Identifying causal associations in tweets using deep learning: Use case on diabetes-related tweets from 2017-2021

Authors

Adrian Ahne\textsuperscript{1,2}, Vivek Khetan\textsuperscript{3}, Xavier Tannier\textsuperscript{4}, Md I. H. Rizvi\textsuperscript{5}, Thomas Czernichow\textsuperscript{2}, Francisco Orchard\textsuperscript{2}, Charline Bour\textsuperscript{6}, Andrew Fano\textsuperscript{3}, Guy Fagherazzi\textsuperscript{6}

1. Paris-Saclay University, UVSQ, Inserm, Gustave Roussy, “Exposome and Heredity” team, CESP, F-94805, Villejuif, France
2. Epiconcept Company, Paris, France, (a.ahne@epiconcept.fr, t.czernichow@epiconcept.fr, f.orchard@epiconcept.fr)
3. Accenture Labs, San Francisco, USA, (vivek.a.khetan@accenture.com, andrew.e.fano@accenture.com)
4. Sorbonne University, Inserm, University Sorbonne Paris Nord, Laboratoire d’Informatique Médicale et d’Ingénierie des Connaissances pour la e-Santé, LIMICS, Paris, France, (xavier.tannier@sorbonne-universite.fr)
5. Indian Institute of Science, Bengaluru, India, (imbugene@gmail.com)
6. Deep Digital Phenotyping Research Unit, Department of Precision Health, Luxembourg Institute of Health, Strassen, Luxembourg, (Guy.Fagherazzi@lih.lu, Charline.Bour@lih.lu)

Corresponding author: Adrian Ahne, a.ahne@epiconcept.fr

Keywords: Causality, deep learning, natural language processing, Diabetes, Social media

Word count: 3999
Abstract

**Objective**
Leveraging machine learning methods, we aim to extract both explicit and implicit cause-effect associations in patient-reported, diabetes-related tweets and provide a tool to better understand opinion, feelings and observations shared within the diabetes online community from a causality perspective.

**Materials and Methods**
More than 30 million diabetes-related tweets in English were collected between April 2017 and January 2021. Deep learning and natural language processing methods were applied to focus on tweets with personal and emotional content. A cause-effect-tweet dataset was manually labeled and used to train 1) a fine-tuned Bertweet model to detect causal sentences containing a causal association 2) a CRF model with BERT based features to extract possible cause-effect associations. Causes and effects were clustered in a semi-supervised approach and visualised in an interactive cause-effect-network.

**Results**
Causal sentences were detected with a recall of 68% in an imbalanced dataset. A CRF model with BERT based features outperformed a fine-tuned BERT model for cause-effect detection with a macro recall of 68%. This led to 96,676 sentences with cause-effect associations. “Diabetes” was identified as the central cluster followed by “Death” and “Insulin”. Insulin pricing related causes were frequently associated with “Death”.

**Conclusions**
A novel methodology was developed to detect causal sentences and identify both explicit and implicit, single and multi-word cause and corresponding effect as expressed in diabetes-related tweets leveraging BERT-based architectures and visualised as cause-effect-network. Extracting causal associations on real-life, patient reported outcomes in social media data provides a useful complementary source of information in diabetes research.
Introduction

In recent years, the amount of available textual data has been growing exponentially through the emergence of social media allowing health researchers to access real-life data, study new risk factors and patient reported outcomes such as mental health problems, cerebral aneurysms, obesity and diabetes, the detection of suicide risk factors or adverse drug reactions. Due to the availability of a large volume of data and the potentiality of discovering unknown etiological results, researchers are interested in identifying causal relations in text (e.g. tweets), specifically causes of health problems, concerns and symptoms. Causal relation extraction in natural language text plays a key role in clinical decision-making, biomedical knowledge discovery or emergency management. Particularly, causal relations on Twitter have been examined for diverse factors causing stress and relaxation, adverse drug reactions or causal associations related to insomnia or headache.

In this paper, we aim to detect potential causes and corresponding effects in diabetes-related tweets collected in the context of the World Diabetes Distress Study (WDDS). The WDDS aims to analyze what is shared on social media worldwide to better understand what people with diabetes and diabetes distress are experiencing. The term diabetes distress (DD) refers to psychological factors such as emotional burden, worries, frustration or stress in the day-to-day management of diabetes. Diabetes distress is associated with poorer quality of life, higher A1C levels and medication adherence. Social media is a useful observatory resource for diabetes issues and could help to contribute directly to public and clinical decision making, given the active online diabetes community.

Most approaches examine explicit causality in text, when cause and effect are explicitly stated, for instance by connective words (e.g. so, hence, because, lead to, since, if-then). An example for an explicit cause-effect pair is “diabetes causes hypoglycemia”. Whereas, implicit causality is more complicated to detect such as in “Cannot sleep #insomnia, #overthinking” with cause “#overthinking” and effect “Cannot sleep”.

Machine and deep learning models have also been applied to extract causal relations. They are able to explore implicit relations and provide better generalisation contrary to rule-based approaches. An interesting approach, leveraging the transfer learning paradigm, and addressing both explicit and implicit cause-effect extraction is provided by Khetan et al. They fine-tuned pre-trained transformer based BERT language models to detect “Cause-Effect” relationships using publicly available datasets such as the adverse drug effect dataset.
In a similar spirit, the objective of the present work is to identify both explicit and implicit multi-word cause-effect relations on noisy, diabetes-related tweets, to aggregate identified causes and effects in clusters and ultimately to visualise these clusters in an interactive cause-effect network.

Materials and Methods

On the basis of diabetes-related tweets, we first preprocessed tweets to only focus on tweets with personal, no-joke and emotional content; secondly, we identified tweets in which causal information (opinion, observation, etc.) is communicated, also referred to as causal tweets or causal sentences, and in a third step, causes and their corresponding effects were extracted. Lastly, those cause-effect pairs were aggregated, described and visualised. The entire workflow is illustrated in Figure 1.
Data collection

Via Twitter’s Streaming Application Programming Interface (API), 32 million diabetes-related tweets in English were collected between April 2017 and January 2021 based on a list of diabetes-related keywords, such as diabetes, hypoglycemia, hyperglycemia and insulin, from all over the world (see Supplementary File S1 for the full list of keywords used). This is an extended dataset of the one used in earlier works. 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.

Data preprocessing

In this work, we applied a preprocessing pipeline, similar to earlier works, to focus on tweets with personal content and remove institutional tweets (organisations, advertisement, news, etc.); identify and exclude jokes; and filter tweets containing emotional elements to adjust the scope of the tweets towards diabetes distress. Besides, questions were removed. This led to 562,013 tweets containing personal, non-joke and emotional content. More details on the preprocessing pipeline are summarized in the Supplementary File S2.

Data annotation

In order to identify causal tweets and cause-effect association, 5,000 randomly chosen diabetes-related tweets were manually labeled. We did not restrict ourselves to a specific area of diabetes-related causal relationships and include potentially all types.

Table 1 illustrates some example tweets. For a more detailed explanation on the annotation please refer to our annotation guidelines in Supplementary File S3.

