated using a χ^2 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 (V.3.6) and the packages scikit-learn (machine learning algorithms and data preprocessing methods) and gensim (text processing, word representation) were used to implement the methods described above.^{25,26} 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 (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: 9663), Los Angeles (N: 5301) 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, 4375 (9%) from type 2 diabetes and 21 747 (47%) diabetes type unknown. Table 1 shows the detailed description of the 30 topics of interest about diabetes.

In most topics, an over-representation of women was 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 6) and ‘diabetic/insulin shock’ (topic 29). Discussions about ‘advocacy for affordable insulin’ (topic 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. The Diabetes Social Media Advocacy (#DSMA) group on Twitter is a

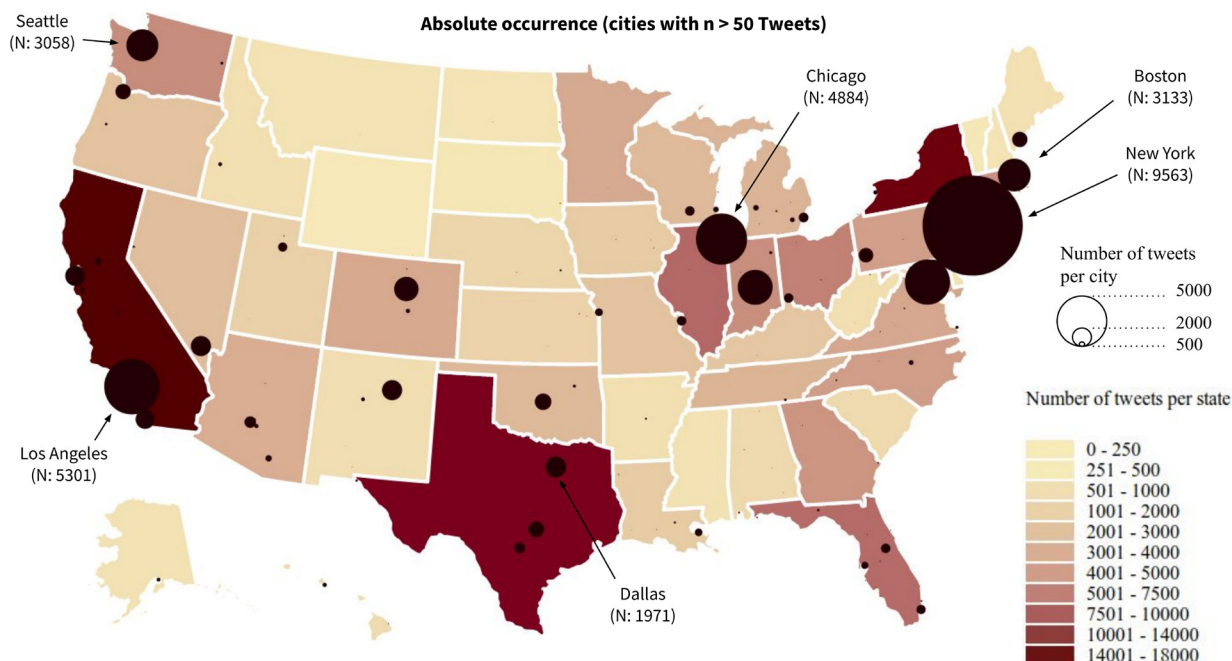


Figure 2 Spatial distribution of the diabetes-related tweets over the USA. The number of tweets in each state is represented by color ranging from few tweets in white-yellow to a lot of tweets in dark red. The size of the black circles is proportional to the number of tweets in a given city.

community of people exchanging about diabetes-related topics, in which the price of insulin is a central concern.

Primary emotions related to the topics of interest

Table 2 illustrates the sentiment and emotions distribution over the 30 topics.

