Googling on Colonoscopy: A Retrospective Analysis of Search Engine Statistics

This article was published in the following Dove Press journal:
Clinical and Experimental Gastroenterology

Purpose: Colonoscopy is a gold standard for screening and diagnosis of colorectal cancer (CRC). The data from the search engine may reveal what information on colonoscopy gains the attention of Internet users. We aimed to investigate Google searches trends and terms related to colonoscopy.

Patients and Methods: We retrieved statistics searches related to colonoscopy using Google Trends (GT) and Google Ads (GA) for the period from April 2016 to March 2020. The GT data was used for the analysis of time and regional search patterns worldwide. GA data for Australia, Canada, Ireland, New Zealand (NZ), Poland, the United Kingdom (UK), and the United States (US) were used to calculate the search volume of categories of queries related to colonoscopy.

Results: Globally, the relative search volume on colonoscopy has increased until the COVID-19 outbreak and revealed seasonal variation: the highest interest was observed in March (CRC awareness month), and the lowest during December (Christmas holidays). The highest number of searches per 1000 Google users-years was done in Poland (59.62) and the lowest in the UK (19.46). Most commonly, Google users searched for details on colonoscopy techniques (Australia, Canada, Ireland, NZ), anesthesia during the procedure (Poland), facility performing colonoscopy (UK, US). In all seven countries, less than 2% of queries concerned with bowel preparation before the procedure.

Conclusion: Before the COVID-19 pandemic, the interest in colonoscopy has increased among Google users. Google users may underestimate the importance of proper bowel preparation.

Keywords: colonoscopy, internet, google, trends, infodemiology

Introduction

Colorectal cancer (CRC) is the third most deadly and fourth commonly diagnosed cancer in the world.1 With two million new cases and almost 1 million deaths in 2018, according to the World Health Organization GLOBOCAN database.2 Of importance, the global number of new cases of, and deaths from gastrointestinal (GI) cancers are predicted to increase by 58% and 73%, respectively, by 2040.3 The implementation of CRC population-based screening contributes to decreasing incidence and mortality rates;4 however, increases in incidence rates have recently been observed in adults of younger age.3

Colonoscopy is instrumental to both diagnosis and CRC screening. For the latter, the participants’ interest and behavior of the screenees are of the most important steps for successful outcomes.5 However, screening programs have already been or will be impeded in the near future by several psychosocial, and
healthcare-related factors, including between others: (i) risk of disease transmission, eg, coronavirus disease 19 (COVID-19); (ii) inadequate number of skilled healthcare professionals; (iii) long learning curves to achieve full professional experience in endoscopic techniques; (iv) increasing numbers of physicians/nurses with burnout syndromes; (v) increasing number of “need-to-screen” individuals worldwide; and more importantly, (vi) patients’ perspective and concerns on colonoscopy-based CRC screening (eg, unpleasant preparation, pain, cost) and/or potential to choose other non-invasive options.

Several investigators have attempted through surveys and questionnaires to document attitudes towards colonoscopy as well as possible concerns about the nature, adverse events, and receptiveness of the procedure. The invasive nature of the procedure, the effects of the laxative bowel preparation, and the fear of discomfort/pain or embarrassment seem to stand out among others. However, these surveys captured only sample opinions from limited groups of individuals and were cross-sectional. Moreover, the survey studies are limited by the response rates, and that results may quickly become outdated. Here, we present an alternative approach based on a massive number of search engine queries on colonoscopy.

