

# **The impact of the COVID-19 pandemic on the self-management of Type 2 Diabetes: An Investigation of Reddit Data**

Meghan Nagpal, Niloofar Jalali, Diana Sherifali, Plinio Morita, Joseph Cafazzo

Submitted to: Interactive Journal of Medical Research  
on: July 22, 2023

**Disclaimer:** © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

## ***Table of Contents***

---

<b>Original Manuscript.....</b>	<b>5</b>
---------------------------------	----------

Preprint  
JMIR Publications

# The impact of the COVID-19 pandemic on the self-management of Type 2 Diabetes: An Investigation of Reddit Data

Meghan Nagpal<sup>1, 2</sup> MSc, PEng; Niloofar Jalali<sup>3</sup> PhD; Diana Sherifali<sup>4</sup> PhD, RN, CDE; Plinio Morita<sup>1, 2, 3</sup> PhD, PEng; Joseph Cafazzo<sup>1, 2, 5, 6</sup> PhD, PEng

<sup>1</sup>Institute of Health Policy, Management, & Evaluation Dalla Lana School of Public Health University of Toronto Toronto CA

<sup>2</sup>Centre for Digital Therapeutics Techna Institute University Health Network Toronto CA

<sup>3</sup>School of Public Health & Health Systems University of Waterloo Kitchener CA

<sup>4</sup>School of Nursing McMaster University Hamilton CA

<sup>5</sup>Institute of Biomedical Engineering University of Toronto Toronto CA

<sup>6</sup>Department of Computer Science University of Toronto Toronto CA

## Corresponding Author:

Meghan Nagpal MSc, PEng

Institute of Health Policy, Management, & Evaluation

Dalla Lana School of Public Health

University of Toronto

Health Sciences Building

4th Floor

Toronto

CA

## Abstract

**Background:** Type 2 Diabetes (T2D) is a chronic disease that can be managed in part through healthy behaviours. However, the COVID-19 pandemic impacted how people managed their condition. Using social media forums and analytics through Patient-Generated Health Data (PGHD) presents an opportunity to understand the health behaviours from the perspective of the patient.

**Objective:** Our objective is to understand how the health behaviours and attitudes of people living with T2D were impacted by the early stages of the COVID-19 pandemic by examining Reddit forums (using PGHD) for people living with T2D.

**Methods:** Data from the Reddit forums related to T2D from January 2018 to early March 2021 were downloaded, and Support Vector Machines (SVMs) were used to classify if a post was made in the context of the pandemic. Latent Dirichlet Allocation (LDA) topic modelling was performed to gather topics of discussion amongst the entire dataset and a subsequent iteration was performed to gather topics of discussion specific to the COVID-19 pandemic. Sentiment Analysis using the Valence Aware Dictionary for sEntiment Reasoning (VADER) algorithm was performed to gauge attitudes towards the pandemic.

**Results:** Of all posts, topics of discussion were classified into themes of Managing Lifestyle, Managing Blood Glucose, Obtaining Diabetes Care, and Coping & Receiving Support. Amongst the COVID-specific posts topics of discussion were Coping with Poor Mental Health, Accessing Doctor & Medications and Controlling Blood Glucose, Changing Food Habits during Pandemic, Impact of Stress of Blood Glucose Levels, Changing Status of Employment & Insurance, Risk of COVID Complications. Overall, posts classified as COVID-related had were associated with lower sentiment than those classified as “noncovid.”

**Conclusions:** Topics of discussion gauged from the Reddit forums provide a holistic perspective of the impact of the pandemic on people living with T2D. Overall, the early stages of the pandemic negatively impacted the attitudes of people living with T2D.

(JMIR Preprints 22/07/2023:51154)

DOI: <https://doi.org/10.2196/preprints.51154>

## Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ Please make my preprint PDF available to anyone at any time (recommended).

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.  
Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <http://www.jmir.org/preprint/51154>, my full manuscript will be made available to all users.



## Original Manuscript

## Original Paper

### The impact of the COVID-19 pandemic on the self-management of Type 2 Diabetes: An Investigation of Reddit Data

Meghan S. Nagpal<sup>1,2</sup>, PEng, MSc; Niloofar Jalali<sup>3</sup>, PhD; Diana Sherifali<sup>4</sup>, RN, PhD, CDE; Plinio P. Morita<sup>1,2,3</sup>, PEng, PhD; Joseph A. Cafazzo<sup>1,2,5,6</sup>, PEng, PhD

<sup>1</sup>Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

<sup>2</sup>Centre for Digital Therapeutics, Techna Institute, University Health Network, Toronto, ON, Canada

<sup>3</sup>School of Public Health and Health Systems, University of Waterloo, Waterloo, ON, Canada

<sup>4</sup>School of Nursing, McMaster University, Hamilton, ON, Canada

<sup>5</sup>Institute of Biomedical Engineering, University of Toronto, Toronto, ON, Canada

<sup>6</sup>Department of Computer Science, University of Toronto

## Correspondence:

### Ms. Meghan S. Nagpal, MSc

Institute of Health Policy, Management and Evaluation

Dalla Lana School of Public Health

University of Toronto

Health Sciences Building, 4<sup>th</sup> floor

155 College St

Toronto, ON, M5T 3M6, Canada

Phone: 1 416 978 4326

Fax: 1 416 978 7350

meghan.nagpal {at} mail.utoronto.ca

## Abstract

### Background

Type 2 Diabetes (T2D) is a chronic disease that can be managed in part through healthy behaviours. However, the COVID-19 pandemic impacted how people managed their condition. Using social media forums and analytics through Patient-Generated Health Data (PGHD) presents an opportunity to understand the health behaviours from the perspective of the patient.

## Objective

Our objective is to understand how the health behaviours and attitudes of people living with T2D were impacted by the early stages of the COVID-19 pandemic by examining Reddit forums (using PGHD) for people living with T2D.

## Methods

Data from the Reddit forums related to T2D from January 2018 to early March 2021 were downloaded, and Support Vector Machines (SVMs) were used to classify if a post was made in the context of the pandemic. Latent Dirichlet Allocation (LDA) topic modelling was performed to gather topics of discussion amongst the entire dataset and a subsequent iteration was performed to gather topics of discussion specific to the COVID-19 pandemic. Sentiment Analysis using the Valence Aware Dictionary for sEntiment Reasoning (VADER) algorithm was performed to gauge attitudes towards the pandemic.

