

## INTERNET SEARCHES FOR “SUICIDE”, ITS ASSOCIATION WITH EPIDEMIOLOGICAL DATA AND INSIGHTS FOR PREVENTION PROGRAMS

Przemysław Maciej Waszak<sup>1,2,3</sup>, Patryk Górski<sup>2</sup>, Janusz Springer<sup>4</sup>, Wioleta Kasprzycka-Waszak<sup>5</sup>, Marta Duży<sup>6</sup> & Paweł Zagożdżon<sup>1</sup>

<sup>1</sup>Department of Hygiene and Epidemiology, Medical University of Gdansk, Poland

<sup>2</sup>T. Bilikiewicz memorial Voivodship Psychiatric Hospital, Gdansk, Poland

<sup>3</sup>Department of Developmental Psychiatry, Psychotic and Geriatric Disorders, Medical University of Gdansk, Gdansk, Poland

<sup>4</sup>Department of Preventive Medicine & Education, Medical University of Gdansk, Poland

<sup>5</sup>Clinical Psychology postgraduate studies, Medical University of Gdansk, Poland

<sup>6</sup>Biomedical engineering, Gdansk University of Technology, Poland

received: 16.3.2018;

revised: 12.10.2018;

accepted: 25.10.2018

### SUMMARY

**Background:** Literature demonstrates that analysis of internet search data is a useful tool in predicting the occurrence of illnesses and health-related behaviors. The aim of the study was to quantitatively present the trends in Google searches for the keyword “suicide” and to analyze its correlation with the number of suicides in Poland.

**Subjects and methods:** We used the Google Trends tool to compile data for years 2004-2016. Statistical analysis was performed for annual, monthly, daily and hourly data. Official data on suicide in Poland were obtained from the Central Statistical Office and the General Police Headquarters of Poland.

**Results:** A gradual decrease in Google Relative Search Volume of the keyword “suicide” was observed in years 2004-2014, despite the significant increase of suicide rate in Poland ( $R=-0.24$ ). Reverse correlation was also found between regional suicide coefficients and search volume ( $R=-0.22$ ). The highest search volumes were recorded in winter months, first days of the week and at night hours ( $p<0.001$ ).

**Conclusions:** Presented results may contribute to more effective suicide prevention programs. By specifying the time intervals in which searching suicide information is the highest, it will become easier get to individuals at risk.

**Key words:** suicide – Internet – epidemiology - Google Trends

\* \* \* \* \*

### INTRODUCTION

In many populations suicide is an increasingly more common cause of death and is now estimated as the fifteenth cause of death worldwide (World Health Organization 2016). Poland is not an exception. According to general suicide rate in 2014 was 1,5 per 10,000 population (Główny Urząd Statystyczny 2016). The issue of suicide started to gain attention in the OCED countries' public health programs, with emphasis on identification of groups at risk and prevention campaigns (Isometsä 2014). Such prevention includes multifaceted aspects and regulations concerning e.g. access to means (methods) of suicide, effective treatment of mental illnesses and proper reporting of suicide in the media etc. (O'Connor et al. 2011). With its global reach and fast access to large amounts of data, the internet provides new supportive tools that can be useful as a form of suicide prevention (Jacob et al. 2014).

Epidemiological data on suicide indicators are limited. Poland's data on fatalities is processed and reported by the Central Statistical Office (GUS, Główny Urząd Statystyczny) and the General Police Headquarters of Poland (GPHP). Literature demonstrated that, the Po-

land's data on suicide attempts was underestimated (Polewka et al. 2004). Similar to other countries, data on nonfatal suicidal behaviors are collected in single, time- and geographic-limited studies (Bernal et al. 2007, Scocco et al. 2008), and for some world populations are virtually unavailable (Vijayakumar 2004).

Another challenge is personal suicide risk profiling. Individuals with increased risk of suicide do not necessarily seek counselling or support from designated agencies and personnel, and on contrary are more likely to search for help online (Harris et al. 2009). Thus, better understanding of online behaviors of suicide information seekers, could contribute to greater effectiveness of prevention strategies (Arora et al. 2016, Parker et al. 2016, Yang et al. 2011).

