

Andrzej Jarynowski^{1,2}, Monika Wójta-Kempa³, Vitaly Belik²

TRENDS IN INTEREST OF COVID-19 ON POLISH INTERNET

TRENDY W ZAINTERESOWANIU TEMATEM COVID-19 W POLSKIM INTERNECIE

¹Interdisciplinary Research Institute, Wrocław, Poland

²System Modeling Group, Institute of Veterinary Epidemiology and Biostatistics, Freie Universität Berlin, Germany

³Department of Public Health, Faculty of Health Science, Wrocław Medical University, Poland

¹Instytut Badań Interdyscyplinarnych we Wrocławiu, Polska

²Grupa Modelowania Systemowego, Instytut Biometrii i Epidemiologii Weterynaryjnej,
Wolny Uniwersytet Berliński, Niemcy

³Zakład Medycznych Nauk Społecznych, Katedra Zdrowia Publicznego,
Uniwersytet Medyczny we Wrocławiu, Polska

ABSTRACT

INTRODUCTION. This study aims to investigate the social interest in SARS-CoV-2 and COVID-19 in the Internet media during the epidemic in Poland and serves as a complement to longitudinal surveys monitoring public perception in real time.

METHODS. We quantitatively analyzed “coronavirus” digital footprints on the Internet, in particular, on Google, Twitter, YouTube, Wikipedia and electronic media represented by EventRegistry, from 15.01 to 07.04.2020 (before and after official introduction of COVID-19 to Poland on 04.03.2020). We present results on trend analysis supported by statistical techniques.

RESULTS. We identified six major temporal clusters of the interest on the topic COVID-19: 1) Chinese, 2) Italian, 3) Waiting, 4) Mitigations, 5) Physical distancing and Lockdown, 6) Anti-crisis shield. There was an exponential increase of the interest when the Polish government “declared a war against disease” around 11/12.03.2020 with a massive response activity. Later on, there was a decay in the interest with additional phases related to physical distancing and an anti-crisis legislation act associated with pronounced local peaks. We have found that declarations of response strategies by the Polish prime minister or the minister of health gathered the highest attention of Internet users. So already enacted or in force events do not affect the interest to such an extent. We have observed very weak or even negative correlations between a colloquial search term “antiviral mask” in Google, encyclopaedic definition in Wikipedia “SARS-CoV-2” as well as official COVID-19 incidence series, implying different mechanisms governing the search for knowledge, panic related behaviour and actual burden of COVID-19.

CONCLUSIONS. Traditional and social media do not only reflect reality, but also create it. Interest in COVID-19 (measured by topic intensity on the Internet) is unrelated to the officially notified viral pressure (measured by incidence) and probably physical risk of acquiring infection in Poland. As traditional media are ahead of social media in time, we advise to choose traditional news media for a quick dissemination of information, however for a greater impact, social media should be used. Otherwise public information campaigns might have less impact on society than expected.

Keywords: *Internet, medical sociology, epidemiological monitoring, SARS-CoV-2, COVID-19, media*

STRESZCZENIE

WSTĘP. Prezentowane badanie ma na celu zilustrowanie społecznego zainteresowania użytkowników Internetu tematami związanymi z koronawirusem SARS-CoV-2 i chorobą Covid-19 podczas epidemii w Polsce. Może ono wnieść dodatkową wartość do podłużnych badań sondażowych nad społeczną percepcją ryzyka mierzoną w czasie rzeczywistym.

METODY. Wykorzystano analizę ilościową śladów cyfrowych zainteresowania „Koronawirusa” w polskim Internecie (głównie w Google i obsługiwanych przez Twitter, YouTube, Wikipedia i media elektroniczne reprezentowane przez EventRegistry) w okresie 15.01-07.04.20 (przed i po oficjalnym potwierdzeniu zawleczenia zakażenia, które miało miejsce 04.03.20). W tym celu wykorzystujemy analizę trendów przy wsparciu prostych technik analizy statystycznej.

WYNIKI. Zidentyfikowaliśmy sześć głównych grup czasowych na temat „Koronawirusa”: 1) chińska, 2) włoska, 3) oczekiwania, 4) epidemii i restrykcji, 5) zamknięcia i dystansowania, 6) „Tarczy antykryzysowej”. Po wypowiedzeniu „wojny” przeciw COVID-19 przez Polskę nastąpił gwałtowny wzrost zainteresowania ze szczytem 11/12.03.20. Później nastąpił zanik zainteresowania z dodatkowymi fazami: dystansowania i tarczą antykryzysową z lokalnymi plikami. Dowiedliśmy, że to deklaracje restrykcji (a nie wydanie rozporządzeń czy wprowadzenie ich w życie) przez polskiego Premiera lub Ministra Zdrowia przyciągnęły największą uwagę internautów. Zauważyliśmy bardzo niskie lub nawet ujemne korelacje między wyszukiwaniem potocznego terminu „maska antywirusowa” w Google, encyklopedyczną definicją z Wikipedii „SARS-CoV-2”, a także oficjalną raportowaną surową zapadalnością zakażenia, co oznacza, że istnieją różne mechanizmy rządzące poszukiwaniem wiedzy, zachowaniami przed-panicznymi a rzeczywistym ryzykiem wynikającym z potencjalnego zakażenia.

WNIOSKI. Tradycyjne i społecznościowe media nie tylko odzwierciedlają rzeczywistość, ale także ją tworzą. Zainteresowanie tematem COVID-19 w Polsce nie ma związku z oficjalnie odnotowywaną presją wirusa w Polsce i być może faktycznym fizycznym ryzykiem nabycia SARS-CoV-2. Jako że media tradycyjne wyprzedzają media społecznościowe, zalecamy kierowanie konkretnych komunikatów (np. dotyczących sprawozdawczości epidemiologicznej) w tradycyjny sposób w celu szybkiego rozpowszechnienia informacji, jednak dla większego efektu kampanii informacyjnych należy korzystać przede wszystkim z mediów społecznościowych. W przeciwnym razie publiczna sprawozdawczość i kampanie informacyjne mogą mieć mniejszy wpływ na społeczeństwo niż oczekiwano.

Słowa kluczowe: *Internet, socjologia medycyny, monitorowanie epidemiologiczne, SARS-CoV-2, COVID-19, media*

INTRODUCTION

Although a large part of the Polish population has heard about coronaviruses for the first time in the beginning of 2020, in reality they face less dangerous coronaviruses causing simple cold all the time. Only the emergence of a novel strain – SARS-CoV-2 from Wuhan gave the word “Coronavirus” (in Polish “Koronawirus”) a new dramatic meaning. Before the first reported case, less than a half of surveyed Poles believed that “Coronavirus” is the most important topic in the second half of February 2020 (1). The disease was not detected in Poland until 04.03.20. This topic did not receive much attention until the official introduction and only started to drive media and social life afterwards. During opinion polls performed on 09-10.03.20, 63% of respondents reported that this is a serious threat for Poland, 40% that this is a serious threat for their family and 73% that this is a serious threat for the Polish Economy (2). Opinion polls performed on 25-27.03.20 showed that almost 90% surveyed Poles declared “Koronawirus” as the most important current threat for them (3). The actual numbers should be carefully interpreted due to different methodologies used in each of these studies, so longitudinal studies such as ours are urgently needed.

WPROWADZENIE

Mimo iż duża część polskiego społeczeństwa po raz pierwszy usłyszała o koronawirusie na początku 2020 roku, w rzeczywistości z mniej niebezpiecznymi jego odmianami wywołującymi przeziębienia, styka się niemal cały czas. Pojawienie się nowego szczepu - SARS-CoV-2 z Wuhan dało pojęciu „koronawirus” nowe, dramatyczne znaczenie. Przed pojawieniem się pierwszego przypadku, w połowie lutego 2020 roku, mniej niż połowa ankietowanych Polaków była przekonana, że „koronawirus” to ważny społecznie temat (1). Pierwszy potwierdzony przypadek zarejestrowano dopiero 4.03.2020 r. Do tego czasu temat potencjalnych zakażeń nie zdobył szerszego zainteresowania opinii społecznej i mediów. Podczas sondażu przeprowadzonego w dniach 9-10.03.2020 63% respondentów określiło „koronawirusa” jako poważne zagrożenie dla Polski i tylko 40% uznało go za poważne zagrożenie dla swojej rodziny, zaś 73% wskazało, jako zagrożenie dla polskiej gospodarki (2). Dopiero sondaże przeprowadzone pomiędzy 25-27.03.2020 uwiaryściły, że „koronawirus” stał się najpoważniejszym bieżącym zagrożeniem dla 90% uczestników badania (3). Należy jednak powyższe liczby porównywać z ostrożnością, gdyż badania zostały wykonane z wykorzysta-

The Internet and its wide resources are the most common way to get information. The perception of many new phenomena is moderated or even created by the media. Recently, media activities are being analyzed worldwide for better understanding of perception and spread of diseases (4-6). In 2020 Internet media (both internet of “traditional” TV, radio and press media, which satisfy standard sender-receiver communication paradigm as well as new “social-content” media, which form nonlinear communication patterns) can serve as a valuable source of information as well as disinformation about the COVID-19 (7). New field of research - infodemiology and infoveillance (8) – should be employed for standard surveillance purposes within the emerging field of “digital epidemiology” and in the area of public health.

In this study, by quantitative analysis of “digital traces/footprints” on the Internet (social-content and traditional media), we concentrate on the number and nature of events of interest measures such as information queries and information propagation. We see “social perception” of COVID-19 pandemic and SARS-CoV-2 virus as a complicated process, when subjects (people) give or read the meaning of the participating objects through communication behaviors. Seeking for information in a source (we use the term “interest”) is the important step to find the definition of a situation. We may observe “massive attack on information” when the phenomenon is new and/or important for society. Thus social interest is an immanent part of perception (also risk perception) and it can be easily measured by “digital footprints”.

To the best of our knowledge there were no previous studies quantitatively linking the Internet activities and interest in infectious diseases in Poland (9, 10), except for our own studies (11, 12) and some straightforward attempts such as correlating the incidence of influenza with Google Trends (4). Moreover, during the first few weeks of COVID-19 pandemic, a tremendous amount of related studies appeared targeting English and Asian languages (13) with an underrepresentation of further European languages. Thus the present study is the first exploratory attempt to fill this gap continuing a preliminary research on the interest in “Koronawirus” in Poland before the first reported case (11) representing a more general approach and highlighting the role of social sciences in public health during COVID-19 pandemic (12).

We collected data on information-seeking behaviors (digital footprints) and compared them with real time COVI-19-related events. These efforts are a contribution to public health surveillance (13) by collecting and analyzing health-related information.

niem różnych metodologii, tym bardziej jest potrzeba badań podłużnych (jak nasze).

Internet i jego szerokie zasoby są najłatwiejszym sposobem pozyskania informacji. Postrzeganie (szczególnie) nowych zjawisk jest moderowane lub nawet kreowane przez media. Ostatnio szeroką popularność zyskują analizy aktywności użytkowników mediów, dają one bowiem wiedzę na temat postrzegania i rozprzestrzeniania się chorób (4-6). W 2020 roku media (zarówno tradycyjne, jak radio, telewizja czy prasa opierające się na standardowej komunikacji nadawca -odbiorca, jak i nowe „społecznościowo-kontentowe media”, kształtujące nieliniarne formy komunikacji) służą jako wartościowe źródło informacji jak i dezinformacji o Covid-19 (7). Wyłaniający się obecnie nowy obszar badań – infodemiologia i wywiad informacyjny (8) - powinny stać się uzupełnieniem standardowych działań wywiadowczych (rozpoznawczych) w ramach „cyfrowej epidemiologii” i w obszarze zdrowia publicznego.