Labeling cause-effect pairs is a complex task. To verify the reliability of the labeling, two authors labeled 500 tweets independently and we calculated Cohen’s kappa score, a statistical measure expressing the level of agreement between two annotators. We obtained a score of 0.83, which is interpreted as an almost perfect agreement according to Altman and Landis. Disagreements were discussed between the two authors and one author labelled additional 4,500 tweets, resulting in 5,000 labeled tweets.
| Tweet                                                                 | Cause                  | Effect                  | C.A.* | Explanation                                           |
|----------------------------------------------------------------------|------------------------|-------------------------|-------|-------------------------------------------------------|
| Diabetes causes me to have mood swings                               | Diabetes               | mood swings             | 1     | Possible causal association                            |
| I just want to eat. I hate #diabetes                                | #diabetes              | hate                    | 1     | Possible causal association related to diabetes distress |
| Scary, have a diabetic daughter but I read thousands of people a year die in the UK just from flu so why panic over corona. |                        |                         | 0     | Non-diabetes or diabetes distress related relationship. “flu” is not diabetes-related |
| I’m back! Had two strokes and recover now. Have high blood pressure and diabetes. -) |                        |                         | 0     | Unclear cause-effect relationship                      |
| Not sure if I’ve been up since 3:30 to watch Titanic or because of my anxiety over my glucose test is what keeps me up 😊 | glucose test           | anxiety                 | 1     | Chaining cause-effect relationship                     |
| My 14 year old daughter is Type 1 = malfunctioning pancreas, meaning not enough insulin being made to regulate 😥 | Type 1                 | malfunctioning pancreas; not enough insulin | 1     | Negation                                              |
| wondering why i felt so bad and then realized i haven't given myself my insulin since early this morning 😥 | insulin                | felt like shit          | 0     | Negation                                              |

Table 1: Sample tweets in different label scenarios. The tweets are fictive to ensure privacy but represent similar real tweets
*C.A.: causal association

Models

A first model was trained to predict if a sentence contains a potential cause-effect association (causal sentence) and a second model extracted the specific cause and associated effect from the causal sentence. Thus, the first model acts like a barrier and filters non-causal sentences out. These tweets may have either a cause, an effect, none of them, but not both. To simplify the model training, we hypothesized that cause-effect-pairs only occur in the same sentence and we removed all sentences with less than 6 words due to a lack of context. For this reason we operated on a sentence and not tweet level.

Additional challenges in our setting were that *causes* and *effects* could be multi-word entities and the language used on Twitter is non-standard with frequent slang and misspelled words.
Causal sentences detection

The identification of causal sentences is a binary classification task. The pre-trained language model BERTweet served as foundation for the model architecture capable of handling the non-standard nature of Twitter data. A feed-forward network is built on top of the BERTweet architecture consisting of two fully connected layers (FCLL) with dropout layers with probability 0.3, finalized by a softmax layer which translates the model predictions into probabilities, see Figure 2. The initial training set of 5,000 tweets was imbalanced and, after splitting the tweets into sentences, resulted in 7,218 non causal sentences and 1,017 causal sentences. To adjust for the data imbalance, class weights were included as parameters in the categorical cross entropy loss function to penalise mispredictions for causal sentences stronger. Parameters for the model training were the adam optimizer with epsilon of 1e-8, a scheduler with linearly decreasing learning rate and 0 warm up steps, learning rate of 1e-3 and we trained for 35 epochs with early stopping. Training data was stratified and separated in 90% training and 10% for the test set. Further, 20% from the training set were extracted for validation. Batch size for training and validation was 16 and 32 for the test set.

![Figure 2: Model architecture - Causal sentence detection](image)

Data augmentation through active learning

Data imbalance on the one hand and on the other hand the limited number of positive training examples for each cause-effect pair, due to the fact that causes and effects could potentially be related to any concept in the diabetes domain, drove us to adopt an active learning approach to increase the
training data. Active learning is a sample selection approach aiming to minimize the annotation cost while maximising the performance of ML-based models.\textsuperscript{36} It has been widely applied on textual data.\textsuperscript{37,38} The training data was increased in several iterations as illustrated in Figure 3.

![Active learning loop to augment the training set in a time-consuming fashion](image)

**Figure 3:** Active learning loop to augment the training set in a time-consuming fashion

The first iteration started by training the causal sentence classifier on sentences from the 5,000 tweets. The trained classifier was then applied on 2,000 randomly selected unlabeled tweets resulting in a set of causal sentences and a set of non-causal sentences. The sentences predicted as causal sentences were examined manually and possible misclassifications were corrected to ensure clean positive training samples. The non-causal sentence set remained untouched. As a consequence potential misclassifications remained in the non-causal sentence set, which should then be considered noisy. Both the causal and non-causal sentence set were then combined and added as new training data to the already labeled data, leading to an updated training set of 7,000 tweets. This process was iterated four times, and allowed us to augment the labelled data much faster and efficiently than without active learning as it enables us to focus on the few positive samples. The final training set was used to train the classification model and the cause-effect extraction model.

### Cause-effect pairs

We cast the identification of cause-effect pairs as an event extraction, or named entity recognition task, i.e assigning a label cause or effect to a sequence of words. The manually labeled *causes* and *effects* were encoded in a IO tagging format based on the common tagging format BIO (Beginning, Inside, Outside), introduced by Ramshaw and Marcus.\textsuperscript{39} Here, “I-C” denotes inside the cause and “I-E” inside the effect. Those two tags were completed by the outside tag “O” symbolizing that the
word is neither cause nor effect. The IO tagging scheme for the example sentence with cause Prediabetes and effect change my lifestyle is summarized:

Sentence : Prediabetes, forces, me, to, change, my, lifestyle
IO tags: I-C O O O I-E I-E I-E

Note that a word can be both cause or effect depending on the context. For instance “Prediabetes” in “Prediabetes forces me to change my lifestyle” takes the role of a cause, whereas in “Limited exercising may lead to prediabetes” it is a possible effect. Moreover, the task is complex and considered open-domain as causes and effects are not restricted to one specific topic, but can be related to any concept in our target domain (diabetes). As a consequence, the creation of a representative training set is challenging, as most cause-effect pairs occur rarely.

This complexity drove us to test several model architectures, compare Figure 4 for an overview:

- **BERT_FFL**: Pre-trained BERTweet language model and on top two feed forward layers with a dropout of 0.3 before, followed by a softmax layer. For the model training the cross entropy loss function is selected and weighted by the class weights to penalise mispredictions for causes and effects stronger.

- **BERT_CRF**: This architecture is equivalent to BERT_FFL except the softmax layer is replaced by a Conditional Random Field (CRF) layer. CRF models the label sequence jointly instead of decoding each label independently by considering the correlations between labels in neighborhoods and jointly decodes the best chain of labels for a given input sentence. The features for the CRF are the outputs of the last linear layer. To implement the CRF function, the pytorch extension pytorch-crf is utilized.

- **WE_BERT_CRF**: Single CRF layer with BERTweet embeddings as features augmented by discrete features such as if the word is lowercase, digit or the word length. The CRF function is implemented with the python package sklearn-crfsuite based on CRFsuite. As parameters for the CRF function, the default algorithm “Gradient descent using the L-BFGS method” was chosen and coefficients for L1 and L2 regularization were 0.1.

- **FastText_CRF**: Similarly to WE_BERT_CRF, with the difference that BERTweet embeddings were replaced by FastText embeddings in the feature vector for each word. FastText vectors trained on similar diabetes-related tweets, which were well adapted to our use case.
Clustering of causes and effects

A large part of causes and effects can be regrouped into similar concepts (clusters) to facilitate analyses and allow effective network analyses. We chose a semi-supervised, time-efficient approach in which 1,000 causes and 1,000 effects were randomly chosen and two researchers manually grouped these into clusters such as “diabetes”, “death”, “family”, “fear”, hereinafter referred as “Parent clusters” to simplify understanding. The remaining causes and effects were then automatically compared to each element of all clusters, using cosine similarity, and associated to the cluster containing the most similar element. Experimentally a similarity threshold of 0.55 was determined; if a cause/effect had a similarity smaller than this threshold for all elements, a new cluster was created for this cause / effect. This process led to 1,751 clusters. To remove noisy clusters through potential misclassifications, only clusters with a minimal number of 10 cause/effect occurrences were considered for the following analyses, resulting in 763 clusters. Note, the order of documents might affect the results, as different clusters might have been created. Please refer to Supplementary File S4 for an overview over the 100 most frequent clusters (automatically added clusters have “Other” as “Parent cluster”).
In a last step, these clusters were visualised in an interactive cause-effect network developed in D3 in order to obtain a deeper idea about the cause-effect association. Python (version 3.8.8) and the deep learning framework pytorch (version 1.8.1) were used to implement the above-mentioned methods. The algorithms are open sourced under the following address: https://github.com/WDDS/Causal-associations-diabetes-twitter/

Results

The following results were obtained from 482,583 sentences which were obtained from splitting the 562,013 personal, emotional, non-joke tweets into sentences; excluding questions; and including only sentences with more than 5 words.