## Results

Of all posts, topics of discussion were classified into themes of Managing Lifestyle, Managing Blood Glucose, Obtaining Diabetes Care, and Coping & Receiving Support. Amongst the COVID-specific posts topics of discussion were Coping with Poor Mental Health, Accessing Doctor & Medications and Controlling Blood Glucose, Changing Food Habits during Pandemic, Impact of Stress of Blood Glucose Levels, Changing Status of Employment & Insurance, Risk of COVID Complications. Overall, posts classified as COVID-related had were associated with lower sentiment than those classified as “noncovid.”

## Conclusion

Topics of discussion gauged from the Reddit forums provide a holistic perspective of the impact of the pandemic on people living with T2D. Overall, the early stages of the pandemic negatively impacted the attitudes of people living with T2D.

## Keywords

Diabetes Mellitus, Type 2; COVID-19; Pandemics; Social Media; Health Behavior; Health Knowledge; Attitudes; Practice

## Introduction

### Background

Diabetes is a serious metabolic condition in which the body experiences elevated blood glucose levels which can result in serious complications such as cardiovascular disease (CVD), kidney disease, stroke, eye disease, foot ulcers, nerve damage, and amputation. The World Health

Organization (WHO) [1] states that high blood glucose levels are the third leading cause of premature mortality. Type 2 Diabetes (T2D) is characterized by the body's resistance or insufficient production of insulin. Research suggests that the risks of further complications for people living with T2D can be mitigated through self-management, including managing glucose levels, optimizing nutrition, regular physical activity, and taking prescribed medications [3, 4].

## **Managing T2D During COVID-19 Pandemic**

With the emergence of the COVID-19 pandemic in March 2020, near the entirety of the global population was impacted by social distancing measures that included business closures, remote school and work measures, prohibition of large crowds, limited socialization and virtual outpatient diabetes care delivery. Beyond the fear and anxiety that manifested [12], individuals living with T2D were at an increased risk for complications from COVID-19, as the virus thrives in an environment of high blood glucose [13]. However, because T2D is a disease managed by lifestyle behaviours, COVID-19 restriction measures reduced physical activity and increased other unhealthy behaviours amongst those living with T2D and other chronic conditions [12].

## **Rationale & Study Objectives**

The objectives of this study were to synthesize emerging themes of peer discussion on social media for people coping with T2D and understand how they were affected and how diabetes management changed during the early part of the pandemic.

## **Methods**

### **Overview**

Latent Dirichlet Allocation (LDA) topic modelling [15] and sentiment analysis (SA) with the



Valence Aware Dictionary for sEntiment Reasoning (VADER) algorithm [17] were performed to achieve the study objectives.

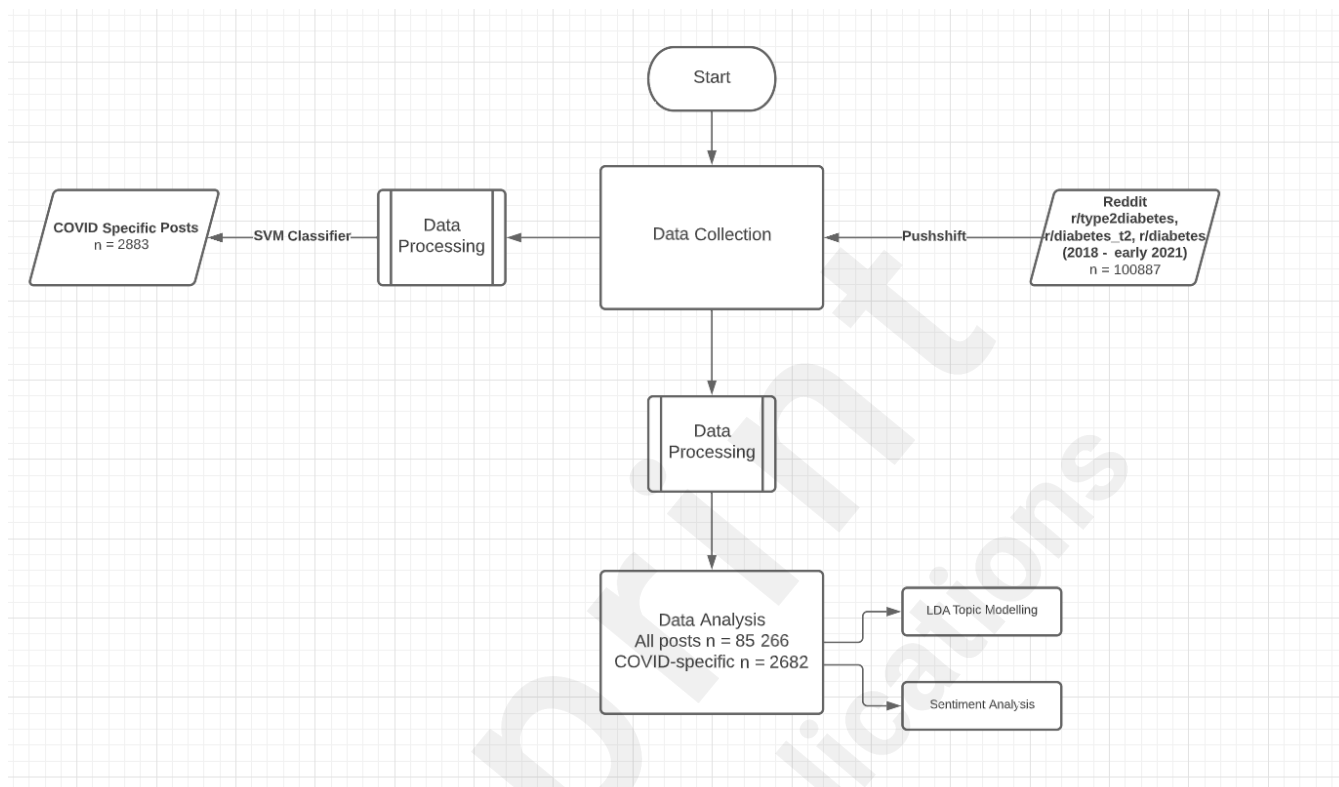


Figure 1 Overview of Methodology

## Data Collection

For this study, three communities on Reddit were examined: r/type2diabetes, r/diabetes\_t2, and r/diabetes [18 - 20]. From the r/diabetes community, only posts that were tagged with the “flair” “Type 2 Diabetes” were examined. In a period from January 1, 2018, to March 5, 2021, 100 887 posts from these three communities were collected from Reddit using the Application Programming Interface (API) Pushshift [21].