With emergence of Google Trends, analysis of large sets of online search data has become simpler than ever. The effectiveness of this tool was demonstrated in predicting the occurrence of infectious diseases (Carneiro & Mylonakis 2009), mental illness and health behaviors (Nuti et al. 2014). A recent review provided a comprehensive summary of the available evidence on the correlation between suicides and Google Trends (Tran et al. 2017). Majority of the published studies demonstrated a medium

to large correlation effect, however their results have been commented as having low-validity (Tran et al. 2017).

The purpose of this study is to present the amount of suicide information seeking in the Polish language: its trends, most frequently searched phrases and correlations with official data on suicide.

## SUBJECTS AND METHODS

We used the Google Trends tool (Google Inc., California, USA, [www.google.com/trends](http://www.google.com/trends)) to analyze the data searches using the Google search engine. This tool allows obtaining quantitative data of particular keyword searches and grouping it into specified geolocations and time intervals. Qualitative data on suicide and suicide rates were obtained from public statistics provided by the GUS (<http://stat.gov.pl/>) and the GPHP (<http://www.statystyka.policja.pl/st/wybrane-statystyki/zamachy-samobojcze>). Data were grouped country-wide and accordingly each Voivodship (province).

The authors have followed the guidelines on using Google Trends tool in health sciences, published in PLOS One (Nutti et al. 2014). The analysis used data from Google Trends from the very beginning of Google's reporting period from January 2004, to December 2016. The study covered the history of all searches for "suicide (polish language – samobójstwo)" in Google search (all searches, regardless of category – default setting). The searches for other words describing suicidal actions/tendencies such as "suicide attempt (próba samobójcza)", "self-harm (samookaleczenie)", "suicide thoughts (myśli samobójcze)" etc. has been excluded from the analysis due to the small size and incompleteness of these inquiries. Only Polish language searches from Polish territory were included in the study. Statistical analyses were performed for annually, monthly, daily and hourly searches trends. The analysis also included the data of most common phrases con-

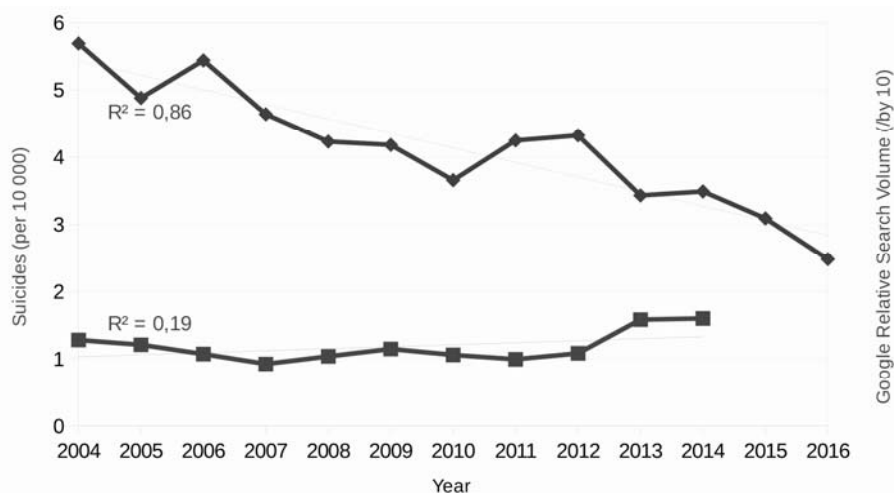
taining keyword "suicide." In addition, data on regional interest were also used (by Voivodship). Data were collected and analyzed during 27.02.2017 - 15.03.2017.

