GOING WHERE PATIENTS ARE, PREDICTING PREDIABETES THROUGH FACEBOOK POSTINGS: A PILOT STUDY PROTOCOL

Xiaomeng Xu, PhD (CORRESPONDING AUTHOR)
Department of Psychology
Idaho State University
921 S 8th Ave, Stop 8112
Pocatello, ID 83209
(208) 282-3541
Fax: (208) 282-5411
xuxiao@isu.edu

Michelle L. Litchman, PhD, FNP-BC, FAANP
University of Utah
College of Nursing
10 S 2000 E
Salt Lake City, UT 84012
Michelle.Litchman@nurs.utah.edu

Perry Gee, PhD, RN
Dignity Health
Nursing Research and Analytics
185 Berry St
San Francisco, CA 94107
Perry.Gee@DignityHealth.org

Webb Whatcott, AA, PharmD Candidate
University of Utah
College of Pharmacy
1473 University Village
Salt Lake City, UT 84108
Webb.Whatcott@pharm.utah.edu

Loni Chacon, RN
Idaho State University
Department of Family Medicine, Clinical Research Center
465 Memorial Drive
Pocatello, ID 83201-4008
chacloni@isu.edu

John Holmes, PharmD
Idaho State University
Department of Family Medicine
465 Memorial Drive
Pocatello, ID 83201-4008
holmjohn@isu.edu

Sankara Subramanian Srinivasan, PhD, MBA
University of Utah - David Eccles School of Business
Spencer Fox Eccles Business Building
1655 E. Campus Center Dr.,
Salt Lake City, UT 84112-9304
sankar@eccles.utah.edu
Abstract

Background: The field of infodemiology uses healthcare trends found in public networks, such as social media, to track and quantify the spread of disease. Type 2 diabetes is on the rise worldwide and there may be utility in using social media to identify prediabetes through behavior exhibited through social media platforms such as Facebook.

Objective: This pilot study is designed to investigate the social media behavior of those with prediabetes, before and after diagnosis. Pre- and Post-diagnosis Facebook content (posts) of such individuals will be used to create a taxonomy of prediabetes indicators and identify if prediabetes indicators can predict actual diagnosis of prediabetes.

Methods: This study is a single center exploratory retrospective study design. The investigators will analyze Facebook posts 3 months before and 3 months after prediabetes diagnosis. Using predictive modeling, these prediabetes indicators will be tested to determine if a participant's own trajectory towards prediabetes can be confirmed.

Results: The study is anticipated to be complete by the end of 2018. Results will include both quantitative and qualitative data about participants and the similarities and differences between coded social media posts.

Conclusions: This pilot study is the first step in creating a taxonomy of social media indicators for prediabetes. Such a taxonomy would allow for researchers and health care professionals to identify those at greater risk for having prediabetes via their social media postings.

Key Words: infodemiology, prediabetes, Facebook, social media, protocol
Introduction

There are 86 million adults with prediabetes, with numbers rising in epidemic proportions [1]. The Centers for Disease Control and Prevention (CDC) has emphasized the need to confront prediabetes through National Diabetes Prevention Programs [1]. In 2014, the CDC reported that of the 37% of US adults who had prediabetes, 29% of those were unaware they had it and were unlikely to seek out treatment [1]. When prediabetes is not diagnosed and treated accordingly, it can increase an adult’s risk of heart disease, stroke, and type 2 diabetes, which have their own economic and quality of life burdens. Enhanced epidemiologic approaches are needed to identify large populations who are unaware they have prediabetes to maximize the efficacy of diabetes prevention programs.

Nearly two-thirds of American adults use social media [2]. Infodemiology, which is an epidemiologic approach that uses social media and other online sources to examine the spread or incidence of disease [3, 4], is one way to examine prediabetes among individuals who have not yet been diagnosed. Infodemiology techniques have been used in the past to successfully predict important issues of public health such as the likelihood of depression resulting in suicide among military service personnel [5, 6].

Infodemiology is one methodological approach to examine healthcare trends [3, 4]. This approach has been successful in identifying infectious diseases, such as influenza [7, 8], Zika [9], and suicide in veterans [5, 6]. Using infodemiology for surveillance (i.e., infoveillance) can be very effective in identifying many real-world health trends, however, has not been used to identify chronic conditions. Infectious diseases present with a sudden onset of symptoms, while the development of chronic conditions are more subtle. Given the association between prediabetes and lifestyle, and the fact that many individuals share their lifestyle on Facebook, this study uniquely examines the posts of those with prediabetes in the 3 months before and after their diagnosis. This innovative approach sets the stage to screen large populations for prediabetes who may have otherwise had a missed or delayed diagnosis.

Prediabetes

Only 6% of primary care providers (PCPs) can accurately identify all prediabetes indicators [10], potentially resulting in missed or delayed diagnosis. Further, with the primary care shortage of 1 physician for every 2,500 patients, lack of access to PCPs limits the capacity for the current healthcare system to make appropriate and timely diagnosis [11]. Thus, a predictive model of prediabetes using social media postings will allow us to identify individuals in need of screening whom PCPs may have missed. Further, we will also be able to identify individuals who have not yet seen a healthcare provider, yet are at risk for prediabetes.