In online supplementary material 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 1) with 72% of emotions related to *joy* and about inspiring relatives living with diabetes (topic 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 29) with an over-representation of *anger* (45% of all emotions) and *fear* (46%) and 'diabetes distress' (topic 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 percentage 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 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', a gender neutral term for a person who is giving money, often to buy insulin, in exchange for some favor or something in return. 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 10 'advocacy for affordable insulin', 22: 'insulin' and 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 online supplementary material S5). 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: 'glycemic instability' ($p < 0.05$) 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)'

Table 1 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(%)			DiabetesType(%)			Tweet description
1	Support/Solidarity in diabetes community	happy, birthday, day, #dsma, #t1d, insulin	29 M	42 F	29 U	35 T1	6 T2	59 U	► Happy birthday wishes ► Affection / Support / Solidarity messages
2	Inspiring relatives living with diabetes	love, t1d, amazing, #t1d, awesome	32 M	47 F	21 U	80 T1	6 T2	15 U	► Inspiring friends and families when living with diabetes or helping to live a better life ► Pointing to people campaigning for diabetes issues and affordable insulin
3	Sharing hope and encouraging	hope, well, #dsma, get, soon, better	37 M	37 F	26 U	42 T1	14 T2	44 U	► Encouraging people with diabetes ► Hoping to find a cure for diabetes
4	Diabetes awareness / Support / Donation	help, love, please, us, let, awareness	30 M	43 F	27 U	55 T1	6 T2	39 U	► Supporting each other within the diabetes community ► Raising awareness for diabetes and its complications ► Donations to get insulin
5	DSMA* enjoying online support	#dsma, good, :, glad, tonight, love	28 M	46 F	26 U	92 T1	2 T2	6 U	► Friendly exchanges within the DSMA* online community ► Being thankful for the people's effort and activity ► Questions within the DSMA online community
6	Sharing diabetes related stories	good, bad, news, really, know, like	37 M	32 F	30 U	32 T1	12 T2	56 U	► People telling stories about diabetes or diabetes treatment ► Messages about daily struggle with diabetes
7	T1D hashtags	#t1d, #diabetes, #type1diabetes, love	25 M	53 F	22 U	80 T1	15 T2	5 U	Mainly short, incoherent messages with several hashtags relating to type one diabetes, such as #t1d, #t1dlife, #t1dmom, or #Dexcom
8	Diabetes care	care, take, health, insulin, one, taking	33 M	42 F	25 U	43 T1	5 T2	52 U	► Talking about diabetes care or mental health with diabetes ► Indignation about the healthcare system
9	Bloodsugar palette (beauty products)	#bloodsugar, got, palette, excited, love,	14 M	63 F	23 U	6 T1	6 T2	88 U	"Spam" about the beauty product "Blood Sugar Palette" palette
10	Advocacy for affordable insulin	#insulin4all, insulin, good, thank, love, us	18 M	70 F	13 U	95 T1	0 T2	5 U	► Encouraging people advocating for affordable insulin ► Talking about healthcare activism
11	Life with and without diabetes	would, love, could, insulin, like, one	37 M	38 F	25 U	36 T1	13 T2	51 U	► Hypothetical life without diabetes ► Frustration about how people without diabetes poorly understood their conditions
12	Life with type one diabetes	#t1d, #diabetes, love, #type1diabetes	25 M	51 F	24 U	72 T1	16 T2	12 U	► Messages about Dexcom devices ► Seeking for encouragement ► Life with type one diabetes
13	Glucose guardian	glucose, guardian, need, love, father	23 M	46 F	30 U	5 T1	1 T2	94 U	Looking for glucose guardians. This is a gender neutral term for a person who is giving money, often to buy insulin, in exchange for some favor or something in return.
14	Chatting about insulin	love, insulin, #dsma good, glucose, one	34 M	39 F	27 U	35 T1	6 T2	59 U	Talking about insulin