Presented results were given in the Google Relative Search Volume (GRSV), a standardized unit introduced by Google and defined "numbers showing how often a word was searched for the highest point of graph in a given time and region. A value of 100 indicates the highest popularity of the word. By contrast, 0 indicates that the popularity was less than 1% of the highest value." (<https://trends.google.pl>)

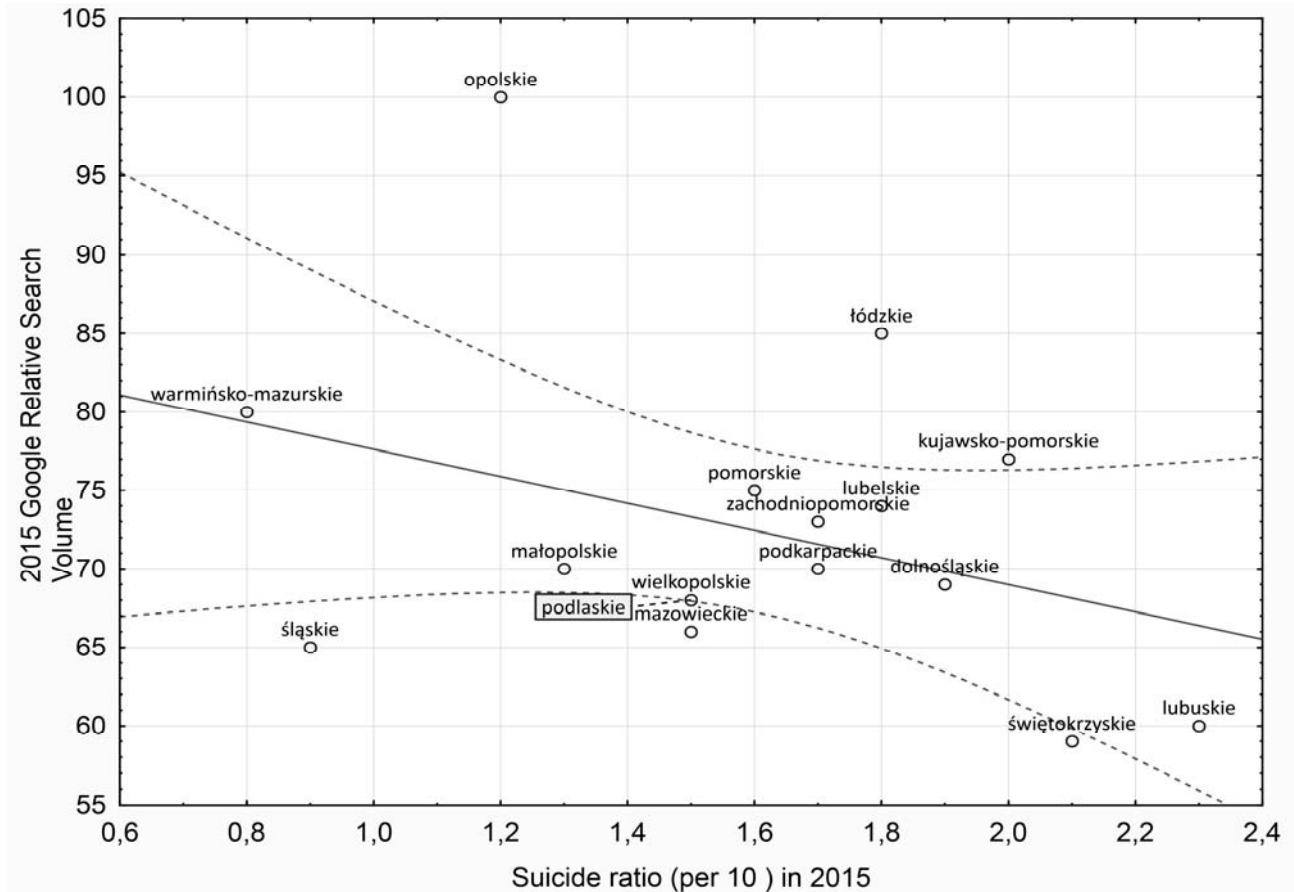
Data was collected into an Excel spreadsheet (Microsoft, Redmond, Washington, USA). Statistical analyses were performed using Statistica 10 software (StatSoft, Tulsa, Oklahoma, USA). For each variable a statistical distribution test was performed. If the variable fit the normal (Gaussian) distribution criteria, parametric tests were used for calculations. If the variable did not meet the above criterion, nonparametric tests were used. For comparisons between groups the parametric ANOVA or ANOVA according to Kruskal-Wallis, were used respectively. Calculation of correlation of variables was calculated using a ranged Spearman test (all correlations data was non-parametric). The level of statistical significance was set at  $p < 0.05$  threshold.