## Classification of Posts

Manual examination of the dataset found that the very first post made about the COVID-19 pandemic was made on February 28, 2020 and coincidentally, the last post about the pandemic in the dataset was made on February 28, 2021. In total, there were 48 988 posts. Within the dataset, terms

related to the COVID-19 pandemic were manually searched for. These terms include: COVID; Coronavirus; Pandemic; Social Distancing; Lockdown; Quarantine; Toilet paper; Unemploy(ed); Work(ing) from home; Telehealth; Vaccine; Sanitizer; Mask.

Posts that contained those terms in the text body were manually evaluated for context and labeled as “covid” or “noncovid”. In total, 9803 posts were manually classified, with 2065 labelled as “covid” and 7738 labelled as “noncovid” and subsequently classified with the support vector machines (SVM). An additional 818 posts that were made in the context of the COVID-19 pandemic were identified, bringing the total number of pandemic-specific posts to 2883. The remainder of unlabelled posts made in the identified pandemic period were labeled as “noncovid.”

## Data Analysis

The Latent Dirichlet Allocation (LDA) topic modelling algorithm [31] was used in two iterations to obtain topics of discussion by obtaining clusters of words belonging to a single topic. In total of 85,266 posts analyzed, with 2682 being specific to the COVID-19 pandemic. The first iteration consisted of performing topic modelling on the entire processed dataset of 85,266 posts made between January 1, 2018 and March 5, 2021. The second iteration of topic modelling was performed on a subset of this dataset, the 2682 posts made specifically in the context of the COVID-19 pandemic.

Sentiment Analysis (SA) was performed to understand the subjective emotions, or sentiment, associated with each post, with a normalized compound score between -1 and +1. Thresholds for classifying a text as per VADER are as follows [32]: **Positive:** compound score  $\geq +0.05$ ; **Neutral:** compound score between -0.05 and +0.05; and **Negative:** compound score  $\leq -0.05$ .

## Results

### Topics of Discussion

The final processed dataset consisted of 85,266 posts made between January 1, 2018, and March 3, 2021. The topics were manually labelled and placed into umbrella themes as follows:

*Table 2 Themes, topics, and associated words for all posts*

Theme	Topic Name	Words
Managing Blood Glucose	Using Blood Glucose Meters	meter, insurance, reading, check, strip, libre, sensor, finger, cover, cgm, free, pay, app, stick, test, track, buy, freestyle, monitor, difference
	Controlling Blood Glucose	glucose, blood, high, glucose, level, normal, low, range, control, number, spike, check, average, mgdl, raise, point, fast, affect, measure, test
	Controlling A1C Levels	ac, month, weight, year, lose, start, med, diagnose, metformin, week, diet, stop, lb, change, medication, drop, loss, exercise, pound, control
	Rationale for Glucose Variability	insulin, type, body, resistance, gt, disease, people, medication, risk, cure, increase, reverse, cell, case, pancreas, term, treatment, long, produce, diabetic
Managing Lifestyle	Exercising for Diabetes Care	work, time, exercise, walk, bit, long, pretty, day, start, minute, run, ill, thing, week, stress, home, stay, big, lot, hit
	Timing Meals & Snacks	fast, hour, meal, bg, metformin, morning, time, number, week, mg, night, eat, effect, day, start, low, dinner, breakfast, sleep, reading
	Eating Low Carbohydrate / Ketogenic Diet	diet, low, keto, carb, fat, body, protein, calorie, lot, liver, carbs, healthy, process, cholesterol, reduce, energy, high, carbohydrate, intake, muscle
	Controlling Carbohydrates & Alcohol	eat, carbs, food, drink, water, cut, meal, carb, avoid, lot, spike, soda, fiber, drinking, limit, small, alcohol, gram, sweet, ate
	Options for Food	bread, rice, add, cheese, egg, potato, meat, cream, fruit, veggie, chicken, taste, pasta, recipe, snack, butter, nut,

Coping & Receiving Support	Attitudes towards Disease	coffee, salad, milk change, life, great, lot, thing, work, dr, make, learn, hope, luck, live, easy, control, love, healthy, manage, hard, lifestyle, sound
	Caregiving for Family & History	feel, time, family, year, felt, thought, happen, care, make, deal, mom, friend, turn, kind, super, shit, die, guess, hard, suck
Obtaining Diabetes Care	Interacting with Care Team	doctor, test, result, advice, talk, doc, give, diagnose, endo, read, appointment, wait, endocrinologist, follow, diagnosis, time, thought, adjust, mention, idea
	Educating for Self-Management	people, post, health, question, medical, understand, research, information, study, read, answer, agree, link, support, comment, patient, diabetic, base, opinion, group
	Managing Complications / Co-morbidities	issue, problem, symptom, pain, foot, effect, bad, eye, year, damage, heart, experience, kidney, stomach, hand, stop, happen, shot, neuropathy, vision

A subset of 2682 posts made between February 28, 2020, and February 28, 2021 were classified as COVID-specific. The topics were manually labelled as follows:

Table 4 Topics and associated words for COVID-related posts

Topic Name	Words
Coping with Poor Mental Health	feel, day, week, work, weight, lose, walk, bit, symptom, bad, lot, gym, night, happen, end, great, covid, felt, ill, diagnosis
Accessing Doctor & Medications and Controlling Blood Glucose	doctor, test, ac, year, insulin, month, metformin, diagnose, low, glucose, diet, week, med, exercise, change, level, stop, reading, check, medication
Changing Food Habits during Pandemic	eat, food, carbs, lot, meal, low, carb, diet, hard, hour, thing, water, keto, fast, cut, easy, glucose, add, stuff, rice

Impact of Stress on Blood Glucose Levels	blood, glucose, high, time, stress, number, long, exercise, bg, body, problem, morning, normal, control, sleep, make, level, effect, change, kind
Changing Status of Employment & Insurance	work, home, hospital, today, time, year, talk, give, advice, wait, insurance, visit, guess, state, order, meter, live, strip, situation, friend
Risk of COVID Complications	covid, type, health, care, risk, sick, control, issue, virus, disease, diabetic, question, pandemic, mask, hand, case, infection, wear, patient, home

## Sentiment Analysis

Using the VADER algorithm, the compound sentiment score of each post was determined. For all posts in the dataset, the sentiment was on average positive with a compound score of 0.23 (95% CI, 0.23 – 0.24).

## Comparison of Sentiment of COVID versus “Noncovid” Posts

The sentiment was compared amongst the posts that were classified as “covid” and “noncovid.” The mean compound score of COVID-related posts still falls in the threshold of positive (0.12, 95% CI 0.01 – 0.15), but its intensity score is still lower than that of “noncovid” posts (0.25, 95% CI 0.24 – 0.25).