Internet plays a major role in patient self-education on health issues. The Web’s main advantages are: immediate access to interesting content, numerous websites on health, and the opportunity to interact with other users. The activity of the Internet users related to health issues became of useful source for research. Search engine statistics provide a unique insight in the interest of the users. Globally, Google is the most popular search engine used by over 90% of Internet users. To date, two studies utilized Google data to investigate the interest of users in colonoscopy. Both focused on Google users living in the United States and associations between Colon Cancer Awareness Month (March), and peaks on search volumes. Therefore, an essential gap exists in the global interest of Google users in colonoscopy. To date, no study presented the global trends, regional interest in colonoscopy as well as the content of the searches. We hypothesized that such analysis might reveal whether Google users are using more commonly the search engine to seek for information on colonoscopy. Moreover, investigation of words using in the queries may reveal what aspects of colonoscopy are the main concerns of Google users. This is very important because colonoscopy is an intimate and invasive procedure related to an unpleasant experience. We assume that Google queries may involve problems that are rarely disclose to physicians.

Here, we go further and analyze the searches related to colonoscopy in the broader context. Therefore, we aimed to investigate: a) worldwide trends of interest of Google users in colonoscopy b) regional differences in search volume on colonoscopy, and c) the main concerns of Google users related to colonoscopy.

Materials and Methods
Data from Google Trends
Data over a 4-year period (April 2016-March 2020) were collected from Google Trends (GT) and Google Ads (GA), on 29th April 2020. The data collection and followed processing protocols used are described in previous studies.

GT presents search statistics for a given search term in the chosen timeframe and region (trends.google.com/trends/). The search volume is presented as an index called relative search volume (RSV) ranged between 0–100. RSV equals 100 represents the peak on interest, while 0 complete lack of interest in the analyzed search term in the given timeframe and region. RSV is adjusted to the number of Google users in a given timeframe, and region. Therefore, RSV measures the intensity of queries rather than crude search volume. GT recognizes input as “search terms” or “topic”. When typing search input in the GT engine, the tool may suggest a topic for the analysis. The topic is a universal method to compare queries in all available languages, while search terms represent typed input.

For instance, the search term “gastroscopy” will generate the highest interest in English-speaking countries. Matching “gastroscopy” with the topic “Gastroscopy” allows us to compare queries in non-English in statistics. GT ignores duplicated queries made from the same IP address in a short period. We attached a screenshot of GT and described the settings in Figure S1.

We set worldwide as a region of analyzed queries. We used the GT option to exclude countries with low search volume. The region with low search volume is susceptible to irregular variation. The search term “colonoscopy” with the topic “Colonoscopy” were thereafter matched. Time trends of RSV in the last four years and interest by region were downloaded. All calculations were performed using
the R-programming language version 3.6.3 (R Foundation, Vienna, Austria) in May 2020. Time trends were visualized using the ggplot2 R package.27 The trend curve was fitted by using the local polynomial regression model provided by function geom_smooth(). Changes of RSV per year were calculated before the COVID-19 pandemic (April 2016–February 2020).

We checked seasonality of the period preceding the pandemic by fitting the time trend, an exponential smoothing state-space model with Box-Cox transformation, autoregressive-moving average errors, trend, and seasonal components (TBATS).28 We searched for months when the RSV is the highest and the lowest by decomposing the time trend before COVID-19 pandemic. The data called interest by region ranks countries based on the relative frequencies of queries related to analyzed input. Here, RSV equals 100 represents the country where Google users generate relatively the most queries (adjusted to the population of Google users). We investigated the association between RSV on colonoscopy in each country and mean disability-adjusted life year’s (DALY) rate of inflammatory bowel diseases (IBD) and colorectal cancer from the Global Burden of Disease study from the years 2016–2017.29 We performed the R Spearman rank correlation test between RSV on colonoscopy, and DALY’s of IBD, and colorectal cancer in each country. Moreover, we visualized countries with population-based or structured opportunities colorectal cancer screenings using colonoscopy or non-colonoscopy as primary techniques.30

The study process GT and GA data do not include human subjects. Therefore, the design of the study did not require Ethical Committee approval. The study does not violate the terms of the use of GT and GA.

Data from Google AdWords

Because Google Trends do not present detailed information on keywords used in the query, we expanded the analysis by data coming from GA. However, due to the limited language skills in our team to English and Polish, we included six English-speaking countries and Poland.