Continued

Table 1 Continued

15	Insulin and insulin pump complications	insulin, pump, love, hope, got, new, feel	31 M	47 F	22 U	53 T1	1 T2	45 U	► Unhappiness about having to manage the insulin pump ► Insulin had no effect or was bad
16	Diabetes in family	family, runs, friends, know, type, history	29 M	42 F	29 U	35 T1	6 T2	59 U	► Worries about getting diabetes because of family predisposition ► Family support
17	Diaversary	years, ago, two, type, one, insulin	36 M	42 F	22 U	49 T1	25 T2	26 U	Celebrating birthday of diabetes diagnosis
18	Day-to-day stories about diabetes	day, today, insulin, glucose, good, last	29 M	47 F	24 U	45 T1	8 T2	48 U	Day-to-day stories about diabetes or insulin adjustment
19	Confusion between T1D and T2D	type, one, two, good, love, people, feel	32 M	40 F	28 U	53 T1	40 T2	7 U	Frustration about people's inability to distinguish between type 1 and type two diabetes
20	Diagnosis	diagnosed, type, old, two, since, year, one	32 M	47 F	21 U	62 T1	17 T2	21 U	Talking about their own diabetes diagnosis or the one of their child
21	Glycemic instability	glucose, blood, high, sugar, insulin, levels	34 M	38 F	28 U	19 T1	17 T2	63 U	► Difficulty to keep blood sugar levels stable ► Suffering from low or high blood sugar - Reliefs about keeping blood sugar levels normal
22	Insulin	insulin, good, like, need, people, feel	37 M	38 F	25 U	35 T1	5 T2	60 U	► Talking about expensive insulin and people who cannot afford treatment ► Insecurity about insulin
23	Diabetes pop star Nick Jonas	nick, jonas, cried, remember, found	16 M	61 F	24 U	53 T1	2 T2	45 U	Talking about pop star Nick Jonas who has type one diabetes and wrote a song about it
24	Misunderstandings of diabetes	people, know, like, one, understand	34 M	39 F	28 U	39 T1	11 T2	51 U	► Complaining about the public view on diabetes ► Unclear relationship for many people between food and blood sugar / diabetes
25	Frustration with insulin prices	insulin, afford, cost, cannot, insurance	36 M	44 F	20 U	55 T1	0 T2	44 U	► Frustration with high insulin prices ► Insurance as elementary factor to get insulin
26	OGTT**	glucose, test, hour, today, tomorrow	10 M	66 F	24 U	8 T1	2 T2	90 U	Dreading glucose drink and glucose tests (mostly during pregnancy)
27	Parents and diabetes	mom, dad, got, died, insulin, one, bad, lost	33 M	43 F	24 U	28 T1	10 T2	62 U	Parents-children stories related to diabetes
28	Diabetes Distress	feel, like, hate, sick, really, bad, feeling	31 M	41 F	28 U	39 T1	8 T2	53 U	► Sharing feelings of sadness and depressive symptoms ► Feeling scared or anxious about insulin use ► Hate towards own diabetes
29	Diabetic/Insulin Shock	shock, insulin, went, going, coma, would	44 M	31 F	25 U	13 T1	2 T2	85 U	► Going into diabetic shock after eating sweet food ► Sharing experience of going into insulin shocks
30	Diabetes-related comorbidities	anxiety, depression, pain, neuropathy, heart, cancer	34 M	35 F	31 U	25 T1	10 T2	65 U	Talking about diabetes-related complications such as depression, anxiety, heart diseases, nerve pain, chronic pain and other health issues

For each topic the following information is provided: the topic label, the most frequent words, gender (M=men, F=women, U=unknown) and type of diabetes (T1=type 1 diabetes, T2=type 2 diabetes, U=unknown) distribution and a tweet description.

DSMA, Diabetes Social Media Advocacy online group; OGTT, oral glucose tolerance test.

Table 2 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		0.72 0.19 0.02 0.04 0.02 0.02
2	Inspiring relatives living with diabetes	3.3 (1552)	0.67		0.36 0.31 0.21 0.05 0.03 0.05
3	Sharing hope and encouraging	1.2 (542)	0.54		0.45 0.45 0.01 0.05 0.01 0.03
4	Diabetes awareness / support / donation	3.2 (1499)	0.48		0.35 0.38 0.4 0.12 0.05 0.08
5	DSMA enjoying online support	3.1 (1459)	0.47		0.48 0.26 0.06 0.09 0.04 0.07
6	Sharing diabetes-related stories	3.3 (1513)	0.36		0.3 0.24 0.06 0.17 0.08 0.15
7	T1D hashtags	1.8 (812)	0.33		0.37 0.38 0.06 0.08 0.07 0.05
8	Diabetes care	2.9 (1353)	0.28		0.05 0.44 0.01 0.05 0.03 0.43
9	Blood sugar palette (beauty products)	1.8 (818)	0.26		0.28 0.49 0.03 0.13 0.02 0.04
10	Advocacy for affordable insulin	1.1 (505)	0.22		0.24 0.29 0.09 0.15 0.13 0.1
11	Life with and without diabetes	2.7 (1239)	0.19		0.31 0.26 0.06 0.16 0.08 0.13
12	Life with T1D	3.3 (1514)	0.18		0.3 0.27 0.04 0.16 0.11 0.11
13	Glucose guardian	0.7 (343)	0.16		0.37 0.25 0.01 0.25 0.06 0.05
14	Chatting about insulin	5.0 (2302)	0.14		0.32 0.23 0.04 0.18 0.12 0.11
15	Insulin and insulin pump complications	4.4 (2036)	0.09		0.35 0.23 0.03 0.18 0.13 0.09