## Sentiment by Theme

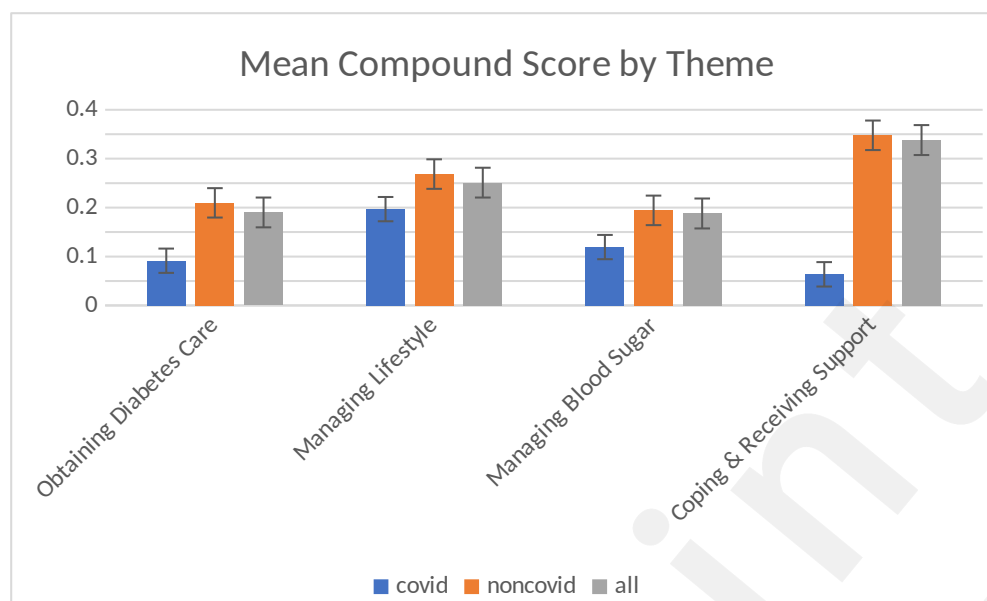


Figure 7 Average compound score by theme

Coping & Receiving Support had the highest mean compound score (0.34, 95% CI 0.33 - 0.35), followed by Managing Lifestyle (0.25, 95% CI 0.25 - 0.26), Obtaining Diabetes Care (0.19, 95% CI 0.18 - 0.20), and Managing Blood Glucose (0.19, 95% CI 0.18 - 0.19). All mean compound scores fall in the threshold to be classified as positive, with Coping & Receiving Support demonstrating the strongest positive intensity score.

## Sentiment by Theme Amongst COVID-Classified and “Noncovid” Posts

Mean compound scores were further analyzed amongst posts classified as COVID-related or as “noncovid.” Coping & Receiving Support had the highest mean compound score (0.35, 95% CI 0.33 - 0.36), followed by Managing Lifestyle (0.27, 95% CI 0.26 - 0.28), Obtaining Diabetes Care (0.21, 95% CI 0.20 - 0.22), and Managing Blood Glucose (0.19, 95% CI 0.18 - 0.20). However, for COVID-classified posts Coping & Receiving Support had the lowest mean compound score (0.06, 95% CI -0.02 - 0.13), which is barely above the threshold to classify a post as positive. Continuing to

rank from lowest to highest, Obtaining Diabetes Care had the second-lowest mean compound score (0.09, 95% CI 0.05 - 0.13), followed by Managing Blood Glucose (0.12, 95% CI 0.08 - 0.16), and Managing Lifestyle (0.20, 95% CI 0.15 - 0.24).

## Sentiment by COVID-specific Topics

Sentiment was analyzed again amongst the topics generated for the COVID-specific posts.

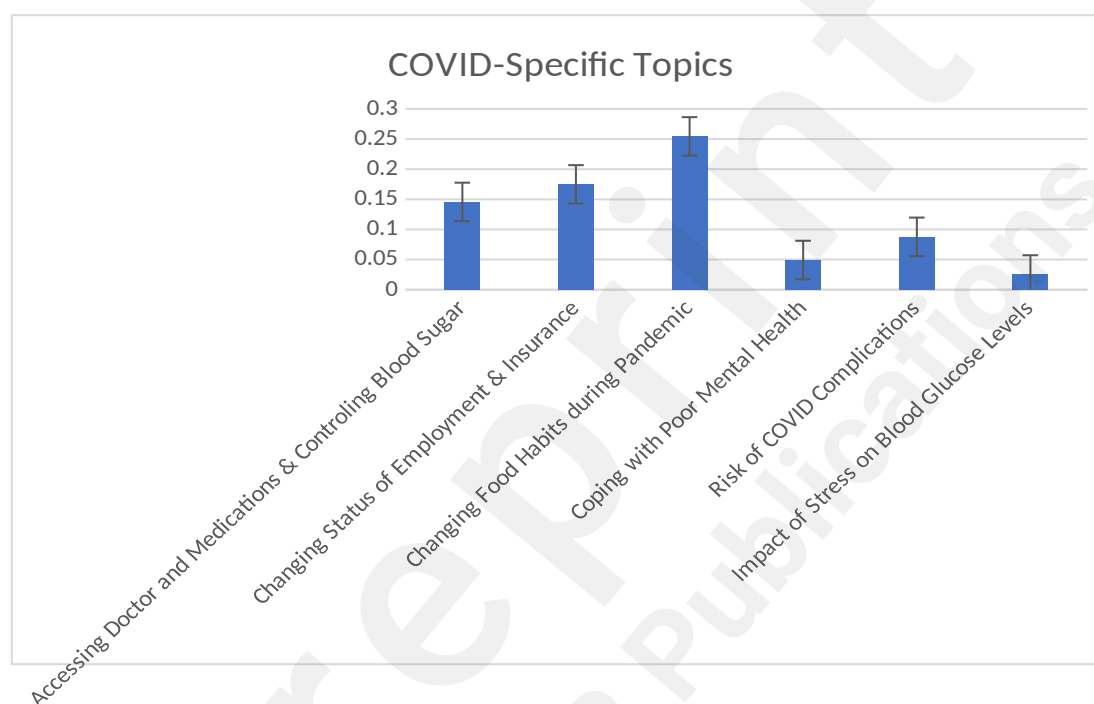


Figure 8 Mean compound score of COVID-specific topics

Impact of Stress on Blood Glucose Levels and Coping with Poor Mental Health had mean compound scores that fell in the threshold of being classified as neutral (0.03, 95% CI -0.03 - 0.08 and 0.04, 95% CI -0.01 - 0.11 respectively). Risking COVID Complications, Accessing Doctor and Medications & Controlling Blood Glucose, and Changing Status of Employment & Insurance were ranked next from lowest to highest with mean compound scores of 0.09 (95% CI 0.03 - 0.14), 0.14 (95% CI 0.09 - 0.20), and 0.17 (95% CI 0.13 - 0.22) respectively. And finally, Changing Food Habits had the highest mean compound score of 0.25 (95% CI 0.2 - 0.31).