Model training and performance

Causal sentences

The performances to detect causal sentences for the imbalanced dataset are illustrated in Table 2 for each round of the active learning loop, with each round having been trained on more data.

| Round | N° sent. train | N° sent. test | Accuracy | Precision | Recall | F1-Score |
|-------|----------------|---------------|----------|-----------|--------|----------|
| 0     | 6,024          | 837           | 64.5     | 58.0      | 67.4   | 53.8     |
| 1     | 7,536          | 1,047         | 67.7     | 61.2      | 71.6   | 58.4     |
| 2     | 8,804          | 1,223         | 67.7     | 60.3      | 66.3   | 56.3     |
| 3     | 10,284         | 1,429         | 65.4     | 60.0      | 68.8   | 54.8     |
| 4     | 11,861         | 1,648         | 71.0     | 61.0      | 67.8   | 58.3     |

Table 2: Performance measures (macro) for each round of more training data

Highest accuracy was reached in round 4 with 71%, whereas the F1-Score of round 1 and round 2 were almost equal with 58.4% respectively 58.3%. We applied the model of round 4 on all remaining tweets, as it was trained on the largest training data set, including difficult causal examples missed by earlier models, and is thus better at identifying complex causal sentences. The active learning strategy led us to increase the training data much quicker than without active learning and without loss in performance. This led to a clean database of 265,328 causal sentences with most noisy sentences removed.
Cause and Effect detection

The active learning strategy led to an extended dataset of 2,118 causal sentences, i.e. containing both cause and effect, which was split in 90% train and 10% test and 20% of the training set served as validation set. The performance of the different cause-effect models are listed in Table 3.

| Models          | Prec | Rec | F1  |
|-----------------|------|-----|-----|
| **BERT_FFL**    |      |     |     |
| I-C             | 0.48 | 0.46| 0.47|
| I-E             | 0.20 | 0.48| 0.29|
| O               | 0.91 | 0.77| 0.83|
| macro           | 0.53 | 0.57| 0.53|
| **BERT_CRF**    |      |     |     |
| I-C             | 0.59 | 0.20| 0.29|
| I-E             | 0.0  | 0.0 | 0.0 |
| O               | 0.83 | 0.99| 0.90|
| macro           | 0.47 | 0.39| 0.40|
| **WE_BERT_CRF** |      |     |     |
| I-C             | 0.63 | 0.61| 0.62|
| I-E             | 0.49 | 0.49| 0.49|
| O               | 0.93 | 0.93| 0.93|
| macro           | 0.68 | 0.68| **0.68**|
| **FastText_CRF**|      |     |     |
| I-C             | 0.59 | 0.57| 0.58|
| I-E             | 0.45 | 0.38| 0.41|
| O               | 0.92 | 0.94| 0.93|
| macro           | 0.65 | 0.63| 0.64|

Table 3: Performance measures for each of the four architectures

The best performing model was the CRF model with BERT embedding features (WE_BERT_CRF) with a precision, recall and F1 of 0.68. Surprisingly, it outperforms fine-tuning a BERT model, which is considered the gold standard of current NER tasks. A potential explanation for that is that BERT-based models make local decisions at every point of the sequence taking the neighboring words into account before its decision. In a situation like ours, with strong uncertainty on all elements, due to the complexity of the task, a single CRF layer model leveraging BERT features, making global decisions using the local context of each word, maximizes the probability of the whole sequence of decision better.

Moreover the CRF model with simpler FastText models achieved strong results as well with one reason being probably that the word embeddings were specifically trained on this diabetes corpus.
In consequence, the WE_BERT_CRF model was applied on all causal sentences leading to a dataset of 96,676 sentences with cause and associated effect predicted.

Cause-effect description

Table 4 provides an overview over the largest clusters, containing either cause or effect, on the left side and on the right side the most frequent cause-effect associations, excluding the largest cluster “Diabetes” as it will be studied separately. The cluster “Diabetes” is the largest one with 66,775 occurrences of “Diabetes” as either cause or effect (ex.: #diabetes, diabetes, diabetes mellitus) followed by “Death” with 16,989 (ex.: passed away, killed, died, suicide, etc.) and “Insulin” (ex.: insulin, insulin hormone, etc.) with 14,148. From the 30 largest clusters, 6 refer to nutrition, 4 to diabetes and 3 clusters to each of insulin, emotions and the healthcare system.

The largest cluster “Diabetes” mainly occurs as a cause and its most frequent effects (“Death”, “fear”, “sick”) are visualised in Figure 5. From the 30 most numerous effects for “Diabetes”, 6 were related to “Nutrition” and 5 to “Complications & comorbidities” and 3 to each of “Diabetes distress”, “Emotions” and “Healthcare system”.

The interactive visualisation in D3 with filter options was published under https://observablehq.com/@adahne/cause-and-effect-associations-in-diabetes-related-tweets. We invite the interested reader to play with the graph to enhance understanding. Figure 6 provides an example graph of this visualisation showing only cause-effect relationships with at least 250 occurrences to ensure readability. It is striking that “death” seems to play such a central role as effect with various causes (“unable to afford insulin”, “rationing insulin”, “finance”, “insulin”, “Type 1 diabetes (T1D)”) pointing at it. Other central nodes are “Type 1 diabetes” acting as cause for “insulin pump”, “insulin”, “hypoglycemia (hypo)”, “sickness”, “finance” and emotions “anger” and “fear”, where latest has the strongest association; or the node “Insulin” mostly relating as cause to “sickness”, “medication”, “finance”, “death”, or “hypoglycemia” and “fear” and “anger”.

13
| Parent cluster | cluster | N   | cause                          | effect | N   |
|---------------|---------|-----|-------------------------------|--------|-----|
| Diabetes      | diabetes| 66,775 | unable to afford insulin      | death   | 1,246 |
| Death         | death   | 16,989 | insulin                       | death   | 1,156 |
| Insulin       | insulin | 14,148 | type 1 diabetes               | fear    | 1,054 |
| Diabetes      | type 1 diabetes | 11,693 | type 1 diabetes               | death   | 999  |
| Emotions      | fear    | 10,160 | rationing insulin             | death   | 805  |
| Glycemic variability | hypoglycemia | 9,547 | type 1 diabetes               | insulin | 751  |
| Symptoms      | sick    | 6,549  | OGTT*                         | sick    | 584  |
| Nutrition     | overweight | 5,186 | type 1 diabetes               | hypoglycemia | 578 |
| Diabetes      | type 2 diabetes | 4,909 | insulin                       | hypo    | 545  |
| Complications & comorbidities | neuropathy | 4,481 | insulin                       | fear    | 534  |
| Healthcare system | medication | 4,389 | type 1 diabetes               | insulin pump | 436 |
| Diabetes Technology | insulin pump | 4,307 | finance                       | death   | 423  |
| Nutrition     | nutrition | 4,230 | type 1 diabetes               | sick    | 400  |
| Emotions      | anger   | 4,149  | insulin                       | sick    | 385  |
| Health        | OGTT*   | 4,053  | insulin                       | finance | 367  |
| Blood pressure | hypertension | 3,782 | type 1 diabetes               | anger   | 356  |
| Healthcare system | finance | 3,767 | insulin                       | medication | 305 |
| Nutrition     | reduce weight | 3,589 | insulin                       | anger   | 296  |
| Insulin       | unable to afford insulin | 3,381 | OGTT*                         | fear    | 293  |
| Nutrition     | diet    | 3,325  | type 2 diabetes               | death   | 293  |
| Emotions      | sadness | 3,153  | type 2 diabetes               | fear    | 290  |
| Glycemic variability | hyperglycemia | 3,144 | hypertension                  | death   | 286  |
| Diabetes      | suffer  | 3,132  | overweight                    | death   | 280  |
| Diabetes Distress | depression | 2,810 | type 1 diabetes               | finance | 277  |
| Healthcare system | hospital | 2,721 | hypoglycemia                  | insulin | 272  |
| Diabetes Distress | stress  | 2,681  | hypoglycemia                  | sick    | 263  |
| Nutrition     | sugar   | 2,369  | affordable insulin            | death   | 262  |
| Nutrition     | fasting | 2,363  | insulin                       | insulin pump | 255 |
| Insulin       | rationing insulin | 2,244 | complications                 | death   | 248  |
| Health        | gestational diabetes | 2,076 | insulin                       | sadness | 240  |

Table 4: Left side column shows the most frequent clusters (causes and effects) with the number of occurrences. The last column shows the most frequent cause-effect relationships excluding the cluster “Diabetes”.