## RESULTS

In the years 2004-2014 a significantly higher number of suicide deaths was reported in Poland (2004:1.28; 2014:1.60 deaths per 10 000) with the greatest increase during 2012-2014. Also two minor, declines in suicide rates were recorded between 2004-2007 and 2009-2011. At the same time (2004-2014) a gradual decrease of Google searches for "suicide" term was noted (Figure 1). Correlation between those values was  $R = -0.24$  (Spearman), but the  $p$ -value was not statistically significant.



**Figure 1.** Suicide death ratios per 10 000 population (data from General Police Headquarters of Poland) and Google Relative Volume Searches (divided by 10 to increase visibility) over period 2004-2016. Correlation efficient  $R = -0.24$ ,  $p > 0.05$  (Spearman). The R values for each trend lines are also given



**Figure 2.** Regional suicide rates (2015 data from Central Statistical Office) and Google Relative Search Volume (2015 data) for word "suicide (samobójstwo)" by Voivodships (Spearman  $R=-0.22$ ,  $p>0.05$ ).

Another negative correlation was found when comparing Voivodships' GRSV "suicide" data with the corresponding 2015 suicide ratios (per 10 000). In the Warmińsko-Mazurskie Voivodship, where the suicide rate in 2015 was the lowest in Poland (0.8/10 000), the GSRV coefficient was 80. Whereas in the province with the greatest ratio of suicides deaths in Poland (Lublin Voivodship, 2.3/10 000) the associated GSRV was 60. One exception was the Opolskie Voivodship, where despite the highest GSRV in Poland (100), the suicide rate was relatively high (1.2). In the data grouped by Voivodship, the Spearman correlation coefficient was  $R=-0.22$ , with  $p>0.05$  (Figure 2).