Social Media

Rates of online social media use among adults is on the rise, with 8 in 10 adults using Facebook [12]. The social media digital divide is closing as more adoption is increasing in baby boomers and older adults [2]. Further, social media use extends across ethnicity groups [2, 13] due to mobile phone access. With the high rates of prediabetes and social media use, there is a high probability of a significant population of individuals with diagnosed and undiagnosed prediabetes using social media. This study will examine Facebook users with prediabetes and investigate their posts and Facebook behavior before and after diagnosis.

Social media provides an outlet for individuals to share information about their lifestyle and health behaviors. This includes diet quality, activity, and sleep quality, all of which are associated with prediabetes [14-16]. For example, someone might post excessively about the movies and television shows they are watching, suggesting inactivity. While someone might post about food they are eating, others might acknowledge unhealthy habits by sharing websites related to unhealthy eating.
Proposed Research
The proposed study is rooted in populomics. Populomics is the study of social interactions that either result in disease or protect health on a population level [17]. This multidisciplinary field incorporates the study of population level risk characterization (e.g., prediabetes behaviors or indicators) through the use of information technology [17]. This knowledge is then used to support public health interventions [17]. Using social media data, we will address the need for early diagnosis in prediabetes in order to prevent progression to diabetes.

We propose to address gaps in prediabetes infodemiology by examining indicators of prediabetes through examination of social posts. Participants diagnosed with prediabetes will be recruited from a Family Medicine Clinic in Pocatello, ID. Participants will be asked to share their social media profiles with the researchers and data for the 3 months prior through the 3 months post diagnosis will be coded. Our overall hypothesis is that individuals with prediabetes will have indicators of prediabetes on social media postings prior to diagnosis. The purpose of our study is two-fold: (1) develop a taxonomy of prediabetes indicators and (2) explore if social media data associated with those diagnosed with prediabetes can predict one's own trajectory towards prediabetes.

Methods

Design
This is an exploratory retrospective study that will examine Facebook posts among individuals with prediabetes. Comparisons of prediabetes indicators before and after prediabetes diagnosis will be analyzed.

Participants
The setting is Health West ISU (Idaho State University), ISU’s family medicine clinic/community health center which is located in Pocatello, ID. The sample will be 20 adults with prediabetes.

To be included, participants must be between the ages of 18 and 89, be able to read/write English, and be Facebook users who have been using Facebook for at least 3 months prior to their prediabetes diagnosis. Participants will be excluded if they have type 2 diabetes and/or any other major health condition (e.g., cancer, pregnancy) or life situation (e.g., incarcerated) that could be a confound and/or significantly alter the content of Facebook posts.

Procedures
The study has been approved by the Human Subjects Committee at Idaho State University and Health West, Inc. has provided a letter of support.

Participants will be recruited through ISU’s Clinical Research Center which is located at Health West ISU. Electronic Medical Record (EMR) queries will be utilized to identify potential participants who would qualify for the study. Potential participants will be adults between 18 and 89 years of age who do not have type 2 diabetes but do have a record of prediabetes (diagnosis and/or Hemoglobin A1C value of 5.7-6.4) at some point since 2015. Per protocols of past studies conducted by ISU’s Clinical Research Center, potential participants will be mailed a recruitment letter providing them with a brief description of the study and a return response card (with instructions that the card be returned by a certain date). The letter will include a statement that the clinic study coordinator will call the potential participant if the card is not returned by the predetermined date indicating that they are not interested. The letter will include an explanation of the call and can be avoided by returning the card (selecting the response that they are not interested in participating) or calling the staff to decline participation. The study coordinator will call potential participants who return the response card indicating interest in the study or who do not return the response card by the requested date. Potential participants who express interest in the study will have the protocol of the study explained to them and will take part in a screening via a telephone call to determine eligibility. This screening will
verify that participants are eligible for the study. Eligible participants will be scheduled for a clinic visit with the study coordinator.

During the clinic visit, participants will complete formal written informed consent and a set of questionnaires (see measures below). These questionnaires will be programmed with Qualtrics and participants will complete the questionnaires online on one of the Clinical Research Center’s computer. Since participants will need to be Facebook users to be eligible for this study, we do not foresee any issues with participants completing questionnaires online, however paper copies of these questionnaires will be kept ready in case any participant does not want to complete the questionnaires online or technical difficulties arise. Participants will be thanked for their participation, provided with a $25 gift card.

**Figure 1. Data Collection Timeline**

One of the items on the questionnaires asks participants to provide us with their social media profile name on Facebook. After their clinic visit, we will send each participant a friend request (from a skeletal Facebook account created for this study with a generic name and no mention of prediabetes, ISU, or other research-related topics). We will then code all Facebook posts made in the 3 months before through the 3 months after their prediabetes diagnosis. Once all Facebook posts have been coded, we will unfriend the participant and mail them a second $25 gift card. Figure 1 presents the timeline of data collection.

