An infodemiological study using search engine query data to explore the temporal variations of depression in Finland

Jonas Tana (R.N., M.A.)

Information Studies, Åbo Akademi University, School of Business and Economics; Arcada – University of applied sciences, Department of Health and Welfare

Jonas Tana, Arcada, Institutionen för hälsa och välfärd, Jan-Magnus Jansson's plats 1, 00560, Helsingfors, FINLAND. Email: jonas.tana@arcada.fi

Abstract

A majority of healthcare is undertaken by individuals without the involvement or knowledge of healthcare professionals. People often try to treat health problems themselves, often first consulting the Internet, with search engines as natural starting points. Health information seeking conducted in search engines generate big data, data that can provide valuable insights into patterns of symptoms and disease, especially for stigmatizing or sensitive health topics, like mental health problems. The aim of this article is to utilize search engine query data for trends and temporal variations of depression in Finland. The key findings of this study show that depression related Internet search query volumes are slightly increasing during the time periods studied. The results also show that search query volumes follow seasonal patterns with peaks during autumn and spring and troughs during the summer months and mid-winter. Of all weekdays, Sundays have the highest search volume for depression related queries. These results present new meaningful insights into the epidemiology of depression, as it can give insights into the part of the population that does not present for treatment or professional help. It can also help health professionals and other officials to understand broader patterns of mental illness when planning services and campaigns.

Keywords: depression, search engine, information seeking behavior, consumer health information, technology

Introduction

The developments in information and communication technology (ICT) in the last decade have led to an emerging behaviour in regards to health behaviour [1]. Because of the easy accessibility, the Internet has become a primary and critical tool when it comes to health related behaviour. Health information seeking is common everyday behaviour, and millions of people surf the Internet every day seeking information ranging from health promotional activities to symptoms, diagnoses and treatments [2-3]. It is estimated that between 70-90% of healthcare is undertaken by individuals without the involvement of healthcare professionals, suggesting that before people seek treatment from healthcare professionals they will try to treat the problems themselves, often first consulting the Internet [4-6]. The Internet also provides a possibility for anonymity, and is a less invasive or stigmatizing way of finding information about sensitive health topics, like mental health problems [8-10]. The most popular search engine in the world, Google, which in Finland had a mar
ket share of 96% in 2014 [10] serves as an important gateway to health information for both clients and patients. Google has reported that four thousand searches per second are health related [1,11]. According to statistics, 66% of the Finnish population aged 16-89 has searched for information related to disease, nutrition or health during the past three months. The percentage is higher, 88%, for the younger demographic segment of ages 25-34 [12]. The Finnish online medical information service Health Library (Terveyskirjasto.fi, administered by The Finnish Medical Society, Duodecim), a popular and authoritative website with information about symptoms, diagnoses and treatments, had over one million monthly and nearly forty thousand daily visitors in 2011. The use of the service has increased steadily since its launch in 2007 [13].

Health information seeking conducted in popular web search engines generate large datasets, or big data, that can provide valuable and unique insights into patterns of symptoms and disease as well as information about health-related behaviour on the internet [2]. The application of Internet data in health care has led to the rise of a new interdisciplinary research discipline as well as methodology called infodemiology, which is defined as the study of the determinants and distribution of health information in an electronic medium with the ultimate goal to inform public health and public policy [6,2].

The aim of this observational infodemiological study is to analyse and evaluate search engine query data for temporal variations and trends of five different depression related queries in Finland. The objective is to provide Internet-based evidence for the seasonality of depression in Finland. Investigations into the monthly and daily variations of search engine query volume for depression in Finland has, to the author’s knowledge, not been performed. In this study, the following questions are addressed: Is the query volume for depression related queries increasing or decreasing? Are there monthly variations in search volume for depression in Finland? Are search volumes higher during any certain day of the week?

**Depression**

Depression is regarded by the WHO to be one of the most burdensome diseases in the world and deemed a public health priority as one in seven people suffer from a severe mood disorder during their lives. Each year about 7% of the world population suffer from a major depression, 25% if anxiety and lighter forms of depression are included [14-15]. Depression is also associated with a higher and significant risk of suicide, as on average 52% of all suicide victims have suffered from depression [16-17]. In Finland, depression is one of the biggest public health problems and it is estimated that at least 5% of the adult population suffer from depression every year [16, 18]. However, the incidence of depression in Finland has in some areas decreased during the time period 2004 – 2015 in Finland [19].