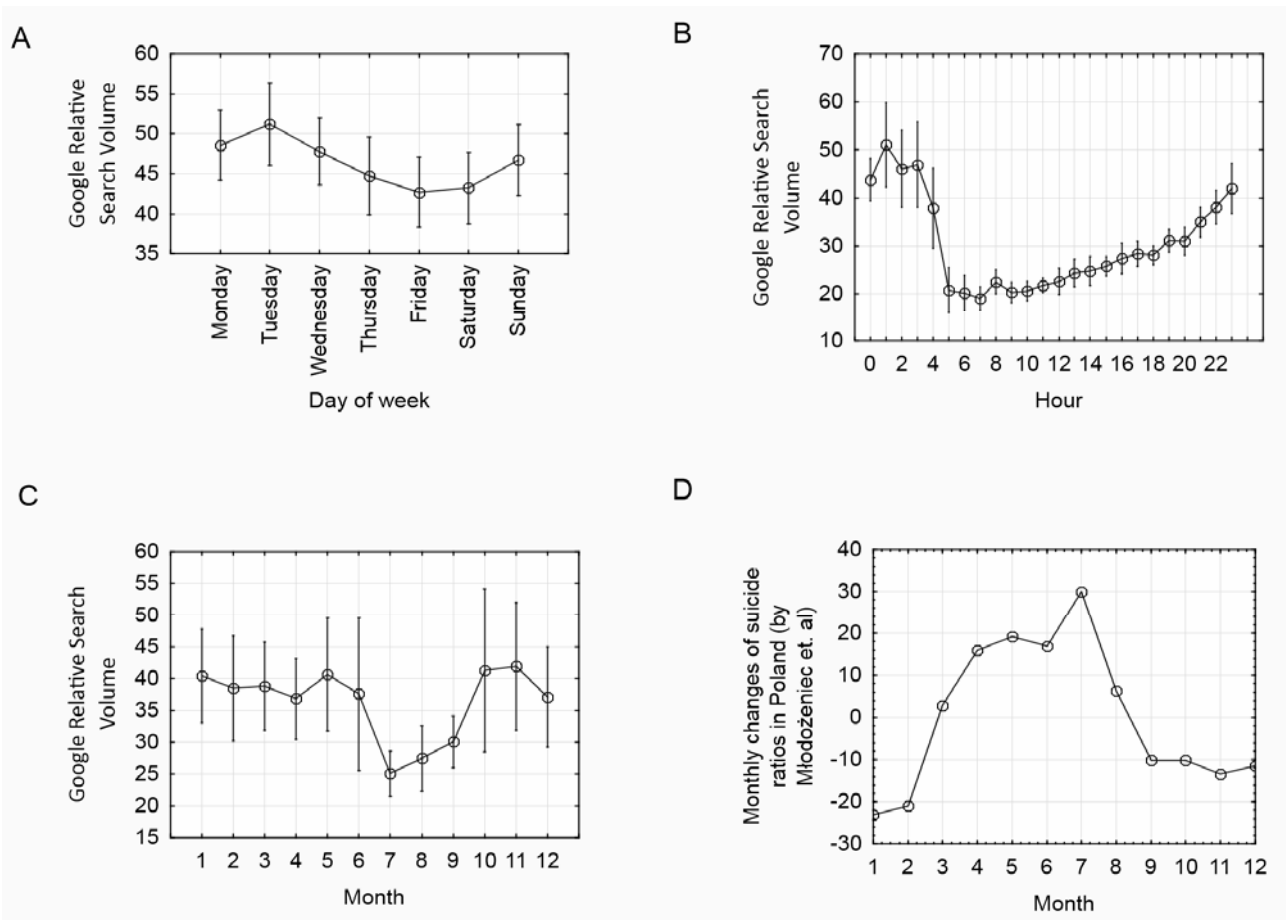
For temporal analysis, most inquiries were recorded in January, May, October and November and the least during the summer months (Figure 3C). Daily trends showed the highest GSRV for "suicide" on Tuesdays and a gradual decrease on the following days of the week (Figure 3A). Hourly analysis (Figure 3B) showed the highest searches are during nighttime, with peak from 1:00 to 3:00 AM ( $p<0.001$ ).

During entire analyzed period the three most searched phrases were "(he) committed suicide (popełnił samobójstwo)", "how to commit suicide (jak popełnić samobójstwo)," and "death (śmierć)". The less frequently searched phrases were "suicide + [city name]" (Figure 4).

## DISCUSSION

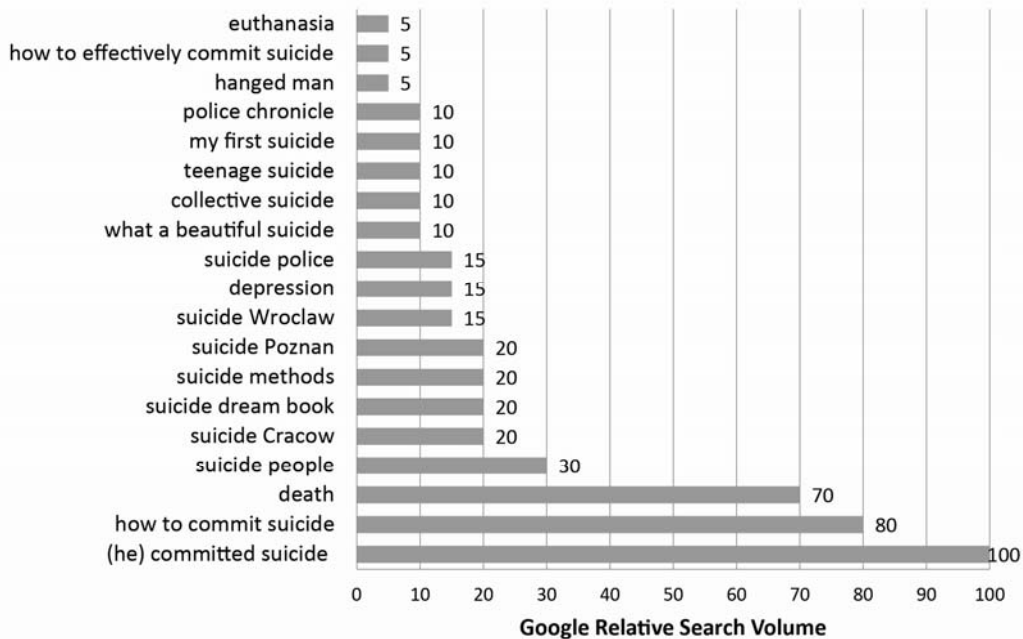
Previously conducted demonstrated the usefulness of predicting suicide ratios by measuring internet searches volume. In the United Kingdom the predictions based on the searches for 'depression' and 'suicide' keywords in years 2004-2013 have been applied (Arora et al. 2016). The research team constructed a prognostic model using custom-designed mathematical algorithms. On this basis, authors found that the increase in volume of search for term 'depression' was associated with fewer suicides. In contrast, more frequent search for 'suicide' keyword was associated with a greater increase of suicide rates (Arora et al. 2016). Similar conclusions were reached by French researchers, who even suggested ways of using Google Trends to model real-time suicide risk (Fond et al. 2015).

