

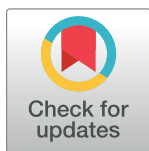
RESEARCH ARTICLE

Increasing digitalization is associated with anxiety and depression: A Google Ngram analysis

Gisbert Wilhelm Teepe^{1*}, Edda Magareta Glase², Ulf-Dietrich Reips³

1 Department of Management, Economics and Technology, Centre for Digital Health Interventions, Eidgenössische Technische Hochschule (ETH) Zurich, Zurich, Switzerland, **2** Chair of Organizational Studies, University of Konstanz, Konstanz, Germany, **3** Department of Psychology, University of Konstanz, Konstanz, Germany

* gteepe@ethz.ch



Abstract

The prevalence of **anxiety disorders and depression** are rising worldwide. Studies investigating risk factors on a societal level leading to these rises are so far limited to social-economic status, social capital, and unemployment, while most such studies rely on self-reports to investigate these factors. Therefore, **our study aims to evaluate the impact of an additional factor on a societal level, namely digitalization, by using a linguistic big data approach.** We extend related work by using the **Google Books Ngram Viewer** (Google Ngram) to **retrieve and adjust word frequencies from a large corpus of books** (8 million books or 6 percent of all books ever published) and to subsequently **investigate word changes in terms of anxiety disorders, depression, and digitalization.** Our analyses comprise and compare data from six languages, British English, German, Spanish, Russian, French, and Italian. We also retrieved **word frequencies for the control construct “religion”.** Our results show an **increase in word frequency for anxiety, depression, and digitalization over the last 50 years** ($r = .79$ to $.89$, $p < .001$), **a significant correlation between the frequency of anxiety and depression words** ($r = .98$, $p < .001$), **a significant correlation between the frequency of anxiety and digitalization words** ($r = .81$, $p < .001$), and **a significant correlation between the frequency of depression and anxiety words** ($r = .81$, $p < .001$). **For the control construct religion, we found no significant correlations for word frequency over the last 50 years and no significant correlation between the frequency of anxiety and depression words.** Our results showed a **negative correlation between the frequency of depression and religion words** ($r = -.25$, $p < .05$). We also improved the method by excluding terms with double meanings detected by 73 independent native speakers. Implications for future research and professional and clinical implications of these findings are discussed.

OPEN ACCESS

Citation: Teepe GW, Glase EM, Reips U-D (2023) Increasing digitalization is associated with anxiety and depression: A Google Ngram analysis. PLoS ONE 18(4): e0284091. <https://doi.org/10.1371/journal.pone.0284091>

Editor: Omar Enzo Santangelo, Regional Health Care and Social Agency of Lodi, ITALY

Received: October 11, 2022

Accepted: March 23, 2023

Published: April 7, 2023

Copyright: © 2023 Teepe et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its [Supporting information](#) files. Data from Google Ngram is freely available to anyone.

Funding: GWT is partially funded by CSS Insurance, Switzerland. CSS Insurance had no role in the study design, data collection, data analysis, interpretation, writing, or reviewing and approving the manuscript for publication. EMG is funded by AgenturQ, Germany. AgenturQ had no role in the study design, data collection, data analysis, and

Introduction

Anxiety and depression are common health issues affecting hundreds of millions of people worldwide [1]. The already rising prevalence of mental health issues [2] is expected to rise

interpretation writing the manuscript, or reviewing and approving the manuscript for publication.

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: GWT is affiliated with the Centre for Digital Health Interventions (www.c4dhi.org), a joint initiative of the Department of Management, Technology, and Economics at ETH Zurich and the Institute of Technology Management at the University of St. Gallen, which is funded in part by the Swiss health insurer CSS. EMG and UDR do not declare any conflicts of interest. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

further due to the ongoing COVID-19 pandemic [3]. This increase in prevalence is met by already thinly stretched resources for treating mental health [4].

Efforts to improve mental healthcare either focus on the improvement of face-to-face therapy [5], matching the correct treatment to the person [6], or the development of digital health interventions that are proposed to increase the reach or improve the effectiveness of treatments [7]. Other works try to identify underlying neurological functions or dysfunctions causing these illnesses to improve treatments [8]. These efforts have yielded a significant improvement in care over the last decades [5] and contributed to a better understanding of underlying dysfunctions. However, despite these improvements, systems struggle to sustain the steady influx of new cases, leading several to call for a stronger focus on preventing anxiety and depression [9].