## Discussion

## Topics Discussed on Reddit Communities

Our study determined that T2D subreddits were indeed used for peer support to discuss themes related to diabetes management in depth. Oyeboade et al. [6] and Griffis et al. [8] both determined that themes of discussion of diabetes management on online platforms pertained to nutrition, lifestyle, symptoms, treatments, research, risk factors, and supplements/remedies. Additionally, we found, topics of discussion also included glycemic control and support-seeking behaviours.

From the LDA topic modelling algorithm, topics discussed on Reddit communities for people managing T2D include glycemic control, exercise, and nutrition, indicating users are discussing health behaviours. However, additionally, users are also discussing their attitudes towards their disease, their interactions with their families and caregivers, and management of comorbidities and also sharing educational resources. This is an indication that online forums provide opportunities for holistic discussion that considers the entire person and the social and psychological factors that impact the management of diabetes. This attributes to the fact that the possibilities for discussion are endless and such discussions may not necessarily be captured in a structured clinical setting.

## COVID-Specific Discussions on Reddit Forums

Topics gauged from covid-specific posts were about accessing medical care, the impact of stress on glycemic control, changing food habits, poor mental health, and fear of complications due to COVID. However, additionally, we found that the changing status of employment & insurance status was also a topic of discussion amongst the COVID-specific posts. This again furthers our finding that discussions amongst people living with T2D on Reddit offer insight from a holistic perspective that considers aspects of a person's life in the context of their disease.



## Sentiment Analysis

Amongst all posts in the dataset, the mean compound score of sentiment as per the VADER algorithm was 0.23 (95% CI, 0.23 – 0.24). As that score is above the threshold of 0.05, we can conclude that, on mean, the posts made on the Reddit communities for people living with diabetes were positive.

Between the COVID-related and “noncovid” posts, there was a significant difference in sentiment. Overall, the mean compound score of posts labelled as “noncovid” was 0.24 (95% CI 0.24 – 0.25) and posts labelled as COVID had a mean compound score of 0.12 (95% CI 0.01 – 0.15), which is significantly lower. While we cannot conclude that the COVID-labelled posts were overall negative in tone, the intensity is lower.

Amongst the “noncovid” posts, Coping & Receiving Support was associated with the highest mean compound score (0.35, 95% CI 0.33- 0.36). However, amongst the COVID-related posts, Coping & Receiving Support had the lowest mean compound score (0.06, 95% CI -0.01 - 0.13), barely above the threshold to be classified as positive. This is an indication that the pandemic negatively impacted people living with the disease.

Specifically looking at the topics generated only amongst the COVID-related posts, the topic with the highest mean compound score was Changing Food Habits (0.25, 95% CI 0.20 – 0.31), indicating that even though eating patterns were impacted during the pandemic, there was peer support and camaraderie amongst discussions about food choices. Further within the COVID-related posts, the Impact of Stress on Blood Glucose Levels and Coping with Poor Mental Health had the lowest mean compound scores (0.03, 95% CI -0.03 - 0.08 and 0.05, 95% CI -0.01 – 0.11 respectively). These low sentiment scores suggest that the pandemic resulted in anxiety that affected glycemic control and resulted in poorer mental health outcomes.

## Impact of Pandemic on Managing T2D

Our study found themes about managing T2D during the pandemic from Reddit communities and associated sentiment towards these themes. Below we will discuss findings gauged from our study, specifically regarding the impact of the COVID-19 pandemic on the management of T2D.

### Access to Diabetes Care during Pandemic

During the COVID-19 pandemic, accessing care was perceived as a significant barrier to managing T2D as more clinical visits were done virtually. Khader, Jabeen, and Namaju [33] found in a study of 1582 participants in India that the frequency of clinical visits was reduced in 87.28% of participants and 87.81% of participants who did not have access to health services. Al-Sofiani et al. [24] found in a study of 568 participants in the Arab Gulf region that lack of communication with healthcare providers was associated with higher odds of depression and anxiety amongst people with diabetes. This was further confirmed through our sentiment analysis. Amongst all posts, the mean compound score of posts that fell into the theme of Obtaining Diabetes Care was 0.19 (95% CI, 0.18 – 0.20), but amongst the COVID-specific posts, the mean compound score was significantly lower at 0.09 (95% CI, 0.05 - 0.13).

Further topic modelling amongst the COVID-specific posts also identified Access to Doctor & Controlling Blood Glucose as a topic of discussion. While barriers to accessing healthcare providers existed, another barrier was presented through the fear of acquiring a COVID infection and avoiding hospitalization in potentially dangerous situations. Further research would need to examine the impact of the lessening of in-person healthcare visits amongst people living with T2D.

### Impact to Employment during Pandemic

Efforts to curb the spread of the virus resulted in employers requiring employees to work from home [35]. As this was a disruption in daily life, it can be inferred from Reddit discussions that

this disruption impacted the health behaviours of those managing T2D.

In addition to work-from-home measures, the pandemic also resulted in the loss of employment due to the shutting down of businesses as a means to curb the spread of the virus. Unemployment already poses the issue of loss of income and standard of living and decreased sense of self-purpose [39], potentially impacting health behaviours due to increased stress. However, many people unemployed also were impacted by changes to their health insurance because of their job loss [40, 41]. From the Reddit discussions, it could be inferred that a loss of insurance posed a barrier for people living with T2D to get medication and blood glucose meters and strips, affecting glycemic control.

## **Glycemic Control during Pandemic**

Managing Blood Glucose was a major theme of discussion throughout our dataset. Glycemic control is a major component in the self-management of T2D and nonetheless, it would be expected to be a major theme of discussion amongst peers. Specifically, during the pandemic, it was found through our analysis that users of the Reddit communities for managing T2D claimed that the stress of the pandemic had an impact on blood glucose levels. Few laboratory studies have demonstrated that psychological stressors have been linked to hyperglycemia [42 – 44].