Continued

Table 2 Continued

No.	Topic label	% (N)	SA score	SA distr. neg=(-1,1)=pos	Primary emotions (%)					
					joy	love	surprise	sad	anger	fear
16	Diabetes in family	0.9 (436)	0.07		0.22	0.26	0.03	0.18	0.1	0.22
17	Diaversary	3.6 (1685)	0.07		0.28	0.2	0.06	0.21	0.11	0.14
18	Day-to-day stories about diabetes	5.7 (2666)	0.05		0.28	0.17	0.04	0.21	0.15	0.15
19	Confusion between T1D and T2D	3.9 (1804)	0.04		0.22	0.21	0.05	0.22	0.14	0.17
20	Diagnosis	2.2 (1023)	0.02		0.24	0.21	0.06	0.21	0.13	0.16
21	Glycemic instability	5.1 (2363)	0.01		0.22	0.19	0.04	0.24	0.15	0.17
22	Insulin	8.6 (3970)	0.01		0.27	0.18	0.05	0.2	0.14	0.16
23	Diabetes pop star Nick Jonas	0.5 (218)	-0.07		0.19	0.19	0.01	0.49	0.04	0.09
24	Misunderstandings of diabetes	7.0 (3244)	-0.08		0.16	0.17	0.04	0.29	0.14	0.2
25	Frustration with insulin prices	6.3 (2916)	-0.12		0.15	0.17	0.03	0.24	0.21	0.2
26	OGTT	4.3 (2012)	-0.13		0.2	0.1	0.02	0.31	0.18	0.19
27	Parents and diabetes	4.0 (1859)	-0.14		0.16	0.17	0.04	0.29	0.14	0.21
28	Diabetes distress	4.5 (2067)	-0.19		0.14	0.11	0.04	0.24	0.27	0.21
29	Diabetic / Insulin shock	0.6 (284)	-0.33		0.03	0.02	0.03	0.45	0.46	
30	Diabetes-related comorbidities	3.8 (1760)	-0.36		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 analysis score (SA score), the sentiment analysis distribution (SA distr.) and the frequency of primary emotions.
DSMA, Diabetes Social Media Advocacy online group; OGTT, oral glucose tolerance test; T1D, type 1 diabetes; T2D, type 2 diabetes.

($p < 0.001$) 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 and insulin pump complications', were more frequently ($p < 0.001$).

DISCUSSION

Our findings suggest that Twitter is a useful tool to capture key diabetes-related topics and the emotions associated with those topics. Our key findings suggest that there is a lot of support and solidarity among the diabetes online community with numerous tweets related to *joy* and *love* emotions observed. In contrast, Twitter users expressed fear, anger and sadness related to insulin pricing and diabetes-related complications and comorbidities, as well as considerable frustration about the inability of people to distinguish between type 1 and type 2 diabetes.

To our knowledge, this is the first study using social media data to capture information regarding key diabetes-related concerns. We are, therefore, unable to compare and contrast the results of this study with others. Nonetheless, the importance of understanding emotions and self-control (regulation of thoughts, emotions and behavior) for health outcomes in people with diabetes has been previously documented in several other studies.²⁷ Hagger *et al* reviewed diabetes distress among adolescents with type 1 diabetes and found a substantial proportion experienced elevated diabetes distress and that it is often associated with suboptimal glycemic control.²⁸ Richman *et al* also suggested that positive emotions such as hope and curiosity may play a protective role in the development of disease.²⁹ Ogbera *et al* showed that higher levels of emotional distress are associated with poor self-care in type 2 diabetes.³⁰ Another 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*, 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.³¹

Insulin pricing

We found that insulin pricing was a major concern among tweets shared in the USA (18% of all tweets were related to insulin pricing) and is associated with both positive (*joy*, *love*) and negative (*sadness*, *anger*, *fear*) emotions. People frequently shared their frustration with insulin prices, access to insulin and identifying sources of insulin including 'glucose guardians' or donations, which represent major obstacles for people with diabetes.^{32,33} Positive emotions are present when it comes to solidarity in

the fight for affordable insulin in the community. We observed associations of topics addressing insulin pricing to be more frequent in cities with high mean incomes. This does not necessarily indicate that people living in cities with a high mean household income feel more concerned about insulin prices, but rather they probably have a greater ability to tweet around this issue. A large number of tweets geolocated in cities with a high mean household income included the hashtag '#insulin4all', a campaign that unites the diabetes community around the access to treatment for everyone.³⁴

It is known that there are key challenges for a global and fair access to insulin.³⁵ 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.