*OGTT: Oral glucose tolerance test
Figure 5: Most frequent effects for the largest cluster “Diabetes”

Figure 6: Cause-effect network with a minimum number of associations (edges) of 250. Accessible under: https://observablehq.com/@adahne/cause-and-effect-associations-in-diabetes-related-tweets
Discussion

Principal results

Our findings suggest that it is feasible to extract both explicit and implicit cause and associated effects from diabetes-related Twitter data. We demonstrated that by adopting the transfer learning paradigm and fine-tuning a pre-trained language model we were able to detect causal sentences. Moreover, we have shown that simply fine-tuning a BERT-based model does not always outperform more traditional methods such as relying on conditional random fields in the case of the cause-effect pair detection. The precision, recall and F1 numbers, given the challenging task and the imbalanced dataset, were satisfying. The semi-supervised clustering and interactive visualisation enabled us to identify “Diabetes” as the largest cluster acting mainly as the cause for “Death” and “fear”. Besides, a central cluster was detected in “Death” acting as an effect for various causes related to insulin pricing, a link already detected in earlier works.

Comparison with the literature

Several former works have addressed causality on Twitter data. Doan and al. focused on three health-related concepts such as, “stress”, “insomnia”, “headache” as effects and identified causes using manually crafted patterns and rules. However they only focused on explicit causality and excluded causes and effects encoded in hashtags and synonymous expressions. On the contrary, we tackled both explicit and implicit causality, including causes and effects in hashtags, and exploiting synonymous expressions through the use of word embeddings. Kayesh et al. proposed an innovative approach, a novel technique based on neural networks which uses common sense background knowledge to enhance the feature set, but they focused on the simplified version of explicit causality in tweets. Bollegala et al. developed a causality-sensitive approach for detecting adverse drug reactions from social media using lexical patterns and in consequence aiming at explicit causality. Dasgupta et al. proposed one of the few deep learning approaches, due to the unavailability of appropriate training data, leveraging a recursive neural network architecture to detect cause-effect relations from text, but also only targeted explicit causality. A Bert-based approach tackling both explicit and implicit causality is provided by Khetan et al. who used already existing labeled corpora not based on social media data. Recently they further extended their work of explicit and implicit causality understanding in single and multiple sentence but in clinical notes.

To the best of our knowledge, this is the first paper investigating both explicit and implicit cause-effect relationships on diabetes-related Twitter data.
Strengths and Limitations

The present work demonstrates various strengths. First, by leveraging powerful language models we were able to identify a large number of tweets containing cause-effect relationships which enabled us to the detect cause-effect associations in 20% (96,676 / 482,583) of the sentences, contrary to other approaches which were able to identify causality in less than 2% of tweets. Second, contrary to most previous work, we tackled both explicit and implicit causal relationships, an additional explanation for the higher number of cause-effect associations we obtained compared to other studies focusing only on explicit associations. Third, relying fully on automatic machine learning algorithms avoided us from defining manually crafted patterns to detect causal associations. Fourth, operating on social media data that is expressed spontaneously and in real-time offers the opportunity to gain knowledge from an alternative data source which might complement traditional epidemiological data sources.

A strong limitation is that cause-effect relations are expressed in tweets and this cannot be used for causal inference as the Twitter data source is uncertain and the information shared can be opinion or observation. Another shortcoming is that the performance of our algorithms to detect cause-effect pairs is not perfect. But the overall process and the vast amount of data minimizes this issue. The lack of recall is counterbalanced by the sheer amount of data and the lack of precision is counterbalanced by the clustering approach in which non-frequent causes or effects are discarded. Labeling causes and effects in a dataset is a highly complicated task and we would like to emphasize that mislabelings in the dataset may occur. Enhancing data quality certainly is a strong point to address to further improve performance. The causal association structures learnt by the model from the training set, might not generalise completely when applied on the large amount of Twitter data. Besides, the active learning strategy certainly added noise to the model, as only positive samples were corrected, which could be improved in future investigations. Moreover, we would like to highlight that the diabetes related information shared on Twitter, may not be representative for all people with diabetes. For instance we observed a bigger cluster of causes/effect related to type 1 diabetes compared to type 2 diabetes, which is contrary to the real world. A potential explanation for that is the age distribution of Twitter users. But due to the large number of tweets analyzed, a significant variability in the tweets could be observed.

Conclusion

In this work, we developed an innovative methodology to identify possible cause-effect relationships among diabetes-related tweets. This task was challenging due to addressing both explicit and implicit causality, multi-word entities, the fact that a word could be both cause or effect, the open domain of
causes and effects, the biases occurring during labeling of causality, and the relatively small dataset for this complex task. We overcame these challenges by augmenting the small dataset via an active learning loop. The feasibility of our approach was demonstrated using modern BERT-based architectures in the preprocessing and causal sentence detection. A combination of BERT features and CRF layer were leveraged to extract causes and effects in diabetes-related tweets which were then aggregated to clusters in a semi-supervised approach. The visualisation of the cause-effect network based on Twitter data can deepen our understanding of diabetes, in a way of directly capturing patient-reported outcomes from a causal perspective.

Acknowledgements

None

Competing Interests

None

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 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.
References

1. Gao J, Zheng P, Jia Y, et al. Mental health problems and social media exposure during COVID-19 outbreak. *PloS One.* 2020;15(4):e0231924. doi:10.1371/journal.pone.0231924

2. Chen M, Mangubat E, Ouyang B. Patient-reported outcome measures for patients with cerebral aneurysms acquired via social media: data from a large nationwide sample. *J NeuroInterventional Surg.* 2016;8(1):42-46. doi:10.1136/neurintsurg-2014-011492

3. Nguyen QC, Brunisholz KD, Yu W, et al. Twitter-derived neighborhood characteristics associated with obesity and diabetes. *Sci Rep.* 2017;7(1):16425. doi:10.1038/s41598-017-16573-1

4. Fodeh S, Li T, Menczynski K, et al. Using Machine Learning Algorithms to Detect Suicide Risk Factors on Twitter. In: *2019 International Conference on Data Mining Workshops (ICDMW).* ; 2019:941-948. doi:10.1109/ICDMW.2019.00137

5. Kayesh H, Islam MdS, Wang J. A Causality Driven Approach to Adverse Drug Reactions Detection in Tweets. In: Li J, Wang S, Qin S, Li X, Wang S, eds. *Advanced Data Mining and Applications.* Lecture Notes in Computer Science. Springer International Publishing; 2019:316-330. doi:10.1007/978-3-030-35231-8_23

6. Yang J, Han SC, Poon J. A Survey on Extraction of Causal Relations from Natural Language Text. *ArXiv210106426 Cs.* Published online January 16, 2021. Accessed June 30, 2021. http://arxiv.org/abs/2101.06426

7. Doan S, Ritchart A, Perry N, Chaparro JD, Conway M. How Do You #relax When You’re #stressed? A Content Analysis and Infodemiology Study of Stress-Related Tweets. *JMIR Public Health Surveill.* 2017;3(2):e35. doi:10.2196/publichealth.5939