Overall, Managing Blood Glucose was the theme with the lowest mean sentiment score (0.19, 95% CI 0.18 - 0.19) and that score was reduced amongst COVID-related posts (0.12, 95% CI 0.08 - 0.16). This correlation already suggests that attitudes towards managing blood glucose are associated with lower sentiment than other themes and that the pandemic further decreased attitudes towards managing blood glucose. Specifically, amongst the COVID-related posts, the topic Impact of Stress on Blood Glucose Levels also had the lowest sentiment score (0.03, 95% CI -0.03 - 0.08), suggesting that users of the Reddit communities had increased anxiety about glycemic control.

Another finding was that the topic associated with accessing a doctor also included terms

related to controlling blood glucose. This suggests that reduced access to a healthcare provider may have impacted glycemic control amongst people living with T2D through reduced access to measuring A1c or to medications during the pandemic. Changes to employment also were associated with reduced access to health insurance, which was further extrapolated to reduced access to medications and blood glucose meters & strips, affecting glycemic control.

## **Lifestyle Management during Pandemic**

In our dataset, lifestyle management was the most frequent theme of discussion, with a positive mean sentiment score of 0.25 (95% CI, 0.25 - 0.26). Amongst the COVID-specific posts, that sentiment score dropped to 0.20, (95% CI, 0.15 - 0.24), but was overall positive. Even amongst the COVID-specific posts, Changing Food Habits during the Pandemic was associated with the highest mean sentiment (0.25, 95% CI 0.20 – 0.31). This may be an indication that increased unhealthy food consumption may have been a coping mechanism associated with positive emotions through the stressful time.

Interestingly, in the topic Coping with Poor Mental amongst the COVID-specific posts, the terms “gym” and “walk” were included. This suggests that a source of reduced physical activity was attributed to gym closures in the initial months of the pandemic and people relying on walking as a means of physical activity. Furthermore, physical activity is also another means of coping with stressors and many people were unable to be physically active while staying at home. Further investigation by clinicians is needed to help people living with T2D cope with changes, such as pandemic lockdowns so that they can continue to maintain their lifestyle habits during times of stress.

## **Mental Health during the Pandemic**

The mental health impact of this pandemic is expected to be long-term due to the extreme measures that were necessary to prevent the spread of the virus and the resulting economic

implications [45]. People living with T2D were no exceptions to this. In our study, we found that while Coping & Receiving Support was associated with the highest mean sentiment overall in our dataset (0.33, 95% CI 0.33 - 0.35), amongst the COVID-related posts, it was associated with the lowest mean sentiment (0.06, 95%CI -0.01 – 0.13). This is a suggestion that while the Reddit forums for managing T2D have been helpful overall for people coping with the disease, the pandemic negatively impacted how people were coping and their attitudes towards their disease. Specifically, within the COVID-related posts, discussions about Coping with Poor Mental Health and Risk of COVID Complications were associated with lower mean sentiment scores (0.05, 95% CI -0.01 - 0.11 and 0.09, 95% CI 0.03 - 0.14, respectively). While our study demonstrated that people living with T2D were utilizing peer support to cope with the stressors of the pandemic, it also demonstrates that users were negatively impacted by the psychological stressors of the pandemic. Moreover, the implications of potentially being exposed to the virus posed another source of anxiety given the vulnerability to serious complications of COVID amongst people living with T2D. As discussed in this paper, implications of poor mental health and anxiety towards COVID exposures impacted glycemic control and lifestyle management, which are necessary to prevent complications from diabetes. Furthermore, the anxiety of COVID exposure also posed as a barrier for people living with T2D to seek help from urgent care centres and hospitals when necessary.

## Conclusion

### Themes of Discussion

Through LDA topic modelling from 85,266 posts made between January 2020 and early March 2021 on Reddit forums for people managing T2D, we found fourteen topics of discussion that fell into four themes: Managing Blood Glucose, Managing Lifestyle, Coping & Receiving Support, Obtaining Diabetes Care. These findings are in line with recommended health behaviours to manage T2D that include glycemic control, healthy nutrition, increasing physical activity, reducing sedentary

time, and taking prescribed medications. Additionally, discussions also encompassed attitudes towards the disease, interactions with family members and caregivers, managing comorbidities, and sharing educational resources. This attests that while Reddit forums for diabetes management are a great support network for people managing T2D, it also is an indication that discussions are holistic in nature and that they go beyond treating the symptoms of the disease. Amongst COVID-specific posts, topics of discussion were: Coping with Poor Mental Health, Accessing Doctor & Medications and Controlling Blood Glucose, Changing Food Habits during Pandemic, Impact of Stress on Blood Glucose Levels, Changing Status of Employment & Insurance, and Risk of COVID Complications

### **Impact to Attitudes towards Managing T2D**

Posts classified as COVID-related had a significantly lower mean sentiment score than those posts classified as “noncovid”, suggesting that the pandemic negatively affected the attitudes of those living with T2D. Amongst posts classified as COVID-related, those with topics of mental health and the impact of stress on glycemic control were associated with lower sentiment, whereas posts relating to changing food habits were associated with higher sentiment.

### **Significance of Findings**

Our study demonstrates that this analysis of social media is a simpler and effective to understand the behaviours and attitudes of people managing chronic diseases, such as diabetes when compared to other qualitative inquiry methods. Topics of discussion and sentiment can potentially be generated in real-time and over a long period, directly from the perspective of the patient. This is also unlike clinical visits, in which data is generated from the perspective of the clinician and only presents a snapshot of a particular moment. The benefit of social media posts is that the possibilities for discussion are infinite and are more holistic, potentially capturing more content than what would have otherwise been missed in structured surveys or interviews.

This study has significance on both macro and micro levels. On a macro scale, harnessing

data from social media forums can provide an epidemiological landscape of behavioural trends of people living with T2D and potentially guide public health policy and help formulate treatment guidelines, especially during public health emergencies such as the pandemic.

## Limitations of Study

While topic modelling and sentiment analysis was performed on the dataset, the researchers did not thoroughly examine each post for context and cannot comment on the quality of discussions made on the forum. We cannot confirm that the users who made the posts were all people living with T2D. We only examined the three Reddit communities mentioned in the study and not other subreddits COVID-19 or mental health.

Furthermore, there is no information about the demographics of users specific to the diabetes forums of Reddit and we assumed that the demographics were similar to the demographics of all Reddit users. Under this assumption, there could have been a sampling bias in our results as Reddit users are mostly male and fall into the age demographic of 18 – 49 [46, 47]. By country, the United States has the largest number of Reddit users, with other users mostly residing in developed, English-speaking countries [48]. As COVID implications, such as case numbers, lockdown policies, and attitudes towards pandemic, varied through different geographical areas, even within countries, it was beyond the scope of this study to correlate our findings with those implications as we did not have geographical information of the users themselves.