Opposite to the cited literature, we found a reverse association – the more frequently "suicide" was searched online, the lower real-world suicide statistics generally were. This trend is particularly significant, when general annual data and data for particular Voivodships was compared. Although these statistics did not reach the significance threshold, their importance was also mirrored by the divergence between monthly Google searches and the monthly changes in suicide death ratios, reported in previous epidemiological studies (Młodożeniec et al. 2010).



**Figure 3.** Average Google relative search time interval graphs for: A) Days of the week (differences  $p > 0.05$ ); B) Hours ( $p < 0.001$  for the difference between 0-3 and 5-15 hours); C) Months (differences  $p > 0.05$ ); D) Monthly indicators of suicides in Poland – graph based on data from Młodożeniec et al. (15)

## Top searched phrases 2004-2016



**Figure 4.** Googles’ the most frequently searched phrases containing word “suicide (samobójstwo)” and related terms in 2004-2016 in Poland

The seasonality of suicide and their epidemiology in Poland were thoroughly investigated by Młodożeniec et al. (Młodożeniec et al. 2010). Based on GUS statistics, the authors have found that the summer months had markedly higher suicide rates than the rest of the year (at least in the period of 1999-2003, analyzed in the study). This observation corresponds with reversed Google search trends, which were the lowest from July to September (Figure 3C, 3D).

Data recently published by the GPHP suggested the lowest prevalence of suicides on Wednesdays and Sundays, and the highest on Mondays and Fridays (<http://statystyka.policja.pl/st/wybrane-statystyki/zamachy-samobojcze>), which only partially corresponded with volume of Google searches. No data is available regarding the suicide deaths in Poland in shorter time intervals (e.g. hours) is available. It is therefore unclear whether it can be assumed that suicides also occur when the search rates were lowest during morning hours. Perhaps in the course of further research it will be possible to investigate the above addressed questions, together with creating a model for predicting future suicide rates in Poland.

Google Trends allows tracking statistics by administrative regions (e.g. Voivodships, states, counties etc.). American authors have studied recent deaths records related to alcohol, drugs and suicide in particular US states and created prediction model using Google Trends (Parker et al. 2016). Their result was subsequently compared to an alternative model that took into account traditionally analyzed data: unemployment and income differences, but this traditional model was found to be less accurate (Parker et al. 2016). Other authors also found a correlation between the number of searches for suicide-related phrases and the historical suicide data for all 50 US states (Gunn & Lester 2013). The above assumptions may also be useful for estimating regional differences in suicide rates. In our study, we observed a clear but statistically insignificant reverse correlation between suicide rates and the volume of suicide-searches in the regions. However, it is not clear why there are so many disparities in suicide rates among some of the Voivodships (Figure 3).