Poprzez ilościową analizę „ślądu cyfrowego” w mediach społecznościowych i tradycyjnych autorzy koncentrują się na liczbie i naturze mierzonych zjawisk związanych z zainteresowaniem pandemią, takich jak wyszukiwanie i rozpowszechnianie informacji. Zakładamy przy tym, że społeczna percepcja choroby COVID-19 i wirusa SARS-CoV-2 jest skomplikowanym procesem, w którym jednostki nadają lub odczytują znaczenie obiektów (zdarzeń) w trakcie zachowań komunikacyjnych. Poszukiwanie informacji u określonych „dystrybutorów” informacji (źródeł), w celu definicji sytuacji jest najważniejszym elementem szerszego konstruktu społecznego nazywanego „zainteresowaniem”. Masowe zachowania komunikacyjne w celu poszukiwań informacji (obecnie bardzo często w Internecie) obserwujemy, gdy zjawisko jest nowe i/ lub ważne dla społeczeństwa. Społeczne zainteresowanie (traktowane jako zachowanie komunikacyjne) jest immanentną częścią społecznej percepcji (w tym percepcji ryzyka) i naszym zadaniem jest próba jego zmierzenia za pomocą owego „ślądu cyfrowego”.

Według wiedzy autorów nie istnieją wcześniejsze prace badawcze łączące zachowania użytkowników polskiego Internetu z zainteresowaniem chorobami zakaźnymi (9,10), za wyjątkiem naszych badań (11,12) i prób korelacji zapadalności na grypę i zachorowania grypopodobne z zapytaniem w Google Trends (4). W czasie początków trwania pandemii koronawirusa ukazało się wiele opracowań w języku angielskim i językach azjatyckich (13), niewiele zaś w językach z obszaru Europy. Z tego powodu można traktować niniejsze opracowanie jedną z pierwszych prób uzupełnienia luki w opisie społecznego zainteresowania koronawirusem (kontynuując analizy przed pojawieniem się pierwszego (potwierdzonego) przypadku w Polsce

We argue that presented tools are able to track population's behavior and monitor or even predict abnormal behaviors (like restrictions or harmful beliefs caused by fake news). Such data analytics may provide near real time outbreak information in various formats, independently from official statistics or surveys reports. We aimed to describe phases of the interest in COVID-19 on the Polish Internet, as well as to indicate the most important topics of social perception of the new disease.

METHODOLOGY

We primarily analyzed quantitative digital footprint data on the Internet during 15.01–07.04.20 including their representativeness. The only exception is Wikipedia data, which is only available from 10.02.20 due to the virus name change by WHO. The unit of our analysis is daily interest, defined in a different form for each platform. As a main keyword for investigated platforms (10) we chose a colloquial term “Koronawirus”/“Coronavirus” due to its high penetration in the society. Other related keywords in use are much less popular, except for Wikipedia, where the biomedical term SARS-CoV-2 was frequently chosen (12). Data from each platform was collected for comparative purposes, to describe the problem as broadly as possible:

- a) from Google (social-content media/ information providing) we collected a daily relative search activity (interest) for “Koronawirus” in terms of queries intensity and Google news picked by Google from Google Trends. The measured interest is normalized to 100 by Google, where 100 represents the maximal share of the investigated keyword over the investigated time period. Google does not disclose exact search volume. However, according to our own heuristic calculations (by comparing series to see relative relations (14)) based on Google Ads service which gives just a range of queries for selected periods, we can provide at least an order of magnitude of the volume. Google has a 95% share among Polish Internet users (around 26 Millions) with over 8 billion entries per month and is the undisputed leader on the Polish Internet market (15).
- b) from YouTube (social-content/ commentary media) we collected a daily relative search volume of “Koronawirus”. The measured Interest is as well normalized to 100 in the same way as in Google searches and news. YouTube has a 68% share among Polish Internet users with about 700 million entries per month (15). In addition, streams from the YouTube mobile app should be

(11)), które reprezentuje szersze podejście i uwypukla rolę nauk społecznych w obszarze zdrowia publicznego w czasie trwania pandemii COVID-19.

Autorzy zgromadzili dane z zakresu zainteresowania („ślad cyfrowy”) i porównali je z pojawiającymi się w czasie rzeczywistym przypadkami COVID-19. Działania te stanowią wkład w działalność „wywiadu” zdrowia publicznego (13) poprzez zbieranie i opracowanie informacji związanych z zachowaniami zdrowotnymi. Autorzy wyrażają przekonanie, iż wykorzystane narzędzia pozwalają śledzić zachowania populacyjne oraz monitorować lub nawet przewidzieć zachowania nieprawidłowe (np. niestosowanie się do wprowadzonych restrykcji czy wiara i rozpowszechnianie nieprawdziwych informacji, tzw. „fake news”). Takie analizy mogą przynieść informacje o realnych zagrożeniach epidemiologicznych, niezależnie od oficjalnych statystyk czy sondaży opinii. Cel ów można osiągnąć m.in. przez opis faz zainteresowania COVID-19 w polskim internecie oraz wskazanie najważniejszych wątków społecznej percepcji nowej choroby zakaźnej.

METODOLOGIA

Autorzy przeanalizowali ilościowe dane pochodzące z internetowego śladu cyfrowego wraz z oceną ich reprezentatywności w okresie 15.01–07.04.2020 (dla Wikipedii dane dostępne są od 10.02.2020 ze względu na zmianę nazwy wirusa i wywoływanej przez niego choroby). Jednostką analizy są dzienne wskaźniki zainteresowania na każdej z platform. Hasłem kluczowym jest kolokwialne słowo „Koronawirus” (10), które ma największą penetrację w społeczeństwie. Inne hasła były znacznie rzadziej wykorzystywane w przestrzeni medialnej, za wyjątkiem Wikipedii, gdzie pojawił się encyklopedyczny termin SARS-CoV-2 (12). Dane z każdej platformy zgromadzono dla celów porównawczych, aby opisać problem z jak najszerszej perspektywy:

- a) z Google (medium informacyjne i społeczno-kontentowe) zbierano dzienne szeregi czasowe z hasłem „Koronawirus” w zapytaniach i newsach za pomocą narzędzia Google Trends. Mierzona intensywność nie jest podawana przez Google w liczbach bezwzględnych, lecz jedynie na wykresach normalizowanych w skali 0-100, gdzie 100 oznacza maksymalne zainteresowanie hasłem w danym przedziale czasowym. Aby oszacować rząd wielkości liczby zapytań zastosowano heurystyczny algorytm (14) porównujący różne wielkości z przybliżonych danych udostępnianych przez Google Ads. Google jest niekwestionowanym liderem rynku internetowego z ponad 8 miliardami wejść miesięcznie i jest wykorzystywane przez ok

taken into account, as it is the most popular app on Poles' smartphones (15).

- c) from Twitter (social-content/ commentary media) we collected a daily number of Tweets with #Koronawirus hashtag using the Twitter API (the only one allowing us to measure the active Internet usage – production and retweeting of information). Twitter in Poland has a relatively low popularity (~3 million registered users or less than 8% of the country population) and is mainly used by expats, journalists and politicians (16).
- d) from EventRegistry (traditional/ information providing media) we collected a daily number of news/articles with the topic "Koronawirus". This source gives us a representative sample of articles with the highest reach among the Internet users, because it gives priority to digital versions of other broadcasting channels, including television, radio or newspapers.
- e) from Wikipedia (social-content/ information providing media) we collected a daily number of page visits for "SARS-CoV-2" articles. Wikipedia has an Internet coverage of 57% with over 350 million page visits per month among Polish Internet users (14) and is distinguished by a significant overrepresentation of users with tertiary education inhabiting big cities (affinity index > 120 (17)).

For comparison with standard COVID-19 surveillance variables, we used case series published by the European Centre for Disease Prevention and Control (ECDC). We used daily number of registered COVID-19 cases (incidence) and 14-days notification rate (per 100 000 inhabitants).

RESULTS AND DISCUSSION

Phases of interest

To compare the interest in COVID-19 on different Internet platforms we visualized the available queries and interest measurements as time series all together and marked events important to the Polish public (Fig. 1). The largest peak of interest in our analysis for most platforms on 11/12.03.2020 (except the news which peaked at 16.03.2020) was on the day of the pandemic declaration by WHO and the declaration of a massive mitigation program (18) in Poland.

The topic "Koronawirus" appeared on the Polish Internet for the first time in the middle of January 2020. Before the official disease introduction to Poland three phases of interest could be distinguished. After the outbreak the next three stages could be considered (Fig. 1):

26 milionów mieszkańców Polski, co stanowi 95% użytkowników Internetu (15).

- b) z YouTube (medium komentatorskie i społeczno-kontentowe) zbierano dzienne szeregi czasowe ze względną intensywnością wyszukiwania filmów za pomocą hasła „Koronawirus”. Zainteresowanie jest również znormalizowane do 100 jak to ma miejsce w przypadku Googla. YouTube jest rozpowszechnione wśród 68% użytkowników Internetu i ma około 700 milionów wejść miesięcznie (15). Jest też najpopularniejszą aplikacją mobilną w Polsce (15) i należy więc zwłaszcza uwzględnić ruch (wyszukiwania) ze smartphonów.
- c) z Twittera (medium komentatorskie i społeczno-kontentowe) zbierano dzienne zliczenia tweetów z hashtagiem „Koronawirus” przy wykorzystaniu API (application programming interface, która pozwala mierzyć szereg zmiennych dotyczących aktywności na tej platformie). Twitter jest relatywnie mało popularny w Polsce i ma mniej więcej 3 miliony użytkowników reprezentujących jedynie 8% populacji i jest głównie używany przez ekspatów, dziennikarzy i polityków (16).
- d) z EventRegistry (medium informacyjne i tradycyjne) wybrano dzienne zliczenia artykułów z tematem „Koronawirus”. Za pomocą tego źródła otrzymano reprezentatywną próbę artykułów z największymi zasięgami wśród internautów, gdyż selekcja preferuje nadawców fizycznych: telewizyjnych, radiowych i prasowych.
- e) z Wikipedii (medium informacyjne i społeczno-kontentowe) zbierano dzienne zliczenia wejść na stronę „SARS-CoV-2”. Z Wikipedii korzysta 57% użytkowników internetu w Polsce co przekłada się na ponad 350 mln wejść miesięcznie (14). Użytkownicy to głównie mieszkańcy dużych miast z wyższym wykształceniem (affinity index > 150 (17)).

W celu porównania ze standardową sprawozdawczością rządową, wykorzystano szeregi czasowe publikowane przez Europejskie Centrum Zapobiegania Chorobom (ECDC). Wykorzystano także dzienną liczbę nowych zarejestrowanych przypadków COVID-19 (zapadalność) i 14-dniową skumulowaną zapadalność (obie w przeliczeniu na 100 000 mieszkańców).