GA was designed for selecting keywords targeting Internet users for e-commerce campaigns. The GA Keyword Planner (https://ads.google.com/aw/keywordplan er) generates a list of keywords related to the input. GA enables to collect data for the last 4 years. After typing in “colonoscopy”, set language: “English”, a list of keywords was collected for six English-speaking countries: Australia, Canada, Ireland, New Zealand, the United Kingdom, and the United States. Furthermore, we type colonoscopy in Polish (pl. “kolonoskopia”), set the language “Polish”, and collected a list of keywords for Poland. In all cases, we set search networks as “Google and search partners” and used all proposed keywords. A screenshot of the search engine settings is presented in Figure S2 and the data processing protocol in Figure S3. All keywords generated are presented in the dataset attached to the supplementary material.

We collected data for the last 4 years from April 2016 to March 2020. Data generated from GA include keywords and the estimated value of search volume per data point (a month or longer period). The search volume is expressed as a range between two exponents of ten (eg, 100–1000). M.K. analyzed the list of keywords and W.M., and A.K. verified the analysis.

We distinguished the following categories of the keywords: patient features (eg, “female colonoscopy”), indications (eg, “colonoscopy for anaemia”), preparation (eg, “citrafleet colonoscopy”), anesthesia (all keywords related to sedation, anaesthesia or diminishing pain during the procedure, eg, “colonoscopy sedation”), colonoscopy findings (eg, “descending colon polyp”), complications (eg, “loose stools after colonoscopy”), screening program (eg, “cdc colon cancer screening”), costs (eg, “colonoscopy cost 2018”), searching facility (eg, “colonoscopy clinic near me”), different colonoscopy techniques (eg, “chromo colonoscopy”, “colonoscopy biopsy”), non-invasive alternative (eg, “fit testing colon cancer”), and virtual colonoscopy (subcategory of non-invasive alternative, eg, “CT colonoscopy”). The keywords could have none or more than one category (eg, “colonoscopy London cost”, categories: searching facility and cost). Keywords that were unrelated to colonoscopy (eg, “gastroscopy”) were excluded.

Thereafter, the mean search volume for all keywords in the analyzed period was calculated. Further, we sum the total number of searches in each category. All search volumes were expressed as a number of queries per 1000 Google-user years. The number of Google users in each analyzed country was displayed in GA. Finally, all categories as a percentage of the total number of searches in each country were calculated. We compared search volumes expressed as the number of queries per 1000 Google-user years between each country, and for each keywords categories by using the Kruskal–Wallis test, with a post-hoc pair-wise Mann–Whitney U-test with
Bonferroni correction. P values of <0.05 were considered to indicate a significant difference.

Results
Globally, the RSV on colonoscopy increased by 7.54 RSV/year to March 2020 when it dropped by ~70%, Figure 1. Moreover, the time trend revealed a 12-month seasonal pattern: the interest was higher during March and the lowest during December. The burden of IBD, and CRC were positively associated with the interest in colonoscopy in analyzed countries (respectively: Rs = 0.65; Rs = 0.57; all p < 0.001), Figure 2.

We included 98–99% of the keywords generated by the Keyword Planner (Figure S3): 308 from 313 for Australia, 308 from 311 for Canada, 319 from 321 for Ireland, 320 from 325 for New Zealand, 437 from 440 for Poland, 305 from 307 for the United Kingdom, and 314 from 316 for the United States. The highest number of searches related to colonoscopy per 1000 Google users-years was observed in Poland (59.52), and the lowest in the United Kingdom (19.46), Table 1. The detailed results of the post-hoc analyses are presented in Table S1. The rate of searches related to analyzed categories differed between most of the countries. Users from Poland and the United Kingdom more frequently searched for information related to facilities performing a colonoscopy, anesthesia, and costs (Poland). In contrast, users from Ireland and New Zealand searched for information on patient features, indications, findings, and colonoscopy techniques more commonly. Interestingly, in all countries, less than 2% of queries were related to preparations before the procedure. In the United States, the search rate on colonoscopy preparation was equal to 0.29%, which was the lowest rate among all keywords categories among all analyzed countries.