8. Cocos A, Fiks AG, Masino AJ. Deep learning for pharmacovigilance: recurrent neural network architectures for labeling adverse drug reactions in Twitter posts. *J Am Med Inform Assoc JAMIA.* 2017;24(4):813-821. doi:10.1093/jamia/ocw180

9. Doan S, Yang EW, Tilak S, Li P, Zisook D, Torii M. Extracting health-related causality from twitter messages using natural language processing. *BMC Med Inform Decis Mak.* 2019;19(Suppl 3):79. doi:10.1186/s12911-019-0785-0

10. Fagherazzi G. Challenges and perspectives for the future of diabetes epidemiology in the era of digital health and artificial intelligence. *Diabetes Epidemnl Manag.* 2021;1:100004. doi:10.1016/j.deman.2021.100004

11. Fagherazzi G, Ahne A, Guillot C, et al. Étude mondiale de la détresse liée au diabète : le potentiel du réseau social Twitter pour la recherche médicale. *Rev DÉpidémiologie Santé Publique.* 2018;66:S197-S198. doi:10.1016/j.respe.2018.04.002

12. Fisher L, Hessler DM, Polonsky WH, Mullan J. When Is Diabetes Distress Clinically Meaningful?: Establishing cut points for the Diabetes Distress Scale. *Diabetes Care.* 2012;35(2):259-264. doi:10.2337/dc11-1572

13. Fisher L, Polonsky WH, Hessler DM, et al. Understanding the sources of diabetes distress in adults with type 1 diabetes. *J Diabetes Complications.* 2015;29(4):572-577. doi:10.1016/j.jdiacomp.2015.01.012

14. Coccaro EF, Lazarus S, Joseph J, et al. Emotional Regulation and Diabetes Distress in Adults With Type 1 and Type 2 Diabetes. *Diabetes Care.* 2021;44(1):20-25. doi:10.2337/dc20-1059

15. Carper MM, Traeger L, Gonzalez JS, Wexler DJ, Psaros C, Safren SA. The differential associations of depression and diabetes distress with quality of life domains in type 2 diabetes. *J Behav Med.* 2014;37(3):501-510. doi:10.1007/s10865-013-9505-x

16. Cummings DM, Lutes L, Littlewood K, et al. Regimen-Related Distress, Medication
Adherence, and Glycemic Control in Rural African American Women With Type 2 Diabetes Mellitus. *Ann Pharmacother.* 2014;48(8):970-977. doi:10.1177/1060028014536532

17. Fisher L, Mullan JT, Skaff MM, Glasgow RE, Arean P, Hessler D. Predicting diabetes distress in patients with Type 2 diabetes: a longitudinal study. *Diabet Med J Br Diabet Assoc.* 2009;26(6):622-627. doi:10.1111/j.1464-5491.2009.02730.x

18. Pandit AU, Bailey SC, Curtis LM, et al. Disease-related distress, self-care and clinical outcomes among low-income patients with diabetes. *J Epidemiol Community Health.* 2014;68(6):557-564. doi:10.1136/jech-2013-203063

19. 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 Diabetes Res Care.* 2020;8(1):e001190. doi:10.1136/bmjdrcc-2020-001190

20. Balkhi A, Reid A, McNamara J, Geffken G. The diabetes online community: The importance of forum use in parents of children with type 1 diabetes. *Pediatr Diabetes.* 2013;15. doi:10.1111/pedi.12110

21. Khoo CSG, Chan S, Niu Y. Extracting causal knowledge from a medical database using graphical patterns. In: *Proceedings of the 38th Annual Meeting on Association for Computational Linguistics - ACL ’00.* Association for Computational Linguistics; 2000:336-343. doi:10.3115/1075218.1075261

22. Kayesh H, Islam MS, Wang J. On Event Causality Detection in Tweets. *ArXiv190103526 Cs.* Published online January 11, 2019. Accessed January 20, 2021. http://arxiv.org/abs/1901.03526

23. Khoo C, Chan S, Niu Y. The Many Facets of the Cause-Effect Relation. In: Green R, Bean CA, Myaeng SH, eds. *The Semantics of Relationships: An Interdisciplinary Perspective.* Information Science and Knowledge Management. Springer Netherlands; 2002:51-70. doi:10.1007/978-94-017-0073-3_4

24. Xu Y, Mou L, Li G, Chen Y, Peng H, Jin Z. Classifying Relations via Long Short Term Memory Networks along Shortest Dependency Paths. In: ; 2015:1785-1794. doi:10.18653/v1/D15-1206

25. Wang L, Cao Z, Melo G de, Liu Z. Relation Classification via Multi-Level Attention CNNs. In: ; 2016:1298-1307. doi:10.18653/v1/P16-1123

26. Ponti EM, Korhonen A. Event-Related Features in Feedforward Neural Networks Contribute to Identifying Causal Relations in Discourse. In: ; 2017:25-30. doi:10.18653/v1/W17-0903

27. Khetan V, Ramnani R, Anand M, Sengupta S, Fano AE. Causal BERT: Language Models for Causality Detection Between Events Expressed in Text. In: Arai K, ed. *Intelligent Computing.* Lecture Notes in Networks and Systems. Springer International Publishing; 2021:965-980. doi:10.1007/978-3-030-80119-9_64

28. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *ArXiv181004805 Cs.* Published online May 24, 2019. Accessed November 13, 2020. http://arxiv.org/abs/1810.04805

29. Vaswani A, Shazeer N, Parmar N, et al. Attention Is All You Need. *ArXiv170603762 Cs.* Published online December 5, 2017. Accessed June 17, 2021. http://arxiv.org/abs/1706.03762

30. Gurulingappa H, Rajput AM, Roberts A, Fluck J, Hofmann-Apitius M, Toldo L. Development of a benchmark corpus to support the automatic extraction of drug-related adverse effects from medical case reports. Published online 2012. doi:10.1016/j.jbi.2012.04.008

31. *Twitter Privacy.*; 2021. Accessed June 15, 2021. https://twitter.com/en/privacy

32. Cohen J. A Coefficient of Agreement for Nominal Scales. *Educ Psychol Meas.* 1960;20(1):37-46. doi:10.1177/001316446002000104

33. Altman D. *Practical Statistics for Medical Research.* Chapman & Hall; 1991.

34. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159-174.

35. Nguyen DQ, Vu T, Nguyen AT. BERTweet: A pre-trained language model for English
54. Polonsky WH, Anderson BJ, Lohrer PA, et al. Assessment of diabetes-related distress. *Diabetes Care*. 1995;18(6):754-760. doi:10.2337/diacare.18.6.754

55. Polonsky WH, Fisher L, Earles J, et al. Assessing psychosocial distress in diabetes: development of the diabetes distress scale. *Diabetes Care*. 2005;28(3):626-631. doi:10.2337/diacare.28.3.626

56. Princeton University. *About WordNet*. Princeton University; 2010. https://wordnet.princeton.edu/ (accessed: 22/08/2021)

57. Miller GA. WordNet: A Lexical Database for English. In: *Communications of the ACM*. Vol 38.; 1995:No. 11: 39-41.
Supplementary File S1: List of diabetes-related keywords for the Twitter API tweet extraction

| English keywords          |
|--------------------------|
| glucose                  |
| #glucose                 |
| insulin pump             |
| #insulinpump             |
| type 1                   |
| #type1                   |
| #bloodsugar              |
| #bloodglucose            |
| #diabetes                |
| #type1                   |
| #t2d                     |
| #type2                   |
| #dsma                    |
| #type2diabetes           |
| #diabetic                |
| #gbdoc                   |
| #hba1c                   |
| #cgm                     |
| #diabetes                |
| #diabetes                |
| #t2d                     |
| #type2                   |
| #dsma                    |
| #type2diabetes           |
| #diabetic                |
| #gbdoc                   |
| #hba1c                   |
| #cgm                     |
| #fingerprick             |
| #changingdiabetes        |
| #thisisdiabetes          |
| #lifewithdiabetes        |
| #diabetesadvocate        |
| #stopdiabetes            |
| #diabetesproblems        |
| #justdiabeticthings      |
| #1dlookslikeme           |
| #diaversary              |
| #diabetestest            |
| #t2dlookslikeme          |
| pwd                      |
| #duckfiabetes            |
| #GBDoc                   |
| #pwd                     |
| #kissmyassdiabetes       |
Supplementary File S2: Preprocessing pipeline