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, and in real-time, in what can be considered as an open digital space with flat role hierarchy for information sharing and development of online communities. This potentially minimizes 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 large variability in their profiles. The methodologies developed in this study present an innovative way to concentrate on relevant (personal, emotional) geolocated tweets (USA), to identify topics of interest and emotions shared within topics. This approach is able to capture trends in the online diabetes community as well as socioeconomic factors that can be associated with social media data at the ecological level. This new way of capturing data supplements the detection of topics which are less medically oriented.

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 social media profiles, a key strength in digital epidemiology.³⁶ While we did observe large variability in our Twitter profiles, we found an over-representation of people with type 1 diabetes and women in our study when compared with known diabetes epidemiology literature.³⁷ The greater representation of type 1 diabetes may be explained by the younger demographics of Twitter users.³⁸ Alternatively, type 1 diabetes may have more involved care, more devices, more challenging medication and more frustrations to report on Twitter as compared with type 2 diabetes. Regardless, our results should be interpreted in the context of the Twitter population only. Second, 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 often impossible to define the sex or type of diabetes. Third, we were unable to account for several clinical and environmental factors that may help to 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. Fourth, the geolocation of tweets was partially based on locations the users provided, which might not be their true location. Fifth, emotion detection is still a challenge in the machine learning field due to the occurrence of sarcasm and irony. It is one of the open research questions. Last, causal inference between the mean household income per city and the topics of interest of people residing in the corresponding city cannot be made as it is subject to ecological fallacy.

A 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, anger and fear. We have shown that using social media posts to capture emotions and concerns of individuals in real life is feasible, and is an efficient way of augmenting psychosocial, behavioral and epidemiological research. Our work should encourage future studies to consider social media data as online supplementary information.

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. Social media data will help develop policies and interventions that consider key concerns among people with diabetes to ultimately improve health outcomes.

Author affiliations

¹Center for Research in Epidemiology and Population Health (CESP), INSERM, University Paris Saclay, Villejuif (Paris), Île-de-France, France

²Epicconcept Company, Paris, France

³LIMICS, INSERM U1142, Sorbonne University, Paris, Île-de-France, France

⁴Luxembourg Institute of Socio-Economic Research, Esch-sur-Alzette, Luxembourg

⁵Department of Allied Health Sciences, UConn Center for mHealth & Social Media, University of Connecticut, Storrs, Connecticut, USA

⁶Division of Transplantation, Department of Surgery, Emory University School of Medicine, Emory University Hospital, Atlanta, Georgia, USA

⁷Digital Epidemiology Hub, Department of Population Health, Luxembourg Institute of Health, Strassen, Luxembourg

Contributors GF 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: GF designed the research; AA and GF conducted the research; FO and

AA collected the data; AA, FO and GF analyzed data; AA, GF, FO, TC, XT interpreted the data; AA and GF drafted the article; GF, FO, TC, XT, BB, JLH, CP and SP revised the manuscript critically; GF had primary responsibility for the final content of the manuscript. All authors read and approved the final manuscript.

Funding 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 or the interpretation of the results of the present study. AA, FO and TC are supported by Epicconcept Company. Epicconcept was involved in the data collection and writing of the report. No study sponsor influenced the decision to submit the paper for publication.

Map disclaimer The depiction of boundaries on the map(s) in this article do not imply the expression of any opinion whatsoever on the part of BMJ (or any member of its group) concerning the legal status of any country, territory, jurisdiction or area or of its authorities. The map(s) are provided without any warranty of any kind, either express or implied.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. All code is publicly available in this github repository (<https://github.com/WDDS/Tweet-Diabetes-Classification>).