Discussion
We analyzed the timeline and geographical pattern of the topic “Colonoscopy” using GT as well as the main concerns related to the procedure of Google users living in seven Western world countries.

The global interest in colonoscopy has recorded an increase in the last few years before the COVID-19
Figure 2 Data from Google Trends: interest by region. A correlation plot of interest in “Colonoscopy” among Google users and: (A) burden of Inflammatory Bowel Diseases (IBD). (B) burden of Colorectal Cancer (CRC).

Abbreviations: DALYs, disability-adjusted life year; RSV, relative search volume.
Table 1: Comparison Between the Analyzed Countries Regarding Colonoscopy-Related Searches from April 2016 to March 2020

<table>
<thead>
<tr>
<th>Country</th>
<th>Australia</th>
<th>Canada</th>
<th>Ireland</th>
<th>NZ</th>
<th>Poland</th>
<th>UK</th>
<th>USA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google users (mln)</td>
<td>18.0</td>
<td>31.0</td>
<td>4.5</td>
<td>3.9</td>
<td>657,320</td>
<td>25.9</td>
<td>59.5</td>
</tr>
<tr>
<td>Total number</td>
<td>3,589,940 (49.86) (100.0%)</td>
<td>3,859,105 (31.12) (100.0%)</td>
<td>756,990 (42.05) (100.0%)</td>
<td>27,980 (1.81) (4.26%)</td>
<td>36,120 (23.4) (5.50%)</td>
<td>9,000 (0.64) (1.51%)</td>
<td>16,300 (1.06) (2.48%)</td>
</tr>
<tr>
<td>Patients features</td>
<td>58,430 (0.81) (1.63%)</td>
<td>63,030 (0.51) (1.63%)</td>
<td>36,780 (2.04) (4.86%)</td>
<td>38,620 (0.37) (0.63%)</td>
<td>43,780 (2.46) (4.86%)</td>
<td>32,830 (0.64) (1.29%)</td>
<td>20,760 (0.35) (0.63%)</td>
</tr>
<tr>
<td>Indications</td>
<td>76,300 (1.06) (2.13%)</td>
<td>71,940 (0.58) (1.86%)</td>
<td>48,370 (2.69) (6.39%)</td>
<td>118,675 (7.15) (1.92%)</td>
<td>133,200 (7.15) (1.92%)</td>
<td>84,035 (1.67) (2.69%)</td>
<td>82,095 (1.66) (2.69%)</td>
</tr>
<tr>
<td>Preparation</td>
<td>31,170 (0.43) (0.87%)</td>
<td>20,760 (0.17) (0.54%)</td>
<td>14,460 (0.80) (1.90%)</td>
<td>20,345 (1.20) (0.33%)</td>
<td>26,375 (1.45) (4.01%)</td>
<td>10,825 (0.20) (0.33%)</td>
<td>84,255 (1.67) (2.69%)</td>
</tr>
<tr>
<td>Anesthesia</td>
<td>34,165 (0.47) (0.95%)</td>
<td>19,170 (0.57) (1.64%)</td>
<td>22,770 (1.54) (3.66%)</td>