WYNIKI I DYSKUSJA

Fazy zainteresowania

Dla porównania zainteresowania wokół COVID-19 na różnych platformach została stworzona wizualizacja dostępnych pomiarów zainteresowania w postaci szeregów czasowych wraz z zaznaczonymi najważniejszymi wydarzeniami. Największy pik

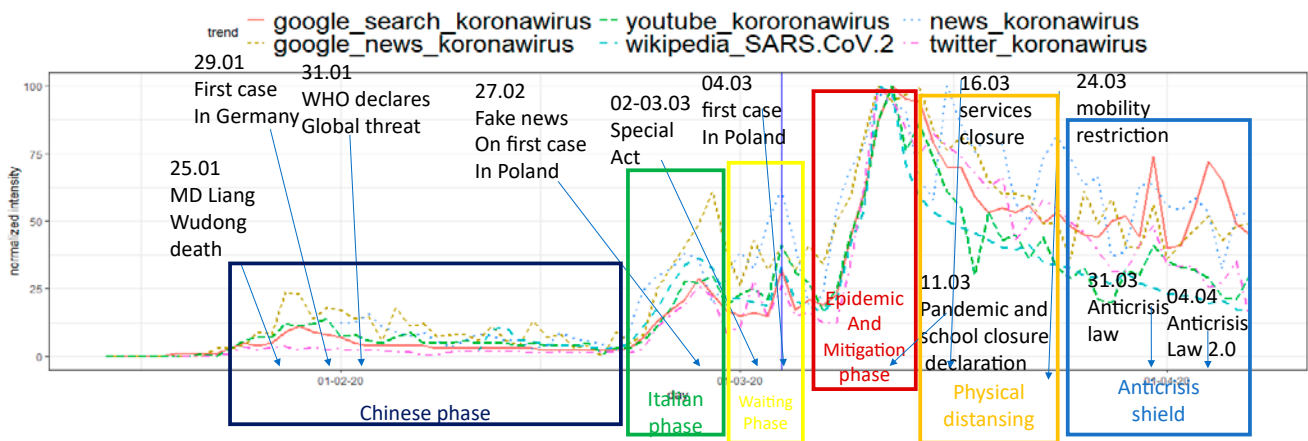


Fig. 1. The intensity of the topic “Koronawirus” on various media platforms during 15.01-07.04.20. Time series were normalized to 100 by a maximal value for a given series. Six phases of interest: 1) Chinese phase (COVID-19 emerges in Asia (19)); 2) Italian phase (second wave of COVID-19 (19)); 3) Waiting phase (awareness building and waiting for a first confirmed case in Poland); 4) Epidemic and Mitigation phase; 5) Physical distancing and Lockdown phase; 6) Anticrisis shield. Official disease introduction is marked by the vertical line.

Ryc. 1. Zestawienie częstości wyszukiwań frazy “Koronawirus” w różnych mediach (Google, Youtube, Wikipedia, Twitter) według wydzielonych faz zainteresowania. Wykres przedstawia szeregi czasowe znormalizowane do 100 według wartości maksymalnej dla każdej serii. Na wykresie zaznaczono 6 faz zainteresowania 1) faza “chińska” (pojawienie się COVID-19 w Chinach (19)), 2) faza „włoska” (druga faza związana z przywleczeniem wirusa do Włoch (19)), 3) faza oczekiwania (rozbudzenie świadomości epidemiologicznej i oczekiwanie na pierwszy przypadek w Polsce); 4) faza epidemii i restrykcji; 5) Faza fizycznego dystansu i obostrzeń; 6) Wprowadzenie Tarczy Antykryzysowej. Linia pionowa oznacza pierwszy potwierdzony przypadek zakażenia w Polsce (4.03.20).

1. China-related phase: from the end of January and the beginning of February when the epidemic was announced and confirmed in China. We observe a small peak around 25.01.20 (e.g. death of Liang Wudong) and around 29.01.20 (e.g. first case in Germany).
 2. Italy-related phase: from the middle of February till the end of February (when the number of infections rapidly increased in Italy). We see a clear peak around 27.02.20 (e.g. fake news about possible introduction of the disease to Poland (20)).
 3. Waiting phase: beginning of March (waiting for introduction – decreased attention). We see a small peak on the confirmed introduction day (04.03.20).
 4. Mitigation phase: around 11/12.03.20 – mitigation (declaration of main restrictions) and epidemic phase with the highest interest in “Koronawirus”.
 5. Physical distancing and lockdown phase: the middle of March with small peaks corresponding to new restrictions.
 6. Anticrisis act (in Polish media called “Anticrisis shield”) phase with high peaks corresponding to anticrisis shield declarations: from the end of March and the beginning of April.
- zainteresowania dla większości platform miał miejsce 11/12.03.2020 (za wyjątkiem artykułów, których pojawiło się najwięcej 16.03.2020), co było zapewne wynikiem ogłoszenia przez WHO pandemii SARS-CoV-2 na świecie i wiąże się z deklaracjami masowych restrykcji (18) w Polsce.
- Temat „Koronawirusa” pojawił się pierwszy raz w polskim Internecie w połowie stycznia 2020. Przed potwierdzeniem przypadku „0” w Polsce można wyróżnić trzy pierwsze fazy zainteresowania. Kolejne trzy fazy zaobserwowaliśmy już po wybuchu epidemii, w marcu i kwietniu (Ryc. 1):
1. Faza chińska: od ogłoszenia epidemii w Chinach do początku/połowy lutego. Obserwujemy małe wzrosty zainteresowania w okolicach 25.01 (m.in. śmierć sygnalizatora dra Lianga Wudonga) i 29.01 (m.in. pierwszy przypadek w Niemczech).
 2. Faza włoska: od połowy do końcówki lutego (kiedy liczba zakażeń we Włoszech gwałtownie wzrosła). Widzimy ewidentny pik 27.02 związany najprawdopodobniej z fake newsem o pierwszym potwierdzonym zakażeniu w Polsce (20).
 3. Faza oczekiwania: początek marca (wyczekiwania pierwszych przypadków i zmniejszona atencja). Niewielki pik miał miejsce w dniu potwierdzenia pierwszego przypadku (04.03).
 4. Faza epidemii i restrykcji: okres wokół szczytów 11/12.03 oraz 16.03. Obserwujemy największe za-

Queries in Google

Interest in novel Coronavirus on Google can be measured by the number of search queries (Fig. 2). There were around 5 million searches of term “Koronawirus” in March 2020 (Google does not provide absolute search numbers). As YouTube belongs to Google corporation, it is a part of Google Trends results and for our analysis, we selected videos with a keyword “Koronawirus”.

5. Faza „Zostań w domu” i fizycznego dystansowania to druga połowa marca i początek zaniku zainteresowania.
6. Faza tarczy antykrzysowej. Przełom marca i kwietnia to wyciszenie się zainteresowania z wyrażnymi pikami związanym z deklaracjami wdrożenia tarczy antykrzysowej.

Zapytania w Googlu

Zainteresowanie nowym Koronawirusem można mierzyć liczbą zapytań (Ryc. 2). Google Trend podaje, iż w Polsce mieliśmy około 5 milionów zapytań słowem Koronawirus w marcu 2020 (dokładne wartości są tajemnicą handlową Google). W związku z tym, że platforma YouTube należy do korporacji Google, wchodzi więc w skład narzędzia Google Trends i uwzględniliśmy również filmy ze słowem kluczowym „Koronawirus”.

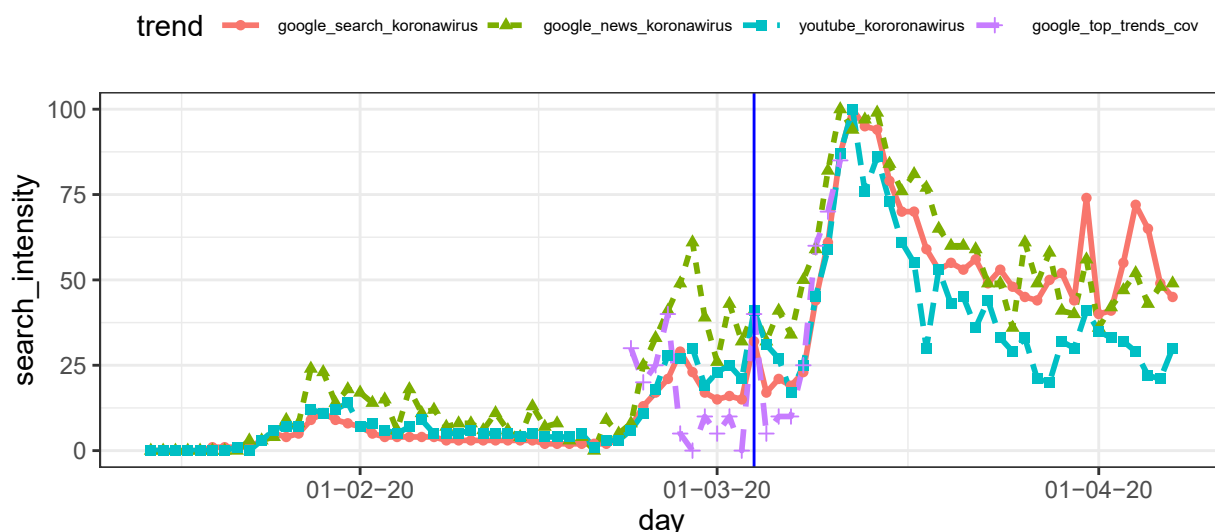


Fig. 2. The intensity of search queries and news with the word “Koronawirus” in Polish Google and search queries with the word “Koronawirus” in Polish YouTube (15.01-07.04.20) and percentage of “Koronawirus” related topics in the Google Top Trend Topics (23.02-11.03.20) both generated using Google Trends tool. Official disease introduction in Poland is marked by the vertical line.

Ryc. 2. Linie trendu wygenerowane przy użyciu Google Trend ukazujące intensywność wyszukiwań dotyczących pojęcia “Koronawirus” w polskim Google, artykułów internetowych i treści na Youtube oraz wyszukiwania związane z “koronawirusem” w Google Top Trend Topics w okresie 23.02-11.03.20. Linia pionowa oznacza datę pierwszego potwierzonego przypadku zakażenia w Polsce (4.03.2020).

It is important to note, that until the first confirmed case in Poland there were no Coronavirus related searches in the top 25 Google queries at all. Although “Koronawirus”-related queries were observed in daily top trending searches before the introduction of the disease to Poland, they start to dominate daily top trending searches only after the massive mitigation measures were announced such as school/university and border closures between 09–12.03.20 (Fig. 2).

Trzeba podkreślić iż „Koronawirus”, ani związane z nim tematy nie pojawiły się ani razu w top 25 dziennych zapytaniach aż do pierwszego potwierzonego przypadku w Polsce. Mimo że tematy około- koronawirusowe pojawiały w Top 20 trendach (czy w najbardziej przyrastających popularnością tematów) przed 4.03, to jednak tematy związane z pandemią zaczęły dominować dopiero po ogłoszeniu masowych obo-

After the first confirmed case in the country, we can see a peak on that day (04.03.20) and substantial growth after important measures were implemented by Polish authorities (09-12.03.20).

strzeń (jak zamknięcie szkół, uczelni i granic) między 9-12.03.2020..