The psychosocial aspect of presented data is particularly interesting, but broader than the scope of this article. The comprehensive annual reports on social well-being in Poland - "The Social Diagnosis" (edited by Czapiński and Panka), note that Poles show successively less and less frequent occurrence of suicidal ideations (Czapiński & Panek 2015). Nevertheless, the suicide rates were rising. The reverse correlation between online search about suicide and the actual suicide deaths described in this study may be a signal of changing trend in committing these actions – they may have been committed abruptly, perhaps impulsively, without previous anticipation (Rimkeviciene et al. 2015). On the other hand, the mentioned reverse correlation may indirectly be an indication of reduction of suicidal tendencies or/and finding help by searching

online for information on these topics (Bell et al. 2017, Thornton et al. 2017). These issues go far beyond the scope of this publication. Further qualitative studies are needed to verify the above premises.

Limitations of this study include, above all the aggregate nature of this data which makes it impossible to refer these finding to individuals or even local groups such as cities. Perhaps restriction to search only the word "suicide" made insufficient to capture the link between online searches and suicide rates. In other study, researchers from Taiwan tested a much larger set of search phrases in the for correlation with suicide ratios (Yang et al. 2011). Their results have showed a strong association with words such as "domestic violence", "divorce" or "depression" (Yang et al. 2011). The Taiwanese authors have also indicated fixed latency intervals for certain keywords, which occurred between search pikes and the subsequent increase of suicides (Yang et al. 2011). Failure to include such factors might be another potential limitation of our study.

Another possible limitation of our study is the potentially low validity of the Google Trends tool (Tran et al. 2017). This study summarized the previously published studies on suicide data using Google Trends and criticized the faulty statistical methods described. Tran et al. stated that total number of significant associations matched the number of expected Type I statistical errors. They also commented on the proprietary structure of Google Trends algorithms and tendency to overfitting, when a large number of potential predictors has been used to predict only a small number of cases. Thus, a simple correlation method of two data series used in our study (the Google Trends and suicide rates) may not be appropriate to derive conclusions. Further studies should use more sophisticated statistical tools to adjust for potential biases e.g., autocorrelational pattern or temporal trends (Tran et al. 2017)

The authors believe that the presented (although limited) findings may contribute to a more effective epidemiological prediction of suicide as well as to planning of effective preventive interventions. Mathematical models based on the Google Trends engine may contribute to development of suicide estimates before the official data is published. By profiling the specific time intervals of peak suicide information search, targeted crisis intervention and mental health promotion can be addressed. Further research, also using other platforms and tools, is necessary to sufficiently assess the online activity associated with suicidal tendencies.

## CONCLUSIONS

The presented results may contribute to more effective suicide prevention programs. By specifying the time intervals in which searching suicide information is the highest, it should be easier reach individuals at risk. Further research, using other tools, are needed in order to better estimate the online activity associated with suicidal tendencies.

**Acknowledgements:** None.

**Conflict of interest:** None to declare.

### **Contribution of individual authors:**

Przemysław Maciej Waszak - study design, data collection, first draft, approval of the final version, statistical analysis

Patryk Górski - study design, data collection, first draft, approval of the final version, statistical analysis

Janusz Springer - first draft, approval of the final version, statistical analysis

Wioleta Kasprzycka-Waszak - data collection, first draft, approval of the final version,

Marta Duży - data collection, first draft, approval of the final version

Paweł Zagożdżon - study design, first draft, approval of the final version.