Tweets are noisy, unstructured and contain numerous misspelled or non-standard English words. To reduce noise in the dataset a similar preprocessing pipeline as in earlier works was adopted.\textsuperscript{19} First, retweets and duplicates were removed to obtain a database with 7.7 million unique tweets. Secondly, to increase the relevance of the analyzed tweets we determined only tweets with personal content where feelings, emotions and opinions could be shared by people with or talking about diabetes. As a consequence, institutional tweets referring to commercial, news or health information, were considered out of scope for this study and excluded. It has already been shown that diabetes-related tweets can be grouped into several clusters such as commercial, health information, social intervention by Beguerisse-Diaz et al.\textsuperscript{49} Contrary to Jonsen et al.\textsuperscript{50} who identified personal content based on personal pronouns like ‘I’, ‘me’, ‘us’, we leveraged the transfer learning paradigm and fine-tuned an already pretrained transformer-based language model to detect personal content in tweets.\textsuperscript{29} As pre-trained language model served BERTweet\textsuperscript{35}, which was trained on 850 million English tweets (16 billion word tokens ~ 80GB) collected from January 2012 to August 2019 following the RoBERTa pre-training procedure.\textsuperscript{51} We undertook the same preprocessing steps to our tweets that were used to pretrain the model such as tokenization, translating emotion icons into text strings and converted user mentions and web/url links into special tokens. To use the model and fine-tune it for a binary sentence classification a linear layer was added on top of the last Transformer layer of the BERTweet model using the transformers package of Huggingface.\textsuperscript{52} The model was then fine-tuned with an extended data set, of the one provided by Ahne et al. leading to a total of 4303 tweets (1539 personal, 2764 institutional), to account for a possible temporal divergence of the way people tweet.\textsuperscript{19} The model performance to identify tweets with personal content was: accuracy of 91.2\%, f1 of 88.5\%, precision of 86.2\% and recall of 90.9\%. The trained model is then applied on all unique tweets resulting in a total of 2.5 million tweets with personal content.

Moreover, jokes around diabetes are common on Twitter and considered out of scope for this study as well. Similarly to the personal content classifier, BERTweet was fine-tuned to detect if a tweet is a joke. For this purpose a joke tweet dataset from earlier works was extended to a total of 1648 tweets (486 jokes, 1162 non jokes).\textsuperscript{19} The performance to identify if a tweet is a joke was: accuracy of 90.4\%; f1 of 84.2\%, precision of 78.5\%, recall of 90.8\%. Applying the joke classifier on all tweets with personal content led to a dataset of 1.8 million personal, non-joke tweets.

A particular focus of this study lied on studying diabetes distress and thus psychological factors and emotions. To capture those psychological factors and emotions in tweets, only tweets containing an emotional element such as emojis and emoticons or emotional words were kept, analogous to earlier works.\textsuperscript{19} In addition, emotional words were identified based on a combination of the psychologue Parrot’s hierarchical classification of emotions with the six primary emotions (joy, love, surprise,
sadness, anger, fear) and emotional words present in common questionnaires to study diabetes distress such as the Problem Areas in Diabetes scale (PAID) and Diabetes Distress Scale (DDS).\textsuperscript{53–55} Those emotional words were augmented by synonyms using the WordNet database to obtain a more extensive list.\textsuperscript{56,57} This led to 562,013 tweets containing personal, non-joke and emotional content.
Supplementary File S3: Annotation guidelines

Objective
The aim of this labeled corpus is to provide a training data set for detecting possible cause-effect-pairs in diabetes and diabetes distress related tweets. Diabetes distress regroups all psychological factors related to the day-to-day disease management such as emotional burden, stress, anxiety, emotions, etc.

Data
Between April 2017 and January 2021 diabetes related tweets have been extracted using the Twitter API based on list of diabetes-related keywords, such as “diabetes”, “insulin”, “hypoglycemia”, “#T1D”, “#DSMA”, “Type 2”, “#diabeteslookslikeme”, compare Supplementary File 1 for the full list. Based on the extracted tweets a random subsample of 5000 tweets has been selected for annotation purposes.

File structure - Columns:

- **Text [String]**: extended tweet message “Full_text” in tweet object
- **Intent [String]**: Intent of the tweet. If several intents, they are separated by a semicolon (“;”)
  
  Can take the following values:
  
  - q: Question in tweet
  - ms: multiple sentences in tweet
  - mC: multiple causes in tweet
  - mE: multiple effects in tweet
  - msS: multiple sentences in tweet with cause and effect in a single sentence
  - neg: A negation which negating the meaning of the cause or effect in a sentence
  - joke: A joke, an irony or sarcasm is in the tweet

- **Cause [String]**: Words describing the causes. A cause can be composed of several words. If several causes occur in a tweet then they are separated by a semicolon (“;”)

- **Effect [String]**: Words describing the effect. An effect can be composed of several words. If several effects occur in a tweet then they are separated by a semicolon (“;”)

- **Causal association [0,1]**: Binary variable of a cause-effect pair occurs in a tweet where 0 means no cause-effect pair and 1 means there is a cause-effect pair
Definition of a cause-effect relationship and annotation rules

The following tweet examples are fictive to ensure privacy.

Non-diabetes or diabetes distress related relationships

The focus on this corpus lies on cause-effect relationships related to diabetes and diabetes distress. For this reason tweets like the following are not labeled as causal. The possible cause here might be “flu” and the effect “die”, but “flu” is out of scope in this project.

| Tweet                                                                                           | Intent | Cause                  | Effect                  | C.A.* |
|-------------------------------------------------------------------------------------------------|--------|------------------------|-------------------------|-------|
| Scary, i have a 13 year old diabetic daughter however i read 4 thousand or more people a year die in UK just from flu. so why this fuss & panic over corona . I read lots and had nightmares last night ! ! This is ridiculous | mS     |                        |                          | 0     |
| Schools are closed to prevent passing the virus , yet ALL DAY LONG they are in the store with parents , putting me and MY HEALTH at risk ! WHY, WHY, WHY? |        |                        |                          | 0     |

* causal association

In the second example “heart disease” is not labeled as the cause as it is out of scope.

Examples for possible causal associations

| Tweet                                                                                           | Intent | Cause                  | Effect                  | C.A.* |
|-------------------------------------------------------------------------------------------------|--------|------------------------|-------------------------|-------|
| Diabetes causes me to have mood swings. :/                                                     |        | Diabetes               | mood swings             | 1     |
| years of diabetes and all I got is a Spidey-sense like ability to notice any abnormal sensations in my body and about 7 new kinds of anxiety I didn't know existed before I got diagnosed . it sucks , but I'm much stronger because of it . | mS; mE | diabetes               | abnormal sensations in my body;anxiety | 1     |
| Gestational Diabetes is shit . The poking , urge to eat the foods that are no good for me . When I am in need of more insulin my body alarms are all going off , making me tired , headache , blurred vision . | mS; mE | in need of more insulin | tired;headache; blurred vision | 1     |
| After 10 years of injections and finger pricks, I have finally gotten an Insulin pump and glucose monitor . Finally I can start to manage my diabetes even better and improve my health . Waited so long | mS;mC; mE | Insulin pump;glucose monitor | manage my diabetes even better;improve my health | 1     |

* causal association
The above examples also show that several causes can lead to an effect, and inversely also one cause can lead to several effects.