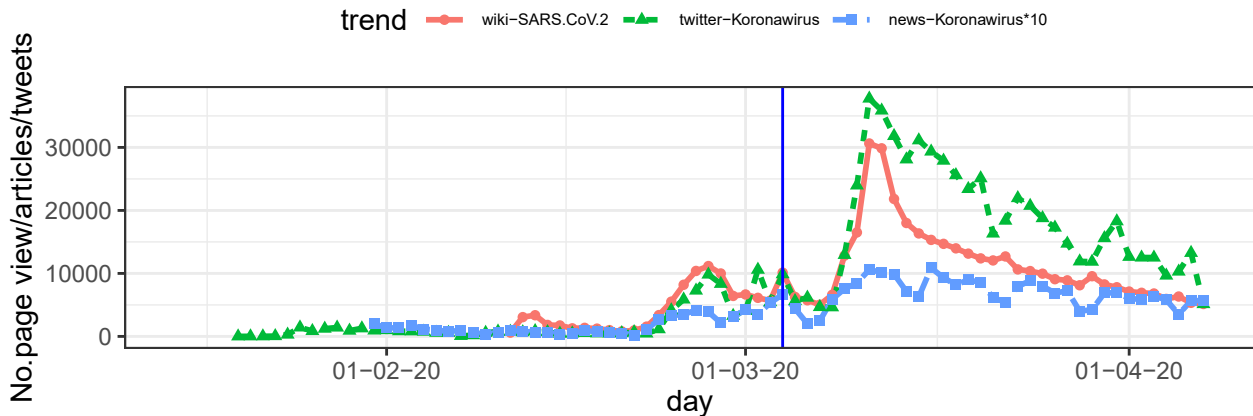


Fig. 3. The intensity of the topic “Coronavirus” in Google for Poland, Italy, Spain, USA, UK and Sweden (15.01-07.04.20) generated using the Google Trends tool. The official disease introduction in Poland is marked by the vertical line.

Ryc. 3. Porównanie linii trendów intensywności wyszukiwań fraz ze słowem “Koronawirus” w Google w Polsce, Włoszech, Hiszpanii, USA, Wielkiej Brytanii i Szwecji. Dane wygenerowano przy użyciu Google Trend. Linia pionowa oznacza pojawienie się pierwszego potwierzonego przypadku zakażenia w Polsce.

Search activity for information on the disease in Poland was still much lower than in countries with a higher global connectivity (21), where the epidemic arrived earlier (Fig. 3) with some exceptions, such as Sweden (a country with a significantly different approach towards epidemic control). On the last day of the investigated period – 07.04.20 the official prevalence “14-days notification rate” in Poland was 9.4 per 100 000 inhabitants, which was below the European Union average (69 per 100 000).

The most important outcome shows the intensity of the search for self-protective measures (Fig. 4, 5). On Figure 4 we can see that the amount of queries regarding protective masks reached around 50 thousand queries daily. The line on Figure 4 shows that Internet users mostly sought information about a product “protective mask” (11) twice: the first peak in end of February and the second in the late March reflecting the interest in protective masks when the necessity of protective masks was discussed. Decrease of interest in the beginning of April may be due to the success of information campaigns on their alleged low effectiveness (in protecting users) or simply due to the lack of the masks on the market. This is a very interesting fact, because before the introduction of SARS-CoV-2 in Poland (before 4.03.20) Polish Ministry of Health did not recommend masks as individual protection measures. It seems that at first Poles looked for risk mitigation measures to confront the politician’s opinion with the

Aktywne wyszukiwania informacji o „Koronawirusie” w Polsce były mniej intensywne niż w krajach „usieciowionych” globalnie (21), do których zakażenia dotarły wcześniej (Ryc. 3), poza paroma wyjątkami, jak Szwecja (kraj o istotnie innym podejściu do kontroli epidemii). Należy jednak podkreślić, że na ostatni dzień badanego okresu (07.04) 14-dniowa skumulowana liczba przypadków wyniosła 9,4 na 100 000 mieszkańców, co było wciąż poniżej średniej w Unii Europejskiej (69 na 100 000).

Do najważniejszych zapytań należą według Google Trends środki ochrony indywidualnej (Ryc. 4, 5). Na rycinie 4 możemy zauważyć, iż dobowe zapytania dotyczące maseczek sięgały nawet 50-ciu tysięcy razy. Internauci wyszukiwali informacji o produkcie *maseczka ochronna* (11) dwukrotnie, pod koniec lutego (zainteresowanie zakupem) i w drugiej połowie marca (dyskusje na temat ich skuteczności i konieczności ich noszenia przez zdrowych obywateli). Spadek zainteresowania maseczkami na początku kwietnia mógł wynikać z medialnej dyskusji na temat domniemanej skuteczności (zwłaszcza nikłej ochrony dla osoby noszącej bez odpowiedniego przeszkolenia) lub po prostu z powodu braku towaru w aptekach stacjonarnych. Warto przypomnieć, że do potwierdzenia pierwszego przypadku Minister Zdrowia Łukasz Szumowski (podał się do dymisji w sierpniu 2020) nie rekomendował używania maseczek jako środka ochrony indywidualnej dla ogółu populacji. Wygląda na to, że Polacy chcieli niezależnie skonfrontować opinie polityków

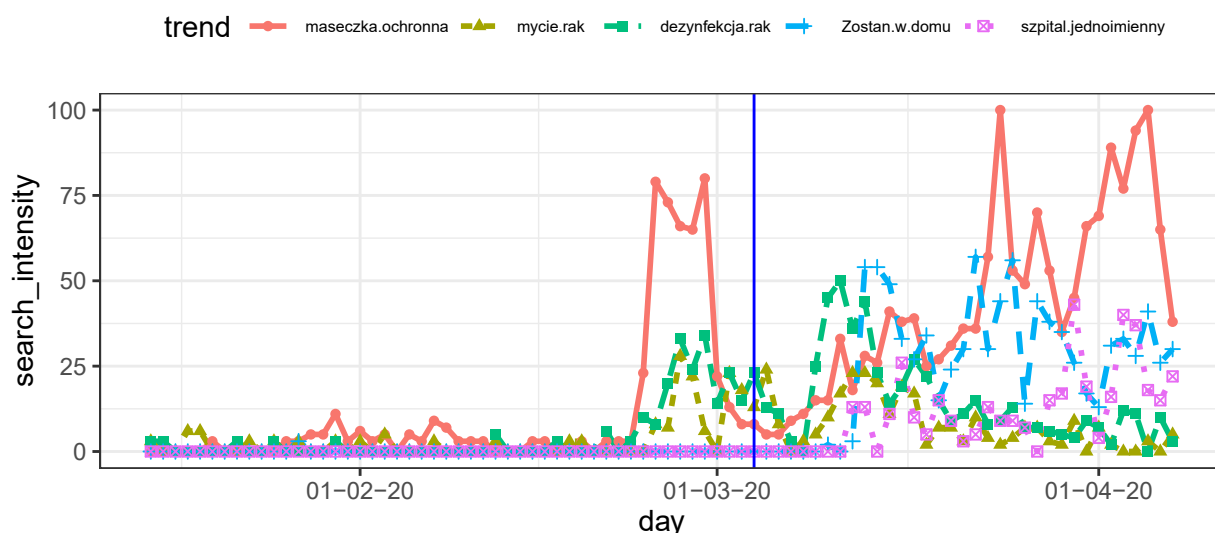


Fig. 4. The intensity of queries with phrases “protective mask”, “hand washing”, “hand disinfection”, “Stay at Home”, “COVID-19 hospital” in Polish Google (15.01-07.04.20) generated using the Google Trends tool. The official disease introduction is marked by the vertical line.

Ryc. 4. Linie trendu intensywności wyszukiwań fraz “maseczka ochronna”, “mycie rąk”, “dezynfekcja rąk”, “Zostań w domu”, “szpital jednoimienny” w polskim Google w okresie 15.01-07.04.20 wygenerowane przy użyciu Google Trends. Linia pionowa oznacza datę pierwszego potwierzonego przypadku zakażenia w Polsce.

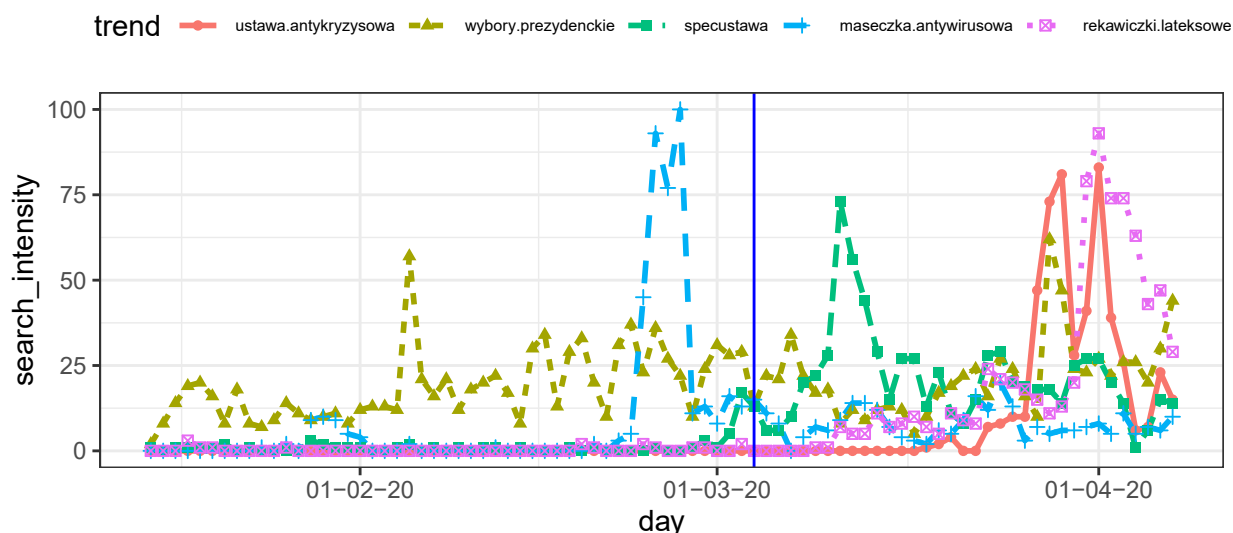


Fig. 5. The intensity of queries with phrases “anticrisis law”, “presidential election”, “special act on COVID-19”, “antiviral mask”, “latex gloves” in Polish Google (15.01-07.04.20) generated using the Google Trends tool. Official disease introduction is marked by the vertical line.

Ryc. 5. Linie trendu intensywności wyszukiwań fraz “ustawa antykryzysowa”, “wybory prezydenckie”, “specustawa”, “maseczka antywirusowa”, “rękawiczki lateksowe” w polskim Google w okresie 15.01-07.04.20 wygenerowane przy użyciu Google Trends. Linia pionowa oznacza datę pierwszego potwierzonego przypadku zakażenia w Polsce.

knowledge retrieved from the Internet community. The special attention must be paid to “antiviral mask” (Fig. 5) – an unprofessional term, but very popular at the beginning of the epidemic (11).