<td>111,640 (6.70) (12.55%)</td>
<td>133,200 (7.15) (1.92%)</td>
<td>100,300 (0.42) (0.87%)</td>
<td>107,700 (2.04) (3.66%)</td>
</tr>
<tr>
<td>Colonoscopy findings</td>
<td>55,185 (0.77) (1.54%)</td>
<td>60,580 (0.49) (1.57%)</td>
<td>36,720 (2.04) (4.85%)</td>
<td>20,345 (1.20) (0.33%)</td>
<td>26,375 (1.45) (4.01%)</td>
<td>10,825 (0.20) (0.33%)</td>
<td>84,255 (1.67) (2.69%)</td>
</tr>
<tr>
<td>Complications</td>
<td>66,285 (0.92) (1.85%)</td>
<td>93,280 (0.75) (2.42%)</td>
<td>37,130 (2.04) (4.01%)</td>
<td>26,375 (1.45) (4.01%)</td>
<td>26,375 (1.45) (4.01%)</td>
<td>10,825 (0.20) (0.33%)</td>
<td>84,255 (1.67) (2.69%)</td>
</tr>
<tr>
<td>Screening program</td>
<td>42,875 (0.60) (1.19%)</td>
<td>69,650 (0.70) (2.42%)</td>
<td>27,190 (1.51) (3.59%)</td>
<td>21,780 (1.14) (2.04%)</td>
<td>21,780 (1.14) (2.04%)</td>
<td>10,825 (0.20) (0.33%)</td>
<td>84,255 (1.67) (2.69%)</td>
</tr>
<tr>
<td>Costs</td>
<td>46,330 (0.64) (1.29%)</td>
<td>25,635 (0.21) (0.66%)</td>
<td>22,900 (1.27) (2.04%)</td>
<td>16,810 (1.09) (2.56%)</td>
<td>16,810 (1.09) (2.56%)</td>
<td>10,825 (0.20) (0.33%)</td>
<td>84,255 (1.67) (2.69%)</td>
</tr>
<tr>
<td>Searching facility</td>
<td>59,905 (0.83) (1.67%)</td>
<td>64,080 (0.51) (1.66%)</td>
<td>27,630 (1.53) (3.65%)</td>
<td>65,340 (6.31) (12.35%)</td>
<td>65,340 (6.31) (12.35%)</td>
<td>10,825 (0.20) (0.33%)</td>
<td>84,255 (1.67) (2.69%)</td>
</tr>
<tr>
<td>Different colonoscopy techniques</td>
<td>97,435 (1.35) (2.71%)</td>
<td>138,875 (1.12) (3.60%)</td>
<td>62,680 (3.48) (8.28%)</td>
<td>132,845 (1.28) (2.51%)</td>
<td>132,845 (1.28) (2.51%)</td>
<td>10,825 (0.20) (0.33%)</td>
<td>84,255 (1.67) (2.69%)</td>
</tr>
<tr>
<td>Non-invasive alternative</td>
<td>19,250 (0.27) (0.54%)</td>
<td>59,360 (0.48) (1.54%)</td>
<td>18,740 (1.04) (2.48%)</td>
<td>17,710 (1.04) (2.48%)</td>
<td>17,710 (1.04) (2.48%)</td>
<td>10,825 (0.20) (0.33%)</td>
<td>84,255 (1.67) (2.69%)</td>
</tr>
<tr>
<td>Virtual colonoscopy</td>
<td>40,440 (0.56) (1.31%)</td>
<td>44,520 (0.36) (1.15%)</td>
<td>14,700 (0.82) (1.94%)</td>
<td>15,150 (0.98) (2.04%)</td>
<td>15,150 (0.98) (2.04%)</td>
<td>10,825 (0.20) (0.33%)</td>
<td>84,255 (1.67) (2.69%)</td>
</tr>
</tbody>
</table>