**Measures**

During the clinic visit, participants will complete a set of questionnaires. These include modified versions of the Facebook Intensity Scale [18], the Diabetes Online Community Engagement Scale [19] - modified to be about the prediabetes online community (as participants will have prediabetes), and the Computer Mediated Social Support Scale [20]. Minor wording modifications were made to make the questionnaires appropriate for prediabetes (instead of diabetes) and social media. Additionally we will ask participants about their use of health applications as past research has shown that these can be helpful in the context of chronic health management [21, 22]. Participants will also be asked to provide demographic information and information, including their profile name on Facebook.

For each participant, we will record their total number of Facebook friends, the number of family members on their Facebook profile, their posted relationship status (if any), and the date they joined Facebook or made their first Facebook post (if joined date is not available). Each Facebook post from the three months before through the three months after prediabetes diagnosis will also be coded. This coding includes meta-data on the post such as the time, date, and type of post (e.g., text, video), the viral nature of the post (e.g., number of likes, if post was shared from/to another page), and the nature of comments made about the post (e.g., whether social support was provided and if so type of social support). Additionally, the post’s content will be coded so we can record mentions of symptoms (e.g., hunger, fatigue, negative mood), lifestyle factors (e.g., exercise, eating, alcohol,
smoking, self-care), medical experiences (e.g., treatment, interaction with health care providers),
health tracking (e.g., physical activity monitor, glucose), the health of others known to the participant,
and additional content (e.g., current affairs, religion, games). Any photos posted by the participant will
also be coded on a number of measures including social variables (e.g., if photo was of a group, if
participant was part of the group, if participant was tagged) and health variables (e.g., if photo was
health related, if the photo was of food, the type and composition of the food).

Results

Anticipated Timelines

We anticipate that all clinic visits will be complete by summer of 2018, all coding of Facebook
posts should be complete by the fall of 2018, and analyses of the data should be complete by the end
of 2018. To date we have enrolled 15 out of the 20 anticipated participants and coded 3 participants’
data.

Planned Data Analyses

A conventional content analysis approach will be used to directly code the Facebook text data
[23]. An abstraction of the Facebook postings by several independent researchers will be completed
to develop a category scheme to enhance the study reliability [24]. The research team will compare
results and will work together to assure high intercoder reliability. This scheme will be used to facilitate
the coding of the study data (text, photographs, images, and videos). A comprehensive codebook will
be developed that will include the definitions for the categories used in the coding [25]. The threshold
for items becoming major themes will be those in which more than fifty percent of the postings are
identifying the same issue [26]. A thematic approach will be used for the data analysis; the Facebook
postings will be read and repeatedly re-read for content. Exemplar quotations will be established for
each category and theme [23]. Quotes that were common across the participants will be isolated and
presented in the findings to show a connection with the data. The findings will be confirmed by sharing
with volunteer participants. Once the results are confirmed a detailed presentation of the findings in a
publishable format will be developed to share the results of the research.

Additionally, a quantitative analysis approach will be used to help characterize the participants
and their data (e.g., demographics, means and SDs for scores on questionnaires, means and SDs for
number of Facebook friends etc.). Paired samples t-tests will also be used to examine potential
differences between the three months pre-diagnosis and the three months post-diagnosis (e.g.,
number of posts, tags, comments, type of post etc.).

Discussion

Principal Considerations

This pilot study aims to enroll 20 participants. Data from these participants and their 6 months
of Facebook behavior should provide an adequate initial taxonomy of prediabetes indicators. These
data will then support future studies with larger samples including testing the utility of the taxonomy on
predicting prediabetes status.

Knowledge Translation

It is anticipated that this study will lead to future research in the field of infodemiology specific
to furthering the examination of prediabetes as well as other health issues. Examining social media
data may support precision health efforts. Precision medicine is more than -omics [27], and
examination of “big data”, such as social media postings, may support innovative interventions in
which individualized care can be provided through identification of social health behaviors. Perhaps
one day, the social behavior that people exhibit online could be predictive of serious health issues,
especially diseases that are preventable such as type 2 diabetes. To which tailored messaging, which
has shown to improve outcomes in those with chronic conditions [29], could be used for primary and secondary prevention. Social media technologies that are commonly used by millions could be crucial in reversing the trends of prediabetes, obesity, and other health concerns in the U.S. and globally.

Conclusions

The current project aims to develop an initial taxonomy for prediabetes indicators among Facebook users and to help us better understand the social media postings of those with prediabetes. These prediabetes indicators and initial taxonomy are crucial to supporting future larger-scale studies that can advance this programmatic line of research. The ultimate goal of this research will be to develop an automated method to identify social media users who are likely to have prediabetes. This would be especially helpful in the cases of those who have prediabetes but do not know of their health condition, as identification can lead to recommendations (e.g., suggesting they be tested) and efforts that can prevent the progression of prediabetes to diabetes.

Acknowledgements

We would like to thank Idaho State University's Office of Research for their support on this project including a Developing Collaborative Partnerships Grant (LIGC05).

Conflicts of Interest

The authors have no conflicts to disclose
References

5. Hesse CM, Bryan C, Rose AE. Indicators of Suicide Found on Social Networks: Phase 1. DEFENSE MANPOWER DATA CENTER SEASIDE CA, 2015.