People were searching for information on further epidemiological topics related to infection and protective measures as well: “hand washing”, “hand disinfection”, “Stay at Home”, “COVID-19 hospital” (Fig. 4). The most popular epidemiological queries such as “latex gloves” reached around 100 thousand

z wiedzą pozyskaną ze społeczności internautów. Warto zwrócić uwagę na nieprofesjonalne wyrażenie „maseczki antywirusowe” (Ryc. 5), które było bardzo popularne zwłaszcza na początku epidemii (11).

Użytkownicy mediów szukali informacji na temat innych epidemiologicznych pojęć związanych z zakażeniami i ochroną przed nimi, m.in. „mycie rąk”, „dezynfekcja rąk”, „Zostań w domu”, „szpital jednoimienny” (Ryc. 4). Najpopularniejszym wyrażeniem były „rękawiczki lateksowe” z ok. 100 tysiącami za-

queries daily (Fig. 5) at the peak. It should be noticed that professional terms such as “hand hygiene” practically does not appear in queries (below the noise level of 1% of relative search volumes compared to other epidemiological terms (Fig. 4, 5)). Queries related to hand hygiene had a peak in the late February and the beginning of March.

However, the most popular terms are related to political issues such as anticrisis shield, which reached around 200 thousands queries daily (Fig. 6) at the peak.

pytań dziennie w piku (Ryc. 5). Należy zauważyć, że profesjonalne terminy, takie jak „higiena rąk” praktycznie nie pojawiały się w zapytaniach (poniżej poziomu szumu czyli 1% względnych wyszukiwań w porównaniu do maksimum innych epidemiologicznych wyrażień) (Ryc. 4, 5)). Zapytania związane z szeroko pojętą higieną rąk miały szczyty zainteresowania pod koniec lutego i na początku marca.

Jednakże najpopularniejsze wyszukiwania były związane z obszarem polityki, jak na przykład „tarcza antykryzysowa”, która osiągała nawet 200 tysięcy wyszukiwań dziennie (Ryc.6).

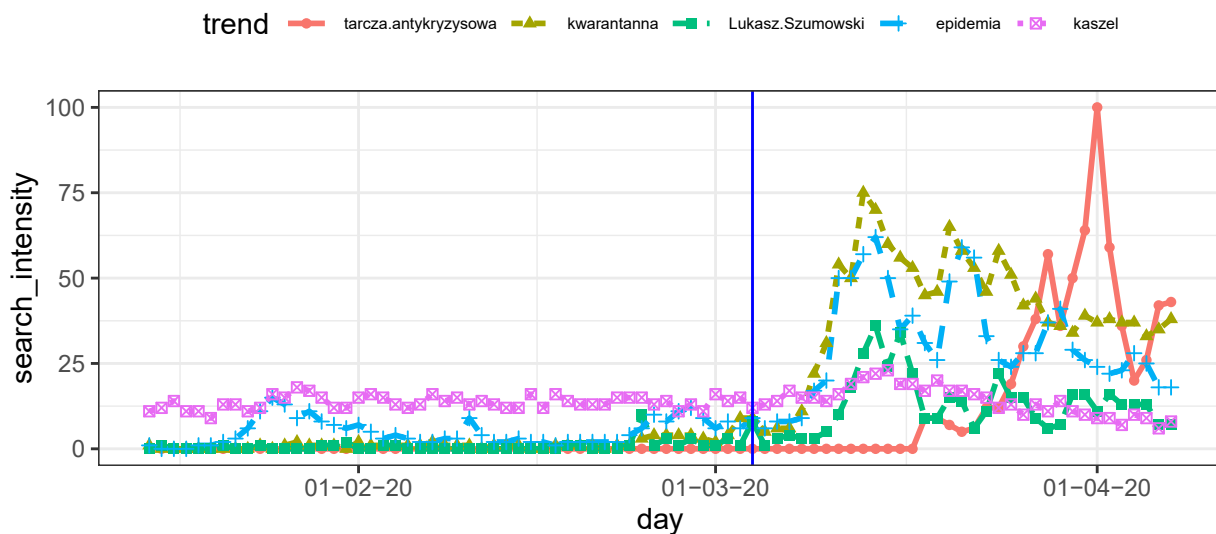


Fig. 6. The intensity of queries with phrases “anticrisis shield”, “quarantine”, “minister of health”, “epidemic”, “cough” in Polish Google (15.01-07.04.20) generated using the Google Trend tool. Official disease introduction is marked by the vertical line.

Ryc. 6. Linie trendu intensywności wyszukiwań fraz “tarcza antykryzysowa”, “kwarantanna”, “Łukasz Szumowski”, “epidemia”, “kaszel” w polskim Google w okresie 15.01-07.04.20 wygenerowane przy użyciu Google Trends. Linia pionowa oznacza datę pierwszego przypadku zakażenia w Polsce.

Interest in ‘coronavirus’ on other platforms

Considering other platforms, there were 492 584 total page views on the Wikipedia article “SARS-Cov-2” during 10.02-07.04.20 and 704 548 tweets with #Koronawirus hashtag during 15.01-07.04.20.

Additionally, during 15.01-07.04.20, 29 885 representative articles (news) were selected from EventRegistry. In the traditional media the weekly seasonality of articles (less articles during weekends on average) is especially prominent (Fig. 7).

Comparison of different platforms

Traditional news agencies (well represented in EventRegistry) as well as Google searches could precede and form more distinct and sharp peaks of interests than social media platforms (Fig. 1, 2, 7). Moreover, discussions on social-content media (YouTube and Wikipedia) demonstrate smoother

Zainteresowanie koronawirusem na innych platformach

Analizując zainteresowanie zjawiskiem na innych platformach zauważyliśmy, że fraza „SARS-Cov-2” miała łącznie 492 584 odsłony w Wikipedii w dniach 10.02-07.04.20. W dniach 15.01-07.04.20. wysłano 704 548 tweetów z hashtagem #Koronawirus. Dodatkowo, za pomocą narzędzia EventRegistry wybrano w przywołanym okresie 29 885 reprezentatywnych artykułów (newsów) na portalach internetowych. Należy podkreślić dobrze widoczną tygodniową sezonowość artykułów (więcej wyszukiwań w dni robocze niż w weekendy) w mediach tradycyjnych.

Porównanie różnych platform

Tradycyjne agencje informacyjne (dobrze reprezentowane w ramach EventRegistry), jak również zapytania w Googlu formują ostrzejsze piki i wyprze-

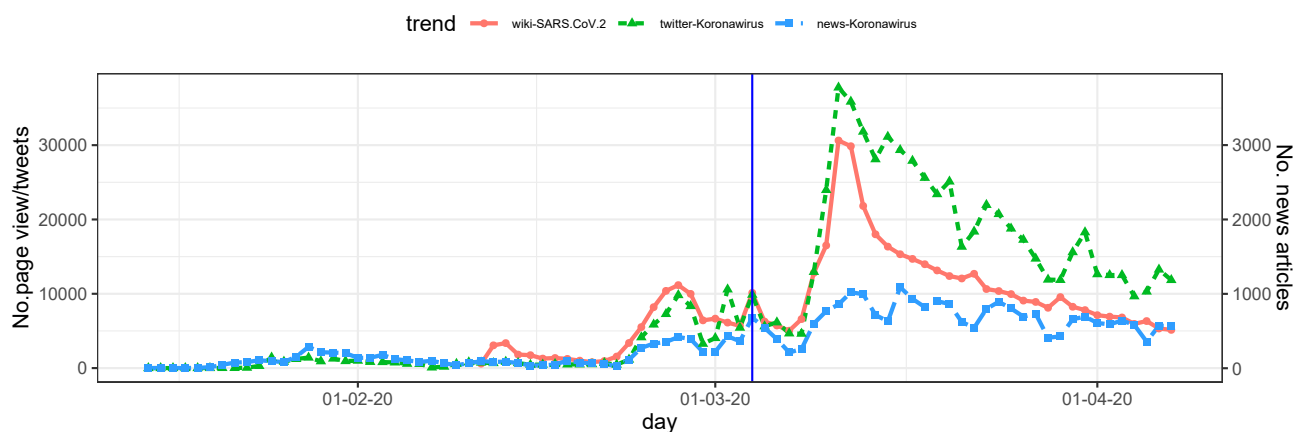


Fig. 7. The intensity of topic “Koronawirus” on various media platforms. For Twitter: Number of tweets per day with the Koronawirus hashtag in Polish language (15.01-07.04.20). For Wikipedia: The number of views of the article “SARS-CoV-2” (10.02-07.04.20). For the news: Number of articles (15.01-07.04.20) generated using EventRegistry. Official disease introduction is marked by the vertical line.

Ryc. 7. Porównanie linii trendu intensywności wyszukiwania tematu “Koronawirus” na różnych platformach. Dla Twittera zaprezentowano liczbę dziennych tweetów w języku polskim z hasztagiem “Koronawirus” (w okresie 15.01-07.04.20). Dla Wikipedii - liczba odsłon artykułu “SARS-CoV-2” (w okresie 10.02-07.04.20). Intensywność wyszukiwania artykułów internetowych określono przy użyciu EventRegistry w okresie 15.01-07.04.20. Linia pionowa oznacza datę pojawienia się pierwszego potwierzonego przypadku zakażenia w Polsce.

interest trajectories (Fig. 7) than information providers (news from EventRegistry, Google).

We detected time lags between different platforms for given topics (e.g. traditional news media are ahead of commentary media in the case of fake news on a possible disease introduction) in phases 1–4 (Fig. 8). It could imply that social media disseminate information via word-of-mouth spreading mechanisms. On the other hand, traditional media have journalists hired to search and select most interesting topics quickly (22). It leads to faster response to events by news agencies (represented by EventRegistry) than social media. The strongest lag (1 day) is observed between news by EventRegistry (the most ahead of time media) and Twitter (the most delayed media) (Fig. 8).

To quantify the similarity in interest on different platforms we calculated Pearson correlations coefficients of the time series (Fig. 9). The majority of measured terms are positively correlated. “Protective mask” as a signal of fear/perception of risk is less correlated with other variables, more related to information needs. “Protective mask” and “washing hands” pair is much less correlated than “protective mask” and “hand disinfection” pair, which could mean that people search for “strong” professional solutions rather than simple and practical ones. Moreover, high amounts of search queries such as “antiviral mask” (there is no such medical term) implies that people were searching for protection measures (that people think are professional) in pseudo-professional terms (11). This *ad hoc-made* neologism or distortion reflects the strong fear-based attitude of common media users. In

dzają zainteresowanie na mediach społecznościowych (Ryc. 1, 2, 7). Warto również podkreślić, że dyskusje w mediach społeczno-kontentowych (YouTube i Wikipedia) charakteryzują się wypłaszczonymi trajektoriami (Ryc. 7) w porównaniu z gwałtownym przebiegiem zainteresowania zarejestrowanym przez EventRegistry i Google.