Notes: Data are presented as the total number of queries (number of queries: 1000 Google-user years) (percentage of total number).

Abbreviations: CRC, colorectal cancer; COVID-19, coronavirus disease 19; DALY, disability-adjusted life years; GA, Google Ads; GI, gastrointestinal; GT, Google Trends; IBD, inflammatory bowel diseases; RSV, relative search volume; NZ, New Zealand; UK, the United Kingdom; US, the United States.

outbreak. However, it should be noted that RSV is adjusted to the number of Google searches per country. Thus, the increase cannot be explained solely by broader access to the Internet. We hypothesize that the increase in searches for RSV is driven by RSV outbreaks, particularly in the United States, where the increase is most pronounced. Interestingly, the search volume for RSV is highest in the United States, where RSV outbreaks are most frequent.
procedure might be underestimated by the public no matter how critical it is to the quality of colonoscopy. On the one hand, before colonoscopy, users might have received detailed information on preparation; thus, they do not need to use Internet sources to educate. On the other hand, one can assume that previous reports on colonoscopy-related perception and fears have led to action and that the quality and detail of send-out information has improved,\textsuperscript{15,35} thus, patient do not need to use Internet sources for anything further on this subject. Nevertheless, previous studies suggest that in 15–35% of colonoscopies bowels are inadequately prepared,\textsuperscript{36–39} and 7.9% of patients perceive unpleasantness of preparation as a barrier to colonoscopy screening.\textsuperscript{13} Therefore, there is still a lot to improve in aspects of bowel cleansing before the procedure.

To date, no study utilized freely available Google data to investigate global interest in colonoscopy. The methods we used are cost-free and have the potential to provide a background for future studies. Here, we showed that the global searches on colonoscopy have a reasonable timeline pattern. Furthermore, the interest in colonoscopy seems to be higher in countries with a higher burden of IBD and CRC. We used Google Ads, the tool for e-marketing to investigate statistics representing over 55 million of Google queries done by users from seven countries. We hypothesize that patients are not fully aware of how proper preparation improves the quality of the procedure and their satisfaction. Therefore, the Internet sources on colonoscopy should always mention the pivotal role of the preparation before colonoscopy; up to date, most Google users focus on different aspects and may omit essential information. From the physicians’ perspective, the home message is that the interest in colonoscopy among Google users increases. Therefore, health education via online sources might be the most efficient. The approach we presented may help health policymakers assess how Internet users responded to an information campaign: whether an interest increase or not. Moreover, the queries analysis might indicate the users’ common concerns (eg, in Poland cost and anesthesia). The information from searches might be translated to improve health services. We believe that our study will be inspiring for the next investigations on patients’ activity on Internet. Future studies should focus on how the information from the Internet affects bowel preparation before the colonoscopy. Could the online sources motivate patients for more scrupulous implementations of the bowel cleaning procedure? How

knowledge from Internet affect the decision to take part in the screening program? Finally, there is a gap of knowledge on the quality of the source for patients about colonoscopy.

The study has several limitations. Firstly, Google is the dominant search engine across the world, but its market share differs. For instance, in most European countries, Google is used by over 90% of Internet users, but in the United States by 80–85\textsuperscript{.22} Regrettably, other search engines do not openly present statistics as Google does. Secondly, we were able to analyze keywords written in two languages; thus, the keywords were limited to seven countries. Thirdly, the list of keywords generated by GA differ between countries, and the search volume is expressed as a range, not the exact number. Therefore, the results should be treated with caution. Lastly, GA does not include all queries related to the given keywords; only these the most common. Therefore, there is a long tail in the distribution of keywords, which includes rare and complicated phrases that are not available for the analysis. Finally, Google does not disclose statistics on Google users such as age and gender in keeping with personally identifiable data confidentiality. Therefore, we cannot conclude which population generates the most queries and whether the statistics represent a general population. Previous studies suggested that younger generations and women tend to use the Internet more as a source of medical knowledge.\textsuperscript{40,41} For this reason, we may assume that these groups are overrepresented among Google users typing colonoscopy-related queries.

**Conclusion**

Globally, the interest of Google users in colonoscopy increases until the COVID-19 pandemic when sharply decreased. The people from regions with a higher burden of IBD and CRC generated more queries on colonoscopy. Less than two percent of queries concerned preparation before the procedure. Google users may underestimate the importance of proper bowel preparation.

**Acknowledgment**

The data is presented in the supplementary material.

**Disclosure**

AK has received material support for research from SynMed, IntroMedic, GivenImaging Ltd and has been in advisory meetings for DrFalkPharmaUk, Tillots while he has received honoraria from DrFalkPharmaUK and reports
no other potential conflicts of interest for this work. The other co-authors report no conflicts of interest for this work.

References