Even though access to care has improved, the majority of people with depression or mood disorders remain outside of care [15,18,5]. Because of the complexity, stigma, and barriers to care that mental illness present, individuals have been shown more likely to seek information about their problems online, with search engines as natural starting points [20-21]. Therefore, knowledge about health information seeking behaviour related to depression has the potential to present new meaningful insights into the epidemiology of depression. It can also help health professionals and other officials to understand broader patterns of mental illness [20].

**Temporal variations of depression**

The influence of seasons on mood disorders is controversial [21]. Many studies of seasonal variation of depression and depressive symptoms report considerable changes in incidence and severity of psychiatric disorders during the year. These temporal variations, or seasonality, follow seasonal patterns and vary between different months during the year [23-25]. Two different patterns of seasonality have been identified, a winter peak and an autumn-spring peak. Some studies, based on both self-report questionnaires and diagnostic interviews, have shown recurring seasonal peaks for depression usually during the mid-winter months [23-24].
Other researchers have found that depressive symptoms and suicides peak in the spring and autumn, especially in the more northern parts of the hemisphere [17,22,26-30]. Researchers have therefore questioned the common belief in mid-winter as the most common season for mental disorders [28]. Research on distribution of depression for different weekdays has not been identified, but the distribution of suicide during the week has in previous research shown peaks on Monday, Tuesday and Sunday. Brådvik [17] found that there was a preponderance of suicide on Sundays in Sweden. Drawing conclusions about the temporal variations of depression can, because of the conflicting and inconsistent results, be challenging. The difficulties in drawing conclusions of the temporal variations of depression have been suggested to be due to the limitations in methodology in tracking and monitoring depression and seasonality in populations [24,20]. One potential method for observing population interest in mental health problems is monitoring the Internet.

Utilizing search engine data to study seasonal variations of depression is not a novel method. Ayers et al [20] studied internet search queries for several different mental health problems, of which one was depression, specifically in regards to seasonal patterns and correlations in mental health information seeking in the U.S and Australia. Yang et al [9] investigated large-scale global seasonal patterns of depression using Internet search query data. The findings indicated that Internet searches for depression from people in higher latitudes, universally across countries and regardless of language, were more vulnerable to seasonal change with peaks in the winter and troughs in the summer [9,20]. Both Ayers et al. [20] and Yang et al. [9] concentrated on patterns for the northern and southern hemispheres and their inverted correlation. The studies did not concentrate on specific countries and search query analysis was based on a single query term for depression. Neither did the studies concentrate on the monthly or daily variations in query volume. With the increasing burden of disease due to mental disorders worldwide, broad knowledge of the epidemiology of these disorders are of increasing interest [14]. Finland, with its northern situation, is an ideal country to study temporal variations for depression related questions, as there are major variations in the number of hours of daylight during four seasons and the weather parameters are rather homogenous throughout the whole country [29].

Method and material

Infodemiological data usually consist of real-time health related metrics harvested from the Internet. The amount of data generated by users on the internet has made measurable what was previously immeasurable [31-32]. Infodemiology as a method offers a novel set of tools for researchers to systematically mine, aggregate, and analyse data to get insights into public health. The methodological advantages are that metrics are available in real time and can be collected automatically and inexpensively [32]. The method has been used to predict influenza outbreaks and disease monitoring, in detecting misinformation on the internet as well as in surveillance purpose of public health concerns [6,32-33]. Because of the complexity and stigmatizing nature of mental health problems, data from search engines can give alternative insights into the problem [20]. For this study, data was retrieved and downloaded from Google Trends (http://www.google.com/trends), an open online search tool that allows users to see how often specific keyword, subjects and phrases have been queried in the Google search engine over a specific time. Data was retrieved in September 2016. The data is non-personally identifiable and does not reflect or reference an individually identifiable user and thus the ethical concerns regarding this study are minimal.

Google Trends works by analysing a portion of Google searches to compute how many searches have been conducted for the search-term entered, relative to the total number of searches performed in Google over the same time period across chosen regions of the world. Query volumes are then automatically normalized and numbers are provided as relative search volumes (RSV). This relative search volume is the query share of a particular term for a given location and time period, normalized by the highest query share of that term over the time-series [33]. RSV = 100 is thus the period with
the highest search proportion for queries, and $\text{RSV} = 50$ is 50\% of the highest search proportion.