## Using Reddit Data

The data collected for this project was through the Pushshift API. Over 100 published research studies have already used Reddit data harnessed through this API [49]. Proferes et al. [51] conducted a systematic analysis of 727 manuscripts that used Reddit as a data source and found that less than 15% of manuscripts mentioned any form of ethical review.

While the username of the post authors was obtained, they were assumed to be pseudonyms of

the user and not their actual names. However, we do acknowledge that some users may have used integrated their real name in their username, but in this study, the usernames were not analyzed to confirm so. Reddit also does not provide identifiable characteristics of individual users, such as their name, gender, or geographical location. However, we do acknowledge that some users may put identifiable information within their text. For this study, this identifiable information was not harnessed or analyzed.

## Acknowledgments

MSN received guidance on the methodology from NJ and PPM. DS served as the clinical advisor for this study. All listed authors approved the final version of this paper.

## Conflicts of Interest

None to declare

## List of Abbreviations

API = Application Programming Interface

CVD = Cardiovascular Disease

LDA = Latent Dirichlet Allocation

NLTK = Natural Language Toolkit

PGHD = Patient-Generated Health Data

SA = Sentiment Analysis

SVM = Support Vector Machine

T2D = Type 2 Diabetes

TF-IDF = Term Frequency – Inverse Document Frequency

VADER = Valence Aware Dictionary for sEntiment Reasoning

WHO = World Health Organization

## References

1. World Health Organization. Global health risks: mortality and burden of disease attributable to selected major risks. World Health Organization; 2009.
2. Houlden RL. 2018 Diabetes Canada Clinical Practice Guidelines: Introduction. Canadian Journal of Diabetes. 2018;42(Suppl1):S1–5.
3. Sherifali D, Viscardi V, Bai J-W, Ali RMU. Evaluating the Effect of a Diabetes Health Coach in Individuals with Type 2 Diabetes. Can J Diabetes. 2016 Feb;40(1):84–94.
4. Diabetes Canada Clinical Practice Guidelines Expert Committee. Diabetes Canada 2018



- Clinical Practice Guidelines for the Prevention and Management of Diabetes in Canada. Canadian Journal of Diabetes. 2018;42(Suppl1):S1–325.
5. Abbar S, Mejova Y, Weber I. You Tweet What You Eat: Studying Food Consumption Through Twitter. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. New York, NY, USA: Association for Computing Machinery; 2015. p. 3197–206. (CHI '15).
  6. Oyeboode O, Orji R. Detecting Factors Responsible for Diabetes Prevalence in Nigeria using Social Media and Machine Learning. In: 2019 15th International Conference on Network and Service Management (CNSM). ieeexplore.ieee.org; 2019. p. 1–4.
  7. Lincke A, Lundberg J, Thunander M, Milrad M, Lundberg J, Jusufi I. Diabetes Information on Social Media [Internet]. Proceedings of the 11th International Symposium on Visual Information Communication and Interaction - VINCI '18. 2018. Available from: <http://dx.doi.org/10.1145/3231622.3232508>
  8. Griffis H, Asch DA, Andrew Schwartz H, Ungar L, Buttenheim AM, Barg FK, et al. Using Social Media to Track Geographic Variability in Language About Diabetes: Infodemiology Analysis. JMIR Diabetes. 2020;5(1):e14431.
  9. Matsumoto K, Ryu M, Yoshida M, Kita K. Emotion Analysis on Weblog of Lifestyle Diseases. In: Proceedings of the 2019 International Symposium on Signal Processing Systems. New York, NY, USA: Association for Computing Machinery; 2019. p. 142–8. (SSPS 2019).
  10. Liu Y, Yin Z. Understanding Weight Loss via Online Discussions: Content Analysis of Reddit Posts Using Topic Modeling and Word Clustering Techniques. J Med Internet Res. 2020 Jun 8;22(6):e13745.
  11. Fung IC-H, Tse ZT, Fu K-W. The use of social media in public health surveillance. Western Pacific Surveillance and Response Journal. 2015;6(2):3–6.
  12. Mental Health and COVID-19 [Internet]. World Health Organization. World Health Organization; [cited 2021Aug23]. Available from: <https://www.who.int/teams/mental-health-and-substance-use/covid-19>
  13. COVID-19 and diabetes [Internet]. International Diabetes Federation - Home. International Diabetes Federation ; 2021 [cited 2021Aug23]. Available from: <https://www.idf.org/aboutdiabetes/what-is-diabetes/covid-19-and-diabetes/1-covid-19-and-diabetes.html>
  14. Palmer K, Monaco A, Kivipelto M, Onder G, Maggi S, Michel J-P, et al. The potential long-term impact of the COVID-19 outbreak on patients with non-communicable diseases in Europe: Consequences for Healthy ageing. Aging Clinical and Experimental Research. 2020;32(7):1189–94.
  15. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. the Journal of machine Learning research. 2003 Mar 1;3:993-1022.
  16. Topic Modeling. Accessed August 24, 2021. <http://mallet.cs.umass.edu/topics.php>.
  17. "VaderSentiment." PyPI. Accessed August 24, 2021. <https://pypi.org/project/vaderSentiment/>.
  18. r/type2diabetes [Internet]. reddit. [cited 2021Aug24]. Available from: <https://www.reddit.com/r/type2diabetes/> 98.
  19. r/diabetes\_t2 [Internet]. reddit. [cited 2021Aug24]. Available from: [https://www.reddit.com/r/diabetes\\_t2/](https://www.reddit.com/r/diabetes_t2/) 99.
  20. r/diabetes [Internet]. reddit. [cited 2021Aug24]. Available from: <https://www.reddit.com/r/diabetes/>
  21. Pushshift. pushshift/api: Pushshift API [Internet]. GitHub. [cited 2021Aug24]. Available from: <https://github.com/pushshift/api>
  22. Source code for gensim.parsing.preprocessing [Internet]. gensim.parsing.preprocessing. [cited 2021Aug24]. Available from:

- <https://tedboy.github.io/nlps/modules/gensim/parsing/preprocessing.html>
23. spaCy 101: Everything you need to know · spaCy Usage Documentation [Internet]. spaCy 101: Everything you need to know. [cited 2021Aug24]. Available from: <https://spacy.io/usage/spacy-101>
  24. Source code for nltk.corpus [Internet]. nltk.corpus - NLTK 3.6.2 documentation. [cited 2021Aug24]. Available from: <https://www.nltk.org/modules/nltk/corpus.html>
  25. Singh S. What is Tokenization: Methods to Perform Tokenization [Internet]. Analytics Vidhya. 2019 [cited 2021Aug24]. Available from: <https://www.analyticsvidhya.com/blog/2019/07/how-get-started-nlp-6-unique-ways-perform-tokenization/>
  26. Stemming and lemmatization. [cited 2021Aug24]. Available from: <https://nlp.stanford.edu/IR-book/html/htmledition/stemming-and-lemmatization-1.html>
  27. Source code for nltk.stem.wordnet [Internet]. nltk.stem.wordnet - NLTK 3.6.2 documentation. [cited 2021Aug24]. Available from: <https://www.nltk.org/modules/nltk/stem/wordnet.html>
  28. sklearn.feature\_extraction.text.TfidfVectorizer [Internet]. scikit. [cited 2021Aug24]. Available from: [https://scikit-learn.org/stable/modules/generated/sklearn.feature\\_extraction.text.TfidfVectorizer.html](https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html)
  29. 1.9. Naive Bayes [Internet]. scikit. [cited 2021Aug24]. Available from: [https://scikit-learn.org/stable/modules/naive\\_bayes.html](https://scikit-learn.org/stable/modules/naive_bayes.html)
  30. 1.4 Support Vector Machines [Internet]. scikit. [cited 2021Nov25]. Available from: <https://scikit-learn.org/stable/modules/svm.html>
  31. Blei DM. Probabilistic topic models. Communications of the ACM. 2012 Apr 1;55(4):77-84.
  32. Bonta V, Janardhan NK. A comprehensive study on lexicon based approaches for sentiment analysis. Asian Journal of Computer Science and Technology. 2019;8(S2):1-6.
  33. Khader M, Jabeen T, Namoju R. A cross sectional study reveals severe disruption in glycemic control in people with diabetes during and after lockdown in India. Diabetes & Metabolic Syndrome: Clinical Research & Reviews. 2020;14(6):1579-1584.
  34. Al-Sofiani ME, Albunyan S, Alguwaihes AM, Kalyani RR, Golden SH, Alfadda A. Determinants of mental health outcomes among people with and without diabetes during the COVID -19 outbreak in the Arab Gulf Region. Journal of Diabetes. 2021;13(4):339-52.
  35. Savić D. COVID-19 and work from home: Digital transformation of the workforce. Grey Journal (TGJ). 2020 Jun 1;16(2):101-4.
  36. Killgore WD, Cloonan SA, Taylor EC, Dailey NS. Loneliness: A signature mental health concern in the era of COVID-19. Psychiatry research. 2020 Aug 1;290:113117.
  37. Fana M, Pérez ST, Fernández-Macías E. Employment impact of Covid-19 crisis: from short term effects to long terms prospects. Journal of Industrial and Business Economics. 2020 Sep;47(3):391-410.
  38. Congressional Research Service. Unemployment Rates During the COVID-19 Pandemic. Congressional Research Service; 2021.
  39. Unemployment and mental health [Internet]. Institute for Work & Health. 2009 [cited 2021 Aug 24]. Available from: [https://www.iwh.on.ca/sites/iwh/files/iwh/reports/iwh\\_issue\\_briefing\\_mental\\_health\\_2009.pdf](https://www.iwh.on.ca/sites/iwh/files/iwh/reports/iwh_issue_briefing_mental_health_2009.pdf)
  40. Gangopadhyaya A, Garrett AB. Unemployment, health insurance, and the COVID-19 recession. Health Insurance, and the COVID-19 Recession (April 1, 2020). 2020 Apr 1.
  41. Garrett AB, Gangopadhyaya A. How the COVID-19 recession could affect health insurance coverage. Available at SSRN 3598558. 2020 May 4.
  42. Goetsch VL, VanDorsten B, Pbert LA, Ullrich IH, Yeater RA. Acute effects of laboratory

- stress on blood glucose in noninsulin-dependent diabetes. *Psychosomatic medicine*. 1993 Nov.
43. Faulenbach M, Uthoff H, Schwegler K, Spinass GA, Schmid C, Wiesli P. Effect of psychological stress on glucose control in patients with type 2 diabetes. *Diabetic Medicine*. 2012 Jan;29(1):128-31.
  44. Bruce DG, Chisholm DJ, Storlien LH, Kraegen EW, Smythe GA. The effects of sympathetic nervous system activation and psychological stress on glucose metabolism and blood pressure in subjects with type 2 (non-insulin-dependent) diabetes mellitus. *Diabetologia*. 1992 Sep;35(9):835-43.
  45. Iqbal SZ, Li B, Onigu-Otito E, Naqvi MF, Shah AA. The long-term mental health effects of COVID-19. *Psychiatric Annals*. 2020Dec;50(12):522-5.
  46. Dean B. Reddit Usage and Growth Statistics: How Many People Use Reddit in 2021? [Internet]. Backlinko. 2021 [cited 24 August 2021]. Available from: <https://backlinko.com/reddit-users>
  47. Agrawal A. The user demographics of Reddit: The Official App [Internet]. Medium. 2021 [cited 24 August 2021]. Available from: [https://medium.com/@sm\\_app\\_intel/the-user-demographics-of-reddit-the-official-app-7e2e18b1e0e1](https://medium.com/@sm_app_intel/the-user-demographics-of-reddit-the-official-app-7e2e18b1e0e1)
  48. Reddit: traffic by country | Statista [Internet]. Statista. 2021 [cited 24 August 2021]. Available from: <https://www.statista.com/statistics/325144/reddit-global-active-user-distribution/>
  49. Baumgartner, J., Zannettou, S., Keegan, B., Squire, M., & Blackburn, J. (2020). The Pushshift Reddit Dataset. *Proceedings of the International AAAI Conference on Web and Social Media*, 14(1), 830-839. Retrieved from <https://ojs.aaai.org/index.php/ICWSM/article/view/7347>
  50. API Terms of Use [Internet]. reddit. 2016 [cited 2021Sep29]. Available from: <https://www.reddit.com/wiki/api-terms>
  51. Proferes N, Jones N, Gilbert S, Fiesler C, Zimmer M. Studying reddit: A systematic overview of disciplines, approaches, methods, and Ethics. *Social Media + Society*. 2021;7(2):205630512110190.
  52. Reddit user agreement [Internet]. reddit. [cited 2021Sep23]. Available from: <https://www.redditinc.com/policies/user-agreement-september-12-2021>