### **References**

1. Arora VS, Stuckler D & McKee M: Tracking search engine queries for suicide in the United Kingdom, 2004-2013. *Public Health* 2016; 137:147-53
2. Bell J, Mok K, Gardiner E & Pirkis J: Suicide-Related Internet Use Amongst Suicidal Young People in the UK: Characteristics of Users, Effects of Use, and Barriers to Offline Help-Seeking. *Archives of Suicide Research* 2018; 22:263-277
3. Bernal M, Haro JM, Bernert S, Brugha T, de Graaf R, Bruffaerts R et al.: Risk factors for suicidality in Europe: Results from the ESEMED study. *Journal of Affective Disorders* 2007; 101:27-34
4. Carneiro HA & Mylonakis E: Google Trends: A Web-Based Tool for Real-Time Surveillance of Disease Outbreaks. *Clinical Infectious Diseases* 2009; 49:1557-1564
5. Czapiński J & Panek T: Social Diagnosis 2015 Objective And Subjective Quality Of Life In Poland, *Quarterly of University of Finance and Management in Warsaw, Warsaw* 2015
6. Fond, G, Gaman A, Brunel L, Haffen E & Llorca PM: Google Trends®: Ready for real-time suicide prevention or just a Zeta-Jones effect? An exploratory study. *Psychiatry Research* 2015; 228:913-917
7. Główny Urząd Statystyczny, *Rocznik Demograficzny* 2016. Warszawa, 2016
8. Gunn JF & Lester D: Using google searches on the internet to monitor suicidal behavior. *Journal of Affective Disorders* 2013; 148:411-412
9. Harris K, McLean J & Sheffield J: Examining Suicide-Risk Individuals Who Go Online for Suicide-Related Purposes. *Archives of Suicide Research* 2009; 13:264-276
10. Isometsä E: Suicidal behaviour in mood disorders-Who, When, and Why? *Canadian Journal of Psychiatry* 2014; 59:120-130
11. Jacob N, Scourfield J & Evans R: Suicide prevention via the Internet: a descriptive review. *Crisis* 2014; 35:261-267
12. Młodożeniec A, Brodniak WA, Polewka A & Bembenek A: Seasonality of suicide in Poland. Analysis of the Main Statistical Office data for the years 1999-2003. *Psychiatria Polska* 2010; XLIV:61-69
13. Nuti SV, Wayda B, Ranasinghe I, Wang S, Dreyer RP, Chen SI et al.: The use of google trends in health care research: A systematic review. *PLoS ONE* 2014; 9:e109583
14. O'Connor RC, Platt S & Gordon J: *International Handbook of Suicide Prevention: Research, Policy and Practice*, John Wiley & Sons, Ltd, West Sussex, 2011
15. Parker J, Cuthbertson C, Loveridge S, Skidmore M & Dyar W: Forecasting state-level premature deaths from alcohol, drugs, and suicides using Google Trends data. *Journal of Affective Disorders* 2016; 213: 9-15
16. Polewka A, Kroch S & Chrostek-Maj J: Suicidal behavior and suicide attempts in adolescents and young adults--epidemiology, risk factors, prevention and treatment. *Przegląd Lekarski* 2004; 61:261-264
17. Rimkeviciene J, Ogorman J & De Leo D: Impulsive suicide attempts: A systematic literature review of definitions, characteristics and risk factors. *Journal of Affective Disorders* 2015; 171:93-104
18. Scocco P, de Girolamo G, Vilagut G & Alonso J: Prevalence of suicide ideation, plans, and attempts and related risk factors in Italy: Results from the European Study on the Epidemiology of Mental Disorders - World Mental Health study. *Comprehensive Psychiatry* 2008; 49:13-21
19. Thornton L, Handley T, Kay-Lambkin F & Baker A: Is A Person Thinking About Suicide Likely to Find Help on the Internet? An Evaluation of Google Search Results. *Suicide and Life-Threatening Behavior* 2017; 47:48-53
20. Tran US, Andel R, Niederkrotenthaler T, Till B, Ajdacic-Gross V & Voracek M: Low validity of Google Trends for behavioral forecasting of national suicide rates. *PLoS ONE* 2017; 12:1-26
21. Vijayakumar L: Suicide prevention: the urgent need in developing countries. *World Psychiatry* 2004; 3:158-159
22. World Health Organization, *World Health Statistics - Monitoring Health For The SDGs*, World Health Organization, Geneva, 2016
23. Yang AC, Tsai SJ, Huang NE & Peng CK: Association of Internet search trends with suicide death in Taipei City, Taiwan, 2004-2009. *Journal of Affective Disorders* 2011; 132:179-184

Correspondence:

Przemysław Waszak, MD  
Department of Hygiene & Epidemiology, Medical University of Gdansk  
Ulica Dębinki 7, Gdansk, Poland  
E-mail: p.waszak@gumed.edu.pl