Data in this study is based on the five following queries relating to depression: “masennus” (finnish for depression), “masennus oireet” (“depression symptoms” in eng.), “masennustesti” (“depression test” in eng.), “depression” and “depression test” (the Swedish as well as English term for depression and depression test). The terms, which according to the Google Trends database were the top queries for depression related search terms, were chosen to represent an extensive sample of queries related to depression in general as well as symptoms and treatment, in both official languages, as well as English, in Finland. The search queries where specified to Finland as a geographic area to prevent the mixing of search queries that originated from other regions. Search engine query data was downloaded for three different time periods: January 2004 to December 2015, January 2010 to December 2015, and the complete year 2015. These different time periods were chosen to get a comprehensive temporal analysis of depression related information seeking that corresponds with the rising query volumes in Google. Reported improvements in the Google Trends algorithm also motivated to select three different time periods [34]. Google Trends provides downloadable data in comma-separated value (.csv) for all given terms and time periods. The data contain a relative search volume for each month for the time periods 2004-2015 and 2010-2015. To get an average RSV for the different weekdays, data for each month of 2015 was downloaded. This monthly data contains an average RSV for each weekday. Data was analysed in MS Excel 2016. For all five queries during the different time periods an average relative search volume per month was calculated, and an overall average RSV per month for all query terms per month was then determined. An average RSV for each weekday in 2015 was also calculated.

**Results**

The comparisons of the relative search volume for the five different search queries in Google Trends Search Volume Index Graph show that there are differences in relative search volume between the five query terms during the time period 2004-2015. As figure 1 shows, the term masennus (eng. depression) has a higher RSV compared to the other four search terms. This is consistent for all three different time periods, and shows that masennus is the most sought query term of the five terms included in this study.

The combined average relative search volume for all five search queries for the time period 2004-2015 is presented in Figure 2. It shows a slight increase in search volume since a trough in 2007. As Figure 3 shows, the same increase in average relative search volume for all the terms is also noticeable for the time period 2010-2015.

![Figure 1. Screen shot from Google Trends Search Volume Index graph comparing the queries masennus (blue), masennus oireet (red), masennus testi (yellow), depression (green) and depression test (lilac) during the time period January 2004 - December 2015.](image-url)
Figure 2. Combined average RSV for all five search queries 2004-2015 shows a slight increase in Internet search query volume since 2007.

Figure 3. Combined average RSV for all five search queries 2010-2015 shows a slight increase in Internet search query volume shows a slight increase in RSV since 2010.
The average monthly search volumes during all time-periods are presented in Figure 4. As can be seen it show that search queries in Finland follow seasonal variations. The annual curve follows a bimodal pattern with two peaks, one in spring and one in autumn. There is a clear trough for the summer months. The highest combined average relative search volume for the five terms related to depression during the three different time periods is presented in March (RSV=61) followed by February (RSV=59) and September (RSV=59). The lowest average RSV is found in in July (RSV=48) and June (RSV=48), followed by May (RSV=49).

The average relative search volume for all query terms during weekdays in 2015 (Table 1) shows that the highest query volume is presented during Sundays (RSV=61), followed by Monday (RSV=56). The lowest average RSV can be found during Friday (RSV=52).

Figure 4. Combined average monthly RSV for all five search queries during all time periods (2004-2015, 2010-2015, and 2015) combined shows peaks during the darker seasons and lows during the lighter seasons.

Table 1. Average relative search volume (RSV) per weekday for all query terms during the year 2015.

<table>
<thead>
<tr>
<th>Weekday</th>
<th>RSV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sunday</td>
<td>61</td>
</tr>
<tr>
<td>Monday</td>
<td>56</td>
</tr>
<tr>
<td>Wednesday</td>
<td>56</td>
</tr>
<tr>
<td>Saturday</td>
<td>56</td>
</tr>
<tr>
<td>Tuesday</td>
<td>55</td>
</tr>
<tr>
<td>Thursday</td>
<td>54</td>
</tr>
<tr>
<td>Friday</td>
<td>52</td>
</tr>
</tbody>
</table>
Discussion