To effectively improve the prevention of anxiety and depression, and to further advance their treatment, a promising approach could lie in identifying risk factors. Risk factors for anxiety and mood disorders on the individual level have been studied. Examples of such risk factors are major life events [10], chronic stress and acute stress [11], comorbidities with physical illnesses [12], loneliness [13,14], trauma [11], and childhood abuse [15]. Some risk factors, such as loneliness and alcohol misuse, have been hypothesized to be modifiable [16].

Studies investigating risk factors on a societal level have focused on social-economic status [17], social capital [18], unemployment [19], migration [20], and climate change [21]. The World Health Organization (WHO) summarizes these findings by stating that “*mental health and many common mental disorders are shaped to a great extent by the social, economic, and physical environments in which people live*” [17, p. 8]. Compton and colleagues [22] argue that these social determinants for mental health, exemplified in their work by income inequality and poor education, need to be addressed by public policies and social norms.

Different works have investigated risk factors on the individual level associated with or accelerated by ongoing digitalization. While the scientific management and information system community seems to be highlighting the positive effects of digitalization on the economy and innovation, work from the clinical and social sciences also highlights some potential negative effects. Digitalization may potentially be associated with high levels of uncertainty as it is unclear what is to be expected after the digitalization-related changes take place and whether those changes will influence the individual's employment or private life. These uncertainties could, in turn, lead to increased anxiety, a feeling of worry, or discomfort.

A digital source of uncertainty and worries is the use of social media. Several studies investigated the impact of social media and social media use on mental health. Different works have investigated risk factors on the individual level, with results indicating that not the quantity of use but the purpose for using social media is relevant [14,23], and some studies even show no clear evidence for a problematic relationship between social media and mental health [24]. The authors explained this finding about general skepticism toward social media with the fact that new forms of media have always sparked the worry of being dangerous.

Furthermore, *technostress*, or the experience of stress due to an inability to adapt to new technologies [25], manifests in various conditions such as job dissatisfaction [26], perceptions of being drained by the use of technological devices [27], feelings of exhaustion, loss of motivation, frustration, and burnout [28]. Further problems associated with digitalization, such as replacement fear [29], stressful hyper-connectivity [30], an increase in a cognitively demanding and simultaneous decrease in manual jobs [31], and digitalization anxiety [32], may further cause or aggravate mental health problems.

Other studies reported fear and other negative outcomes due to digitalization. Kim and colleagues [33] and Turel and colleagues [34] found that exposure to digitalization in the work environment can exhaust employees to the point of poor mental health, sometimes observed

as job burnout. Ferguson and colleagues [35] reported a negative spillover effect of unpleasant technology-related constant connectivity on family or other live domains, causing work-family conflicts. Furthermore, on an emotional level, Tarafdar and colleagues [36] reported that technology use evoked negative attitudes and emotions.

However, evidence investigating the association between ongoing digitalization and mental health is rare, specifically before the COVID-19 pandemic. In a Scopus search using the search term (*"Digitalization" OR "Digitalisation"*) AND *"Mental Health"*, we found 86 publications (out of 890 screened publications, date of search 1st of June 2022) that had some connection to digitalization or mental health but not one examined the connection between ongoing digitalization and anxiety or depression on a broader level. While these changes associated with digitalization may potentially affect large numbers of individuals in different situations, the methods used to investigate the influence of these changes remained the same. Researchers rely on longitudinal studies, survey self-reports, questionnaires [37], and interviews [38]. So far, few studies have leveraged the vast and various data available (i.e., big data) to investigate the connection between digitalization and anxiety and depression. This gap in research is surprising because there is, for example, other work highlighting that digitalization may lead to higher unemployment, especially in populations with a lower social-economic status that has been shown to be additionally affected more strongly by mental health problems [39].