Zaobserwowaliśmy opóźnienie między poszczególnymi platformami (przykładowo tradycyjne agencje medialne wyprzedziły media komentatorskie w przypadku fake newsa o domniemanym pierwszym przypadku SARS-CoV-2 w Polsce) w fazach 1-4 (Ryc. 8). Sugeruje to, iż media społecznościowe korzystają z mechanizmu sieciowej propagacji informacji. Z drugiej strony tradycyjne media zatrudniają dziennikarzy, którzy wyszukują szybko najciekawsze informacje i podają je do dalszej dystrybucji (22). To mogło spowodować szybszą reakcję na wydarzenia (tym bardziej fikcyjne) przez agencje informacyjne (reprezentowane przez zbiór EventRegistry) w porównaniu z reakcją mediów społecznościowych. Najbardziej widoczne - jednodniowe -opóźnienie występuje między EventRegistry właśnie (medium najszybciej reagujące) a Twitterem (najbardziej „opóźnionym” medium) (Ryc. 8).

Dla porównania podobieństwa między zainteresowaniem na różnych platformach policzono współczynnik korelacji Pearsona dla serii czasowych (Ryc. 9). Większość mierzonych zmiennych jest skorelowana pozytywnie. „Maseczka ochronna” jako sygnał strachu/percepcji ryzyka /jest słabiej skorelowany z większością innych zmiennych, bardziej związanych z potrzebami informacyjnymi Polaków. Para „Maseczka

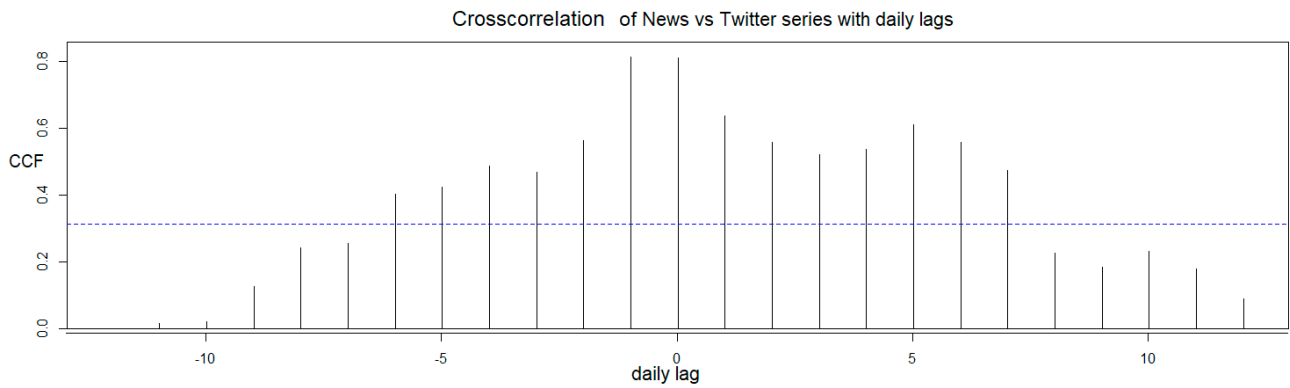


Fig. 8. Crosscorrelation (with daily lags) between daily series of news article counts from EventRegistry (news) and Tweets numbers (31.01–14.03.20) on COVID-19 related topics.

Ryc. 8. Korelacje z opóźnieniami (w dniach) pomiędzy liczbą artykułów cyfrowych a liczbą tweetów (31.01–14.03.20) na temat „Koronawirusa”.

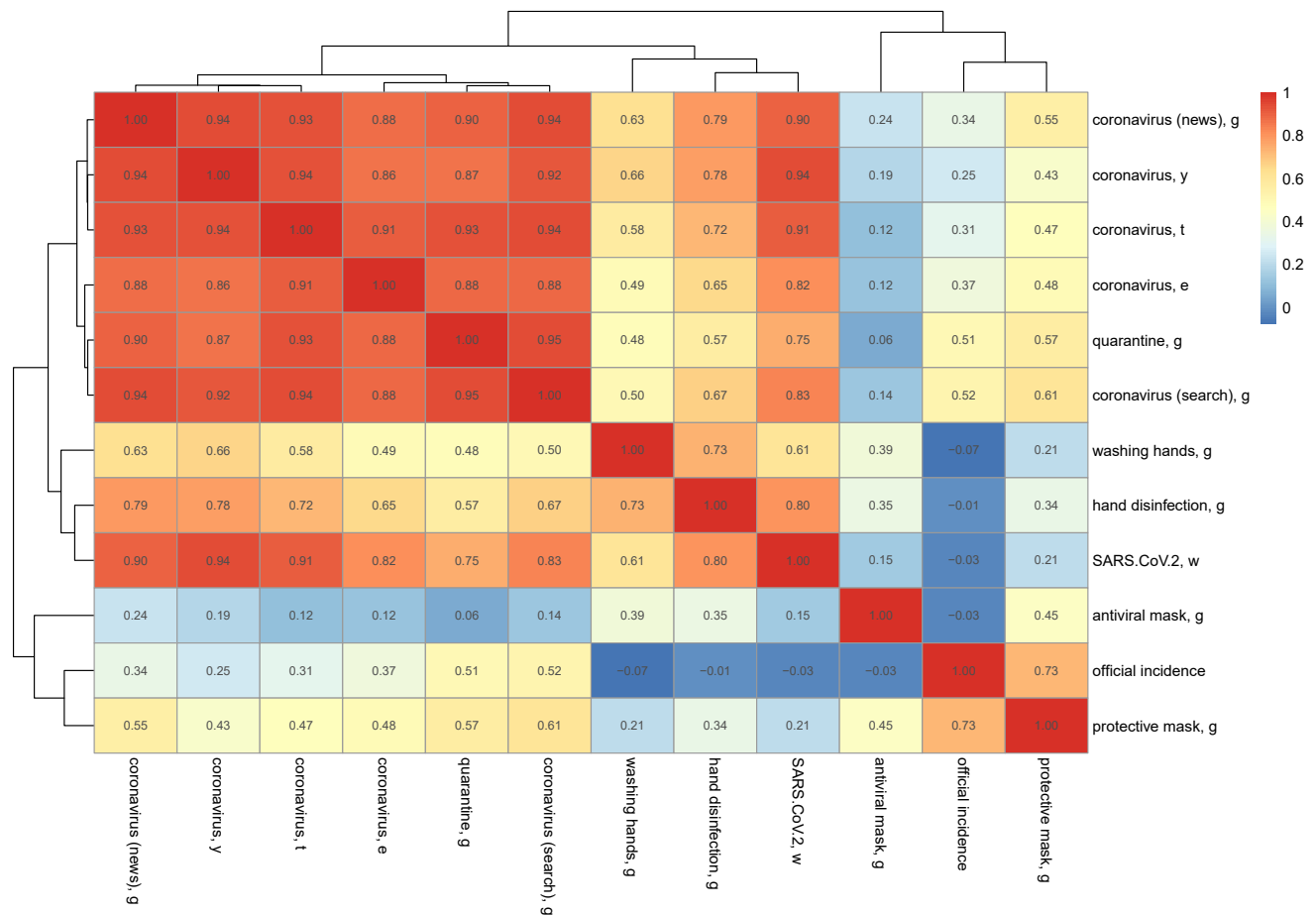


Figure 9. The Pearson correlation matrix and the corresponding hierarchical clustering for the terms “Koronawirus”/“Coronavirus” and related epidemiological queries on various media platforms for pairwise observations during 15.01–07.04.20 period (g – Google, w – Wikipedia, y – Youtube, t – Twitter, e – EventRegistry) and the official COVID-19 incidence. With the significance level 0.05 all correlations were significant except for correlations related to “antiviral mask, g”, “official incidence” and the pairs “protective mask, g”/“washing hands, g” and “protective mask, g”/“SARS-CoV-2, w”. Color scale corresponds to the correlation strength.

Rycina 9. Macierz korelacji Pearsona dla hierarchicznie zgrupowanych do obserwacji parami zainteresowanie terminami “Koronawirus”/“Coronavirus” i innych zapytań epidemiologicznych realizowanych na różnych platformach medialnych w okresie 15.01–07.04.20 (g – Google, w – Wikipedia, y – Youtube, t – Twitter, e – EventRegistry) i oficjalnej zapadalności COVID-19. Przy poziomie istotności 0.05 wszystkie korelacje były znaczące, z wyjątkiem korelacji związanych z “maską antywirusową, g”, “oficjalną zapadalnością” i parami “maska ochronna, g”/“mycie rąk, g” i “maska ochronna, g”/“SARS-CoV-2, w”. Skala kolorystyczna odpowiada sile korelacji.

our opinion it is not simply "Chinese whisper"- effect (popular childrens' game). Language diffusion of non-professional term on the Internet on such a short time horizon must have an emotional background. Daily number of reported incident COVID-19 cases correlates only weakly positively or even negatively (Fig. 9) with the measured interest values. Thus the interest in Poland is significantly uncorrelated with the actual burden of the disease (measured by the incidence), which contrasts situations in some other countries (19). There is a moderate positive correlation between protective masks search queries and the incidence (note that masks were declared to be obligatory after the end of the investigated period).

CONCLUSIONS

Observing the amount of the Internet queries we have to state that citizens have treated the Internet as a very important source of knowledge. Our approach targeted a wide range of general population with a relatively high coverage of Internet users with quite a significant audience variability across platforms. It's important to mention that our analysis operated on the subjective population interest and there is no direct translation between scientific evidence (e.g. the question of the effectiveness of protective masks in infection prevention) and a colloquial knowledge concerning infectious disease and fear of acquiring infection (23). The conclusions of our study are as follows:

1. The retrieved and analyzed data revealed the following six phases of interest in SARS-CoV-2 on the Polish Internet. These phases could be named as: 1. Chinese phase, 2. Italian phase, 3. Waiting phase 4. Epidemic and Mitigation phase, 5. Physical distancing and Lockdown phase, 6. Anticrisis shield. In future we should overlap and confront them to the whole process of emergence and extinction of the pandemic.
2. We noticed the differences in platform usage dependent on knowledge (information) providers because different media (platforms) have separated informative functions – in the following phases of epidemics. Information media (Wikipedia and Google) do not display some peaks of interest (e.g. for the special act against COVID-19), because probably the awareness about the virus and the knowledge on the disease has already been saturated. After the "need for definition phase" (which covers phases 1 and 2) people seem to be interested mainly in the updates (news information on COVID-19) on Twitter, YouTube or other electronic media. This has covered the phase, which might be called "need for norms

ochronna" i „mycie rąk” jest słabiej skorelowana niż para „Maseczka ochronna” i „dezynfekcja rąk”, co może świadczyć, iż ludzie chętniej poszukiwali silnych profesjonalnych rozwiązań niż prostych, choć również skutecznych. Ponadto, wyszukiwanie „maseczki antywirusowej” (nie istnieje taki medyczny termin) oznacza dużą popularność pseudo-profesjonalnych rozwiązań (o których ludzie myślą, iż są profesjonalne). Ten wymyślony/zniekształcony *ad hoc* neologizm jest pochodną atmosfery strachu budowanej przez media w fazie włoskiej. Naszym zdaniem to nie jest prosty mechanizm „głuchego telefonu”, a raczej odzwierciedlenie propagacji emocji w Internecie w krótkiej perspektywie czasowej. Dzienna liczba notyfikacji COVID-19 koreluje słabo dodatnio lub nawet ujemnie (Ryc. 9) z większością mierzonych zmiennych. Stąd należy zauważyć, iż zainteresowanie koronawirusem w Polsce nie jest skorelowane z rzeczywistą presją wirusa w społeczeństwie (mierzą zachorowalnością), co kontrastuje z obserwacjami z innych krajach (19). Jednocześnie mamy do czynienia ze słabą, ale dodatnią korelacją między zapytaniami dotyczącymi wyszukiwania maseczek ochronnych a zapadalnością (należy zauważyć, że obowiązek noszenia maseczek został ogłoszony dopiero po zakończeniu badanego okresu).