Implicit relations

The cause-effect relationship is not stated by a causal link word

| Tweet | Intent | Cause     | Effect       | C.A.* |
|-------|--------|-----------|--------------|-------|
| I was sent to the penalty box to fix a low blood sugar #diabetes #NHLPlayoffs |        | #diabetes   | low blood sugar | 1 |

* causal association

Unclear cause - effect relationships

In tweets in which there is a possible cause and a possible effect but it is not clear if the “cause” had an influence on the “effect”, the tweet is labeled as non-causal.

| Tweet | Intent | Cause       | Effect               | C.A. * |
|-------|--------|-------------|----------------------|--------|
| I’m back! Had two strokes recovering now my legs do not want to move. Have high blood pressure and diabetes. So all of you out there please watch you're blood pressure. It matters. |        |                          |         | 0 |
| My dad has diabetes, cancer, heart problems, and a weak immune system. |        |                          |         | 0 |

* causal association

The possible cause is “High blood sugar and diabetes” and the possible effect is “stroke”. But it can not be concluded that the stroke was provoked by the high blood sugar or diabetes.

Several chaining cause-effect relationships: A -> B -> C

If in a tweet we have two relationships: event A causes event B and at the same time event B causes event C, then we labelled the relationship that is closest to our objective to study diabetes and diabetes distress:

| Tweet | Intent | Cause       | Effect  | C.A. * |
|-------|--------|-------------|---------|--------|
| Not sure if I’ve been up since 3:30 for Titan or because my anxiety over my glucose test is keeping me up 😫 Bahh | glucose test | anxiety | 1 |
I am also a diabetic with all this worry & stress, is adding to my sugar levels to rise.

Excess insulin from eating too many carbs spikes insulin, making you hungry. Belief me

| Tweet | Intent | Cause | Effect | C.A. |
|-------|--------|-------|--------|------|
| My 14 year old daughter. Type 1 (malfunctioning pancreas, aka not enough insulin being made to regulate) | mE | Type 1 | malfunctioning pancreas, not enough insulin | 1 |
| I'm a Type 1 Diabetic, out of work and unable to afford my insulin 😩 | | | out of work unable to afford my insulin | 1 |
| I was wondering why I felt like shit and then I realized I haven't given myself my insulin since early this morning. stupid.. | neg | insulin | felt like shit | 1 |
| Don't hate your diabetes; instead, find ways to love it and get rid of it over time. It helps | neg | diabetes | hate | 1 |
| My friend "gave" herself diabetes by not doing what her doctor told her LOSE WEIGHT! | neg | LOSE WEIGHT | diabetes | 1 |

* causal association

The last example shows when a tweet is labeled as negation. The negation “haven’t given myself” alters the meaning of the causal relationship “insulin” -> “felt like shit”.

Diabetes Distress

In labeling this dataset, a special focus was lying on diabetes distress (psychological factors related to the day-to-day disease management, such as anxiety, stress, emotions, etc.). For this reason we labeled possible causal associations related to diabetes distress as well:
| Tweet                                                                 | Intent | Cause          | Effect | C.A. |
|----------------------------------------------------------------------|--------|----------------|--------|------|
| I do I just want to go to the kitchen and eat . I hate #diabetes    | msS    | #diabetes      | hate   | 1    |
| I have gestational diabetes and im very much bothered               |        | gestational    | bothered | 1   |
| Kent’s just angry because his diabetes is flaring up again          |        | diabetes       | angry  | 1    |

* causal association

Jokes
As jokes were also labeled tweets containing ironic or sarcastic elements.

| Tweet                                                                 | Intent | Cause          | Effect | C.A. |
|----------------------------------------------------------------------|--------|----------------|--------|------|
| This tweet is so dumb it gave me diabetes                           | joke   |                |        | 0    |
| I love lifestyle choices become a non smoker and a temporary diabetic , why the hell not *irony out* | joke   |                |        | 0    |
| Thanks sweetie And I think I've developed diabetes from your sweetness | joke   |                |        | 0    |

* causal association

Frequently used abbreviations related to diabetes

| Abbreviation | Explanation                                                                 |
|--------------|-----------------------------------------------------------------------------|
| lows, going low | low blood sugar                                                              |
| #gbdoc, #doc, #dsma | diabetes related online groups on social media to exchange about the disease |
| dexcom, Freestyle Libre | continuous glucose monitoring tools helping to monitor blood sugar levels levels |
| cgm, CGM    | continuous glucose monitoring                                                |
| DKA, dka    | Diabetic Ketoacidosis                                                        |
| 3 hours     | 3 hour glucose test for gestational diabetes                                |
| LCHF        | low carb high fat diet: The diet, because of its low requirement for insulin, has been recognised by the Swedish government as being suitable for people with type 2 diabetes and as helpful to individuals looking to lose weight or maintain a healthy weight. |
| BS          | blood sugar                                                                  |
Supplementary File S4: Most frequent clusters

The synonyms were manually added for the initial clusters. Clusters whose parent cluster is “Other”, are automatically added clusters that were predicted from the cause-effect classifier.