To the best of the author’s knowledge, this study is the first of its kind in analysing temporal variations of online search behaviour for depression related queries in Finland. It is also the first study investigating the differences in Internet search query volume during different weekdays for depression related information in Finland. The key findings of this study show that depression related Internet search query volumes are slightly increasing. This is interesting and contradictory to the fact that the incidence of diagnosed depression has decreased during the time-period 2004 – 2015 in Finland [19]. This may have many possible explanations, one being the fact that people try to treat health problems themselves by first seeking information about their state before visiting a physician. The stigmatizing factors, as well as the combination of treatment avoidance due to shame and denial also have to be taken into account. It could also mean that alternative and complementary treatments are getting a stronger foothold of depression treatment, an explanation suggested by Oliphant [5]. The increasing use of information and communications technology (i.e. smartphones), both generally and for seeking health related information on the Internet during the studied time period also need to be taken into account. Since 2005 the use of Internet in the last three months has grown from seventy three per cent to eighty eight per cent [12,35]. The results also show that search query volumes follow seasonal patterns with peaks during autumn and spring and troughs during the summer months and mid-winter. This supports previous studies where search queries have been shown to follow patterns with seasonal variations [20,9]. However, in previous research, the highest relative search volumes have been identified during the darkest months of the year in the northern hemisphere [20,9] whereas the results of this study show higher query volumes for autumn and spring. A peak in March has not been identified in previous infodemiological research on depression. These findings conform with previous conclusions of seasonality in diagnosed depression, hospital admissions and suicide rates in the Scandinavian countries and the northern hemisphere, with peaks during spring and autumn [22,25-26,29]. The highest search volume for the different weekdays can be found during Sunday. This finding is in line with previous research findings where Sunday has been found to be the weekday when most suicides are committed [17]. A possible explanation for this have been suggested by Brådvik [17], who suggests that the reason could be the low activity or being at home on Sundays that may trigger inner tension and lead to mood disorders. Similar search query analysis of terms relating to depression for the different weekdays have not, to the knowledge of the author, been conducted in Finland or anywhere else in the world, and therefore this result cannot be compared to other previous research results.

This kind of Internet search query analysis can assist health care providers in designing temporally optimal informational interventions. Having insights into the first step in health related behaviours can be useful in improving public health. This is especially relevant in sensitive and stigmatized problems like mental health problems [36]. Investigation of Internet search trends may complement traditional approaches and aid future research in many health related disciplines and with the appropriate utilization Internet search records can establish an effective, reliable prediction system for many medical illnesses or psychiatric emergencies. The increased interest in depression related information on the Internet should make professionals particularly observant during spring. A major challenge in mental health is how not only assess but also treat mental illness among individuals who do not present for treatment or cannot be reached by more traditional surveys. The Internet is a stigma and cost-reducing tool to help screen and treat those who search for treatment but may not bring problems to the attention of their clinicians [20].

Limitations

This study and the method utilized has some limitations that need to be addressed. Individual search queries for depression cannot accurately reflect the actual motivational factor, influence or intention, behind seeking information about depression and not all depression related seeking is clinical. Alternative nonclinical expla-
nations such as news events or media that trigger an interest for mental health information needs to be taken into account. However, as previous studies employing search query analysis has stated, it is reasonable to presume that the reason people seek health information about depression on the internet is because they, or people they know, may be experiencing symptoms of depression [9].

Additionally, search query analysis cannot capture demographic profiles or user characteristics, which has limiting effects on the ability to draw conclusions about population behaviour in general. However, statistics in Finland point out that nearly all age-by-demographic population categories seek online health information, which suggest that search trends may indicate trends in the health of populations. Moreover, the aim of analysing search engine query volumes is not to replace, but to complement traditional surveillance methods of mental health. There are also limitations with the Google Trends tool itself. As previous research has reported [20], there is a lack of detailed information on the method by which Google generates this search data and the specific algorithms it employs to analyse it.

Conclusions

This study shows that monitoring and analysing search query volume can provide different insights into national trends regarding mental health than more traditional surveillance methods offer. Internet query surveillance using search engine data can guide the development of more traditional surveillance systems. The benefits of this kind of monitoring are the cost effectiveness and anonymous nature that analysing search queries presents. The popularity of search engines, the increasing availability of health-related information on the Internet and individuals’ responsibility for health-related matters have the potential to create valuable insights to a broad range of public health concerns that can be assumed to correspond with the health of a population [11]. Preventive measures should be developed to reach people who are seeking depression related information, as identification and treatment could influence depression and suicide rates.

While this study concentrates on health information seeking by analysing search queries related to depression, it could be fruitful to analyse what is being published, about depression on the Internet with different query terms related to depression. This could allow for interesting insights into the dynamics and interactions between health information provision and health information seeking behaviour.

Acknowledgements

Funding for this research has been granted by Fonden för Teknisk Undervisning och Forskning and A.F. Lindstedts och Svenska Handelsinstitutets fond för handelsutbildning.

References