WNIOSKI

Obserwacje liczby wyszukiwań internetowych w temacie koronawirusa i pojęć pochodnych prowadzą do konstatacji, że Internet jest bardzo ważnym źródłem wiedzy dla Polaków. Nasze badania objęły szeroką część populacji Polaków korzystających z rozmaitych platform informacyjnych. Należy podkreślić, iż przyjęta metodologia odnosiła się do subiektywnie zdefiniowanego zainteresowania i nie ma bezpośredniego przełożenia między dowodami naukowymi (np. kwestii skuteczności noszenia maseczek w zapobieganiu zakażeniom) a potoczną wiedzą dotyczącą chorób i strachu przed zarażeniem. Wyniki naszych prac prowadzą do następujących wniosków:

1. Uzyskane i przeanalizowane dane ujawniły sześć faz społecznego zainteresowania tematyką SARS-CoV-2 w polskim internecie. Poszczególnym fazom zainteresowani nadaliśmy kolejno nazwy: 1. faza chińska, 2. faza włoska, 3. faza oczekiwania, 4. faza epidemii i obostrzeń, 5. faza fizycznego dystansu (lockdown), 6. faza tarczy antykryzysowej. W przyszłości warto nałożyć i skonfrontować wyżej wymienione fazy zainteresowania z całym procesem powstawania i wygasania pandemii.
2. Zaobserwowano różnice w intensywności korzystania z różnych platform informacyjnych z uwagi na odmienne funkcje informacyjne tych źródeł informacji w różnych stadiach epidemii. Media in-

- phase” (phases 4, 5), because during this time the most searched topics were protection measures and risk mitigation behaviors. That is the reason why the anti-crisis shield issues gathered more interest in the information related media (Google, news articles) as well as on the politically oriented Twitter than on other social media.
3. Risk mitigation strategies have gathered the highest interest on all investigated platforms. What is important, the dates of the declaration of measures’ introduction, not the dates of their enacting or acting in force (18), were responsible for the visible peaks of the interest. This suggests that governmental decisions were instantly picked up by social networks implying a high level of social uncertainty. This raises questions about the role of the media in managing the epidemic and whether the ordered measures were accompanied with proper information campaigns or the information needs were largely unmet.
 4. The focus on the social activity posed in this study revealed trends in a common understanding of self-protective measures. This lay-referral framing system of opinions and beliefs (23) should draw the attention of public health professionals towards evaluation of potential epidemiological risk of infection. Our analysis showed that Poles looked for professional (or quasi-professional) measures of self-protection: “antiviral mask” or “hand disinfection” (instead of simple “washing hands”). We were able to identify at least several instances of misinformation, cognitive errors or fake news, which significantly impact the interest on “Koronawirus” topic in Poland. They fall on a fertile ground, due to the script mechanism of simplifying reality, where simpler explanations are more easily absorbed by people (11). Especially since epidemiologists and public health professionals do not know enough about COVID-19 yet, and recommendations and opinions of authorities such as WHO or Ministry of Health are constantly being updated and often being contradictory to previous opinions (e.g. effectiveness of protective masks) causing a cognitive dissonance in the public. As a good example we could take the popularity of a product, which does not exist on the medical market such as an “antiviral mask”. We show that scientifically confirmed and effective infection control strategies such as keeping high hygienic standards are not popular among Internet users and pseudo-professional solutions were much more attractive until official recommendations were introduced.
- formacyjne (Wikipedia i Google) nie odnotowały niektórych plików zainteresowania (odnoszących się np. ustaw specjalnych dotyczących walki z COVID-19), najprawdopodobniej z powodu nasycenia populacji wiedzą. Po fazie 1 i 2, ilustrującej silne potrzeby definicji sytuacji, zainteresowanie przekształciło się w proces bieżącego uzupełniania wiedzy np. na Twitterze czy YouTube. Owe kolejne fazy zainteresowania (fazy 4 i 5) można nazwać fazą tworzenia norm społecznych, gdzie ważnym elementem było poszukiwanie potwierdzenia sensowności lub konieczności używania środków ochrony osobistej i przestrzegania zachowań mitygacyjnych. W tej fazie znajduje się również pik zainteresowania tarczą antykryzysową, który zaobserwowano zarówno w mediach informacyjnych (Google, wiadomości i artykuły), jak i politycznie zagnieżdżonym Twitterze, a w mniejszym stopniu w mediach społecznościowych.
3. Największym zainteresowaniem użytkowników polskiego Internetu cieszyły się - na wszystkich analizowanych platformach - strategie i środki mitygacyjne. Co ważne, piki zainteresowania były widoczne w momencie ogłoszenia deklaracji ich wprowadzenia, a nie w terminie ich wejścia w życie (18). To oznacza, że sieci społeczne natychmiast przefiltrowały deklaracje rządowe, co wskazuje na wysoki stopień niepewności społecznej. Wywołuje to pytania o realną rolę mediów w zwalczaniu epidemii i czy wprowadzeniu obowiązku zachowań mitygacyjnych towarzyszyła odpowiednia kampania informacyjna, czy też potrzeby informacyjne ludności pozostawały w wysokim stopniu niezaspokojone.
 4. Koncentracja na analizie aktywności społecznej w mediach ujawniła kierunek społecznego myślenia o środkach zabezpieczających przed zakażeniem. Zidentyfikowany w naszym badaniu laicki system odniesień i przekonań (23) powinien być przez profesjonalistów zdrowia publicznego rozpatrywany w kontekście ryzyka zakażenia. Nasze badania pokazały, że Polacy chętnie szukali zabezpieczeń profesjonalnych lub quasi-profesjonalnych: „maseczka antywirusowa” czy „dezynfekcja rąk” (zamiast prostszego „mycia rąk”). W trakcie zbierania danych zidentyfikowaliśmy również przynajmniej kilka przypadków dezinformacji, błędów poznawczych i fake newsów, które znacząco wpłynęły na poziom zainteresowania koronawirusem w wyszukiwarkach internetowych. Padły one na żyzny grunt poznawczego mechanizmu upraszczania rzeczywistości i szybko nabrały charakteru wiralnego (11). Ta sytuacja jest zrozumiała, gdy epidemiolodzy i specjaliści zdrowia publicznego nie posiadają pełnej wiedzy na

5. Low interest of the general public in COVID-19 before the first reported case could be associated with a low epidemiological awareness of average Poles, especially when COVID-19 pandemic is massively discussed in traditional and social media by a relatively small but influential group of people (12). It's worth to mention, that a positive effect of the interest in "Koronawirus" could be a probable increase in epidemiological and hygienic knowledge in Polish population which is probably (11) below European average (even health literacy for non-infectious diseases does not differ significantly in comparison to other European countries (24)).
6. Given the scope of public interest especially during the later phases and the most popular platforms accessed it seems that the Ministry of Health's Public Relation Team should have concentrated more on traditional media and inform public opinion quicker instead of using Twitter. This is also supported by other research confirming the mechanism of stability seeking in the general population during the current outbreak in Poland (25). Stable and traditional communication strategy of the Minister of Health could improve the risk perception management (26).

Our study is subject to some methodological limitations. The population structure of the Internet users may lead to under/over-estimation of observed phenomena. We cover ca. 28 million Polish language speaking residents of Poland representing over 85% of the literate population (15). The Polish Internet users are younger (especially among content creators (27)) with females prevailing (especially among content consumers (14)) as compared to the average population. Our results do not describe the interest of the oldest groups, due to their digital exclusion (or preference to non-Internet media), health literacy limitations, etc. (24). Although the collected data can be successfully treated as a "digital proof", the research on this phenomena were conducted *in statu nascendi*, therefore the conclusions of our analysis should be verified by in-depth triangulational studies (28).

It is important to mention that the collected data can be biased e.g. due to content presenting algorithms on the media platforms. For instance, technological giants Google and Twitter are supposed to implement fact verification algorithms to filter out false information and promote official information (e.g. showing widgets related to COVID-19) and we can rely only on their search engines after filtration. This may jeopardize the reproducibility of our results. In our study each considered Internet platform is

temat COVID-19, a wytyczne instytucji takich jak WHO czy Ministerstwo Zdrowia są stale uzupełniane i często sprzeczne z poprzednimi opiniami i rekomendacjami, wywołując dysonans poznawczy w społeczeństwie (tak jak w przypadku oceny skuteczności noszenia maseczek w wydaniu ministra zdrowia Łukasza Szumowskiego). Dobrym przykładem laickiego systemu odniesień w obliczu zagrożenia zakażeniem jest popularność wyszukiwań „maseczki antywirusowej” jako produktu zabezpieczającego przed zakażeniem (nie istnieje taki termin medyczny). Pokazuje on, że w społeczeństwie naukowo potwierdzone i skuteczne strategie kontroli epidemii, takie jak standardy higieniczne, nie są popularne wśród użytkowników Internetu i często rozwiązania pseudo-profesjonalne były bardziej atrakcyjne, przynajmniej do czasu wprowadzenia oficjalnych rekomendacji i kampanii informacyjnych.

5. Niewielkie zainteresowanie COVID-19 przed pojawieniem się pierwszego potwierdzonego przypadku zakażenia może być związane z niską świadomością epidemiologiczną Polaków, przy czym należy rozróżnić nieliczną, ale głośną grupę aktywistów od reszty populacji (12). Pozytywnym efektem zainteresowania „koronawirusem” może być zwiększenie wiedzy epidemiologicznej i higienicznej, które w polskim społeczeństwie wydają się być na niskim poziomie w porównaniu ze społeczeństwami europejskimi (11).
6. Wydaje się, iż w interesie publicznym Minister Zdrowia i jego współpracownicy powinni częściej i bardziej efektywnie korzystać z tradycyjnych mediów informacyjnych, a nie z Twittera. Tym bardziej, że potwierdzono już w Polsce (25) mechanizm poszukiwania stabilizacji społecznej za pośrednictwem reprezentantów władzy. Wykorzystanie tradycyjnych i osadzonych w kulturze mediów informacyjnych pozwoliłoby na poprawę zarządzania ryzykiem epidemicznym.

Nasze analizy zawierają pewne ograniczenia metodologiczne. Ujęta w badaniu populacja użytkowników Internetu może prowadzić do prze- lub niedoszacowania obserwowanych zjawisk. W pracy objęliśmy ok. 28 mln mieszkańców Polski posługujących się językiem polskim, co reprezentuje ponad 85% populacji (15). Na tle populacji generalnej użytkownicy polskiego Internetu są nieco młodszy (szczególnie wśród tych, którzy tworzą treści (27)), z przewagą kobiet (szczególnie wśród odbiorców treści medycznych (14)). W naszych badaniach nie znalazły odzwierciedlenia zachowania starszych grup wiekowych, z uwagi na ich sieciowe wykluczenie wynikające bądź z braku dostępu, braku kompetencji cyfrowych, czy z ich preferencji informa-

described separately and has its own target audience and communication style. The most representative data comes from Google Trends due to the highest share of Polish Internet users and EventRegistry article selection due the highest reach in the general population. Despite some disadvantages and exploratory nature of this study, it provides us with an opportunity to analyse a huge amount of digital footprint data at low cost and in a short time.