| Nº  | Parent cluster       | cluster          | synonyms                                                                 | N    |
|-----|----------------------|------------------|--------------------------------------------------------------------------|------|
| 0   | Diabetes             | diabetes         | diabetic, #diabetic, #diabetes mellitus, diabetics, DIABETIC, #Diabetic   | 66775|
| 1   | Death                | death            | die, passed away, gave life, kill, killing, dead, shorter lifespan, loose live, died, dying, commit suicide, losing father | 16989|
| 2   | Insulin              | insulin          | insulin hormone, supplies, HUMALOG bottle                                | 14148|
| 3   | Diabetes             | type 1 diabetes  | T1D, type 1, #type1, #type1diabetes, juvenile diabetes                  | 11693|
| 4   | Emotions             | fear             | anxiety, terrifed, scared, anxious, concern, dread, scary, worried, nervous, distress, panic, hysteria, terrified, crested, traumatized | 10160|
| 5   | Glycemic variability | hypoglycemia     | low glucose, low blood sugar, low, hypo, go low, glucose down, blood sugar down, lowered BS, sugar drop | 9547 |
| 6   | Symptoms             | sick             | headache, dizzy, threw up, vomiting, painful, puke, feeling sick, shaky, nausea, coughing, diarrhea, dry mouth, fever, sweat | 6549 |
| 7   | Nutrition            | overweight       | obese, eat too much, weight gain, fat, gained pounds, obesity            | 5186 |
| 8   | Diabetes             | type 2 diabetes  | T2D, type 2, #type2, #type2diabetes, T2 diabetic, t2d, TYPE TWO          | 4909 |
| 9   | Complications & comorbidities | neuropathy | amputation, feet amputation, lost feet, lost leg, leg amputation, nerve death, nerve damage, lost hand, diabetes neuropathy | 4481 |
| 10  | Healthcare system    | medication       | meds, diabetes meds, drug, antibiotics, pills, drugs, dose, medicine, metformin prescribed, prescription, cheapest medicines | 4389 |
| 11  | Diabetes Technology  | insulin pump     | injecting insulin, injection, inject, needle, pump, finger prick, shot, | 4307 |
| 12  | Nutrition            | nutrition        | vegan, vegetarian, eat carbs, carbohydrates, no chocolate, can't eat donuts, food, salad, noodles, appetite fish, NEEDED meal, seafood, milk, entire meal, broccoli | 4230 |
| 13  | Emotions             | anger            | rage, outrageous, frustration, hate, angry, jealous, jealousy, raging, pissed, pissed off, frustrating | 4149 |
| 14  | Health               | OGGT             | glucose test drink, glucose test, 3 hour test, ogtt, diabetic drink, horrific drink | 4053 |
| 15  | Blood pressure       | hypertension     | high blood pressure, BP                                                 | 3782 |
| 16  | Healthcare system    | finance          | wages, student loans, GoFundMe, expenses, costly, expensive care, pay, price gouging, spend money, debt, insulin price, insulin prices, donations, Healthcare Cost | 3767 |
| 17  | Nutrition            | reduce weight    | lost pounds, lose weight, #loseweight                                   | 3589 |
| 18  | Insulin              | unable to afford insulin | can't afford insulin, no access to affordable insulin, could not afford insulin, can't afford meds, could not buy insulin, bankrupt, financially unstable | 3381 |
| 19  | Nutrition            | diet             | diabetic diet, Keto diet, carnivore diet, keto, plant based diet, high fat diet, change diet, #lowcarb, LCHF, dietary needs, Low Carb | 3325 |
| 20  | Emotions             | sadness          | cry, sad, sucks, loneliness, lonely, sadly, CRIED, despair, hurtful, hurting, psychological grief, disappointing | 3153 |
| 21  | Glycemic variability | hyperglycaemia    | high blood sugar, high glucose, high glucose levels, spike glucose, higher blood glucose, blood glucose up, blood glucose levels up, elevated #BP | 3144 |
| 22 | Diabetes | suffer | suffering, TERRIBLE PAIN, HURT | 3132 |
| 23 | Diabetes Distress | depression | depressed, depressing, lose hope, mentally ill, hopeless, antidepressants, psychologically fragile | 2810 |
| 24 | Healthcare system | hospital | surgery, syringes, doctor, appointment, checkup, medical attention, medical treatment, ER, ICU, hospitalize, ambulance, doc, surgeries, GP practice, clinical psychologist, Caregiver | 2721 |
| 25 | Diabetes Distress | stress | mood disorder, stressed, stressful, MOOD SWINGS, tense | 2681 |
| 26 | Nutrition | sugar | sweets, candy, waffle, soda, Sugar, CAKE, artificial sweeteners, CRAVE SWEETS, milkshakes, LOVE SUGAR | 2369 |
| 27 | Nutrition | fasting | starvation, not eating | 2363 |
| 28 | Insulin | rationing insulin | shortage insulin, denying insulin, lack insulin, ration insulin, EXPIRED INSULIN | 2244 |
| 29 | Health | gestational diabetes | pregnancy, pregnant | 2076 |
| 30 | Health | prediabetes | pre diabetic, borderline diabetic | 1932 |
| 31 | Diabetes Distress | feel bad | feel awkwarded, disgusting, appetite, grumpy | 1861 |
| 32 | Complications & comorbidities | retinopathy | horrible vision, bad eyesight, vision decline, lost sight, blind, diabetes retinopathy | 1750 |
| 33 | Complications & comorbidities | high risk | risk | 1663 |
| 34 | Healthcare system | insurance | company, pharma, health insurance, coverage, Medicare, #BigPharma #Insulin, medicaid | 1627 |
| 35 | Complications & comorbidities | coma | unconscious, pass out, Diabetic Coma | 1540 |
| 36 | Complications & comorbidities | heart attack | cardiovascular, cardiovascular disease, diabetic heart attack, CHF | 1511 |
| 37 | Health | insulin resistance | | 1505 |
| 38 | Complications & comorbidities | complications | diabetes complication, issues | 1443 |
| 39 | Other | struggle | | 1357 |
| 40 | Complications & comorbidities | nephropathy | kidney damage, diabetes kidney failure, Nephrologist #Diabetes #Nephrologist,kidney failure | 1338 |
| 41 | Emotions | joy | feel good, feel better, relief, happy, proud | 1279 |
| 42 | Other | constant pain | | 1240 |
| 43 | Other | r't know language | | 1234 |
| 44 | Diabetes Distress | fatigue | no power, without energy, exhausted, tired, lethargic, burnout, exhaustion | 1190 |
| 45 | Pandemic | covid | corona, coronavirus, virus, covid pandemic business, Corona, vaccine, worrying COVID, severity COVID, VIRUS, #CoronavirusPandemic, SARS CoV 2 INFECTION, vaccinated | 1073 |
| 46 | Lifestyle | physical activity | exercising, walking, sport, exercises, walk, gym, gyms | 1066 |
| 47 | Healthcare system | politics | health system, NHS, #brexit, brexit, Canada, capitalism, government, EU, administration, frustrated #NHS, economy, CANADA, CAPITALISM | 985 |
| 48 | Insulin | access insulin | no insulin, don't have insulin, without insulin, #insulin4all, #Insulin4all #Diaversary | 956 |
| 49 | Diabetes Technology | continuous glucose monitor | freestyle libre, #freestylelibre, monitoring, #dexcom, CGM, cgm | 934 |
| 50 | Insulin | affordable insulin | afford insulin | 915 |
| 51 | Complications & comorbidities | shock | dlaBeTiC SHOCK | 908 |
| 52 | Pandemic | home | staying home, quarantine, shutdown | 864 |
| 53 | Diabetes | management | control diabetes, uncontrol | 825 |
| 54 | Complications & comorbidities | infection | wound, wounds, inflammation | 822 |
| 55 | Symptoms | Insomnia | can't sleep, awake, wake, sleepy, asleep | 737 |
| 56 | Health | lost job | without work, laid off | 723 |
| 57 | Complications & comorbidities | diabetic ketoacidosis | keto acidosis, #ketoacidosis, diabetic ketoacidosis, DKA, #ketoacidosis, keto acidosis, DAIBETIC KETO ACIDS high, #KETO #NSNG | 710 |
| 58 | Health | immune system | | 705 |
| 59 | Nutrition | eating healthy | #healthy #meal | 608 |
| 60 | Blood pressure | hypotension | low blood pressure | 587 |
| 61 | Family | family | brother, daughter, grandpa, dad, mom, grandma, parent | 581 |
| 62 | Complications & comorbidities | renal failure | diabetes renal failure, dialysis | 554 |
| 63 | Complications & comorbidities | legs swollen | foot swelled | 544 |
| 64 | Diabetes | reverse diabetes | reversed, cured overnight | 489 |
| 65 | Diabetes community | support | #dsma, help, raise awareness, supporters | 483 |
| 66 | Lifestyle | lifestyle | environment | 453 |
| 67 | Emotions | love | like | 451 |
| 68 | Other | cold | | 443 |
| 69 | Complications & comorbidities | cholesterol | | 392 |
| 70 | Other | meat fake meat | | 376 |
| 71 | Other | starve | | 363 |
| 72 | Diabetes Distress | isolation | alone, live alone, distrust | 358 |
| 73 | Other | cut back rice | | 348 |
| 74 | Other | taking necessary precautions | | 346 |
| 75 | Complications & comorbidities | pancreas | diabetes pancreas | 340 |
| 76 | Health | PCOS | pcos, PCOs, Pcos | 333 |
| 77 | Other | ass alive | | 318 |
| 78 | Other | arthritic | | 312 |
| 79 | Other | dangerous | | 301 |
| 80 | Other | vulnerable | | 299 |
| 81 | Other | seizures | | 285 |
| 82 | Other | acting | | 274 |
| 83 | Other | needles sensation | | 273 |
| 84 | Other | slightly concerned | | 267 |
| 85 | Insulin | insulin spike | insulin jump | 263 |
| 86 | Complications & comorbidities | liver failure | diabetes liver failure | 256 |
| 87 | Other | fucked drunk | | 253 |
| 88 | Complications & comorbidities | stomach | | 246 |
| 89 | Symptoms | thirsty | thirst, dehydrated, THIRSTY | 245 |
| 90 | Other | stop talking | | 244 |
| ID | Category                  | Term                        | Notes                                      |
|----|---------------------------|-----------------------------|--------------------------------------------|
| 91 | Nutrition                 | alcohol beer, BEER, alcoholism |                                           |
| 92 | Glycemic variability      | A1C a1c predict HbA 1c      |                                           |
| 93 | Health                    | genetic genes, Genetics, hereditary, genetically modified, shit genetics |                                           |
| 94 | Complications & comorbidities | cancer chemo, #Cancer     |                                           |
| 95 | Other                     | hair fall                   |                                           |
| 96 | Other                     | crash hard                  |                                           |
| 97 | Other                     | shut kidneys                |                                           |
| 98 | Other                     | isolated ' society          |                                           |
| 99 | Other                     | afraid ingredients          |                                           |
| 100| Other                     | surgeon excited             |                                           |