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Guy Fagherazzi <http://orcid.org/0000-0001-5033-5966>

REFERENCES

- 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. *Lancet Diabetes Endocrinol* 2017;5:935–6.
- 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* 2019;42:1170–7.
- 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;42:841–8.
- 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* 2019;42:1669–74.
- Polonsky WH, Fisher L, Earles J, *et al*. Assessing psychosocial distress in diabetes: development of the diabetes distress scale. *Diabetes Care* 2005;28:626–31.
- 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* 2018;23:667–80.
- 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;33:835–43.
- 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;35:259–64.
- Polonsky WH, Anderson BJ, Lohrer PA, *et al*. Assessment of diabetes-related distress. *Diabetes Care* 1995;18:754–60.
- Bevan A. T1D exchange: an online community for people Touched by type 1 diabetes. *AADE Pract* 2017;5:44–6.
- 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;15:408–15.
- Number of monthly active Twitter users worldwide from 1st quarter 2010 to 1st quarter 2019 (in millions). Statista. Available: <https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/> [Accessed 28 Oct 2019].
- Twitter privacy policy, Twitter URL, 2018. Available: <https://twitter.com/en/privacy> [Accessed 23 Jul 2019].

- 14 Gerrod Parrott W. *Emotions in Social Psychology: Essential Readings*. Psychology Press, 2001.
- 15 Princeton University. Princeton University 'About WordNet, 2010. Available: <https://wordnet.princeton.edu/> [Accessed 10 Apr 2019].
- 16 Comesaña M, Soares AP, Perea M, et al. Erp correlates of masked affective priming with emoticons. *Comput Human Behav* 2013;29:588–95.
- 17 Wolny W. Emotion analysis of Twitter data that use Emoticons and Emoji Ideograms. *25th international conference on information systems development*.
- 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 Sci* 2016;5.
- 20 Tamersoy A, De Choudhury M, Chau DH. Characterizing smoking and drinking abstinence from social media. *HT ACM Conf Hypertext Soc Media* 2015;2015:139–48.
- 21 Eric C. Vader: a parsimonious rule-based model for sentiment analysis of social media text. Ann Arbor, MI: Eighth International Conference on Weblogs and Social Media (ICWSM-14), 2014.
- 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: 281–97.
- 23 Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* 1987;20:53–65.
- 24 U.S. census bureau; 2013-2017 American community survey 5-year estimates using American FactFinder. Available: <http://factfinder.census.gov> [Accessed 12 Aug 2019].
- 25 Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in python. *J Mach Learn Res* 2011;12:2825–30.
- 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 2010:45–50.
- 27 Hughes AE, Berg CA, Wiebe DJ. Emotional processing and self-control in adolescents with type 1 diabetes. *J Pediatr Psychol* 2012;37:925–34.
- 28 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.
- 29 Richman LS, Kubzansky L, Maselko J, et al. Positive emotion and health: going beyond the negative. *Health Psychol* 2005;24:422–9.
- 30 Ogbera A, Adeyemi-Doro A. Emotional distress is associated with poor self care in type 2 diabetes mellitus. *J Diabetes* 2011;3:348–52.
- 31 Nguyen QC, Brunisholz KD, Yu W, et al. Twitter-derived neighborhood characteristics associated with obesity and diabetes. *Sci Rep* 2017;7:16425.
- 32 Conner F, Pfister E, Elliott J, et al. Unaffordable insulin: patients pay the price. *Lancet Diabetes Endocrinol* 2019;7:748.
- 33 Fralick M, Kesselheim AS. The U.S. Insulin Crisis - Rationing a Lifesaving Medication Discovered in the 1920s. *N Engl J Med* 2019;381:1793–5.
- 34 T1International insulin4all. t1international. Available: <https://www.t1international.com/insulin4all/> [Accessed 16 Mar 2020].
- 35 Luo J, Gonsalves G, Greene J. Insulin for all: treatment activism and the global diabetes crisis. *Lancet* 2019;393:2116–7.
- 36 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.
- 37 International Diabetes Federation. *IDF diabetes atlas*. 9th edn. Brussels, Belgium, 2019. <https://www.diabetesatlas.org/en/>
- 38 Percentage of U.S. adults who use Twitter as of February 2019, by age group. Statista. Available: <https://www.statista.com/statistics/265647/share-of-us-internet-users-who-use-twitter-by-age-group/> [Accessed 28 Oct 2019].