In conclusion, Internet media analysis could fill the gaps in socio-medical research on collective actions during such important public health emergencies as epidemics of infectious diseases (8, 13). Moreover, as the lockdown was lifted, the second wave in the Fall 2020 becomes probable, with a wider geographical spread (29), which is expected to be reflected in the Internet media (possibly in syndromic terms). Thus the Internet analysis could complement epidemiological modelling (30) to inform implementation of non-pharmaceutical interventions – a major control tool until preferably via vaccination the herd immunity of the population is achieved if at all possible.

ACKNOWLEDGMENTS:

We thank PNFN (2019-21), NCN (2016/22/E/HS2/00034), FU Berlin (FU AvH: 08166500) and the European Commission (EOSC-48) for partial financial support. We thank Łukasz Krzowski, Mariusz Duplaga, Daniel Płatek, Ireneusz Skawina, Andrzej Buda, and Marcus Doherr for fruitful discussions. Code and data are available at <https://github.com/ajarynowski/koronawirus>.

REFERENCES

1. IBRIS. Koronawirus ponad polityka; 2020. Accessed: 2020-03-04. <https://wiadomosci.onet.pl/tylko-w-onecie/wybory-prezydenckie-2020-sondaz-andrzej-duda-prowadzi-potrzebna-iitura/cmk8ssw>.
 2. IPSOS. Sondaż Ipsos o koronawirusie. Dwie trzecie widzi duże zagrożenie, ale nie panikujemy; 2020. Accessed: 2020-03-12. <https://oko.press/sondaz-ipsos-o-koronawirusie/>.
 3. Monitor. Koronawirus odbiera pracę i pustoszy portfele Polaków, pandemia COVID-19 budzi większy lęk o finanse niż o zdrowie; 2020. Accessed: 2020-04-13. <https://alebank.pl/koronawirus-odbiera-prace-i-pustoszy-portfele-polakow-pandemia-covid-19-budzi-wiekszy-lek-o-finance-niz-o-zdrowie/>.
 4. Gałuszka M, et al. Medykalizacja w kulturze strachu: przykład medialnej ekspozycji grypy A/H1N1. *PrzełSocjol* 2017;66(1):53–81.
- cyjnych (24). Mimo iż prezentowanie badania mogą być z powodzeniem traktowane jako „śląd cyfrowy” epidemii, badaliśmy zjawisko *in statu nascendi*, dlatego wyniki powinny być w przyszłości zweryfikowane w pogłębionych badaniach triangulacyjnych (28).
- Warto również zwrócić uwagę, że zbierane dane mogą ulegać zniekształceniom w wyniku działania algorytmów stosowanych przez administratorów poszczególnych platform. Na przykład Google i Twitter stosują algorytmy weryfikacji w celu odfiltrowania fałszywych treści i promocji informacji oficjalnych (np. widgetów informacyjnych o COVID-19) i badacze mogą korzystać tylko z oczyszczonych już danych. Przynosi to ryzyko, iż przy każdym kolejnym badaniu wyniki mogą być nieco różne, gdyż algorytmy na bieżąco moderują dostępne zbiory danych i ich skład może się przy każdym eksporcie różnić. Każda z wykorzystanych w badaniach platform została opisana oddzielnie, ze wskazaniem typu odbiorców i stylu komunikacyjnego. Najbardziej wartościowe dane pochodzą z Google Trends z uwagi na odzwierciedlenie zachowań znakomitej większości użytkowników Internetu w Polsce oraz EventRegistry, który obejmuje praktycznie całą populację internautów. Pomimo niedoskonałości i eksploracyjnego charakteru niniejszych badań, stanowią one doskonałą okazję do ukazania „ślada cyfrowego” epidemii w stosunkowo krótkim czasie i przy minimalnych kosztach.
- Podsumowując, analizy mediów internetowych mogą z powodzeniem wypełnić lukę w socjomedycznych badaniach zachowań zbiorowych podejmowanych w czasie rozwoju i trwania zagrożeń dla zdrowia publicznego, takich jak epidemie chorób zakaźnych. Ponadto, po zniesieniu lockdownu i przy przewidywanej na jesieni 2020 roku drugiej fali epidemii, prawdopodobnie o bardzo szerokim zasięgu geograficznym (29), będziemy mieli do czynienia z drugą falą zainteresowania, mającą odzwierciedlenie w poszukiwaniu informacji. Dlatego analizy zachowań internautów mogą uzupełniać modelowanie epidemiologiczne (30), w celu usprawnienia nefarmakologicznych interwencji – jako podstawowego narzędzia kontroli – do czasu osiągnięcia drogą szczepień zbiorowej odporności, o ile to w ogóle możliwe.
- Podziękowania: Autorzy dziękują PNFN (2019-21), NCN (2016/22/E/HS2/00034), FU Berlin (FU AvH: 08166500) oraz Komisji Europejskiej (EOSC-48) za częściowe sfinansowanie niniejszych badań. Dziękujemy również Łukaszowi Krzowskiemu, Mariuszowi Dupladze, Danielowi Płatkowi, Ireneuszowi Skawinie, Andrzejowi Budzie oraz Marcusowi Doherr za owocne dyskusje w tematyce artykułu. Kody i dane dostępne są na <https://github.com/ajarynowski/koronawirus>.

5. Gonsalves G, Staley P. Panic, paranoia, and public health—the AIDS epidemic’s lessons for Ebola. *New England J Medicine* 2014;371(25):2348–2349.
6. Wójcik S, Duplaga M, Grysztar M, Pałka P. Web browser as a tool for predicting the incidence of influenza. *Piel Zdr Publ* 2018;8(2):83–88.
7. Effenberger M, Kronbichler A, Shin JI, Mayer G, Tilg H, Perco P. Association of the COVID-19 pandemic with Internet Search Volumes: A Google Trends™ Analysis. *International J Infec Dis*. 2020 Apr 17.
8. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. *J Med Internet Res* 2009;11(1):e11.
9. Samaras L, García-Barriocanal E, Sicilia MA. Syndromic surveillance using web data: a systematic review. In: *Innovation in Health Informatics* 2020. p. 39–77.
10. 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(10).
11. Jarynowski A, Wójta-Kempa M, Belik V. Percepcja „koronawirusa” w polskim Internecie do czasu potwierdzenia pierwszego przypadku zakażenia SARS-CoV-2 w Polsce. *Piel Zdr Publ* 2020;10(2).
12. Jarynowski A, Wójta-Kempa M, Płatek D, Czopek K. Attempt to understand public health relevant social dimensions of COVID-19 outbreak in Poland. *Society Register* 2020, 4(3), 7-44.
13. Latif S, Usman M, Manzoor S, Iqbal W, Qadir J, Tyson G, Castro I, Razi A, Boulos MN, Crowcroft J. Leveraging Data Science To Combat COVID-19: A Comprehensive Review. Preprint 2020.
14. Jarynowski A, Belik V. Choroby przenoszone drogą płciową w dobie Internetu i e-zdrowia: kalkulatory ryzyka. Kraków: Biblioteka Jagiellońska 2018, p 101–111.
15. PBI. Polski internet w styczniu 2020; 2020. Accessed: 2020-03-04. <http://pbi.org.pl/badanie-gemius-pbi/polski-internet-w-styczniu-2020/>.
16. Sotrender. Twitter w Polsce – podsumowanie 2018 roku; 2019. Accessed: 2020-03-04. <https://www.sotrender.com/blog/pl/2019/02/twitter-w-polsce-podsumowanie-2018-roku-infografika/>.
17. PBI. Wikipedia i jej użytkownicy; 2018. Accessed: 2020-03-12. http://pbi.org.pl/wp-content/uploads/2017/09/2017-09-26-Wikipedia_analiza.pdf.
18. Pinkas J, Jankowski M, Łukasz Szumowski, Lusawa A, Zgliczyński WS, Raciborski F, et al. Public Health Interventions to Mitigate Early Spread of SARS-CoV-2 in Poland. *Medical Science Monitor* 2020;26.
19. Strzelecki A. The second worldwide wave of interest in coronavirus since the COVID-19 outbreaks in South Korea, Italy and Iran: A Google Trends study. *Brain Behav Immun*. 2020;S0889-1591(20).
20. Wyboreza. Chybicka: W Łodzi potwierdzono koronawirusa”; 2020. Accessed: 2020-03-12. <https://wiadomosci.gazeta.pl/wiadomosci/7,114884,25741826,chybicka-w-lodzi-potwierdzono-koronawirusa-minister-zdrowia.html>.
21. Jia JS, Lu X, Yuan Y, Xu G, Jia J, Christakis NA. Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature*. 2020; 29:1-5.
22. Schultz F, Utz S, Göritz A. Is the medium the message? Perceptions of and reactions to crisis communication via twitter, blogs and traditional media. *Public relations review* 2011;37(1):20–27.
23. Taranowicz I. Zdrowie i sposoby radzenia sobie z jego zagrożeniami. Analiza socjologiczna Wrocław: Oficyna Wydawnicza Arboretum. 2010.
24. Duplaga M. Determinants and Consequences of Limited Health Literacy in Polish Society. *International Journal of Environmental Research and Public Health*. 2020;17(2):642.
25. Karwowski M, Kowal M, Groyecka A, Białek M, Lebuda I, Sorokowska A, et al. When in Danger, Turn Right: Covid-19 Threat Promotes Social Conservatism and Right-wing Presidential Candidates. Preprint 2020;.
26. Duplaga M. Perception of the Effectiveness of Health-Related Campaigns among the Adult Population: An Analysis of Determinants. *International journal of environmental research and public health* 2019;16(5):7 Physical Review X 2011;1(1): 011001.
27. Smahel D, MacHackova H, Mascheroni G, Dedkova L, Staksrud E, Olafsson K, Livingstone S, Hasebrink U. EU Kids Online 2020: survey results from 19 countries.
28. Jarynowski A, Paradowski MB, Buda A. Modelling communities and populations: an introduction to computational social science. *Studia metodologiczne* 2019;39:117–139.
29. Jarynowski A, Wójta-Kempa M, Płatek D, Krzowski Ł, Belik V. Spatial Diversity of COVID-19 Cases in Poland Explained by Mobility Patterns - Preliminary Results 2020; <http://dx.doi.org/10.2139/ssrn.3621152>
30. Belik V., Geisel T., Brockmann D. Natural human mobility patterns and spatial spread of infectious diseases.

Received: 3.05.2020

Accepted for publication: 21.07.2020

Otrzymano: 3.05.2020 r.

Zaakceptowano do publikacji: 21.07.2020 r.

Address for correspondence:

Institute for Veterinary Epidemiology and

Biostatistics (FU-Berlin)

Königsweg 67

14163 Berlin (Germany)

Adres do korespondencji:

Instytut Biometrii i Epidemiologii Weterynaryjnej

(FU-Berlin)

Königsweg 67

14163 Berlin (Niemcy)