To fill the gap, we aim to use the vast amount of data available from Google Books Ngram Viewer (henceforth: Google Ngram) to investigate the impact of digitalization on anxiety and depression by reviewing the historically disjointed works of literature on anxiety and depression and digitalization. Google Ngram is a large corpus of books (more than 8 million books or more than 6 percent of all books ever published) that has been used in related linguistic big data approaches to investigate societal trends and changes [e.g., 40,41]. While online self-report data used in other studies may be hampered by methodological issues and, for example, privacy concerns [42], Google Ngram relies on data from published scientific and fiction (fiction is only available for the English corpus) books and can thus be considered a less biased source in that sense. By targeting the effect that digitalization may have on anxiety and depression and using this novel approach, we aim to advance the social science branch in the following ways.

First, in line with its surge in public attention, we investigate the relative frequency of anxiety, depression, and digitalization terms over the last 50 years. We hypothesize that the relative frequency of each of these terms will increase over the years covered. Second, we aim to provide further evidence for the co-occurrence of depression and anxiety. Due to the often-reported comorbidity of both mental health diseases [43], we expect these word lists to be strongly associated. Third, we aim to investigate the ongoing digitalization's impact on anxiety and depression. Existing research on this influence has focused on technostress [e.g., 44], challenging hyper-connectivity [30], and digitalization anxiety [32]. These findings indicate a possible link between digitalization and increasing anxiety and depression rates. We argue that these mental disorders are multidimensional and therefore propose that they need to be investigated using various approaches. Here we present and analyze data from Google Ngram. By investigating the possible link between anxiety, depression, and digitalization, we aim to contribute to a broader understanding of how societal changes may influence these mental health disorders. We hypothesize that word lists semantically describing these disorders are strongly associated with one another. Last, we aim to advance methodological approaches to investigate the prevalence and causes of mental disorders by refining existing linguistic big data approaches [41], thereby contributing to developing complementary approaches to self-reports in clinical and social psychology research. We argue that the frequency of words in both scientific research, fictional literature, and news could be a relevant measure because

research is driven by addressing pressing real-world problems, literature is influenced by Zeitgeist, and news reports address the most demanding societal issues.

Materials and methods

We aim to investigate the general usage of word frequencies for anxiety, depression, and digitalization over the last 50 years via the Google Ngram database. Google Ngram is a digital repository for analyzing change through changes in relative word frequencies over a specific period available online [45]. It offers a big data perspective and allows for studying changes in various topics [46]. To this aim, we extracted anxiety and depression word lists from eight successive versions of the ICD-8 (International Classification of Diseases) to ICD-11 [47–49]. By using the ICD, we set as a criterion the widely applied international standard classification for diagnoses regarding mental and physiological health. The ICD classification allowed us to interpret and compare different health symptoms over a specific period. The digitalization word lists were derived from Brewster and Murray-Smith [50] and Wolf and Bartelheimer [51], covering the last 50 years. Following the procedure by Younes and Reips [41], we generated a set of 26 words most representative of the concepts of anxiety and depression and a set of 33 most representative words describing digitalization resulting in two separate lists (S1 and S2 Tables in [S1 File](#)).

Following the guidelines by Younes and Reips [41] (Procedure IV), we searched for synonyms using the standard online dictionary Roget's Thesaurus (standard reference for English). We included all synonyms labeled “*most relevant*”, excluded all synonyms with a broader semantic meaning (e.g., “*interest*”, “*platform*”), and only focused on 1-grams because “*particular two-word combinations would be too infrequent in the corpus to show change over time*” [52, p. 1724]. This process yielded 85 words translated into five languages (German, Spanish, Russian, French, and Italian) and cross-validated by native speakers who are also proficient in English. We chose these six languages (British English, German, Spanish, Russian, French, and Italian) as they are the only languages spoken on the European continent available on Google Ngram. In the subsequent step, 73 native speakers translated the word lists by themselves without pre-existing translation using independent back-translation [53,54] to optimize the translatability. We merged all translations, excluded terms with double meanings detected by native speakers, and generated two final sets of 59 words that exist in all six languages. These lists are displayed in the (S1 and S2 Tables in [S1 File](#)).

We analyzed word inflections to quantify the final word lists using the tag “_INF” for each word. This method finds and compares inflection frequencies for words (Younes and Reips [41]). This consistency check confirmed that our originally selected word lists displayed the highest frequency. Additionally, we double-checked the word lists for correctness. Following Lin and colleagues [55], we compared the most frequent word in each corpus with the most frequent noun in each language to account for the influx of data (British English = “*time*”, “*people*”; German = “*Zeit*”, “*Jahre*”; Spanish = “*parte*”, “*años*”; Russian = “*эго*”, “*он*”; French = “*temps*”, “*partie*”; Italian = “*parte*”, “*tempo*”, p. 171). Lastly, we calculated summed z-scores for each language and each word list.

Additionally, we compared the generated word lists with a control list, namely the religion word list [41,56]. In doing so, we followed the approach by Younes and Reips [41] and were able to control for a general misleading trend counterfactually.

We used the final set of words as the Google Ngram search terms. We collected data on the word frequency per year in percent. Google Ngram divides the number of a word's yearly appearances by the total number of words in the corpus in that year and generates charts of those word frequencies [41], resulting in a longitudinal analysis over a period. We extracted

the Google Ngram results using the Git Hub repository econpy/google-ngrams available under (<https://github.com/econpy/google-ngrams>, MIT Licence). We retrieved all available data up to 2019, the latest year the Google Ngram data is available. Thus, our investigation only considers the time before the ongoing COVID-19 pandemic and for languages spoken on the European continent. Our study investigates changes over the 50 years before 2020 in Europe.

Results

We calculated correlation coefficients between the word frequencies in percentage and years. By doing so, we assumed that the more frequent a word was, the larger its proportional influence was per year. Changes in frequencies from 1970 until 2019 over all languages and for each language can be found in Fig 1.

Separate correlations were calculated for each language and the overall combined frequency scores of terms from all languages. Table 1 illustrates the result of the analysis over all languages, and the results for each language can be found in S4–S10 Tables in S1 File.

Anxiety

We found a strong positive significant correlation between anxiety and years over all languages ($r = .87, p < .001$) and for each individual language: British English ($r = .81, p < .001$), German ($r = .88, p < .001$), Spanish ($r = .82, p < .001$), Russian ($r = .93, p < .001$), French ($r = .93, p < .001$), and Italian ($r = .62, p < .001$).

Depression

We found a strong positive significant correlation between depression and years over all languages ($r = .94, p < .001$) and positive correlations of varying strength for the individual

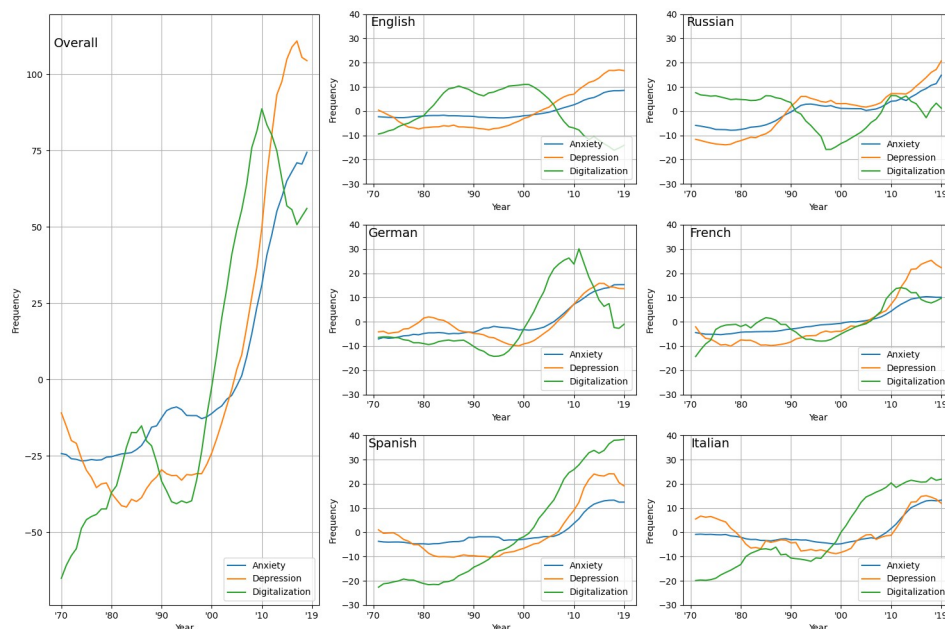


Fig 1. List-averaged word frequencies from 1970 until 2019 for each category and language. Frequencies (in percent) are z-transformed, most common word controlled, and summed over all words.

<https://doi.org/10.1371/journal.pone.0284091.g001>